JACS Au
AN OPEN ACCESS JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/jacsau
Article

Machine Learning Optimizing Enzyme/ZIF Biocomposites for Enhanced Encapsulation Efficiency and Bioactivity

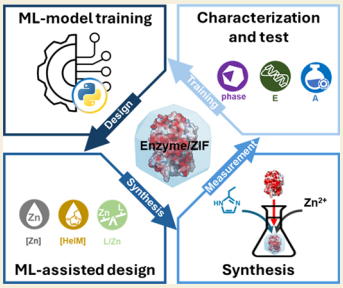
Weibin Liang,* Sisi Zheng, Ying Shu, and Jun Huang*

Cite This: JACS Au 2024, 4, 3170−3182

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: In this study, we present the first example of using a machine learning (ML)-assisted design strategy to optimize the synthesis formulation of enzyme/ZIFs (zeolitic imidazolate framework) for enhanced performance. Glucose oxidase (GOx) and horseradish peroxidase (HRP) were chosen as model enzymes, while $Zn(eIM)_2$ (eIM = 2-ethylimidazolate) was selected as the model ZIF to test our ML-assisted workflow paradigm. Through an iterative ML-driven training-design-synthesis-measurement workflow, we efficiently discovered GOx/ZIF (G151) and HRP/ZIF (H150) with their overall performance index (OPI) values (OPI represents the product of encapsulation efficiency (E in %), retained enzymatic activity (A in %), and thermal stability (T in %)) at least 1.3 times higher than those in systematic seed data studies. Furthermore, advanced statistical methods derived from the trained random forest model qualitatively and quantitatively reveal the relationship among synthesis, structure, and performance in the enzyme/ZIF system, offering valuable guidance for future studies on enzyme/ZIFs. Overall, our proposed ML-assisted design strategy holds promise for accelerating the development of enzyme/ZIFs and other enzyme immobilization systems for biocatalysis applications and beyond, including drug delivery and sensing, among others.



KEYWORDS: *enzyme/ZIF biocomposite, biocatalysis, machine learning, NMR spectroscopy, encapsulation efficiency, ATR-FTIR spectroscopy*

INTRODUCTION

In recent years, enzyme/MOF biocomposites (MOF = metal−organic framework), produced through biomimetic mineralization, have garnered considerable scientific attention due to their simple synthesis and promising performance.[1−6] These biocomposites exhibit exceptional encapsulation efficiency, retained enzymatic bioactivity, and protective effects against stressors.[1−5] However, experimental findings illustrate that achieving these desired properties is not guaranteed,[4,7−10] as structural factors such as crystallinity,[11] structural defects,[12] chemical texture,[9,10] and spatial localization of the enzyme[13,14] all influence the performance of a biocomposite. Importantly, all these structural factors are highly dependent on the synthesis conditions of the enzyme/MOF. These conditions involve the choice of organic ligand,[8−10,15] synthesis component concentrations,[13] ligand-to-metal molar ratio, etc.[7,14]

In ZIFs (ZIF = zeolitic imidazolate framework),[13,16−18] modulation of synthesis conditions can alter the formation of polymorphic phases. A high concentration of ZIF precursors and/or a high ligand-to-metal molar ratio in the enzyme/ZIF synthesis is reported to favor the porous kinetic phase formation (e.g., sod-$Zn(mIM)_2$[19,20] and rho-$Zn(eIM)_2$;[20] mIM = 2-methylimidazolate and eIM = 2-ethylimidazolate). These porous kinetic phases are theoretically more ideal in biocatalysis applications as they offer intrinsic porosity to facilitate mass diffusion. For example, GOx@sod-$Zn(mIM)_2$ (GOx = glucose oxidase) has been demonstrated to exhibit superior bioactivity compared to GOx@dia-$Zn(mIM)_2$.[14] However, exceptions exist; for instance, enzymes encapsulated in a crystallographically dense ZIF-L ($Zn(mIM)_2(HmIM)_{1/2}\cdot(H_2O)_{3/2}$)[21] are reported to exhibit bioactivity.[22−24] These conflicting results may arise from structural defects within ZIF-L. The non-native defect-induced meso- and/or macro-channels in ZIF-L enable mass diffusion.

Even when the enzyme/ZIFs share an identical crystallographic phase, distinct performances may be observed.[7,14] A recent report illustrates that GOx@sod-$Zn(mIM)_2$ synthesized under a low HmIM/Zn molar ratio (BZIF-8-B, $\frac{[HmIM]}{[Zn]} = 4$) exhibits significantly higher retained bioactivity compared to its counterpart synthesized using a high HmIM/Zn molar ratio (BZIF-8-S, $\frac{[HmIM]}{[Zn]} = 12$).[14] Transmission electron microscopy (TEM) measurements reveal that the enhanced bioactivity of BZIF-8-B is attributed to the abundant structural defects within the sample.

Received: June 6, 2024
Revised: July 31, 2024
Accepted: August 1, 2024
Published: August 12, 2024

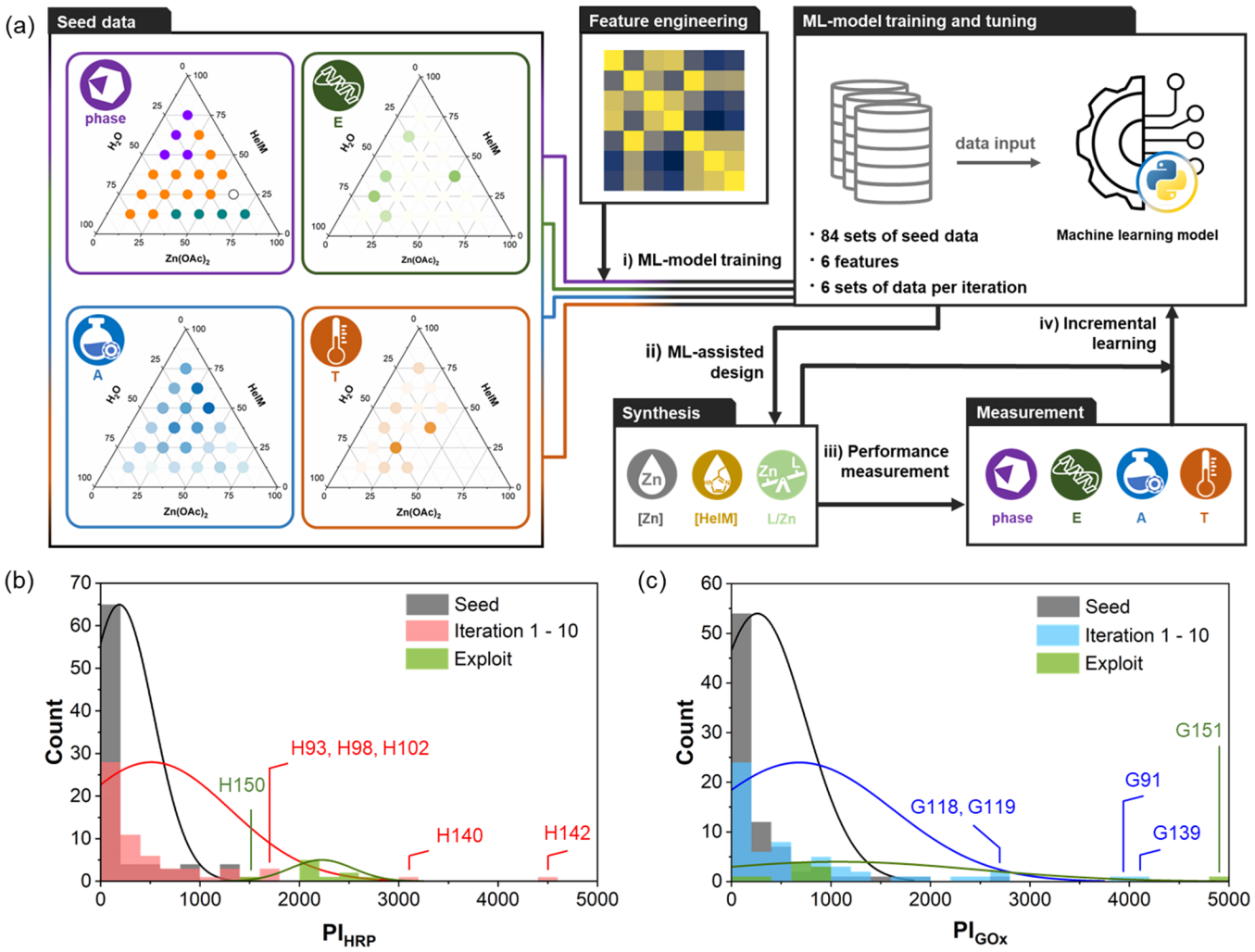


Figure 1. (a) Schematic workflow depicting the iterative ML-assisted enzyme/ZIF discovery process employed in this study. (b, c) Histograms showing the distributions of HRP/ZIFs (b) and GOx/ZIFs (c) according to their PI values. Samples from the seed data set and exploitation stage are shown in gray and green, respectively. Those from the iteration stage are indicated in red for HRP/ZIFs and blue for GOx/ZIFs.

The variation in the synthesis conditions is reported to result in distinct formation mechanisms of enzyme/ZIFs. An operando TEM study reveals that BSA@sod-Zn(mIM)$_2$ (BSA = bovine serum albumin) formed with a low HmIM/Zn molar ratio ($\frac{[HmIM]}{[Zn]}$ = 4) follows a biomimetic mineralization mechanism, while a combination of coprecipitation and biomimetic mineralization mechanisms is observed in the condition of a high HmIM/Zn molar ratio ($\frac{[HmIM]}{[Zn]}$ = 35).[25] The diverse physical properties and performance results of the aforementioned enzyme/ZIFs[7,14] can be partially explained by their distinct formation mechanisms,[14,25] which arise from variations in their synthesis conditions.[7,14]

Derived from the existing enzyme/ZIF examples,[2−5,14] we realize a robust correlation emerges among their synthesis conditions (e.g., concentration of ZIF precursor), structure (e.g., polymorphic phases, structural defect, and crystallinity), and performance (e.g., *encapsulation efficiency (Ein %), retained enzymatic activity (Ain %), and thermal stability (Tin %)*E, A, and stability). However, a comprehensive study to unveil this synthesis−structure−performance relationship in both qualitative and quantitative terms is absent in the literature. This research gap may be attributed to the formidable challenge of conducting a thorough study of enzyme/ZIFs synthesized with an orthogonal combination of different synthesis parameters, a task deemed laborious, time-consuming, and resource-intensive.

Over the past decade, machine learning (ML) has emerged as a powerful tool for advancing materials discovery across scientific domains.[26−37] Its application greatly improved the efficiency of identifying materials with specific target properties. In reticular chemistry, researchers have utilized ML-driven approaches to tailor synthesis conditions for materials, optimizing properties like phase purity,[38] crystallinity,[39] and water sorption performance.[40] ML has also proven particularly valuable in screening MOF candidates for propane/propylene separation[41] and exploring innovative synthesis conditions for the discovery of new materials.[42] A comprehensive ML-assisted study on enzyme/ZIF systems to improve immobilization performance (E, A, and protective effect) is currently absent in the literature.[43] This untapped potential holds the promise of identifying optimal synthesis conditions for high-performance biocomposites as well as unveiling hidden synthesis-structure−property patterns in the enzyme/ZIF system.

In our previous research, we report the better bioactivity exhibited by esterase@rho-Zn(eIM)$_2$ in catalyzing trans-

https://doi.org/10.1021/jacsau.4c00485
JACS Au 2024, 4, 3170−3182

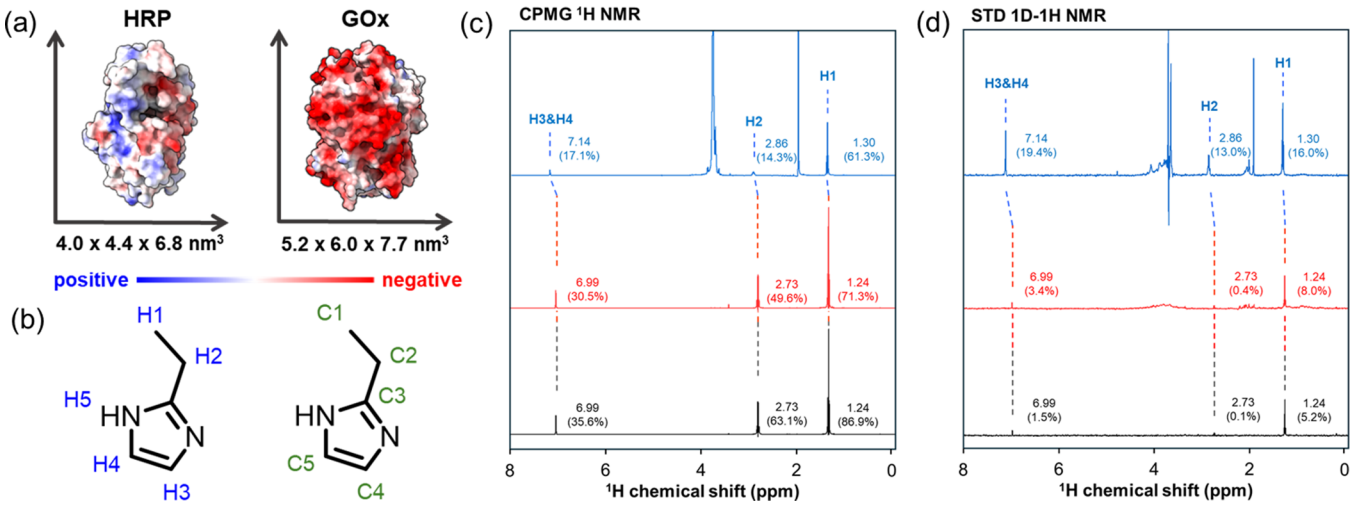


Figure 2. (a) Molecular representations of HRP (PDB entry 1h5d) and GOx (PDB entry 1cf3) showing the Coulombic surface of the enzymes. The color coding on the protein surface indicates the surface charge, with positively charged and negatively charged regions shown in blue and red, respectively. (b) Molecular representation of HeIM, with proton labeling (H1−H5) shown in blue and carbon labeling (C1−C5) in green. (c) 1D ^{1}H CPMG NMR spectra of HeIM (400 μM in D_2O, black), HeIM/HRP (400 μM HeIM and 25 μM HRP in D_2O, red), and HeIM/GOx (400 μM HeIM and 25 μM GOx in D_2O, blue). The peaks in the 1D ^{1}H CPMG spectra are assigned to the protons in HeIM, with their chemical shifts provided (in ppm). The quantitative comparison of the integrated peak areas of the δ_{1H} signals in the CPMG spectra and those in a one-pulse 1D ^{1}H NMR measurements are displayed as percentages in brackets (in %, $\frac{I_{CPMG}}{I_0}$). (d) 1D ^{1}H STD NMR spectra (on-resonance) of HeIM (400 μM in D_2O, black), HeIM/HRP (400 μM HeIM and 25 μM HRP in D_2O, red), and HeIM/GOx (400 μM HeIM and 25 μM GOx in D_2O, blue). The peaks in the 1D ^{1}H STD spectra (on-resonance) are assigned to the protons in HeIM, with their chemical shifts provided (in ppm). The quantitative comparison of the integrated peak areas of the δ_{1H} signals in the on- and off-resonance spectrum are displayed as percentages in brackets (in %, $\frac{I_{on-resonance}}{I_{off-resonance}}$).

esterification reactions when compared to esterase@sod-Zn(mIM)$_2$.[8] However, the hydrophobic nature of rho-Zn(eIM)$_2$ hinders bioactivity in GOx@rho-Zn(eIM)$_2$.[8]

Here, we report an ML-driven iterative training-design-synthesis-measurement workflow that targets optimizing [Zn] and [HeIM] in an enzyme/Zn(eIM)$_2$ synthesis to maximize its performance index (PI) value. PI is calculated as the product of encapsulation efficiency (E in %) and retained enzymatic bioactivity (A in %) (PI = $E \times A$). The random forest (RF) algorithm was employed in predicting the enzyme/ZIF's structure and performance (Figures S7 and S8). Trained by the data in seed and iteration studies, the RF model accurately predicts the phase (>75%), encapsulation efficiency ($R^2 > 0.9$), and retained bioactivity ($R^2 > 0.8$) of enzyme/Zn(eIM)$_2$ biocomposites. Importantly, the ML-assisted design samples (G151 and H150) exhibited significantly higher PI values compared with those in the seed data. Moreover, statistical analyses derived from the trained RF model (e.g., PCP and t-SNE studies) uncover synthesis−structure, synthesis−performance, and structure-performance relationships in the enzyme/ZIF system, complementing experimental observations in the literature and providing new insights that may aid in future enzyme/ZIF synthesis endeavors.

Overall, the proposed ML-based methodology is envisioned to accelerate the design of high-performing enzyme/ZIFs in biocatalysis applications. It is also noteworthy to mention that the proposed ML-based methodology is versatile and can be adapted to optimize the encapsulation of various biomacromolecules in ZIFs, including DNA, RNA, glycans, cells, and beyond.

RESULTS AND DISCUSSION

As shown in Figure 1a, commencing with a seed data set and feature engineering, the proposed ML workflow encompasses four stages: (1) ML model learning: training an ML model using experimental data; (2) ML-assisted design: utilizing the Bayesian optimization (BO) algorithm in the trained ML model to propose new synthesis recipes for enzyme/ZIFs with an enhanced PI value; (3) synthesis, characterization, and test: synthesizing the proposed enzyme/Zn(eIM)$_2$, followed by powder X-ray diffraction (PXRD) measurement, protein assay, and activity assay to obtain characterization (phase) and performance (E, A, and thermal stability (T in %)) data; (4) incremental learning: integrating the synthesis recipes and characterization/performance data of enzyme/Zn(eIM)$_2$ obtained after each iteration back into the existing data set, providing feedback to the ML model to initiate a new iteration cycle. It is worth noting that the spatial localization of an enzyme in the biocomposite was not the focus of this study and has not been examined. Thus, we termed our biocomposite in this study as enzyme/ZIF to avoid confusion.

Seed Data

To initiate the ML-assisted enzyme/Zn(eIM)$_2$ design workflow depicted in Figure 1, our first step was to acquire the seed data set.

The seed data set for ML model training can be derived from literature data, experimental results, or a combination of both.[27,36,44,45] Given limited data on enzyme/Zn(eIM)$_2$,[8,18] we decided to experimentally generate seed data sets. Two chemically distinct enzymes—GOx and horseradish peroxidase (HRP)—were selected for the synthesis of biocomposites, with the aim of examining our ML-assisted enzyme/ZIF design paradigm. HRP is a single-chain glycoprotein containing four

https://doi.org/10.1021/jacsau.4c00485
JACS Au 2024, 4, 3170−3182

disulfide bridges, with a molecular mass of ~44 kDa[46] and a dimension of 4.0 × 4.4 × 6.8 nm^3 (Figure 2a);[47] while GOx is a dimeric glycoprotein with a molecular mass of 160 kDa[48] and a dimension of 5.2 × 6.0 × 7.7 nm^3 (Figure 2a).[47] The isoelectric points for GOx and HRP are 4.2[49] and 3−9,[50] respectively.

In a biomimetic mineralization process, a strong ligand−protein and/or metal−protein affinity is reported to be vital for establishing a concentrated local ligand−metal ion environment on the protein surface. This environment renders the protein molecules as the nucleation sites, enabling and/or accelerating MOF formation around them.[1]

To investigate the propensity of GOx and HRP to trigger biomimetic mineralization processes, we sought to examine the HeIM binding to GOx or HRP using 1D ^{1}H Carr−Purcell−Meiboom−Gill (CPMG) and 1D ^{1}H saturation transfer difference (STD) NMR techniques. Both CPMG and STD methods are widely employed in studying the binding of small molecules to proteins, particularly in the field of drug design and discovery.[51]

As shown in Figure 2b−d, the CPMG and STD NMR results collectively substantiate a high binding affinity of HeIM to GOx and a weak HeIM−HRP interaction (Figures S1−S4). This is evidenced by the downfield shifts of all δ_{1H} signals from HeIM in the HeIM/GOx sample (Figure 2c,d), the decreased integrated areas of all δ_{1H} peaks in CPMG NMR measures for HeIM/GOx (Figure 2c), and the appearance of all HeIM protons in the on-resonance STD spectrum of HeIM/GOx (Figure 2d). Based on the CPMG and STD NMR results, it is anticipated that the formation mechanisms for GOx/Zn(eIM)$_2$ and HRP/Zn(eIM)$_2$ would be distinct. Consequently, the optimal synthesis conditions for these biocomposites are expected to be different.

To obtain the seed data set, we experimentally synthesized a total of 168 samples: 84 for GOx/ZIFs and 84 for HRP/ZIFs. The synthesis conditions covered diverse combinations of [Zn] (25 to 1125 mM) and [HeIM] (150 to 4500 mM) (Figures S5 and S6). The structure data for enzyme/ZIFs were obtained via a comparison of their PXRD patterns with the simulated rho-Zn(eIM)$_2$,[20] ana-Zn(eIM)$_2$,[20] qtz-Zn(eIM)$_2$,[52] and zinc hydroxyl acetate.[53] Subsequently, performance data (E, A, and T) of each biocomposite were assessed through Bradford[54] and activity assays. Summarizing all synthesis, structure, and performance information, we presented comprehensive ternary figures for the seed studies in Figures S5 and S6, forming the foundation for the subsequent ML model training step.

Upon examining the seed data set (Figures S5 and S6), we realized that a significant portion of enzyme/ZIFs exhibited relatively low E, A, and T values. The mean values for GOx/ZIFs were calculated to be E = 35.7 ± 29.2%, A = 7.6 ± 12.5%, and T = 2.8 ± 9.0%, while those for HRP/ZIFs were E = 31.5 ± 34.7%, A = 7.1 ± 11.5%, and T = 11.5 ± 18.8%. These modest outcomes can be partially attributed to the constrained chemical space explored during the systematic screening. Additionally, we found that the phase and performance results of the biocomposites vary, according to their synthesis conditions. This observation supports our hypothesis that the structure/performance of an enzyme/ZIF can be fine-tuned by simply adjusting the [Zn] and [HeIM] in its synthesis.

Another notable discovery in the seed data is that variation results are observed for GOx/ZIFs and HRP/ZIFs in phase, E, A, and T data under any given synthesis condition (Figure S5 versus Figure S6). Such differences in structure and performance data between GOx- and HRP-based biocomposites stem from the distinct molecular size, surface charge, surface functionality patterns, and HeIM binding strength in GOx and HRP (Figure 2). The chemical nature of a protein molecule was reported to be a pivotal factor in influencing the structure and performance of enzyme/ZIF biocomposites.[55−57] For example, the morphology and performance of CAT@ZIF-8 (CAT = catalase) and GOx@ZIF-8 have been reported to intricately link to the functionalization and/or folding nature of CAT and GOx.[57]

ML Model Selection and Evaluation

With the seed data in hand, we then shifted our focus to identify synthesis features that influence the structure and performance outcomes of an enzyme/ZIF (Figures S10 and S11) and ML models applied in the subsequent studies (Figures S7 and S8).

Synthesis feature engineering was carried out based on literature reviews,[13,58] relative importance analysis (Figure S10), and heat map analysis (Figure S9). Four distinct ML algorithms (gradient boosting (GB), support vector machine (SVM), neural network (NN), and RF)[36] were trained and tested to identify a reliable ML model in predicting the phase and performance of enzyme/ZIFs (Figures S7 and S8). The RF model showed the best performance in predicting the structure and performance among the studied ML algorithms and was selected for subsequent studies.

Iterative ML-Assisted Enzyme/ZIF Design

With the synthesis features and RF model selected, we then trained the RF model by feeding it with 84 sets of seed data for each enzyme. In this study, we exclude the T data in the calculation of PI values and the BO-assisted sampling process (Figures S7 and S8).

In the literature, the optimization of synthesis formulation of an enzyme/ZIF is generally carried out based on intuition and experience.[7,8,10,13] However, intuition and experience may introduce unconscious biases that may favor specific conditions and could result in oversight of unconventional conditions that hold untapped effectiveness. In ML, the BO algorithm is generally applied for experiment planning to achieve optimization goals with the minimum number of trials.[36,59]

In this study, we coupled BO with the RF model to plan our synthesis experiments during the iteration stages, with the objective of balancing exploration (searching in regions of the chemical space where the enzyme/ZIFs' PI values are uncertain) and exploitation (searching in regions where the enzyme/ZIFs' PIs are expected to be optimal based on the current knowledge). Typically, 30 synthesis proposals (from a total chemical space comprising approximately 12,300 experiments) were suggested by the RF-BO algorithm. BO-suggested biocomposites with the top six PI values were subjected to synthesis and structure/performance tests. Thereafter, the synthesis conditions for the new enzyme/ZIFs, along with their corresponding structure and performance data, were integrated with the existing data set to retrain the RF model, initiating a new iteration cycle (Figure 1a). We concluded the iteration stage after ten iteration runs. This was attributed to (1) the convergence of PI_{max} values for the enzyme/ZIFs suggested by the BO-RF algorithm; (2) a reduction in fluctuations observed in PI_{max} values for the BO-suggested enzyme/ZIFs; and (3) the limitation of a fixed budget allocated for optimization efforts

https://doi.org/10.1021/jacsau.4c00485
JACS Au 2024, 4, 3170−3182

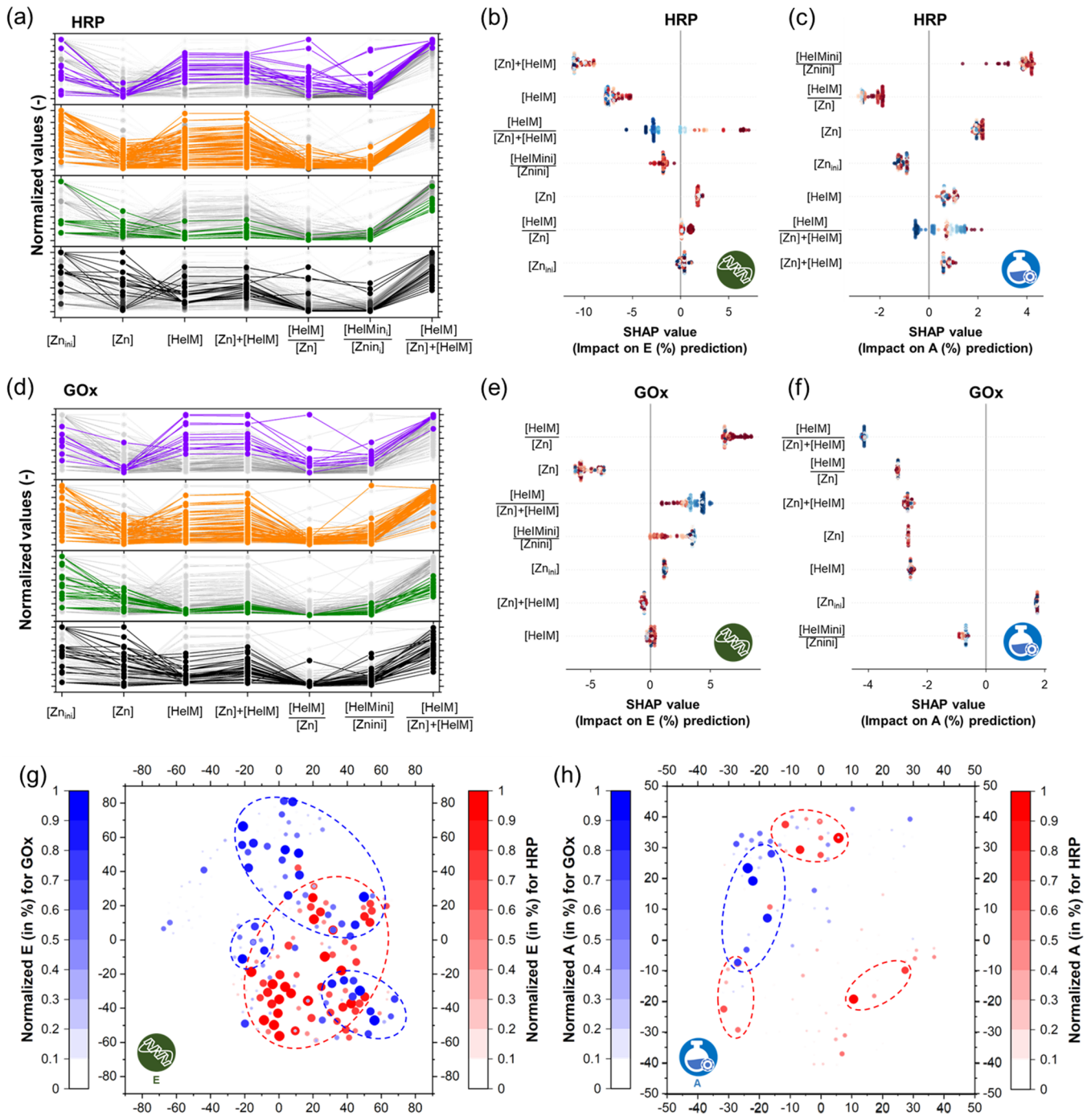


Figure 3. PCPs visualizing the synthesis features of all HRP/ZIF (a) and GOx/ZIF (d) samples in both the seed and iteration stages. The enzyme/rho-Zn(eIM)$_2$, enzyme/qtz-Zn(eIM)$_2$, enzyme/zinc hydroxyl acetate, and solution phases are colored purple, orange, green, and black, respectively. PCPs for all studied data are shown as gray lines and spots in the background. SHAP values for HRP/ZIF (b and c) and GOx/ZIF (e and f) to quantify the importance of synthesis features on the prediction of the encapsulation efficiency (E) and retained bioactivity (A). (g, h) Two-dimensional t-SNE visualization showing all 288 distinct synthesis recipes for GOx/ZIF (blue) and HRP/ZIF (red). Prior to dimension reduction, normalization is applied to all synthesis features. The color and size coding of the data points corresponds to the experimental values of E (g) and A (h) with deeper shades and larger spots indicating higher values. Optimal synthesis regions, conducive to achieving high-performance values, are outlined by circles with dashed lines.

(Figure S11). We propose that alternative iteration-stopping criteria for active learning can be devised and implemented in future studies.

The refined RF model (after 10× iterations) exhibits commendable performance in predicting phase E and phase A for both GOx/ZIFs and HRP/ZIFs. The accuracy for phase prediction is notably high, reaching 83.1% for GOx/ZIF and 75.1% for HRP/ZIF. Strong correlation coefficients ($R^2 > 0.8$) were observed in the cases of E and A, indicating a reliable predictive capability of the refined RF model (Figure S12).

The efficacy of our ML-assisted enzyme/ZIF design strategy in exploring untapped synthesis regions is evidenced in the comparison of synthesis recipes in the seed data and iterations in t-SNE plots (t-SNE = t-distributed stochastic neighbor

https://doi.org/10.1021/jacsau.4c00485
JACS Au 2024, 4, 3170−3182

embedding; Figure S15). As shown in Figure S15, our approach effectively explores synthesis regions that were previously untapped during systematic searches. To further demonstrate the advantage of our proposed ML-assisted enzyme/ZIF design strategy, we plotted the histogram of enzyme/ZIFs in the seed and iteration stages based on their PI values. As shown in Figure 1b,c, the proposed ML-assisted enzyme/ZIF design strategy facilitates the identification of numerous GOx/ZIFs (Figure 1c, e.g., G91, G118, G119, and G139) and HRP/ZIFs (Figure 1b, e.g., H93, H98, H102, H140, and H142) with enhanced PI values. The median PI values of enzyme/ZIFs discovered in the iteration stage exhibit a substantial increase as compared to those in the seed data set ($\overline{PI}_{iteration}$ = 681.4 versus $\overline{PI}_{seed}$ = 260.3 for GOx/ZIFs; $\overline{PI}_{iteration}$ = 511.9 versus $\overline{PI}_{seed}$ = 186.9 for HRP/ZIFs). The best-performing GOx/ZIF and HRP/ZIF discovered within the iteration stage show remarkable PI values of 4046 (G139) and 4436.7 (H142), respectively, surpassing both the $\overline{PI}_{mean}$ and PI_{max} observed in their representative seed data set (Figure 1b,c). In quantitative terms, enzyme/ZIFs discovered in the iteration stage (comprising 41.6% of all samples in the seed and iteration stages) are notably overrepresented as top performers, accounting for 70.0 and 63.3% of the top twentieth percentile of GOx/ZIFs and HRP/ZIFs, respectively, when sorted by PI values.

In all, results from statistical (Figure 1b,c) and t-SNE (Figure S15) analyses highlight the efficacy and effectiveness of our ML-driven approach in exploring and exploiting the optimal synthesis conditions for enzyme/ZIFs. Our strategy achieves improving overall enzyme/ZIF performance with less than 70% additional data beyond the initial systematic screen and approximately 0.49% of the total design space.

Understanding the Synthesis Driving Enzyme/ZIF Performance

In the subsequent stage, we intended to employ parallel coordinate plots (PCPs) to examine the impact of synthesis features on enzyme/ZIF phase formation (Figures 3a,d, S13, and S14).

In previous studies, we and others have suggested that a high [Zn], [ligand], and/or a high $\frac{[ligand]}{[Zn]}$ tend to promote the formation of a porous kinetic ZIF phase. Conversely, a low [Zn], [ligand], and/or a low $\frac{[ligand]}{[Zn]}$ is more likely to result in the formation of a dense thermodynamic ZIF phase.[13,17] Our findings of this study are partially in line with this published hypothesis while providing new insights. For instance, by examining the phase distribution of all samples within both the seed and iteration data sets in PCPs (Figure 3a,d), we observed that low [Zn], high [HeIM], high [Zn] + [HeIM], and/or high $\frac{[HeIM]}{[Zn]+[HeIM]}$ values tend to promote the formation of the porous rho-Zn(eIM)$_2$ phase. Conversely, to obtain a thermodynamic product (qtz-Zn(eIM)$_2$), a synthesis condition characterized by a low [Zn], a low $\frac{[HeIM]}{[Zn]}$, and/or a high $\frac{[HeIM]}{[Zn]+[HeIM]}$ ratio is preferred. Surprisingly, we found that variations in $\frac{[HeIM]}{[Zn]}$ and [HeIM] have minimal influence on determining the formation of rho-Zn(eIM)$_2$ and qtz-Zn(eIM)$_2$, respectively, which contradicts the hypothesis presented in the literature. While variations exist, the overall trend regarding the impact of synthesis features on determining the phase of GOx/ZIF and HRP/ZIF remains consistent.

With the new knowledge obtained from PCP analysis (Figure 3a,d), our vision of enzyme/ZIF synthesis is refreshed and advanced. For instance, to achieve crystalline enzyme/ZIF products, we recommend that a low concentration of [Zn] and a high concentration of [HeIM] should be employed in the synthesis. Additionally, if the kinetic phase is the desired outcome, we suggest prioritizing a high $\frac{[HeIM]}{[Zn]+[HeIM]}$ ratio rather than a high $\frac{[HeIM]}{[Zn]}$ value in an enzyme/ZIF synthesis.

In the subsequent analysis, we focus on examining the influence of each synthesis feature on the performance of the enzyme/ZIF biocomposite. Unlike the enzyme/ZIF phase, the impact of synthesis features on the enzyme/ZIF performance is an underexplored area. We sought to quantify the impact of each synthesis feature on the prediction of E and A in their respective RF models by calculating Shapley additive explanations (SHAP) values (Figure 3b,c,e,f). Although the SHAP values are not directly derived from experimental data, the strong fitting of our RF models to the enzyme/ZIF's E and A values (Figure S12) suggests that they could effectively reflect the synthesis−performance relationship in the enzyme/ZIF. As depicted in Figure 3b,c,e,f, different synthesis features exhibit distinct importance rankings in predicting the performance of an enzyme/ZIF. Such a synthesis−performance relationship is enzyme-dependent and can be attributed to the distinct chemical characteristics of HRP and GOx (Figure 2a), which may alter the kinetics and/or mechanisms of enzyme/ZIF formation.[7,56,57] This concept has been illustrated experimentally and reported by Doonan and his team. The authors have demonstrated that the surface charge of a protein, as indicated by its zeta-potential value or isoelectric point, governs the success or formation kinetics of biocomposite synthesis. Specifically, proteins with a negative surface charge tend to accelerate the formation of enzyme/ZIF biocomposites, while a suppression effect is observed for proteins with a positively charged surface.[55]

For HRP, [Zn] + [HeIM] ranks as the most important synthesis feature in predicting E, with a negative SHAP value (Figure 3b). This implies that E_{HRP} increases with a decrease of [Zn] + [HeIM]. Conversely, in the case of GOx/ZIF, the most important synthesis feature governing E_{GOx} prediction is $\frac{[HeIM]}{[Zn]}$ (Figure 3e), with a positive SHAP value. This indicates that a higher $\frac{[HeIM]}{[Zn]}$ molar ratio favors a higher E_{GOx}. In terms of A, $\frac{[HeIM_{ini}]}{[Zn_{ini}]}$ and $\frac{[HeIM]}{[Zn]+[HeIM]}$ rank the most important synthesis features in A_{HRP} and A_{GOx} prediction, respectively. Specifically, $\frac{[HeIM]}{[Zn]}$ (with a negative SHAP value) contributes significantly to the A prediction, ranking second in importance for both GOx/ZIF and HRP/ZIF (Figure 3c,f). This observation aligns well with the experimental findings reported in the literature. For example, lipase@ZIF-8[7] and GOx@ZIF-8[14] synthesized under a low $\frac{[HeIM]}{[Zn]}$ molar ratio ($\frac{[HeIM]}{[Zn]}$ = 4) were reported to exhibit higher A values as compared to their counterparts synthesized under a high $\frac{[HeIM]}{[Zn]}$ condition.

Although there is no systematic study that quantitatively correlates properties between enzymes and enzyme/ZIFs, existing qualitative evidence suggests that variations in the

https://doi.org/10.1021/jacsau.4c00485
JACS Au 2024, 4, 3170−3182

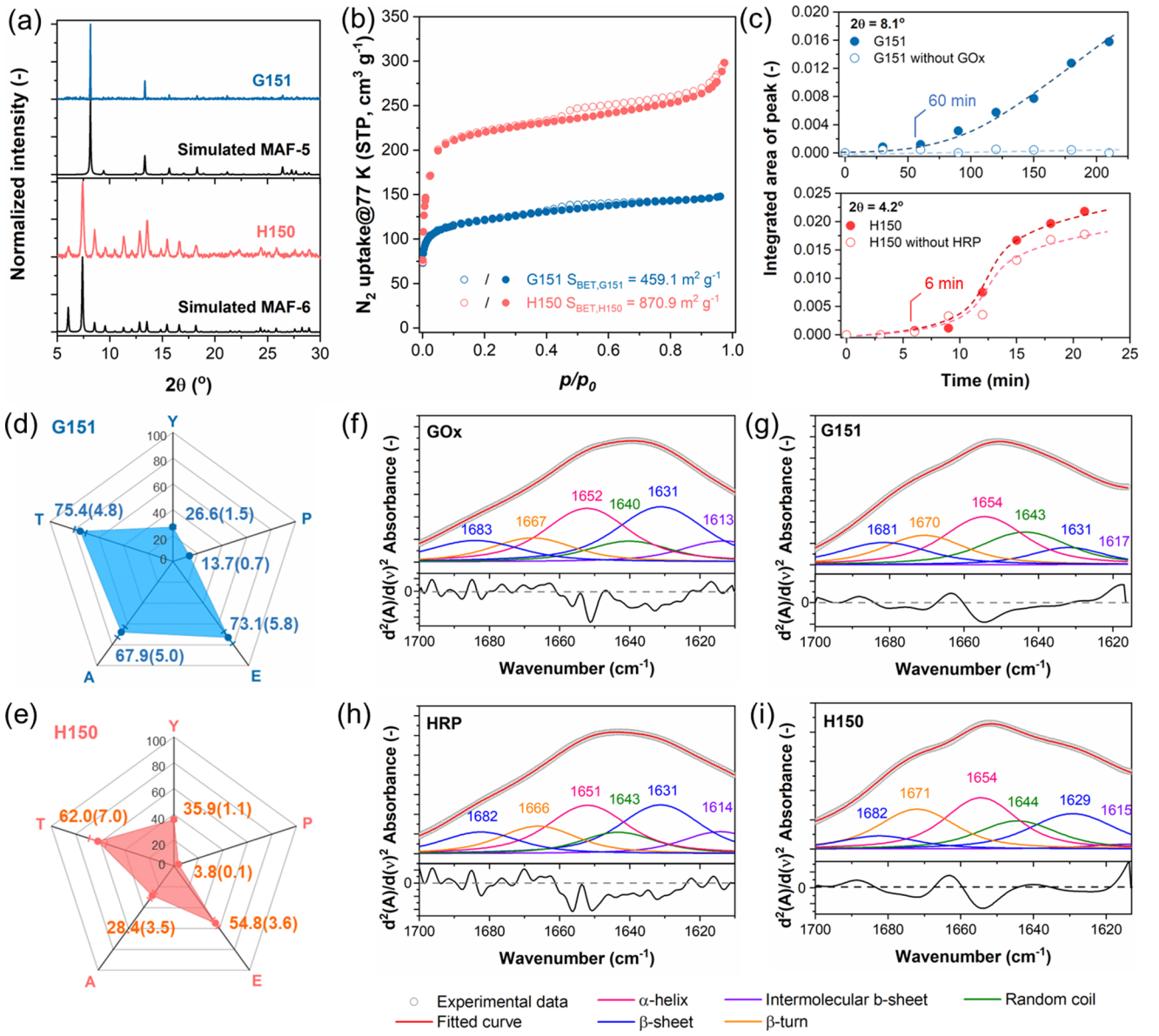


Figure 4. (a) Simulated PXRD patterns for MAF-5 and MAF-6, along with the experimental PXRD patterns for H150 and G151. (b) 77 K N_2 sorption isotherms of G151 (blue) and H150 (red). Closed and open symbols represent adsorption and desorption isotherms, respectively. (c) Growth of characteristic diffraction peaks for G151 (2θ = 8.1°; filled blue circle and solid blue line) and its control sample synthesized without the presence of GOx (open blue circle and dashed blue line), depicted in the upper position of the combined figure. Additionally, H150 (2θ = 4.2°, filled red circle and solid red line) and its control sample synthesized without the presence of HRP (open red circle and dashed red line) are illustrated in the lower position of the combined figure. (d, e) Radar plots summarizing the key immobilization parameters (*Y* in %, *P* in %, *E* in %, *A* in %, and *T* in %) for G151 (d) and H150 (e). Error bars were generated from at least three replicates. (f–i) ATR-FTIR spectra for GOx (f), G151 (g), HRP (h), and H150 (i), showing the structural content of α-helix (pink), β-sheet (blue), intermolecular β-sheet (purple), β-turn (orange), and random coil (green). The structural contents were determined by the Voigt function fitted to OMNIC software.

surface chemical patterns,[55] degrees of enzyme folding structure,[56,57] and enzyme stability toward synthesis conditions[60] all contribute to the diverse structure and performance of resulting enzyme/ZIFs. To clearly depict and contrast the synthesis recipes and performance data of HRP/ZIFs and GOx/ZIFs, we have consolidated them into t-SNE plots (Figure 3g,h). The hot synthesis zones, characterized by high-performance values (*E*, *A*), are notably distinct for GOx/ZIFs and HRP/ZIFs. This observation underscores the unique and tailored synthesis requirements for optimizing the performance of each enzyme/ZIF biocomposite. Additionally, from another perspective, these findings affirm the validity and versatility of our ML-assisted enzyme/ZIF design approach in effectively identifying high-performing enzyme/ZIFs using different enzymes.

In the next step, we summarized all data from the seed and iteration stages in box plots according to the phase of the samples, aiming to study the structure–performance relationship in enzyme/ZIFs.

As depicted in Figure S16, rho-Zn(eIM)$_2$ and ana-Zn(eIM)$_2$ were determined to be the most desired immobilization platforms for HRP and GOx, respectively. In the literature, amorphous ZIF (aZIF) synthesized by reacting Zn^{2+} and mIM^- was reported to be an ideal immobilization platform for enzymes (GOx, HRP, lactate oxidase, and *Candida antarctica* lipase B), exhibiting high *E*, *A*, and *T* values.[11,61] However, in

https://doi.org/10.1021/jacsau.4c00485
JACS Au 2024, 4, 3170−3182

our case, although the amorphous enzyme/ZIFs exhibited relatively high PI values (Figure S16), their thermal stability was found to be poor. We attribute this discrepancy in experimental observations to the different choices of organic ligands in the biocomposite.

In all, the presented PCP, SHAP, and t-SNE analyses (Figure 3) align closely with the experimental findings reported in the literature and provide valuable new insights into the synthesis−structure, synthesis−performance, and structure−performance relationships in enzyme/ZIFs. These observations advance our understanding of the synthesis-driving enzyme/ZIF performance and are envisioned to contribute to the design of high-performing enzyme/ZIF biocomposites in future studies.

Characterization and Performance Test for H150 and G151

The successful development of a trained RF model to predict the structure/performance of enzyme/ZIFs encouraged us to exploit the synthesis conditions for high-performing HRP/ZIFs and GOx/ZIFs. To achieve this, we initially clustered and filtered the biocomposites predicted by the trained RF model based on their synthesis conditions and PI values using the density-based spatial clustering of applications with noise (DBSCAN) method. From the DBSCAN-filtered data set, the top ten enzyme/ZIF candidates based on their PI ranking were selected for experimental synthesis and performance testing (E, A, and T). Based on the OPI values (overall performance index, OPI = $E \times A \times T$), we selected H150 and G151 as the best-performing biocomposites (Figure 1b,c). Importantly, the OPI values for G151 and H150 are 5 times and 1.3 times higher than the OPI values of the best samples in their respective seed data sets ($OPI_{G151} = 3.74 \times 10^5$ vs $OPI_{GS9} = 7.58 \times 10^4$; $OPI_{H150} = 9.6 \times 10^4$ vs $OPI_{H31} = 7.28 \times 10^4$). PXRD measurements confirmed the crystalline phases for G151 and H150 to be ana-Zn(eIM)$_2$ and rho-Zn(eIM)$_2$, respectively (Figure 4a).

Comparing the operando X-ray diffraction results between the H150/G151 synthesis solutions and those without enzyme addition revealed that H150 formed via a coprecipitation pathway, while G151 formed through a biomimetic mineralization mechanism (Figures 4d, S23, and S24). In the G151 synthesis solution, particles formed immediately upon mixing all synthesis components in H_2O (Figure S23), with G151 crystallizing after ∼60 min (Figure 4c). Conversely, when GOx was absent from the synthesis recipe, the Zn^{2+}-HeIM mixture remained as a solution (Figures 4c and S23), indicating the crucial role of GOx in the formation of G151. Interestingly, the presence of HRP did not affect the crystallization kinetics of rho-Zn(eIM)$_2$ (Figure 4c). We rationalized that the different formation mechanisms observed for H150 and G151 stemmed from the varying affinity of HeIM for HRP and GOx (Figure 2). As shown in the PCP analysis (Figure 3a,d), a high-$\frac{[HeIM]}{[Zn]+[HeIM]}$ condition in the enzyme/ZIF synthesis is crucial for obtaining crystalline GOx/ZIF and HRP/ZIF. In addition, high crystallinity is important for achieving high thermal stability and thus a high OPI value in the biocomposite as evidenced by statistical studies (Figure S16). Based on the information obtained from Figures 2, 3a,d, 4c, and S16, we propose that if an organic ligand has a high affinity toward a protein surface, a high local ligand concentration around the protein surface will be generated. This will result in a highly localized $\frac{[HeIM]}{[Zn]+[HeIM]}$ environment around the protein surface, triggering ZIF crystallization around the protein surface. In this case, the biomimetic mineralization process occurs even at a relatively lower $\frac{[HeIM]}{[Zn]+[HeIM]}$ concentration in the synthesis medium. Otherwise, a high-$\frac{[HeIM]}{[Zn]+[HeIM]}$ condition is necessary to facilitate ZIF formation and coprecipitate the enzyme to form the enzyme/ZIF biocomposite. In this regard, we suggest that the proposed CPMG/STD NMR methods (Figure 2) may serve as a straightforward approach to estimate the propensity of a protein to biomineralize a ZIF-based biocomposite with a specific organic ligand or vice versa.

The porosity of the studied samples was determined by 77 K N_2 sorption analysis (Figure 4b), yielding Brunauer−Emmett−Teller (BET) specific surface areas of 459.1 and 870.9 $m^2\ g^{-1}$ for G151 and H150, respectively. Structural defects in G151 and H150 were evidenced by the presence of expanded pores in these samples, as confirmed by the pore size distribution studies (Figure S19) and TEM images (Figure S29). The presence of intrinsic and defect-induced porosities in G151 and H150 (Figures 4b and S19) benefits substrate and product diffusion through the biocomposites in an enzymatic reaction.

To examine the tertiary structure of immobilized enzymes and unveil the structure−bioactivity relationship, we measured and compared the deconvoluted ATR-FTIR spectra for GOx, G151, HRP, and H150 (Figures 4f−i, S28, and Table S1). In the literature, ATR-FTIR spectroscopy is widely adopted to investigate the tertiary structure of proteins, particularly focusing on the characteristic amide I region (1700−1600 cm^{-1}).[62] Conformational change in a protein upon immobilization can be sensitively reflected in the change of the characteristic stretching of its secondary structural elements (α-helix, β-sheet, β-turn, and random coils).[62−68] As shown in Figure 4f and Table S1, lyophilized GOx shows six peaks that were assigned to α-helix (1652 cm^{-1}, 26.7%), β-sheet (1631 and 1683 cm^{-1}, 40.5%), intermolecular β-sheet (1613 cm^{-1}, 10.6%), β-turn (1667 cm^{-1}, 11.9%), and random coil (1640 cm^{-1}, 10.4%). Quantitative analysis reveals that GOx appears in a less aggregated form in/on G151, evidenced by a significant decrease in the intermolecular β-sheet content for GOx within the biocomposite (from 10.6 to 0.2%, Table S1). Additionally, compared to free GOx, there is a reduction in β-sheet content (from 40.5 to 26.2%) and an increase in random coil content (from 10.4 to 22.7%) for GOx in/on G151, suggesting a slight unfolding of GOx upon immobilization. However, the α-helix stretching of GOx in G151 remains relatively constant in terms of stretching position (1652 versus 1654 cm^{-1}) and content (26.7% versus 31.5%), implying that the tertiary structure of GOx in/on G151 is largely retained (Figure 4g,h and Table S1). A similar trend was observed for HRP versus H150 (Figure 4h,i and Table S1).

The presence of expanded porosities in G151 and H150 (Figures 4b and S19), along with the preservation of the tertiary structure of the immobilized enzyme in H150 and G151 (Figures 4f,g,i), explains the observed enzymatic activity (Figure 4d,e).

As compared to free enzymes, both H150 and G151 exhibited better stability upon thermal treatment (60 °C, 60 min; Figures 4e,f, S25, and S26). The immobilization parameters for G151 and H150 are summarized in Figure 4e,f. The yield (Y in %, based on Zn), enzyme loading (P in wt %), E, A, and T were determined, respectively, to be 26.6 ±

https://doi.org/10.1021/jacsau.4c00485
JACS Au 2024, 4, 3170−3182

1.5%, 13.7 ± 0.7%, 73.1 ± 5.8 wt %, 67.9 ± 5.0%, and 75.4 ± 4.8% for G151 (Figure 4e). For H150, *Y*, *P*, *E*, *A*, and *T* were measured to be 35.9 ± 1.1%, 3.8 ± 0.1 wt %, 54.8 ± 3.6%, 28.4 ± 3.5%, and 62.0 ± 7.0%, respectively (Figure 4f). Enzyme loadings in G151 and H150 were corroborated via TGA analysis (Figure S20). Control experiments confirmed that enzyme leaching from G151 and H150 was negligible as no bioactivity was detected in the supernatant from the aqueous suspension of G151 and H150 (Figure S27).

CONCLUSIONS

In conclusion, we have developed and applied an ML-assisted design protocol to optimize enzyme/ZIF synthesis for the first time in the literature (Figure 1). The robustness and versatility of this ML design methodology are demonstrated by its success in optimizing enzyme/ZIF biocomposites with two chemically and physically distinct model enzymes, GOx and HRP. Through an iterative ML-assisted training-design-synthesis-measurement workflow, we efficiently identified enzyme/ZIFs (G151 and H150) with OPI values that are at least 1.3 times higher than those obtained from seed data studies.

We also realized that the optimized synthesis recipes and formation mechanisms of ZIF-based biocomposites are enzyme-dependent. G151 was formed under a low-$\frac{[HeIM]}{[Zn]}$ (3.5) condition via a biomimetic mineralization mechanism, while H150 was formed under a high-$\frac{[HeIM]}{[Zn]}$ (17) condition via a coprecipitation pathway. 1D ^{1}H CPMG and STD NMR measurements suggest that the distinct formation mechanism and optimized synthesis protocol observed for G151 and H150 are attributed to the different interaction affinity between the HeIM ligand and the protein (Figure 2). HeIM can strongly bind to GOx and be concentrated on the protein surface, while weak HeIM−HRP binding is observed. This information refreshes and advances our understanding of enzyme/ZIF synthesis, providing valuable guidelines for future studies.

In our perspective, the proposed ML-based optimization strategy is not confined to optimizing the synthesis features and target set investigated in this study. Instead, it can be easily tailored to optimize various other synthesis features (e.g., ligand composition, synthesis time, and enzyme dosage) with different optimization targets (e.g., release profile and sensing property) in ZIF-based systems. There is another clear opportunity to adopt the ML-assisted workflow and synthesis−structure−performance relationship insights developed from the present study to rationally optimize the enzyme immobilization using other reticular framework materials, e.g., enzyme@MOF,[2,6,69] enzyme@COF,[70] and enzyme@HOF,[71] enabling the realization of new biocatalytic materials with enhanced immobilization performance. By efficiently planning experiments and establishing robust prediction models, we envision that ML will accelerate the discovery, optimization, and fundamental studies of enzyme/ZIFs, leading to a deeper understanding of their properties and ultimately broadening the range of enzyme/ZIF applications.

EXPERIMENTAL SECTION

Syntheses

All chemicals and solvents were purchased from commercial sources and used as received without further purification. Milli-Q water was used in all experiments.

Seed Data Generating for HRP/ZIF and GOx/ZIF

In a typical set of seed data for enzyme/ZIF synthesis, a desired volume of aqueous stock solution of $Zn(OAc)_2{\cdot}2H_2O$ (reagent grade, Sigma-Aldrich), HeIM (98%, Sigma-Aldrich), enzyme (glucose oxidase (GOx), glucose oxidase from *Aspergillus niger*, Type VII, lyophilized powder, ≥100,000 units/g solid (without added oxygen), Sigma-Aldrich; horseradish peroxidase (HRP), peroxidase from horseradish, Type VI, essentially salt-free, lyophilized powder, ≥250 units/mg solid (using pyrogallol), Sigma-Aldrich), and H_2O were mixed. The concentration of the enzyme solution was 8 mg mL^{-1}. The enzyme concentration was set at 1 mg mL^{-1} for all of the syntheses. In an enzyme/ZIF synthesis, various mixing orders of Zn^{2+}, ligand, and enzyme can be employed.[12,60] While the choice of mixing order has been reported to influence the enzyme/ZIF performance,[12,60] it is not the focus of this study. In the present study, we adhere to a synthesis protocol involving the mixing of HeIM and enzyme in H_2O prior to the addition of Zn^{2+}. To maintain consistency across different samples, we kept the enzyme dosage at 1 mg mL^{-1} during the synthesis. Additionally, we set the synthesis temperature at 4 °C and adhered to a 24 h synthesis duration.

To enable variation on [Zn] and [HeIM] in the seed data, we use four different concentrations of $Zn(OAc)_2{\cdot}2H_2O$ and HeIM stock solutions for four seed data sets. The choice of concentration was based on the experience we gained from our previous study[8] and the water solubility of the chemicals. The concentrations of the stock aqueous solution of $Zn(OAc)_2{\cdot}2H_2O$ for $GOx_{seed}1/HRP_{seed}1$, $GOx_{seed}2/HRP_{seed}2$, $GOx_{seed}3/HRP_{seed}3$, and $GOx_{seed}4/HRP_{seed}4$ samples were set at 1.0, 0.5, 0.2, and 1.5 M, respectively. The concentrations of the stock aqueous solution of HeIM for $GOx_{seed}1/HRP_{seed}1$, $GOx_{seed}2/HRP_{seed}2$, $GOx_{seed}3/HRP_{seed}3$, and $GOx_{seed}4/HRP_{seed}4$ samples were set at 6.0, 3.0, 1.2, and 5.0 M, respectively. The mixing order for the synthesis components was H_2O, enzyme stock solution (62.5 μL), HeIM stock solution, and then $Zn(OAc)_2{\cdot}2H_2O$ solution. After the addition of $Zn(OAc)_2{\cdot}2H_2O$, the synthesis mixture was vortexed for 30 s. Thereafter, the reaction mixture was left under static conditions at 4 °C for 24 h. The volume of the synthesis was 0.5 mL. After synthesis, the precipitates were collected by centrifugation (148,000 rpm for 3 min) and washed three times with Milli-Q water. The collected enzyme/ZIF was resuspended in Milli-Q water (0.5 mL) and subjected to the Bradford assay and activity assay for encapsulation efficiency (*E* in %) and retained enzymatic bioactivity (*A* in %) determination, respectively. Thereafter, PXRD was conducted on air-dried enzyme/ZIFs. The PXRD patterns were compared to the simulated data to identify the MOF phases in samples. Results were then summarized as ternary plots of $[Zn(OAc)_2]$ versus [HeIM] versus $[H_2O]$.

Synthesis of MAF-6

MAF-6 was synthesized according to the reported protocol.[16] In a typical synthesis, a concentrated aqueous ammonia solution (25% NH_4OH, Supelco; 40 mL) of ZnO (Merck; 162.8 mg) was added dropwise into a methanol (Merck; 30 mL) and cyclohexane (Sigma-Aldrich; 2 mL) solution of HeIM (384 mg) under stirring conditions at room temperature. After a 30 min reaction, the resulting precipitate was collected via centrifugation and washed three times with H_2O to remove any unreacted precursors. Thereafter, the MAF-6 particles were dried under vacuum at room temperature.

Characterization

Powder X-ray Diffraction (PXRD). PXRD measurements were performed on a Rigaku SmartLAB SE powder diffractometer using Cu Kα (λ = 1.5406 Å) radiation.

Attenuated Total Reflection Fourier Transform Infrared Spectroscopy (ATR-FTIR) Analysis. ATR-FTIR measurements were carried out on a Thermo Nicolet iS50 infrared spectrometer. Repeated spectra were collected in the spectral range from 4000 to 700 cm^{-1}, averaging 128 scans at 4 cm^{-1} spectral resolution. ATR-FTIR spectra were collected and analyzed using OMNIC software. Prior to the peak deconvolution study, the ATR-FITR spectrum of the enzyme/ZIF was subtracted by that of MAF-6. Thereafter, second

https://doi.org/10.1021/jacsau.4c00485
JACS Au 2024, 4, 3170−3182

derivatives were computed on the subtracted spectra by using a Savitzky−Golay algorithm with 25 smoothing points. The number and frequencies of the individual band components in the spectral range of 1610−1700 cm^{-1} were determined based on the secondary derivative analysis. Six bands (intermolecular β-sheet (1600−1619 cm^{-1}), β-sheet (1620−1639 and 1680−1700 cm^{-1}), random coil (1640−1649 cm^{-1}), α-helix (1648−1660 cm^{-1}), and β-turn (1660−1679 cm^{-1})) were targeted according to the literature.[63−65,67,68,72] In a deconvolution process, bands were allowed to move in a ±2 cm^{-1} range from their initial position. The full width at half-heights (fwhh) of the bands were fitted in a range from 0 to 50 cm^{-1}. Deconvolution was fitted by using a Voigt function (OMNIC 9.8, Thermo Fisher Scientific). The secondary structure content was estimated from the relative band areas.

In Situ Small-Angle X-ray Diffraction (SAXRD). In situ SAXRD experiments were carried out on an Anton Paar SAXS point diffractometer. In a typical experiment, a synthesis solution of ZIF or enzyme/ZIF was prepared in a 2 mL Eppendorf centrifuge tube and injected immediately after preparation into a quartz capillary (Ø = 1 mm) placed in the X-ray beam. The SAXRD measurements were started 1 min after the synthesis solution was prepared.

Quantification Analysis of Enzyme Loading

To quantitatively measure the loading of enzymes in enzyme/ZIF samples, a dye-binding assay (Bradford assay) was conducted on the acid-digested enzyme/ZIF sample. In a typical experiment, 10 μL of enzyme/ZIF suspension was mixed with 100 μL of aqueous 0.05 M HCl in a 96-well plate. Thereafter, 150 μL of Bradford reagent (Sigma-Aldrich) was introduced and stored at room temperature for 30 min prior to UV−vis analysis (SPECTROstar Nano spectrometer, BMG LABTECH). The maxima at 595 nm were recorded and compared to the calibration curve to calculate the enzyme encapsulation efficiency (E in %) in enzyme/ZIF samples. Protein assay experiments were repeated three times. E was calculated for each enzyme/ZIF by the following equation:

$$\text{Encapsulation efficiency}(E) = \frac{m_{\text{enzyme in enzyme/ZIF}}}{m_{\text{enzyme used in the enzyme/ZIF synthesis}}}\%$$

Enzymatic Assay for GOx and GOx/ZIF

The enzymatic activity of GOx and GOx/ZIF was evaluated by using OPD (OPD = *o*-phenylenediamine) as the substrate. The GOx solution or biocomposites dispersed in water which contained 80 μg of GOx (10 μL of stock solution containing 0.1 mg mL^{-1} GOx) was added and mixed with 200 μL of GOx stock assay solution (200 μL) in a 96-well plate. The final volume of the assay solution was 210 μL. The final assay solution was performed in Tris−HCl buffer (pH 7.0, 0.1 mM) with a GOx concentration of 4.76 μg mL^{-1}, an OPD concentration of 2.28 mg mL^{-1}, a glucose concentration of 5.71 mg mL^{-1}, and an HRP concentration of 63.5 μg mL^{-1}. The evolution of absorbance at 418 nm was monitored using a SPECTROstar Nano spectrometer (BMG LABTECH). Assay experiments were repeated three times.

Enzymatic Assay for HRP and HRP/ZIF

The activity of HRP and HRP/ZIF was determined by measuring the rate of decomposition of hydrogen peroxide with an OPD as the hydrogen donor. In a typical assay, solution A was prepared by mixing 10 mL of Tris−HCl buffer (100 mM, pH 7.0), 100 μL of H_2O_2 88 mM in H_2O, and 50 μL of OPD (7.2 mg mL^{-1} in H_2O). In a 96-well plate, 10 μL of HRP (0.01 mg mL^{-1}) or HRP/ZIF suspension (0.01 mg mL^{-1}, based on HRP) was mixed with 200 μL of solution A. Thereafter, the absorbance of the solution was immediately monitored at 418 nm by using a SPECTROstar Nano spectrometer (BMG LABTECH). The initial reaction rate of the enzymatic assay was used to represent the bioactivity of the free enzyme and enzyme/ZIFs. The retained enzymatic bioactivity (A in %) was calculated for each enzyme/ZIF by the following equation:

$$\text{Retained enzymatic bioactivity}(A) = \frac{A_{\text{enzyme/ZIF}}}{A_{\text{free enzyme}}}\%$$

Thermal Stability Test for Enzyme/ZIFs

In a typical thermal stability test, 100 μL of enzyme (0.1 mg mL^{-1} GOx or 0.01 mg mL^{-1} HRP in H_2O) or enzyme/ZIF (0.1 mg mL^{-1} GOx/ZIF (based on GOx) or 0.01 mg mL^{-1} HRP/ZIF (based on HRP) suspension in H_2O) was subjected to thermal treatment (60 °C, 60 min). Our selection of 60 °C was rational. The melting temperatures for GOx and HRP were reported to be 60 and 45 °C, respectively.[73,74] After thermal treatment, the bioactivities of GOx, HRP, and enzyme/ZIF were subjected to bioactivity assay tests. Thermal stability (T in %) was calculated for each enzyme/ZIF by the following equation:

$$\text{Thermal stability}(T) = \frac{A_{\text{free enzyme or enzyme/ZIF after thermal treatment}}}{A_{\text{free enzyme or enzyme/ZIF without thermal treatment}}}\%$$

Solution-State Nuclear Magnetic Resonance (NMR)

All solution-state NMR experiments were conducted on a Bruker Avance III 600 MHz NMR spectrometer equipped with a proton-optimized triple resonance (H/C/N) CryoProbe. All NMR data were analyzed using Bruker Topspin v4.2.0. Three samples were prepared for 1D 1H CPMG and STD-NMR measurements: (1) 400 μM HeIM in D_2O; (2) 25 μM GOx and 400 μM HeIM in D_2O; and (3) 25 μM HRP and 400 μM HeIM in D_2O.

CPMG spectra were acquired with a total echo time of 600 ms, a relaxation delay of 5 s, and an acquisition time of ∼0.4 s. A total of 128 scans were collected with 16 dummy scans.

For GOx, the on- and off-resonance frequencies for STD measurements were set at −395 and 20,000 Hz, respectively. For HRP, the on- and off-resonance frequencies for STD measurements were set at −860 and 20,000 Hz, respectively. The saturation time was set at 3 s. A total of 128 scans were collected with 16 dummy scans. All measurements were performed at 298 K.

77 K N_2 Sorption Analysis

Gas adsorption isotherms were obtained on an Autosorb IQ physisorption analyzer. Approximately 40 mg of the sample was placed into a glass analysis tube and degassed under a dynamic vacuum for 16 h at 120 °C prior to measurement. Nitrogen (N_2) adsorption and desorption isotherms were measured at 77 K. The isotherms were then analyzed to determine the BET surface area using the ASiQwin software. Pore size distribution was modeled from the N_2 adsorption isotherms using the NLDFT model.

Thermal Gravimetric Analysis (TGA)

TGA measurements were conducted on a TA STD650 instrument. In a typical analysis, approximately 2 mg of samples were placed on an alumina pan for TGA measurement. The TGA measurements were conducted in an oxidative atmosphere (air). The weight change of the samples was recorded from 50 to 700 °C at a ramp rate of 10 °C min^{-1}.

ML Model Selection and Development

ML Algorithm. Four ML algorithms, namely, GB, SVM, NN, and RF, were tested in this study. Within each algorithm framework, four separate models are trained on the seed data set, each aimed at predicting different outcomes: phase, E, A, and T. Specifically, the phase prediction employs a classifier model due to the categorical nature of the target variable, while the prediction of E, A, and T uses regression models.

A 10-fold cross-validation approach was implemented for each model. Cross-validation is a widely utilized technique to evaluate a model's capacity to generalize to unseen data. It offers a more robust and consistent estimation of the model's performance compared to a single train/test split. It is especially beneficial when the data set size is limited, a common constraint in laboratory experimental research.

https://doi.org/10.1021/jacsau.4c00485
JACS Au 2024, 4, 3170−3182

In this approach, the seed data set is divided into 10 approximately equal subsets (referred to as folds). Notably, for the phase prediction task where a classification model was employed, a stratified K-fold cross-validation strategy was adopted. This ensures that each fold maintains a proportional representation of samples for each class as in the original data set. The model undergoes training 10 times, each time utilizing a different fold as the test set while the remaining 9 folds for training. Subsequently, 10 sets of statistical metrics are derived from the model prediction, facilitating the calculation of an average score that indicates the model's performance.

In the phase prediction task, the indicative metric used is prediction accuracy, which represents the proportion of accurately classified instances relative to the total number of instances in the data set. Conversely, for the remaining target variables (A, E, and T), the mean squared error (MSE) is employed.

After evaluation of the performance scores of the four ML algorithms, RF is found to be the most accurate model for this study.

It is worth noting that data points with 0% E values were eliminated for the A modeling, while data points with 0% E or 0% A values were eliminated for the T modeling.

ML-Assisted Experiment Planning. The BO algorithm was applied in tandem with an RF model to propose promising enzyme/ZIF candidates. During the iteration stage, enzyme/ZIFs that maximize the expected improvement (EI) acquisition function were given by

$$f(\vec{x}) = Z\sigma(\vec{x})\Phi(Z) + \sigma(\vec{x})\emptyset(Z)$$

$$Z = \begin{cases} \dfrac{(\mu(\vec{x}) - f' - \xi)}{\sigma(\vec{x})}, & \sigma(\vec{x}) > 0 \\ 0, & \sigma(\vec{x}) = 0 \end{cases}$$

where $\mu(\vec{x})$ is the predicted mean PI ($E \times A$) from the RF, f' is the current largest mean PI predicted by the model, $\sigma(\vec{x})$ is the standard deviation of PI from the RF, and Φ and $\emptyset$ are the cumulative and probability density functions of the normal distribution, respectively. ξ (ξ = 0.01 in the present study) is a hyperparameter that controls the balance between exploring untapped regions of the chemical space and exploiting known regions of it to achieve enzyme/ZIFs with a high PI.

Specifically, after generating 30 enzyme/ZIF designs with BO, candidates were ranked by their PI values in descending order and iteratively chosen for the final set of six candidates.

A different down-sampling strategy was implemented during the exploitation stage. Various variables in the synthesis recipe, such as $[Zn_{stock}]$, $[HeIM_{stock}]$, V_{Zn}, and V_{HeIM} (where $[Zn_{stock}]$ represents the concentration of $Zn(OAc)_2{\cdot}2H_2O$ in the Zn aqueous stock solution, $[HeIM_{stock}]$ represents the HeIM concentration in the HeIM aqueous stock solution, V_{Zn} denotes the volume of the Zn aqueous stock solution used in the enzyme/ZIF synthesis, and V_{HeIM} indicates the volume of the HeIM aqueous stock solution used in the enzyme/ZIF synthesis), were included in the trained RF models to predict the E and A values. A constraint was enforced on the RF model to maintain the total volume of the synthesis mixture at 0.5 mL. Candidates with PI values (products of E and A) greater than 1600 were then clustered using a DBSCAN method with a distance threshold of 0.3 and a minimum of three points per cluster. After cluster formation, the enzyme/ZIF with the shortest distance to the centroid position of each cluster in the enzyme/ZIF synthesis feature vector space was selected as a representative candidate for further consideration. The filtered synthesis proposals were then ranked in descending order based on their predicted PI values, and the top ten candidates were selected for synthesis and performance testing.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacsau.4c00485.

ML model selection, feature engineering, RF model analysis, ternary figures for seed data, formation kinetics studies, porosity and pore size distribution, TGA analysis, ATR-FTIR analysis, and supplementary data for activity assays (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Weibin Liang − *School of Chemical and Biomolecular Engineering, The University of Sydney, Darlington, NSW 2008, Australia;* orcid.org/0000-0002-3032-436X; Email: weibin.liang@sydney.edu.au

Jun Huang − *School of Chemical and Biomolecular Engineering, The University of Sydney, Darlington, NSW 2008, Australia;* orcid.org/0000-0001-8704-605X; Email: jun.huang@sydney.edu.au

Authors

Sisi Zheng − *School of Chemical and Biomolecular Engineering, The University of Sydney, Darlington, NSW 2008, Australia;* Present Address: Roam Technologies, Carlton, NSW 2218, Australia (S.Z.)

Ying Shu − *School of Chemical and Biomolecular Engineering, The University of Sydney, Darlington, NSW 2008, Australia*

Complete contact information is available at: https://pubs.acs.org/10.1021/jacsau.4c00485

Author Contributions

All authors have given approval to the final version of the manuscript.

Funding

Australian Research Council Discovery Early Career Researcher Award (DECRA) (DE220100583); Australian Research Council Discovery Project (DP220102851).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the Australian Research Council under the Discovery Early Career Researcher Award (DECRA) (DE220100583) and the Australian Research Council Discovery Project (DP220102851). The authors acknowledge the facilities and the scientific and technical assistance of Sydney Analytical, a core research facility at The University of Sydney. The authors extend their heartfelt gratitude to Dr. Biswaranjan Mohanty from the Sydney Analytical Center for his invaluable assistance and insightful discussions during the NMR data collection and results analysis in this study.

■ REFERENCES

(1) Liang, K.; Ricco, R.; Doherty, C. M.; Styles, M. J.; Bell, S.; Kirby, N.; Mudie, S.; Haylock, D.; Hill, A. J.; Doonan, C. J.; Falcaro, P. Biomimetic mineralization of metal-organic frameworks as protective coatings for biomacromolecules. *Nat. Commun.* **2015**, *6*, 7240.

(2) Liang, W.; Wied, P.; Carraro, F.; Sumby, C. J.; Nidetzky, B.; Tsung, C.-K.; Falcaro, P.; Doonan, C. J. Metal−Organic Framework-Based Enzyme Biocomposites. *Chem. Rev.* **2021**, *121*, 1077−1129.

(3) Doonan, C.; Riccò, R.; Liang, K.; Bradshaw, D.; Falcaro, P. Metal−Organic Frameworks at the Biointerface: Synthetic Strategies and Applications. *Acc. Chem. Res.* **2017**, *50*, 1423−1432.

(4) Chen, G.; Kou, X.; Huang, S.; Tong, L.; Shen, Y.; Zhu, W.; Zhu, F.; Ouyang, G. Modulating the Biofunctionality of Metal−Organic-

Framework-Encapsulated Enzymes through Controllable Embedding Patterns. *Angew. Chem., Int. Ed.* **2020**, *59*, 2867−2874.
(5) Huang, S.; Kou, X.; Shen, J.; Chen, G.; Ouyang, G. Armor-Plating" Enzymes with Metal−Organic Frameworks (MOFs). *Angew. Chem., Int. Ed.* **2020**, *59*, 8786−8798.
(6) Akpinar, I.; Wang, X.; Fahy, K.; Sha, F.; Yang, S.; Kwon, T.-w.; Das, P. J.; Islamoglu, T.; Farha, O. K.; Stoddart, J. F. Biomimetic Mineralization of Large Enzymes Utilizing a Stable Zirconium-Based Metal-Organic Frameworks. *J. Am. Chem. Soc.* **2024**, *146*, 5108−5117.
(7) Maddigan, N. K.; Linder-Patton, O. M.; Falcaro, P.; Sumby, C. J.; Bell, S. G.; Doonan, C. J. Influence of the Synthesis and Storage Conditions on the Activity of Candida antarctica Lipase B ZIF-8 Biocomposites. *ACS Appl. Mater. Interfaces* **2021**, *13*, 51867−51875.
(8) Liang, W.; Flint, K.; Yao, Y.; Wu, J.; Wang, L.; Doonan, C.; Huang, J. Enhanced Bioactivity of Enzyme/MOF Biocomposite via Host Framework Engineering. *J. Am. Chem. Soc.* **2023**, *145*, 20365−20374.
(9) Liang, W.; Xu, H.; Carraro, F.; Maddigan, N. K.; Li, Q.; Bell, S. G.; Huang, D. M.; Tarzia, A.; Solomon, M. B.; Amenitsch, H.; Vaccari, L.; Sumby, C. J.; Falcaro, P.; Doonan, C. J. Enhanced Activity of Enzymes Encapsulated in Hydrophilic Metal−Organic Frameworks. *J. Am. Chem. Soc.* **2019**, *141*, 2348−2355.
(10) Li, Y.-M.; Yuan, J.; Ren, H.; Ji, C.-Y.; Tao, Y.; Wu, Y.; Chou, L.-Y.; Zhang, Y.-B.; Cheng, L. Fine-Tuning the Micro-Environment to Optimize the Catalytic Activity of Enzymes Immobilized in Multivariate Metal−Organic Frameworks. *J. Am. Chem. Soc.* **2021**, *143*, 15378−15390.
(11) Wu, X.; Yue, H.; Zhang, Y.; Gao, X.; Li, X.; Wang, L.; Cao, Y.; Hou, M.; An, H.; Zhang, L.; Li, S.; Ma, J.; Lin, H.; Fu, Y.; Gu, H.; Lou, W.; Wei, W.; Zare, R. N.; Ge, J. Packaging and delivering enzymes by amorphous metal-organic frameworks. *Nat. Commun.* **2019**, *10*, 5165.
(12) Hu, C.; Bai, Y.; Hou, M.; Wang, Y.; Wang, L.; Cao, X.; Chan, C.-W.; Sun, H.; Li, W.; Ge, J.; Ren, K. Defect-induced activity enhancement of enzyme-encapsulated metal-organic frameworks revealed in microfluidic gradient mixing synthesis. *Sci. Adv.* **2020**, *6*, No. eaax5785.
(13) Liang, W.; Ricco, R.; Maddigan, N. K.; Dickinson, R. P.; Xu, H.; Li, Q.; Sumby, C. J.; Bell, S. G.; Falcaro, P.; Doonan, C. J. Control of Structure Topology and Spatial Distribution of Biomacromolecules in Protein@ZIF-8 Biocomposites. *Chem. Mater.* **2018**, *30*, 1069−1077.
(14) Tong, L.; Huang, S.; Shen, Y.; Liu, S.; Ma, X.; Zhu, F.; Chen, G.; Ouyang, G. Atomically unveiling the structure-activity relationship of biomacromolecule-metal-organic frameworks symbiotic crystal. *Nat. Commun.* **2022**, *13*, 951.
(15) Feng, Y.; Cao, X.; Zhang, L.; Li, J.; Cui, S.; Bai, Y.; Chen, K.; Ge, J. Defect engineering of enzyme-embedded metal−organic frameworks for smart cargo release. *Chem. Eng. J.* **2022**, *439*, No. 135736.
(16) He, C.-T.; Jiang, L.; Ye, Z.-M.; Krishna, R.; Zhong, Z.-S.; Liao, P.-Q.; Xu, J.; Ouyang, G.; Zhang, J.-P.; Chen, X.-M. Exceptional Hydrophobicity of a Large-Pore Metal−Organic Zeolite. *J. Am. Chem. Soc.* **2015**, *137*, 7217−7223.
(17) Jian, M.; Liu, B.; Liu, R.; Qu, J.; Wang, H.; Zhang, X. Water-based synthesis of zeolitic imidazolate framework-8 with high morphology level at room temperature. *RSC Adv.* **2015**, *5*, 48433−48441.
(18) Ren, H.; Yuan, J.; Li, Y.-M.; Li, W.-J.; Guo, Y.-H.; Zhang, Y.-B.; Wang, B.-H.; Ma, K.; Peng, L.; Hu, G.; Wang, W.-Q.; He, H.; Chou, L.-Y.; Zeng, M.-H.; Zhang, Y.-B.; Cheng, L. Highly Enantioselective Catalysis by Enzyme Encapsulated in Metal Azolate Frameworks with Micelle-Controlled Pore Sizes. *ACS Cent. Sci.* **2024**, *10*, 358−366.
(19) Park, K. S.; Ni, Z.; Côté, A. P.; Choi, J. Y.; Huang, R.; Uribe-Romo, F. J.; Chae, H. K.; O'Keeffe, M.; Yaghi, O. M. Exceptional chemical and thermal stability of zeolitic imidazolate frameworks. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 10186−10191.
(20) Huang, X.-C.; Lin, Y.-Y.; Zhang, J.-P.; Chen, X.-M. Ligand-Directed Strategy for Zeolite-Type Metal−Organic Frameworks: Zinc(II) Imidazolates with Unusual Zeolitic Topologies. *Angew. Chem., Int. Ed.* **2006**, *45*, 1557−1559.
(21) Chen, R.; Yao, J.; Gu, Q.; Smeets, S.; Baerlocher, C.; Gu, H.; Zhu, D.; Morris, W.; Yaghi, O. M.; Wang, H. A two-dimensional zeolitic imidazolate framework with a cushion-shaped cavity for CO2 adsorption. *Chem. Commun.* **2013**, *49*, 9500−9502.
(22) Cui, J.; Feng, Y.; Lin, T.; Tan, Z.; Zhong, C.; Jia, S. Mesoporous Metal−Organic Framework with Well-Defined Cruciate Flower-Like Morphology for Enzyme Immobilization. *ACS Appl. Mater. Interfaces* **2017**, *9*, 10587−10594.
(23) Zhou, Z.; Chao, H.; He, W.; Su, P.; Song, J.; Yang, Y. Boosting the activity of enzymes in metal-organic frameworks by a one-stone-two-bird enzymatic surface functionalization strategy. *Appl. Surf. Sci.* **2022**, *586*, No. 152815.
(24) Gao, S.; Hou, J.; Deng, Z.; Wang, T.; Beyer, S.; Buzanich, A. G.; Richardson, J. J.; Rawal, A.; Seidel, R.; Zulkifli, M. Y.; Li, W.; Bennett, T. D.; Cheetham, A. K.; Liang, K.; Chen, V. Improving the Acidic Stability of Zeolitic Imidazolate Frameworks by Biofunctional Molecules. *Chem.* **2019**, *5*, 1597−1608.
(25) Ogata, A. F.; Rakowski, A. M.; Carpenter, B. P.; Fishman, D. A.; Merham, J. G.; Hurst, P. J.; Patterson, J. P. Direct Observation of Amorphous Precursor Phases in the Nucleation of Protein−Metal−Organic Frameworks. *J. Am. Chem. Soc.* **2020**, *142*, 1433−1442.
(26) Luo, Y.; Bag, S.; Zaremba, O.; Cierpka, A.; Andreo, J.; Wuttke, S.; Friederich, P.; Tsotsalas, M. MOF Synthesis Prediction Enabled by Automatic Data Mining and Machine Learning. *Angew. Chem., Int. Ed.* **2022**, *61*, No. e202200242.
(27) Tamasi, M. J.; Patel, R. A.; Borca, C. H.; Kosuri, S.; Mugnier, H.; Upadhya, R.; Murthy, N. S.; Webb, M. A.; Gormley, A. J. Machine Learning on a Robotic Platform for the Design of Polymer−Protein Hybrids. *Adv. Mater.* **2022**, *34*, No. 2201809.
(28) Burger, B.; Maffettone, P. M.; Gusev, V. V.; Aitchison, C. M.; Bai, Y.; Wang, X.; Li, X.; Alston, B. M.; Li, B.; Clowes, R.; Rankin, N.; Harris, B.; Sprick, R. S.; Cooper, A. I. A mobile robotic chemist. *Nature* **2020**, *583*, 237−241.
(29) Perera, D.; Tucker, J. W.; Brahmbhatt, S.; Helal, C. J.; Chong, A.; Farrell, W.; Richardson, P.; Sach, N. W. A platform for automated nanomole-scale reaction screening and micromole-scale synthesis in flow. *Science* **2018**, *359*, 429−434.
(30) Wilbraham, L.; Mehr, S. H. M.; Cronin, L. Digitizing Chemistry Using the Chemical Processing Unit: From Synthesis to Discovery. *Acc. Chem. Res.* **2021**, *54*, 253−262.
(31) Caramelli, D.; Granda, J. M.; Mehr, S. H. M.; Cambié, D.; Henson, A. B.; Cronin, L. Discovering New Chemistry with an Autonomous Robotic Platform Driven by a Reactivity-Seeking Neural Network. *ACS Cent. Sci.* **2021**, *7*, 1821−1830.
(32) Minato, T.; Salley, D.; Mizuno, N.; Yamaguchi, K.; Cronin, L.; Suzuki, K. Robotic Stepwise Synthesis of Hetero-Multinuclear Metal Oxo Clusters as Single-Molecule Magnets. *J. Am. Chem. Soc.* **2021**, *143*, 12809−12816.
(33) Jiang, Y.; Salley, D.; Sharma, A.; Keenan, G.; Mullin, M.; Cronin, L. An artificial intelligence enabled chemical synthesis robot for exploration and optimization of nanomaterials. *Sci. Adv.* **2022**, *8*, No. eabo2626.
(34) Cronin, L.; Kitson, P. J. Selection of assembly complexity in a space of tetrapeptides. *Chem.* **2022**, *8*, 1791−1793.
(35) She, S.; Bell, N. L.; Zheng, D.; Mathieson, J. S.; Castro, M. D.; Long, D.-L.; Koehnke, J.; Cronin, L. Robotic synthesis of peptides containing metal-oxide-based amino acids. *Chem.* **2022**, *8*, 2734−2748.
(36) Tao, H.; Wu, T.; Aldeghi, M.; Wu, T. C.; Aspuru-Guzik, A.; Kumacheva, E. Nanoparticle synthesis assisted by machine learning. *Nat. Rev. Mater.* **2021**, *6*, 701−716.
(37) Notin, P.; Rollins, N.; Gal, Y.; Sander, C.; Marks, D. Machine learning for functional protein design. *Nat. Biotechnol.* **2024**, *42*, 216−228.
(38) Du, M.-H.; Dai, Y.; Jiang, L.-P.; Su, Y.-M.; Qi, M.-Q.; Wang, C.; Long, L.-S.; Zheng, L.-S.; Kong, X.-J. Exploration and Insights on Topology Adjustment of Giant Heterometallic Cages Featuring

https://doi.org/10.1021/jacsau.4c00485
JACS Au 2024, 4, 3170−3182

Inorganic Skeletons Assisted by Machine Learning. *J. Am. Chem. Soc.* **2023**, *145*, 23188−23195.
(39) Zheng, Z.; Zhang, O.; Nguyen, H. L.; Rampal, N.; Alawadhi, A. H.; Rong, Z.; Head-Gordon, T.; Borgs, C.; Chayes, J. T.; Yaghi, O. M. ChatGPT Research Group for Optimizing the Crystallinity of MOFs and COFs. *ACS Cent. Sci.* **2023**, *9*, 2161−2170.
(40) Zheng, Z.; Alawadhi, A. H.; Chheda, S.; Neumann, S. E.; Rampal, N.; Liu, S.; Nguyen, H. L.; Lin, Y.-h.; Rong, Z.; Siepmann, J. I.; Gagliardi, L.; Anandkumar, A.; Borgs, C.; Chayes, J. T.; Yaghi, O. M. Shaping the Water-Harvesting Behavior of Metal−Organic Frameworks Aided by Fine-Tuned GPT Models. *J. Am. Chem. Soc.* **2023**, *145*, 28284−28295.
(41) Wang, Y.; Jiang, Z.-J.; Wang, D.-R.; Lu, W.; Li, D. Machine Learning-Assisted Discovery of Propane-Selective Metal−Organic Frameworks. *J. Am. Chem. Soc.* **2024**, *146*, 6955−6961.
(42) Zheng, Z.; Rong, Z.; Rampal, N.; Borgs, C.; Chayes, J. T.; Yaghi, O. M. A GPT-4 Reticular Chemist for Guiding MOF Discovery. *Angew. Chem., Int. Ed.* **2023**, *62*, No. e202311983.
(43) Chai, M.; Moradi, S.; Erfani, E.; Asadnia, M.; Chen, V.; Razmjou, A. Application of Machine Learning Algorithms to Estimate Enzyme Loading, Immobilization Yield, Activity Retention, and Reusability of Enzyme−Metal−Organic Framework Biocatalysts. *Chem. Mater.* **2021**, *33*, 8666−8676.
(44) Bender, A.; Schneider, N.; Segler, M.; Patrick Walters, W.; Engkvist, O.; Rodrigues, T. Evaluation guidelines for machine learning tools in the chemical sciences. *Nat. Rev. Chem.* **2022**, *6*, 428−442.
(45) Zheng, Z.; Zhang, O.; Borgs, C.; Chayes, J. T.; Yaghi, O. M. ChatGPT Chemistry Assistant for Text Mining and the Prediction of MOF Synthesis. *J. Am. Chem. Soc.* **2023**, *145*, 18048−18062.
(46) Welinder, K. G. Amino Acid Sequence Studies of Horseradish Peroxidase. *Eur. J. Biochem.* **1979**, *96*, 483−502.
(47) Lian, X.; Chen, Y.-P.; Liu, T.-F.; Zhou, H.-C. Coupling two enzymes into a tandem nanoreactor utilizing a hierarchically structured MOF. *Chem. Sci.* **2016**, *7*, 6969−6973.
(48) Tsuge, H.; Natsuaki, O.; Ohashi, K. Purification, Properties, and Molecular Features of Glucose Oxidase from Aspergillus niger. *J. Biochem.* **1975**, *78*, 835−843.
(49) Pazur, J. H.; Kleppe, K. The Oxidation of Glucose and Related Compounds by Glucose Oxidase from Aspergillus niger*. *Biochem.* **1964**, *3*, 578−583.
(50) Shannon, L. M.; Kay, E.; Lew, J. Y. Peroxidase Isozymes from Horseradish Roots: I. ISOLATION AND PHYSICAL PROPERTIES. *J. Biol. Chem.* **1966**, *241*, 2166−2172.
(51) Peng, C.; Frommlet, A.; Perez, M.; Cobas, C.; Blechschmidt, A.; Dominguez, S.; Lingel, A. Fast and Efficient Fragment-Based Lead Generation by Fully Automated Processing and Analysis of Ligand-Observed NMR Binding Data. *J. Med. Chem.* **2016**, *59*, 3303−3310.
(52) Beldon, P. J.; Fábián, L.; Stein, R. S.; Thirumurugan, A.; Cheetham, A. K.; Friščić, T. Rapid Room-Temperature Synthesis of Zeolitic Imidazolate Frameworks by Using Mechanochemistry. *Angew. Chem., Int. Ed.* **2010**, *49*, 9640−9643.
(53) Schlur, L.; Carton, A.; Pourroy, G. A new zinc hydroxy acetate hydrogen carbonate lamellar phase for growing large and clean ZnO nanorod arrays. *Chem. Commun.* **2015**, *51*, 3367−3370.
(54) Kielkopf, C. L.; Bauer, W.; Urbatsch, I. L. Bradford Assay for Determining Protein Concentration. *Cold Spring Harb. Protoc.* **2020**, *2020*, No. 102269.
(55) Maddigan, N. K.; Tarzia, A.; Huang, D. M.; Sumby, C. J.; Bell, S. G.; Falcaro, P.; Doonan, C. J. Protein surface functionalisation as a general strategy for facilitating biomimetic mineralisation of ZIF-8. *Chem. Sci.* **2018**, *9*, 4217−4223.
(56) Carpenter, B. P.; Talosig, A. R.; Mulvey, J. T.; Merham, J. G.; Esquivel, J.; Rose, B.; Ogata, A. F.; Fishman, D. A.; Patterson, J. P. Role of Molecular Modification and Protein Folding in the Nucleation and Growth of Protein−Metal−Organic Frameworks. *Chem. Mater.* **2022**, *34*, 8336−8344.
(57) Carpenter, B. P.; Rose, B.; Olivas, E. M.; Navarro, M. X.; Talosig, A. R.; Hurst, P. J.; Di Palma, G.; Xing, L.; Guha, R.; Copp, S. M.; Patterson, J. P. The role of protein folding in prenucleation clusters on the activity of enzyme@metal−organic frameworks. *J. Mater. Chem. A* **2024**, *12*, 813−823.
(58) Carraro, F.; Velásquez-Hernández, M. D. J.; Astria, E.; Liang, W.; Twight, L.; Parise, C.; Ge, M.; Huang, Z.; Ricco, R.; Zou, X.; Villanova, L.; Kappe, C. O.; Doonan, C.; Falcaro, P. Phase dependent encapsulation and release profile of ZIF-based biocomposites. *Chem. Sci.* **2020**, *11*, 3397−3404.
(59) Shahriari, B.; Swersky, K.; Wang, Z.; Adams, R. P.; Freitas, N. D. Taking the Human Out of the Loop: A Review of Bayesian Optimization. *Proc. IEEE* **2016**, *104*, 148−175.
(60) Kaneti, Y. V.; Dutta, S.; Hossain, M. S. A.; Shiddiky, M. J. A.; Tung, K.-L.; Shieh, F.-K.; Tsung, C.-K.; Wu, K. C.-W.; Yamauchi, Y. Strategies for Improving the Functionality of Zeolitic Imidazolate Frameworks: Tailoring Nanoarchitectures for Functional Applications. *Adv. Mater.* **2017**, *29*, No. 1700213.
(61) Zhang, Y.; Xu, L.; Ge, J. Multienzyme System in Amorphous Metal−Organic Frameworks for Intracellular Lactate Detection. *Nano Lett.* **2022**, *22*, 5029−5036.
(62) Vedantham, G.; Sparks, H. G.; Sane, S. U.; Tzannis, S.; Przybycien, T. M. A Holistic Approach for Protein Secondary Structure Estimation from Infrared Spectra in H2O Solutions. *Anal. Biochem.* **2000**, *285*, 33−49.
(63) Serrano, V.; Liu, W.; Franzen, S. An Infrared Spectroscopic Study of the Conformational Transition of Elastin-Like Polypeptides. *Biophys. J.* **2007**, *93*, 2429−2435.
(64) Kong, J.; Yu, S. Fourier Transform Infrared Spectroscopic Analysis of Protein Secondary Structures. *Acta Biochim. Biophys. Sin.* **2007**, *39*, 549−559.
(65) Singh, B. R. *Infrared Analysis of Peptides and Proteins*; American Chemical Society, 1999, vol *750*, ch. 1; pp 2−37.
(66) Adochitei, A.; Drochioiu, G. Rapid Characterization of peptide secondary structure by FT-IR spectroscopy. *Rev. Roum. Chim.* **2011**, *56*, 783−791.
(67) Vass, E.; Hollósi, M.; Besson, F.; Buchet, R. Vibrational Spectroscopic Detection of Beta- and Gamma-Turns in Synthetic and Natural Peptides and Proteins. *Chem. Rev.* **2003**, *103*, 1917−1954.
(68) Goormaghtigh, E.; Cabiaux, V.; Ruysschaert, J.-M. Secondary structure and dosage of soluble and membrane proteins by attenuated total reflection Fourier-transform infrared spectroscopy on hydrated films. *Eur. J. Biochem.* **1990**, *193*, 409−420.
(69) Cases Díaz, J.; Lozano-Torres, B.; Giménez-Marqués, M. Boosting Protein Encapsulation through Lewis-Acid-Mediated Metal−Organic Framework Mineralization: Toward Effective Intracellular Delivery. *Chem. Mater.* **2022**, *34*, 7817−7827.
(70) Zhang, Y.; Xing, C.; Mu, Z.; Niu, Z.; Feng, X.; Zhang, Y.; Wang, B. Harnessing Self-Repairing and Crystallization Processes for Effective Enzyme Encapsulation in Covalent Organic Frameworks. *J. Am. Chem. Soc.* **2023**, *145*, 13469−13475.
(71) Liang, W.; Carraro, F.; Solomon, M. B.; Bell, S. G.; Amenitsch, H.; Sumby, C. J.; White, N. G.; Falcaro, P.; Doonan, C. J. Enzyme Encapsulation in a Porous Hydrogen-Bonded Organic Framework. *J. Am. Chem. Soc.* **2019**, *141*, 14298−14305.
(72) Adochitei, A.; Drochioiu, G. Rapid characterization of peptide secondary structure by FT-IR spectroscopy. *Rev. Roum. Chim* **2011**, *56*, 783−791.
(73) Zoldák, G.; Zubrik, A.; Musatov, A.; Stupák, M.; Sedlák, E. Irreversible Thermal Denaturation of Glucose Oxidase from Aspergillus niger Is the Transition to the Denatured State with Residual Structure *. *J. Biol. Chem.* **2004**, *279*, 47601−47609.
(74) Chattopadhyay, K.; Mazumdar, S. Structural and Conformational Stability of Horseradish Peroxidase: Effect of Temperature and pH. *Biochem.* **2000**, *39*, 263−270.

https://doi.org/10.1021/jacsau.4c00485
JACS Au 2024, 4, 3170−3182