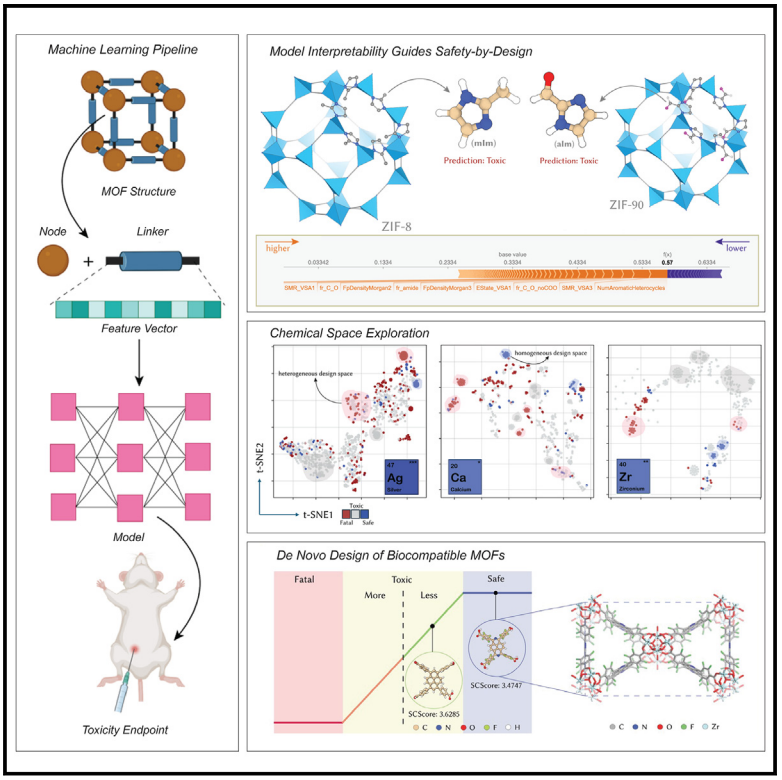
Matter

Article

Guiding the rational design of biocompatible metal-organic frameworks for drug delivery

Graphical abstract



Highlights

- ML-guided computational pipeline for rapid assessment of MOF biocompatibility
- Interpretable ML enables derivation of chemical guidelines for biocompatible MOFs
- High-throughput screening of 86,000 structures for identifying promising candidates
- Rational, *de novo* design of biocompatible MOFs for drug delivery

Authors

Dhruv Menon, David Fairen-Jimenez

Correspondence

df334@cam.ac.uk

In brief

This study presents a machine-learning-guided computational pipeline for the rapid assessment of metal-organic framework (MOF) biocompatibility. Our method enables the high-throughput screening of MOF databases for identifying existing and prospective candidates for drug delivery applications. Moreover, the interpretability of our models facilitates the derivation of chemical guidelines for the rational, *de novo* design of biocompatible MOFs. Our approach offers the possibility of shortening experimental timescales by de-risking structures prior to animal testing, thereby accelerating clinical translation.

3 Understanding
Dependency and conditional studies on material behavior

Menon & Fairen-Jimenez, 2025, Matter *8*, 101958
March 5, 2025
https://doi.org/10.1016/j.matt.2025.101958

CellPress

Matter

CellPress
OPEN ACCESS

Article

Guiding the rational design of biocompatible metal-organic frameworks for drug delivery

Dhruv Menon[1,2] and David Fairen-Jimenez[1,3,*]
[1]The Adsorption & Advanced Materials Laboratory (A²ML), Department of Chemical Engineering & Biotechnology, University of Cambridge, Philippa Fawcett Drive, Cambridge CB3 0AS, UK
[2]Cavendish Laboratory, Department of Physics, University of Cambridge, JJ Thompson Avenue, Cambridge CB3 0DY, UK
[3]Lead contact
*Correspondence: df334@cam.ac.uk
https://doi.org/10.1016/j.matt.2025.101958

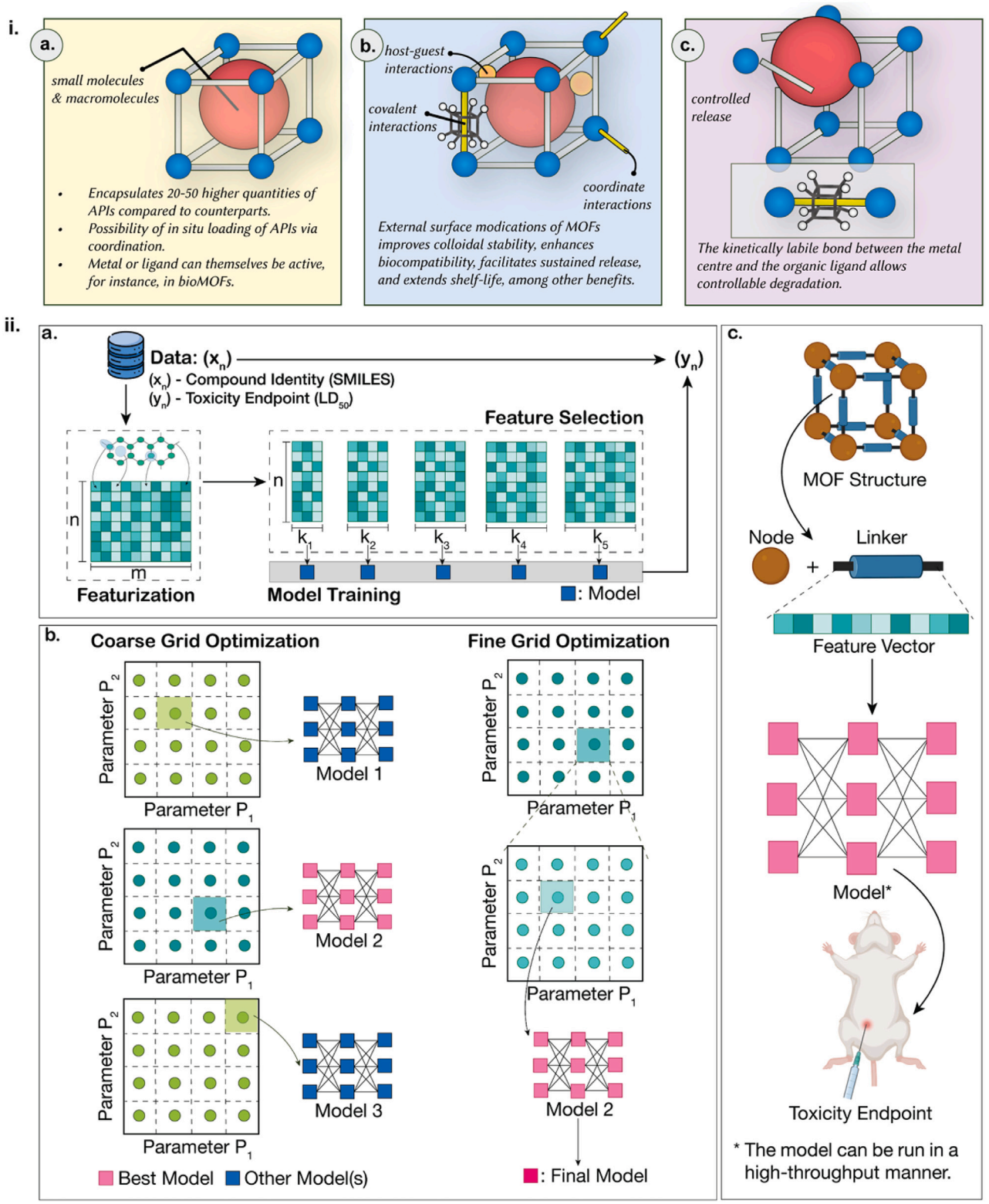
PROGRESS AND POTENTIAL Metal-organic frameworks (MOFs) are rapidly emerging as a promising platform for drug delivery; however, their clinical translation remains hindered by concerns regarding their biocompatibility. Experimental approaches to assess MOF safety are resource intensive, are time consuming, and raise ethical concerns related to extensive animal testing. Moreover, the vast chemical diversity of MOF building blocks makes the navigation of the chemical space intractable to experimental approaches, while classical modeling approaches are unable to capture the behavior of MOFs in complex biological systems. To address this, our study reports a machine-learning-guided computational pipeline for the rapid assessment of MOF biocompatibility based on the toxicity of its building blocks. Using this pipeline, we conducted a high-throughput screening of 86,000 structures from the Cambridge Structural Database, identifying existing and future candidates with minimal toxicity profiles suitable for drug delivery applications. Beyond screening, the models provide insights into the chemical landscape of high-biocompatibility building blocks, enabling the derivation of guidelines for the rational, *de novo* design of biocompatible MOFs. Our framework thus enables the efficient, rapid navigation of the MOF chemical space to shortlist ideal candidates for experimental validation, potentially accelerating the timescales for clinical translation.

SUMMARY

Metal-organic frameworks (MOFs) are promising for drug delivery due to their high drug-loading capacity, tunable porosity, and structural diversity. However, their clinical translation is hindered by concerns over biocompatibility. Unfortunately, experiments are resource and time intensive, while modeling approaches fail to capture the behavior of MOFs in intricate biological systems. Herein, we report a machine learning (ML)-guided computational pipeline for probing MOF biocompatibility. Using a database of over 35,000 organic molecules, our interpretable ML models predict the toxicity of MOF linkers with over 80% accuracy across different administration routes. Furthermore, we cataloged the toxicity of MOF metallic centers and screened 86,000 MOFs from the Cambridge Structural Database, identifying candidates with minimal toxicity profiles. Beyond screening, our models provide insights into chemical features of biocompatible MOFs—enabling *de novo* rational design. This framework expedites the discovery of safer MOFs for drug delivery and deepens the understanding of their underlying chemistry.

INTRODUCTION

For many diseases, a promising prognosis is often linked to the ability to deliver therapeutic agents to specific sites in the body. In chemotherapy, this is often challenging, because cancer drugs are distributed throughout the body, provoking damage to healthy cells and, therefore, side effects.[1] Moreover, some cancer drugs are hydrophobic and require formulation to be used in the clinic, while macromolecules such as RNA—potentially the most effective cancer therapeutics that exist—require a vehicle to avoid *in vivo* degradation.[2,3] Without a doubt, drug delivery—i.e., bringing an agent from outside a biological system to a specific site within the body—has rightfully been identified as a critical area of research and is indeed the bottleneck for the translation of genetic therapies to the clinic. In this context, the engineering of nanomaterials has emerged as an interesting avenue for improved disease diagnosis and treatment.[4] The encapsulation of therapeutic agents in nanoparticles has



Scheme 1. MOFs as emerging drug delivery systems and computational pipeline workflow for assessment of MOF biocompatibility

(i) (a) The versatility of MOFs as drug delivery systems and the advantages they offer over their counterparts (API stands for 'active pharmaceutical ingredient'). (b) External surface modifications of MOFs improve their colloidal stability, biodistribution, and overall biocompatibility. (c) Kinetic lability of the metal-ligand bond and *in vivo* degradation of MOFs facilitate controlled release.

(legend continued on next page)

improved their stability and solubility, augmented their ability to penetrate membranes, and increased their circulation times, enhancing treatment efficacies.[5] Despite these superior capabilities, massive investment, and decades of research, the number of clinically approved nanomedicines is below projections, partly attributed to a gap in outcomes between pre-clinical data (*in vitro* and *in vivo*) and human studies.[5,6] These poor therapeutic outcomes are associated with complex biological barriers that serve to limit the bioavailability of therapeutic agents at diseased sites.[4–6] For instance, initial iterations of nanomaterials engineered for drug delivery had poor cell membrane permeability and struggled with lysosomal entrapment—which in turn affected intracellular functioning.[7] However, an improved understanding of the interactions of nanoparticles with biological systems[8] and clever advances in engineering have seen several nanoparticle formulations being approved for clinical usage.[5] In addition to small-molecule chemotherapeutics, state-of-the-art drug delivery formulations have more recently demonstrated efficacy for the delivery of macromolecules such as proteins and DNA.[7] The current classes of drug delivery formulations can broadly be classified as lipid based, polymer based, and inorganic nanoparticle based. Despite their advantages, however, these formulations either have poor targeting capabilities or face solubility, toxicity, and aggregation concerns.[5,9–12]

In view of these constraints, and owing to their inorganic-organic crystalline nature, metal-organic frameworks (MOFs) have emerged as a new system for drug delivery.[13–15] Composed of metal-containing clusters or metal ions connected via organic linker molecules, MOFs have well-defined, tunable structures with high void space and permanent porosities. While these unique characteristics allow MOFs to be utilized in a variety of applications such as gas storage and separation,[16] sensing,[17] and catalysis,[18] with regard to drug delivery, these materials have significant advantages over currently utilized formulations. The biggest advantage is that the high void space allows the encapsulation of up to 50 times higher quantities of therapeutic agents per unit mass (up to 1.4 g drug per gram MOF), which are released over long timescales (Scheme 1i, a).[19] The open structure and kinetic lability of the coordination bonds of MOFs imply, many times, relatively lower stability compared to conventional porous materials, such as zeolites and carbon nanotubes, but, in drug delivery, they are beneficial, as they will avoid bioaccumulation by facilitating a controlled degradation *in vivo* (Scheme 1i, c). In addition, the unique chemistries of the organic and inorganic components can be leveraged to incorporate multiple functionalities onto the MOF's internal or external surface.[20] An avenue of particular interest is the functionalization of the external MOF surface post-synthesis—sometimes referred to as biofunctionalization[21]—which improves colloidal stability, extends release times, and enhances cellular uptake, in an attempt to achieve targeted drug delivery (Scheme 1i, b).[22] Moreover, these surface-modified MOFs show improved intracellular stability, lower cytotoxicity *in vitro,* and delayed release times.[23] Surface-modified MOFs have shown efficacy in the delivery of small-molecule drugs[24] and monoclonal antibody checkpoint inhibitors in cancer immunotherapy[25] and, more recently, in photodynamic therapy.[26] Despite advancements over the past decade, there are some gaps that are yet to be filled before MOFs can be considered at a clinical level.

Perhaps the most fundamental of these gaps is the inherent biocompatibility of the MOF. Pre-clinical information on proposed materials—in addition to efficacy studies—requires stability, biodistribution, and biocompatibility data.[27] The screening is usually performed in a pipeline, starting with *in vitro* tests (e.g., examination of cytotoxic effects and performance on human cells), followed by animal *in vivo* studies (e.g., distribution, efficacy, and immune response in several animal models); for clinical translation, this will culminate in several stages of clinical trials.[8,27] Most studies pertaining to MOFs are limited to an *in vitro* stage, with very few *in vivo* studies reported to date. The first study was reported by Horcajada et al.[13] in 2010, where, upon intravenous (i.v.) administration of Fe-MOFs to Wistar rats, no immune or inflammatory reactions were detected, hinting that these MOFs are well tolerated. A more in-depth history of *in vivo* studies on MOFs has been presented in an excellent review by Ettinger et al.[28] While these results suggest the applicability of MOFs for drug delivery, the translational gap between animal and human outcomes—with statistics indicating that only one in nine human trials are approved for clinical usage[6]—implies that these studies need to be scaled up or accelerated.

From a theoretical standpoint, with near-infinite possible combinations of metal nodes and organic linkers, there have been over 500,000 hierarchical MOF structures predicted, with more than 100,000 experimentally synthesized.[27] This makes optimum material selection through experimental means intractable due to the associated timescales, costs, and, more importantly, ethical concerns with respect to large-scale animal studies. While chemical intuition and empirical efforts have captured chemical trends influencing MOF biocompatibility, these rules are difficult to generalize for the efficient screening of biocompatible MOF candidates. With these constraints in mind, here, we present a systematic analysis of factors influencing MOF toxicity and attempt to guide the rational design of biocompatible MOFs for drug delivery applications using interpretable machine learning (ML). We would like to note that, in a recent study, a large-scale screening of the literature was conducted to curate research articles pertaining to cytotoxicity of MOFs, wherein cell viability data were extracted and used to train predictive models.[29] However, given the scarcity of comprehensive cell viability data, especially negative results (i.e., reports on toxic MOFs), the generalization capabilities of these models remain

(ii) (a) Data are first processed using conventional techniques and featurized using cheminformatics libraries to accurately capture chemical information. Following this, feature selection is carried out to find the optimum subset of features that can maximize model performance. (b) These subsets are then used to train different models—namely random forests, support vector machines, and gradient boosted machines—on a coarse grid of hyperparameters evaluated across multiple metrics. The best-performing model is further optimized on a fine grid of hyperparameters. (c) The final model is first validated on unseen test data and then used to screen the CSD for potentially biocompatible MOF candidates for drug delivery.
Parts of the scheme were created with BioRender.

unexplored. Herein, inspired by the drug discovery process,[30] we propose a computational pipeline (outlined in Scheme 1ii) capable of screening large MOF libraries for their inherent biocompatibility in a high-throughput manner. Our methodology utilizes the distinctive chemistries of MOF building blocks to guide inferences surrounding prospective biocompatibility. The *in vivo* degradation of the MOF due to the kinetic lability of the metal-ligand bond implies that the toxicity of building blocks would be strongly correlated to its toxicity.[28] Our central premise is that MOFs built from non-toxic precursors or building blocks are *more likely* to be non-toxic than those built on toxic ones. This approach allows one to narrow down the chemical landscape for future experimental navigation, shortlisting potentially optimal materials. Similar approaches leveraging MOF precursor chemistry have been reported by Pétuya et al.,[31] for predicting guest-pore accessibility, and by Batra et al.,[32] for predicting water stability. As elucidated by Bencherif and colleagues,[27] the chemistry of the precursors emerges as a pivotal determinant of MOF toxicity. For the metallic center, correlations of intrinsic chemical features with the toxicity of the MOF are straightforward, while for the organic linker, these correlations are convoluted, owing to conflicting effects of different chemical features. To address the toxicity concerns of the metal center, we curated a database of reported median lethal dosage (LD_{50}) values of corresponding ions in oxidation states commonly found in MOFs, maintaining respective chloride salts as an ionic control. For the organic linker, we developed predictive algorithms (classification models) using a database of LD_{50} values of over 35,000 organic molecules[33] when administered using the intraperitoneal (i.p.) route. For each molecule in the database, a vector of chemical features was calculated, using which a three-class classification model was developed, capable of classifying linkers as "safe," "toxic," or "fatal." In the event the molecule was predicted to be toxic, it was passed to a second two-class classification model for predicting the degree of toxicity. Three unique predictive algorithms were developed, optimized, and evaluated against several metrics. The best-performing model showed an 81% accuracy (receiver operator characteristics-area under the curve [ROC-AUC] score of 0.86) against previously unseen data and showed good generalization capabilities when tested on MOF linkers with known toxicity profiles.

The developed models are interpretable at both global and local levels. At a global level, the judicious selection of ML models facilitated the extraction of chemical features exhibiting robust correlations with toxicity, thereby facilitating the formulation of a general framework to guide the design of MOFs with the heightened prospects of biocompatibility. At a local level, Shapley additive explanations (SHAPs)[34] revealed features that had high positive and negative contributions toward the classification of specific molecules. Using these computational tools, we screened the Cambridge Structural Database (CSD),[35] containing over 86,000 non-disordered MOF structures, hierarchically, initially delineated by the metallic center, subsequently refining the selection based on the organic linker, thereby identifying potentially safe candidates that may be considered for drug delivery applications. The pipeline developed here can be easily transferred to other, similar systems to accelerate de-risking strategies of new materials.

RESULTS

Data processing and featurization

Toxicity data for the organic ligands were obtained from a recently published database containing comprehensive toxicological information spanning over 113,000 organic molecules across 1,474 toxicity endpoints.[33] For predicting the potential toxicity of organic linkers, we chose the largest subsets of the database, reporting quantitative results on the LD_{50} when administered to mice through i.p. or oral routes. We recognize that the route of administration and localization is highly correlated to the observed toxicity,[36] and our choice of data was driven by the need to maximize model performance. The chosen subsets contained quantitative LD_{50} values for 35,299 organic compounds (for the i.p. route) and 22,691 organic compounds (for the oral route) represented in the simplified molecular-input line-entry system (SMILES) format.[37] We performed additional steps to clean the data, by removing missing or duplicate entries. To comprehend the dataset, we used the Globally Harmonized System of Classification and Labelling of Chemicals (GHS) proposed by the United Nations.[38] The GHS classifies substances into five acute toxicity hazard categories based on LD_{50} values when administered through either oral or dermal route (discussed in section S2 of the supplemental information). In the absence of categories for the i.p. route, we used the same categorization as the oral one. Compounds belonging to categories 1 and 2 are usually fatal, categories 3 and 4 are usually toxic, and category 5 are usually safe.[38] Considering the organic linker in MOFs, we would not want to use potentially fatal linkers and would ideally want to use potentially safe linkers. However, as will be discussed in upcoming discussions, most linkers are toxic. In view of this, we propose three customs categories, (1) safe, (2) toxic, and (3) fatal, formed by combining relevant categories as developed in the GHS. Upon curation and processing, we deployed strategies to represent the chemical data in a machine-readable format. A robust and reliable technique of representation is to calculate a list of molecular features that are relevant to the problem at hand—in our case, toxicity.[39] To build accurate models, it is essential to select features that can collectively capture trends in toxicity across different classes of organic linkers.[32] For featurization, we used an open-source library[40] for generating 197 cheminformatics-based descriptors at different hierarchical levels (Figure 1A). These were a combination of physical molecular descriptors—such as molecular weight, fragment-based descriptors—such as the number of carbonyl groups present, and more complex descriptors based on polarity, partial charges, and surface areas. Section S3 of the supplemental information shows the full list of descriptors calculated, with brief descriptions of their physical implications. The choice of descriptors was motivated by the broader goal of capturing experimentally implementable chemical features that are correlated to toxicity. In this regard, descriptors that can encode information about the entire atomic structure—known as "global descriptors"—are more useful than descriptors representing localized regions within a structure—known as "local descriptors."[41] The inclusion of descriptors such as those related to molar refractivity, van der Waals surface areas, and partial charges may seem peculiar at first. These descriptors, however,

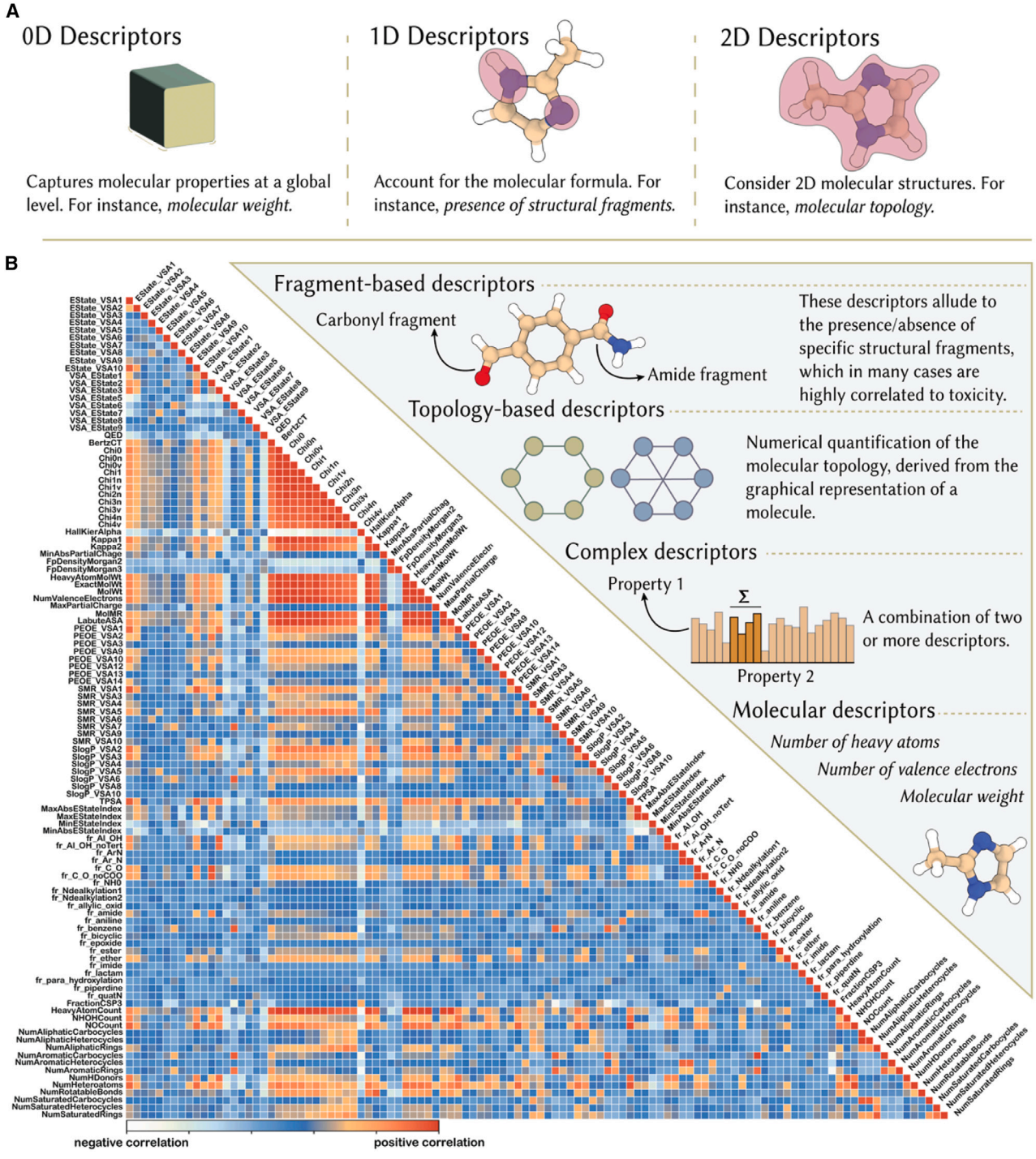
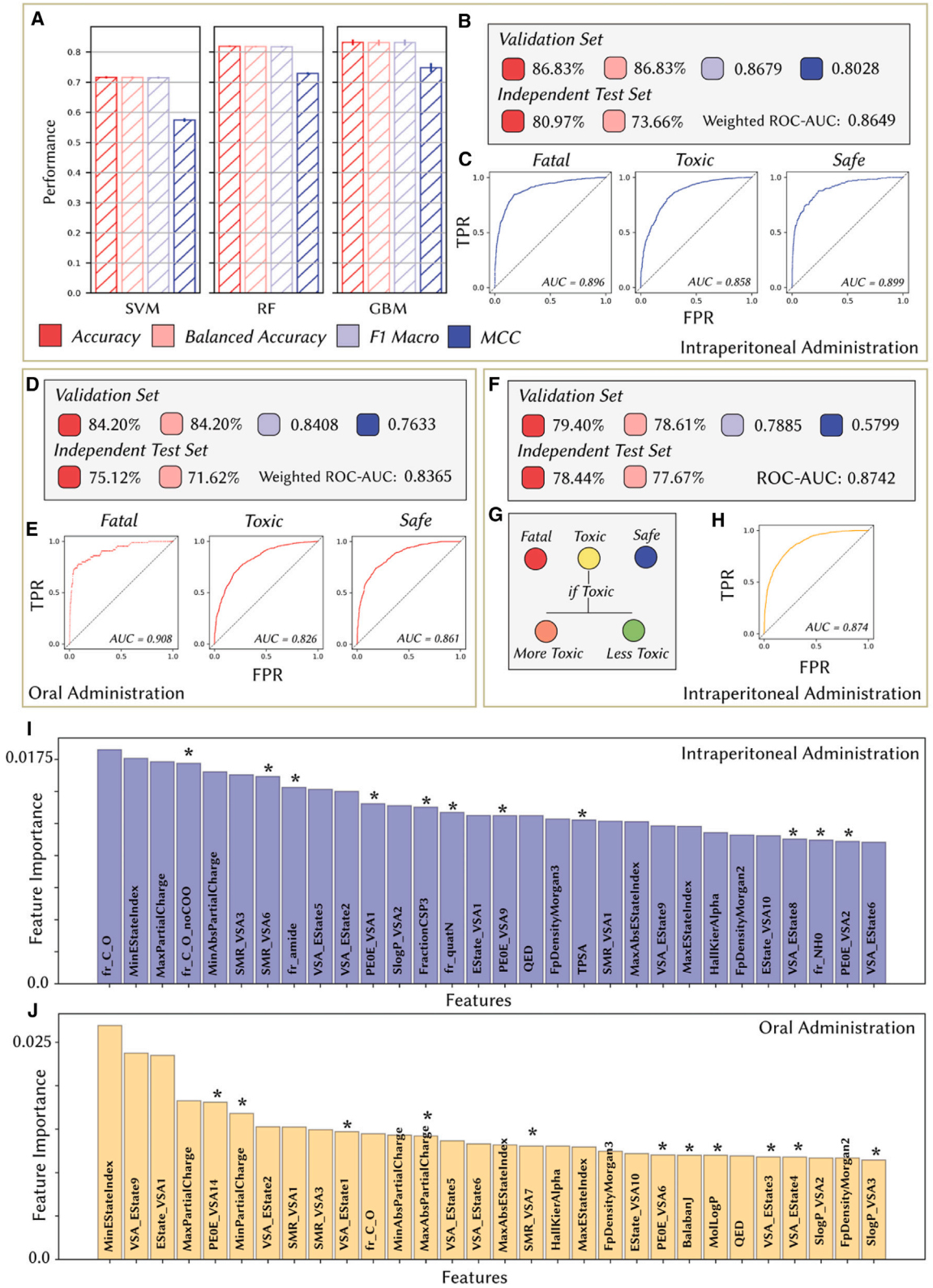


Figure 1. Molecule featurization

(A) Three types of global descriptors were used for the development of the models. The first are 0D descriptors that do not account for molecule structure or atom connectivity, such as molecular weights. The second are 1D descriptors that account only for the molecular formula, such as the presence or absence of specific fragments. Last, 2D descriptors consider 2D molecular structures, such as surface-area-based descriptors. There are other ways of categorizing these descriptors, such as topology based or electronic-state based. We also calculated complex ("abstract") descriptors such as the quantitative estimate of drug-likeness (QED). Inspired by Aloy and colleagues.[44]

(B) A correlation matrix of 110 features. The correlation matrix helps visualize relationships between different features. The color suggests relations between features ranging from strong negative correlations to strong positive correlations.



(legend on next page)

are widely used in computational chemistry, as they introduce orthogonality, i.e., they capture different aspects of the molecule that traditional descriptors cannot—allowing us to train models on uncorrelated data, which is more desirable. In addition, not only does the introduction of orthogonality minimize potential biases that may arise during subsequent feature selection processes but these descriptors also allow us to discriminate between molecules that have small structural differences.[42] While these correlations are complex (Figure 1B)—and not necessarily experimentally accessible—these descriptors help rationalize the ability of molecules to bind to receptors and cross biological membranes.[42,43] Figure S2 shows the variation in these features across different categories of toxicity. From these distributions, it is evident that not all calculated descriptors are correlated to toxicity. In such a scenario, it is important to retain only those features that are relevant to the target. Here, we relied on the simple, effective technique of selecting the "K" ($K \in \{20, 30, \ldots, 180\}$) features most correlated to toxicity, discussed in the following sections—and further in section S5 of the supplemental information.

Figure S1 shows the distribution of data points for the i.p. route across the GHS categories. The distribution suggests that the data are heavily biased, with 3,737 fatal, 29,678 toxic, and only 1,704 safe molecules. Models trained on these biased data would be more likely to predict molecules as toxic, which, while undesirable for incorrectly labeled safe molecules, is a serious concern for incorrectly labeled fatal molecules.[45] To avoid biasing the model predictions, following a train-test split, we chose to equalize the number of datapoints in each category (see section S4 of the supplemental information). There are three ways of achieving this: (1) oversampling the minority category, (2) undersampling the majority category, and (3) a combination of oversampling and undersampling. The first two strategies carry the risk of representation bias—the generated samples are non-representative of the entire population—and aggregation bias—generated samples falsely represent the entire population by ignoring sub-groups of data.[45] Intuitively, a combination of the two strategies should produce the best results.[46] Here, we carried out random undersampling of the majority class to avoid sampling bias and used the adaptive synthetic (ADASYN)[47] sampling algorithm to oversample the minority classes (refer to section S1 of the supplemental information for further details). These strategies produced classes with 5,000 datapoints each. To prevent information leakage, the independent test set was not sampled and was thus heavily imbalanced. To account for this imbalance, appropriate metrics, such as balanced accuracy, were used for evaluating the generalization capabilities of the developed models.

ML model development and optimization

Once data were processed, curated, and featurized, the next step involved the selection of appropriate ML models to be trained. Here, due to their excellent performance on materials data, relative simplicity, and interpretability, we limited our search to support vector machines (SVMs), random forests (RFs), and gradient boosting machines (GBMs),[31,32,48] discussed in sections S1 and S6 of the supplemental information. SVM-, RF-, and GBM-based classifiers were trained on identical datasets, their performances being evaluated across five metrics: (1) accuracy, (2) balanced accuracy, (3) macro-F1 score, (4) Mattheus correlation coefficient (MCC), and (5) ROC-AUC score—discussed in section S1 of the supplemental information. In case it is not clear as to why more than one evaluation metric has been employed, it is because, in several cases, in addition to accuracy, we are concerned about where the model tends to fail. For instance, in the case of organic linker toxicity, apart from the accuracy of the model, we would also want to minimize potentially fatal linkers being falsely classified as safe. In such scenarios, metrics such as macro-F1 tend to shed more insights. Moreover, given the imbalanced nature of real-world data, metrics such as balanced accuracy provide a more reliable assessment of the generalization capabilities of the model.

As discussed previously, to optimize model performance, it is crucial to identify a non-redundant subset of the feature space that is correlated to the target variable, so as to reduce overfitting while improving model interpretability. Here, we utilized univariate statistical tests based on the ANOVA F value[49] to select the K-best features, $K \in \{20, 30, \ldots, 180\}$. The three models implemented on the scikit-learn library[50] in Python were trained on all subsets to find the most optimum, unbiased set of features relative to the model. Using a single 5-fold stratified cross-validation approach, the SVM showed the worst performance across all metrics, with an accuracy of 71.62% (Figure 2A). As described in Scheme 1, this model was trained on a coarse grid of hyperparameters pertaining to the regularization parameter, kernel type, and shape of the decision function. The GBM was optimized on a coarse grid of hyperparameters pertaining to the learning rate, maximum iterations, maximum leaf nodes, minimum samples required to be a leaf node, and regularization parameter; it showed the best performance on the training set with an accuracy of 83.22% (Figure 2A). In turn, the RF was optimized on a coarse grid of hyperparameters pertaining to the number of estimators and maximum depth (refer to section S6 of the supplemental information); it showed performance comparable to that of GBM, with an accuracy of 81.91%. Due to better performance on independent test data and overall better interpretability, we chose the RF-based classifier for subsequent

Figure 2. ML model evaluation

(A) Performance of SVM-, RF-, and GBM-based three-class classification models across four evaluation metrics on the i.p. dataset.

(B and D) The performance of the hyperparameter-optimized RF-based three-class classifier on the training and independent test sets for i.p. and oral routes of administration.

(C and E) ROC-AUC of the three-class classifier on the independent test sets for i.p. and oral routes of administration.

(F and H) The performance of the two-class classifier on the training and independent test sets for i.p. administration.

(G) Sequential classification scheme developed.

(I and J) Ranked feature importance (top 30 features) for the best-performing RF classifiers for the i.p. and oral routes of administration, respectively. The unique features have been marked with stars.

investigations. The RF-based classifier showed the best performance across all metrics on a subset of 110 features selected using the F-value test. The general trend observed was an increase in performance from 20 to 110 features, followed by a decrease in performance up to 170 features, with a slight improvement at 180 features (Figure S6).

The hyperparameter optimization process suggests that at maximum depth >20, with number of estimators >400, the model generally performed well, showing the best performance at a maximum depth of 40 with 700 estimators. Following this, pursuant to Scheme 1, we constructed a finer grid of hyperparameters around those reported above, to fine-tune the model performance. Fine grid optimization suggests that the model performs best at a maximum depth of 32, with 725 estimators, although these improvements were marginal at best. The best-performing classifier had a validation accuracy of 86.83% (macro-F1 of 0.87, MCC of 0.80) on the i.p. dataset and a validation accuracy of 84.20% (macro-F1 of 0.84, MCC of 0.76) on the oral dataset (Figures 2B and 2D). As expected, due to the imbalanced nature of the independent test set (which is what we would expect for real-world data), the performance metrics were slightly lower; however, the models largely retained their generalization capabilities. As pointed out earlier, apart from accuracy, a major concern for toxicity-related models is correctly identifying molecules that are fatal (specificity) and correctly identifying molecules that are not safe (sensitivity).[51] To investigate this, we calculated ROC curves[52]—discussed in section S7 of the supplemental information. In the ROC curve, the y axis represents the true positive rate (TPR; analogous to sensitivity), while the x axis represents the false positive rate (FPR; analogous to 1 − specificity). A perfect model would have a TPR of 1 and FPR of 0, while a random guess would follow the curve denoted by the dashed lines in Figure 2C. In simple terms, the ROC curves help analyze the trade-off between the TPR and the FPR across different classification thresholds. The AUC is often used as a metric to compare ROC curves. Figure 2C suggests that for the independent i.p. test dataset, in a one-vs.-rest setting, the RF-based classifier performs well across classes, showing the best performance for predicting potentially safe molecules correctly (AUC = 0.899) and a relatively poorer performance for predicting potentially toxic molecules correctly (AUC = 0.858). Figure 2D suggests that for the independent oral test dataset, the classifier again performs well across classes, with the best performance in predicting potentially fatal molecules correctly (AUC = 0.908). To assess model performance against external data, we curated a small dataset of 25 Food and Drug Administration (FDA)-approved drugs along with their reported LD_{50} values when administered to mice or rats orally. Our developed three-class classification model for oral administration correctly predicted the toxicity class of 18 of the 25 drugs, showing an accuracy of 72%, which is indicative of good generalization capabilities. Notably, none of the drugs with a fatal profile were incorrectly identified as safe. In section S10 of the supplemental information and Data S4, we provide our curated dataset, along with the relevant data sources and corresponding model predictions.

Furthermore, given that a majority of organic linkers may fall in the toxic category, we included a more robust classification of the "degree" of toxicity to further guide materials selection and design choices. Here, pertaining to categories 3 and 4 of the UN GHS, a two-class classification model was developed for discriminating between molecules that are "more" or "less" toxic (Figure 2G). Figure 2F shows the performance of the classifier on a subset of the i.p. dataset pertaining to the relevant categories. The model performance in this case is slightly poorer compared to the three-class classification model, hinting that it may be more difficult to discriminate between "more" and "less" toxic molecules. Nonetheless, the model had good generalization capabilities against previously unseen data (Figures 2F and 2H). A crucial outcome from the developed models in the present context is the global feature importance rankings with respect to the toxicity endpoints as shown in Figures 2I and 2J. These feature importances would play a key role in guiding linker design choices—e.g., what characteristics should be avoided to design linkers with better toxicity profiles—and are discussed later. As mentioned previously, the induced toxicity is strongly correlated to the route of administration. On comparing the 30 features most correlated to the toxicity endpoint across the i.p. and oral routes of administration, it becomes apparent that not only do several features differ in their relative importance but there also exist features that are unique to the route of administration. For instance, in the case of the i.p. route, we observed that among these 30 features, there were more fragment-based descriptors compared to the oral route. There were also several features such as those pertaining to partial charges, molar refractivity, and metrics like the "quantitative estimate of drug-likeness" (QED) that were prominent across the routes of administration. We would, however, like to note that the relative feature importance is sensitive to correlations between features.

High-throughput screening of MOFs for biocompatibility

The ML model for predicting the toxicity of the organic linker partly addresses the biocompatibility of MOFs. The second aspect—as important as the first one—is the toxicity of the metallic center. The chemical landscape of MOF metal centers is limited, making it easier to map and interpret. However, some crucial considerations should be accounted for prior to making design choices. Here, we would like readers to refer to Figure 3A, where we present a dataset of metal center toxicity, with the corresponding chloride salt as the ionic control. This dataset was constructed for all metal centers having formed MOF structures as reported in a recent review by Zhou and coworkers[53]—with a description of the data curation in section S1 of the supplemental information. For better interpretability, we provide the raw data in section S8 of the supplemental information and Data S1, along with the data sources. In many cases, the mechanisms by which metal ions induce toxicity are well known. Disruption of the metal ion balance in the body (also known as metal ion homeostasis) can lead to metal binding to protein sites outside of their natural binding sites, causing the oxidative deterioration of biological macromolecules—which as such can lead to several diseases.[54] For instance, disruption of Fe and Cu homeostasis has been linked to neurological disorders such as Parkinson's.[55] This disruption can also lead to the formation of free radicals, which in turn can modify DNA bases

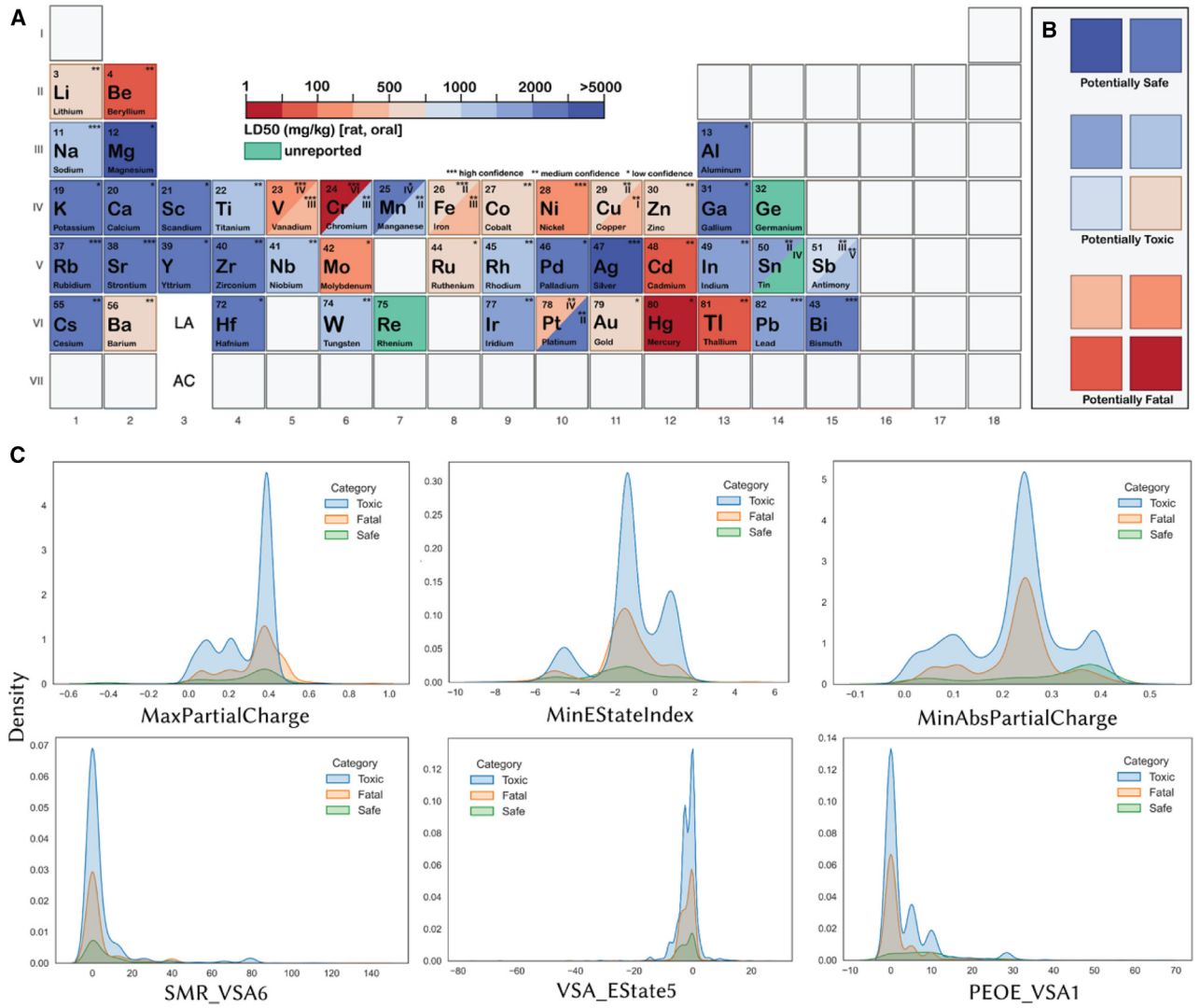


Figure 3. High-throughput screening

(A) Database for the reported LD_{50} values for the corresponding chloride salts taken as the ionic control when administered through the oral route to rats. The elements are color coded according to their toxicity. All possible centers with MOFs reported have been considered for the construction of this database—refer to section S8 of the supplemental information for the raw data.

(B) Based on toxicity values, centers can be classified as potentially safe, toxic, or fatal. For the high-throughput screening, only potentially safe centers were considered.

(C) Distribution of some of the top-ranked features, MaxPartialCharge (maximum partial charge of the molecule), MinEStateIndex (an electrotopological state index), MinAbsPartialCharge (minimum absolute partial charge of the molecule), SMR_VSA6 (a complex descriptor calculated from the molar refractivity and the atomic contributions to the van der Waals surface area), VSA_Estate5 (a complex descriptor calculated from the van der Waals surface area and the atomic contributions to the electrotopological state), and PEOE_VSA1 (a complex descriptor calculated from the partial charges and the atomic contributions to the van der Waals surface area), across the three classes of toxicity as predicted by the ML model for MOFs with potentially safe metal centers.

and enhance lipid peroxidation.[54,56] Indeed, not only the choice of the metal center but also its connectivity in the cluster would play key roles in modulating the safety profile of the resulting structure.

There are several inferences that can be drawn from Figure 3A. The more obvious primary inference is that the oxidation state of the metallic center largely influences toxicity.[57] This is most evident in the case of Cr, with Cr(VI) being fatal, while Cr(III) is borderline toxic. This effect is less pronounced in the case of Mn, Fe, Cu, and Sb. A second inference is that, notwithstanding physicochemical properties, several MOFs currently being explored at an *in vivo* level have potentially safer metal center alternatives. Most current research on MOFs for drug delivery predominantly focuses on materials centered around Fe or Zn, which have demonstrated low toxicity in the short term.[58,59] However, there is a notable gap in investigations regarding the long-term effects and comprehensive animal studies for these materials. In this context, the utilization of metallic centers that

are potentially safer offers a degree of risk mitigation when translating these studies to clinical applications, from a safety-by-design (SbD) perspective. For instance, going by these considerations, Ag-/Ca-centered analogs of Fe-/Zn-centered MOFs may be potentially safer candidates. Nevertheless, this discussion is not straightforward, as several other considerations come into play—for example, the metal coordination, topology, and pore properties of the MOF are as important as the biocompatibility. Arguably, one of the most critical considerations revolves around the stability of MOFs within biological systems. This aspect remains relatively unexplored, especially concerning MOFs with metallic centers that we classified as potentially safe. In such scenarios, however, external surface modification or functionalization emerges as a viable strategy for enhancing material stability.[21] For instance, we have demonstrated that the post-synthetic modification of the MOF external surface with phosphate-functionalized methoxy polyethylene glycol enhances the MOF colloidal stability and provides precise control over drug-release kinetics. This approach effectively overcomes a long-standing issue related to MOF aggregation, which has been a significant bottleneck for the translation of MOFs to the clinic.[23] These strategies can be extended to MOFs containing these potentially safer metallic centers, thereby improving their suitability for drug delivery applications. A second consideration that we have touched upon briefly is the variation in toxicity with the route of administration.

Like the categories of toxicity used in the ML model development, we classified metal centers as potentially safe, toxic, or fatal (Figure 3B). Ag, Mg, Al, K, Ca, Sc, Ga, Rb, Sr, Y, Zr, Pd, Cs, Hf, and Bi have been identified as potentially safe metal centers. With these metallic centers we screened the non-disordered MOF subset—containing over 86,000 entries—of the CSD.[35] We appreciate that, in many cases, due to exceptional properties, researchers are persuaded to use MOFs with metallic centers outside the list of metals considered here. However, our goal here is to find the safest possible MOFs. In our initial screening, we compiled the crystallographic information files (CIFs) of MOFs containing these metals of interest—which were then screened in a high-throughput manner, described in Scheme 1, and are discussed further in section S1 of the supplemental information. Here, it was necessary to deconstruct the MOF structures to find their respective building blocks. For this, we leveraged the *moffragmentor* library developed by Jablonka et al.,[60] which uses a structure graph analysis to identify the building blocks. We passed the compiled list of MOF structures through this library in a high-throughput manner to extract the organic linkers of only those MOFs that have metal centers of interest in the present context, discarding the rest. For example, Ag-centered or Zr-centered MOFs containing Pd in their structure were excluded from consideration. Also, in some cases where the organic linkers used are highly complex, we were not able to successfully fragment the structure. The fragmentation procedure gave us 5,178 MOF linkers.

The organic linkers corresponding to these 5,178 MOFs were then passed to the trained RF-based classifier for the i.p. route to predict their potential toxicity. Four hundred seventy-nine linkers were classified as potentially safe, 3,189 as potentially toxic, and 1,384 as potentially fatal; for approximately 120 linkers, featurization was not successful—possibly due to complexities during the fragmentation procedure. Since these linkers accounted for ca. 2% of all linkers, they were thus ignored from consideration—and we anticipate that this low percentage does not affect the subsequent analysis performed. Furthermore, of the toxic linkers, 2,479 were assigned as "less" toxic, while the remaining were assigned as "more" toxic. In section S9 of the supplemental information, Data S2, and Data S3, we provide the lists of MOFs assigned safe and "less" toxic. Figure 3C shows the distribution of six continuous features (MaxPartialCharge, MinEStateIndex, MinAbsPartialCharge, SMR_VSA6, VSA_Estate5, and PEOE_VSA1) having high correlations to the observed toxicity (for the i.p. route) as identified in Figure 2I. It is evident that these distributions overlap considerably—with rather subtle differences between individual features. It thus becomes difficult to infer useful guiding principles on linker design on the basis of these distributions, relying purely on experimental approaches. This brings us back to our initial point that, although there are chemical trends that guide biocompatibility, these are difficult to generalize. It is hypothesized that these subtle differences add up due to synergistic interactions between several features, resulting in large apparent differences in observed toxicity profiles—something similar to what was proposed by the Lipinski's "rule of five" (Ro5).[61] While these trends may not be necessarily intuitive, they nonetheless suggest regions in the chemical space that are theoretically safe for the design of biocompatible linkers. More recent computational approaches, such as generative models, are particularly good at modeling these distributions by capitalizing on subtle differences between distributions for inverse design.[62,63]

Model interpretation and validation

Throughout our discussions, a recurring theme has been the importance of model interpretability. By analyzing the feature importance, such as those shown in Figures 2I and 2J, we can make several inferences pertaining to chemical trends at a global level. For instance, based on feature importance, the number of carbonyl oxygen fragments (fr_C_O) appears to be highly correlated to the toxicity profile. Among the MOF linkers that were assessed, safe linkers had an average of 1.44, toxic linkers had 1.34, while fatal linkers had 1.11 carbonyl oxygens. Such descriptors that can effectively discriminate between fatal and non-fatal linkers are very useful from a design perspective. Similarly with regard to amide fragments (fr_amide), safe and toxic linkers tend to have a higher number of amide fragments compared to fatal linkers (an average of 0.20 for safe, 0.19 for toxic, and 0.13 for fatal linkers). In such cases, however, it is crucial to not confuse the absence of evidence with evidence of absence. As is evident from these values, in several cases, trends may not be linear. For instance, as can be observed in Figure 3C, fatal molecules may have properties that fall between those of safe and toxic molecules. While fragment-based and general descriptors such as molecular weight are easier to interpret and can directly guide linker design, some descriptors are more "abstract" but nonetheless crucial to consider. These abstract descriptors build off of the seminal Lipinski Ro5,[61] wherein it was proposed that highly desirable drug-like compounds have properties that fall within specific ranges. For instance, compounds having

hydrogen-bond donors >5, an octanol-water partition coefficient >5, a molecular mass >500 Da, and hydrogen-bond acceptors >10 are more likely to have poor permeability and are, thus, less desirable. One such abstract descriptor is the QED developed by Bickerton et al.[64] for evaluating the desirability of drug-like compounds through underlying properties such as solubility, permeability, metabolic stability, and transporter effects. The QED is seen to vary for the different classes of molecules both in the original dataset and for the predictions. Notably, it was observed that molecules with the highest QED values belong predominantly to the toxic category (with an average value of 0.61 compared to safe and fatal molecules having average values of 0.45 and 0.45, respectively). Thus, relying on the QED value appears to be a viable strategy from a biocompatibility perspective—as again, it on average can discriminate between toxic and fatal molecules. Similarly, it has been proposed that increasing the saturation, i.e., the number of sp^3 carbon atoms in a molecule, and reducing the aromatic ring count increases the chances of clinical success.[65,66] It should, however, be noted that there are several exceptions to these "rules," with estimates suggesting that only 51% of FDA-approved small-molecule drugs comply with the Ro5.[67]

Since global feature importance captures feature correlations across the dataset, for individual molecules, they may deviate significantly. It is due to these deviations that it becomes necessary for models to be locally interpretable. In such cases, we may rely on local explanations from the associated SHAP analysis. SHAP is a game-theory-based approach to explaining the output of ML models by decomposing this output into the sum of the impact of individual features, i.e., looking at the marginal contribution of each feature to the prediction.[34] Features with positive SHAP values increase the probability of a particular classification, while those with negative values decrease the probability. The magnitude of the SHAP value highlights the relative importance (or marginal contribution) of a feature toward a particular classification (which may differ from the global feature rankings). As depicted in Figure 4, from SHAP analysis, it can be inferred that the toxicity classification of the 2-methylimidazole (mIm) linker in ZIF-8 has large positive contributions from its electrotopological state (refer to Figure S7), molar refractivity, and absence of fragments such as carbonyl oxygens and aromatic heterocycles. Possibly due to structural similarities, the linker in ZIF-90, too, has contributions to its toxicity classification from similar features (EState_VSA1 and SMR_VSA3). Notably, in such cases, relatively simple functionalizations, such as the addition of an amide functionality, significantly change the toxicity profile. To aid the interpretation of the force plots depicted in Figure 4, we refer the reader to a detailed schematic in Figure S8.

While one aspect of biocompatibility is designing safe MOFs by observing trends in the chemical features of building blocks, another aspect is choosing the safest candidate given a choice. Here, guided by the overarching principles of SbD,[68] we explored the chemical landscape of linker characteristics to make rational decisions for the choice of optimum MOFs. To visualize this landscape, we utilized the t-distributed stochastic neighborhood embedding (t-SNE)[69] approach to project the 110 features onto a two-dimensional map. We chose to visualize the chemical landscape of experimentally reported Ag-MOF linkers (Figure 5A), Ca-MOF linkers (Figure 5B), and Zr-MOF linkers (Figure 5C) as they either represent the largest subsets of the experimentally reported MOFs with a safe metallic center or have been gaining traction as potential drug delivery systems.[15,23] Clusters represent linkers with high chemical similarity and are color coded based on their predicted toxicity. Within the chemical landscape enclosed by the linkers (Figures 5A–5C), we were able to identify "homogeneous" and "heterogeneous" design spaces. Here, we define homogeneous design spaces as clusters with linkers having the same toxicity class, while heterogeneous design spaces are clusters having linkers with different toxicity classes. Intuitively, operating within homogeneous design spaces containing only safe linkers would significantly enhance the chances of designing biocompatible MOFs. Heterogeneous design classes, however, pose SbD challenges. Caution is essential when exploring the chemical space pertaining to such clusters. These observations, however, are based on experimental data, which may not represent the complete chemical landscape.

A noteworthy observation is that a majority of the experimentally reported MOFs predicted to be safe lack porosity, as evident in Figures 5D–5F—hinting toward a biocompatibility-porosity trade-off. This trade-off potentially exists because higher molecular masses (which is often the case for longer linkers) are inversely correlated with molecule desirability.[64] Given the intended use in drug delivery, porosity is a crucial attribute. Among the safe MOFs that do exhibit porosity, a substantial portion are Zr based. This highlights the potential of Zr-MOFs as promising candidates for healthcare applications, primarily attributed to their high biocompatibility and superior pore characteristics. This finding aligns with our experimental validation in previous studies.[23] It is essential to note that the limited porosity observed in most safe MOFs reflects the available experimental data and may not hold true universally. In theory, by employing the principles outlined in this study, it should be feasible to design MOFs with these safe centers while ensuring adequate porosity. In this context, linker modifications may play a key role in modulating the safety profiles of MOFs. In Figure 5G, we demonstrate these modulations for a subset of substituted imidazolates (that form metal-azolate frameworks).[70] Here a knowledge of "toxicophores"—substructures or functional groups that are frequently associated with toxicity—would again help in design choices.[71] For instance, nitro groups and aromatic amines are well-known toxicophores, which have been identified by our classification models as highly toxic.[72,73] While the coordination modes will largely remain the same[70] during such modifications, one has to be mindful of the biocompatibility-porosity trade-off. In other words, while bulky modifications may modulate safety profiles, it would come at the cost of a lower porosity, hampering drug-loading capabilities. Achieving these goals would thus benefit significantly from the application of generative artificial intelligence. In section S9 of the supplemental information, Data S2, and Data S3, in addition to the list of safe and "less" toxic MOFs, we provide the associated pore properties as calculated using Zeo++.[74]

We tried to validate our models against experimentally reported data. Due to the reduced incentive to publish negative results, most of the literature focuses on MOFs that perform well

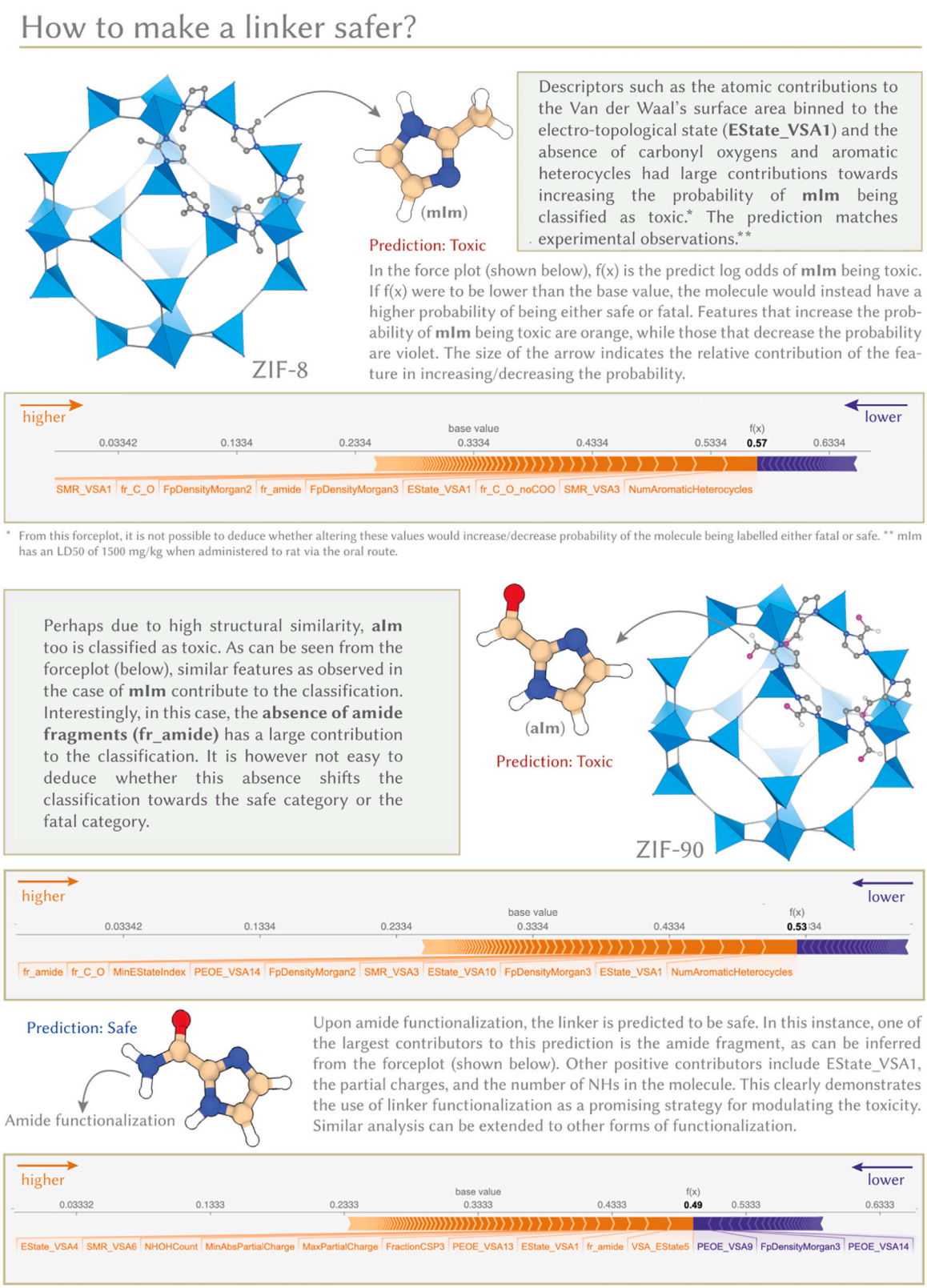
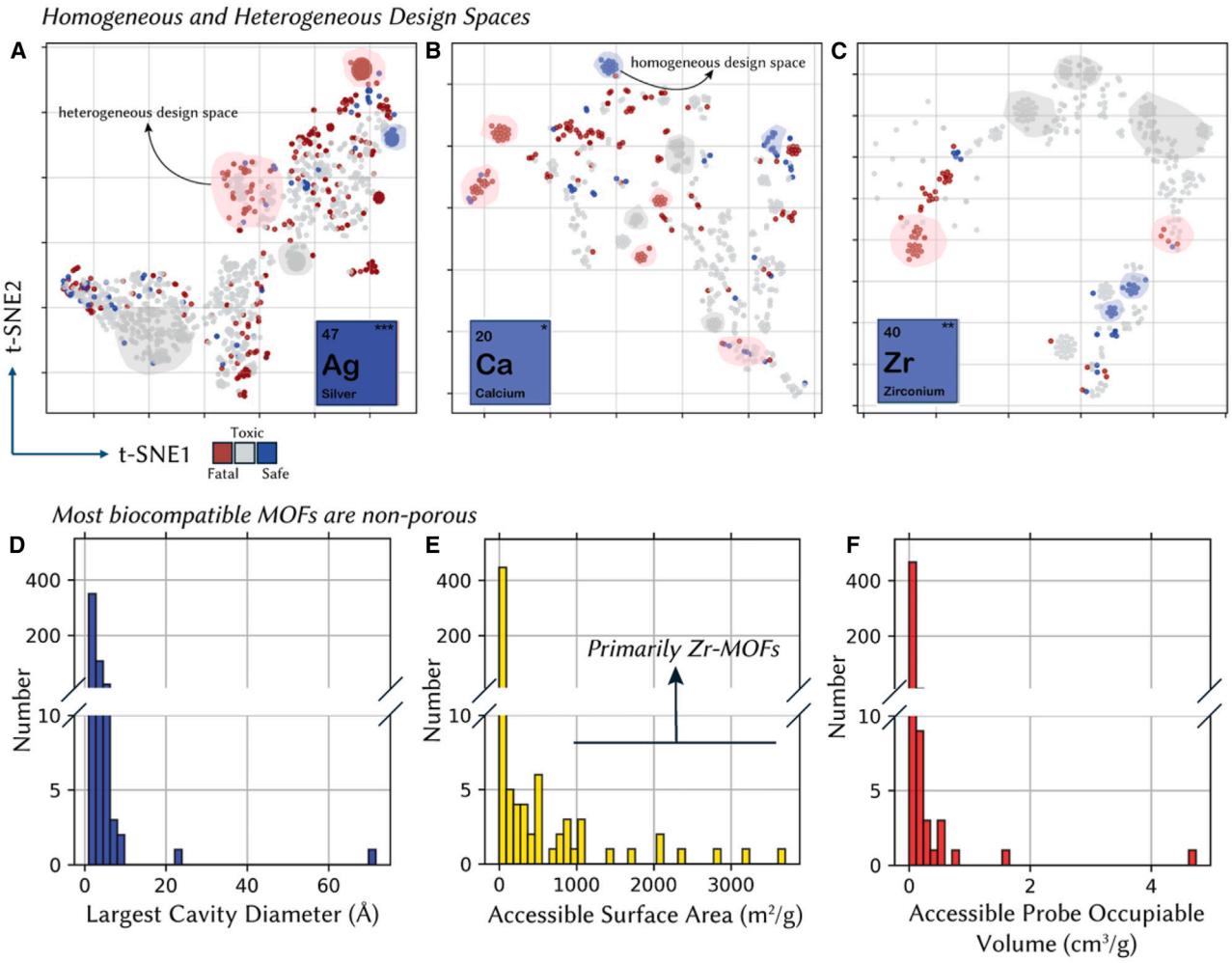


Figure 4. Local interpretability of classifications

Using SHAP analysis, we explore features that make positive and negative contributions to the toxicity of the mIm and aIm linkers of ZIF-8 and ZIF-90, respectively. Based on this analysis, we then demonstrate how a simple functionalization using an amide group may improve the safety of the linker.



(legend on next page)

in vitro and *in vivo*, making the data inherently biased toward highly biocompatible MOFs. Moreover, these analyses are generally performed at small time scales and, typically, do not include immunogenicity studies that could flag potential biocompatibility issues; to advance the translation of MOFs toward the clinic, these will be required. Nonetheless, we thought it would be a useful exercise to analyze whether the model can accurately predict the toxicity of experimentally reported MOFs. Notably our prediction that the mIm linker in ZIF-8 is "less" toxic aligns well with recent work by Alsaiari et al.,[75] wherein it was shown that the pH-triggered degradation of ZIF-8 into mIm facilitated an inflammatory response associated with the production of pro-inflammatory interleukin-6 (IL-6) cytokine. Furthermore, the RF-based three-class classifier for the i.p. route predicted the benzene-1,3,5-tricarboxylic acid linker in Cu-centered HKUST-1 MOF, the tetrakis(4-carboxyphenyl)porphine linker in Zr-centered PCN-222 MOF, and the benzene-1,4-dicarboxylic acid linker in Zr-centered UiO-66 MOF as safe. On the other hand, it classified the 1,3,6,8-tetrakis(p-benzoic acid)pyrene linker in Zr-centered NU-1000 as toxic. Pursuant to these predictions, and the potential toxicity of the metallic centers, HKUST-1 is predicted to be toxic, NU-1000 to be toxic, and PCN-222 and UiO-66 to be largely safe for use in drug-delivery-related applications. Given the chemical similarity between PCN-222 and NU-1000, the observation that NU-1000 is a potentially toxic MOF is certainly interesting. To dig deeper, we, again, performed SHAP analysis to understand features that made large contributions toward the classification (Figure 6A). The analysis indicates that the low QED, planarity of the π system, and features corresponding to the polarity (discussed in Figure 6A) significantly contribute to the toxicity of the linker. Based on the performance of unseen testing data, and these few examples, the model thus proves to be broadly reliable. However, the scarcity of *in vivo* MOF data prevents rigorous testing.

Taking things one step further, a combination of model interpretability and linker modifications by leveraging domain knowledge offers the possibility of the *de novo* design of biocompatible linkers and, by extension, biocompatible MOFs. Considering again the example of the linker with a pyrene backbone in NU-1000, several of the features responsible for the toxicity classification relate to the EState (Figure 6A). Here, modifying the 2,7-positions of pyrenes offers the possibility of modulating electronic properties. For instance, 2,7-diazapyrenes have attracted special interest owing to their intriguing electronic, photophysical, and supramolecular properties.[76] In addition, the introduction of fluorine into complex organic molecules—often as a bioisostere for the hydrogen atom—is widely used in pharmaceutical chemistry due to its lipophilicity and extreme electronegativity.[77,78] Consequently, by replacing the pyrene backbone with 2,7-diazapyrene, coupled with the fluorination of the aromatic groups, we observed that the designed linker was predicted to be safe (Figure 6B). To probe the synthesizability of the organic linker, we calculated the synthetic complexity score (SCScore).[79] The SCScore ranges from 1 to 5, with a higher score implying that the molecule is more challenging to synthesize. Interestingly, the *de novo* designed linker had a similar synthetic complexity compared to the linker present in NU-1000—implying that the linker is synthetically accessible (Figure 6B). Moreover, upon assembling a MOF structure with the oxo-Zr_6 cluster and the designed linker using a topology-guided automated MOF construction algorithm,[80] we obtained an NU-1000-like MOF, which is potentially more biocompatible while retaining porosity and being synthesizable (as estimated from the SCScore) (Figure 6B).

In the bigger picture of the clinical translation of MOFs, biocompatibility is undoubtedly a crucial consideration. An equally important consideration is that of stability, under both storage and physiological conditions. While, again, it is possible to extract chemical trends related to stability based on chemical intuition and experimental data, these may not be generalizable—especially in biological contexts—due to the inherent complexity of biological systems and the kinetic lability of the metal-ligand bond. From a thermodynamic standpoint, MOF degradation would depend on the strength of the metal-ligand coordination bond. With regard to the kinetics of degradation, the connectivity and shielding of the building unit is crucial. The higher the connectivity, the lower the rate of substitution of incoming species and, as a result, the slower the framework decomposition.[81] Similarly, rigid linkers require higher activation energies for decomposition compared to flexible linkers.[81,82] For more in-depth discussions on stability, we refer readers to dedicated reviews on this topic.[81,82] Perhaps the most interesting recent development with regard to the clinical translation of MOFs is the advancement of external surface functionalization—as briefly discussed earlier. These external surface modifications have improved colloidal stability, enable controlled payload release, facilitate targeted delivery, and improve the biocompatibility of MOFs.[23] Current focus lies in the engineering of the external surface modifications for facilitating endosomal escape. With the exception of a few MOFs—such as ZIF-8 and derivates—that show the proton sponge effect at acidic pH leading to the rupture of the endosomal membrane, most MOFs are unable to escape the endosome, serving as another critical roadblock from a translational perspective.[8] We anticipate that advances in external surface engineering can help navigate the "endosomal trap." For a comprehensive discussion of external surface modifications, we refer readers to our recent review.[83]

DISCUSSION

The translation of MOFs into clinical applications requires the design of biocompatible MOFs. Indeed, new drug delivery

Figure 5. Chemical landscape of linker characteristics of experimentally reported MOFs

(A–C) Landscape of chemical features projected on a two-dimensional map for (A) 1,590 Ag-centered, (B) 560 Ca-centered, and (C) 409 Zr-centered MOFs. Clusters indicate linkers sharing high chemical similarity. The space enclosed around the cluster (manually annotated) may serve to guide linker design choices. (D–F) Distributions of (D) largest cavity diameter (LCD), (E) accessible surface area, and (F) accessible probe-occupiable volume of the experimentally reported safe MOFs.

(G) Understanding the role of linker modifications on modulating safety profiles. Here, we consider archetypal substituted imidazoles.

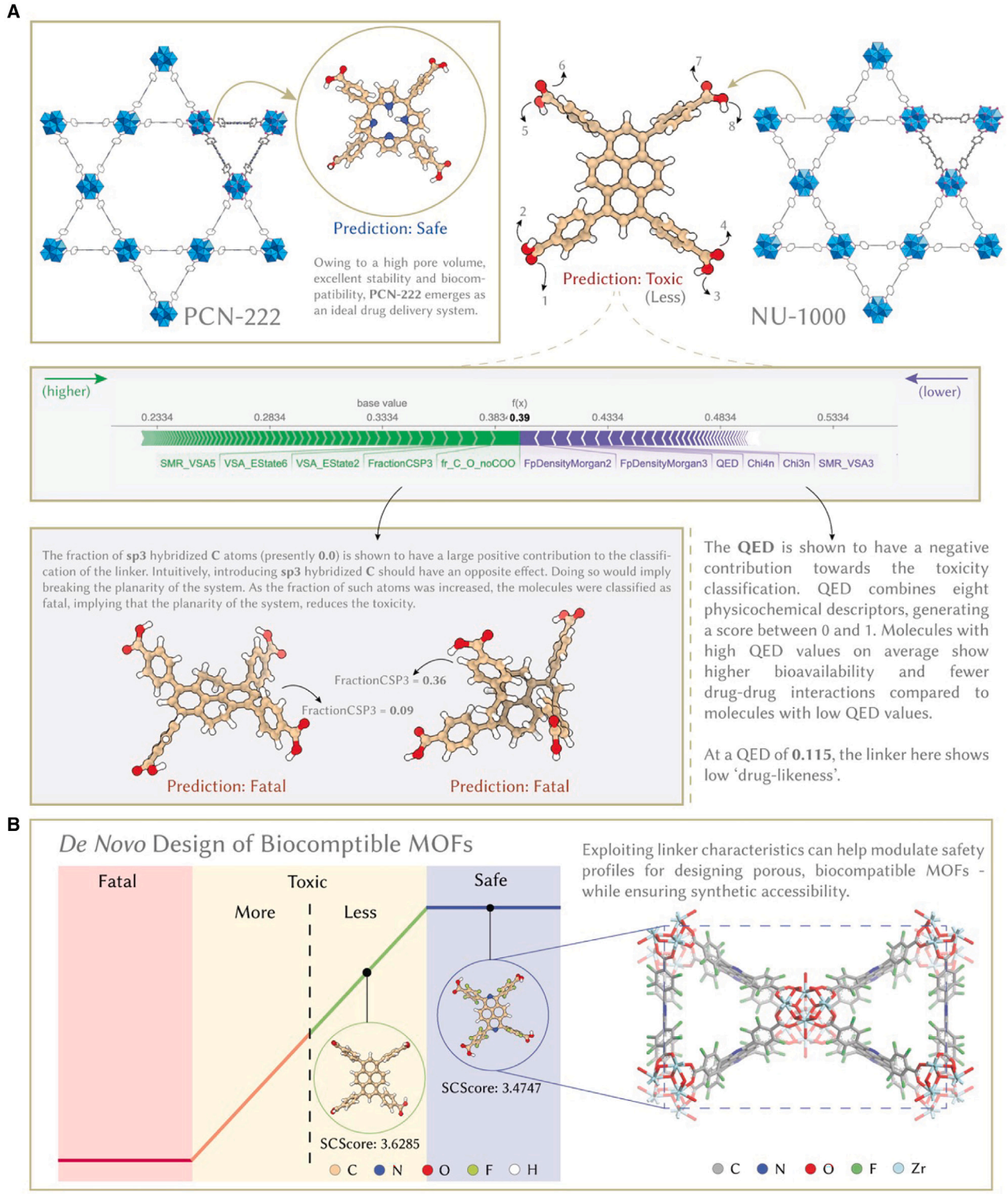


Figure 6. *De novo* design of biocompatible MOFs

(A) Despite a high chemical similarity to the linker of PCN-222, the linker of NU-1000 was classified as toxic. SHAP analysis sheds light on features that made both positive and negative contributions toward this classification. Analyzing these features in the context of the linker paves the way for incorporating strategies that may improve the toxicity profile of the linker.

(legend continued on next page)

systems must be de-risked as soon as possible to facilitate successful clinical translation. While experiments are resource and time intensive, classical modeling approaches fail to capture the behavior of MOFs in biological systems and, therefore, the fundamental understanding of what makes a MOF biocompatible or toxic. Key insights from our investigations suggest that it is feasible to leverage the precursor chemistry to gain predictive insights into MOF biocompatibility. For the metallic centers, we constructed a database of toxicity considering multiple oxidation states normalized to the chloride salt as an ionic control and considering oral administration. For the organic linkers, we developed robust and interpretable ML models based on SVMs, RFs, and GBMs. The best-performing model evaluated against multiple metrics gave an accuracy of over 80% on unseen data. Using the constructed database and ML models, we screened the CSD containing over 86,000 non-disordered MOF structures for potential biocompatible MOFs for drug delivery, identifying safe and "less" toxic candidates, most of which have well-defined synthesis protocols. Notably, our findings, especially for the metallic center toxicity, largely agree with recent research concerned with the clinical potential of MOFs.[28,29,84]

By employing interpretable ML models, we have uncovered crucial chemical features in organic linker molecules strongly linked to toxicity at both global and local levels. These features encompass the presence or absence of specific molecular fragments, surface areas, partial charges, and more abstract characteristics like drug-likeness. This analysis has allowed us to propose general design strategies for creating MOFs with enhanced biocompatibility. At a local level, we were able to analyze the features that made large contributions to the toxicity classification and propose design choices for making the linker safer. In addition, guided by the principles of SbD, we have surveyed the chemical space of experimentally reported MOFs to pinpoint promising candidates and chemical feature domains suitable for drug delivery applications. A crucial finding is that MOFs commonly employed in drug delivery applications may be substituted by potentially safer metallic centers such as Ag, Ca, and Zr. It is important to note that material selection for drug delivery involves various factors, including porosity, stability, and solubility—briefly discussed in the present context—which these MOFs with alternative metallic centers may not fully address. However, techniques such as external surface functionalization could offer effective solutions to mitigate these concerns.

Most experimentally reported MOFs that were predicted to be "safe," except for certain Zr-centered MOFs, lack the desired porosity, posing a challenge for drug delivery applications where ample pore volume is essential. This biocompatibility-porosity trade-off in particular could be a potential translational hurdle. This is because, in many cases, the drugs themselves are also highly toxic. In such cases, looking at quantitative relationships between efficacy and safety—usually captured using the "therapeutic index" (TI)—provides useful guidelines.[85] A strong translational potential necessitates a high TI, which in turn requires a high porosity and low density to accommodate large quantities of drugs in a small dosage of MOF. This emphasizes the potential of Zr-based MOFs for healthcare applications due to their high biocompatibility and superior pore characteristics. It also underscores the necessity for theory-guided rational MOF design using the principles outlined in this study to create highly porous, biocompatible MOFs. Furthermore, the emergence of generative artificial intelligence techniques in materials science will be a valuable tool for achieving this goal.[86,87]

The idea of being able to predict the toxicity of a material from its precursors is, by all means, a tool to select and shortlist materials. Since our models focus on composition, they do not consider factors such as node connectivity or the presence of defects. In realistic scenarios, several factors such as the strength of bonds, particle size, morphology, and external surface chemistry would also contribute to toxicity. Since final MOF degradation is expected during or after the drug-release process, our models here are meant for screening purposes of their building blocks, with the intent of restricting the number of MOF candidates that need to be experimentally tested. This approach will help reduce associated costs, timescales, and experimental burden and, more importantly, address the ethical concerns pertaining to animal studies. These approximations are also the reason we did not develop regression-based models, as quantitative predictions would rarely have any correlation to observations in realistic scenarios.[32] Such approximations are not uncommon to most modeling approaches, such as density functional theory and molecular dynamics, where strategies such as force-field mixing and treating materials as a sum of their parts are routinely employed.[23,88–90]

Having screened over 400 potentially safe and thousands of "low" toxicity candidates for drug delivery, we are optimistic of the applicability of these materials at a clinical level. Looking to the future, there are several ways these findings can be utilized. These MOFs can be further screened using computational approaches for ideal properties such as surface chemistry and interactions with therapeutic agents to shortlist candidates that can be explored in-depth using experimental approaches. Alternatively, operating according to the design guidelines and within the chemical spaces identified suggests the possibility of discovering new MOFs for drug delivery applications. In either case, the experimental outcomes can be used as feedback to further optimize model performance. Looking toward clinical translation, it would be critical to complement the understanding of MOF biocompatibility with stability under biological contexts and optimum drug release characteristics. This is because MOFs with ideal safety profiles may have poor stability or unfavorable loading or release characteristics—leading to a balancing act between several competing factors. That being

(B) Exploiting linker characteristics coupled with model interpretability facilitates the design of porous, biocompatible MOFs while ensuring synthetic accessibility. Here, replacing the pyrene backbone in NU-1000 with a 2,7-diazapyrene backbone, coupled with fluorination of aromatic groups, enhances biocompatibility. The MOF assembled using the oxo-Zr_6 node and this *de novo*-designed linker retains porosity while also being synthetically accessible.

said, the landscape of MOFs is incredibly vast, with only a fraction of it having been experimentally explored. Integrating recent advances in the prediction of MOF stability,[91] along with the powerful capabilities of molecular simulation techniques to calculate loading profiles,[92] it is feasible to accelerate the navigation of the MOF chemical space to expand the library of biocompatible candidates for drug delivery applications.

METHODS

Details regarding the methods can be found in the supplemental information.

RESOURCE AVAILABILITY

Lead contact

Requests for further information or data should be directed to the lead contact, David Fairen-Jimenez (df334@cam.ac.uk).

Materials availability

This study did not generate any new materials.

Data and code availability

The Python code for executing the entire pipeline as outlined in Scheme 1, along with a Jupyter Notebook for training and executing the model, is freely available for general use under the MIT license and is deposited at https://github.com/fairen-group/mof-biocompatibility. In addition, a complete data replication package has been deposited on Zenodo.[93]

ACKNOWLEDGMENTS

D.M. would like to thank Mohammad Reza Alizadeh Kiapi for independently testing and verifying the source code and Dr. Xu Chen and Dr. Mehrdad Asgari for helpful discussions. D.M. acknowledges NanoDTC Cambridge – EPSRC EP/S022953/1. This project has received funding from the EPSRC IRC Hard-to-Treat Cancers (EP/S009000/1).

AUTHOR CONTRIBUTIONS

D.M. and D.F.-J. conceived the project. D.M. developed the computational pipeline under the supervision of D.F.-J. D.M. and D.F.-J. wrote the manuscript.

DECLARATION OF INTERESTS

D.F.-J. is founder and CEO of Vector Bioscience Cambridge, a company working on the commercialization of MOFs for healthcare applications.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.matt.2025.101958.

Received: June 27, 2024
Revised: October 3, 2024
Accepted: December 24, 2024
Published: January 29, 2025

REFERENCES

1. Schirrmacher, V. (2019). From chemotherapy to biological therapy: A review of novel concepts to reduce the side effects of systemic cancer treatment (Review). Int. J. Oncol. *54*, 407–419. https://doi.org/10.3892/ijo.2018.4661.
2. Zhang, Y., Sun, C., Wang, C., Jankovic, K.E., and Dong, Y. (2021). Lipids and Lipid Derivatives for RNA Delivery. Chem. Rev. *121*, 12181–12277. https://doi.org/10.1021/acs.chemrev.1c00244.
3. Paunovska, K., Loughrey, D., and Dahlman, J.E. (2022). Drug delivery systems for RNA therapeutics. Nat. Rev. Genet. *23*, 265–280. https://doi.org/10.1038/s41576-021-00439-4.
4. Poon, W., Kingston, B.R., Ouyang, B., Ngo, W., and Chan, W.C.W. (2020). A framework for designing delivery systems. Nat. Nanotechnol. *15*, 819–829. https://doi.org/10.1038/s41565-020-0759-5.
5. Mitchell, M.J., Billingsley, M.M., Haley, R.M., Wechsler, M.E., Peppas, N.A., and Langer, R. (2021). Engineering precision nanoparticles for drug delivery. Nat. Rev. Drug Discov. *20*, 101–124. https://doi.org/10.1038/s41573-020-0090-8.
6. Blanco, E., Shen, H., and Ferrari, M. (2015). Principles of nanoparticle design for overcoming biological barriers to drug delivery. Nat. Biotechnol. *33*, 941–951. https://doi.org/10.1038/nbt.3330.
7. Qutub, S.S., Bhat, I.A., Maatouk, B.I., Moosa, B., Fakim, A., Nawaz, K., Diaz-Galicia, E., Lin, W., Grünberg, R., Arold, S.T., and Khashab, N.M. (2024). An Amphiphilic Cell-Penetrating Macrocycle for Efficient Cytosolic Delivery of Proteins, DNA, and CRISPR Cas9. Angew. Chem. Int. Ed. Engl. *63*, e202403647. https://doi.org/10.1002/anie.202403647.
8. Linnane, E., Haddad, S., Melle, F., Mei, Z., and Fairen-Jimenez, D. (2022). The uptake of metal–organic frameworks: a journey into the cell. Chem. Soc. Rev. *51*, 6065–6086. https://doi.org/10.1039/D0CS01414A.
9. Yang, W., Liang, H., Ma, S., Wang, D., and Huang, J. (2019). Gold nanoparticle based photothermal therapy: Development and application for effective cancer treatment. Sustainable Materials and Technologies *22*, e00109. https://doi.org/10.1016/j.susmat.2019.e00109.
10. Arias, L.S., Pessan, J.P., Vieira, A.P.M., Lima, T.M.T.d., Delbem, A.C.B., and Monteiro, D.R. (2018). Iron Oxide Nanoparticles for Biomedical Applications: A Perspective on Synthesis, Drugs. Antibiotics *7*, 46. https://doi.org/10.3390/antibiotics7020046.
11. Fenton, O.S., Olafson, K.N., Pillai, P.S., Mitchell, M.J., and Langer, R. (2018). Advances in Biomaterials for Drug Delivery. Adv. Mater. *30*, e1705328. https://doi.org/10.1002/adma.201705328.
12. Sercombe, L., Veerati, T., Moheimani, F., Wu, S.Y., Sood, A.K., and Hua, S. (2015). Advances and Challenges of Liposome Assisted Drug Delivery. Front. Pharmacol. *6*, 286. https://doi.org/10.3389/fphar.2015.00286.
13. Horcajada, P., Chalati, T., Serre, C., Gillet, B., Sebrie, C., Baati, T., Eubank, J.F., Heurtaux, D., Clayette, P., Kreuz, C., et al. (2010). Porous metal–organic-framework nanoscale carriers as a potential platform for drug delivery and imaging. Nat. Mater. *9*, 172–178. https://doi.org/10.1038/nmat2608.
14. Lawson, H.D., Walton, S.P., and Chan, C. (2021). Metal–Organic Frameworks for Drug Delivery: A Design Perspective. ACS Appl. Mater. Interfaces *13*, 7004–7020. https://doi.org/10.1021/acsami.1c01089.
15. Abánades Lázaro, I., and Forgan, R.S. (2019). Application of zirconium MOFs in drug delivery and biomedicine. Coord. Chem. Rev. *380*, 230–259. https://doi.org/10.1016/j.ccr.2018.09.009.
16. Tian, T., Zeng, Z., Vulpe, D., Casco, M.E., Divitini, G., Midgley, P.A., Silvestre-Albero, J., Tan, J.-C., Moghadam, P.Z., and Fairen-Jimenez, D. (2018). A sol–gel monolithic metal–organic framework with enhanced methane uptake. Nat. Mater. *17*, 174–179. https://doi.org/10.1038/nmat5050.
17. Dolgopolova, E.A., Rice, A.M., Martin, C.R., and Shustova, N.B. (2018). Photochemistry and photophysics of MOFs: steps towards MOF-based sensing enhancements. Chem. Soc. Rev. *47*, 4710–4728. https://doi.org/10.1039/C7CS00861A.
18. Yang, D., and Gates, B.C. (2019). Catalysis by Metal Organic Frameworks: Perspective and Suggestions for Future Research. ACS Catal. *9*, 1779–1798. https://doi.org/10.1021/acscatal.8b04515.
19. McKinlay, A.C., Morris, R.E., Horcajada, P., Férey, G., Gref, R., Couvreur, P., and Serre, C. (2010). BioMOFs: Metal–Organic Frameworks for

Biological and Medical Applications. Angew. Chem. Int. Ed. Engl. *49*, 6260–6266. https://doi.org/10.1002/anie.201000048.

20. Kalaj, M., Bentz, K.C., Ayala, S., Palomba, J.M., Barcus, K.S., Katayama, Y., and Cohen, S.M. (2020). MOF-Polymer Hybrid Materials: From Simple Composites to Tailored Architectures. Chem. Rev. *120*, 8267–8302. https://doi.org/10.1021/acs.chemrev.9b00575.

21. Menon, D., and Bhatia, D. (2022). Biofunctionalized metal–organic frameworks and host–guest interactions for advanced biomedical applications. J. Mater. Chem. B *10*, 7194–7205. https://doi.org/10.1039/D2TB00459C.

22. Simon-Yarza, T., Mielcarek, A., Couvreur, P., and Serre, C. (2018). Nanoparticles of Metal-Organic Frameworks: On the Road to In Vivo Efficacy in Biomedicine. Adv. Mater. *30*. https://doi.org/10.1002/adma.201707365.

23. Chen, X., Zhuang, Y., Rampal, N., Hewitt, R., Divitini, G., O'Keefe, C.A., Liu, X., Whitaker, D.J., Wills, J.W., Jugdaohsingh, R., et al. (2021). Formulation of Metal–Organic Framework-Based Drug Carriers by Controlled Coordination of Methoxy PEG Phosphate: Boosting Colloidal Stability and Redispersibility. J. Am. Chem. Soc. *143*, 13557–13572. https://doi.org/10.1021/jacs.1c03943.

24. He, Y., Xiong, T., He, S., Sun, H., Huang, C., Ren, X., Wu, L., Patterson, L.H., and Zhang, J. (2021). Pulmonary Targeting Crosslinked Cyclodextrin Metal–Organic Frameworks for Lung Cancer Therapy. Adv. Funct. Mater. *31*. https://doi.org/10.1002/adfm.202004550.

25. Alsaiari, S.K., Qutub, S.S., Sun, S., Baslyman, W., Aldehaiman, M., Alyami, M., Almalik, A., Halwani, R., Merzaban, J., Mao, Z., and Khashab, N.M. (2021). Sustained and targeted delivery of checkpoint inhibitors by metal-organic frameworks for cancer immunotherapy. Sci. Adv. *7*, eabe7174. https://doi.org/10.1126/sciadv.abe7174.

26. Chen, X., Mendes, B.B., Zhuang, Y., Conniot, J., Mercado Argandona, S., Melle, F., Sousa, D.P., Perl, D., Chivu, A., Patra, H.K., et al. (2024). A Fluorinated BODIPY-Based Zirconium Metal–Organic Framework for *In Vivo* Enhanced Photodynamic Therapy. J. Am. Chem. Soc. *146*, 1644–1656. https://doi.org/10.1021/jacs.3c12416.

27. Wiśniewska, P., Haponiuk, J., Saeb, M.R., Rabiee, N., and Bencherif, S.A. (2023). Mitigating metal-organic framework (MOF) toxicity for biomedical applications. Chem. Eng. J. *471*, 144400. https://doi.org/10.1016/j.cej.2023.144400.

28. Ettlinger, R., Lächelt, U., Gref, R., Horcajada, P., Lammers, T., Serre, C., Couvreur, P., Morris, R.E., and Wuttke, S. (2022). Toxicity of metal–organic framework nanoparticles: from essential analyses to potential applications. Chem. Soc. Rev. *51*, 464–484. https://doi.org/10.1039/D1CS00918D.

29. Wang, Y., He, L., Wang, M., Yuan, J., Wu, S., Li, X., Lin, T., Huang, Z., Li, A., Yang, Y., et al. (2024). The drug loading capacity prediction and cytotoxicity analysis of metal–organic frameworks using stacking algorithms of machine learning. Int. J. Pharm. *656*, 124128. https://doi.org/10.1016/j.ijpharm.2024.124128.

30. Bajorath, J. (2002). Integration of virtual and high-throughput screening. Nat. Rev. Drug Discov. *1*, 882–894. https://doi.org/10.1038/nrd941.

31. Pétuya, R., Durdy, S., Antypov, D., Gaultois, M.W., Berry, N.G., Darling, G.R., Katsoulidis, A.P., Dyer, M.S., and Rosseinsky, M.J. (2022). Machine-Learning Prediction of Metal–Organic Framework Guest Accessibility from Linker and Metal Chemistry. Angew. Chem. Int. Ed. *61*. https://doi.org/10.1002/anie.202114573.

32. Batra, R., Chen, C., Evans, T.G., Walton, K.S., and Ramprasad, R. (2020). Prediction of water stability of metal–organic frameworks using machine learning. Nat. Mach. Intell. *2*, 704–710. https://doi.org/10.1038/s42256-020-00249-z.

33. Wu, L., Yan, B., Han, J., Li, R., Xiao, J., He, S., and Bo, X. (2023). TOXRIC: a comprehensive database of toxicological data and benchmarks. Nucleic Acids Res. *51*, D1432–D1445. https://doi.org/10.1093/nar/gkac1074.

34. Lundberg, S.M., and Lee, S.-I. (2017). A unified approach to interpreting model predictions. Adv. Neural Inf. Process. Syst. *30*.

35. Moghadam, P.Z., Li, A., Wiggin, S.B., Tao, A., Maloney, A.G.P., Wood, P.A., Ward, S.C., and Fairen-Jimenez, D. (2017). Development of a Cambridge Structural Database Subset: A Collection of Metal–Organic Frameworks for Past, Present, and Future. Chem. Mater. *29*, 2618–2625. https://doi.org/10.1021/acs.chemmater.7b00441.

36. Akhila, J.S., Shyamjith, D., and Alwar, M.C. (2007). Acute toxicity studies and determination of median lethal dose. Curr. Sci. *93*, 917–920. http://www.jstor.org/stable/24099255.

37. Weininger, D. (1988). SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J. Chem. Inf. Comput. Sci. *28*, 31–36. https://doi.org/10.1021/ci00057a005.

38. United Nations. (2019). Globally harmonized system of classification and labelling of chemicals (GHS).

39. Wigh, D.S., Goodman, J.M., and Lapkin, A.A. (2022). A review of molecular representation in the age of machine learning. WIREs Comput. Mol. Sci. *12*. https://doi.org/10.1002/wcms.1603.

40. Landrum, G. (2013). RDKit: A software suite for cheminformatics, computational chemistry, and predictive modeling.

41. Himanen, L., Jäger, M.O., Morooka, E.V., Federici Canova, F., Ranawat, Y.S., Gao, D.Z., Rinke, P., and Foster, A.S. (2020). DScribe: Library of descriptors for machine learning in materials science. Comput. Phys. Commun. *247*, 106949. https://doi.org/10.1016/j.cpc.2019.106949.

42. Labute, P. (2000). A widely applicable set of descriptors. J. Mol. Graph. Model. *18*, 464–477. https://doi.org/10.1016/S1093-3263(00)00068-1.

43. Waring, M.J., Arrowsmith, J., Leach, A.R., Leeson, P.D., Mandrell, S., Owen, R.M., Pairaudeau, G., Pennie, W.D., Pickett, S.D., Wang, J., et al. (2015). An analysis of the attrition of drug candidates from four major pharmaceutical companies. Nat. Rev. Drug Discov. *14*, 475–486. https://doi.org/10.1038/nrd4609.

44. Fernández-Torras, A., Comajuncosa-Creus, A., Duran-Frigola, M., and Aloy, P. (2022). Connecting chemistry and biology through molecular descriptors. Curr. Opin. Chem. Biol. *66*, 102090. https://doi.org/10.1016/j.cbpa.2021.09.001.

45. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., and Galstyan, A. (2022). A Survey on Bias and Fairness in Machine Learning. ACM Comput. Surv. *54*, 1–35. https://doi.org/10.1145/3457607.

46. Jablonka, K.M., Ongari, D., Moosavi, S.M., and Smit, B. (2020). Big-Data Science in Porous Materials: Materials Genomics and Machine Learning. Chem. Rev. *120*, 8066–8129. https://doi.org/10.1021/acs.chemrev.0c00004.

47. He, H., Yang, B., Garcia, E.A., and Li, S. (2008). ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In 2008 IEEE International Joint Conference on Neural Networks ((IEEE World Congress on Computational Intelligence) (IEEE)), pp. 1322–1328. https://doi.org/10.1109/IJCNN.2008.4633969.

48. Butler, K.T., Davies, D.W., Cartwright, H., Isayev, O., and Walsh, A. (2018). Machine learning for molecular and materials science. Nature *559*, 547–555. https://doi.org/10.1038/s41586-018-0337-2.

49. Kim, T.K. (2017). Understanding one-way ANOVA using conceptual figures. Korean J. Anesthesiol. *70*, 22–26. https://doi.org/10.4097/kjae.2017.70.1.22.

50. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. *12*, 2825–2830.

51. Swift, A., Heale, R., and Twycross, A. (2020). What are sensitivity and specificity? Evid. Base Nurs. *23*, 2–4. https://doi.org/10.1136/ebnurs-2019-103225.

52. Mandrekar, J.N. (2010). Receiver Operating Characteristic Curve in Diagnostic Test Assessment. J. Thorac. Oncol. *5*, 1315–1316. https://doi.org/10.1097/JTO.0b013e3181ec173d.

53. Yuan, S., Qin, J.-S., Lollar, C.T., and Zhou, H.-C. (2018). Stable Metal–Organic Frameworks with Group 4 Metals: Current Status and Trends. ACS Cent. Sci. *4*, 440–450. https://doi.org/10.1021/acscentsci.8b00073.

54. Jomova, K., and Valko, M. (2011). Advances in metal-induced oxidative stress and human disease. Toxicology *283*, 65–87. https://doi.org/10.1016/j.tox.2011.03.001.

55. Bush, A.I., and Curtain, C.C. (2008). Twenty years of metallo-neurobiology: where to now? Eur. Biophys. J. *37*, 241–245. https://doi.org/10.1007/s00249-007-0228-1.

56. Nelson, N. (1999). Metal ion transporters and homeostasis. EMBO J. *18*, 4361–4371. https://doi.org/10.1093/emboj/18.16.4361.

57. Levard, C., Hotze, E.M., Lowry, G.V., and Brown, G.E. (2012). Environmental Transformations of Silver Nanoparticles: Impact on Stability and Toxicity. Environ. Sci. Technol. *46*, 6900–6914. https://doi.org/10.1021/es2037405.

58. Liu, X., Liang, T., Zhang, R., Ding, Q., Wu, S., Li, C., Lin, Y., Ye, Y., Zhong, Z., and Zhou, M. (2021). Iron-Based Metal–Organic Frameworks in Drug Delivery and Biomedicine. ACS Appl. Mater. Interfaces *13*, 9643–9655. https://doi.org/10.1021/acsami.0c21486.

59. Liu, W., Pan, Y., Xiao, W., Xu, H., Liu, D., Ren, F., Peng, X., and Liu, J. (2019). Recent developments on zinc(ii) metal–organic framework nanocarriers for physiological pH-responsive drug delivery. Medchemcomm *10*, 2038–2051. https://doi.org/10.1039/C9MD00400A.

60. Jablonka, K.M., Rosen, A.S., Krishnapriyan, A.S., and Smit, B. (2023). An Ecosystem for Digital Reticular Chemistry. ACS Cent. Sci. *9*, 563–581. https://doi.org/10.1021/acscentsci.2c01177.

61. Lipinski, C.A., Lombardo, F., Dominy, B.W., and Feeney, P.J. (1997). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv. Drug Deliv. Rev. *23*, 3–25. https://doi.org/10.1016/S0169-409X(96)00423-1.

62. Menon, D., and Ranganathan, R. (2022). A Generative Approach to Materials Discovery, Design, and Optimization. ACS Omega *7*, 25958–25973. https://doi.org/10.1021/acsomega.2c03264.

63. Gómez-Bombarelli, R., Wei, J.N., Duvenaud, D., Hernández-Lobato, J.M., Sánchez-Lengeling, B., Sheberla, D., Aguilera-Iparraguirre, J., Hirzel, T.D., Adams, R.P., and Aspuru-Guzik, A. (2018). Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. ACS Cent. Sci. *4*, 268–276. https://doi.org/10.1021/acscentsci.7b00572.

64. Bickerton, G.R., Paolini, G.V., Besnard, J., Muresan, S., and Hopkins, A.L. (2012). Quantifying the chemical beauty of drugs. Nat. Chem. *4*, 90–98. https://doi.org/10.1038/nchem.1243.

65. Ritchie, T.J., and Macdonald, S.J.F. (2009). The impact of aromatic ring count on compound developability – are too many aromatic rings a liability in drug design? Drug Discov. Today *14*, 1011–1020. https://doi.org/10.1016/j.drudis.2009.07.014.

66. Lovering, F., Bikker, J., and Humblet, C. (2009). Escape from Flatland: Increasing Saturation as an Approach to Improving Clinical Success. J. Med. Chem. *52*, 6752–6756. https://doi.org/10.1021/jm901241e.

67. Zhang, M.-Q., and Wilkinson, B. (2007). Drug discovery beyond the 'rule-of-five. Curr. Opin. Biotechnol. *18*, 478–488. https://doi.org/10.1016/j.copbio.2007.10.005.

68. Yan, L., Zhao, F., Wang, J., Zu, Y., Gu, Z., and Zhao, Y. (2019). A Safe-by-Design Strategy towards Safer Nanomaterials in Nanomedicines. Adv. Mater. *31*, e1805391. https://doi.org/10.1002/adma.201805391.

69. van der Maaten, L., and Hinton, G. (2008). Visualizing Data using t-SNE. J. Mach. Learn. Res. *9*, 2579–2605.

70. Zhang, J.-P., Zhang, Y.-B., Lin, J.-B., and Chen, X.-M. (2012). Metal Azolate Frameworks: From Crystal Engineering to Functional Materials. Chem. Rev. *112*, 1001–1033. https://doi.org/10.1021/cr200139g.

71. Williams, D.P. (2006). Toxicophores: Investigations in drug safety. Toxicology *226*, 1–11. https://doi.org/10.1016/j.tox.2006.05.101.

72. Boelsterli, U.A., and Kashimshetty, R. (2010). Idiosyncratic Drug-Induced Liver Injury: Mechanisms and Susceptibility Factors. In Comprehensive Toxicology (Elsevier), pp. 383–402. https://doi.org/10.1016/B978-0-08-046884-6.01015-0.

73. Nepali, K., Lee, H.-Y., and Liou, J.-P. (2019). Nitro-Group-Containing Drugs. J. Med. Chem. *62*, 2851–2893. https://doi.org/10.1021/acs.jmedchem.8b00147.

74. Willems, T.F., Rycroft, C.H., Kazi, M., Meza, J.C., and Haranczyk, M. (2012). Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials. Microporous Mesoporous Mater. *149*, 134–141. https://doi.org/10.1016/j.micromeso.2011.08.020.

75. Alsaiari, S.K., Nadeef, S., Daristotle, J.L., Rothwell, W., Du, B., Garcia, J., Zhang, L., Sarmadi, M., Forster, T.A., Menon, N., et al. (2024). Zeolitic imidazolate frameworks activate endosomal Toll-like receptors and potentiate immunogenicity of SARS-CoV-2 spike protein trimer. Sci. Adv. *10*, eadj6380. https://doi.org/10.1126/sciadv.adj6380.

76. Mukherjee, A., Akulov, A.A., Santra, S., Varaksin, M.V., Kim, G.A., Kopchuk, D.S., Taniya, O.S., Zyryanov, G.V., and Chupakhin, O.N. (2022). 2,7-Diazapyrenes: a brief review on synthetic strategies and application opportunities. RSC Adv. *12*, 9323–9341. https://doi.org/10.1039/D2RA00260D.

77. Meanwell, N.A. (2018). Fluorine and Fluorinated Motifs in the Design and Application of Bioisosteres for Drug Design. J. Med. Chem. *61*, 5822–5880. https://doi.org/10.1021/acs.jmedchem.7b01788.

78. Yerien, D.E., Bonesi, S., and Postigo, A. (2016). Fluorination methods in drug discovery. Org. Biomol. Chem. *14*, 8398–8427. https://doi.org/10.1039/C6OB00764C.

79. Coley, C.W., Rogers, L., Green, W.H., and Jensen, K.F. (2018). SCScore: Synthetic Complexity Learned from a Reaction Corpus. J. Chem. Inf. Model. *58*, 252–261. https://doi.org/10.1021/acs.jcim.7b00622.

80. Colón, Y.J., Gómez-Gualdrón, D.A., and Snurr, R.Q. (2017). Topologically Guided, Automated Construction of Metal–Organic Frameworks and Their Evaluation for Energy-Related Applications. Cryst. Growth Des. *17*, 5801–5810. https://doi.org/10.1021/acs.cgd.7b00848.

81. He, T., Kong, X.-J., and Li, J.-R. (2021). Chemically Stable Metal–Organic Frameworks: Rational Construction and Application Expansion. Acc. Chem. Res. *54*, 3083–3094. https://doi.org/10.1021/acs.accounts.1c00280.

82. Ding, M., Cai, X., and Jiang, H.-L. (2019). Improving MOF stability: approaches and applications. Chem. Sci. *10*, 10209–10230. https://doi.org/10.1039/C9SC03916C.

83. Chen, X., Argandona, S.M., Melle, F., Rampal, N., and Fairen-Jimenez, D. (2024). Advances in surface functionalization of next-generation metal-organic frameworks for biomedical applications: Design, strategies, and prospects. Chem *10*, 504–543. https://doi.org/10.1016/j.chempr.2023.09.016.

84. Tyagi, N., Wijesundara, Y.H., Gassensmith, J.J., and Popat, A. (2023). Clinical translation of metal–organic frameworks. Nat. Rev. Mater. *8*, 701–703. https://doi.org/10.1038/s41578-023-00608-3.

85. Muller, P.Y., and Milton, M.N. (2012). The determination and interpretation of the therapeutic index in drug development. Nat. Rev. Drug Discov. *11*, 751–761. https://doi.org/10.1038/nrd3801.

86. Yao, Z., Sánchez-Lengeling, B., Bobbitt, N.S., Bucior, B.J., Kumar, S.G.H., Collins, S.P., Burns, T., Woo, T.K., Farha, O.K., Snurr, R.Q., and Aspuru-Guzik, A. (2021). Inverse design of nanoporous crystalline reticular materials with deep generative models. Nat. Mach. Intell. *3*, 76–86. https://doi.org/10.1038/s42256-020-00271-1.

87. Kim, B., Lee, S., and Kim, J. (2020). Inverse design of porous materials using artificial neural networks. Sci. Adv. *6*, eaax9324. https://doi.org/10.1126/sciadv.aax9324.

88. Mancuso, J.L., Mroz, A.M., Le, K.N., and Hendon, C.H. (2020). Electronic Structure Modeling of Metal–Organic Frameworks. Chem. Rev. *120*, 8641–8715. https://doi.org/10.1021/acs.chemrev.0c00148.

89. Wu, X., Choudhuri, I., and Truhlar, D.G. (2019). Computational Studies of Photocatalysis with Metal–Organic Frameworks. Energy &. Environ. Materials *2*, 251–263. https://doi.org/10.1002/eem2.12051.

90. Markopoulou, P., Panagiotou, N., Li, A., Bueno-Perez, R., Madden, D., Buchanan, S., Fairen-Jimenez, D., Shiels, P.G., and Forgan, R.S. (2020). Identifying Differing Intracellular Cargo Release Mechanisms by Monitoring In Vitro Drug Delivery from MOFs in Real Time. Cell Rep. Phys. Sci. *1*, 100254. https://doi.org/10.1016/j.xcrp.2020.100254.

91. Terrones, G.G., Huang, S.-P., Rivera, M.P., Yue, S., Hernandez, A., and Kulik, H.J. (2024). Metal–Organic Framework Stability in Water and Harsh Environments from Data-Driven Models Trained on the Diverse WS24 Data Set. J. Am. Chem. Soc. *146*, 20333–20348. https://doi.org/10.1021/jacs.4c05879.

92. Bernini, M.C., Fairen-Jimenez, D., Pasinetti, M., Ramirez-Pastor, A.J., and Snurr, R.Q. (2014). Screening of bio-compatible metal–organic frameworks as potential drug carriers using Monte Carlo simulations. J. Mater. Chem. B *2*, 766–774. https://doi.org/10.1039/C3TB21328E.

93. Menon, D., & Fairen-Jimenez, D. (2024). Guiding the rational design of biocompatible metal-organic frameworks for drug delivery. https://doi.org/10.5281/zenodo.13771559