Supporting Information for Machine Learning Algorithm Identifies an Antibiotic Vocabulary for Permeating Gram-Negative Bacteria

Rachael A. Mansbach,[†] Inga V. Leus,[‡,§] Jitender Mehla,[‡,§] Cesar A. Lopez,[†] John K. Walker,[¶] Valentin V. Rybenkov,[‡] Nicolas W. Hengartner,[†] Helen I. Zgurskaya,[‡] and S. Gnanakaran[*,†]

†*Department of Theoretical Biology and Biophysics, Los Alamos National Lab, Los Alamos, NM*

‡*Department of Chemistry and Biochemistry, University of Oklahoma, Norman, OK*

¶*School of Medicine, Saint Louis University, St. Louis, MO*

§*These authors contributed equally to this work*

E-mail: gnana@lanl.gov

S1 Methods Details

S1.1 Multinomial logistic regression for fragment selection and external molecule classification

We assume that a large number of the N_f fragments will not be relevant to the response variable and therefore we initially perform feature selection by seeking a set of sparse classifiers that will retain only a small number of features. A "sparse classifier" is a classifier with an additional regularization parameter that can be tuned to find solutions with only a small number of non-zero parameters. In addition to cutting down on the number of relevant variables, such an approach improves interpretability and helps prevent overfitting, a problem in which a model shows high performance on the dataset for which it is trained, but poor performance when asked to generalize to new data points.

To fit initial sparse classifiers for a single run of the Hunting FOX algorithm, we employed stratified 5-fold cross-validation followed by multinomial logistic regression (see also Sec. S2) on 595 molecules for which the MIC ratio $\left(\frac{\mu_{P\Delta 6\text{-Pore}}}{\mu_{P\Delta 6}}\right)$ was measured and identified as a definite value as described in Sec. 2.2. The data was curated and organized by employing the CDD Vault from Collaborative Drug Discovery (Burlingame, CA. www.collaborativedrug.com)[1]. We define the classes as follows, based on the approximate "streaking" in the data, which we expect to indicate drastically different molecular classes. If $\frac{\mu_{P\Delta 6\text{-Pore}}}{\mu_{P\Delta 6}} < 0.2$, Class $\left(\frac{\mu_{P\Delta 6\text{-Pore}}}{\mu_{P\Delta 6}}\right) = 0$; if $0.2 \leq \frac{\mu_{P\Delta 6\text{-Pore}}}{\mu_{P\Delta 6}} < 0.4$, Class $\left(\frac{\mu_{P\Delta 6\text{-Pore}}}{\mu_{P\Delta 6}}\right) = 1$; if $0.4 \leq \frac{\mu_{P\Delta 6\text{-Pore}}}{\mu_{P\Delta 6}} < 0.6$, Class $\left(\frac{\mu_{P\Delta 6\text{-Pore}}}{\mu_{P\Delta 6}}\right) = 2$; if $0.6 \leq \frac{\mu_{P\Delta 6\text{-Pore}}}{\mu_{P\Delta 6}} < 0.8$, Class $\left(\frac{\mu_{P\Delta 6\text{-Pore}}}{\mu_{P\Delta 6}}\right) = 3$; and if $\frac{\mu_{P\Delta 6\text{-Pore}}}{\mu_{P\Delta 6}} > 0.8$, Class $\left(\frac{\mu_{P\Delta 6\text{-Pore}}}{\mu_{P\Delta 6}}\right) = 4$, where $\mu_{P\Delta 6\text{-Pore}}$ is the MIC of the compound in the $P\Delta 6$-Pore mutant and $\mu_{P\Delta 6}$ is the MIC of the compound in the $P\Delta 6$ mutant. The class breakdown is as follows: $\approx 48\%$ of MIC ratios fall into class 0, 10% into class 1, 9% into class 2, 10% into class 4, and 22% into class 4.

We employed the `StratifiedKFold` class with a random seed from the python `scikit-learn` package[2] to split each of the two groups into five training and testing sets wherein the testing sets were disjoint and both training and testing sets approximately reproduced the distribution of different classes of the original group. For each split, we employed the `cvglmnet` function from `glmnet-python` package with `family='multinomial'`, `mtype='grouped'`, and $\alpha = 1$ to fit a sparse multinomial logistic regression model on the training set with regularization parameter λ controlling the strength of the LASSO penalty for sparsification[3]. Employing the built-in cross-validation from the package, we identified $\lambda = \lambda_{min}$ and $\lambda = \lambda_{1se}$ wherein λ_{min} corresponds to the regularization parameter with minimum deviance and λ_{1se} corresponds to the maximum regularization parameter with deviance within one standard deviation of the minimum. The potential benefit of employing λ_{1se} is that it forces the model to be sparser and may therefore reduce overfitting[3], and all trained classifiers we discuss from now on are those employing $\lambda = \lambda_{1se}$. For the fitting of the non-sparse predictive models, the only difference in the fitting procedure besides using different descriptors was that we employed $\alpha = 0$ instead of $\alpha = 1$ in the `cvglmnet` function to change from LASSO to ridge regression.

S2 Theory of multinomial logistic regression

We chose to fit our classifiers by employing the multinomial logistic regression technique, since it is well-known and therefore easily employed, provides an easily-interpretable model and, in addition to projected compound classes, the probabilities of a compound belonging to each class. It is a method employing maximum likelihood estimation to solve a multiclass problem in which, in essence, we seek to find the probability of a particular outcome being in class C, $C \in [0,4]$ given a molecular vector, $\vec{\ell}$,

$$P\left(\text{Class}\left(\frac{\mu_{P\Delta6\text{-Pore}}}{\mu_{P\Delta6}}\right) = C|\vec{\ell}\right) = \begin{cases} 1/\sum_{k=0}^{4} \exp\left(g_k(\vec{\ell}),\right) & C = 0 \\ \exp\left(g_C(\vec{\ell})\right)/\sum_{k=0}^{4} \exp\left(g_k(\vec{\ell}),\right) & \text{else} \end{cases} \tag{1}$$

where $g_C(\vec{\ell})$ is the logit function, which is the particular form of the fit chosen in general logistic regression,

$$g_C(\vec{\ell}) = \ln\left[\frac{P(\text{Class}\left(\frac{\mu_{P\Delta6\text{-Pore}}}{\mu_{P\Delta6}}\right) = C|\vec{\ell})}{P(\text{Class}\left(\frac{\mu_{P\Delta6\text{-Pore}}}{\mu_{P\Delta6}}\right) = 0|\vec{\ell})}\right] = \beta_{C0} + \sum_{k=1}^{p} \beta_{Ck}\ell_k, \tag{2}$$

where β_{Ck} are a set of coefficients to be fitted, and m is the length of the molecular descriptor vector. The logit expresses the log of the ratio of the probabilities of being in class C versus class 0. Employing this expression of the probabilities results in the following expression for the log-likelihood function[3],

$$\mathcal{L} = -\left[\frac{1}{N}\sum_{i=1}^{N}\left(\sum_{k=1}^{K} I(g_i = C)\left(\beta_{0k} + \ell_i^T\beta_k\right) - \ln\left(\sum_{k=1}^{K} \exp\left(\beta_{0k} + \ell_i^T\beta_k\right)\right)\right)\right], \tag{3}$$

where $I(g_i = \mathcal{L})$ is an indicator function, and β_k is the kth column of the $p \times K$ coefficient matrix, where K is the number of classes and p is the length of the molecular descriptor vector $\vec{\ell}$. Then multinomial logistic regression is the procedure of maximizing the log-likelihood of this function, potentially subject to additional constraints such as regularization for sparsification. Specifically, the `cvglmnet` function solves the optimization problem,

$$\begin{aligned} \min_{\{\beta_{0k},\beta_k)\}_1^K} & -\left[\frac{1}{N}\sum_{i=1}^{N}\left(\sum_{k=1}^{K} I(g_i = C)\left(\beta_{0k} + \ell_i^T\beta_k\right) - \ln\left(\sum_{k=1}^{K} \exp\left(\beta_{0k} + \ell_i^T\beta_k\right)\right)\right)\right] \\ & + \lambda\left[(1-\alpha)||\beta||_F^2 + \alpha\sum_{j=1}^{p}||\beta_j||_q\right], \end{aligned} \tag{4}$$

where $\alpha = 1, q = 2$ was used for sparse classification and $\alpha = 0$ for non-sparse classification.

S3 Parallel-tempered well-tempered Metadynamics (PtWt-MET)

In order to enhance the convergence of the membrane-translocation free energies, we have employed a metadynamics protocol making use of both parallel tempered and the well-tempered approach. In a regular metadynamics calculation, a desired set of collective variables (CV) are affected by a bias potential so a free energy landscape may be fully mapped in terms of the relevant reaction coordinates driving a process. The bias potential is built as a sum of Gaussian kernels deposited along the trajectory of the CVs. In the well-tempered approach, the Gaussian heights are affected during the simulation time according to,

$$W\left(k\tau\right) = W_0 \exp\left(\frac{-V\left(s(q(k\tau))\right)}{k_B\Delta T}\right), \tag{5}$$

where W_0 is an initial Gaussian height, ΔT the temperature of annealing, k the strength of the bias potential, $V(s(q), k\tau)$ is the potential measured along the collective variable (CV) s dependent on the molecular coordinates q, k_B is the standard Boltzmann constant, and $\tau = \frac{T+\Delta T}{T}$ is a bias term compensating for the underlying changes in the free energy landscape, defined as the ratio between the temperature of the CVs $(T + \Delta T)$ and the temperature of the system. Often, metadynamics displays reduced convergence due to large system hysteresis or complicated energetic pathways. To overcome this situation, the well-tempered metadynamics approach can be effectively combined with the replica exchange protocol. Thus, a single metadynamics calculation can be performed by N replicates in parallel and at different temperatures, using the same set of CVs. Under these circumstances, the standard parallel-tempered (PT) acceptance probability[4] must be modified to account for the presence of the bias potential in the following manner,

$$\Delta_{ij}^{PTMetaD} = \Delta_{ij}^{PT} + \frac{1}{k_B T_i}\left[V_G^i\left(s(R_i), t\right) - V_G^i\left(s(R_j), t\right)\right] + \frac{1}{k_B T_j}\left[V_G^j\left(s(R_j), t\right) - V_G^j\left(s(R_i), t\right)\right], \quad (6)$$

where V_G^i and V_G^j are bias potentials corresponding to R_i and R_j, the ith and jth replicas, respectively. For a more complete description of the method, we refer the reader to the following references: Laio and Gervasio[5], Barducci et al.[6], Sutto et al.[7], Abrams and Bussi[8], Bussi et al.[9]

S4 Future improvements

One major difficulty with the Hunting FOX algorithm stems from the fact that fragments are combinatorial in nature, which means that to describe a larger compound library a huge number of fragments will be necessary. Due to the sparsity of the representation, it is difficult to apply standard statistical techniques to the regression results. Furthermore, we note that the fragments identified and hence the molecules that the final classifiers identify change not insignificantly when the training/testing split is changed. This observation implies that the current representation is insufficiently robust to changes in the dataset, which we hypothesize is a consequence of (i) magnification of noise due to an overly high-dimensional representation, (ii) lack of incorporation of hierarchical and locational relationships among fragments in the initial construction of the representation, and (iii) uncertainties in experimental measurements of MICs. This is similar to the difficulties encountered when attempting to employ one-hot-encoding to learning on textual data.

With respect to fragment representation, work on word and sentence embeddings provide a good source of how to compress highly sparse, combinatorial representations in a meaningful way; combining, for example, the word2vec[10] or more recent graph2vec[11] approach with an autoencoder[12] would provide a compact representation that is reconstructible with the fragment composition, as well as potentially identifying a latent space of meaningful variables. An additional limitation of the current representation is that it contains geometric information only implicitly, which leads to difficulty in the direct assessment of relevant fragment properties and the potential of losing important information such as enantiomerism. For example, the molecule OU-315 studied in detail is, in fact, a stereoisomer of OU-313, and OU-313 could as easily have been chosen for analysis based on fragment composition. Therefore we also plan to pursue the incorporation of three-dimensional information into our representation.

There are several limitations associated with the use of solely coarse-grained molecular dynamics. Drug pKa can change in the membrane, but this behavior cannot be captured by a fixed-charge model. We note also that the calculated free energies are rather high, which may be related to the unrealistic charge representation. Having said this, the force field is expected to capture the relative behavior of the different drugs and give a reasonable qualitative assessment of the mechanism. In the future, we plan to use a combination of constant-pH atomistic molecular dynamics with a coarse-grained approach to more fully characterize the proposed mechanism.

Additionally, as the discussion in Sec. 3.3 highlighted, our experimental assays are currently insufficient to identify molecules that permeate well but do not have antibiotic action, a necessary development to test the full suite of predicted hits from the Hunting FOX algorithm, as well as a necessary step in the understanding of mechanisms of antibiotic resistance in isolation. Development of direct assays to analyze permeation across the outer membrane of *P. aeruginosa* is in progress.

Table S1: List of unique fragments of radius 1 and unique fragments of radius 2 contained in OU-31237.

SMILES	**Radius**
C=O	1
CC(N)=O	1
CCc	1
CF	1
CNC	1
CO	1
cCN	1
cOC	1
cc(c)C	1
cc(c)O	1
ccc	1
cC(F)(F)F	1
cC(F)(F)F	2
cc(c)C(F)(F)F	2
ccc(cc)C(F)(F)F	2
cccc(c)C	2
ccccc	2
ccc(cc)CN	2
cc(C)cc(c)C	2
cc(c)CNC	2
cCNC(C)=O	2
cCC(=O)NC	2
CC(N)=O	2
cc(c)CC(N)=O	2
ccc(cc)CC	2
cccc(c)O	2
ccc(cc)OC	2
cc(c)OC	2
COc	2

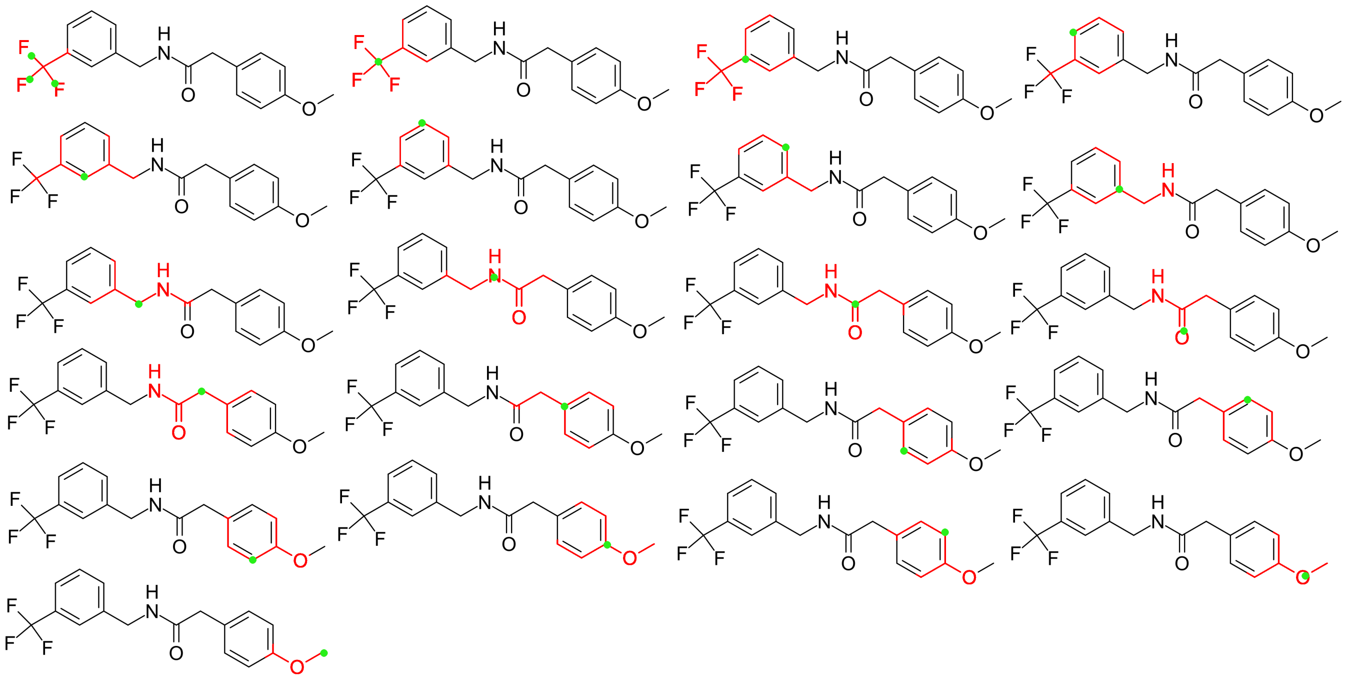

Figure S1: Example of fragment identification on the arbitrarily chosen molecule OU-31237 for fragments of radius $k = 2$. Green dots represent the central atom and red bonds represent the fragments. In the first frame the three green dots demonstrate that there are three central atoms corresponding to the same fragment. Arrows guide the eye in the direction of the sliding window along the molecule.

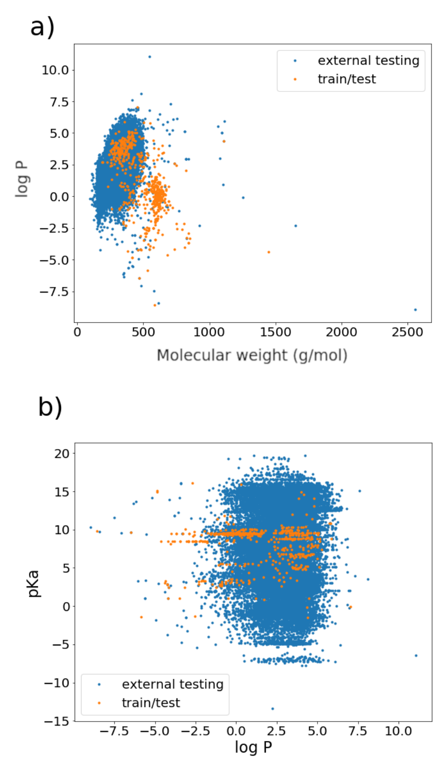

Figure S2: Physicochemical properties of external testing library (blue) and train/test set (orange) for Hunting FOX algorithm. (a) Molecular weight versus log P; (b) pKa versus log P.

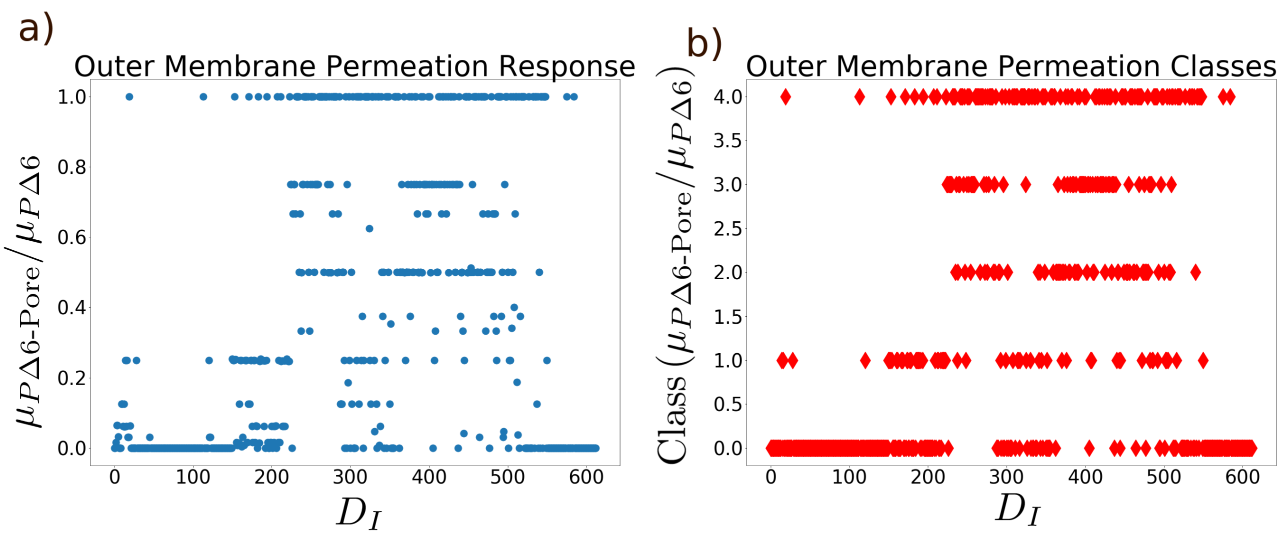

Figure S3: Illustration of class assignment for MIC ratio response variables. In a) we show the actual MIC ratio $\frac{\mu_{P\Delta 6\text{Pore}}}{\mu_{P\Delta 6}}$ for each compound considered in a training or testing simply versus its index D_I. In b) we show the corresponding class assignment. Note the discreteness of the data.

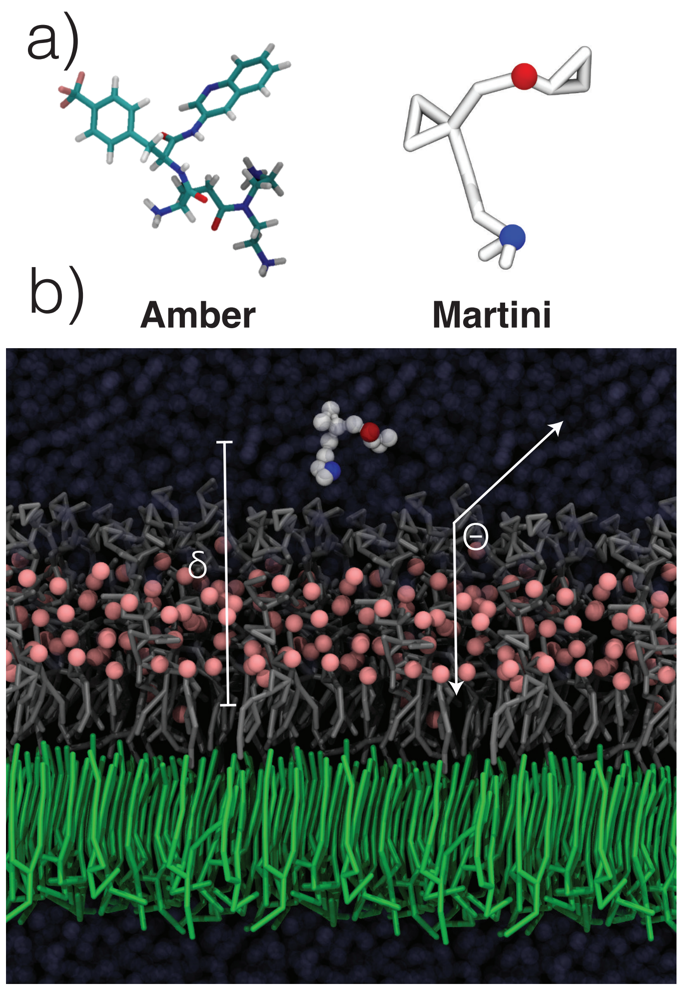

Figure S4: Martini parametrization and metadynamics setup. a) structural representation of OU-315 both at atomic resolution (Amber) and CG (Martini), with parametrization performed by using standard MARTINI procedures[13]. b) Metadynamics system setup using MARTINI. Both angle (θ) and COM-COM bias (δ) CVs are displayed. θ was defined to be the angle formed between the read bead, the blue bead and the COM of the membrane. Color scheme: Gray:LPS, Green:DPPE, Pink spheres: CA++ cations. The drug (OU-315) is displayed as white spheres. Images rendered with VMD[14].

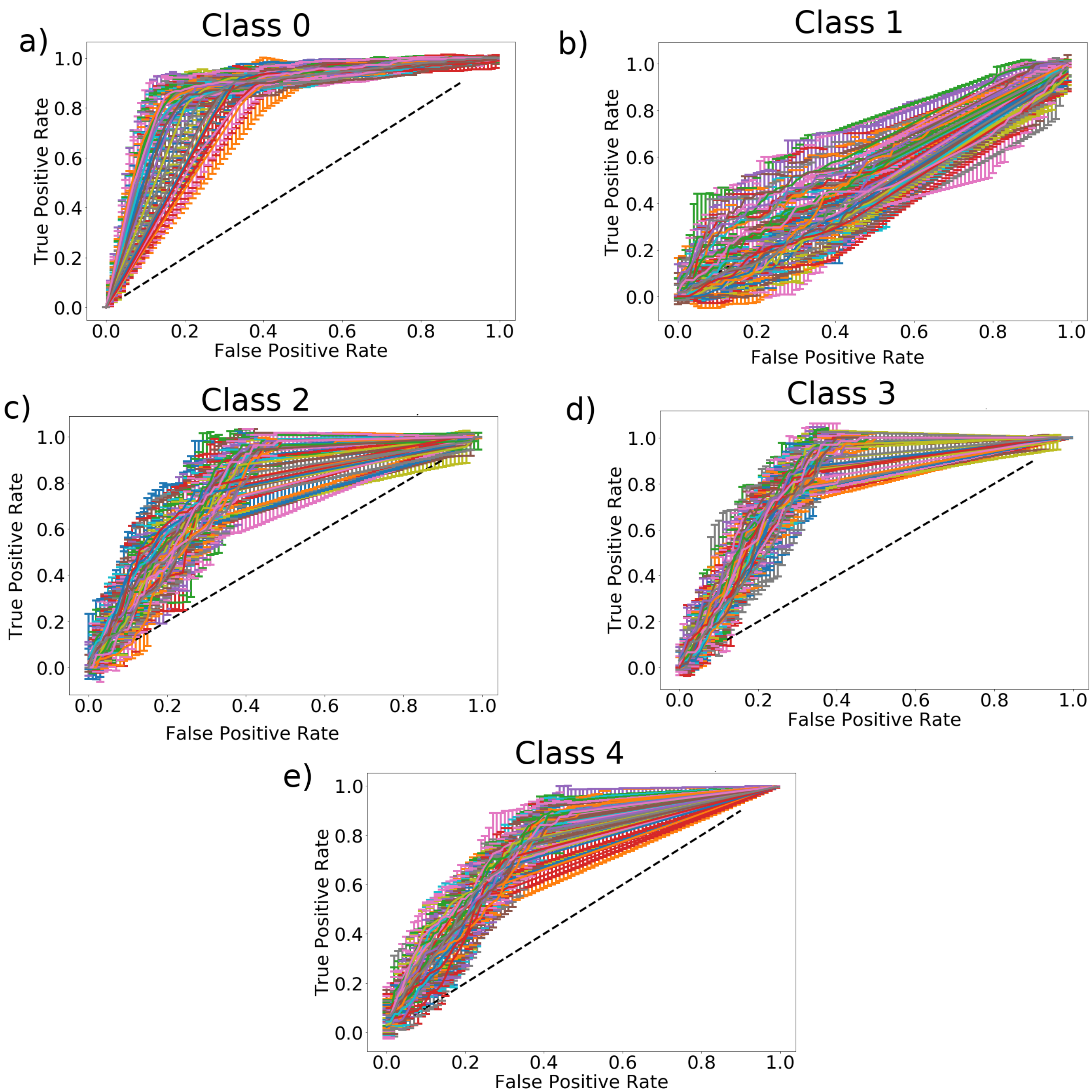

Figure S5: ROC curves for classes 0-4 resulting from classifiers trained on top fragments produced by a single iteration of the Hunting FOX algorithm. Error bars indicate standard deviation across the five disjoint training sets. Different colors indicate different runs. The black dashed line along $y = x$ is the ROC curve of a random classifier.

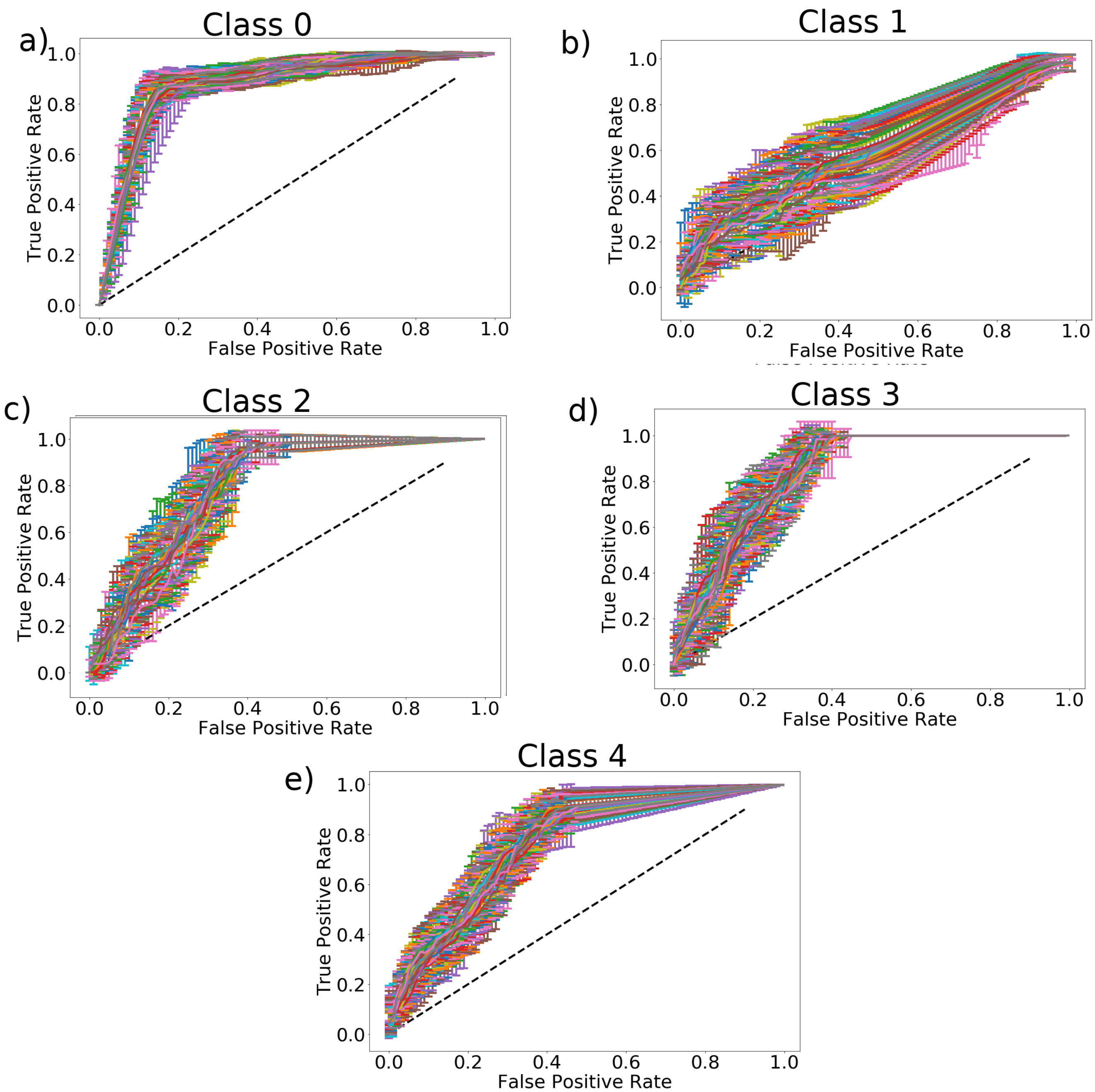

Figure S6: ROC curves for classes 0-4 resulting from classifiers trained on top 9 fragments displayed in Fig. 3. Error bars indicate standard deviation across the five disjoint training sets. Different colors indicate different runs. The black dashed line along $y = x$ is the ROC curve of a random classifier.

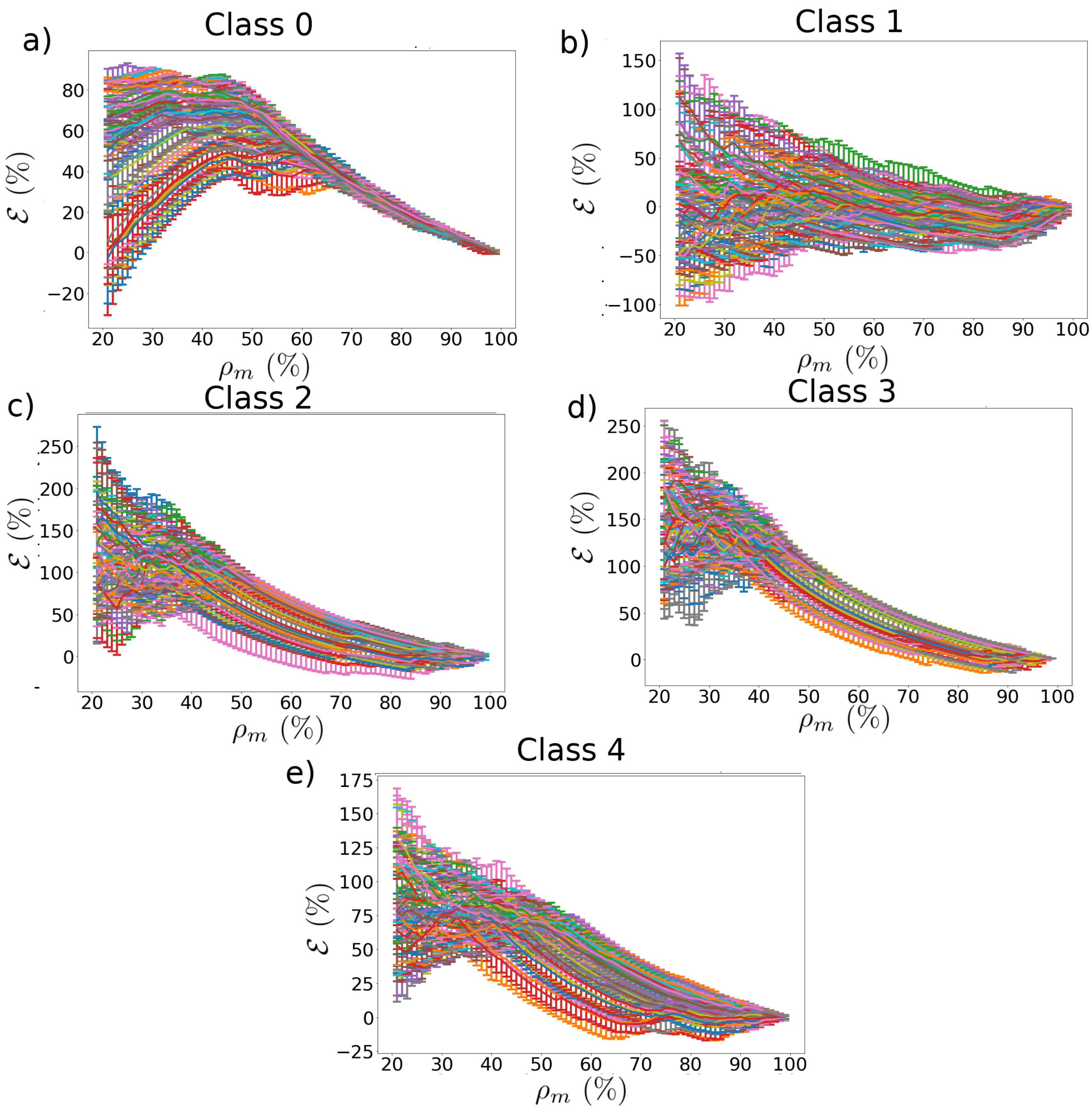

Figure S7: Enrichment versus percent ranking for classes 0-4 resulting from classifiers trained on top fragments produced by a single iteration of the Hunting FOX algorithm. Error bars indicate standard deviation across the five disjoint training sets. Different colors indicate different runs.

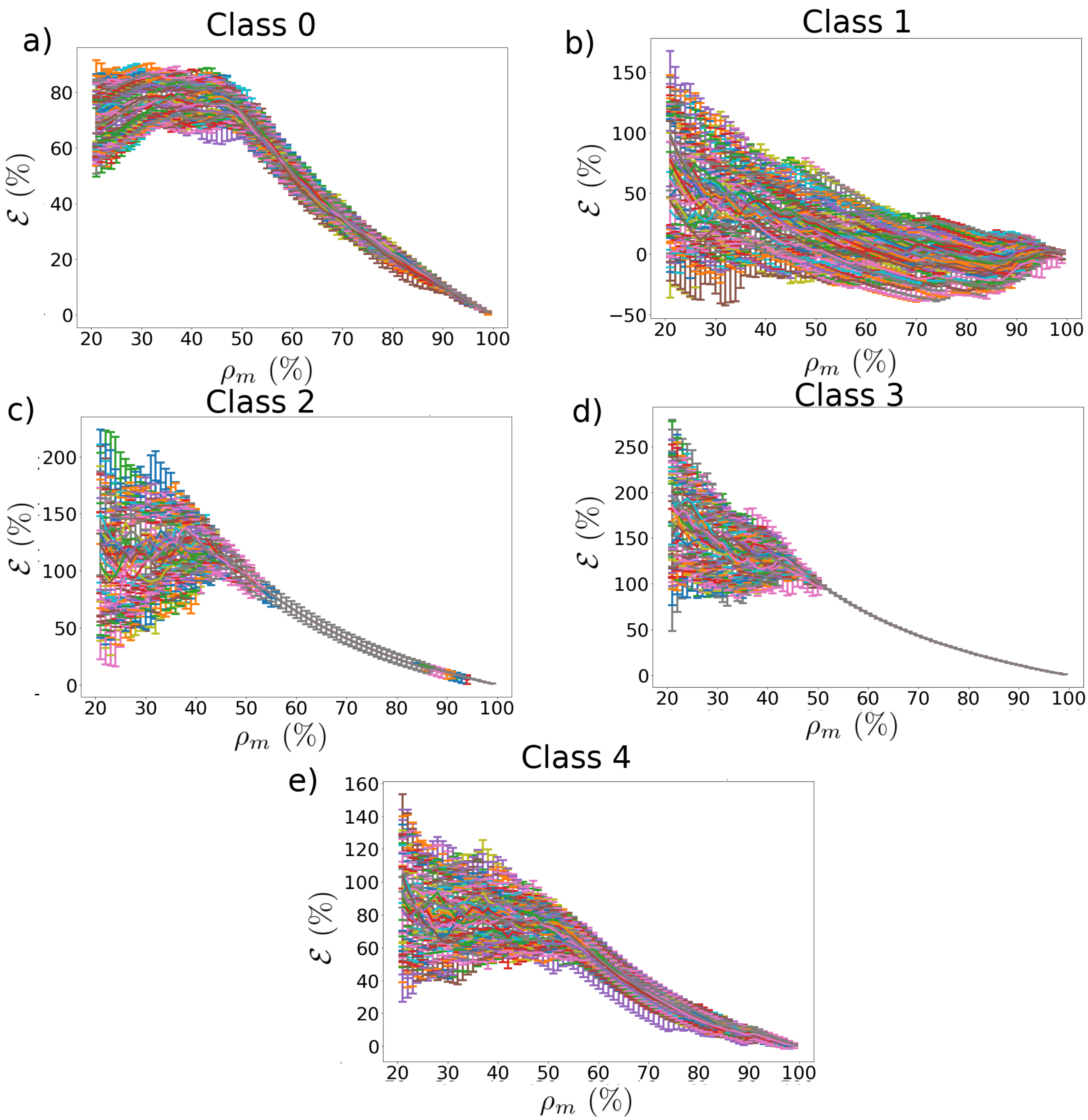

Figure S8: Enrichment versus percent ranking for classes 0-4 resulting from classifiers trained on top 9 fragments displayed in Fig. 3. Error bars indicate standard deviation across the five disjoint training sets. Different colors indicate different runs.

References

(1) Hohman, M.; Gregory, K.; Chibale, K.; Smith, P. J.; Ekins, S.; Bunin, B. Novel web-based tools combining chemistry informatics, biology and social networks for drug discovery. *Drug Discovery Today* **2009**, *14*, 261–270.

(2) Pedregosa, F. et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

(3) Balakumar, B.; Hastie, T.; Friedman, J.; Tibshirani, R.; Simon, N. Glmnet for Python. 2016; http:///www.stanford.edu/~hastie/glmnet_python/.

(4) Garcia, A. E.; Herce, H.; Paschek, D. Chapter 5 Simulations of Temperature and Pressure Unfolding of Peptides and Proteins with Replica Exchange Molecular Dynamics. *Annu. Rep. Comput. Chem.* **2006**, *2*, 83–95.

(5) Laio, A.; Gervasio, F. L. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Rep. Prog. Phys.* **2008**, *71*, 126601.

(6) Barducci, A.; Bonomi, M.; Parrinello, M. Metadynamics. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1*, 826–843.

(7) Sutto, L.; Marsili, S.; Gervasio, F. L. New advances in metadynamics. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 771–779.

(8) Abrams, C.; Bussi, G. Enhanced Sampling in Molecular Dynamics Using Metadynamics, Replica-Exchange, and Temperature-Acceleration. *Entropy* **2013**, *16*, 163–199.

(9) Bussi, G.; Gervasio, F. L.; Laio, A.; Parrinello, M. Free-Energy Landscape for β Hairpin Folding from Combined Parallel Tempering and Metadynamics. *J. Am. Chem. Soc.* **2006**, *128*, 13435–13441.

(10) Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient Estimation of Word Representations in Vector Space. *arXiv preprint arXiv:1301.3781* **2013**, 1–12.

(11) Narayanan, A.; Chandramohan, M.; Venkatesan, R.; Chen, L.; Liu, Y.; Jaiswal, S. graph2vec: Learning Distributed Representations of Graphs. **2017**,

(12) 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.; Aspuru-Guzik, A. Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. *ACS Cent. Sci.* **2018**, *4*.

(13) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. The MARTINI force field: Coarse grained model for biomolecular simulations. *J. Phys. Chem. B* **2007**, *111*, 7812–24.

(14) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.