Applying Support-Vector Machine Learning Algorithms towards Predicting Host-Guest Interactions with Cucurbit[7]uril

Anthony Tabet,[a,b,c,‡] Thomas Gebhart,[d,‡] Guanglu Wu,[a] Charlie Readman,[a] Merrick Pierson Smela,[e] Vijay K. Rana,[a] Cole Baker,[f] Harry Bulstrode,[b] Polina Anikeeva,[c] David H. Rowitch,[b] and Oren A. Scherman*[,a]

[a]Melville Laboratory for Polymer Synthesis, Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, UK
[b]Department of Paediatrics, Addenbrooke's Hospital, University of Cambridge, Cambridge CB2 0QQ, UK
[c]Department of Materials Science & Engineering and Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA USA
[d]Department of Computer Science, University of Minnesota, Minneapolis, MN USA
[e]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA USA
[f]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA USA
‡Authors contributed equally
Email: oas23@cam.ac.uk

S.1 Materials and Methodology

Chemicals and Reagents

DMEM/F12 cell culture media and other culture reagents were purchased from Invitrogen and sterile filtered prior to use. Selumetinib was purchased from Selleck Chemicals. TAK-580 was a gift from Prof. Charles Stiles (Harvard). All deuterated solvents were purchased from Sigma Aldrich.

Density Functional Theory Calculations

Gas phase geometry optimizations and electric field gradient calculations were performed using density functional theory (DFT) with the hybrid B3LYP (Becke, 1993; Stephens, 1994) exchange-correlation energy functional (restricted) in combination with the Pople-type

6- 311++G** basis set for all atoms. Starting structures were first optimized with molecular dynamics (MM2) using Chem3D 16.0 software. All DFT computations were carried out using the Gaussian '09 *ab initio* program package (Frisch, 2009).

Kernel Method

Given some space $\mathcal{X}$, a *kernel* is a two-argument, real-valued function $\kappa : \mathcal{X} \times \mathcal{X} \to R$ such that for any $\mathbf{x}, \mathbf{y} \in \mathcal{X}$,

$$\kappa(\mathbf{x}, \mathbf{y}) = \langle \phi(\mathbf{x}), \phi(\mathbf{y}) \rangle_{\mathcal{V}} \tag{1}$$

for some inner-product space $\mathcal{V}$ such that $\forall \mathbf{x} \in \mathcal{X}$, $\phi(\mathbf{x}) \in \mathcal{V}$. The *feature map* $\phi : \mathcal{X} \to \mathcal{V}$ need not take an explicit representation; it is sufficient to define the kernel function under the condition that $\mathcal{V}$ is an inner-product space. As we will see, a useful property of kernel functions is that they are closed under addition, multiplication, and scaling. In other words, given $\{\kappa_1, \kappa_2, \ldots, \kappa_n\}$ kernels on $\mathcal{X}$, $c_i > 0$, and $\mathbf{x}, \mathbf{y} \in \mathcal{X}$, the following are also valid kernel functions:

$$\kappa(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} c_i \kappa_i(\mathbf{x}, \mathbf{y}) \tag{2}$$

$$\kappa(\mathbf{x}, \mathbf{y}) = \prod_{i=1}^{n} c_i \kappa_i(\mathbf{x}, \mathbf{y}) \tag{3}$$

Note that all of the structural features within the data set are dependent on molecule-specific information like the number of atoms within the molecule or these atoms' electrical potential. Because of this, each feature may vary in dimensionality across samples. For features X3, X6, and X7 (corresponding to features (3), (6), and (7); see Table S1), we take the largest dimension across training samples as the common space. With a common space for each feature, we look to kernelize these features. X3, for example, is a square and positive matrix that describes electrostatic information condensed to each atom in each molecule and has a convenient graphical interpretation. Therefore we chose to kernelize these features via the

Laplacian Kernel:

$$\kappa_L = \exp(\gamma \|\mathbf{x}_i - \mathbf{x}_j\|_1) \tag{4}$$

where $\mathbf{x}_i, \mathbf{x}_j$ are sample i's and j's X3 matrix, respectively and γ is a free parameter. For X6, X7, and the environmental data, we kernelize using the Gaussian Kernel:

$$\kappa_G = \exp(\gamma \|\mathbf{x}_i - \mathbf{x}_j\|_2^2) \tag{5}$$

where $\mathbf{x}_i$ and $\mathbf{x}_j$ are the X6 or X7 features for samples i and j.

For the intrinsic features that contain spatial information (optimized orientation, SCF density for atom centers, and electrostatic properties in atom units), we transform these features to exploit this geometric prior in a multi-scale and orientation-independent manner. Treating each of these spatial features as a point cloud in three dimensions, we place an ϵ ball $B_p(\epsilon)$ around each point p in the space where ϵ controls the radius of the ball. We then allow these $B_p(\epsilon)$ balls to grow by increasing ϵ at the same rate across each point. As we increase epsilon, we look for non-trivial intersections of balls between points i.e. $B_p(\epsilon) \cap B_q(\epsilon) \neq \emptyset$ for all points p and q. If we find a non-trivial intersection, we connect these two points with an edge. As ϵ grows, the number of points connected will increase such that we have complex of connected points (Fig. 12). We track the holes (circles) created in this structure along with the range of ϵ values within which these holes exist. The most distinct holes within these structures are precisely those that exist over the longest range of ϵ values.

We can represent each hole that appears by a point in the two-dimensional plane with the first axis representing the ϵ value at which the hole was created (by the overlap of some ball and thus the addition of a new edge) and the second axis representing the (larger) ϵ value at which that hole was subsumed into a larger hole already in existence. This representation provides us with a quantitative view of the diameter of each structural hole of the molecule without the need to specify a common orientation across molecules. Let $D_i \in R^{l \times 2}$ denote

the (upper half-plane) vector representation of molecule i. We refer to this structure as a *diagram*. The number of holes l that appear throughout the range of ϵ varies from molecule to molecule, so choosing a common vector space in which to embed these diagrams would be lossy with respect to the underlying geometric information. Instead, we can kernelize these diagrams by projecting each point within the diagram onto a family of lines passing through the origin and comparing the distances projected across samples. This projection operation can be interpreted as an aggregate measure over the number and size of holes within them molecular structure. This combined view of the number and associated ϵ lifetimes of these holes provides an orientation-free basis for the width, height, and depth of the spatial molecular information.

Given $\theta \in R^2$ where $\|\theta\|_2 = 1$, let $L(\theta)$ denote the line $\{\lambda\theta : \lambda \in R\}$. Let $\pi_\theta : R^2 \to L(\theta)$ be the orthogonal projection onto $L(\theta)$ and $\pi_\Delta : R^2 \to \Delta$ where $\Delta = \{(x, x) \subset R^2\}$. For D_i, D_j two diagrams, let $\mu_i^\theta = \sum_{p \in D_i} \delta_{\phi_\theta(p)}$ and $\mu_{i\Delta}^\theta = \sum_{p \in D_i} \delta_{\phi_\theta \circ \phi_\Delta(p)}$. Similarly, let $\mu_j^\theta = \sum_{p \in D_j} \delta_{\phi_\theta(p)}$ and $\mu_{j\Delta}^\theta = \sum_{p \in D_j} \delta_{\phi_\theta \circ \phi_\Delta(p)}$. We can define the *Sliced Wasserstein distance* as:

$$\mathrm{SW}(D_i, D_j) = \frac{1}{2\pi} \int_{S^1} \mathcal{W}(\mu_i^\theta + \mu_{j\Delta}^\theta, \mu_j^\theta + \mu_{i\Delta}^\theta) d\theta \tag{6}$$

where

$$\mathcal{W}(\mu, \nu) = \inf_{P \in \Pi(\mu,\nu)} \int \int_{R \times R} |x - y| P(dx, dy) \tag{7}$$

for μ and ν nonnegative measures on R such that $|\mu| = \mu(R)$ and $|\nu| = \nu(R)$ are equal to the same number, and $\Pi(\mu, \nu)$ is the set of measures on R^2 with these marginals. This SW distance is conditionally negative definite on the space of diagrams, so we can define a valid kernel

$$\kappa_{\mathrm{SW}}(D_i, D_j) = \exp\left(-\frac{\mathrm{SW}(D_i, D_j)}{2\sigma^2}\right) \tag{8}$$

where σ is a free parameter called the *bandwidth*.

Support-Vector Machines (SVM)

Given training data $\{(\mathbf{x}_1, y_1), ..., (\mathbf{x}_n, y_n)\} \subset \mathcal{X} \times R$, SVM regression models look to find a function $f(\mathbf{x})$ that has at most ϵ deviation from the actual targets y. In other words, we assume f is of the form

$$f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b \tag{9}$$

where $\mathbf{w} \in R^{|\mathbf{x}|}$ and $b \in R$. To constrain this function to being as flat as possible, we seek a small $\mathbf{w}$ which is equivalent to minimizing $\|\mathbf{w}\|^2$. We do not assume our data may be perfectly approximated by such a linear form, so we allow for some error by introducing slack variables ξ_i, ξ_i^* to cope with otherwise infeasible constraints in optimization with noise, leaving the following optimization problem

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \\ \text{subject to} \quad & \begin{cases} y_i - \langle \mathbf{w}, \mathbf{x}_i \rangle - b & \leq \epsilon + \xi_i \\ \langle \mathbf{w}, \mathbf{x}_i \rangle + b - y_i & \leq \epsilon + \xi_i^* \\ \xi_i \xi_i^* & \leq 0 \end{cases} \end{aligned} \tag{10}$$

The constant $C > 0$ determines the trade-off between the flatness of f and the amount up to which deviations larger than ϵ are tolerated. The primal formulation in Equation 10 is often easier to solve through the dual.

We define the Lagrangian as

$$\begin{aligned} L =& \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*) - \sum_{i=1}^{l}(\eta_i\xi_i + \eta_i^*\xi_i^*) \\ & - \sum_{i=1}^{l}\alpha_i(\epsilon + \xi_i - y_i + \langle \mathbf{w}, \mathbf{x}_i\rangle + b) \\ & - \sum_{i=1}^{l}\alpha_i^*(\epsilon + \xi_i^* + y_i - \langle \mathbf{w}, \mathbf{x}_i\rangle - b) \end{aligned} \tag{11}$$

where $\eta_i, \eta_i^*, \alpha_i, \alpha_i^* \geq 0$ are Lagrangian multipliers.

This dual function has a saddle point with respect to the primal and dual variables at the solution, so the partial derivatives of L with respect to the primal variables will go to zero in optimality:

$$\begin{aligned} \frac{\partial L}{\partial b} &= \sum_{i=1}^{l}(\alpha_i^* - \alpha_i) = 0 \\ \frac{\partial L}{\partial \mathbf{w}} &= \mathbf{w} - \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)\mathbf{x}_i = 0 \\ \frac{\partial L}{\partial \xi_i} &= C - \alpha_i - \eta_i = 0 \\ \frac{\partial L}{\partial \xi_i^*} &= C - \alpha_i^* - \eta_i^* = 0 \end{aligned} \tag{12}$$

Substituting Equations 12 into Equation 11, we arrive at the dual optimization problem

$$\begin{aligned} \text{maximize} \quad & \begin{cases} -\frac{1}{2}\sum_{i,j=1}^{n}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\langle \mathbf{x}_i, \mathbf{x}_j\rangle \\ -\epsilon\sum_{i=1}^{n}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{n} y_i(\alpha_i - \alpha_i^*) \end{cases} \\ \text{subject to} \quad & \sum_{i=1}^{n}(\alpha_i - \alpha_i^*) = 0 \text{ and } \alpha_i, \alpha_i^* \in [0, C] \end{aligned} \tag{13}$$

Here the α_i and α_i^* are variables from the Lagrangian. The representation of f now may be

γ	γ_3	γ_6	ϵ	C	$\lvert\theta\rvert$	σ
1.00E-04	1.00E-05	1.00E-05	1.00E-04	10.0	1	1.0

Optimal hyperparameters from grid search and 5-fold cross-validation over training set

given equivalently as

$$f(\mathbf{x}) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)\langle \mathbf{x}_i, \mathbf{x}\rangle + b. \tag{14}$$

and

$$\mathbf{w} = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)\mathbf{x}_i \tag{15}$$

Note $\mathbf{w}$ is completely described as a linear combination of the training examples $\mathbf{x}_i$. Thus, the complexity of f's representation is independent of the dimensionality of $\mathcal{X}$ and depends only on the number of support vectors. Therefore, we may replace this inner product in the training data with *any* kernel function, giving

$$f(\mathbf{x}) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)\kappa(\mathbf{x}_i, \mathbf{x}) + b. \tag{16}$$

The optimization problem now corresponds to finding the flattest function in *feature space*, allowing the model to approximate an f that is linear in this inner product, but nonlinear with respect to the input space. This transition to a kernelized representation of the input results in a model with much better expressitivity than one with an objective to finding an optimal subspace representation of the input itself. This is especially important in a domain with limited training samples.

1D and 2D NMR

^{1}H NMR, ^{13}C NMR, long-range COSY, HSQC, and HMBC spectra were acquired in a D_2O/DMSO-d6 solution at 298 K, and recorded on a Bruker AVANCE 500 with TCI Cryoprobe system (500 MHz) being controlled by TopSpin2. Pre-saturation of the residual solvent signal was utilized for ^{1}H NMR.

Isothermal Titration Calorimetry

All ITC experiments were carried out on a Microcal ITC200 at 298.15 K in water/DMSO-d6 (2%) solutions. A solution of CB[7] was held in the sample cell, and drug (MLN2480) was held in the injection syringe. The concentration of CB[7] was calibrated by ITC titration with a standard aqueous solution of 1-adamantanamine (ADA). The concentration of drug was calibrated by NMR using 3-(Trimethylsilyl)-1-propanesulfonic acid sodium salt (DSS) as internal standard in 90% H_2O and 10% D_2O. A baseline correction of CB[7] in water/DMSO-d6 was used to offset any heat of mixing. In order to avoid bias or potentially arbitrary offsets caused by manual adjustment of baseline, all raw data (thermograms) of ITC were integrated by NITPIC (v.1.2.2), fitted in Sedphat (v.12.1b), and visualized through GUSSI (v.1.2.1).

Confluence Curves

Time-resolved confluence curves were obtained. Murine v600e cells (Rowitch lab) were cultured in DMEM/F12 with L-Glutamine medium containing glucose (0.3%), penicillin/ streptomycin (50μg/mL), Apo-Transferrin (0.1mg/mL), Progesterone (20nM), sodium selenite (30nM), putrescine (60μM), insulin (25μg/mL), EGF (20ng/mL), and laminin (20ng/mL). Cells were cultured and passaged on laminin coated T25 flasks. Cells were plated onto laminin coated 96-well plates to generate confluence curves at 37 °C with 5% CO_2. These were generated using an IncuCyte® S3 Live Cell Analysis System. Percent confluence was calculated using Standard Analysis (n=4/well; 3 wells/condition).

Mass Transfer Model

We sought to develop a mathematical model for determining how binding of a guest with CB[7] affects the release kinetics from a depot. The below model assumes a spherical hydrogel depot with CB[7] covalently tethered to the matrix. A good guest G establishes equilibrium with bound and unbound states:

$$G + CB \rightleftharpoons GCB \tag{17}$$

Using a pseudo-steady state approximation, a shell balance outside the depot yields:

$$0 = (N * A)|_r - (N * A)|_{r+\Delta r} \tag{18}$$

$$K = \frac{[GCB]}{[CB]c_{surf}} \tag{19}$$

Where N is the flux, A is the surface area, and [ ] indicates concentration. [CB] is free CB[7]. For a sphere, $A = 4\pi r^2$. Dividing by $4\pi\Delta r$ and taking the limit as $\Delta r \to 0$:

$$\frac{d}{dr}\left(Nr^2\right) = 0 \tag{20}$$

Integrating this equation results in the following expression with integration constant α:

$$N = \frac{\alpha}{r^2} \tag{21}$$

Assuming a dilute concentration results in only a Fickian flux J^*:

$$N = xN_{total} + J^* \approx J^* = -D\frac{dc}{dr} \tag{22}$$

Where c is the concentration of G and D is the diffusivity.

Plugging in the flux and integrating yields the following expression:

$$c = \frac{\alpha}{Dr} + \beta \tag{23}$$

Where β is the integration constant. The boundary conditions are:

$$c(r = R) = c_{surf} \tag{24}$$

$$c(r \to \infty) = 0 \tag{25}$$

Where c_{surf} is the surface concentration of G and R is the radius of the depot. Plugging in these boundary conditions yield:

$$c = c_{surf}\left(\frac{R}{r}\right) \tag{26}$$

The flux at the surface is:

$$N = \frac{Dc_{surf}}{R} \tag{27}$$

The amount of solute released over time is:

$$\frac{d(G_{total}V)}{dt} = -NA \tag{28}$$

Rearranging the equilibrium constant:

$$c_{surf} = \frac{[GCB]}{[CB]K} \tag{29}$$

Mass balances on guest and CB[7] yield:

$$[GCB] + [CB] = [CB]_{total} \tag{30}$$

Where $[CB]_{total}$ is the amount of CB[7] added to the system in material synthesis and will be a tuned parameter in this study. Rearranging for c_{surf}:

$$c_{surf} = \frac{G_{total} - c_{surf}}{K([CB]_{total} - G_{total} + c_{surf})} \tag{31}$$

Solving for c_{surf}:

$$c_{surf} = \frac{\sqrt{G_{total}^2 K^2 - 2G_{total}K(K[CB]_{total} - 1) + (K[CB]_{total} + 1)^2} + KG_{total} - K[CB]_{total} - 1}{2K} \tag{32}$$

Plugging in for the rate equation:

$$\frac{d(G_{total})}{dt} = -\frac{ADc_{surf}}{VR} \tag{33}$$

The non-linear initial value problem set by the above two equations was solved using an explicit Euler method with an appropriate step size for numerical stability.

In order to verify the above technique, an analytical solution with no CB[7] was derived by setting the bulk concentration and surface concentration to be equivalent. The resulting differential equation is:

$$V\frac{dc}{dt} = -\frac{ADc}{R} \tag{34}$$

This results in the following analytical solution for the concentration in the depot:

$$\frac{c}{c_0} = exp\left(-\frac{3D}{R^2}t\right) \tag{35}$$

Or equivalently:

$$\bar{c} = exp\left(-\bar{t}\right) \tag{36}$$

Where the characteristic timescale τ is

$$\tau = \frac{R^2}{3D} \tag{37}$$

with c = c_0 $\bar{c}$ and t = τ $\bar{t}$.

S.2 Supporting Tables and Figures

Table S1: Summary of the type of data used to for algorithm training. DFT inputs are simulated data, and environmental inputs come from the literature.[4]

Training Data	
DFT Inputs	***Environmental Inputs***
geometry from optimized orientation (1)	method
occ./virt. eigenvalues condensed to atoms (2)	temperature
all electrons, condensed to atoms (3)	buffer & salt concentration
geometry accompanying electronic data (4)	pH
electrostatic properties (5)	counter-ion
electric field gradient (6)	
gradient eigenvalues (7)	

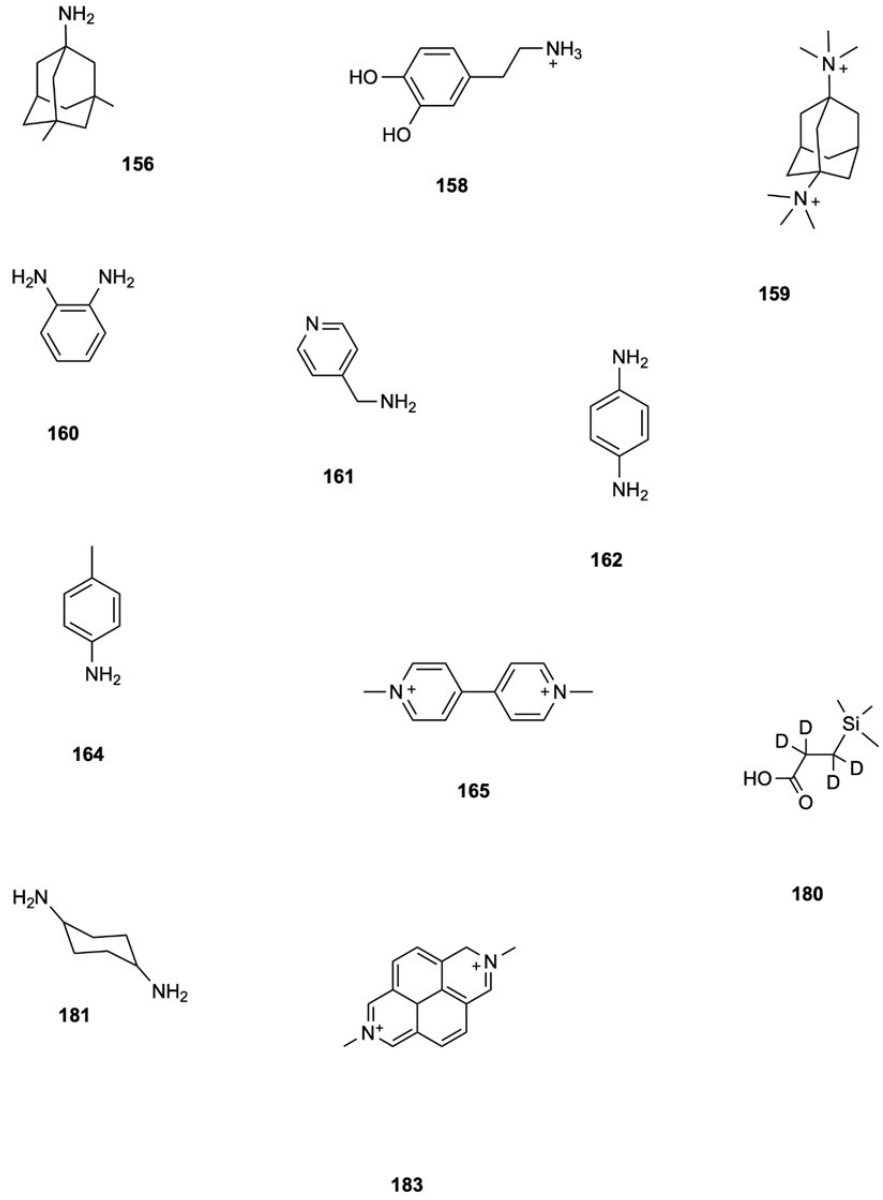

Figure S1: Library of molecules used as the training set. Numbers correspond to entries from Ref. 4.

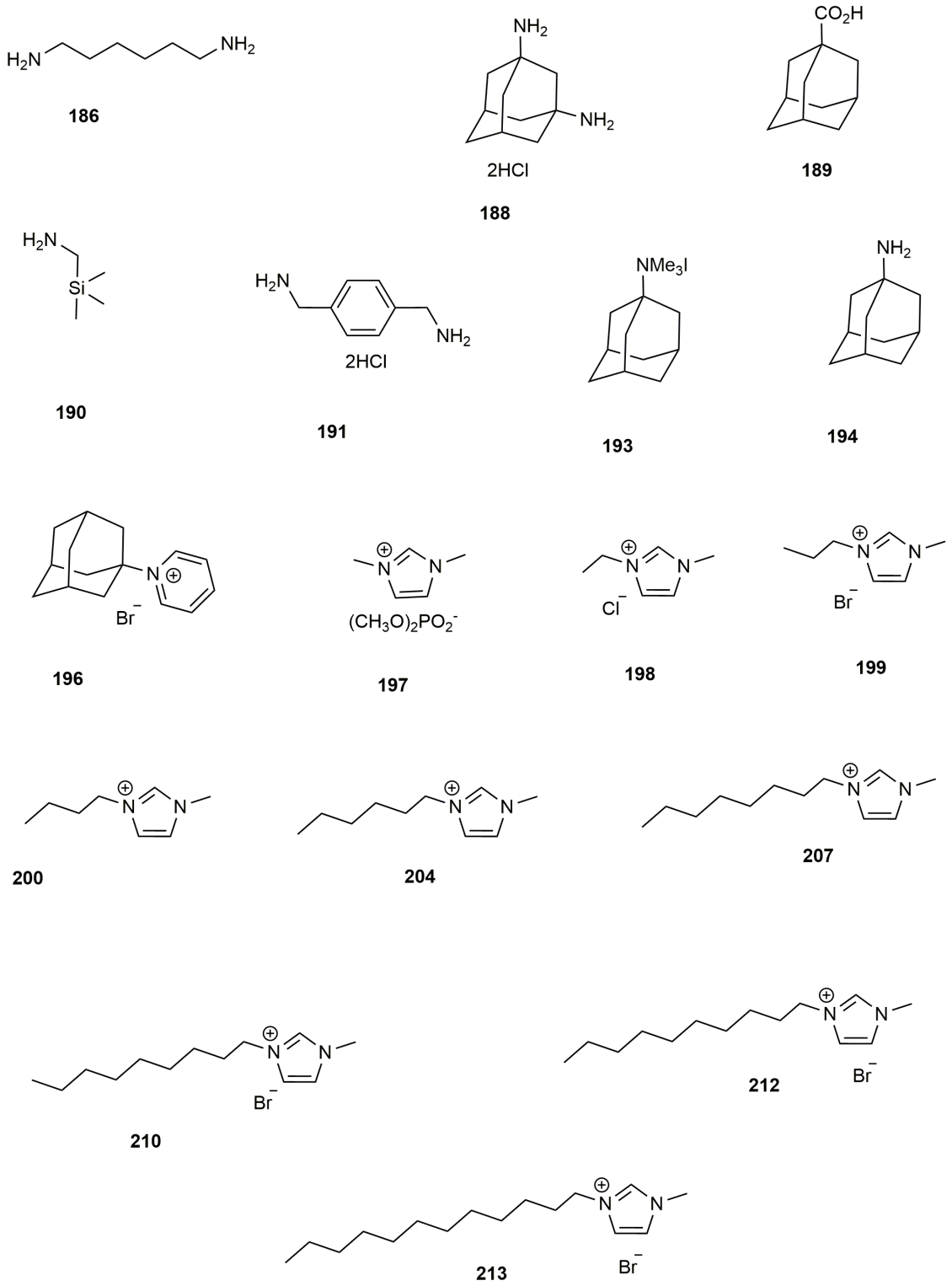

Figure S2: Library of molecules used as the training set. Numbers correspond to entries from Ref. 4.

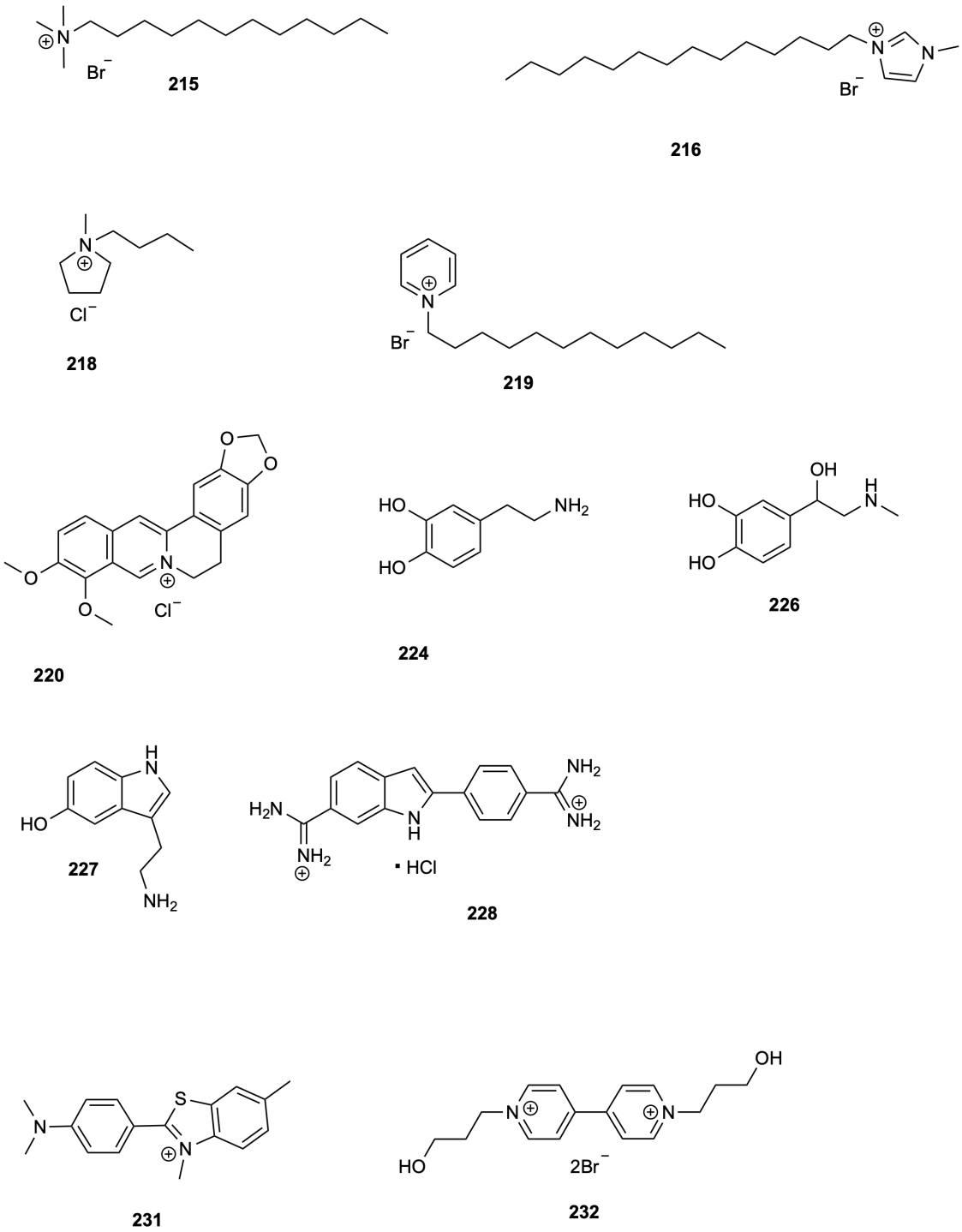

Figure S3: Library of molecules used as the training set. Numbers correspond to entries from Ref. 4. See pre-processing note for 228 and relevant controls in supporting data set.

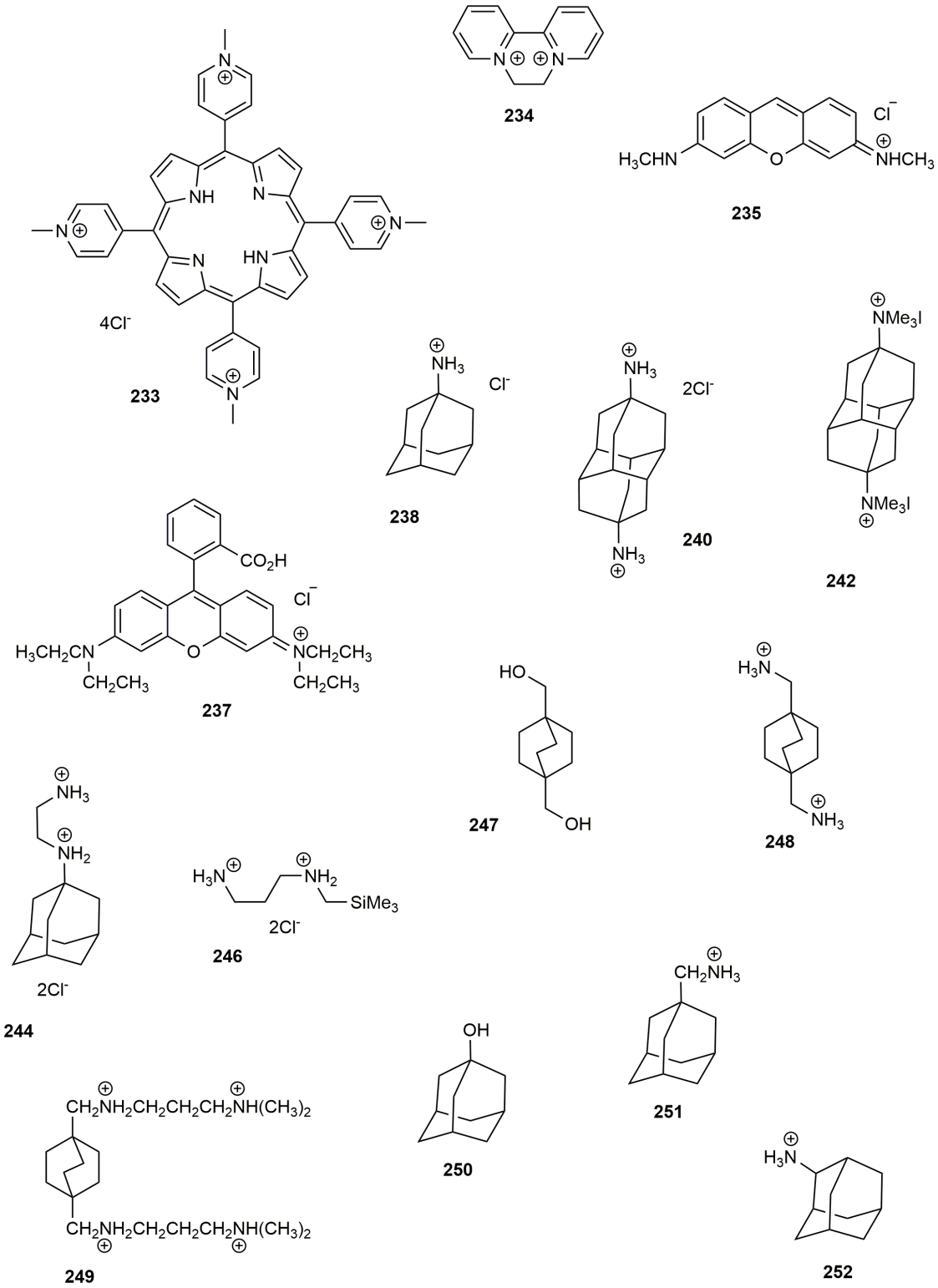

Figure S4: Library of molecules used as the training set. Numbers correspond to entries from Ref. 4.

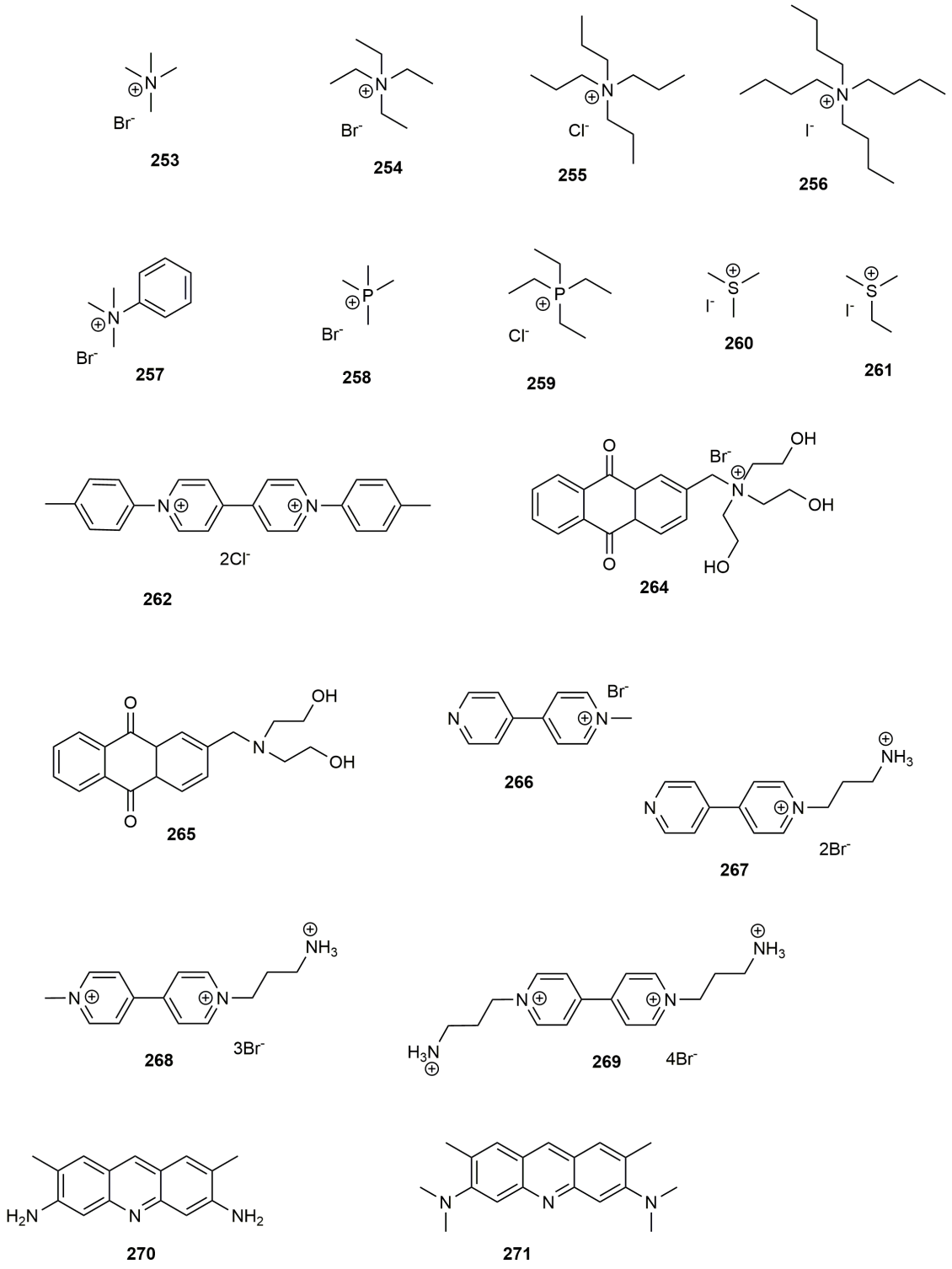

Figure S5: Library of molecules used as the training set. Numbers correspond to entries from Ref. 4.

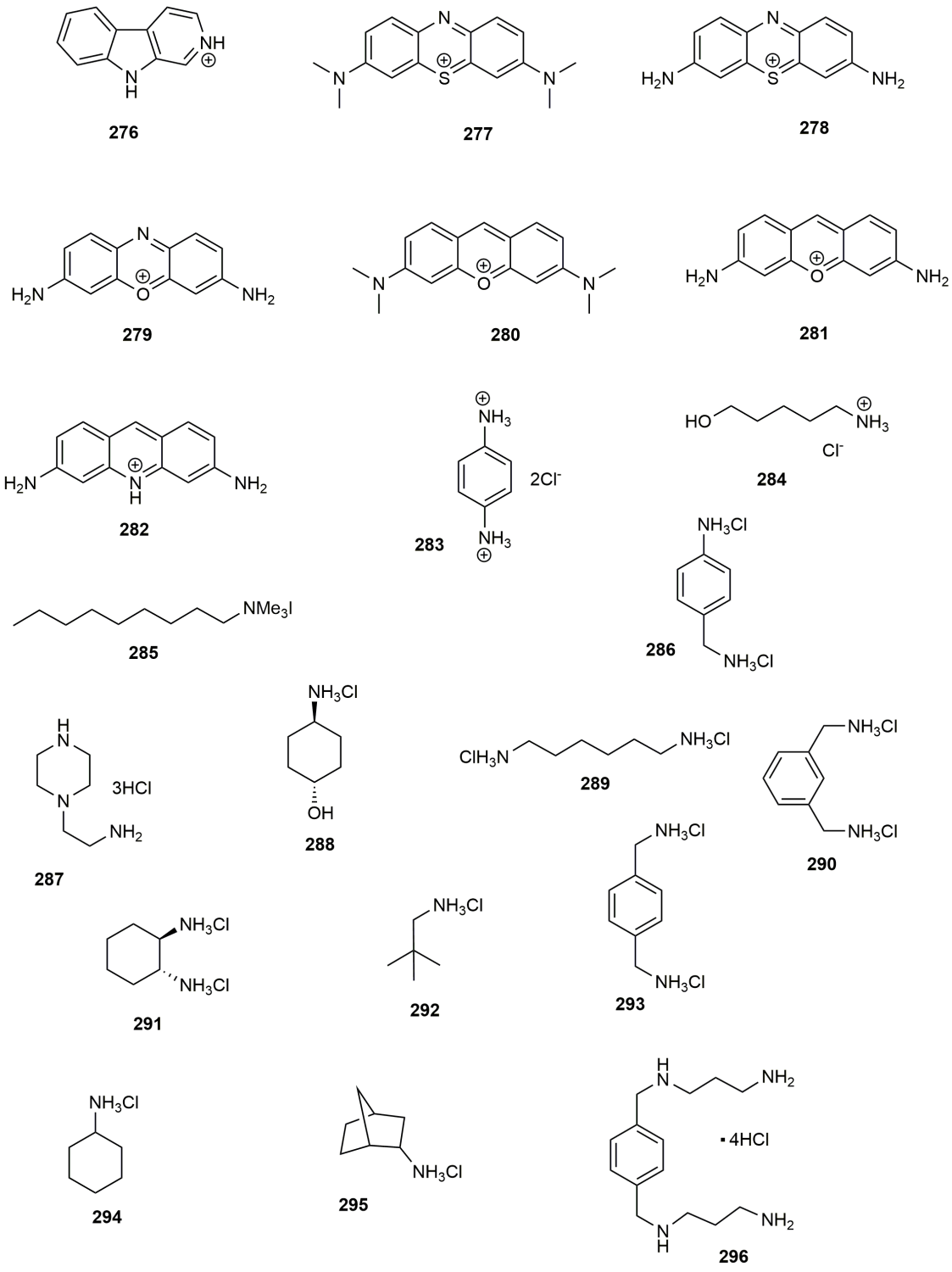

Figure S6: Library of molecules used as the training set. Numbers correspond to entries from Ref. 4.

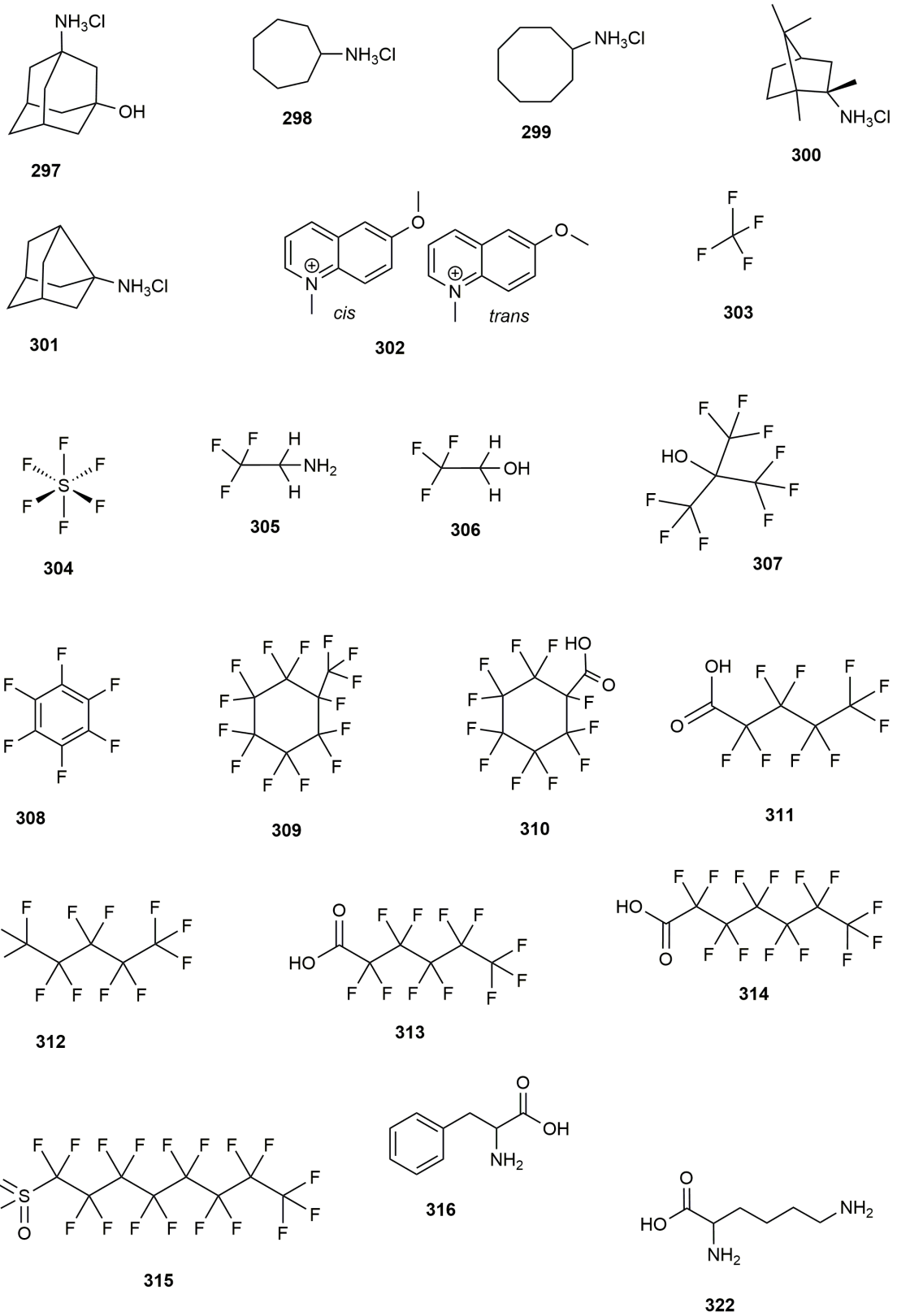

Figure S7: Library of molecules used as the training set. Numbers correspond to entries from Ref. 4.

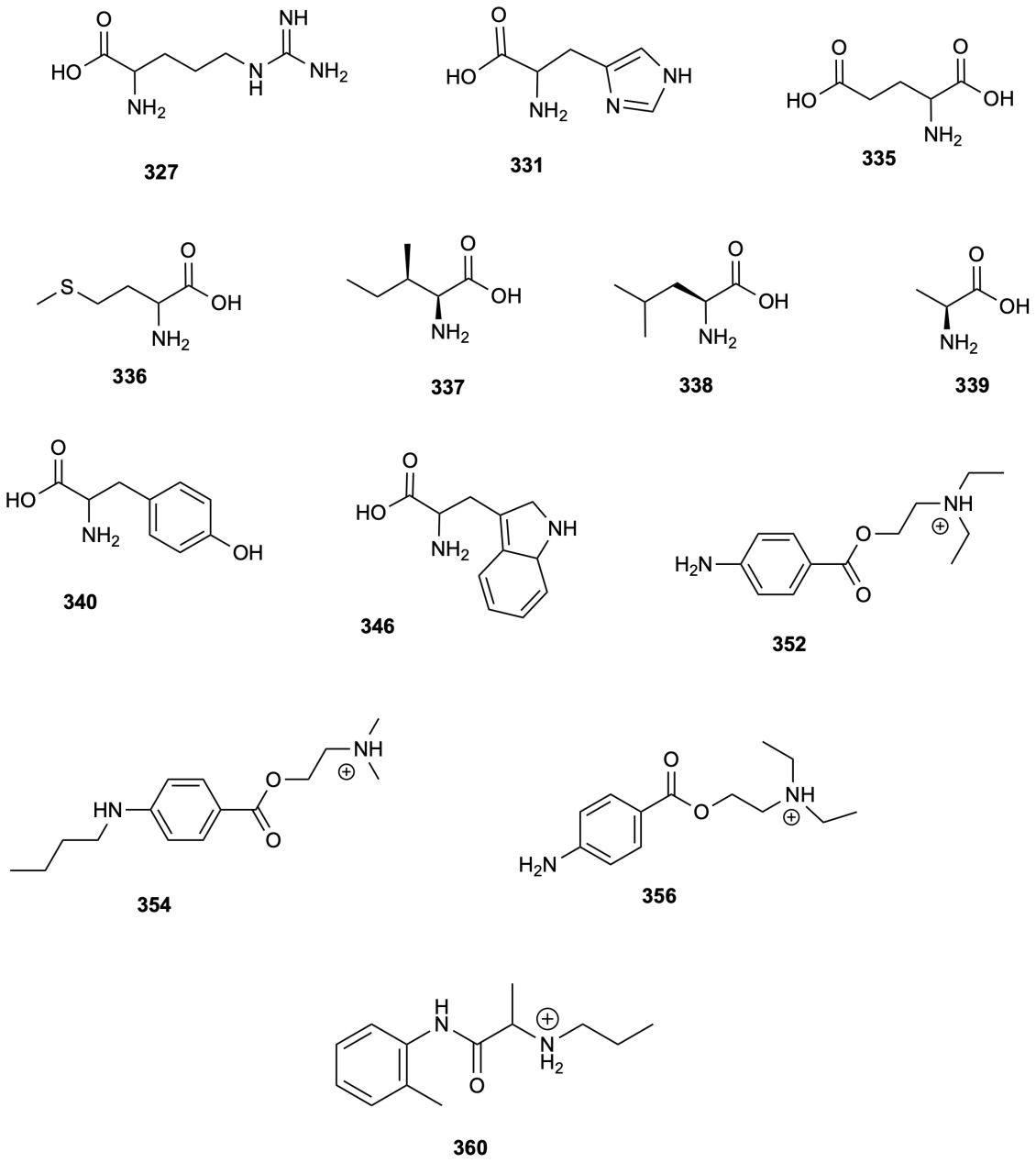

Figure S8: Library of molecules used as the training set. Numbers correspond to entries from Ref. 4.

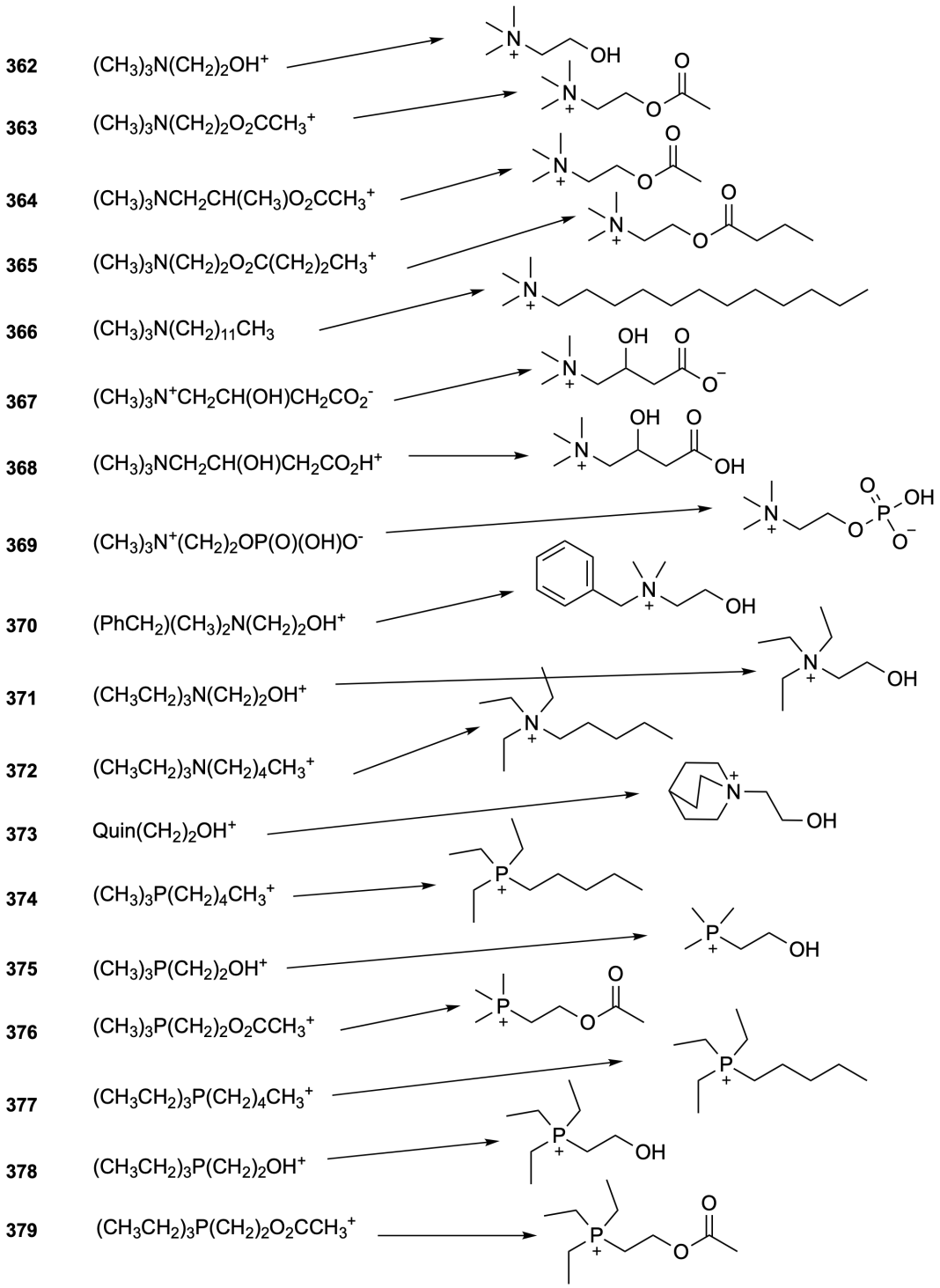

Figure S9: Library of molecules used as the training set. Numbers correspond to entries from Ref. 4.

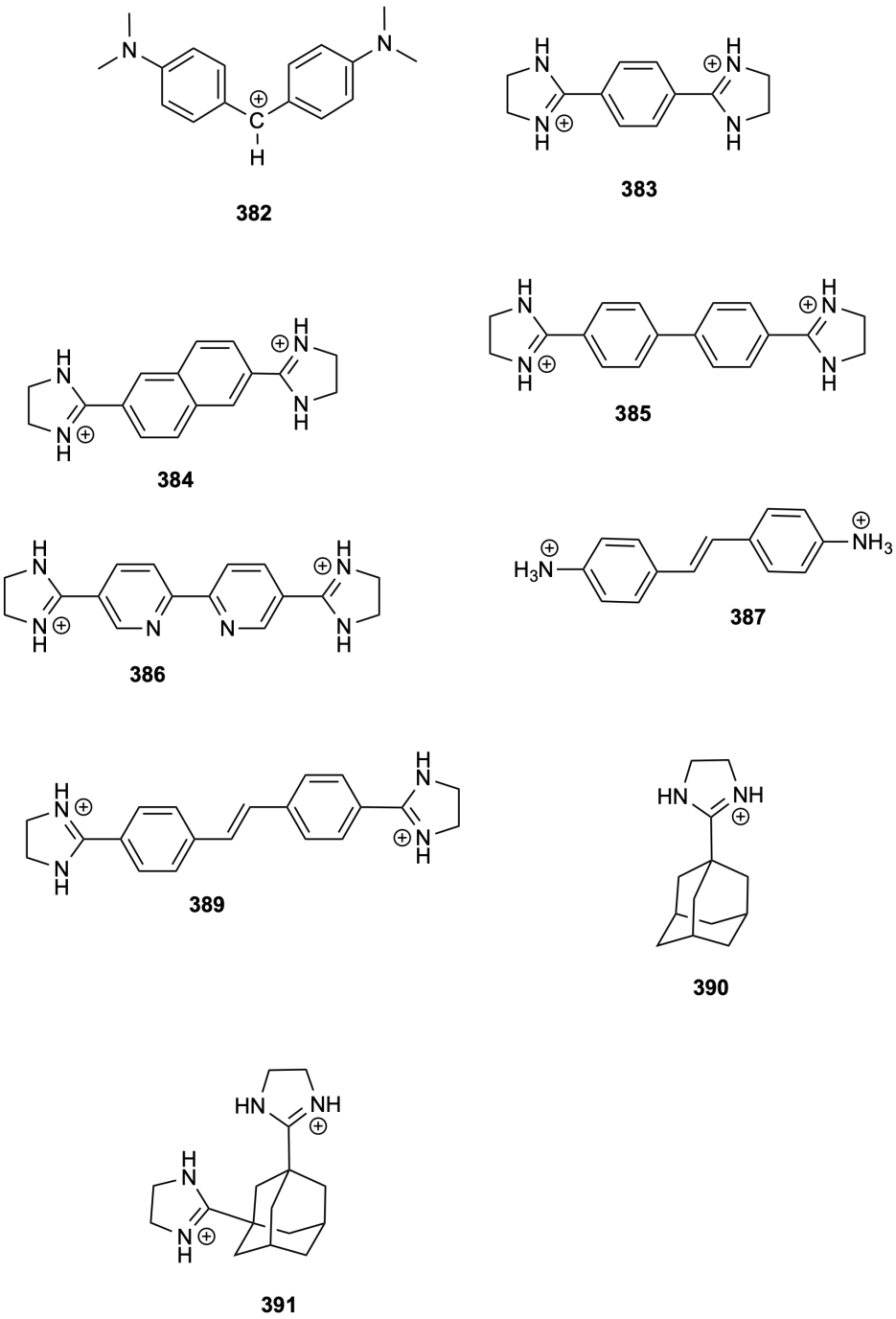

Figure S10: Library of molecules used as the training set. Numbers correspond to entries from Ref. 4. See pre-processing note for 391 and relevant controls in supporting data set.

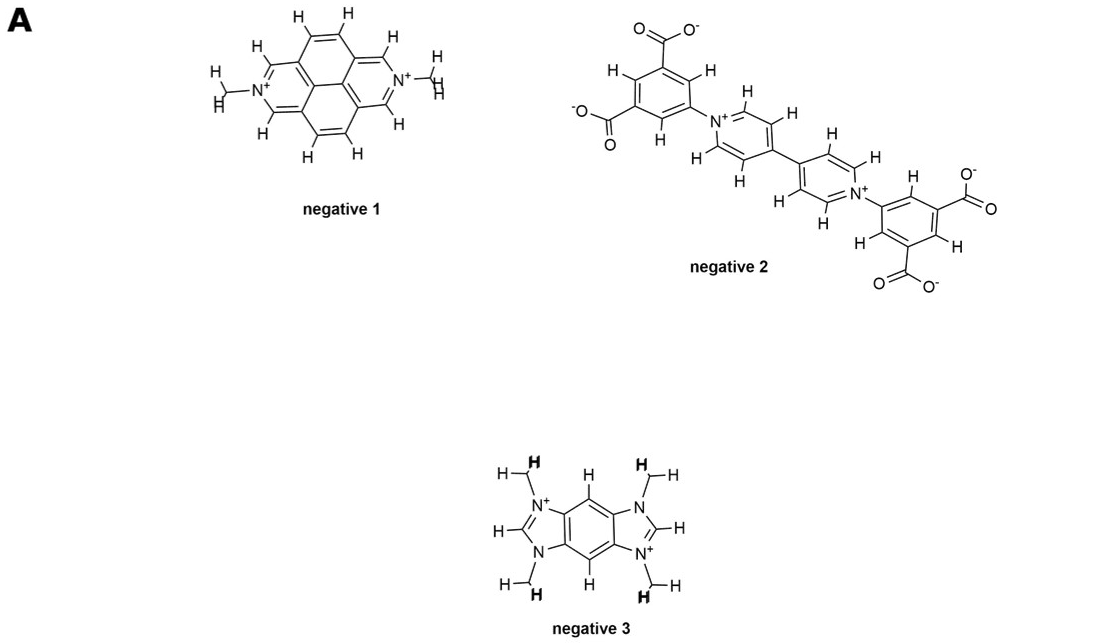

Figure S11: (A) Library of negative controls used as the training set. (B) Chemical structure of cucurbit[7]uril and cucurbit[8]uril.

Table S2: Complete list of family subdivisions used in model limitations studies corresponding to Fig. 4.

Family of molecules	***Molecules***
small arylamines	160, 162, 164, 286
viologen derivatives	165, 232, 266, 267, 268, 269
methylene blue derivatives	235, 270, 271, 277, 278, 279, 280, 281, 282
perfluorinated compounds	303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315
amino acids	316, 322, 327, 331, 335, 337, 338, 339, 340, 346
imidazolium derivatives	197, 198, 199, 200, 204, 207, 210, 212
adamantyl compounds	156, 188, 189, 193, 194, 196, 250, 251, 252, 297, 390, 391

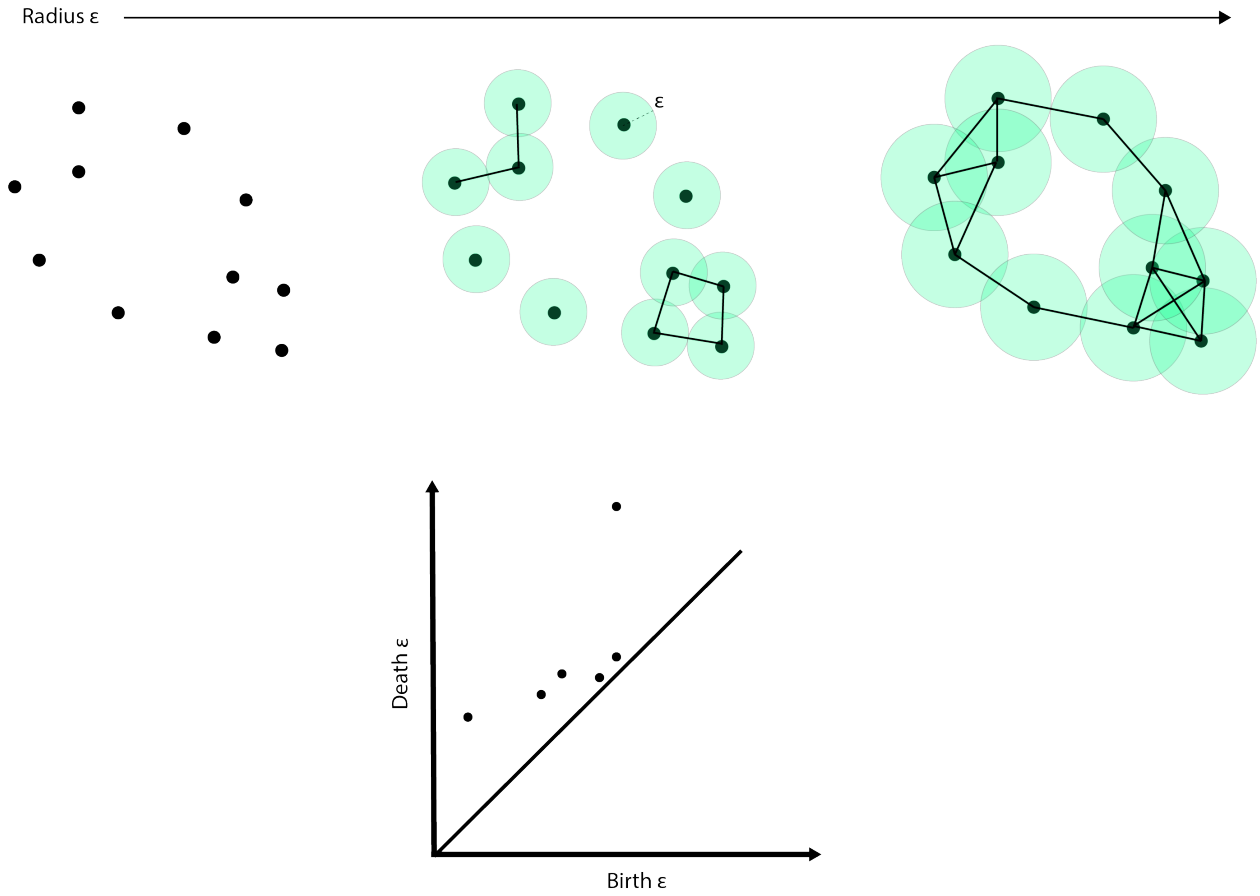

Figure S12: Visualization of a Rips filtration and the associated diagram, allowing the quantitative representation of spatial information in an orientation-independent manner.

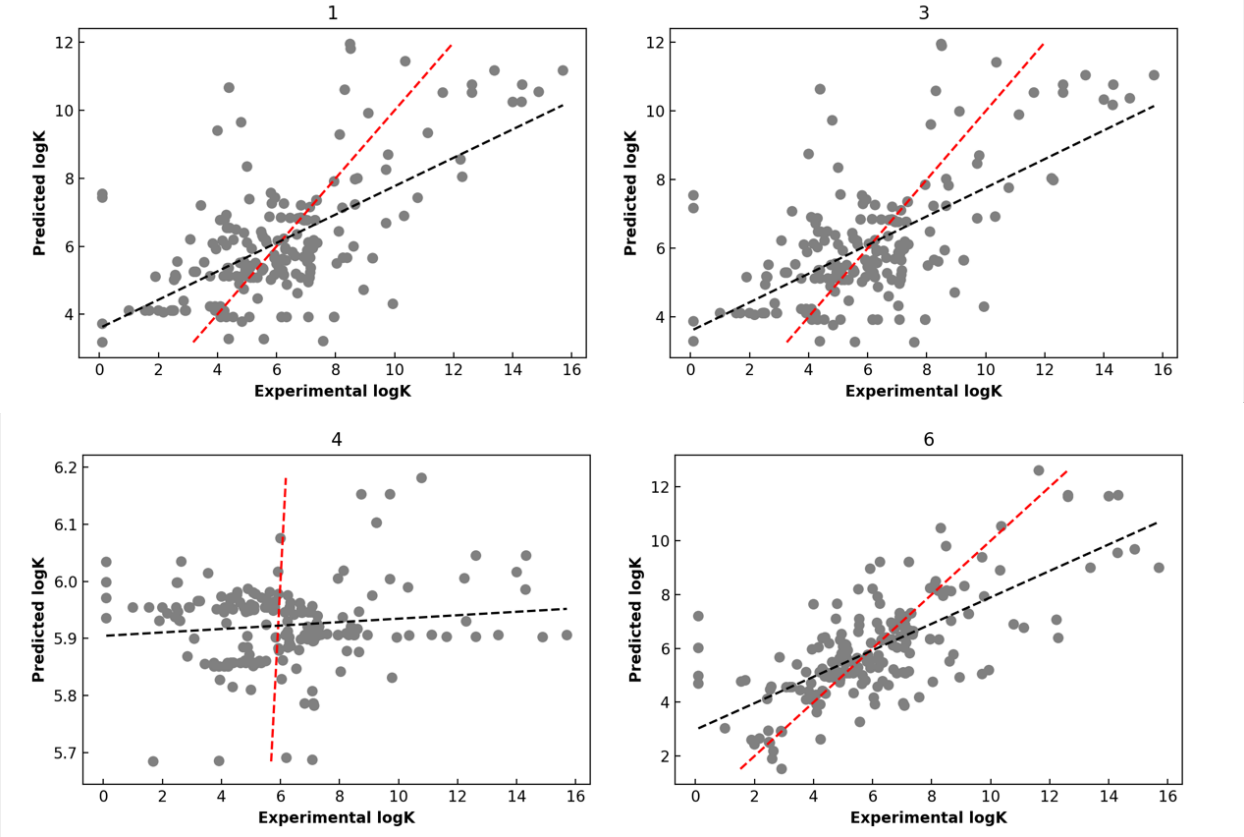

Figure S13: Model development and reduction. Leave one out experiments with different combinations of DFT and experimental parameters.

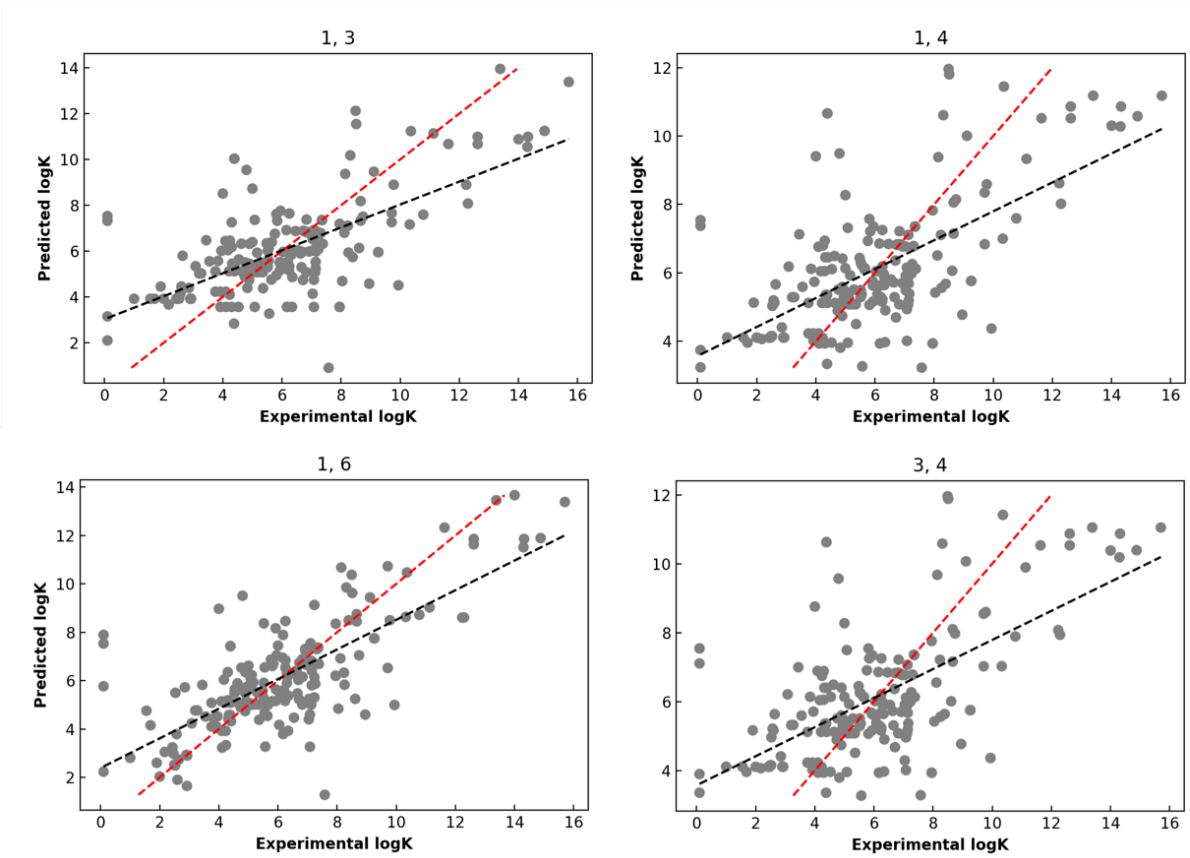

Figure S14: Model development and reduction. Leave one out experiments with different combinations of DFT and experimental parameters.

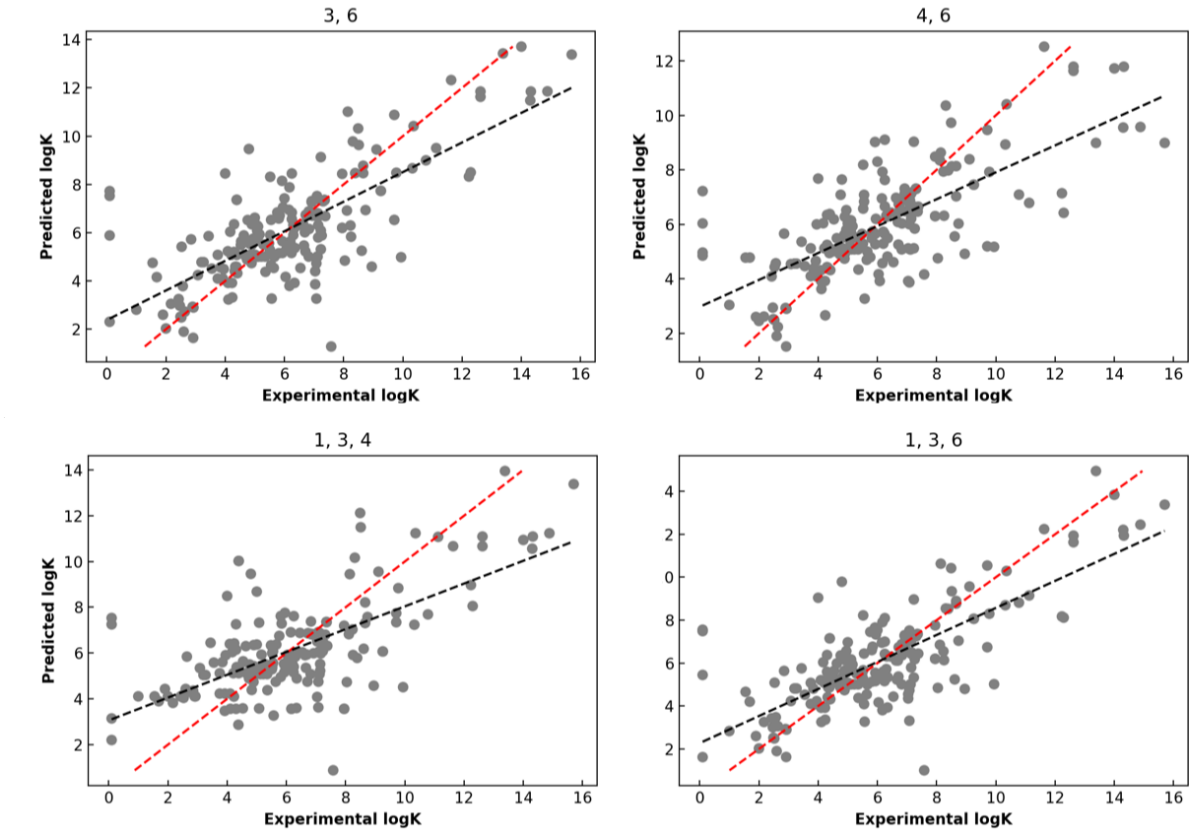

Figure S15: Model development and reduction. Leave one out experiments with different combinations of DFT and experimental parameters.

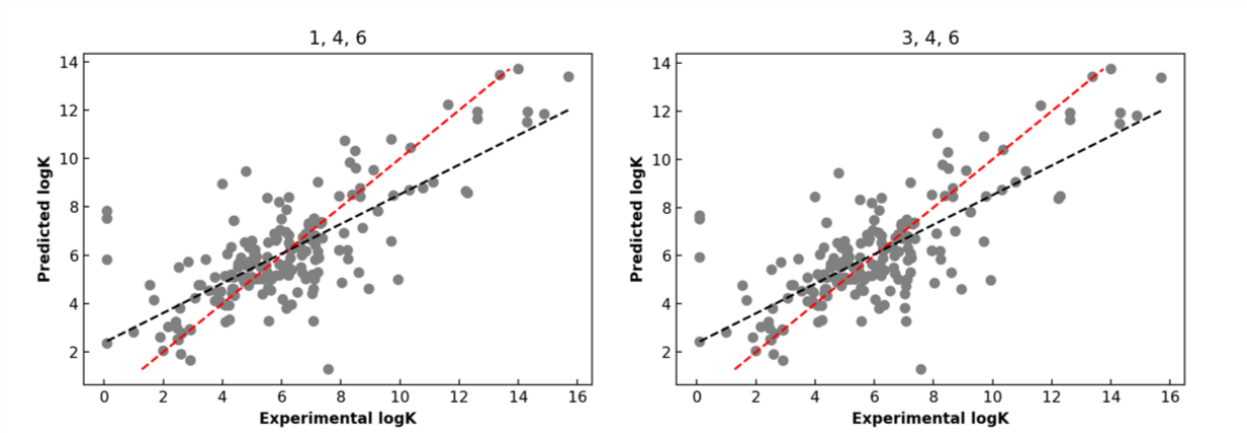

Figure S16: Model development and reduction. Leave one out experiments with different combinations of DFT and experimental parameters.

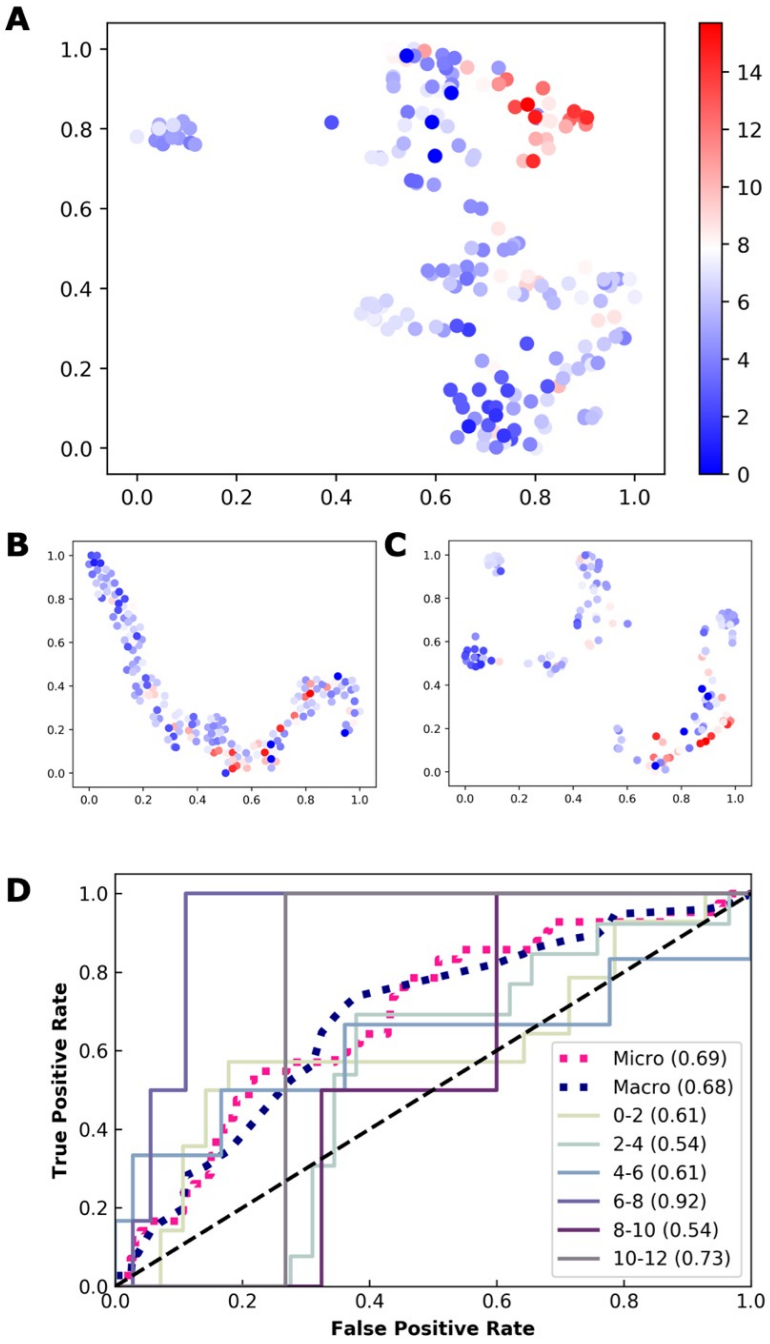

Figure S17: (A-C) Reduced dimensionality of feature sets via t-SNE[16] colored by the true binding affinity. **A**: Kernelized feature set given optimal training parameters. **B**: The un-kernelized features set for optimized DFT orientation. **C**: The kernelized feature set for optimized orientation using κ_{SW}. Note the better separation and clustering between extreme binding affinities after preprocessing optimized orientation. (D) Corresponding ROC-AUC plot from Figures 1-2.

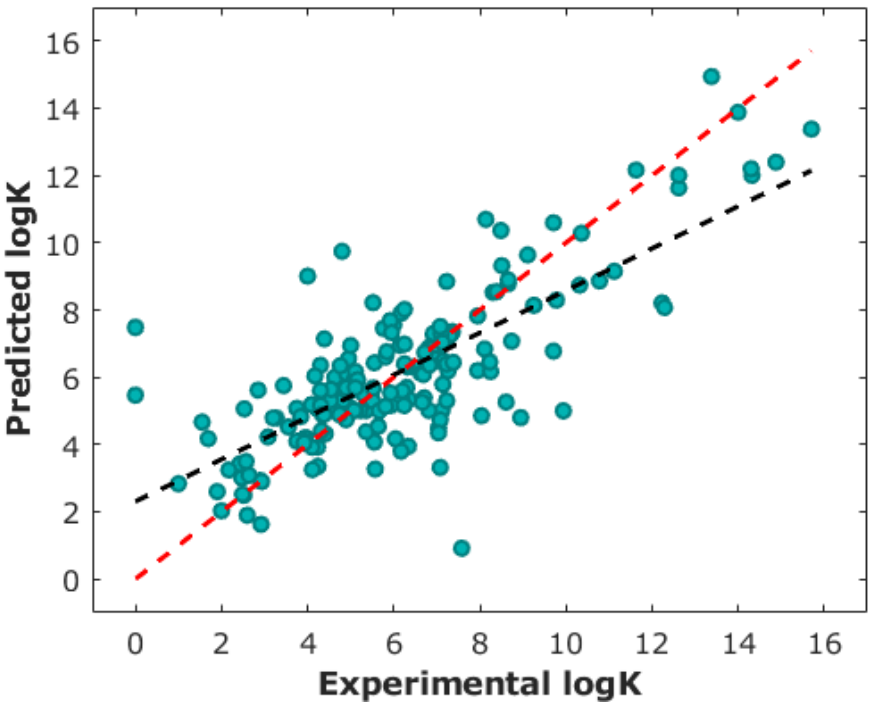

Figure S18: Prediction performance for optimal leave-one-out experiments when duplicates are considered. The line of best fit based on predicted points is shown in black, and the line representing perfect prediction is shown in red.

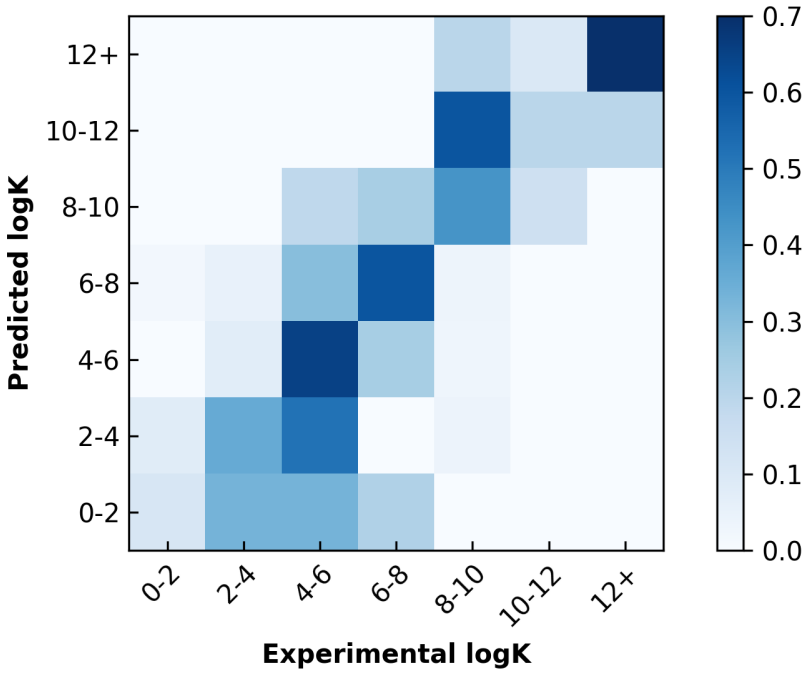

Figure S19: Normalized confusion matrix for the optimal SVM classifier when duplicates are considered.

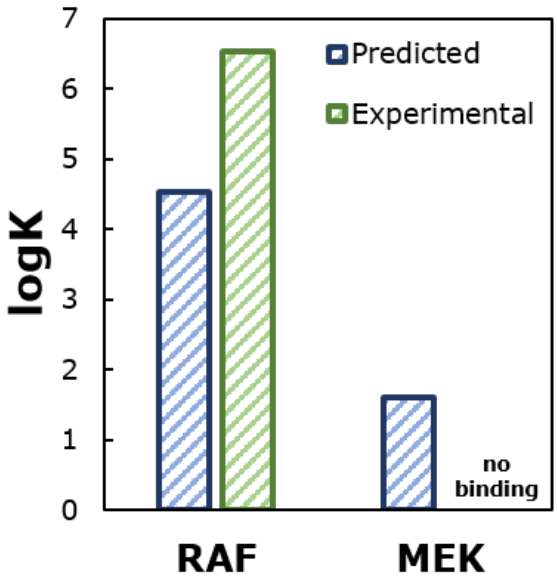

Figure S20: Predicted and experimental logK of RAF and MEK when duplicates are considered.

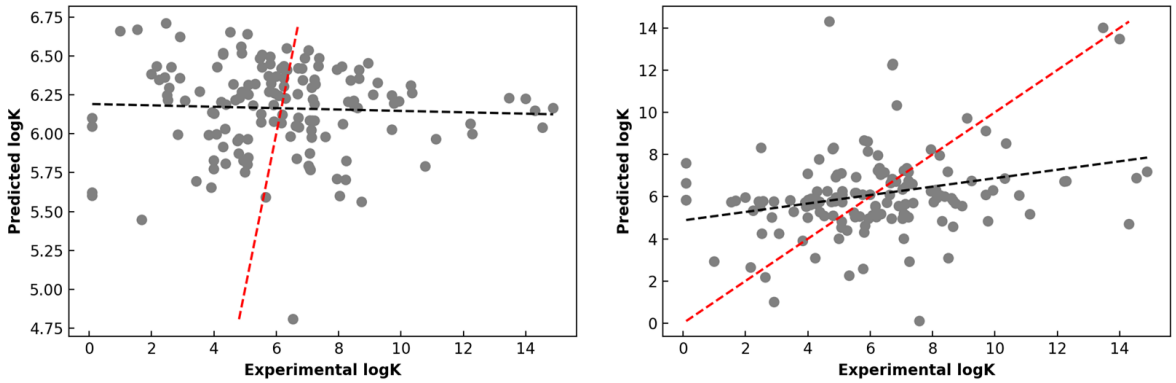

Figure S21: Prediction performance for optimal leave-one-out experiments considering the number of carbons. The line of best fit based on predicted points is shown in black, and the line representing perfect prediction is shown in red. Both the linear regression (left, score = 3.34) and kernelization and analysis in a SVM (right, score = 2.23) had poorer performance than any model considering structural data. For comparison with our main results, the R^2 of the SVM is 0.0769.

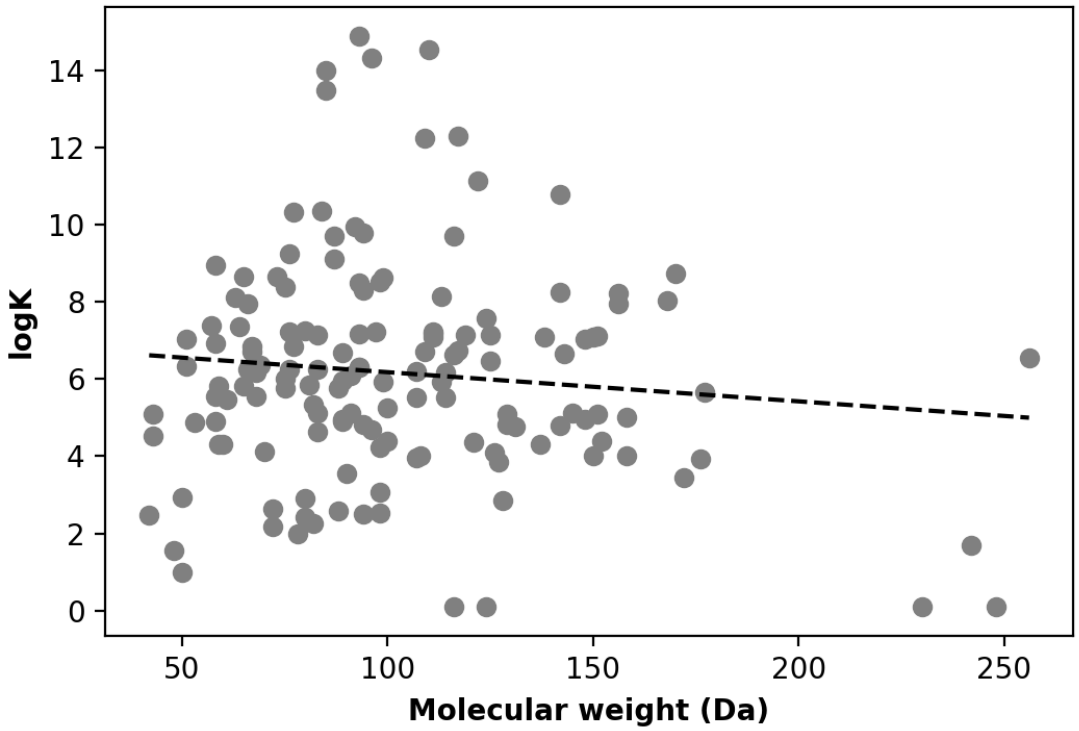

Figure S22: Distribution of molecular weights in the data set.

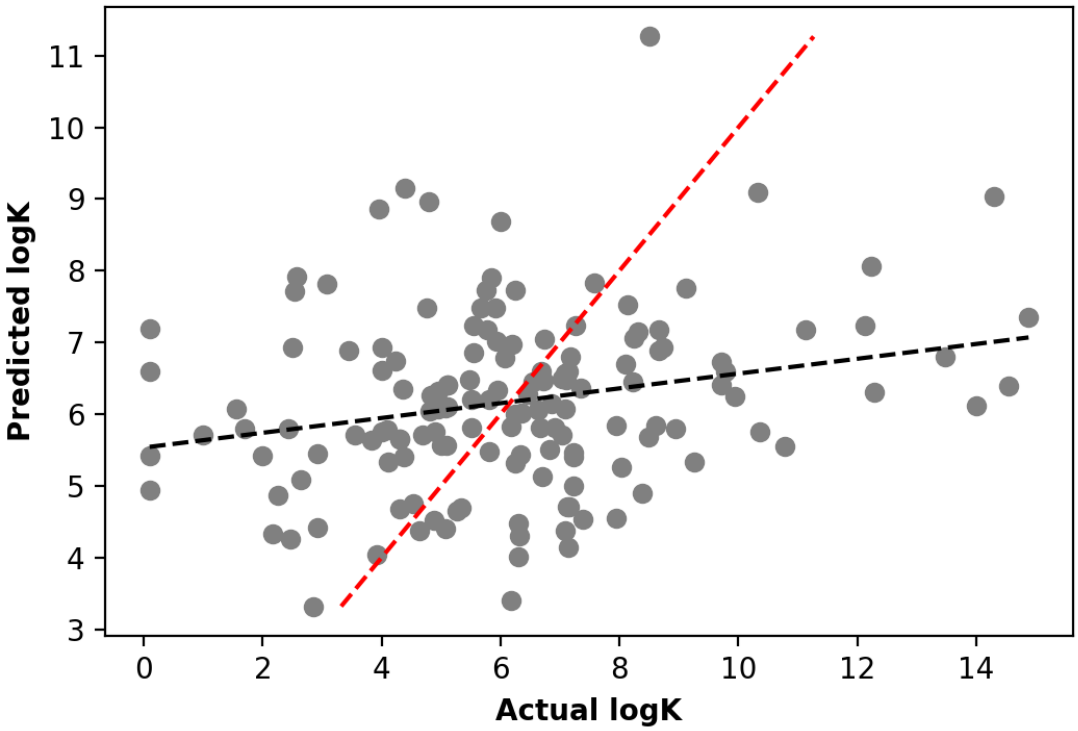

Figure S23: Prediction performance for leave-one-out experiments with random input data. The line of best fit based on predicted points is shown in black, and the line representing perfect prediction is shown in red (score = 2.17, $R^2 = 0.056898$).

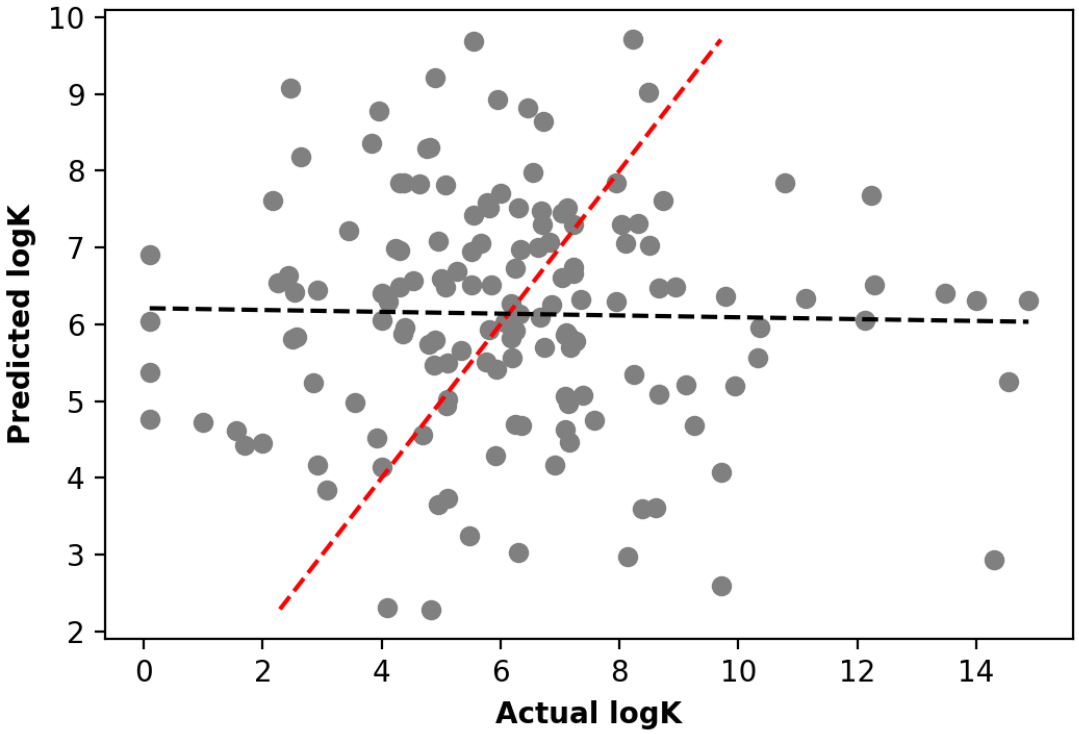

Figure S24: Prediction performance for leave-one-out experiments with shuffled input and output data. The line of best fit based on predicted points is shown in black, and the line representing perfect prediction is shown in red (score = 2.49, R^2 = 0.000497).

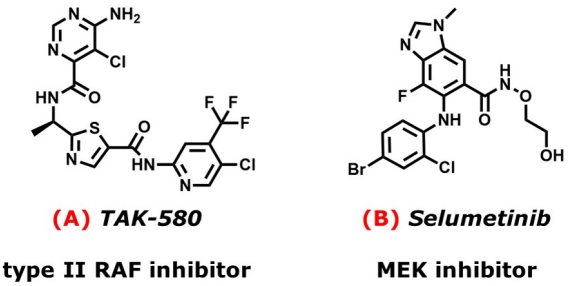

Figure S25: API chemical structures as reported by the manufacturer.

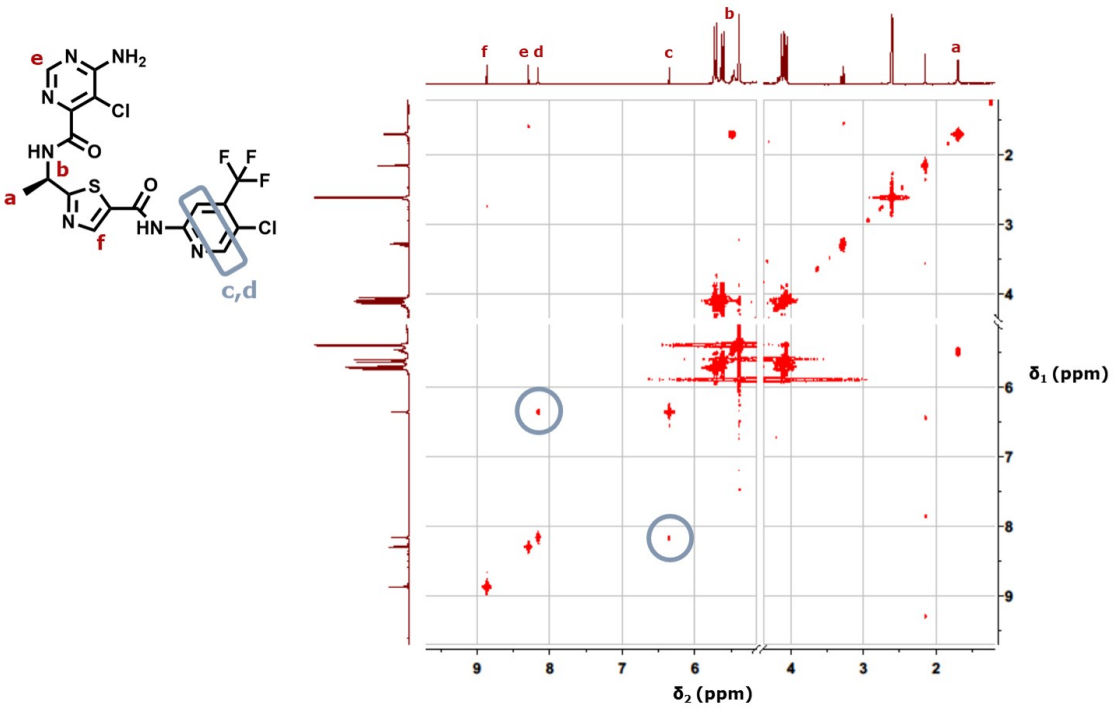

Figure S26: Long-range correlation spectroscopy (COSY) of MLN2480/CB[7] solution. The $^{5}J_{H,H}$ correlation signal peaks at ($\delta_2 = 8.16$ ppm, $\delta_1 = 6.35$ ppm) and ($\delta_2 = 6.35$ ppm, $\delta_1 = 8.16$ ppm) suggest that peaks c and d correspond to the two protons on the trifluoromethyl-substituted ring.

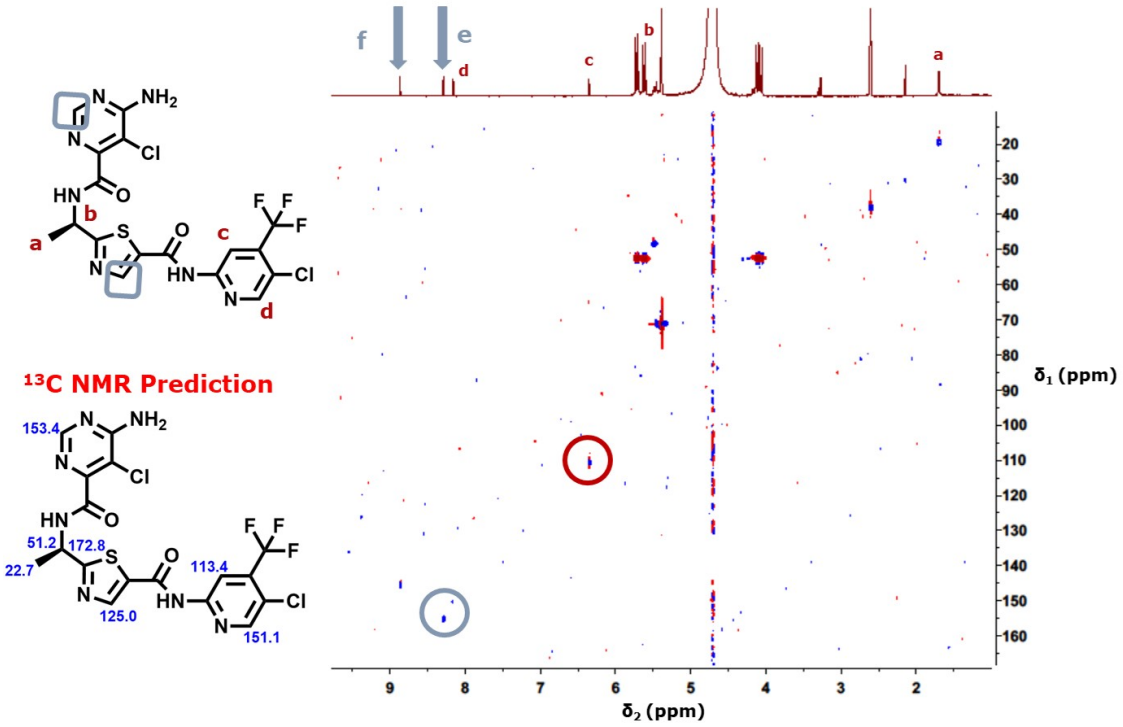

Figure S27: Heteronuclear single quantum coherence spectroscopy (HSQC) of MLN2480/CB[7] solution. The peak at ($\delta_2 = 6.35$ ppm, $\delta_1 = 110.64$ ppm) and computational ^{13}C NMR predictions were used to assign peaks *c* and *d*. The peak at ($\delta_2 = 8.29$ ppm, $\delta_1 = 155.24$ ppm) suggested that proton *e* was the aromatic proton on the pyrimidine ring.

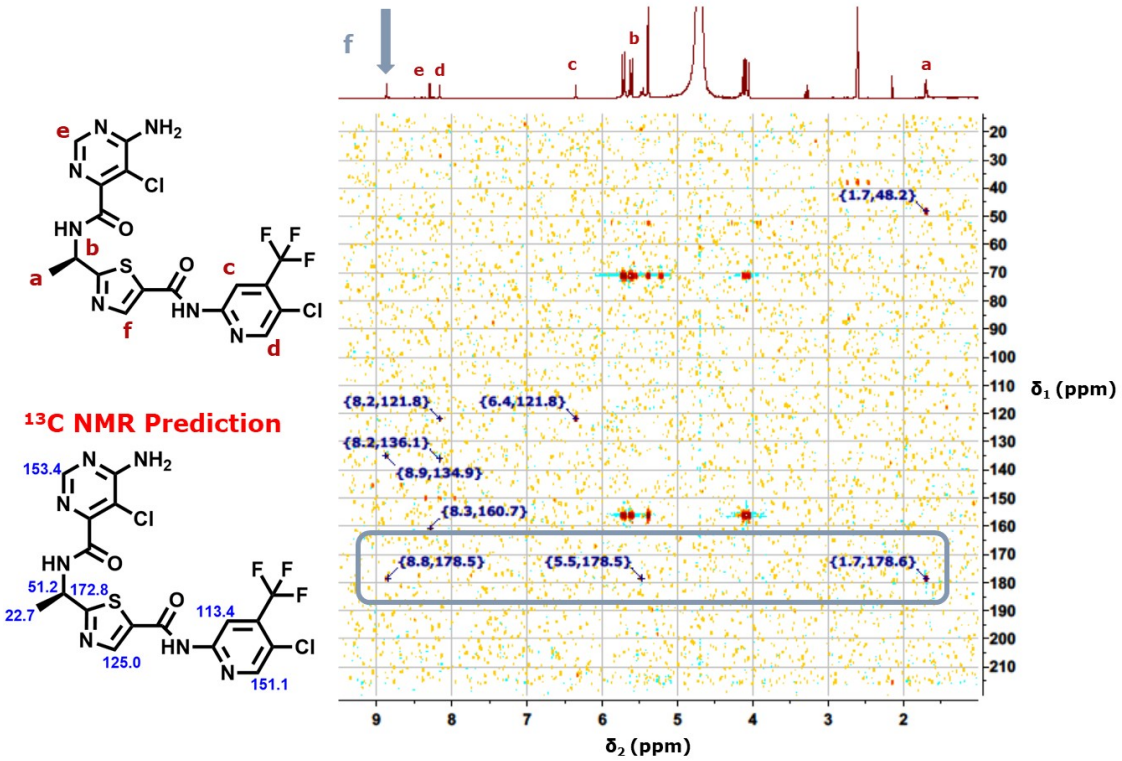

Figure S28: Heteronuclear multiple bond correlation spectroscopy (HMBC) of MLN2480/CB[7] solution. The peaks at (δ_2 = 8.8 ppm, δ_1 = 178.5 ppm), (δ_2 = 5.5 ppm, δ_1 = 178.5 ppm), and (δ_2 = 1.7 ppm, δ_1 = 178.6 ppm) correspond to the carbon predicted at 172.8, which will have a correlation signal with protons *a* and *b*, as well as the aromatic proton in the 4-position on the thiazole, suggesting it is proton *f*. All the remaining correlation signals corresponding to the drug which could be resolved from CB[7] signals are also marked.

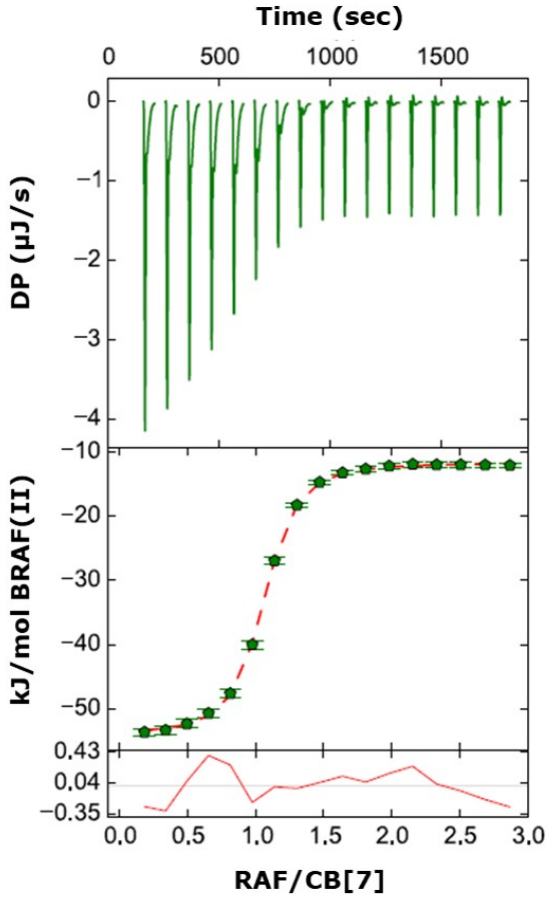

Figure S29: Isothermal titration calorimetry (ITC) was used to quantify the binding affinity of MLN2480 to CB[7]. (A) Illustration depicting the orientation of MLN2480 and the location of static binding with CB[7]. (B) ITC data fit to a one-site binding model. Residuals are shown in the third panel (bottom). ΔG = -37.3 ± 0.1 kJ/mol; ΔH = -42.5 ± 0.2 kJ/mol; $T\Delta S$ = -5.2 ± 0.3 kJ/mol; baseline = -11.6 kJ/mol.

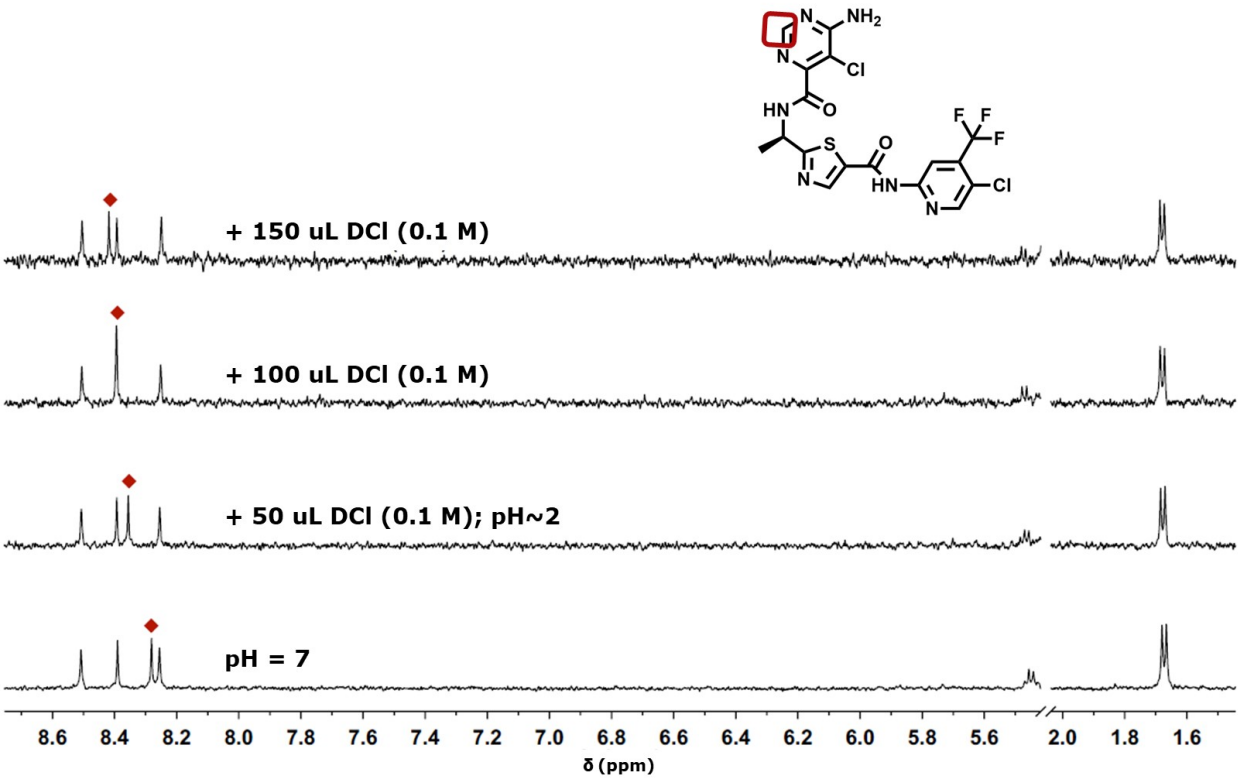

Figure S30: ^{1}H NMR spectra showing MLN2480 in CB[7] at pH = 7 (bottom). DCl was titrated in to lower the pH (bottom to top). The shift in proton *e* suggests that primary amine is uncharged at physiological pH.

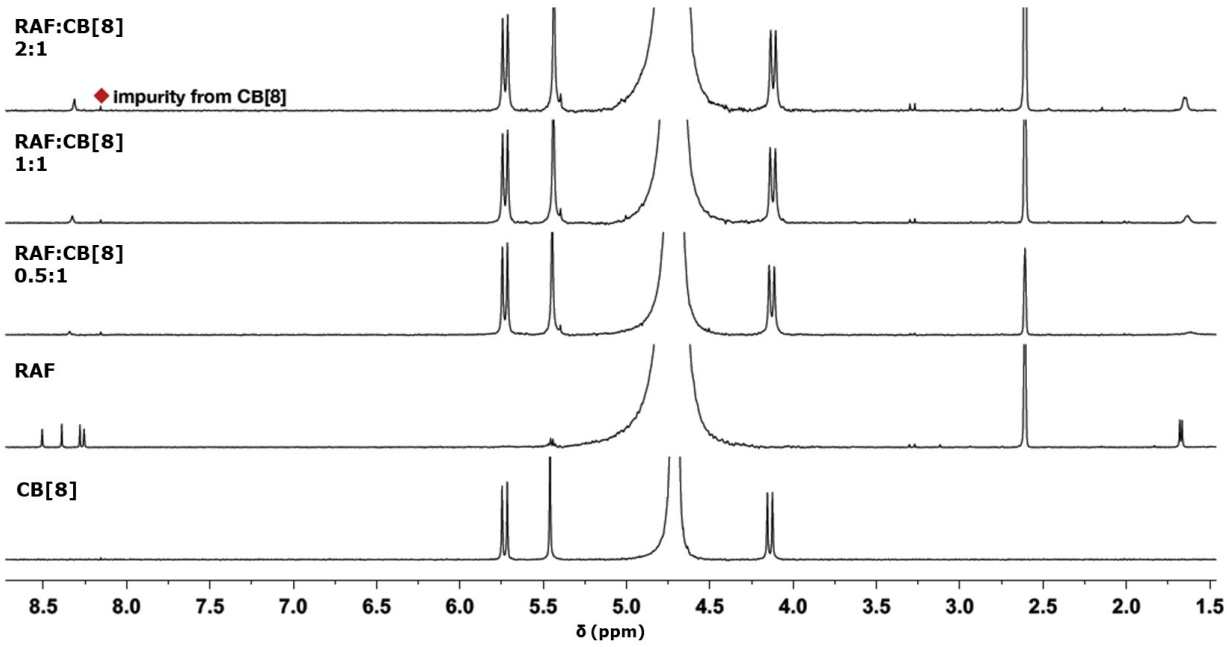

Figure S31: ^{1}H NMR spectra showing titration of MLN2480 in CB[8] at various molar ratios. These data suggest dynamic binding of MLN2480 to CB[8] in a 1:1 fashion with rapid exchange.

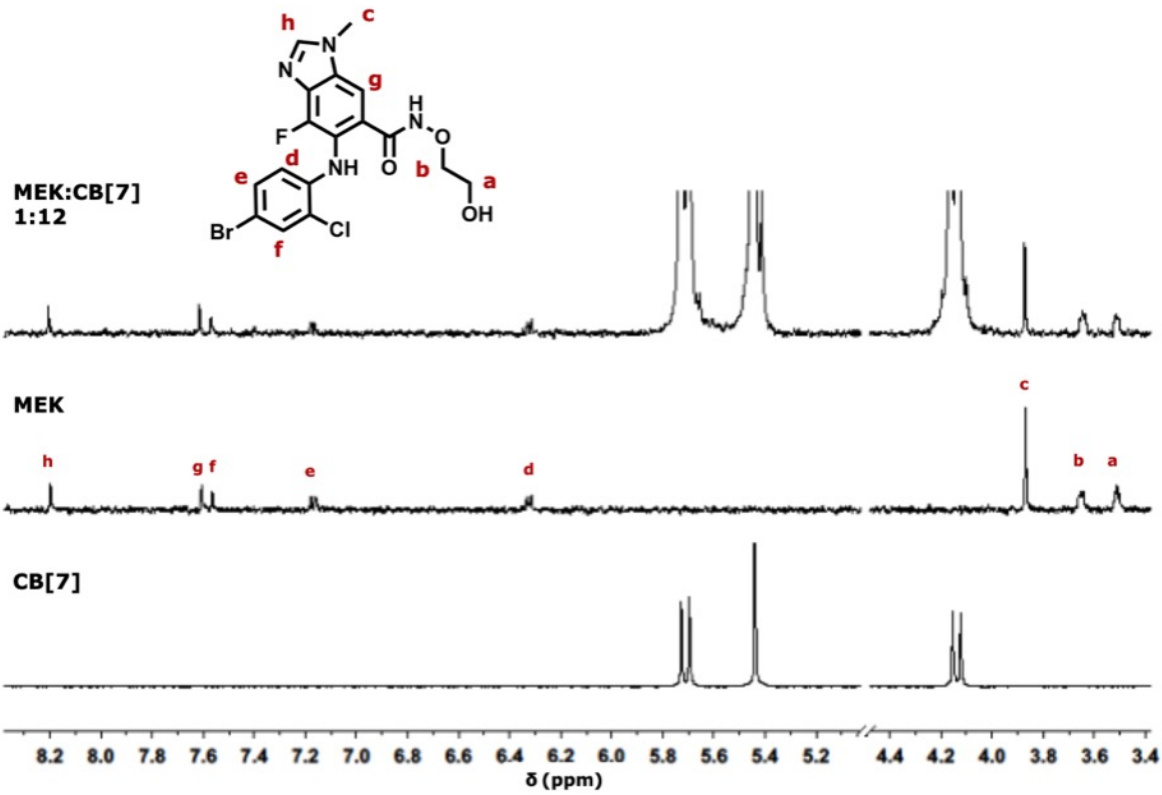

Figure S32: ^{1}H NMR spectra showing addition of AZD6244 in CB[7] with excess CB[7]. Lack of movement of any peaks of the MEK inhibitor strongly suggests that the drug does not interact with CB[7].

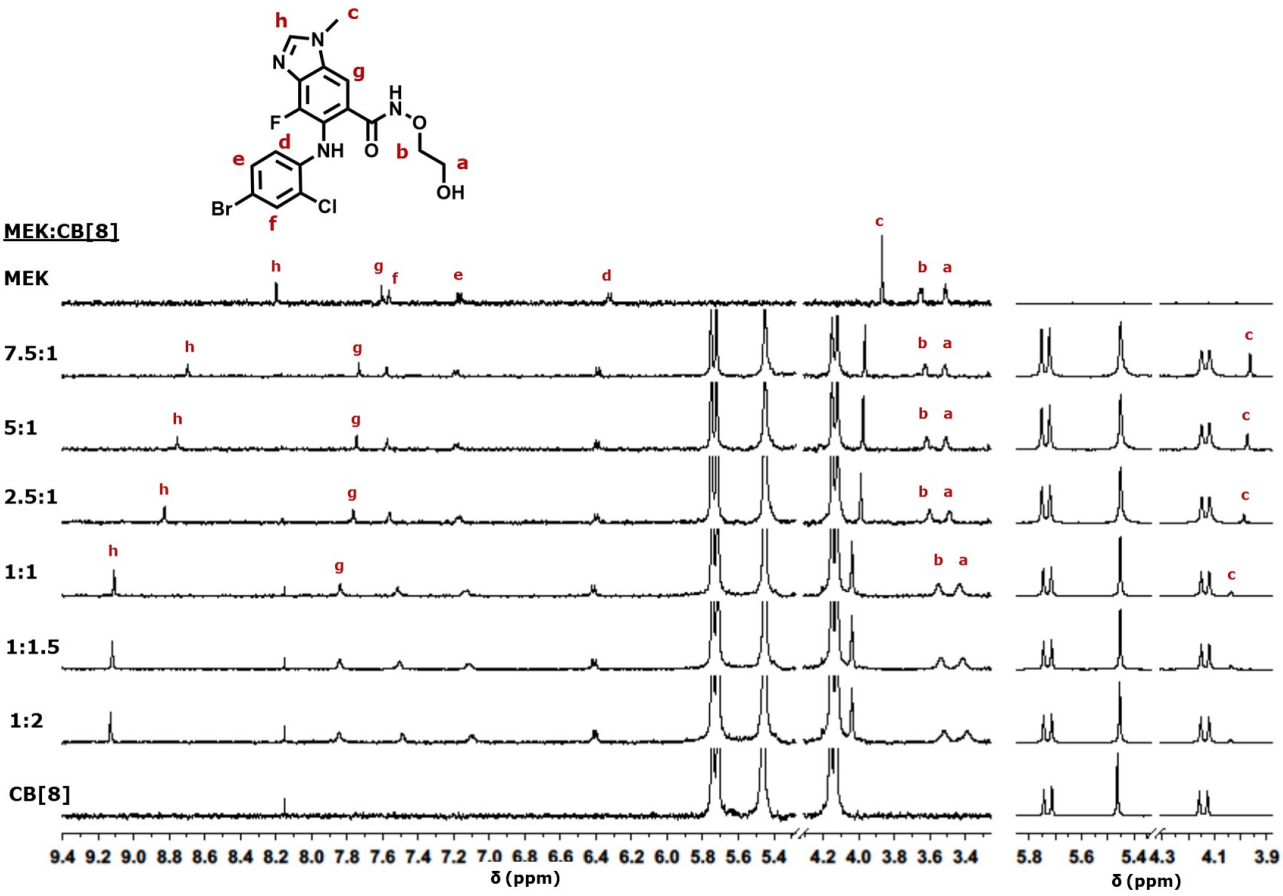

Figure S33: ^{1}H NMR spectra showing titration of CB[8] into AZD6244 at various molar ratios. These data suggest more static binding of AZD6244 to CB[8] compared to MLN2480 in CB[8]. These data also suggest binding occurs in a 1:1 fashion due to the proton shifts until a 1:1 ratio and no further change after continued addition of CB[8].

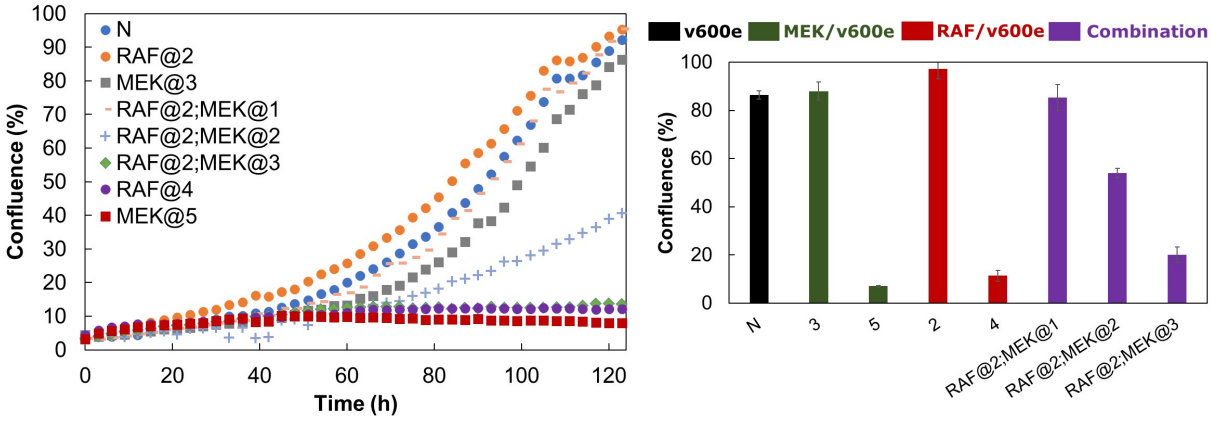

Figure S34: Kinetic growth curves of mouse v600e cells incubated with different concentrations of RAF and MEK. N is positive control (media). Terminal confluence is plotted for each condition.

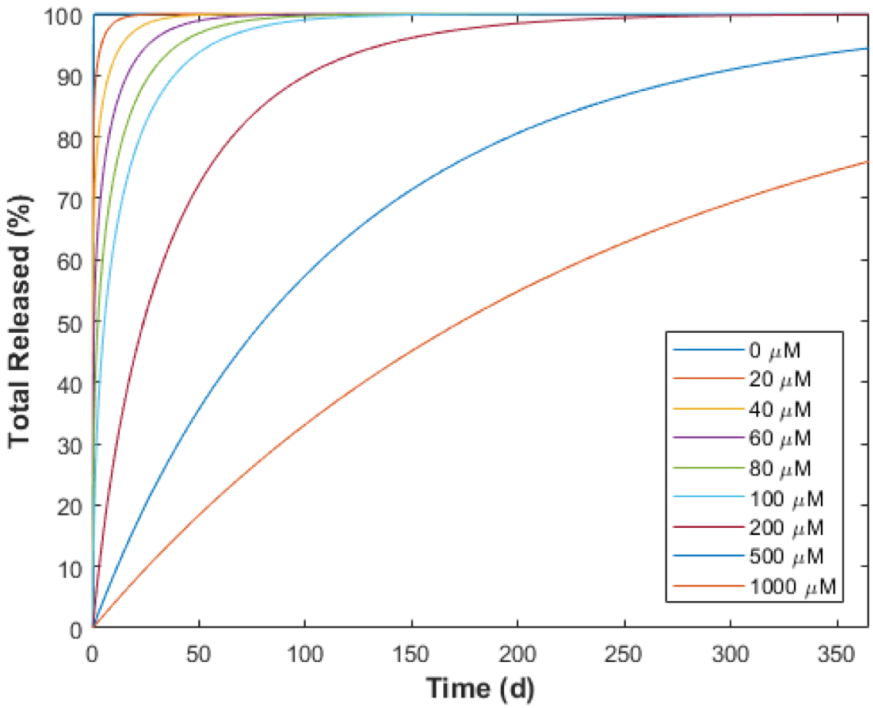

Figure S35: Simulated time-resolved release kinetics of RAF over one year with different concentrations of CB[7].

S.3 Binding Analyses *via* NMR

Determining RAF Binding Orientation

Doublet *a* and quartet *b* were assigned to the methyl and methine protons of the stereocenter. Long-range correlation spectroscopy (COSY) was used to identify which of the remaining four peaks *c,d,e,f* corresponded to the two aromatic protons on the same trifluoromethyl-substituted ring (Fig. S26). Long-range COSY was used to detect proton interactions showing ^{n}J-coupling constants with n $\geq$ 4. Both protons on the trifluoromethyl-substituted ring were the only aromatic protons close enough to generate a correlation signal on RAF. While these data strongly suggests these two protons were *c* and *d*, it is not clear which is which. Heteronuclear single quantum coherence spectroscopy (HSQC) was used to detect correlation signals between these protons and their bound carbons (Fig. S27). Fig. S27 plots ^{13}C signals on the y-axis and ^{1}H signals on the x-axis. ^{13}C NMR prediction suggested that aromatic proton ortho to -CF_3 will have an upfield shift relative to the meta aromatic proton. The correlation signal at (δ_2 = 6.35 ppm, δ_1 = 110.64 ppm) (circled red) suggested the aromatic proton in the ortho-position could be assigned to signal *c*, while the aromatic proton in the meta-position could be attributed to signal *d*. The HSQC correlation peak at (δ_2 = 8.29 ppm, δ_1 = 155.24 ppm) suggested the aromatic proton on the pyrimidine ring was *e*. This was further confirmed with heteronuclear multiple bond correlation spectroscopy (HMBC) in Fig. S28. HMBC identifies heteronuclear correlations between protons and carbons through multiple bonds ($^{2-3}$J). For the assignment of proton *f*, three signals in the HMBC spectrum were identified (Fig. S28). Proton *a* had a $^{3}J_{H,C}$ coupling to the carbon in the 2-position of the thiazole (δ_1 = 178.5 ppm). As expected, the methine proton *b* showed a $^{2}J_{H,C}$ coupling to the same thiazole carbon. Based on the $^{3}J_{H,C}$ coupling value at (δ_2 = 8.8 ppm, δ_1 = 178.5 ppm), it could be concluded that the proton in the 4-position of the thiazole was *f*.

With the peaks assigned, Fig. 6A could be used to elucidate how CB[7] binds with RAF. We hypothesized that the primary amine would be positively charged at physiological pH, and further that since the trifluoromethyl group on the lower ring is bulky, CB[7] would bind at the pyrimidine ring. However, our data suggested that is not the case. Protons *c* and *d* underwent a large upfield shift when CB[7] was added, which is consistent with those protons located *inside* the CB[7] cavity (Wu et al, JACS 2017). Proton *f* underwent a downfield shift, which suggested it is near but *outside* the CB[7] cavity. It is unlikely that such a shift would occur in the chemical orientation drawn in Fig. S25 provided by the manufacturer, at least when it is bound by CB[7]. Furthermore, protons *a*, *b*, and *e* remained unchanged with CB[7], which suggested they were not near CB[7]. Taken in tandem, these observations suggest our original hypothesis was incorrect and CB[7] binds statically at the trifluoromethyl-substituted ring in a 1:1 fashion (Fig. S29).

MEK:CB[7] Interactions

MEK was added to a solution of excess CB[7] to determine whether any interactions were occurring (Fig. S32). Triplets *a* and *b* were assigned to the ethylene glycol protons, singlet *c* was assigned to the methyl protons, and singlet *h* was assigned to the aromatic imidazole proton. Doublet *d* was assigned to the aromatic proton meta of the bromo group, and doublet of doublets *e* was assigned to the proton ortho of the bromo group and ortho to *d*. Peak *f* was assigned as the remaining aromatic proton ortho to the bromo group, and finally singlet *g* was assigned as the aromatic proton para to the fluoro group. Clearly, these ^{1}H NMR data demonstrated that the drug does not bind to CB[7] whatsoever, as evidenced by no change in any peak position (Fig. S32).