

Supplementary Information

CardioGenAI: A Machine Learning-Based Framework for Re-Engineering Drugs for Reduced hERG Liability

Gregory W. Kyro^{1,2,*}, Matthew T. Martin², Eric D. Watt², Victor S. Batista^{1,*}

¹ Department of Chemistry, Yale University, New Haven, Connecticut 06511

² Drug Safety Research & Development, Pfizer Research & Development, Groton, Connecticut 06340

* Indicates corresponding author

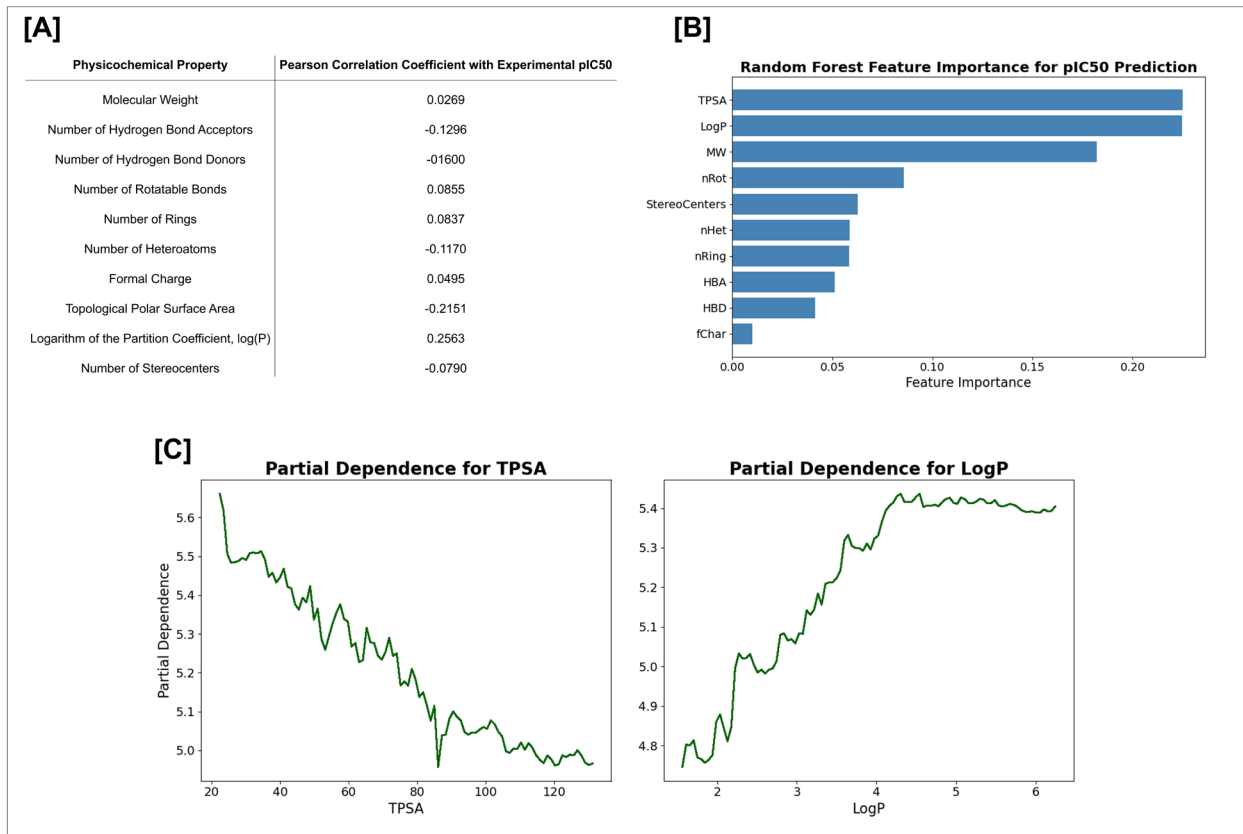


Figure S1. Analysis of the relationships between physicochemical properties and experimental hERG channel pIC₅₀ values. In [A], Pearson correlation is shown between each physicochemical property and experimental pIC₅₀ values obtained from the training set used. There are noteworthy correlations with pIC₅₀ for topological polar surface area (TPSA) and LogP. A random forest model with 100 estimators was fit to the data to predict pIC₅₀ values, and the importance of each feature was then deduced. In [B], the feature importance of each physicochemical property is shown. In [C], the partial dependences for TPSA (Å²) and LogP are shown. For LogP, there is an initial positive trend where an increase in LogP corresponds to an increase in pIC₅₀ values up to a LogP value of approximately 4. For TPSA, as TPSA increases, pIC₅₀ values generally decrease.

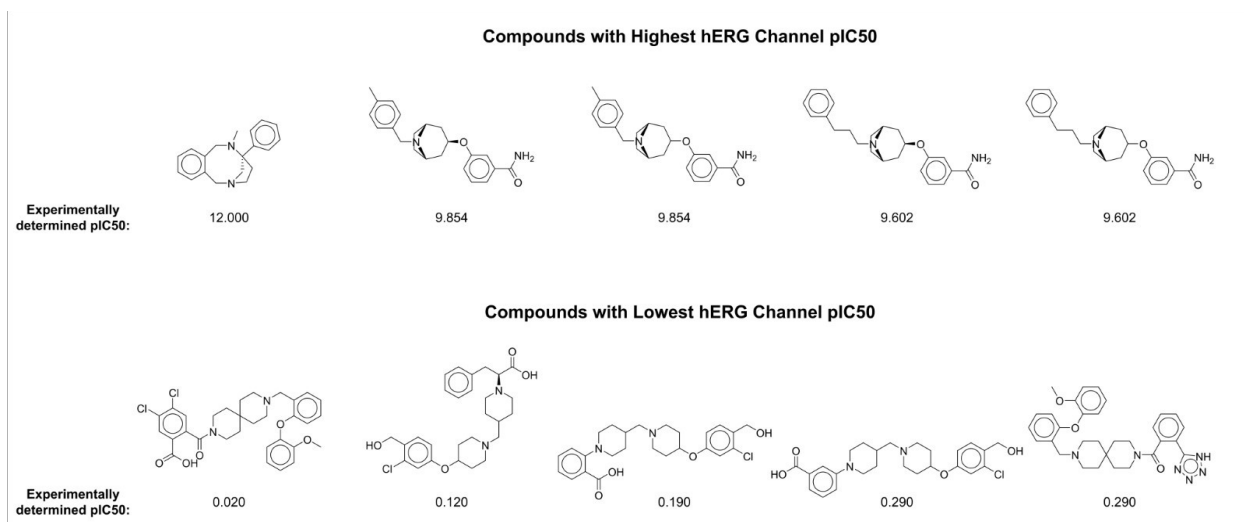


Figure S2. Display of the five compounds with highest hERG channel activity, as well as the five compounds with the lowest hERG channel activity of compounds in the hERG channel training set used. Each molecule is labeled with the corresponding experimentally determined hERG channel pIC₅₀. Molecules are rendered with ChemDraw v23.1.1.

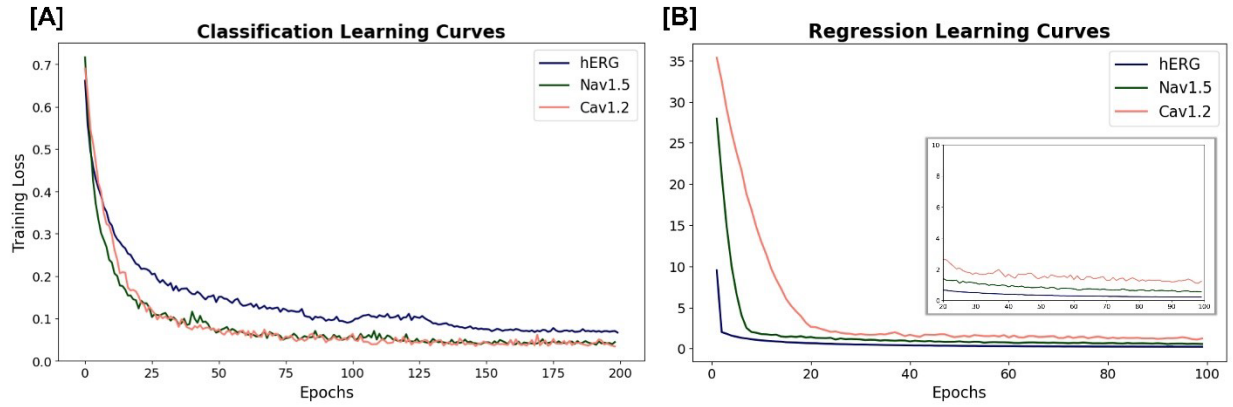


Figure S3. Learning curves for hERG, Nav1.5 and Cav1.2 cardiac ion channel [A] classification and [B] regression models. The classification models were trained with binary cross entropy loss and the AdamW optimizer for 200 epochs. Regression models were trained analogously but using mean squared error loss for 100 epochs.

Table S4. Performance regarding each possible feature-representation combination for binary classification of hERG channel blockers.

Feature Representations	AC	SN	SP	F1	CCR	MCC
Transformer feature vector + Graph + Fingerprint	83.5	86.2	80.3	85.1	83.2	66.7
Transformer feature vector + Fingerprint	80.4	82.9	77.3	82.3	80.1	60.4
Transformer feature vector + Graph	80.0	82.9	76.4	81.9	79.6	59.5
Graph + Fingerprint	78.6	80.9	75.9	80.6	78.4	56.8
Transformer feature vector	77.7	83.3	70.9	80.4	77.1	54.9
Fingerprint	76.6	79.3	73.4	78.8	76.3	52.7
Graph	74.4	87.8	58.1	79.0	73.0	48.6

^a The evaluation set used is that developed by Arab et al.; compounds in the evaluation set have a structural similarity (as determined by pairwise Tanimoto similarity between 2048-bit Morgan fingerprints) no greater than 0.70 to any compound in the corresponding training or validation sets.

^b The top value achieved for each metric is shown in bold.

^c Accuracy (AC), sensitivity (SN), specificity (SP), F1-score (F1), correct classification rate (CCR), and Matthew's correlation coefficient (MCC) are shown.

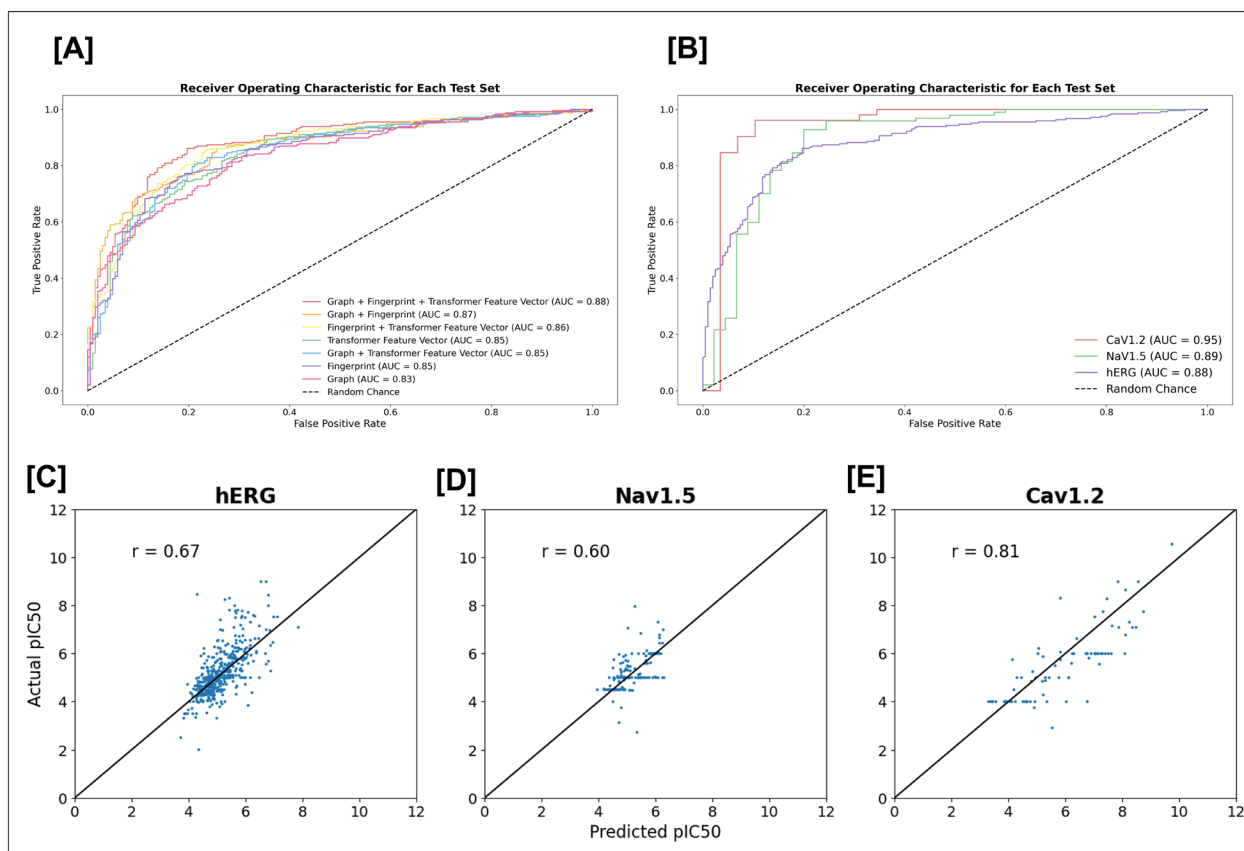


Figure S5. Evaluation of classification and regression models for cardiac ion channel blocker prediction. [A] Receiver operating characteristic (ROC) curve for evaluation on the hERG benchmark presented by Arab et al. Results regarding each feature-representation combination are shown with the corresponding area under the curve (AUC). [B] ROC curve for evaluation on the hERG, Nav_v1.5 and Cav_v1.2 channel benchmarks. Scatter plots depicting actual pIC₅₀ as a function of predicted pIC₅₀, with the corresponding Pearson correlation coefficient (r), shown for evaluation on the [C] hERG, [D] Nav_v1.5, and [E] Cav_v1.2 channel benchmark sets.

Table S6. Performance of the hERG, Nav1.5 and Cav1.2 regression models on the benchmarks presented by Arab et al.

Channel	R²	Q_{F1}²	Q_{F2}²	Q_{F3}²	CCC	r_m²	RMSE	MAE
hERG	0.439	0.439	0.439	-20.922	0.606	0.116	0.741	0.509
Nav1.5	0.323	0.327	0.323	-6.060	0.576	0.060	0.620	0.429
Cav1.2	0.536	0.541	0.536	-2.597	0.784	0.181	1.006	0.821

^a The evaluation sets used are those developed by Arab et al.; compounds in the evaluation set have a structural similarity (as determined by pairwise Tanimoto similarity between 2048-bit Morgan fingerprints) no greater than 0.70 to any compound in the corresponding training or validation sets.

^b The coefficient of determination (R²), predictive squared correlation coefficient for the first fold (Q_{F1}²), predictive squared correlation coefficient for the second fold (Q_{F2}²), predictive squared correlation coefficient for the third fold (Q_{F3}²), concordance correlation coefficient (CCC), squared correlation coefficient for model (r_m²), root mean square error (RMSE), and mean absolute error (MAE) are shown.

Table S7. Pearson correlation between physicochemical properties and predicted pIC₅₀ values of compounds in each respective test set.

Physicochemical Property	hERG	Nav1.5	Cav1.2
Molecular Weight	0.146	0.118	0.444
Number of Hydrogen Bond Acceptors	-0.159	-0.239	0.621
Number of Hydrogen Bond Donors	-0.114	-0.593	-0.002
Number of Rotatable Bonds	0.327	-0.294	0.318
Number of Rings	0.080	0.219	-0.315
Number of Heteroatoms	-0.081	0.095	0.555
Formal Charge	0.192	NaN	0.234
Topological Polar Surface Area, TPSA	-0.273	-0.545	0.581
Logarithm of the Partition Coefficient, log(P)	0.321	0.406	0.233
Number of Stereocenters	-0.169	-0.085	-0.182

^a Compounds in the evaluation set have a structural similarity, as determined by pairwise Tanimoto similarity between 2048-bit Morgan fingerprints, no greater than 0.70 to any compound in the corresponding training or validation sets.

^b NaN indicates that all values in the set are constant. In the case of Formal Charge for the Nav1.5 channel, all compounds in the test set have a value of 0.

^c Pearson correlation coefficient values with a magnitude greater than 3.00 are shown in bold.

Table S8. Y-randomization test applied to each of the classification models.

Channel	AC	SN	SP	F1	CCR	MCC
hERG	55.0	100.0	0.5	70.9	50.2	5.2
Nav1.5	68.3	100.0	0.0	81.2	50.0	0.0
Cav1.2	65.4	100.0	3.4	78.8	51.7	15.0

^a Compounds in the evaluation set have a structural similarity, as determined by pairwise Tanimoto similarity between 2048-bit Morgan fingerprints, no greater than 0.70 to any compound in the corresponding training or validation sets.

^b Accuracy (AC), sensitivity (SN), specificity (SP), F1-score (F1), correct classification rate (CCR), and Matthew's correlation coefficient (MCC) are shown.

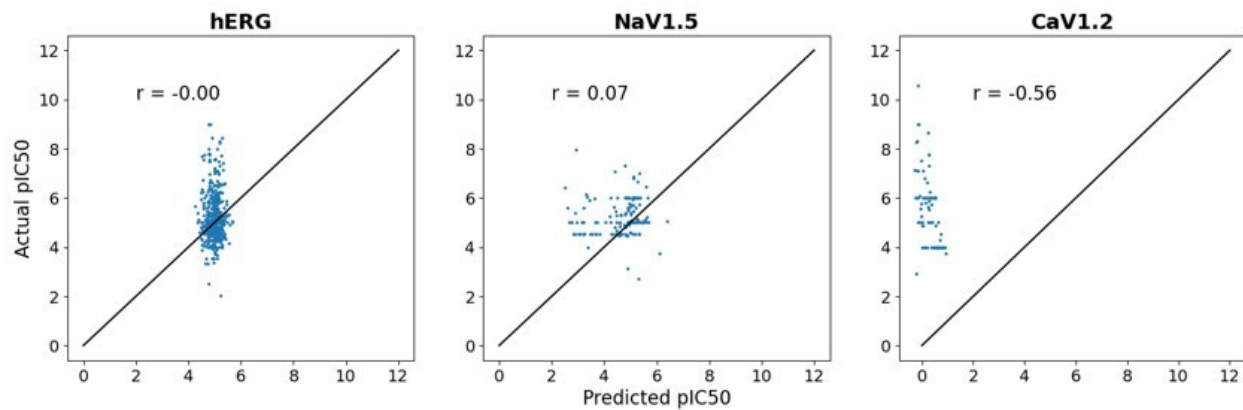


Figure S9. Y-randomization test applied to each of the regression models. Scatter plots depicting actual pIC_{50} as a function of predicted pIC_{50} , with the corresponding Pearson correlation coefficient (r), shown for evaluation on the hERG, $Na_V1.5$, and $Ca_V1.2$ channel benchmark sets.

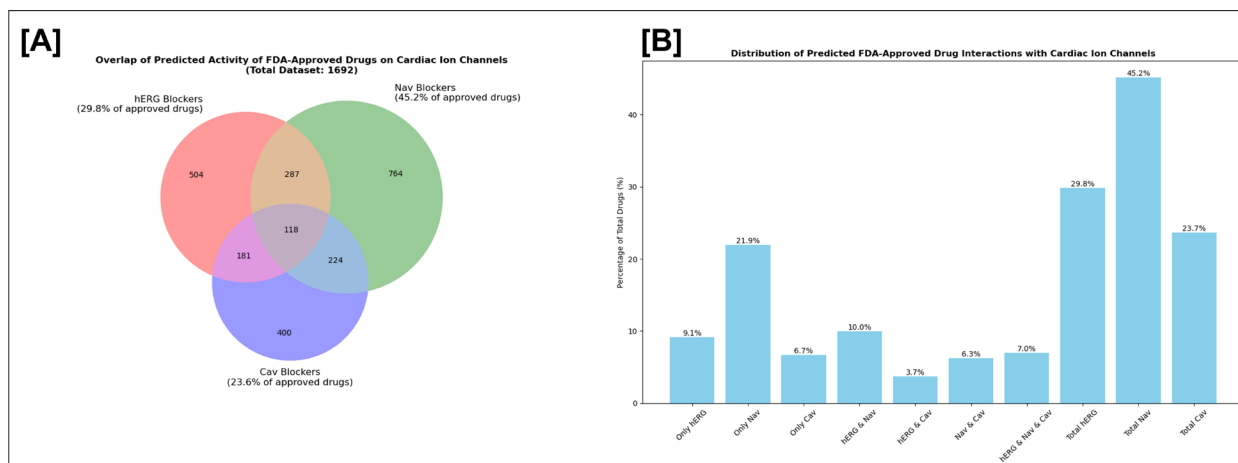


Figure S10. Application of cardiac ion channel classification models to a subset of FDA-approved drugs obtained from DrugCentral. A Venn diagram [A] and bar plot [B] are shown for the screening results of the 1692 total compounds.

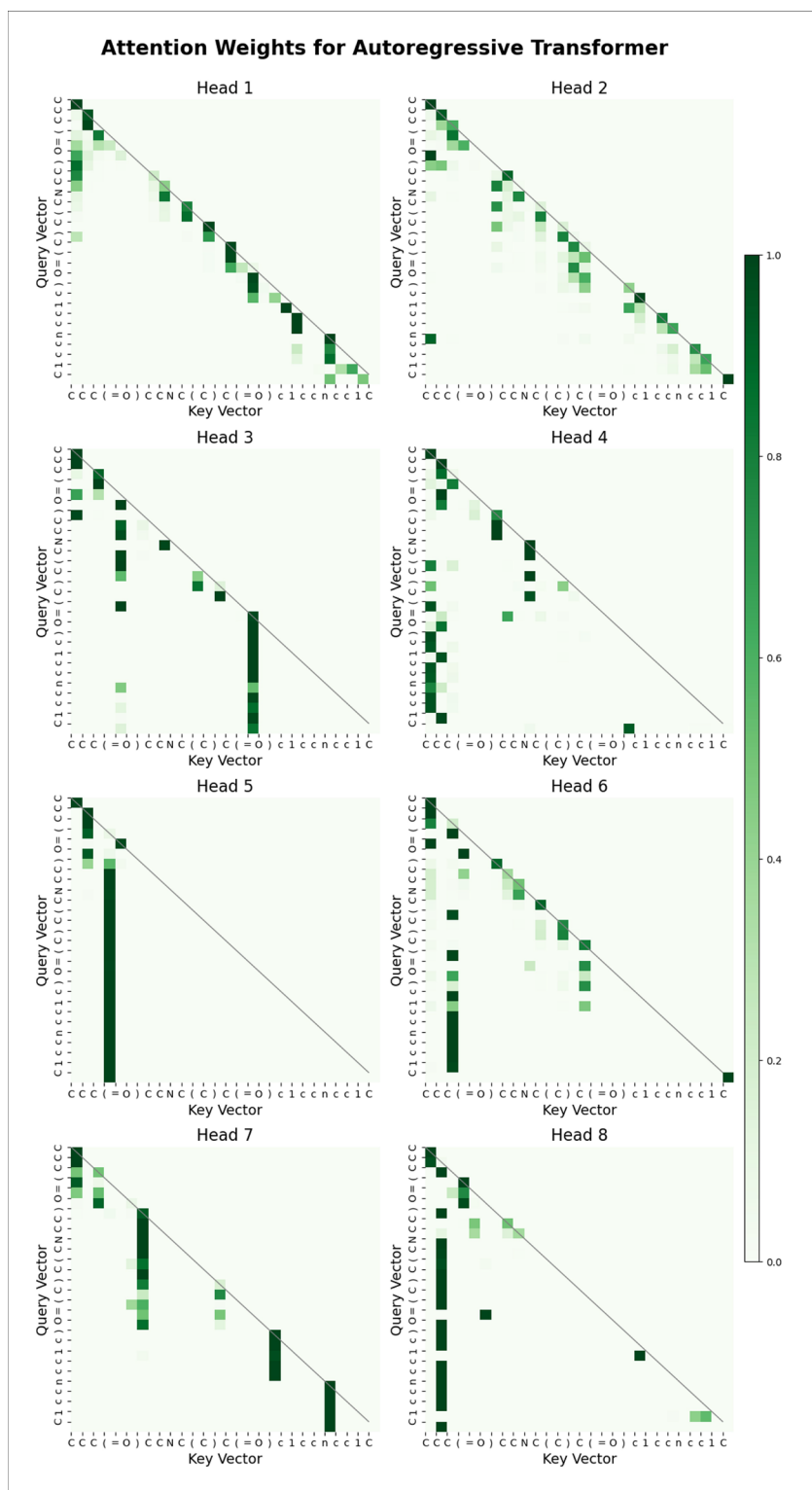


Figure S11. Representative attention maps depicting the distribution of attention weights extracted from the autoregressive transformer trained on SMILES strings, with each subplot corresponding to one of the eight heads in the model's attention layer. Weights are shown for the SMILES string: “CCC(=O)CCNC(C)C(=O)c1ccncc1C”.

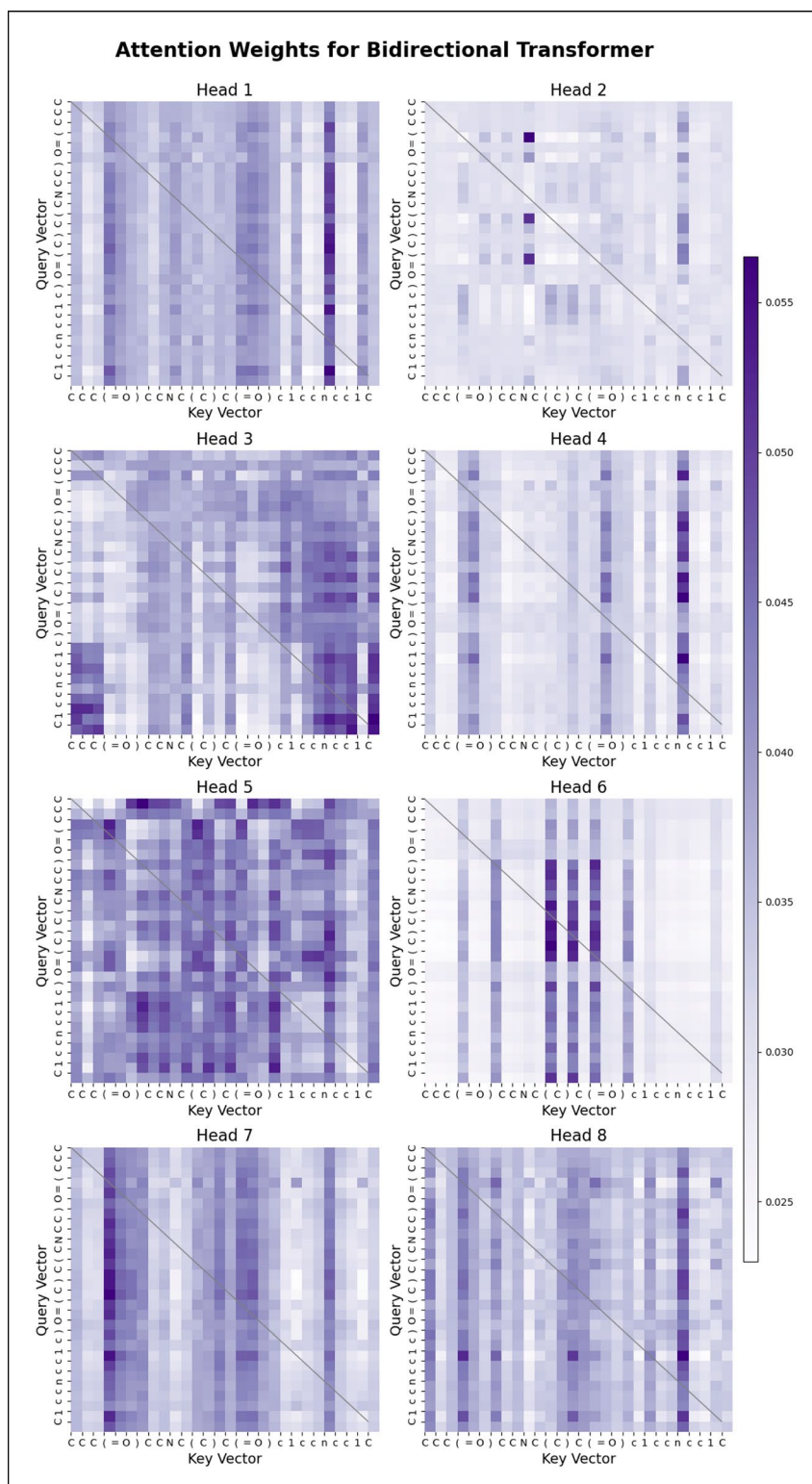


Figure S12. Representative attention maps depicting the distribution of attention weights extracted from the bidirectional transformer trained on SMILES strings, with each subplot corresponding to one of the eight heads in the model's attention layer. Weights are shown for the SMILES string: “CCC(=O)CCNC(C)C(=O)c1ccncc1C”.

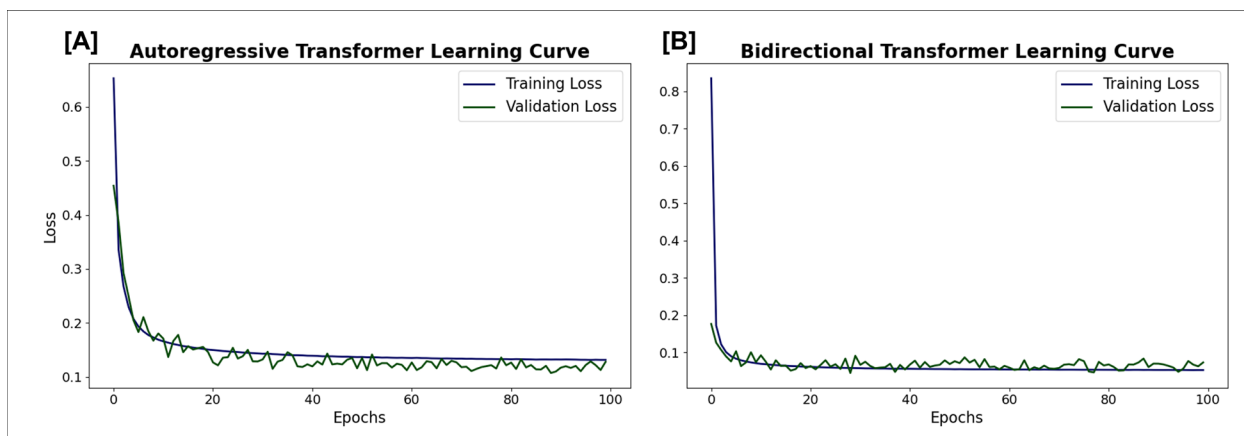


Figure S13. Learning curves for the [A] autoregressive transformer and [B] bidirectional transformer models. The autoregressive transformer is trained for next-token prediction, and the bidirectional transformer is trained for masked-token prediction. Both models are trained with cross entropy loss and the Sophia optimizer for 100 epochs.

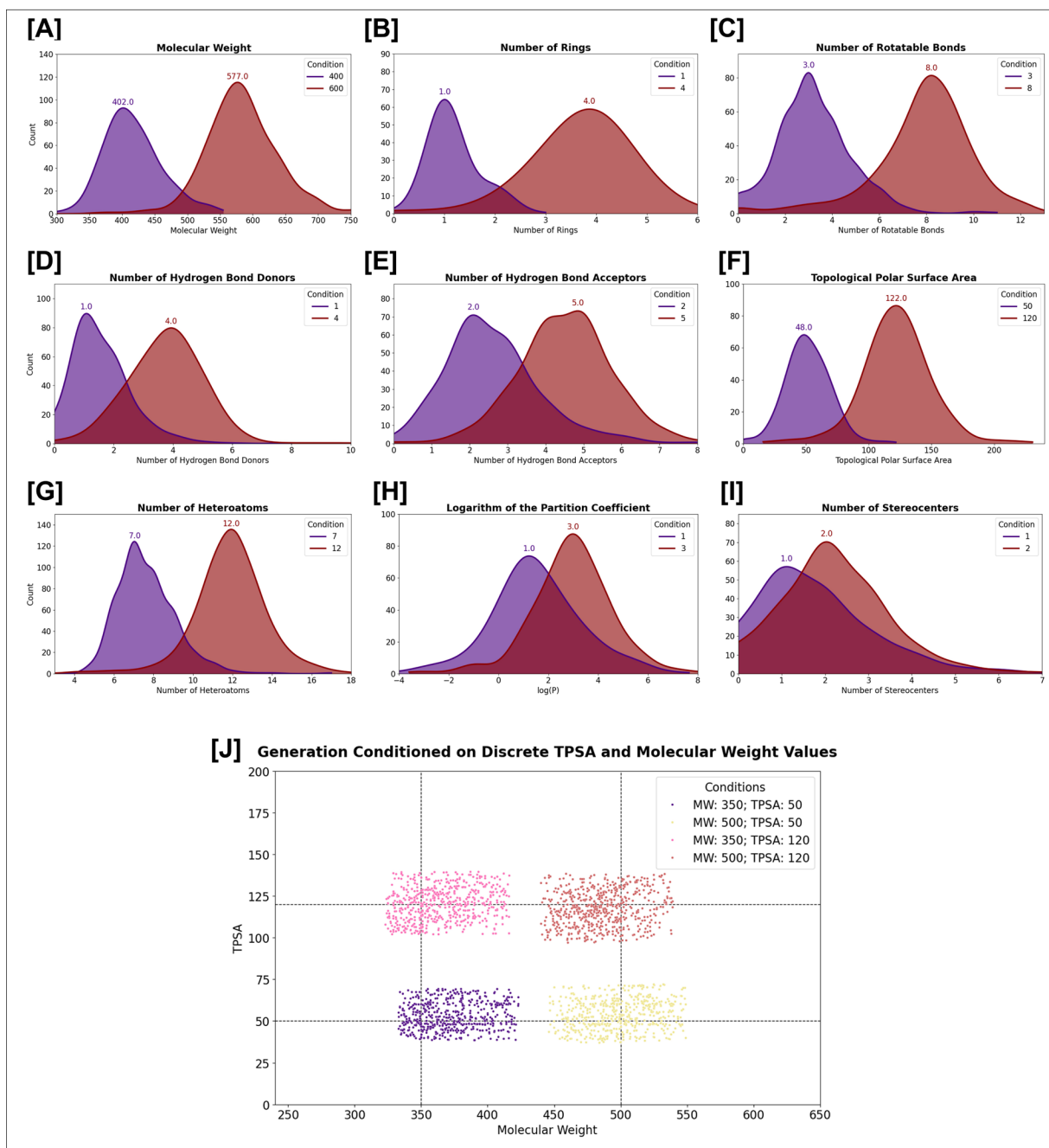


Figure S14. Probing the ability of the generative model to generate distributions of molecules with desired conditions. Results are shown for generations conditioned on different discrete values for [A] molecular weight ($\frac{g}{mol}$), [B] number of rings, [C] number of rotatable bonds, [D] number of hydrogen bond donors, [E] number of hydrogen bond acceptors, [F] topological polar surface area (\AA^2), [G] number of heteroatoms, [H] LogP (logarithm of the partition coefficient), and [I] number of stereocenters. In [J], the generations are conditioned based on different value combinations of topological polar surface area (TPSA) and molecular weight. For each of the four pairs of conditions, values above/below two standard deviations greater/less than the mean value of each metric are excluded to emphasize the locations of the distribution means.

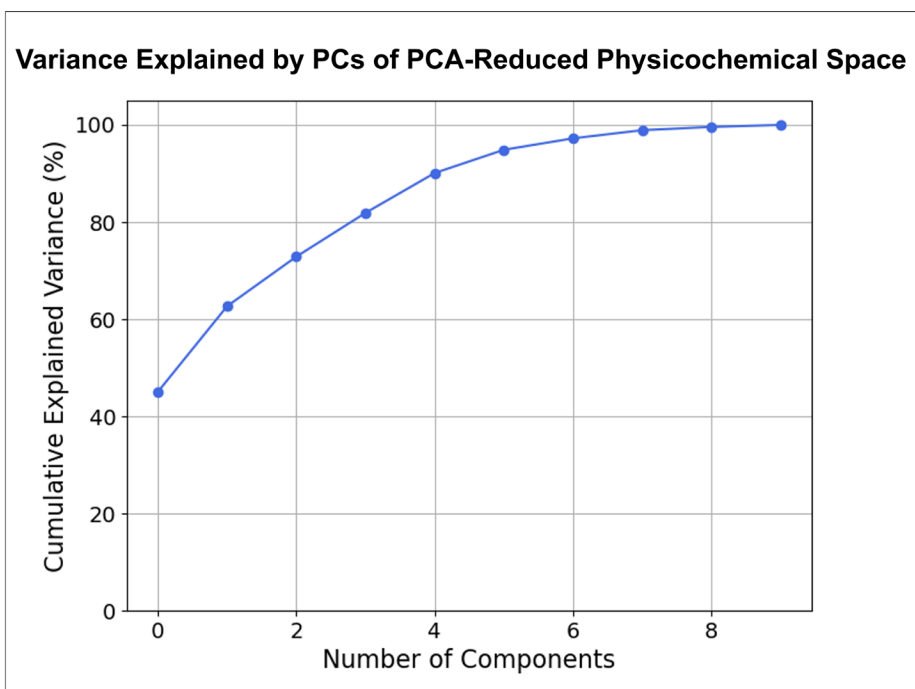


Figure S15. Cumulative variance as a function of principal component for principal component analysis (PCA) of physicochemical-based chemical space. For the input cardiotoxic molecule (pimozide), the generated molecules (100 datapoints), and the molecules in the pretraining set for the autoregressive transformer-based generative model (approximately 5 million datapoints), physicochemical properties are calculated, and PCA is performed to generate a lower-dimensional representation of chemical space.

Table S16. CardioGenAI methodology applied to pimoziide, an FDA-approved antipsychotic drug that has a predicted hERG-channel pIC_{50} of 7.629, and is reported to cause hERG channel blockade-induced arrhythmias (Table 4). 100 molecules are generated, and among them is fluspirilene, a compound that belongs to the same class of drugs as pimoziide and therefore has a similar primary therapeutic mode of action, but exhibits significantly less hERG-channel activity (5.785 pIC_{50}). Included in the table are the five most similar generated compounds to pimoziide in terms of cosine similarity between molecular descriptor vectors.

Similarity Rank	SMILES String	Cosine Similarity	Predicted pIC_{50}
Input Molecule	<chem>O=c1[nH]c2ccccc2n1C1CCN(CCCC(c2ccc(F)cc2)c2ccc(F)cc2)CC1</chem> (pimoziide)	1.000	7.629
1	<chem>Fc1ccc(CCCN2CCc3c[nH]c(n3)C2)c(N2CCC(c3ccccc3)CC2)c1</chem>	0.980	5.367
2	<chem>O=C(Nc1ccccc1)N(CCCN1CCCC1)CC1(c2ccc(F)cc2)CCC1</chem>	0.977	5.904
3	<chem>O=C(NCc1ccc(F)cc1)N(CCCN1CCCC1)c1ccc2ccccc12</chem>	0.976	5.794
4	<chem>O=C1NCN(c2ccccc2)C12CCN(CCCC(c1ccc(F)cc1)c1ccc(F)cc1)CC2</chem> (fluspirilene)	0.948	5.785
5	<chem>Fc1cccc(Cc2ncc(C3(N4CCC(Cc5ccccc5)CC4)CCOCC3)[nH]2)c1</chem>	0.946	5.201

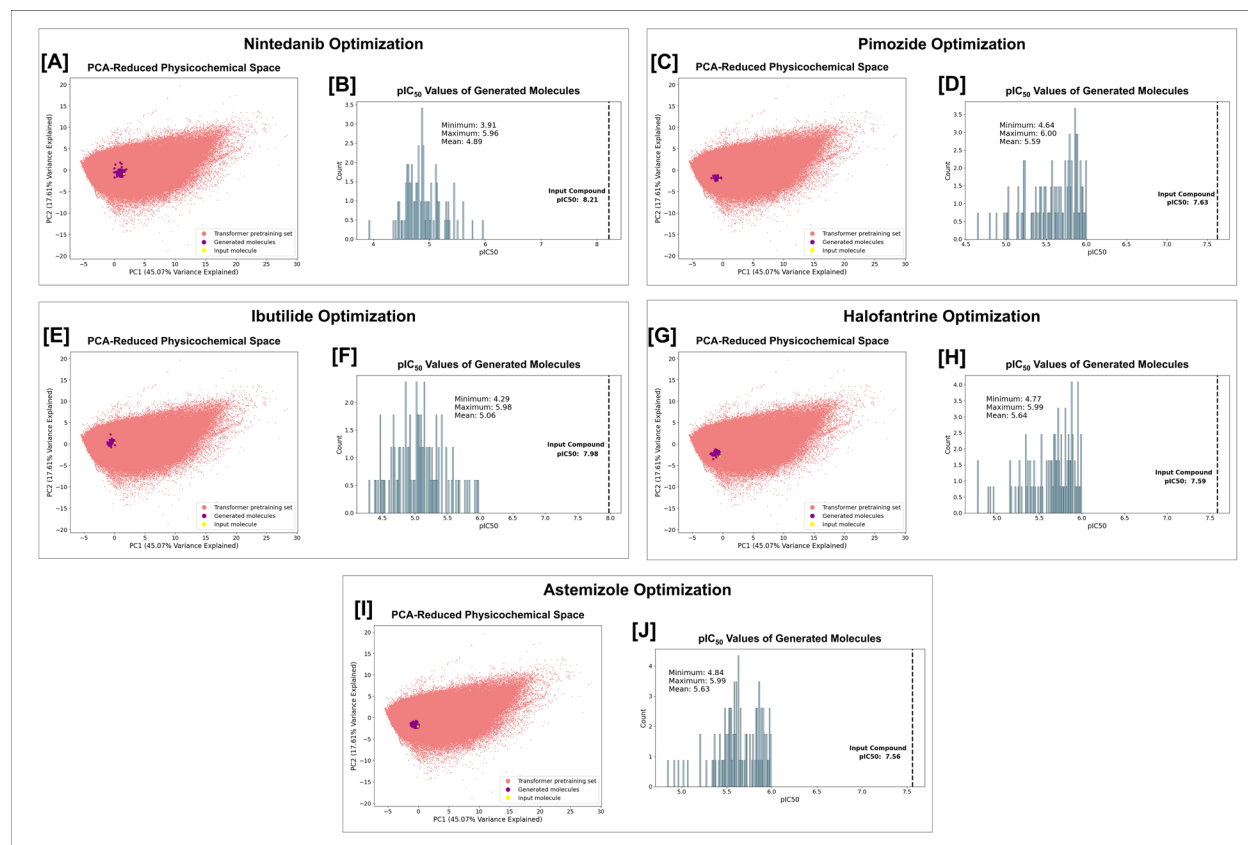


Figure S17. Visualization of the CardioGenAI framework applied to nintedanib [A-B], pimoziide [C-D], ibutilide [E-F], halofantrine [G-H], and astemizole [I-J]. In each application, the specified maximum predicted hERG pIC₅₀ value of any of the generated compounds was set to 6.00. For each optimization, the input molecule, the 100 generated refined molecules, and the molecules in the training set for the transformer-based models (approximately 5 million datapoints), are projected into a principal component analysis (PCA)-reduced physicochemical-based space. The input compound is colored yellow, the generated refined compounds are colored purple, and the compounds in the training set of the transformer-based models are colored red. The first two principal components explain 45.07% and 17.61% of the total variance, respectively. In each case, the CardioGenAI framework is able to identify the region of physicochemical space corresponding to compounds that are similar to the input compound, yet exhibit significantly reduced activity against the hERG channel. The densities of predicted pIC₅₀ values against the hERG channel of the generated refined compounds as compared to that of the respective input compound are shown in [B]. Relevant metrics are shown on each plot.

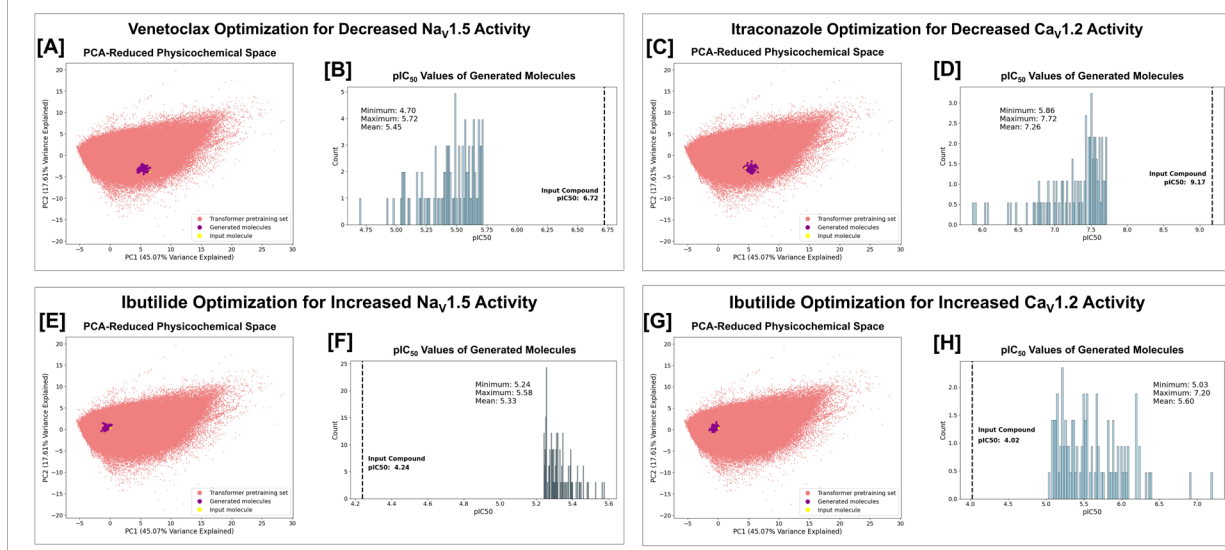


Figure S18. Visualization of the CardioGenAI framework applied to venetoclax [A-B], itraconazole [C-D], and ibutilide [E-H]. In each case, the specified predicted cardiac ion channel pIC_{50} value for each of the generated compounds is set to be at least an improvement of one order of magnitude compared to that of the input compound. For each optimization, the input molecule, the 100 generated refined molecules, and the molecules in the training set for the transformer-based models (approximately 5 million datapoints), are projected into a principal component analysis (PCA)-reduced physicochemical-based space. The input compound is colored yellow, the generated refined compounds are colored purple, and the compounds in the training set of the transformer-based models are colored red. The first two principal components explain 45.07% and 17.61% of the total variance, respectively. For venetoclax, which has a predicted $\text{Na}_v1.5$ pIC_{50} of 6.72, we reduce the $\text{Na}_v1.5$ pIC_{50} by at least one order of magnitude for each generated compound [B]. For itraconazole, which inhibits $\text{Ca}_v1.2$ with a predicted pIC_{50} of 8.17, we reduce the $\text{Ca}_v1.2$ pIC_{50} by at least one order of magnitude for each generated compound [D]. For ibutilide, which has a predicted pIC_{50} for hERG, $\text{Na}_v1.5$, and $\text{Ca}_v1.2$ of 7.98, 4.24 and 4.02, respectively, we independently increase the $\text{Na}_v1.5$ pIC_{50} by at least one order of magnitude for each generated compound [F] and increase the $\text{Ca}_v1.2$ pIC_{50} by at least one order of magnitude for each generated compound [H]. In each case, the CardioGenAI framework is able to identify the region of physicochemical space corresponding to compounds that are similar to the input compound, yet exhibit significantly improved activity against the respective cardiac ion channel. The densities of predicted pIC_{50} values of the generated refined compounds against the respective cardiac ion channel are shown. Relevant metrics are shown on each plot.

Running the CardioGenAI Framework

To optimize a cardiotoxic compound with CardioGenAI, utilize the `optimize_cardiotoxic_drug` function from the `Optimization_Framework` module:

```
from src.Optimization_Framework import optimize_cardiotoxic_drug

optimize_cardiotoxic_drug(input_smiles,
                           herg_activity,
                           nav_activity,
                           cav_activity,
                           n_generations,
                           device)
```

- `input_smiles (str)`: The input SMILES string of the compound that you seek to optimize for reduced cardiac ion channel activity.
- `herg_activity (tuple or str)`: HERG activity for which to filter. If the entry is a string, it must be either 'blockers' or 'non-blockers'. If it is a tuple, it must indicate a range of activity values.
- `nav_activity (tuple or str)`: NaV1.5 activity for which to filter. If the entry is a string, it must be either 'blockers' or 'non-blockers'. If it is a tuple, it must indicate a range of activity values.
- `cav_activity (tuple or str)`: CaV1.2 activity for which to filter. If the entry is a string, it must be either 'blockers' or 'non-blockers'. If it is a tuple, it must indicate a range of activity values.
- `n_generations (int)`: The number of optimized drug candidates to generate. Default is 100.
- `device (str)`: The device to use for the optimization. Must be either 'gpu' or 'cpu'. Default is 'gpu'.

Figure S19. Python function to run the complete CardioGenAI framework.

Performing Inference with the Discriminative Models

To predict activity against the hERG, Nav1.5 and Cav1.2 channels, utilize the `predict_cardiac_ion_channel_activity` function from the `Discriminator` module:

```
from src.Discriminator import predict_cardiac_ion_channel_activity

predict_cardiac_ion_channel_activity(input_data,
                                    prediction_type,
                                    predict_hERG,
                                    predict_Nav,
                                    predict_Cav,
                                    device)
```

- `input_data (str or list)`: The input data for which the discriminative models will process. If the entry is a string, it must be either a SMILES string or a path to a prepared h5 file. If it is a list, it must be a list of SMILES strings.
- `prediction_type (str)`: Either 'regression' or 'classification'. Default is 'regression'.
- `predict_hERG (bool)`: Whether to predict hERG activity. Default is True.
- `predict_Nav (bool)`: Whether to predict Nav1.5 activity. Default is False.
- `predict_Cav (bool)`: Whether to predict Cav1.2 activity. Default is False.
- `device (str)`: The device to use for the inference computations. Must be either 'gpu' or 'cpu'. Default is 'gpu'.

Figure S20. Python function to perform inference with the discriminative models.