

"Illustration reproduced courtesy of Cell Signaling Technology, Inc. (www.cellsignal.com)"

Finding CAMK2G Selective Drug Candidates

Leela S. Dodda

Highlight

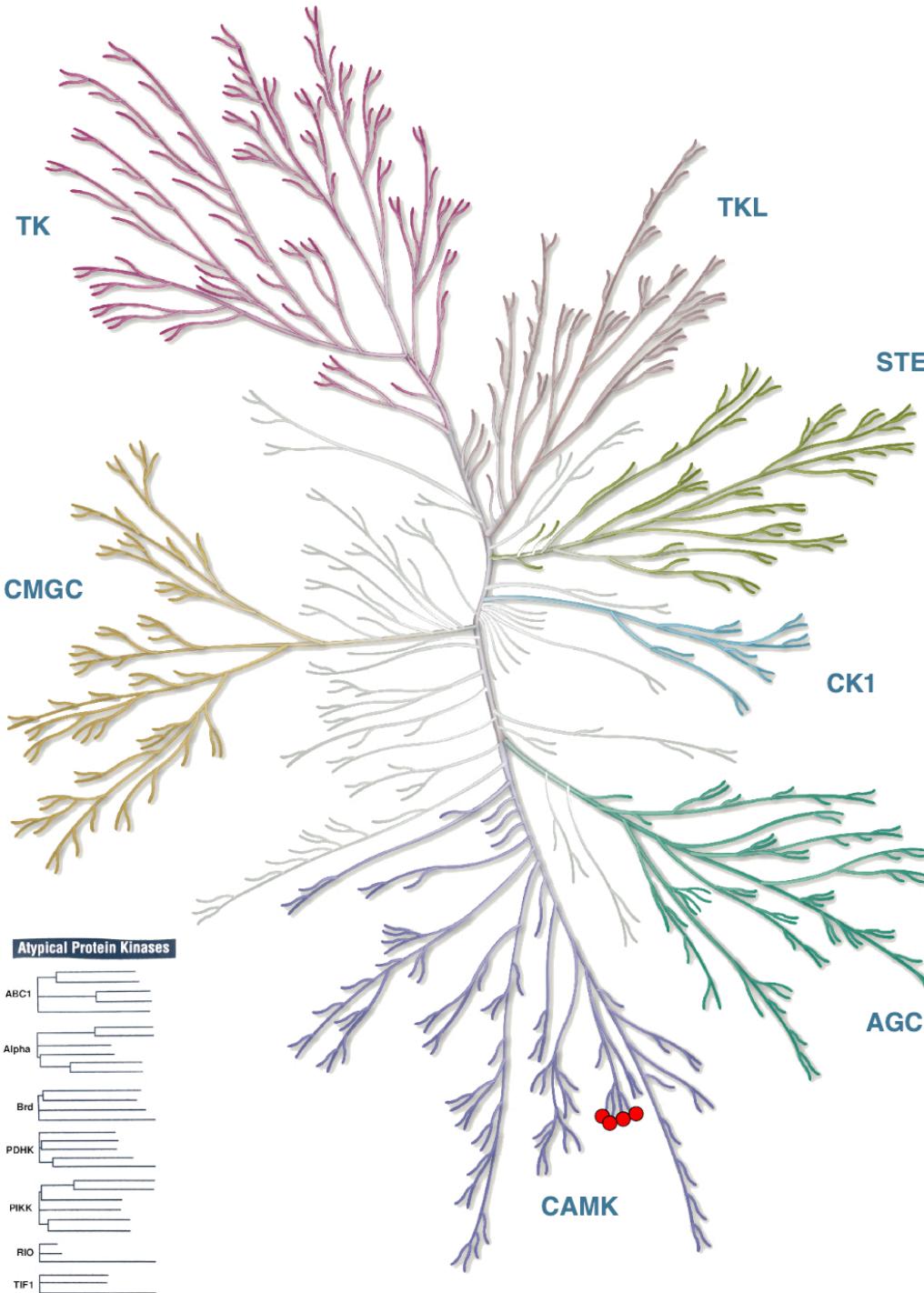
Annotation

- Chain
 - Sequence conflict
 - Modified residue
 - Beta strand
 - Alternative sequence
 - Helix
 - Region
 - Active site
 - Mutagenesis
 - Turn
 - Binding site
 - Natural variant
 - Nucleotide binding
 - Initiator methionine
 - Domain

Amino acid properties

 - Similarity
 - Hydrophobic
 - Negative
 - Positive
 - Aliphatic
 - Tiny
 - Aromatic
 - Charged
 - Small
 - Polar
 - Big
 - Serine Threonine

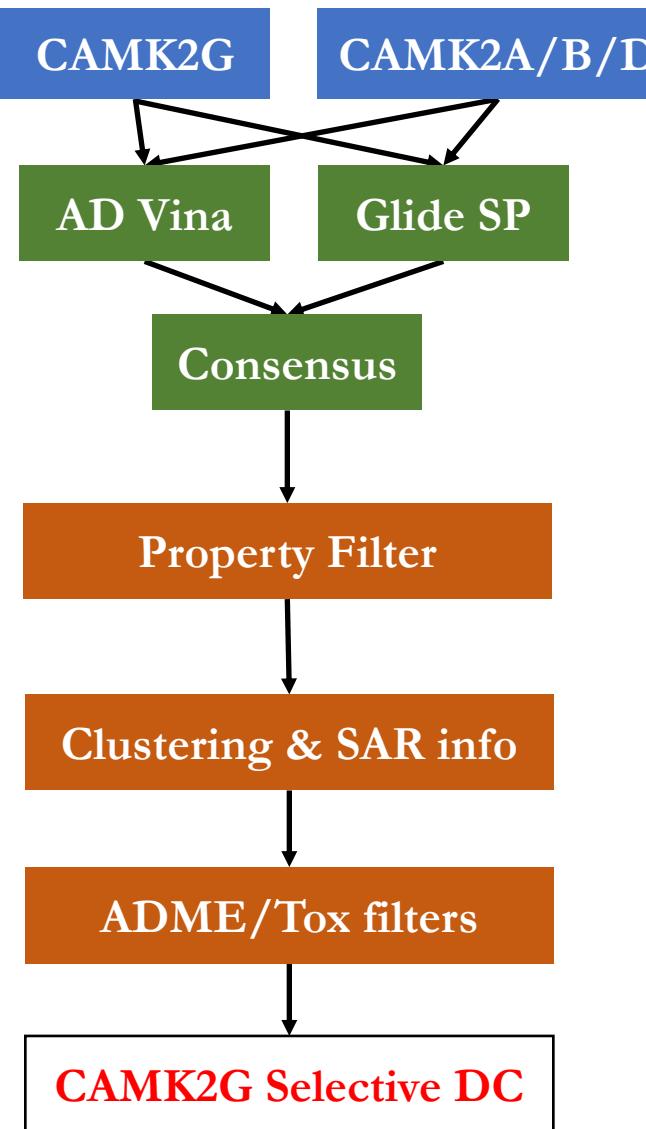
Plan of Action



- Best way to approach the design a selective CAMK2G drug candidate is by combining both ligand-based (LBDD) and structure-based (SBDD) drug design approaches
- Because of the structural similarity, achieving selectivity by targeting highly conserved ATP binding site might be very challenging
- In such cases, allosteric pockets in CAMK2G can be targeted to achieve selectivity as these regions tend to be less conserved than ATP binding site
- SBDD/LBDD plan presented in the following slides are for targeting ATP binding .
- For targeting allosteric binding site, more detailed MD and biophysics based structural analyses need to be performed.

SBDD Pipeline

Target: Find CAMK2G selective Drug Candidate



Dock using appropriate X-tal/MD relaxed structures of different isoforms of CAMK2 protein

Consensus of at Top 10% from at least 2 different scoring functions.. Bias the selection via following criteria to achieve selectivity.

$$LigCoreRMSD_{\gamma \leftrightarrow \alpha\beta\delta} < 2.0 \text{\AA} \quad E_{\gamma}^{int} - E_{\alpha\beta\delta}^{int} < 1.5 \text{kcal/mol}$$

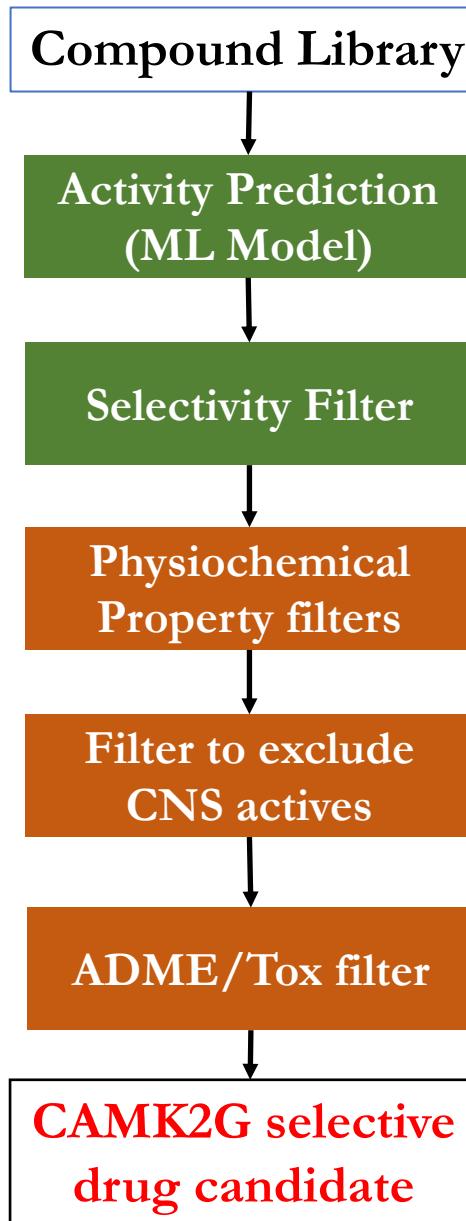
Kinase specific physio chemical property filters. De-prioritize compounds with good BBB penetration. α and β isoforms are specific to brain
([github/rd_filters](#))

Cluster using Cheminformatics and identify compounds/series with SAR. Prioritize according to Synthetic Accessibility score. ([RDKit blogs](#))

Pick the series with fewer ADME/Tox liabilities, along with a backup series. Take into account IP around the compounds during selection process
([github/eToxPred & DeepChem ADME Predictor](#))

LBDD Pipeline

Target: Find CAMK2G selective Drug Candidate



Ensemble scoring with RF regression models trained using MACCS, Pharmacophore & ECFP fps. Data obtained from ChEMBL (~ 50-100 data points per isoform) (**RDKit, Scikit-learn & Deepchem**)

ML model that can predict activity w.r.t CAMK2A/B, and select compounds that have high CAMK2G score and low CAMK2A/B score (**RDKit, Scikit-learn & Deepchem**)

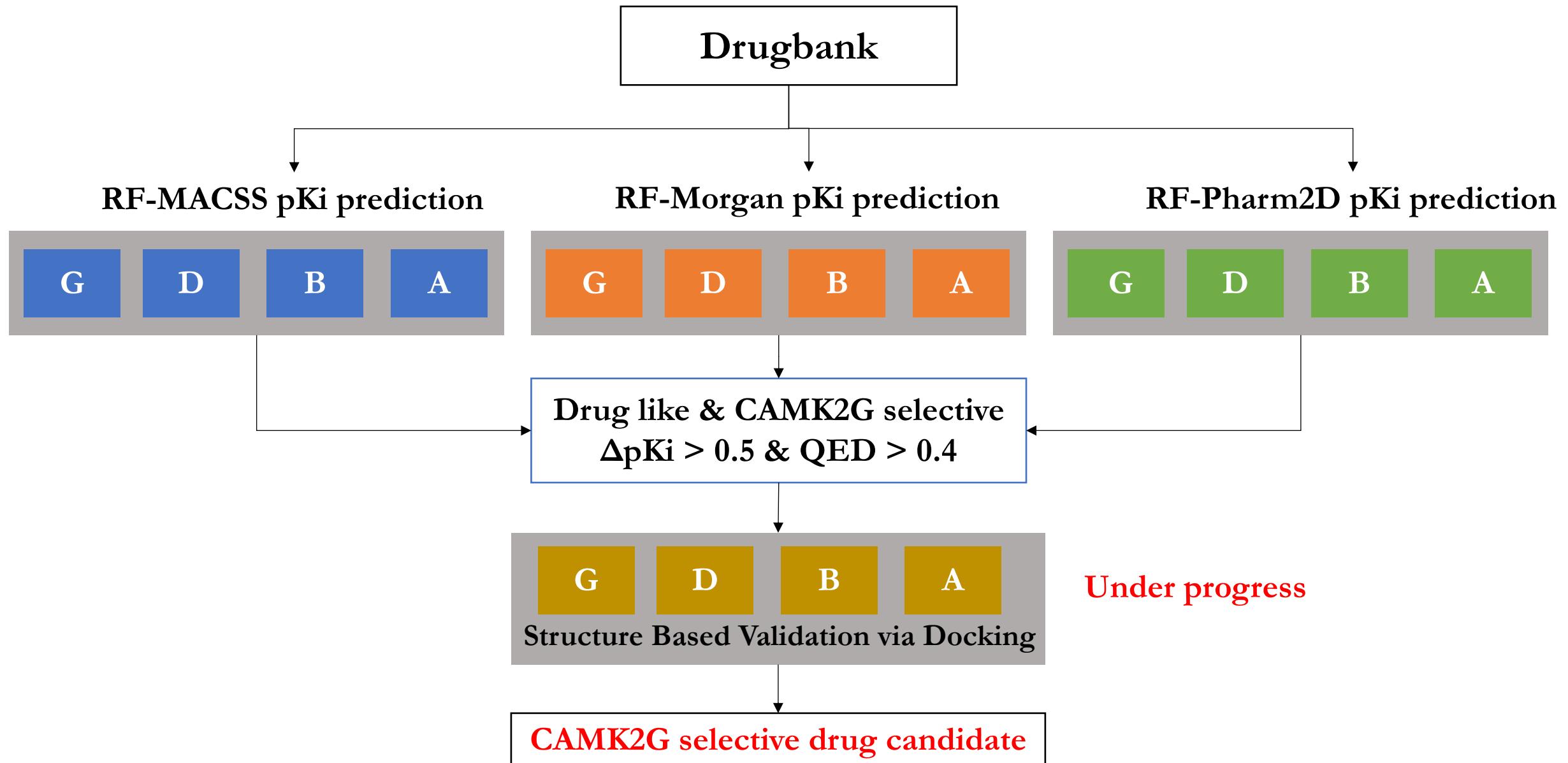
Kinase specific filters ([github/rd_filters](#))

Since CAMK2A/B are brain specific, exclude compounds that can cross Blood Brain Barrier (BBB) ([github/rd_filters](#))

Pick the best series with fewer ADME/Tox liabilities, along with a backup series. Take into account IP around the compounds during selection process (**DeepChem & eToxPred**)

Methodology

Finding CAMK2G Selective DC via Drug Repurposing



Packages used

- KNIME & Python for extracting CHEMBL data
- Sklearn & deepchem for crafting RF models & hyper-parameter optimization
- RDKit is used for featurization of molecules for RF models
- MACCS fp & 2D pharmacophore fingerprint functionalities are implemented into deepchem using RDKit
- Jupyter notebook are used for preparing Drugbank DB for screening using the ML models
- RDKit, Openbabel for preparing the ligands for preparation of ligands for docking
- Smina for docking & Open Drug Discovery Toolkit (ODDT) for analysis of structures

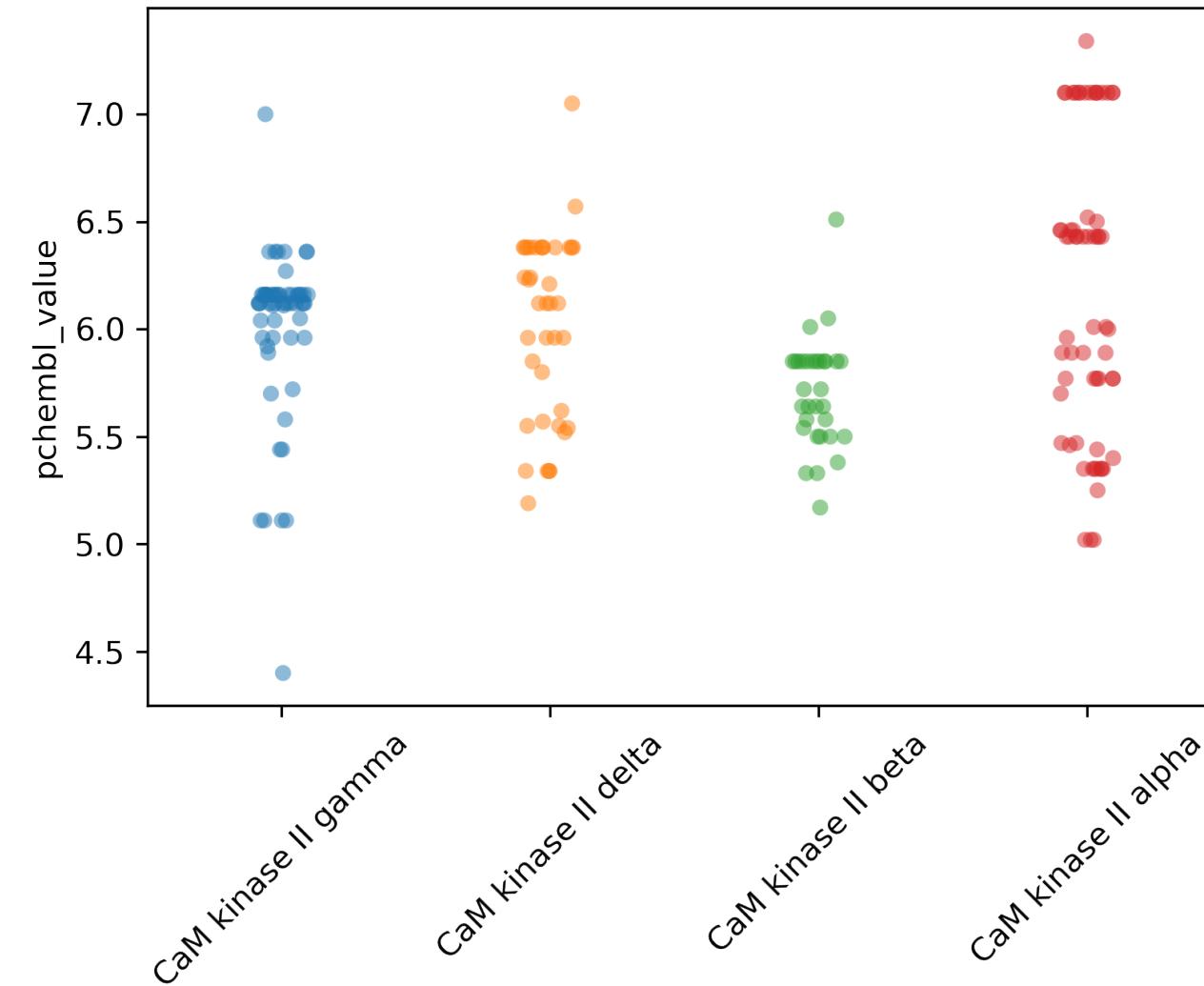
Results & Summary

Which Library to Screen? Drug Bank

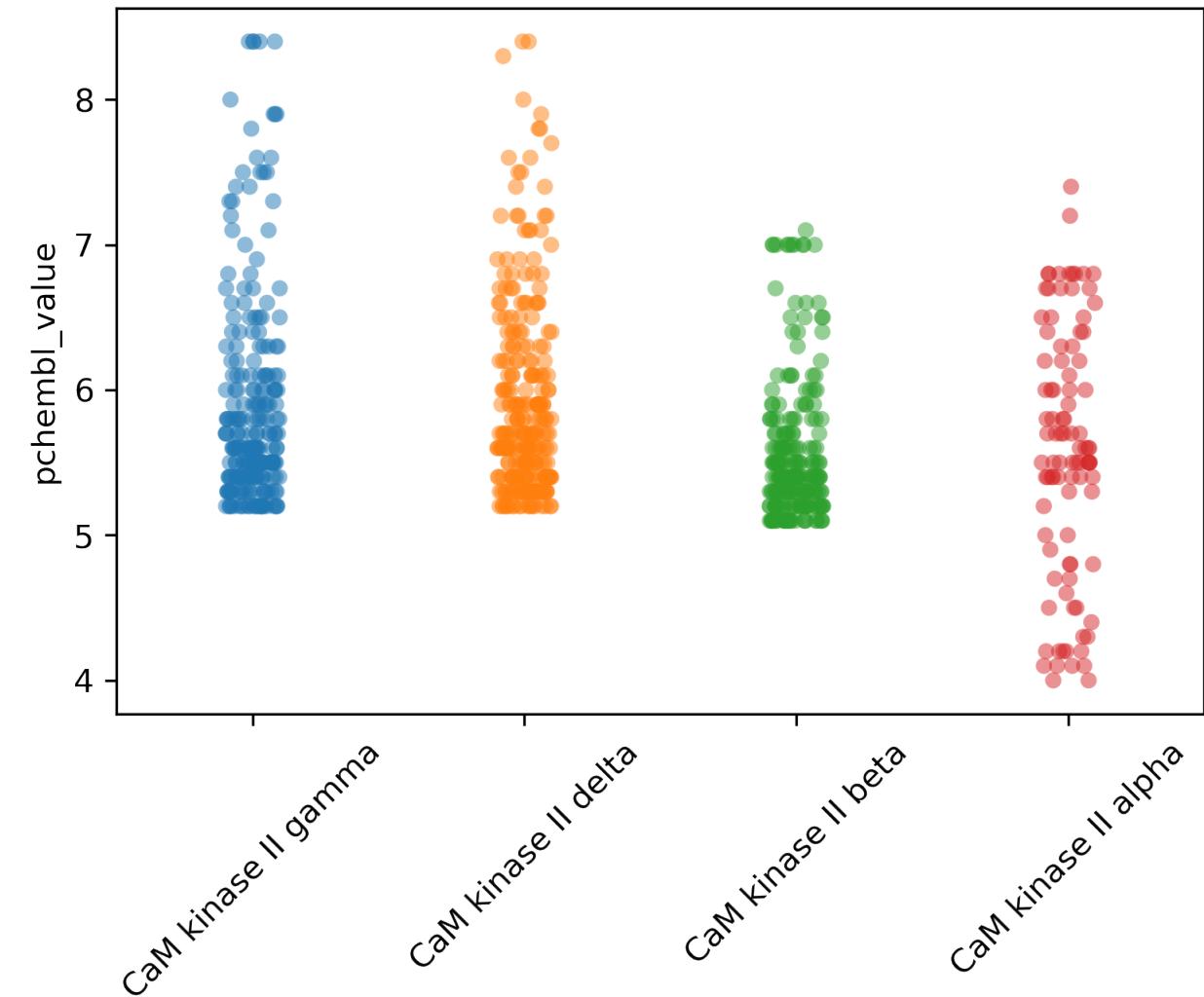
- The aim of this exercise is to find CAMK2G selective drug candidate
- Most if not all the compounds in Drug Bank are drug candidates with acceptable ADME-Tox profiles
- Any hits with this library can thus be readily repurposed for targeting CAMK2G selectively

CHEMBL Data on CAMK2 Isoforms

Functional pKd Data from CHEMBL

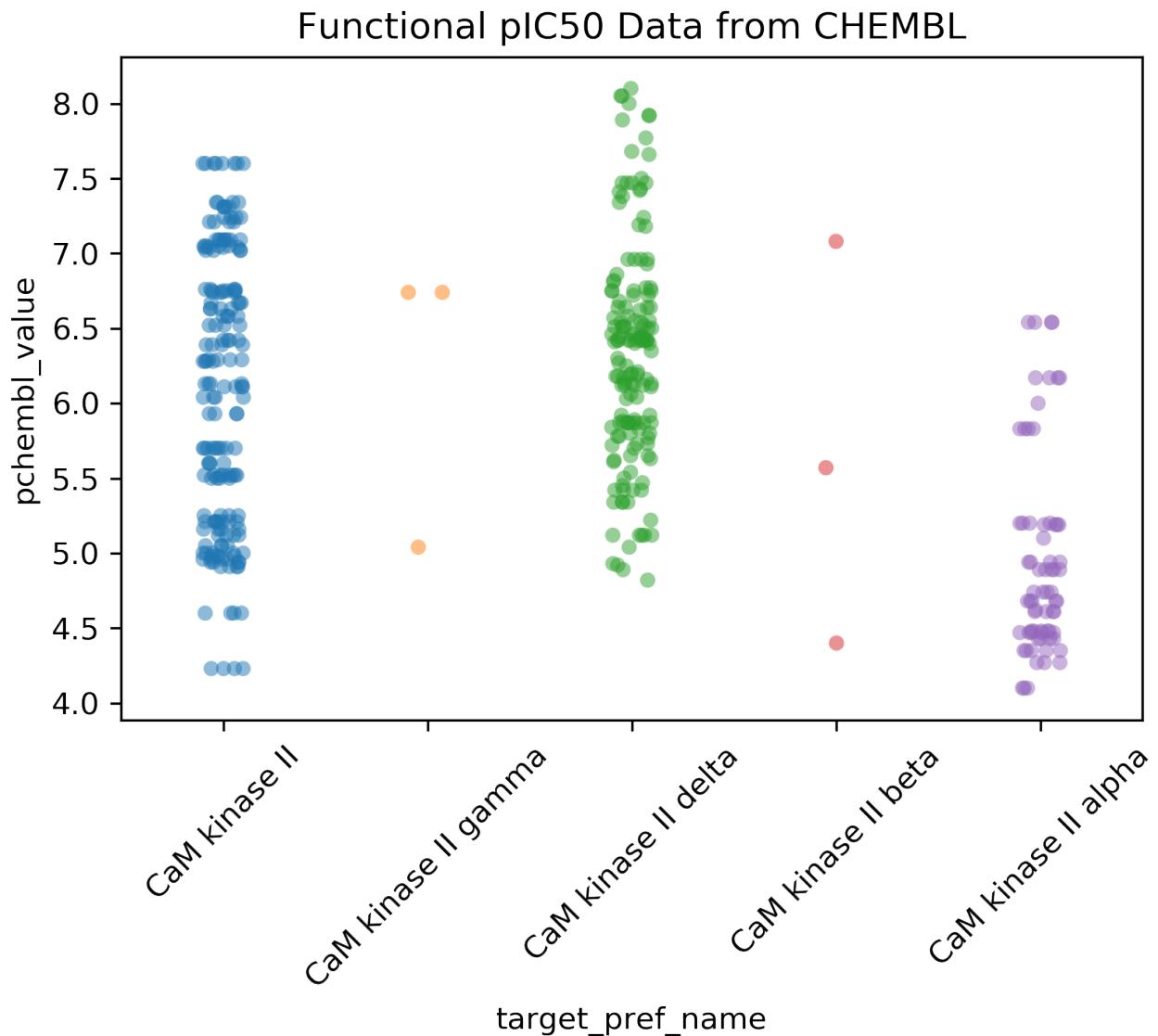


Functional pKi Data from CHEMBL



pK_d and pK_i data are the best ones to use for training/validating LBDD/SBDD approaches

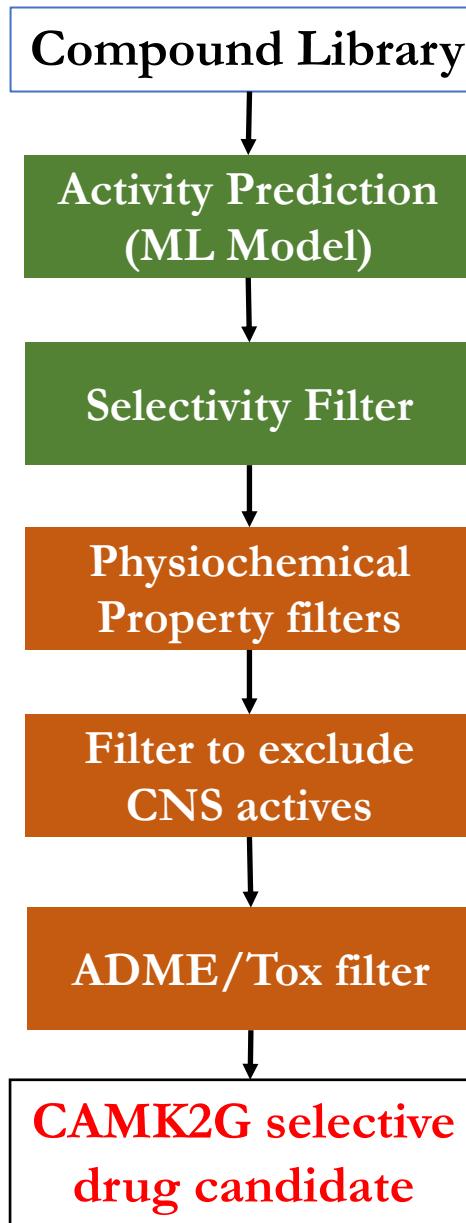
CHEMBL Data on CAMK2 Isoforms



pIC50 data are not the best ones to use as there is not enough data for each isoform

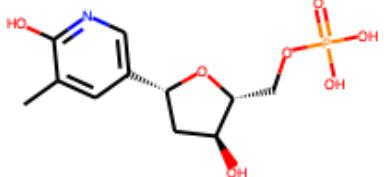
LBDD Pipeline

Target: Find CAMK2G selective Drug Candidate

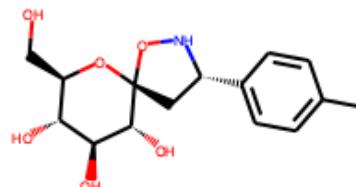


- Drug bank molecules are downloaded and cleaned by removing compounds with MW > 700, and thus remained molecules are used as screening library
- For activity prediction, Random forest models are created with deepchem for each of the 4 isoforms.
 - Before training the models, CHEMBL data was cleaned to remove compounds with PAINS functional groups
 - Models were created with (I) Morgan fingerprint, (II) MACCS fingerprint(II) and (III) 2D pharmacophore finger prints.
 - MACCS fingerprint and 2D pharmacophore finger print functionalities are implemented into deepchem for this exercise.
 - For each model best hyperparameters were obtained using GridSearch CV to find the model with least RMSE and best R2_score

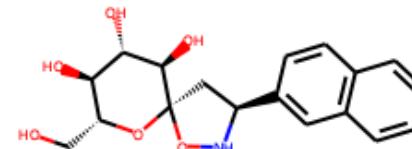
Hits obtained via LBDD approach



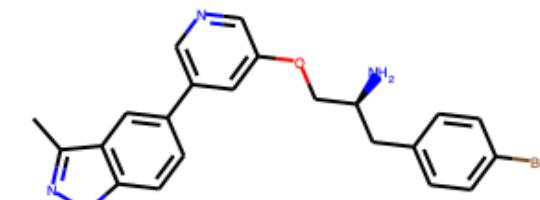
DB04060(QED 0.59)



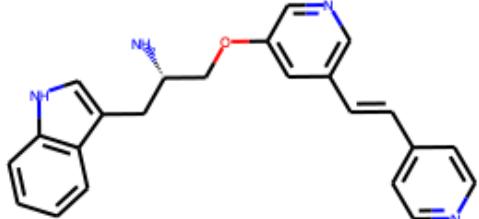
DB08503(QED 0.48)



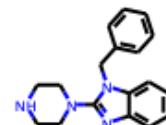
DB08500(QED 0.52)



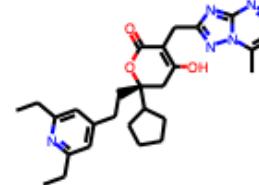
DB08569(QED 0.46)



DB07107(QED 0.51)

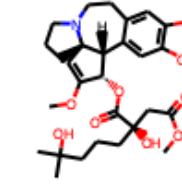


DB12964(QED 0.80)



DB11878(QED 0.43)

Investigational



approved

DB04865(QED 0.45)

- Drugbank library is then screened with each type of models, and for each model compounds with $pKi > 6.5$, $pSel > 0.5$ (with respect to other 3 isoform models) and $QED > 0.4$
- All compounds that satisfy the above criteria from all three models were then selected for further analysis

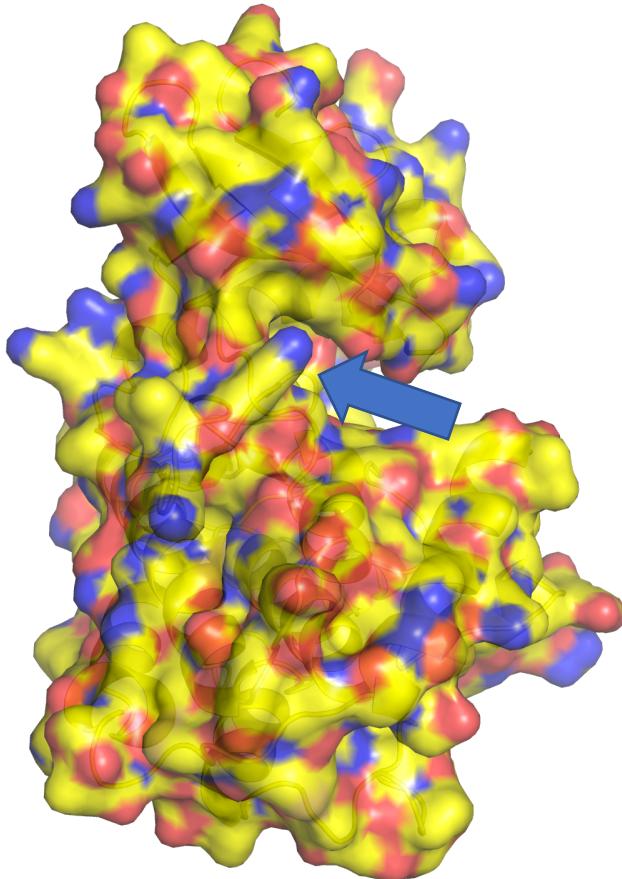
Hits obtained via LBDD approach

DB_ID	GROUP	RO5 VIOL	CHARGE	LOGP
DB08569	experimental	1	1	4.16
DB07107	experimental	1	1	3.12
DB04060	experimental	1	-2	-0.82
DB12964	investigational	1	1	3.01
DB11878	investigational	0	-1	5.13
DB08503	experimental	1	0	-0.33
DB08500	experimental	1	0	0.14
DB04865	approved	0	1	2.09

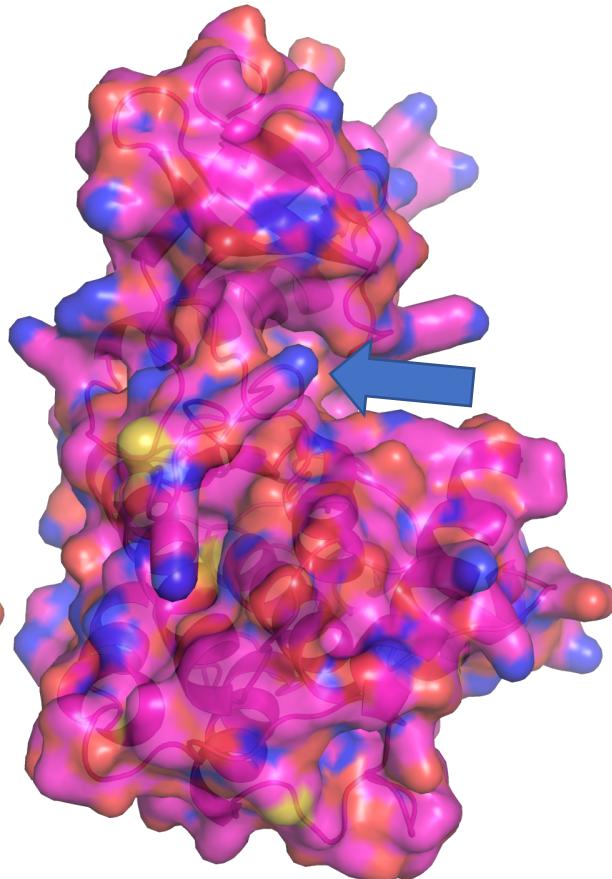
- How can we narrow down to one of the following compounds ?

Docking to rank/prioritize Hits

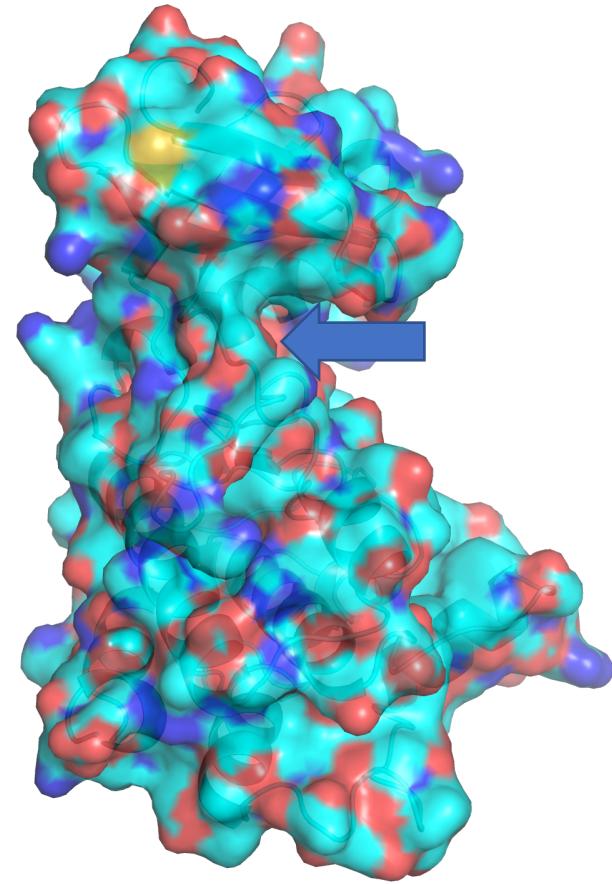
CAMK2A



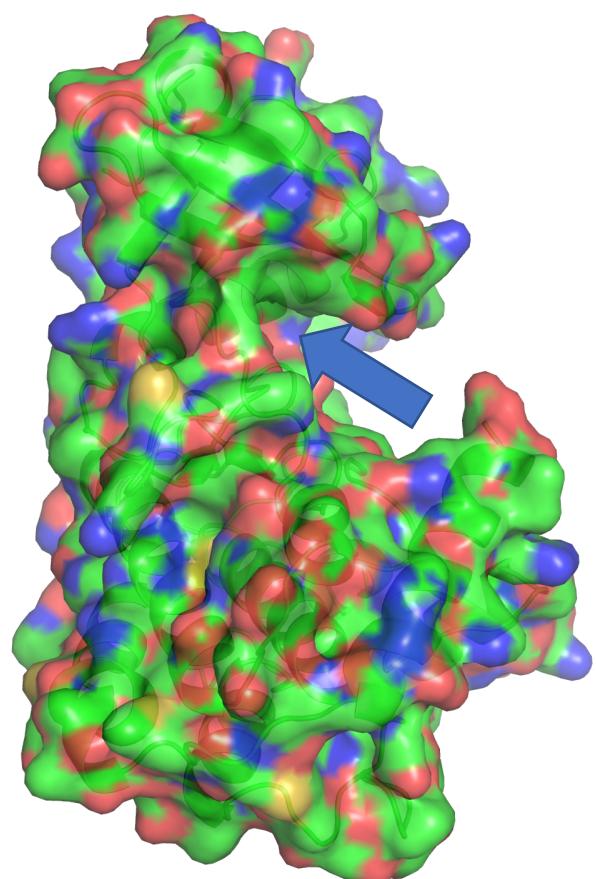
CAMK2B



CAMK2D



CAMK2G



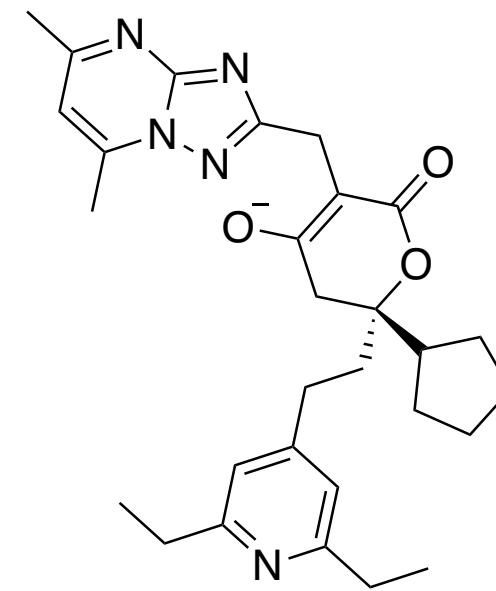
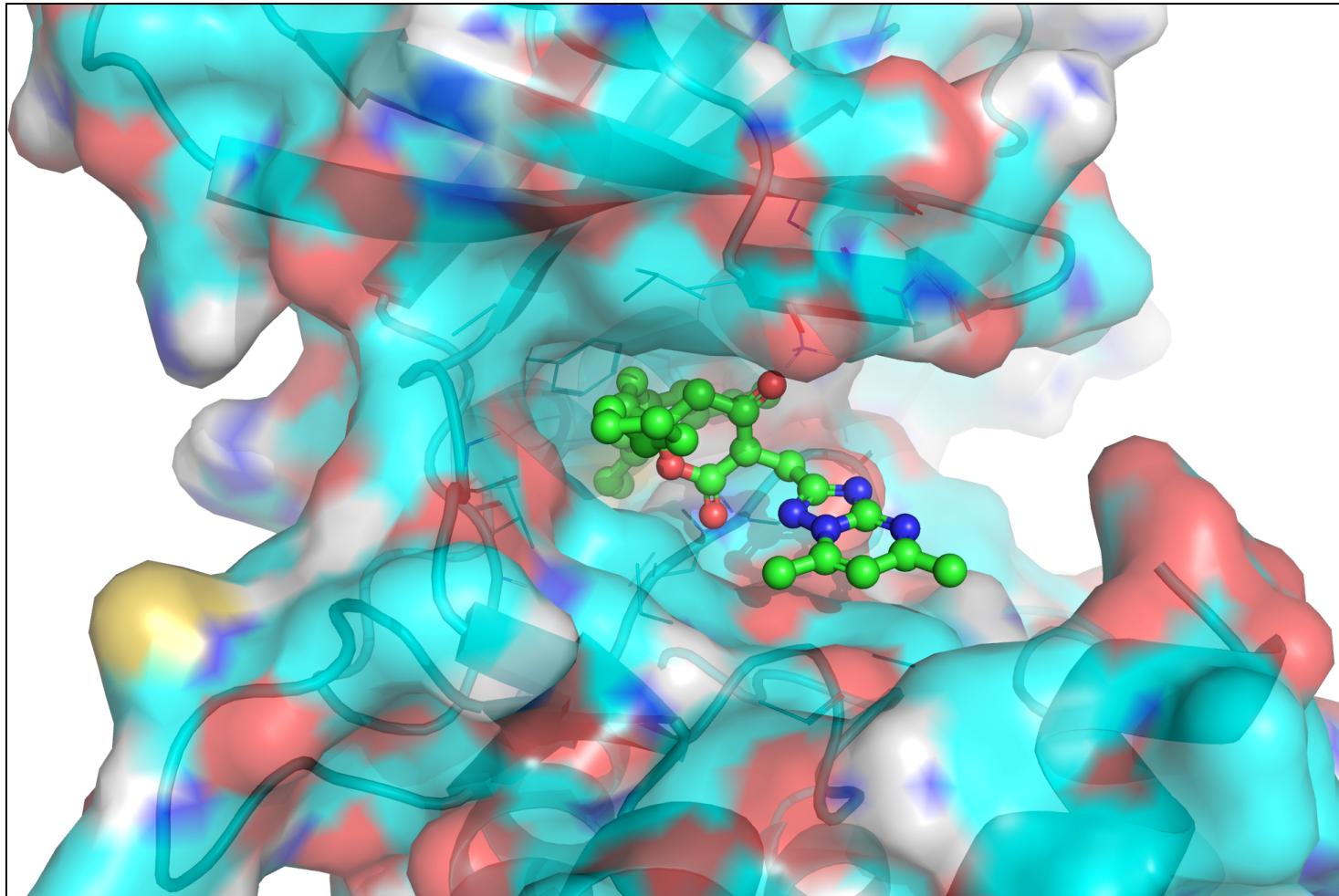
- Models are created using SwissModel webserver, based on existing holo structures
- **A thorough structure based analysis for these hits is still ongoing**

With out SB insights, Which one to pick ?

DB_ID	GROUP	RO5 VIOL	CHARGE	LOGP
DB08569	experimental	1	1	4.16
DB07107	experimental	1	1	3.12
DB04060	experimental	1	-2	-0.82
DB12964	investigational	1	1	3.01
DB11878	investigational	0	-1	5.13
DB08503	experimental	1	0	-0.33
DB08500	experimental	1	0	0.14
DB04865	approved	0	1	2.09

Highlighted molecule has a negative charge and LogP>3 with high MW. These three features probably make it less BBB permeable, and can help with selectivity against CAMK2A/B isoforms

Selected CAMK2G compound



DB11878/ Filibuvir

What more can is needed to improve this work ?

- Craft a better library for both SB and LB approaches
 - Screen against commercially available & virtual kinase like inhibitors
 - Train a RNN smiles generator to generate kinase inhibitor like structures to enrich the library (github: [isayev/ReLeaSE](#))
- Ligand Based Approach
 - More validation around RF-models. To assess if the RF model is actually placing the appropriate importance to the key features. (deepchem & rdkit in combo with lime or eli5)
 - Create 3D pharmacophore models for LBVS using LigandScout / (github: DrrDom/pmapper)
- Structure Based Approach
 - Use expt pKd data to benchmark the scoring functions and ability of docking programs to reproduces X-tal poses.
 - Include protein-ligand interaction fingerprints to select compounds forming specific interactions (ODDT)
 - Use MM/GBSA & MD to further refine poses and rank them appropriately.