Molecular Property Prediction Using Machine Learning

TPD Assay Summit

San Diego, CA

December 8, 2022

Important Notice and Disclaimers

This presentation contains statements that relate to future events and expectations and as such constitute forward-looking statements within the meaning of the Private Securities Litigation Reform Act of 1995. When or if used in this presentation, the words "anticipate," "believe," "could," "estimate," "expect," "intend," "may," "outlook," "plan," "predict," "should," "will," and similar expressions and their variants, as they relate to Nurix Therapeutics, Inc. ("Nurix", the "Company," "we," "us" or "our"), may identify forward-looking statements. All statements that reflect Nurix's expectations, assumptions or projections about the future, other than statements of historical fact, are forward-looking statements, including, without limitation, statements regarding our future financial or business plans; our future performance, prospects and strategies; future conditions, trends, and other financial and business matters; our current and prospective drug candidates; the planned timing and conduct of the clinical trial programs for our drug candidates; the planned timing for the provision of clinical updates and initial findings from our clinical studies; the potential advantages of our DELigase™ platform and drug candidates; the extent to which our scientific approach and DELigase™ platform may potentially address a broad range of diseases; the extent animal model data predicts human efficacy; and the timing and success of the development and commercialization of our current and anticipated drug candidates. Forward-looking statements reflect Nurix's current beliefs, expectations, and assumptions. Although Nurix believes the expectations and assumptions reflected in such forward-looking statements are reasonable, Nurix can give no assurance that they will prove to be correct. Forward-looking statements are not guarantees of future performance and are subject to risks, uncertainties and changes in circumstances that are difficult to predict, which could cause Nurix's actual activities and results to differ materially from those expressed in any forward-looking statement. Such risks and uncertainties include, but are not limited to: (i) risks and uncertainties related to Nurix's ability to advance its drug candidates, obtain regulatory approval of and ultimately commercialize its drug candidates; (ii) the timing and results of clinical trials; (iii) Nurix's ability to fund development activities and achieve development goals; (iv) the impact of the COVID-19 pandemic on Nurix's business, clinical trials, financial condition, liquidity and results of operations; (v) Nurix's ability to protect intellectual property and (vi) other risks and uncertainties described under the heading "Risk Factors" in Nurix's Quarterly Report on Form 10-Q for the fiscal quarter ended August 31, 2022, and other SEC filings. Accordingly, readers are cautioned not to place undue reliance on these forward-looking statements. The statements in this presentation speak only as of the date of this presentation, even if subsequently made available by Nurix on its website or otherwise. Nurix disclaims any intention or obligation to update publicly any forward-looking statements, whether in response to new information, future events, or otherwise, except as required by applicable law.

Certain information contained in this presentation relates to or is based on studies, publications, surveys and other data obtained from third-party sources and the Company's own internal estimates and research. While the Company believes these third-party sources to be reliable as of the date of this presentation, it has not independently verified, and makes no representation as to the adequacy, fairness, accuracy or completeness of, any information obtained from third-party sources. In addition, all of the market data included in this presentation involves a number of assumptions and limitations, and there can be no guarantee as to the accuracy or reliability of such assumptions. Finally, while we believe our own internal estimates and research are reliable, such estimates and research have not been verified by any independent source.

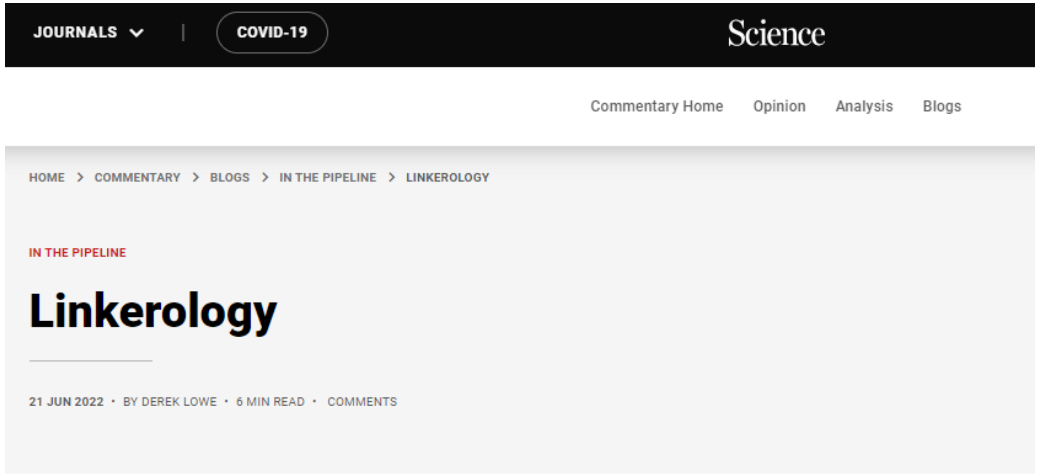
Presentation Outline

- Degrader optimization is difficult due to their unique chemical space and the lack of predictive tools to guide development
- Nurix has successfully developed machine learning models able to accurately predict key properties of degraders, such as solubility and permeability.
- Our models provide not only useful computational tools, but can be leveraged to inform human understanding to assist in degrader development

Degrader Optimization Remains Largely Empirical

...**We're still in the "try 'em and see" stage of optimization**. … It would be much, much nicer if we could stand at the whiteboard or look at a screen, pursing our lips thoughtfully and then pointing purposefully at The Compound To Make, but we ain't there yet.

Derek Lowe
In The Pipeline
6/21/22



I've written a number of times here about bifunctional protein degraders, which have been a big topic in drug discovery for the past few years. There's a new paper that illustrates some of the challenges in this area, and it's worth using as an example.

For those outside the field, the idea behind these things is pretty straighforward, at least in principle. You find a protein that you think is involved in a disease process, one whose activity you would like to dial down. You find a small-molecule ligand that binds that protein - you may already have some inhibitors around, in house or from the literature, and for these purposes your small molecule doesn't even have to be an inhibitor, just a binder. (Of course, the way we run assays means that most of the time we're not set up to detect silent binders, so those are thinner on the ground). Now you break out your synthetic organic chemistry skills and build out a linker group from that known ligand, and at the other end of that linker you attach a known ligand for an "E3 ligase" enzyme. There are several possibilities, but so far the well-established ones for the enzymes cereblon and VHL are the ones that get used the most, by far.

Degrader Molecules Are Outside the Domain of Applicability for Commonly Applied Intuition and Rules

Targeted Protein Degrader molecules occupy a property space well beyond the traditional Lipinski RO5 molecules.

As a result, intuition and rules for predicting physicochemical properties developed based on RO5 small molecules fail when applied to TPD molecules.

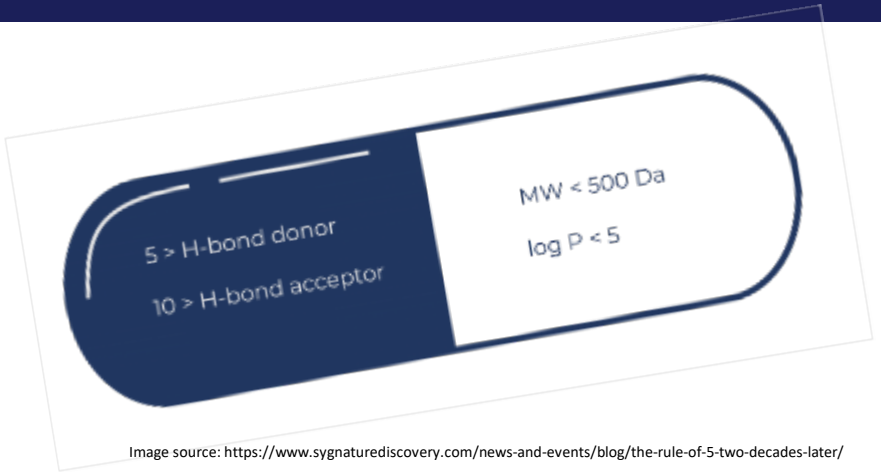
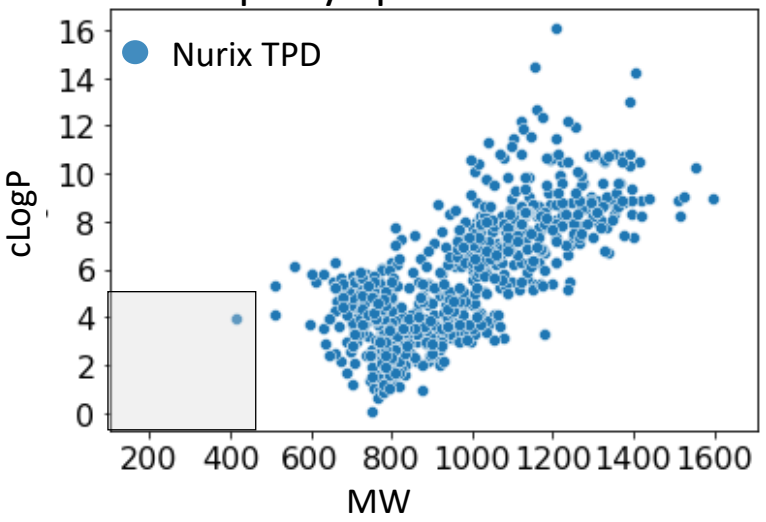
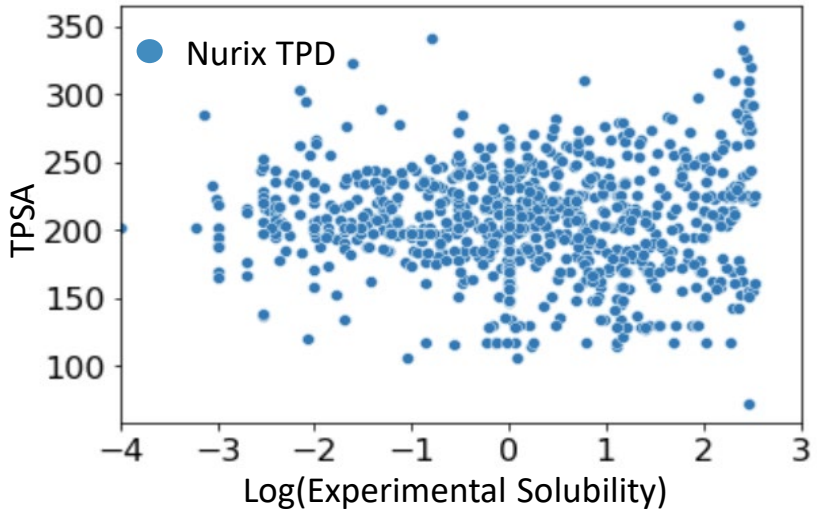


Image source: https://www.sygnaturediscovery.com/news-and-events/blog/the-rule-of-5-two-decades-later/

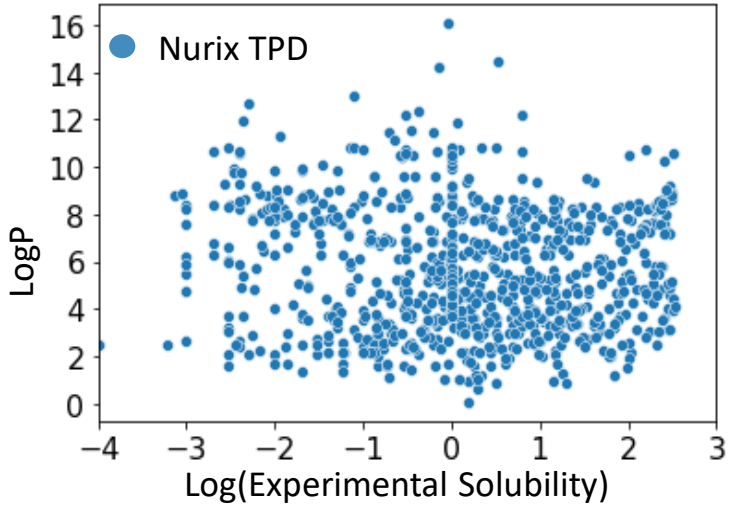
Property Space of TPD



TPSA vs Experimental KSol



CLogP vs Experimental KSol



Degrader Molecules Are Outside the Domain of Applicability for Commonly Applied Intuition and Rules

Targeted Protein Degrader molecules occupy a property space well beyond the traditional Lipinski RO5 molecules.

As a result, intuition and rules for predicting physicochemical properties developed based on RO5 small molecules fail when applied to TPD molecules.

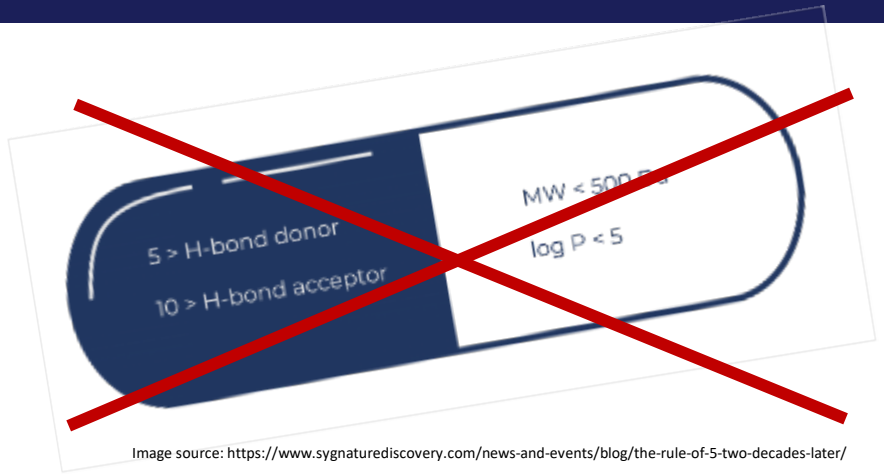
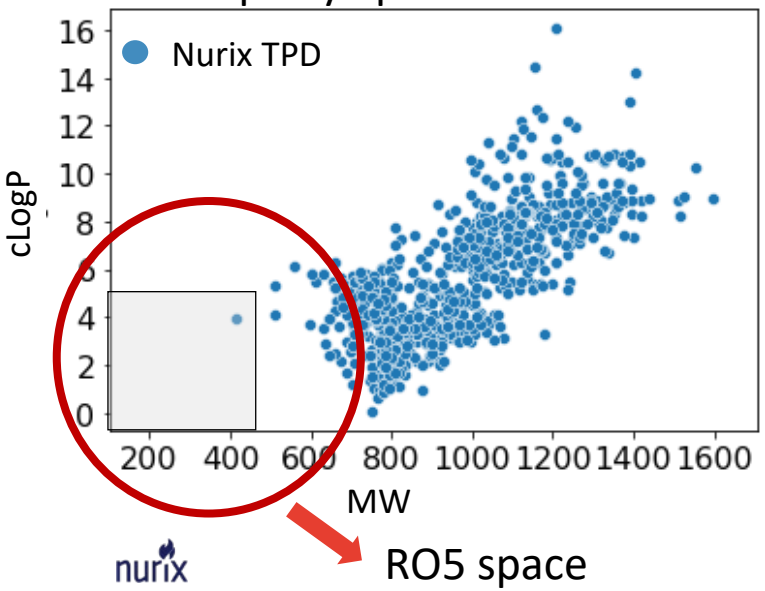
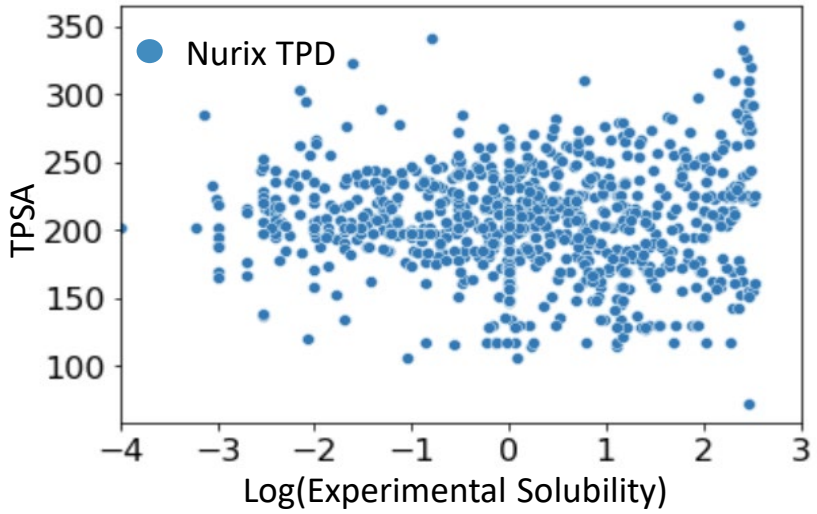


Image source: https://www.sygnaturediscovery.com/news-and-events/blog/the-rule-of-5-two-decades-later/

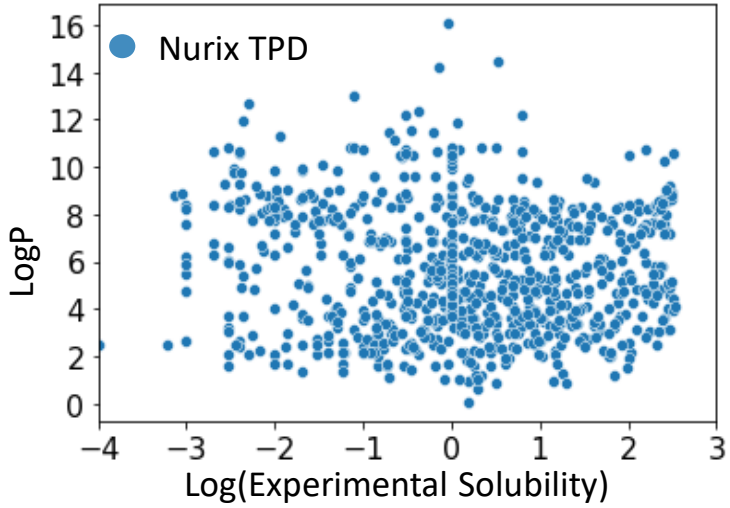
Property Space of TPD



TPSA vs Experimental KSol



CLogP vs Experimental KSol



Poor Correlation Between Ksol and 1D Molecular Properties

Extended view of weak correlation between molecular descriptors and solubility.

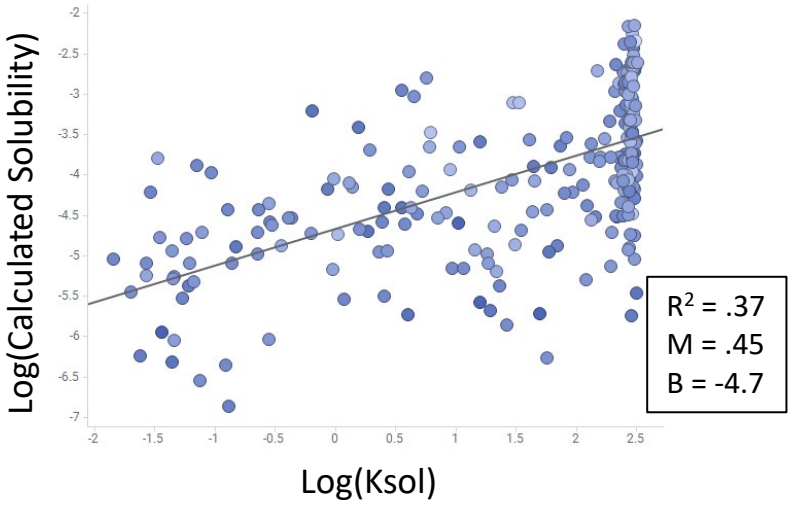
	R-squared
Kinetic Solubility X	1.000000
MolWt	0.117218
LogP	0.130907
NumHAcceptors	0.047313
NumHDonors	0.022847
NumHeteroatoms	0.079500
NumRotatableBonds	0.009230
NumHeavyAtoms	0.117752
NumAliphaticCarbocycles	0.016792
NumAliphaticHeterocycles	0.062368
NumAliphaticRings	0.085996
NumAromaticCarbocycles	0.032745
NumAromaticHeterocycles	0.044656
NumAromaticRings	0.132261
RingCount	0.185393
FractionCSP3	0.002377
TPSA	0.056036

A more thorough evaluation of 1D properties commonly thought to affect or relate to the solubility of small molecules shows very low correlation to experimental data.

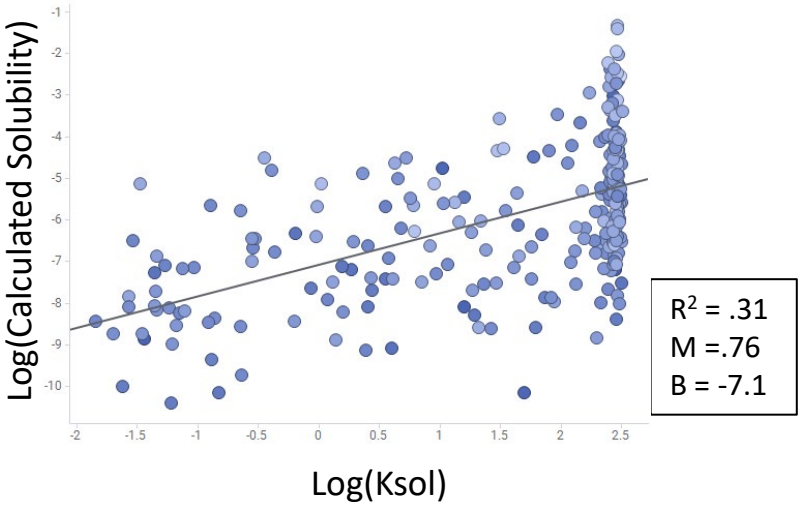
nurix

More Sophisticated Cheminformatics Calculators Similarly Struggle to Generalize to Degraders

Well Used Comp Chem Software #1



Well Used Comp Chem Software #2



Popular CompChem and Cheminformatics software similarly fails to accurately predict relevant properties of Degrader compounds.

Active Literature Addressing Property Prediction in Bro5 Space

ACS Publications

Impact of Dynamically Exposed Polarity on Permeability and Solubility of Chameleonic Drugs Beyond the Rule of 5

Use of 3D Properties to Characterize Beyond Rule-of-5 Property Space for Passive Permeation

How Beyond Rule of 5 Drugs and Clinical Candidates Bind to Their Targets

Bradley C. Doak, Jie Zheng, Doreen Dobritzsch, and Jan Kihlberg*

Cite this: J. Med. Chem. 2016, 59, 6, 2312−2327
Publication Date: October 12, 2015
https://doi.org/10.1021/acs.jmedchem.5b01286
Copyright © 2015 American Chemical Society

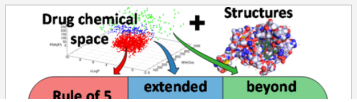
Article Views 13630 | Altmetric 41 | Citations 200

PDF (6 MB) | Supporting Info (1) » | SUBJECTS: Interfaces, Ligands, Macrocycles,

Journal of Medicinal Chemistry

Abstract

To improve discovery of drugs for difficult targets, the opportunities of chemical space beyond the rule of 5 (bRo5) were examined by retrospective analysis of a comprehensive set of structures for complexes between drugs and clinical candidates and their targets. The analysis illustrates the



Chemistry & Biology
Review

CellPress

Oral Druggable Space beyond the Rule of 5: Insights from Drugs and Clinical Candidates

Bradley Croy Doak,[1] Björn Over,[2] Fabrizio Giordanetto,[3] and Jan Kihlberg[1,*]
[1]Department of Chemistry, BMC, Uppsala University, Box 576, 751 23 Uppsala, Sweden
[2]CVMD iMed, AstraZeneca R&D Mölndal, 431 83 Mölndal, Sweden
[3]Medicinal Chemistry, Taros Chemicals GmbH & Co. KG, Emil-Figge-Strasse 76a, 44227 Dortmund, Germany
*Correspondence: jan.kihlberg@kemi.uu.se
http://dx.doi.org/10.1016/j.chembiol.2014.08.013

The rule of 5 (Ro5) is a set of in silico guidelines applied to drug discovery to prioritize compu
increased likelihood of high oral absorption. It has been influential in reducing attrition due to po
kinetics over the last 15 years. However, strict reliance on the Ro5 may have resulted in lost
particularly for difficult targets. To identify opportunities for oral drug discovery beyond the
we have comprehensively analyzed drugs and clinical candidates with molecular weight (M
We conclude that oral drugs are found far bRo5 and properties such as intramolecular hydro
macrocyclization, dosage, and formulations can be used to improve bRo5 bioavailability. Nat
and structure-based design, often from peptidic leads, are key sources for oral bRo5 drugs. T
ar interest for di

Drug discovery beyond the 'rule-of-five'

Ming-Qiang Zhang ✉, Barrie Wilkinson

Show more

+ Add to Mendeley | Share | Cite

https://doi.org/10.1016/j.copbio.2007.10.005 Get rights and content

Although a very useful guideline for orally bioavailable small-molecule drug design, the 'rule-of-five' (also known as 'Lipinski's rule of drug-likeness') has to some extent been overemphasized. Firstly, only 51% of all FDA-approved small-molecule drugs are both used orally and comply with the 'rule-of-five'. This does not even include the increasing number of biologicals of which several have reached 'blockbuster' status. Secondly, it does not cover natural product and semisynthetic natural product drugs, which constitute over one-third of all marketed small-molecule drugs. A more balanced and programmatic approach to drug discovery should be more productive than to rely on an overemphasis of 'rule-of-five' compliance. Rather it should consider proactively the

Opportunities and guidelines for discovery of orally absorbed drugs in beyond rule of 5 space

Vasanthanathan Poongavanam [1, 3], Bradley C. Doak [2, 3], Jan Kihlberg [1] ✉

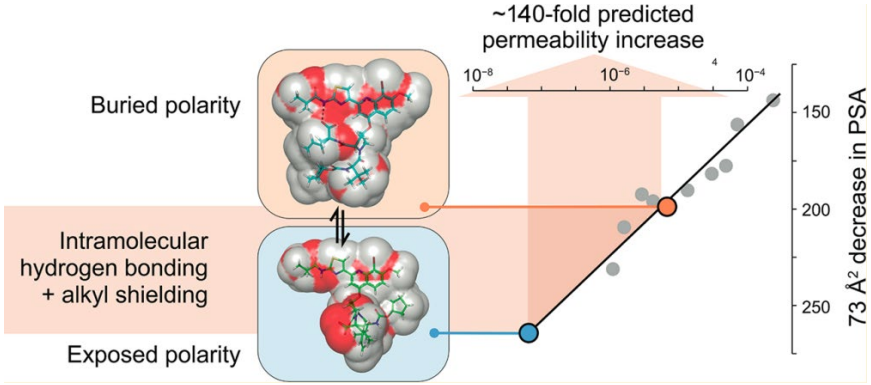
Show more

+ Add to Mendeley | Share | Cite

https://doi.org/10.1016/j.cbpa.2018.05.010 Get rights and conten

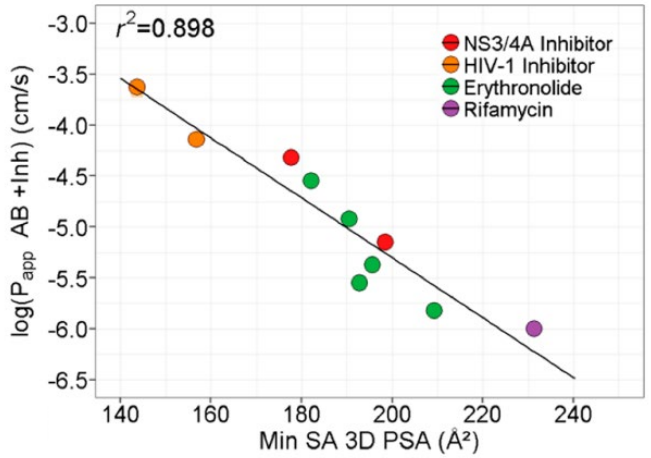
Recent years have seen a dramatic increase in the number of drugs approved in chemical space outside of Lipinski's rule of 5, that is in what has been termed beyond rule of 5 (bRo5) space. The development of three major classes of oral drugs that treat HIV and HCV infections and the growing evidence that novel, difficult targets can be accessed has prompted research into understanding design of drugs displaying cell permeability, solubility and ultimately oral bioavailability in bRo5 space. Studies have found a consistent outer property limit for a reasonable chance of *de novo* designing oral bioavailability. In addition, several property-based guidelines, along with incorporation of chameleonic features, have emerged as strategies to aid design in bRo5 space. A more

Literature Models Fail to Translate to Degrader Experimental Data

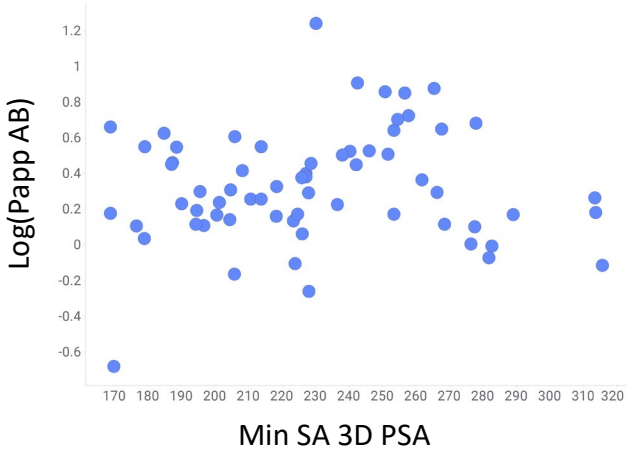


Impact of Dynamically Exposed Polarity on Permeability and Solubility of Chameleonic Drugs Beyond the Rule of 5

Sebastiano, et al., J Med Chem, 2018

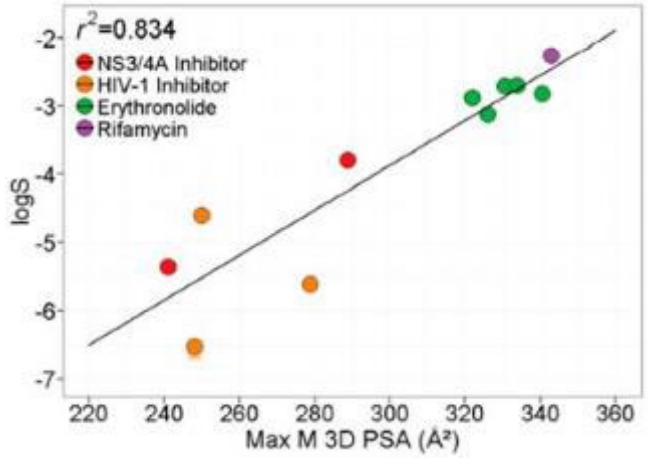


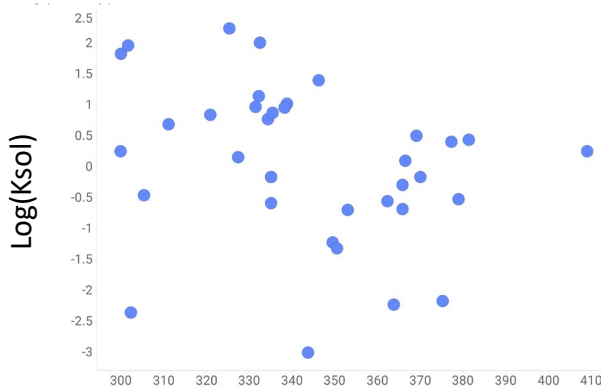
Publication Data



Nurix Data

Literature model for predicting permeability and solubility of BRO5 compounds as a function of 3D polar surface area struggle to track with Nurix experimental data



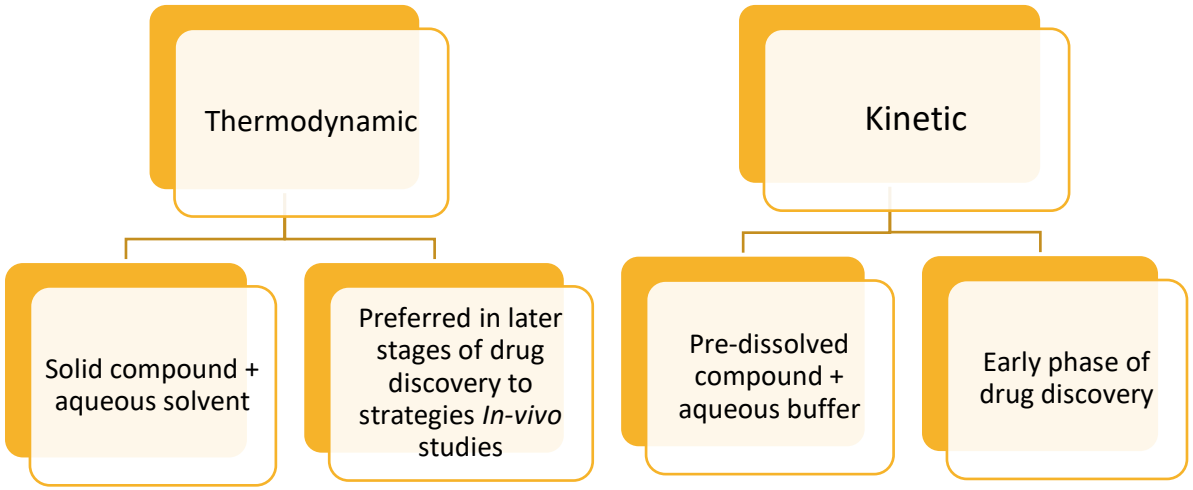


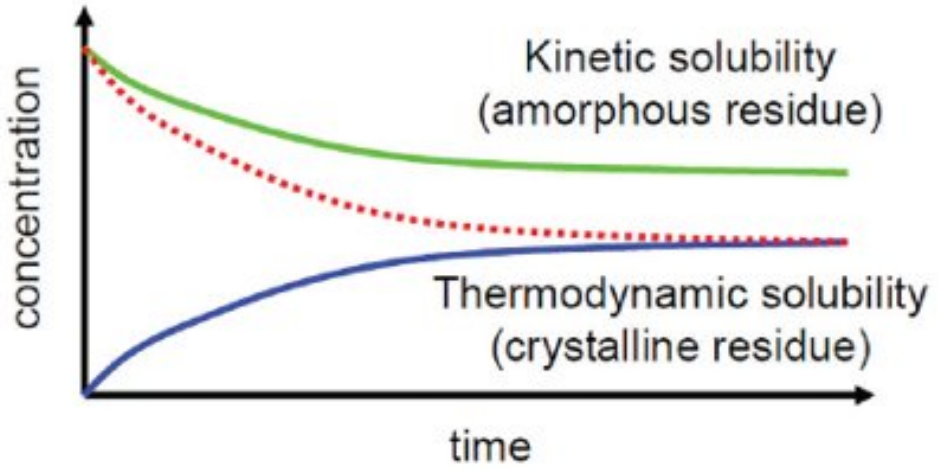
Motivation for Using Machine Learning Models for Property Prediction in the Degrader Design Space

- The rules commonly used to guide optimization campaigns are not applicable to targeted degraders, which occupy a distinct chemical space
- This frustrates rational design and optimization of degrader molecules, and thus necessitates empirical discovery.
- Machine Learning techniques can perceive relationships between structure and properties which are non-obvious, and could lead to useful property prediction.
- We evaluated our ability to develop predictive ML models for a variety of TPD properties, and will share results from Solubility and Permeability

Solubility Assays from a Drug Design Perspective

Kinetic solubility vs thermodynamic solubility

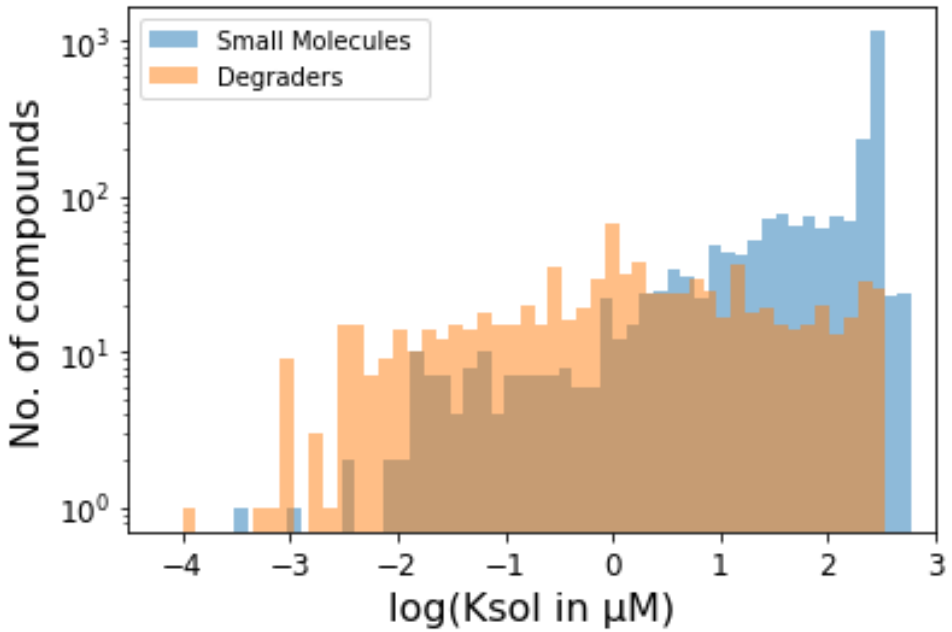




Img source : https://media.americanpharmaceuticalreview.com/m/28/article/36748-fig1.jpg

Kinetic solubility correlates well with discovery tasks and can help guide drug discovery teams in designing optimized compounds through structure modification

Nurix Solubility Dataset and Ways of Framing the ML Problem



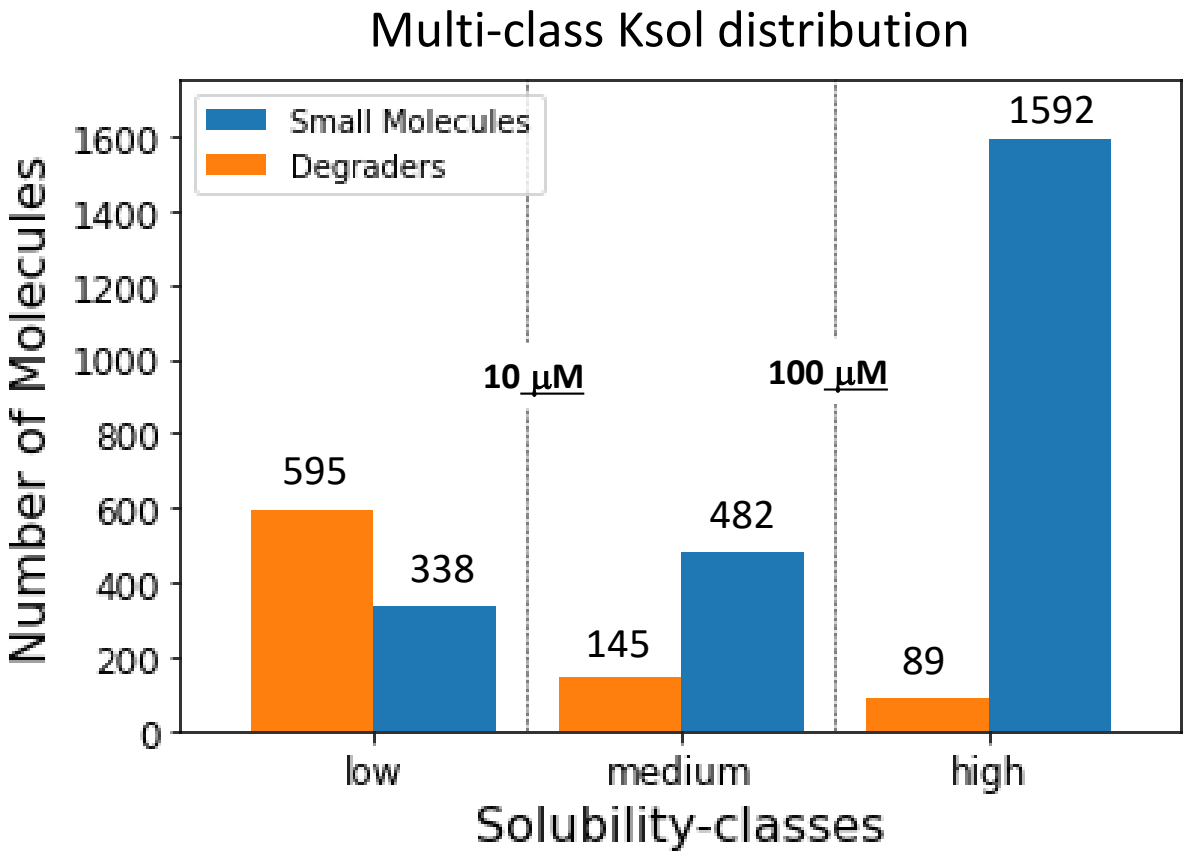
Binary Classification (Soluble/Insoluble)

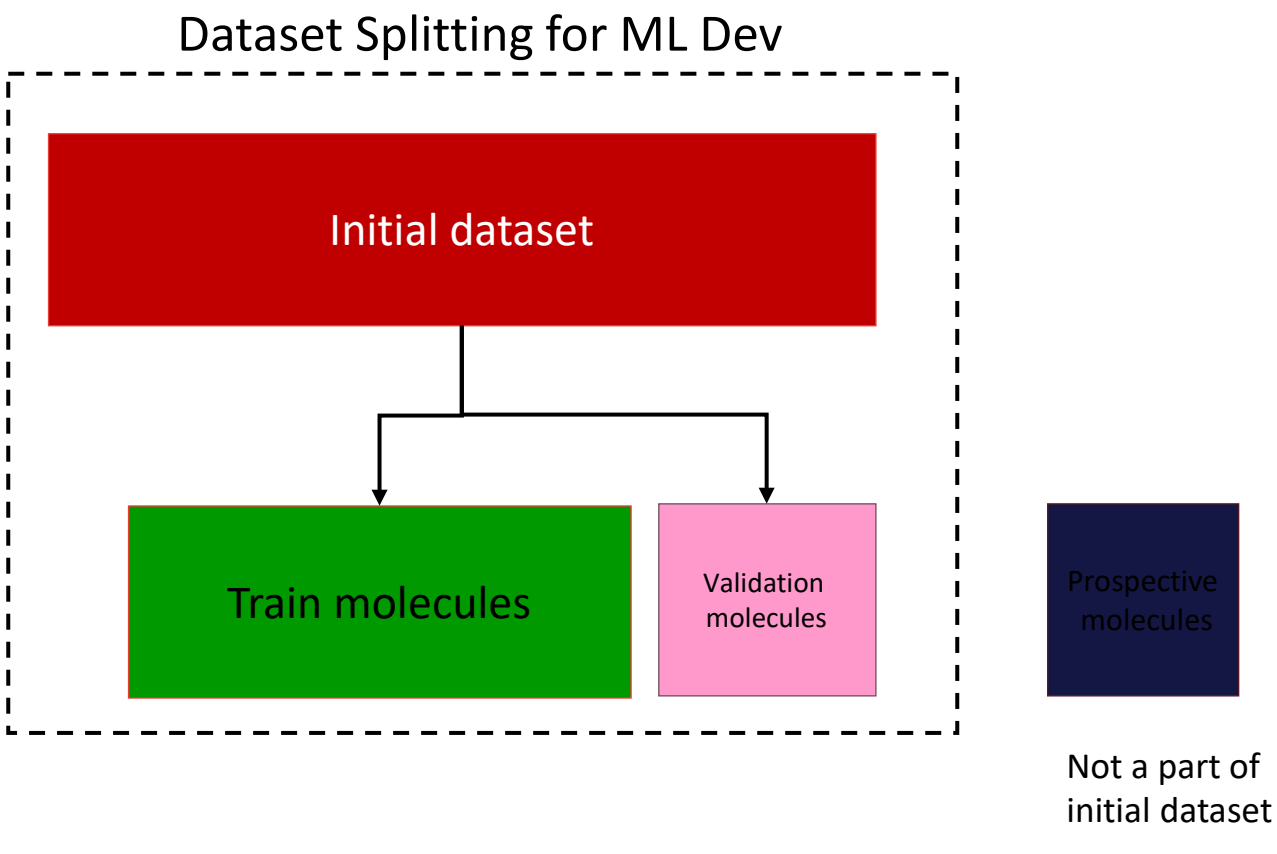
Multiclass Classification (High/Med/Low)

Regression (Predict numeric Ksol value)

Leveraging the Nurix solubility dataset, we evaluated both Regression and Classification models (and a host of implementations and parameter sets) to identify the best performing models.

Dataset Partitioning for Multiclass ML Model





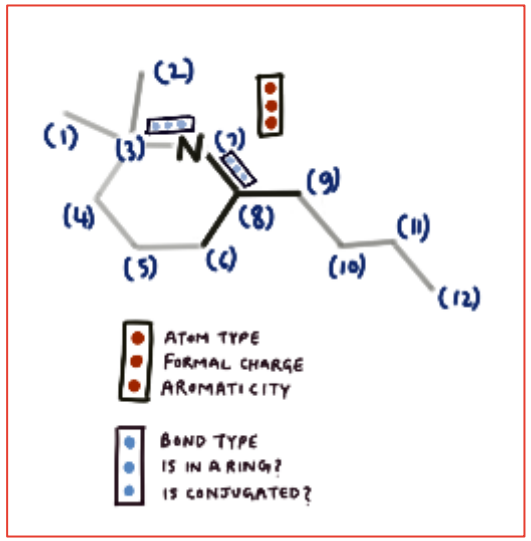
Combined small molecule and degrader dataset to increase generalizability of models

Proper dataset hygiene to ensure model is performant in prospective predictions

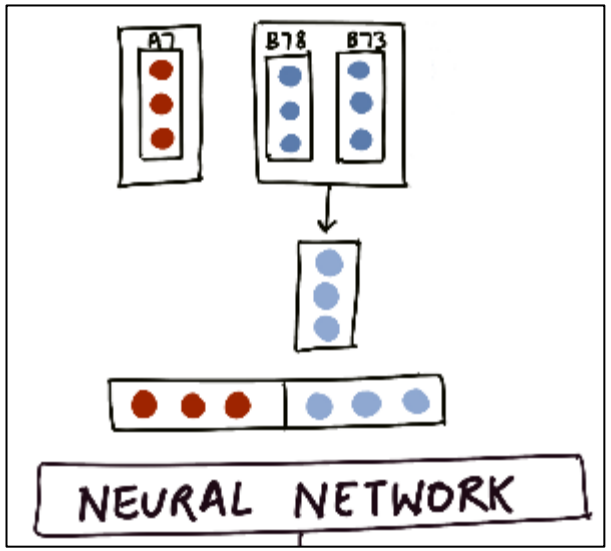
Model Architecture of Best Performing Ksol ML Model

Input: SMILES

C(#N)C1=Cc2n(c(-c3ncc(C(=O)....



Feed model with embeddings

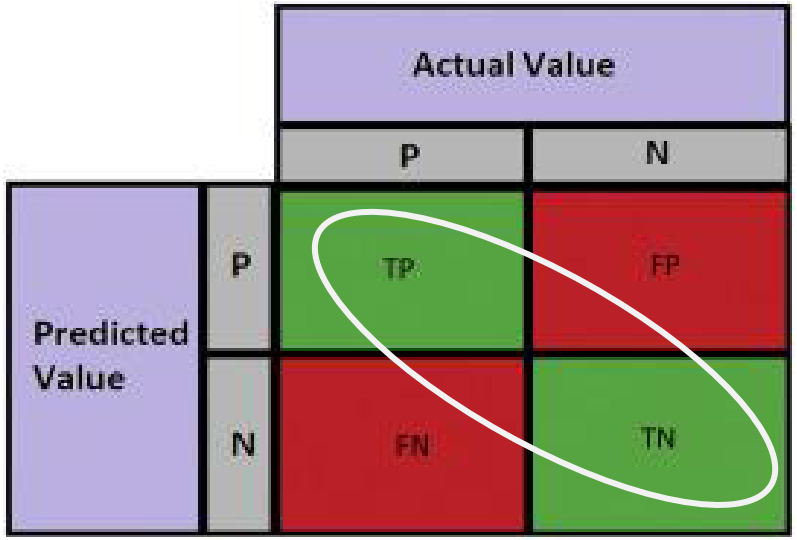


Predict class

- low
- medium
- high

- This is a variant of commonly used molecular property prediction ML model called Chemprop.
- Represents a molecule as a graph made up of atoms and bonds, with associated properties.

Understanding Model Performance: Confusion Matrix



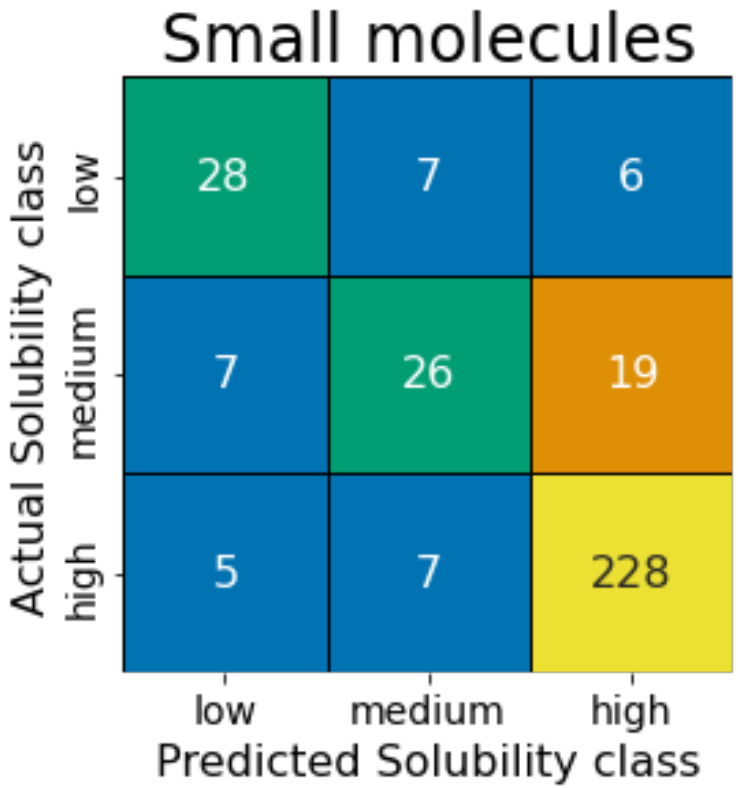
Confusion Matrix

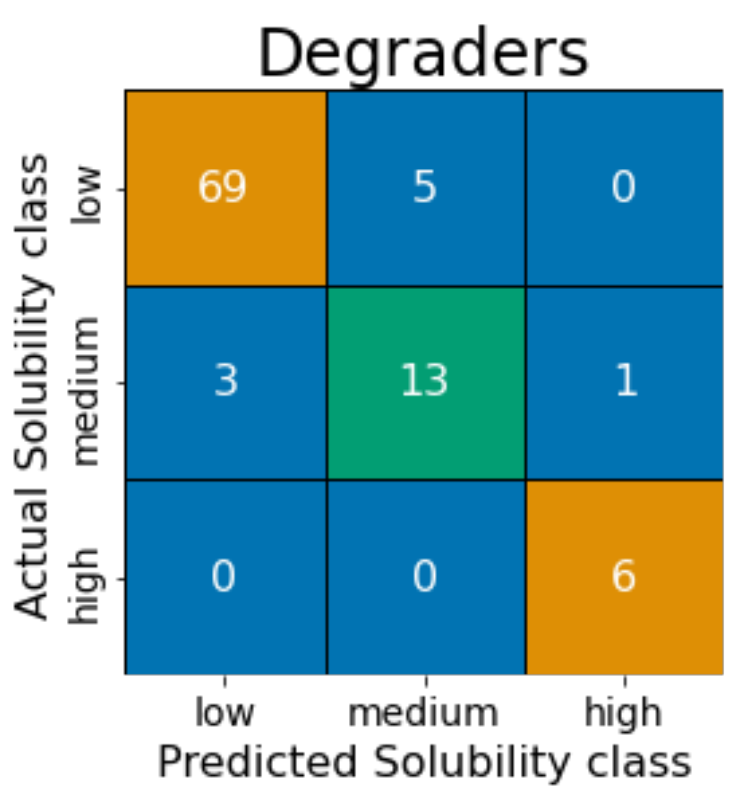
Incorrect Predictions lie in off-diagonal boxes (we want small numbers here)

Correct Predictions lie along the axis (we want big numbers here)

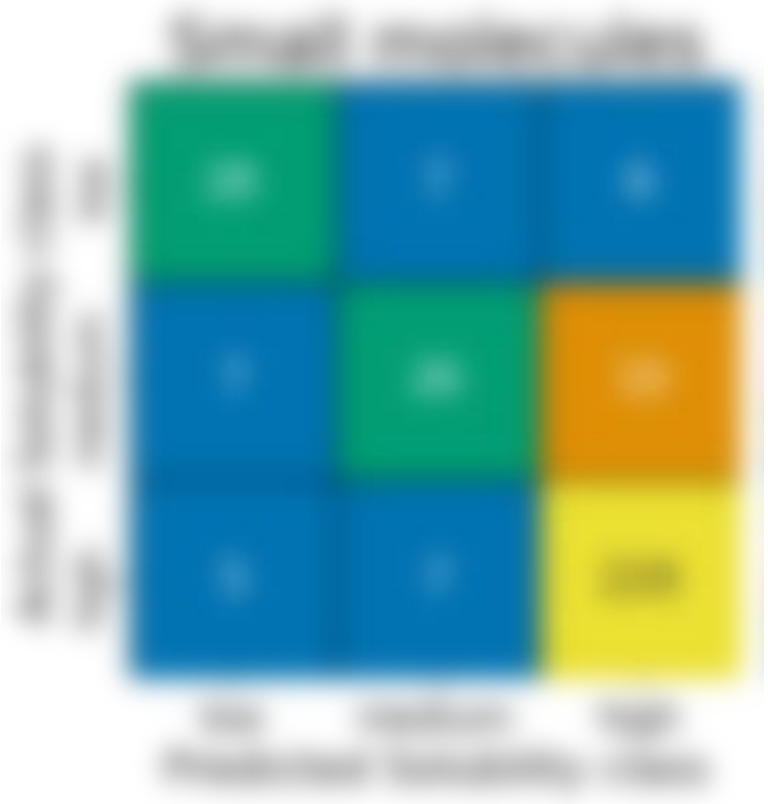
Das, C., Sahoo, A.K. and Pradhan, C., 2022. Multicriteria recommender system using different approaches. In *Cognitive Big Data Intelligence with a Metaheuristic Approach* (pp. 259-277). Academic Press.

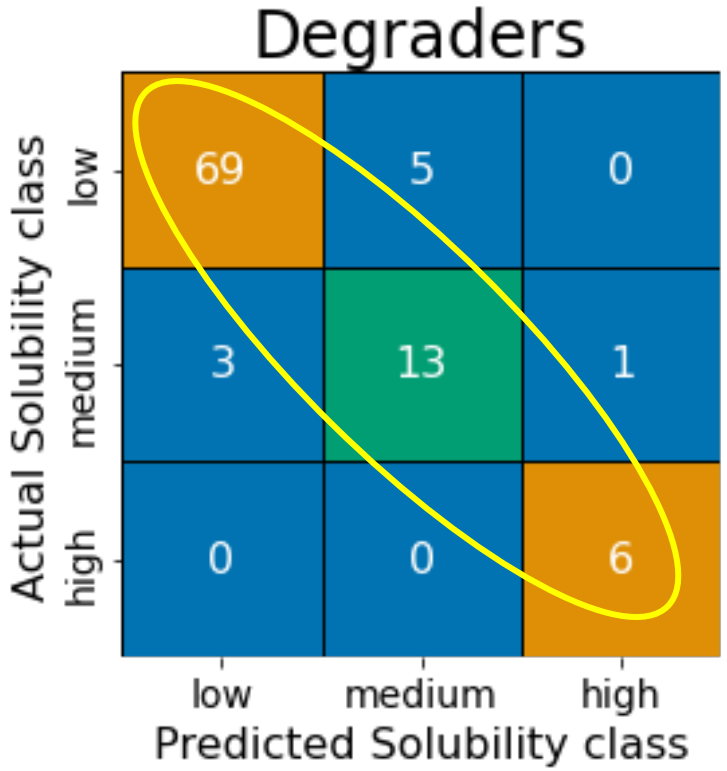
Model Performance on Prospective Dataset





Generalizability Across Diverse Chemical Space



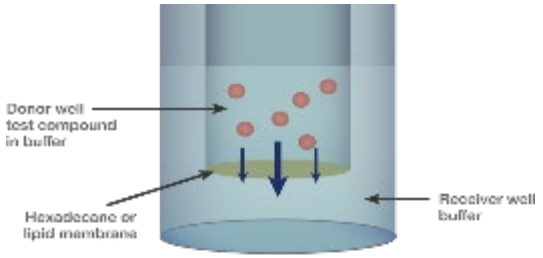


High accuracy in prospective prediction of the true label of TPD compounds

Model shows a high level of generalizability across compound classes, and high accuracy for TPD

nanoBret Target Engagement Assay to Assess Permeability

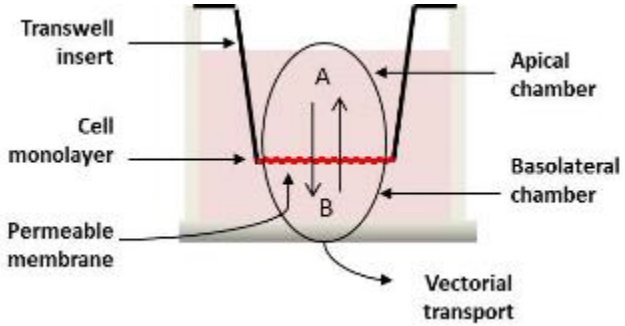
Artificial membrane assays (PAMPA)



Measures passive diffusion across an artificial membrane

Endpoints: logPe, Recovery

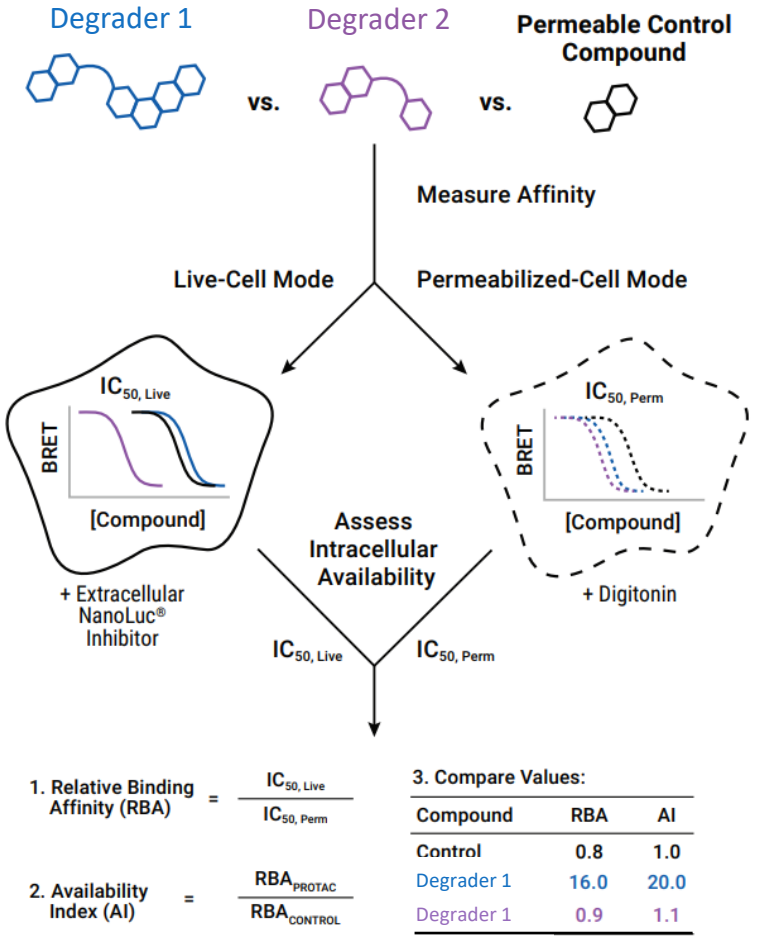
Cell monolayer assays (MDCK, Caco-2)



Directional transport owing to transporter proteins

Endpoints: Papp A->B, Papp B->A, Efflux Ratio, Recovery, Leakage, TEER

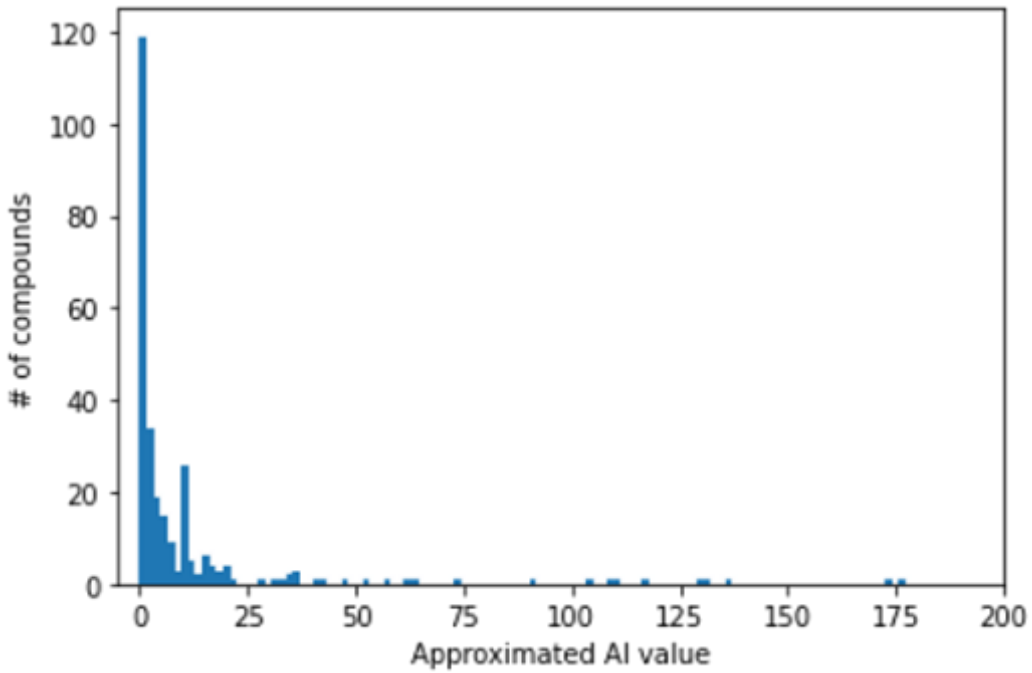
nanoBret Target Engagement assay



Evaluate shift in binding to intracellular target in live vs permeabilized cells

Endpoints: Relative Binding Affinity (RBA), Availability Index (AI)

Nurix Permeability Dataset and Framing the Problem



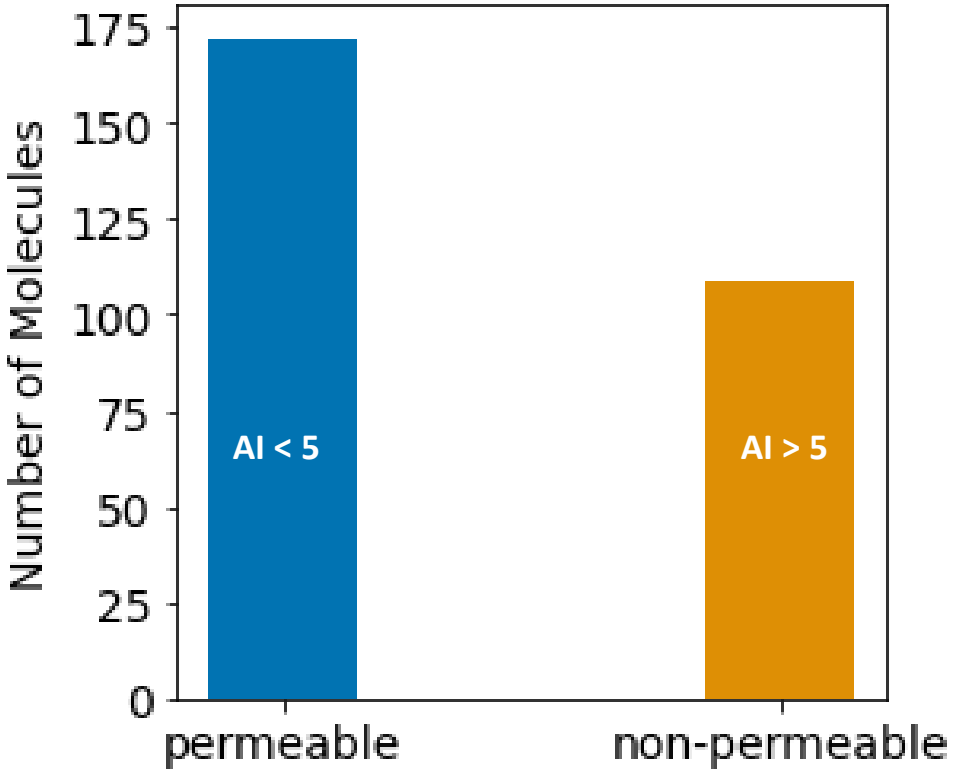
Binary Classification (Permeable/Impermeable)

Multiclass Classification (High/Med/Low)

Regression (Predict numeric TE AI value)

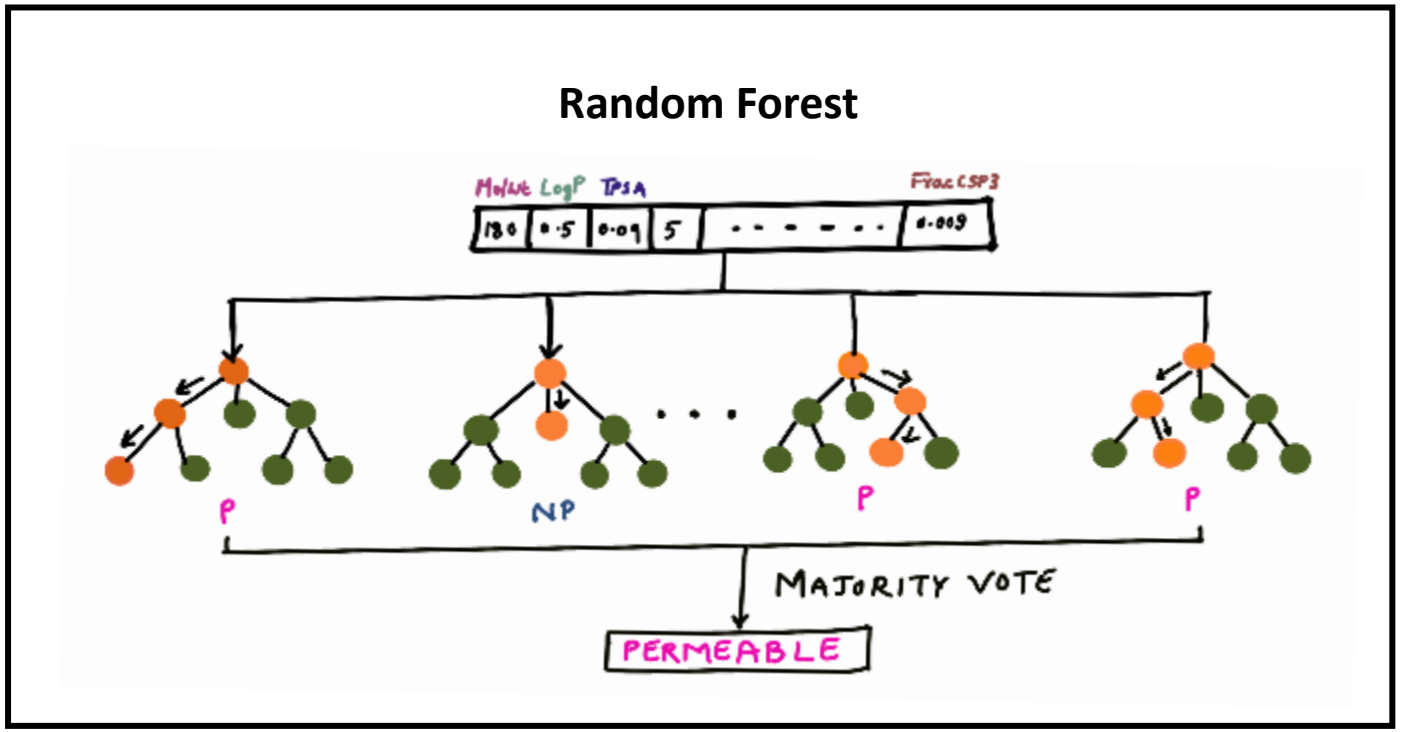
Due to the small size of the Nurix permeability dataset, we evaluated a binary-classification model

Permeability Data as a Binary-Class ML Dataset



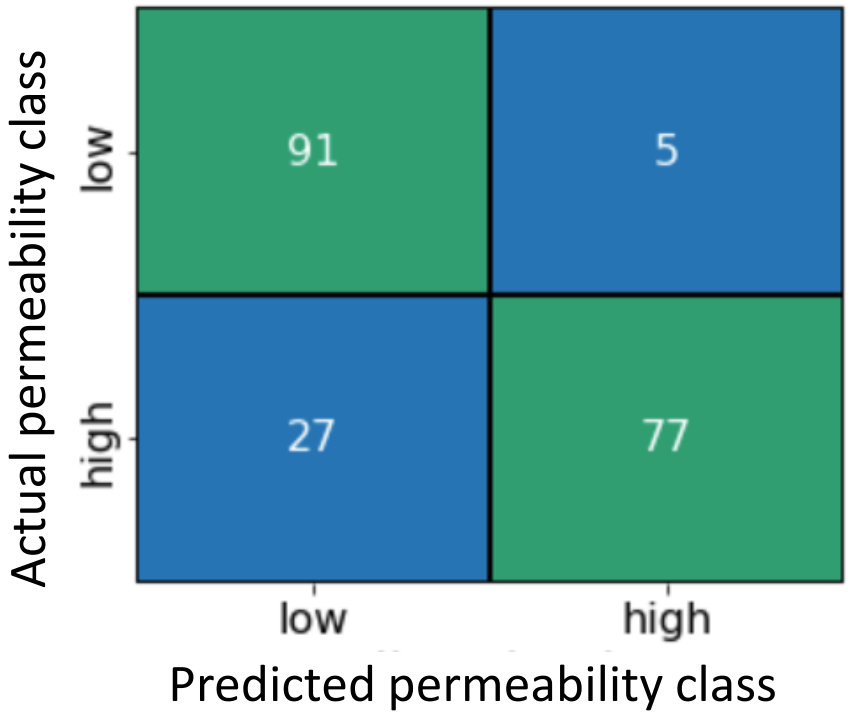
- Dataset distribution : 281 degraders representing multiple ligases
- Data labeling: If **Availability Index (AI) > 5, then non-permeable** else **permeable.**

Model Architecture of Best Performing Permeability ML Model



A simple decision tree that predicts permeability based on molecular properties

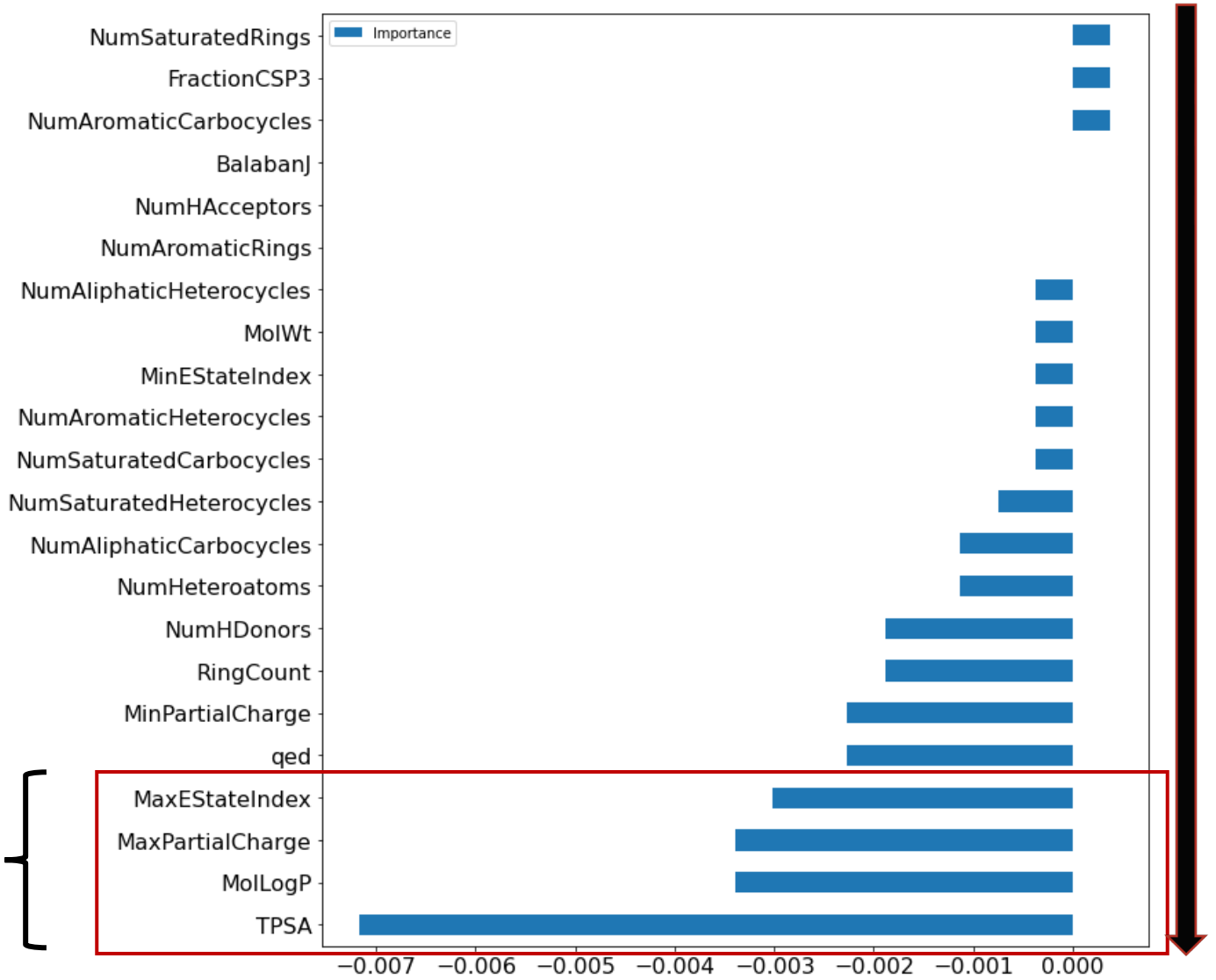
Performance of Permeability Model on Prospective Dataset



Model shows **high levels of generalizability** across different Ligases.

Investigating Model's Perspective for Important Features

- Feature importance tells which features contributed to model's learning process.
- Bars to the left are negatively correlated and to the right are positively correlated to permeability.
- **Feature importance can allow chemists to design better compounds.**



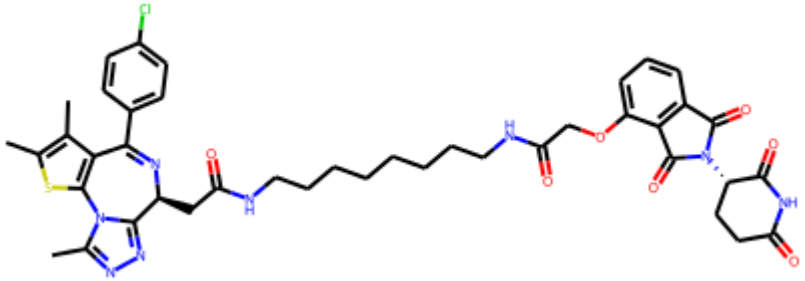
Positive correlation (higher Perm)

Negative correlation (lower Perm)

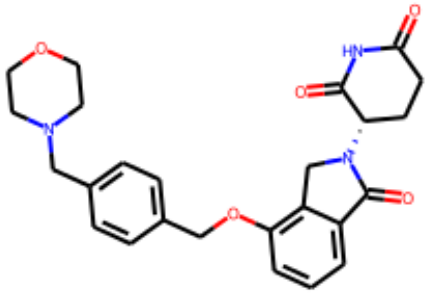
Important features

Permeability with Varying Molecular Properties

TPSA



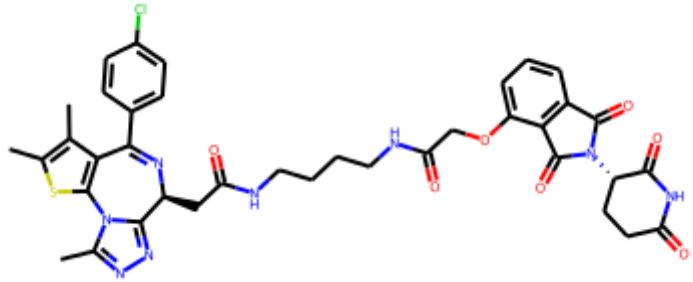
Compound : dBET6
TPSA – 194.05
Predicted as – Poor permeability



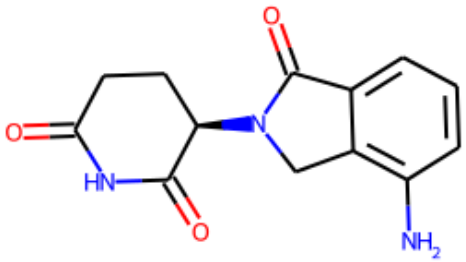
Compound : Iberdomide
TPSA – 88.18
Predicted as – High permeability

Compounds with higher **TPSA** and greater **CLogP** are identified as poorly permeable by the model

CLogP

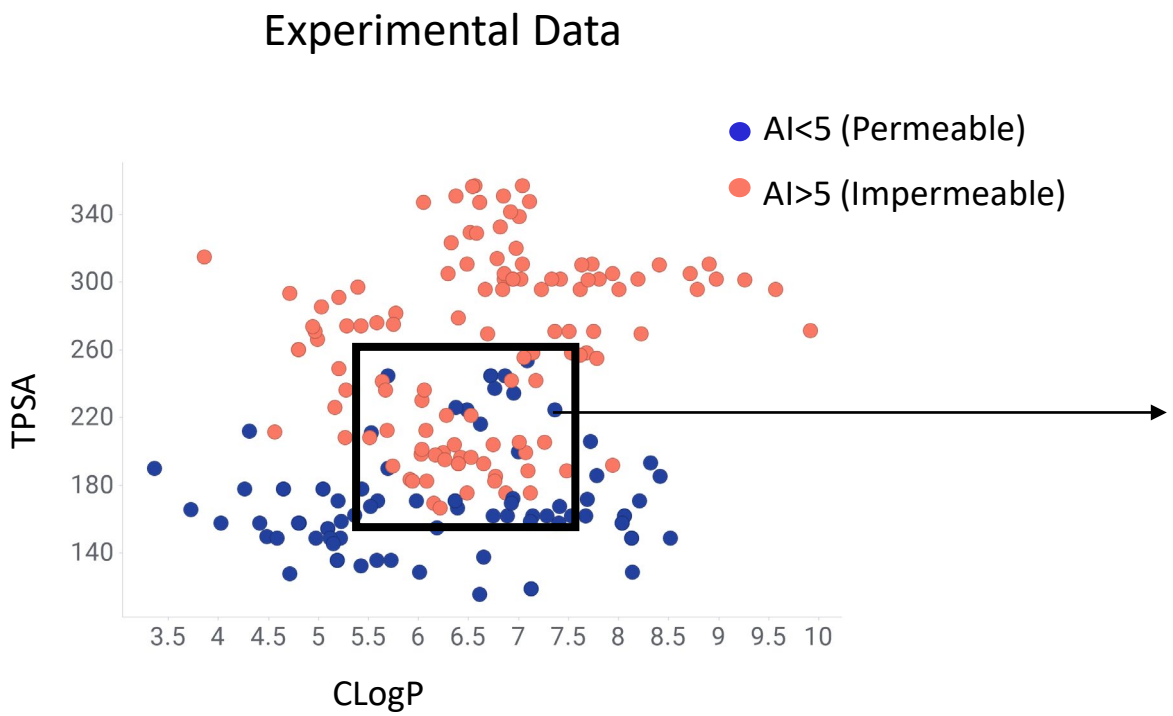


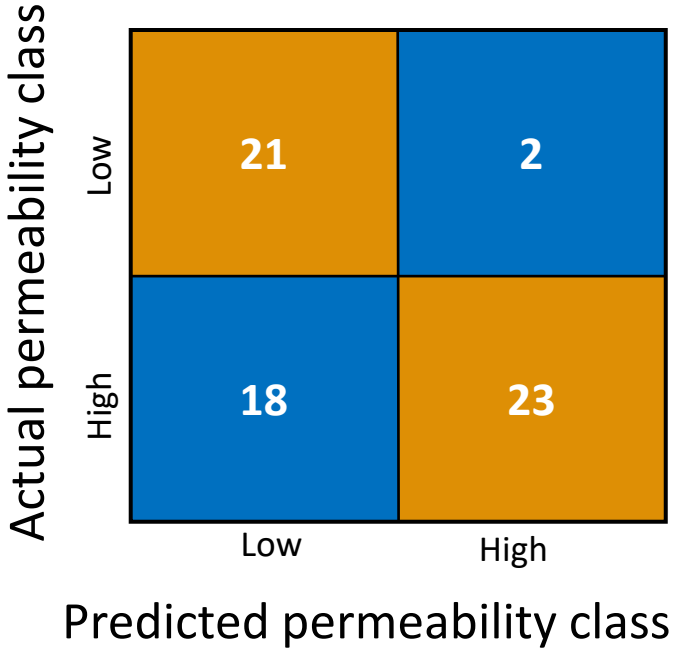
Compound : dBET1
ClogP: 3.68
Predicted as – Poor permeability



Compound : Lenalidomide
ClogP: 0.02
Predicted as – High permeability

Accuracy in Predicting Across Overlapping Property Space





69% accuracy in convoluted property space

Permeability ML model predicts correct class membership despite no obvious separation in property space defined by highest importance features.

Permeability Prediction: Recent and Similar Approaches

- In the following section, we will see some recent efforts published after our method was initialized.
- Similar to our method, most of these efforts make use of molecular properties as features to train the model.
- Surprisingly, the top features identified coincide with our top features from the model.

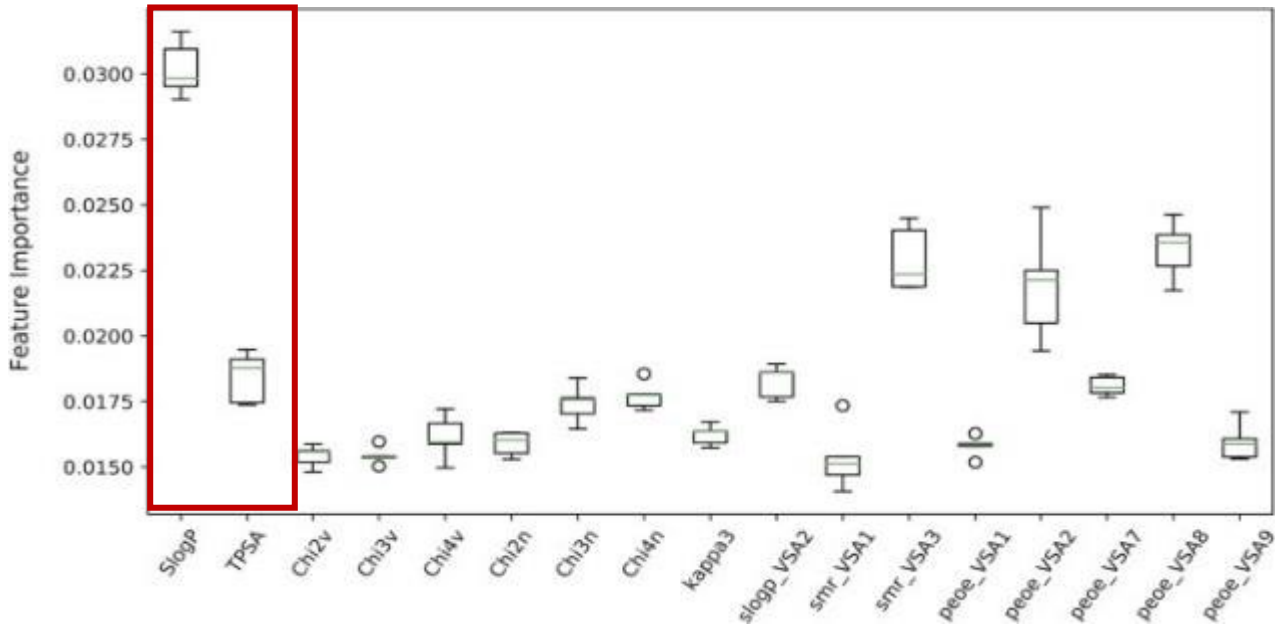
Emerging Consensus Regarding Highest Importance Features

Reliable prediction of Caco-2 permeability by supervised recursive machine learning approaches : Falcón-Cano, G. et.al

Supplementary Table S1. Final list of most important variables sorted by number of occurrences

Descriptors		Number of occurrences
SlogP	Log of the octanol/water partition coefficient (including implicit hydrogens). Captures lipophilicity	8222
SMR	Molecular Refractivity (including implicit hydrogens). Captures polarizability and protonation state	1706
TPSA	Topological Polar Surface Area. Captures polarizability	1269
Hallkier alpha	Hall Kier Alpha value. Captures polarizability	549
Kappa 3	Molecular shape index. Captures flexibility	352

Using *in vitro* ADME data for lead compound selection: An emphasis on PAMPA pH 5 permeability and oral bioavailability Jordan Williams et al.



SlogP and **TPSA** show up as important features from multiple groups' efforts

Important Take Homes

- ML methods can learn interesting patterns from the chemical data where generalized literature-based methods might lack accuracy.
- Most of the published methods for property prediction are built in small molecule chemical space, and thus might not span well for degraders which share a different chemical space.
- We demonstrate that a property prediction ML model built using both small molecules and degraders was able to learn the respective chemical space and thus resulted in a well generalizing model to predict solubility for both type of compounds.
- We show that using basic molecular properties can be used to build a ML model that predicts the permeability of degraders.
- Feature importance of a ML model can give additional insights about which properties were most important in decision making, hence can help chemists to design much better chemical compounds.

Thank you