



**CERTAINTY
DISCOVERY**

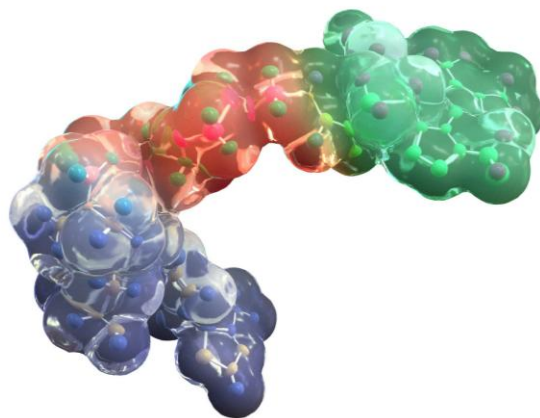
CERTARA[®] | Chemaxon

Closing the DMTA Loop: Interplay Between Experiments and Predictions Drives PROTACs Bioavailability

Filip Miljković

AstraZeneca

Associate Principal AI Scientist



Closing the Design-Make-Test-Analyze Loop: Interplay Between Experiments and Predictions Drives PROTACs Bioavailability

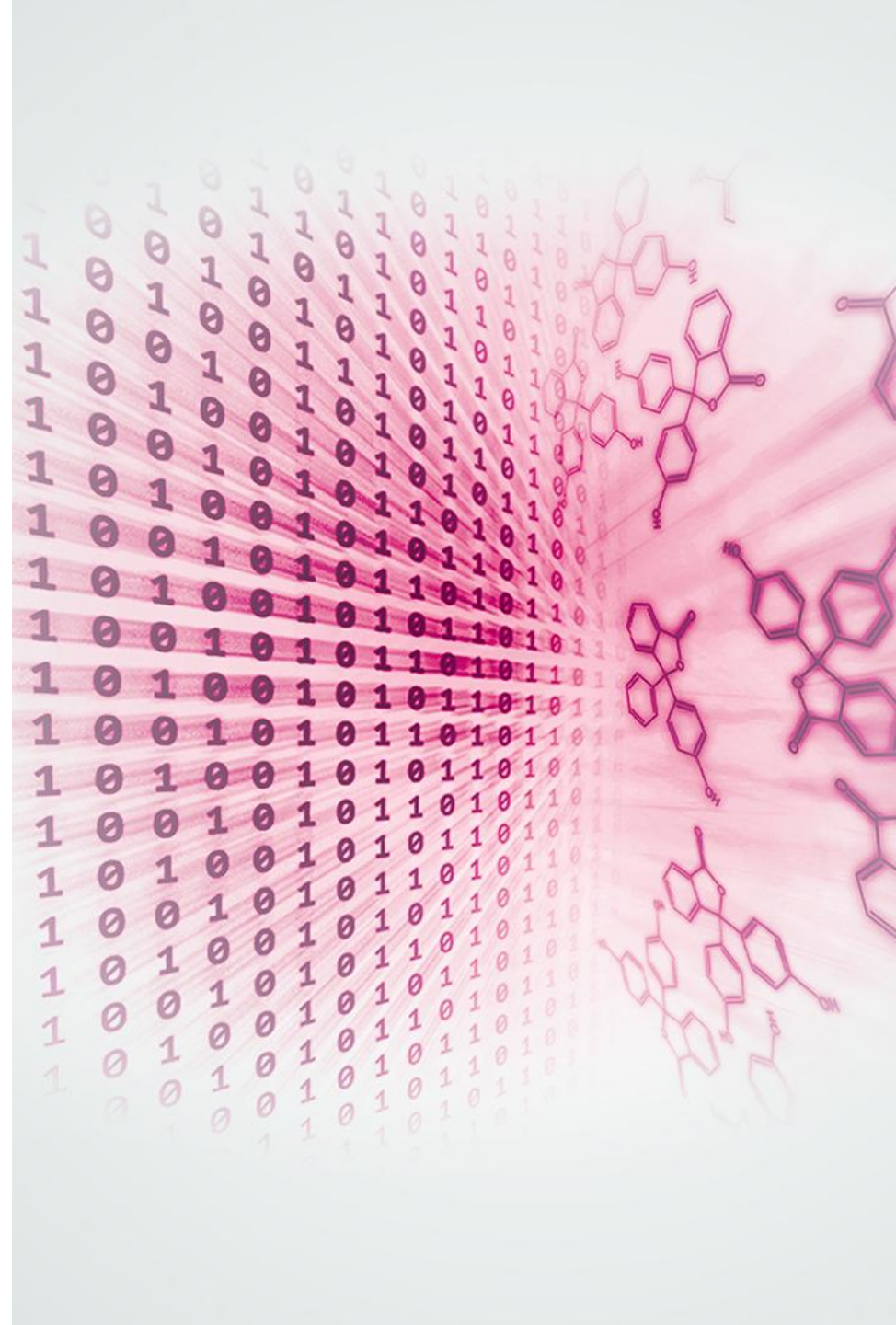
Filip Miljković

Medicinal Chemistry, Research and Early Development, Cardiovascular, Renal and Metabolism (CVRM),
BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden

Certainty Discovery 2025

Scientific Conference and User Group Meeting | Day 1
Frankfurt, Germany
November 4, 2025

Link to paper

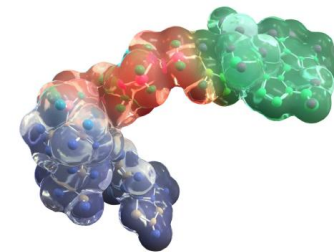


Agenda

- 1 Drug permeability and bioavailability
- 2 ChromLogD and EPSA data
- 3 ChromLogD and EPSA models
- 4 Project examples
- 5 Final remarks



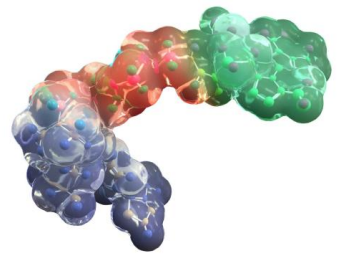
Lipinski's Rule of Five and Beyond



- Impact of physicochemical properties on DMPK parameters **well-established for traditional small molecule space**
- **Lipinski's Rule of Five** has notably impacted the discovery of orally bioavailable drugs, shaping the development of preclinical profiling assays used in DMTA processes
- **bRo5** molecules necessitate assay modifications and adoption of new profiling assays – essential for **refining bioavailability guidelines** and **building efficient prediction models**
- bRo5 molecules can engage in intramolecular interactions, adopting their shape and exposed chemistry to the surrounding, therefore **behaving differently when passing through a phospholipid membrane**



Lipophilicity and Cell Permeability for bRo5



- Lipophilicity and cell permeability are two critical properties that inform on the drug bioavailability
- While traditional small molecules are well catered for, bRo5 molecules pose some challenges:
 - Stability issues
 - Lower solubility
 - Lipophilicity outside the experimentally accessible range (shake-flask method)
- Methods that measure physicochemical properties based on chromatographic retention times are attractive for bRo5:
 - ChromLogD: lipophilicity
 - EPSA: passive cellular uptake (dependent not only on lipophilicity, but also compound size and polarity)



Shake-flask LogD

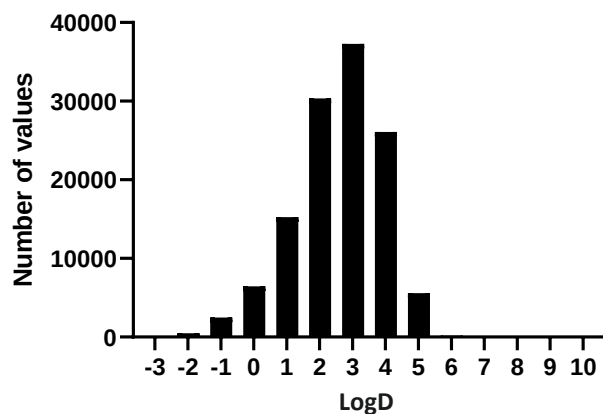


Octanol
Buffer



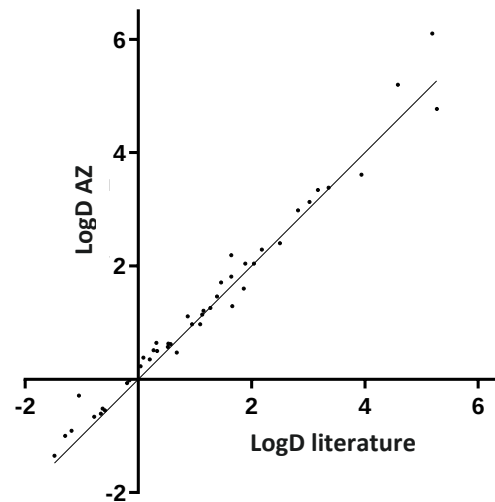
Waste
Buffer
Waste

1. Octanol + Buffer pH 7.4 + compound
2. Shake 3h
3. Phase separation, sampling, analyze

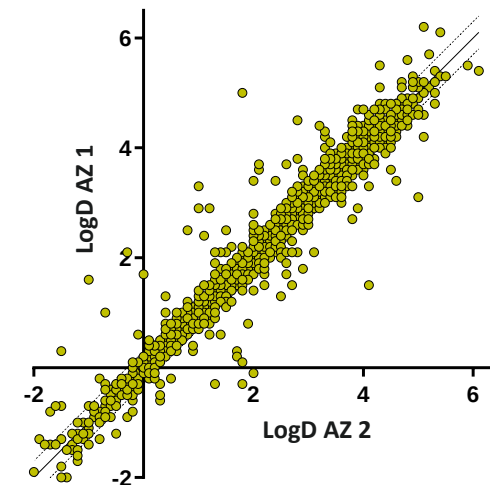


Histogram of the LogD assay results (2013 - Sep 2025) summarizing data for 130K compounds

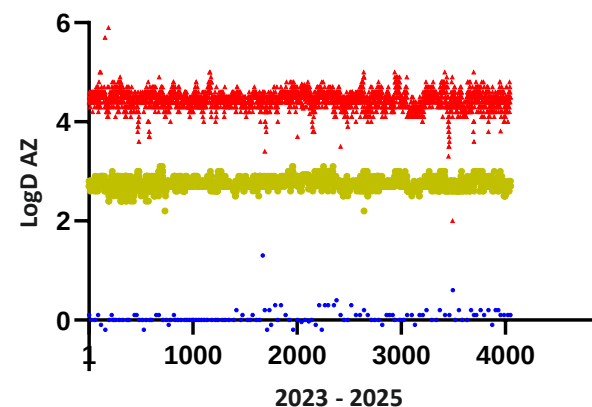
Performance of the automated shake flask LogD assay



Comparison to literature values
(44 compounds⁵, $r^2 = 0.99$)



Replicates from independent runs
(4158 compounds, $r^2 = 0.96$)



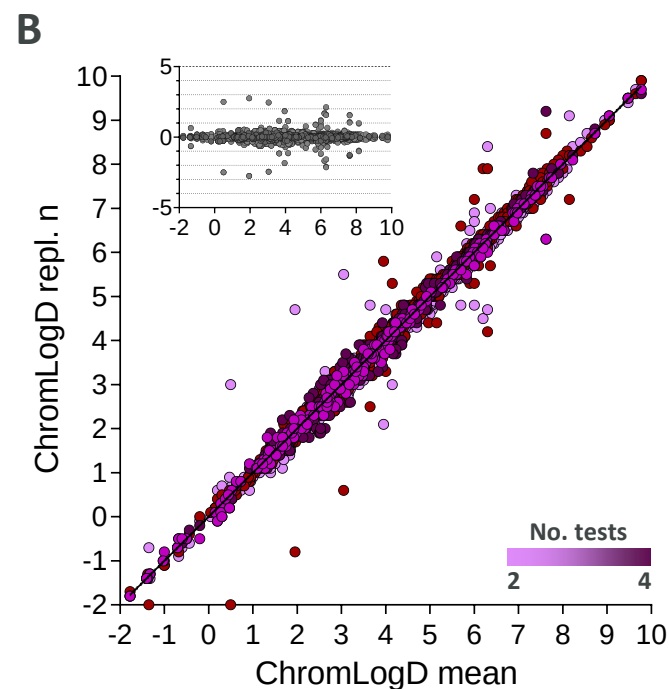
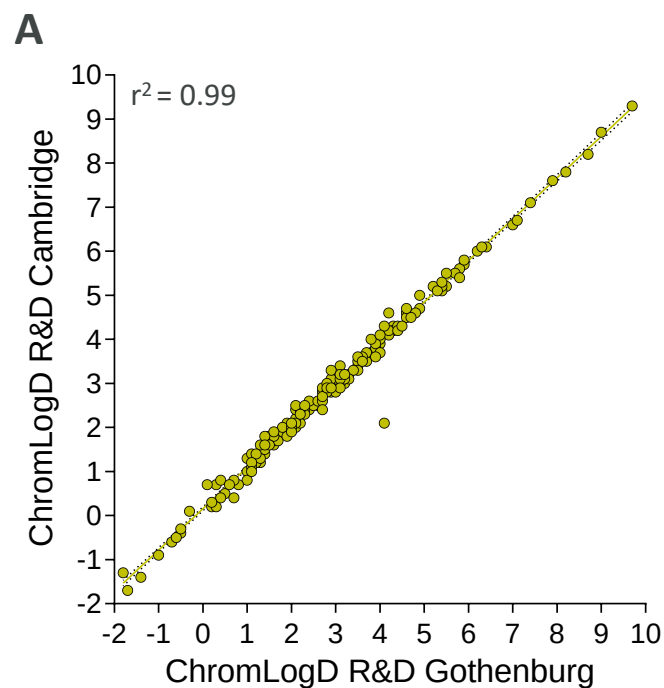
Quality controls (Jan 2023 - Sep 2025)

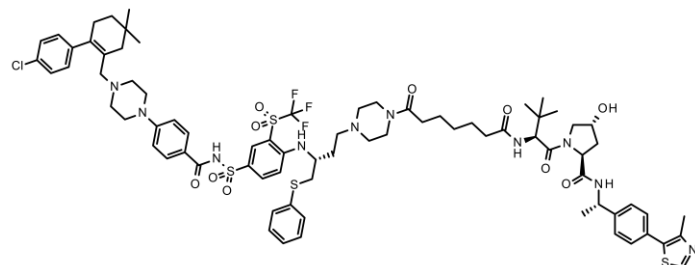
- ◆ Caffeine
- ◆ Cyclobenzaprine
- ◆ Nicardipine



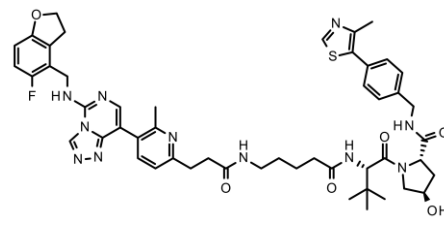
ChromLogD

- ChromLogD (reverse-phase HPLC assay) was optimized and automated to allow high-capacity measurements
- Several measures were taken to validate performance of the assay protocol:
 - Generation of equivalent data for 234 compounds at two R&D sites (**A**)
 - Reproducibility between test occasions for 1560 compounds (24 compounds with $\Delta\text{ChromLogD} > 1$) (**B**)

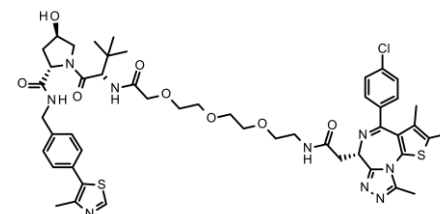




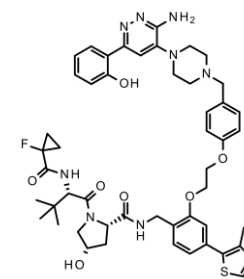
DT-2216 (1)



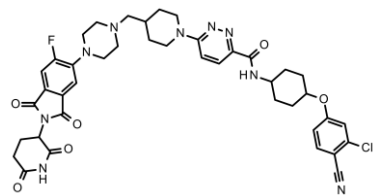
AZ-A (2)



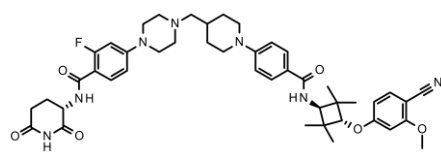
MZ-1 (3)



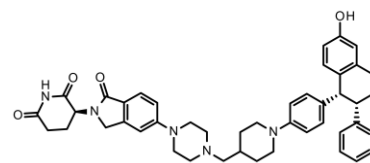
ACBI-1 (4)



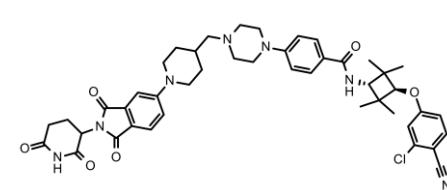
ARV-110 (5)



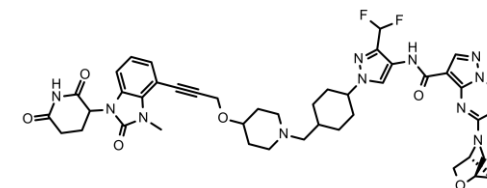
ARV-766 (6)



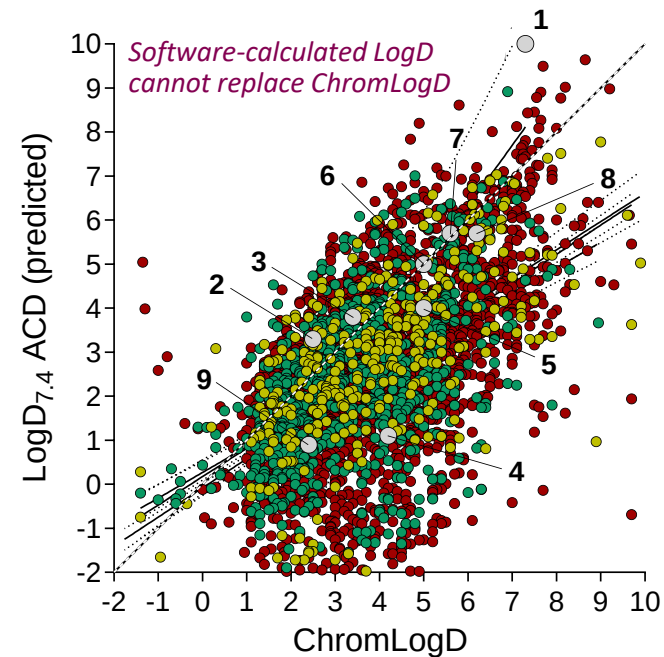
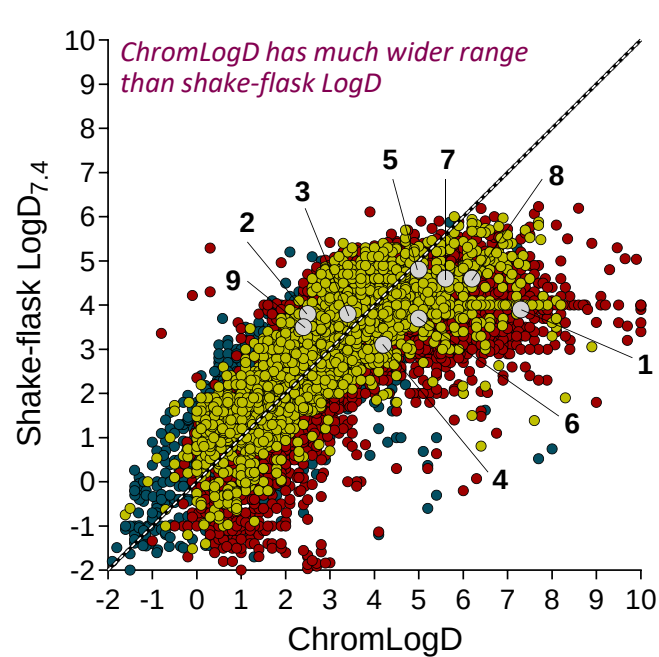
ARV-471 (7)



ARD-2128 (8)



KT-474 (9)

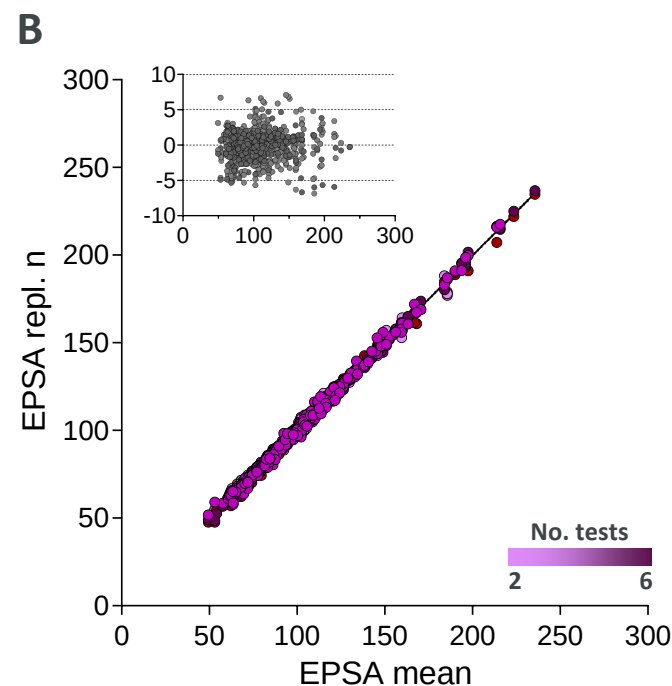
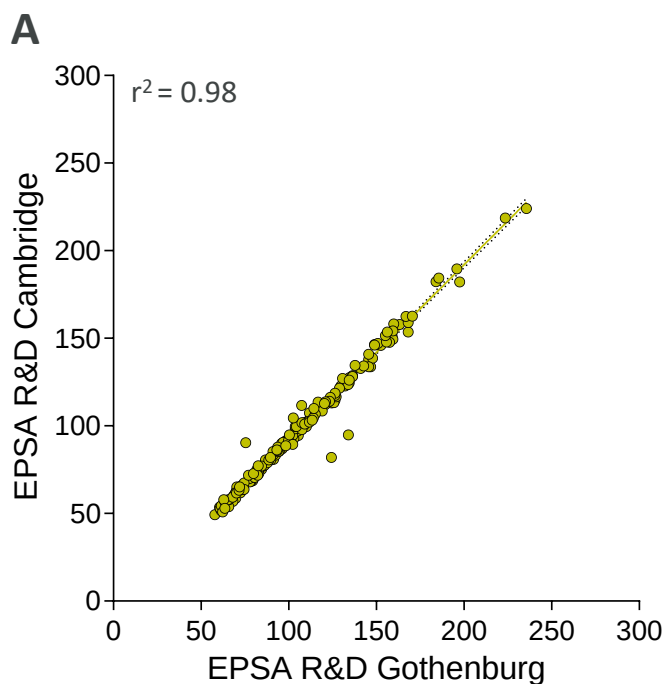


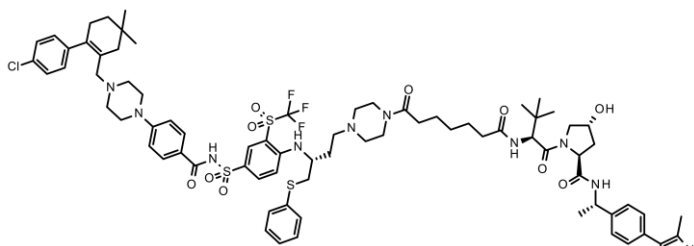
- Compliant with Lipinski's Ro5
- Breaking one rule
- Breaking two or more rules



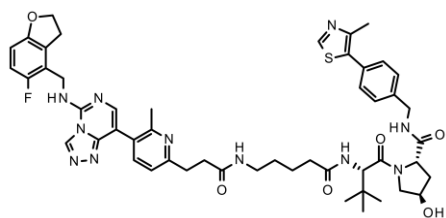
EPSA

- EPSA (SFC assay) allows for intramolecular hydrogen bonding and masking of polar surface area
- Like for ChromLogD, several actions were taken to validate assay performance:
 - Cross-site validation between two R&D sites for a set of 245 compounds (**A**)
 - Reproducibility between test occasions for 304 compounds (the majority within ± 5) (**B**)

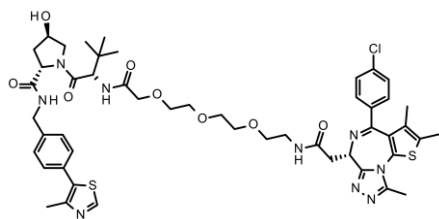




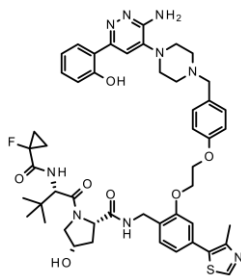
DT-2216 (1)



AZ-A (2)

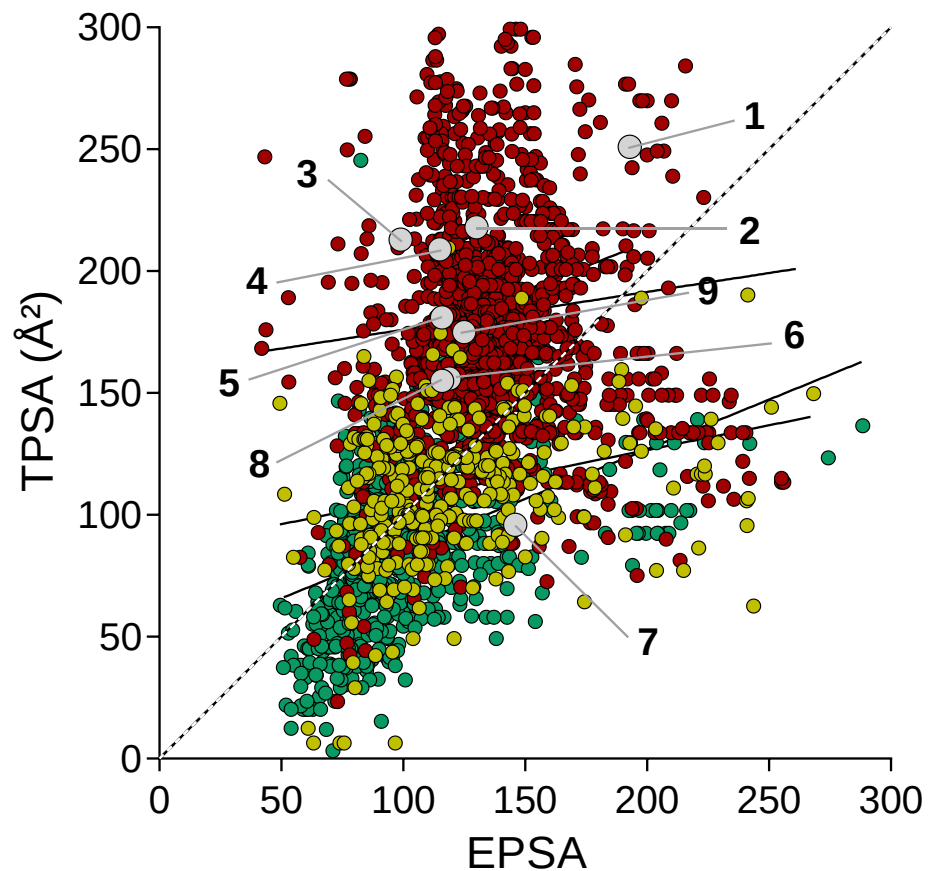


MZ-1 (3)

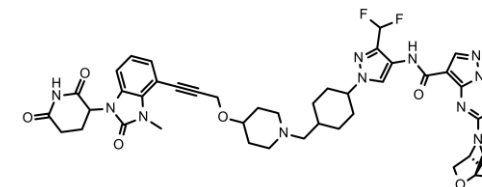


ACBI-1 (4)

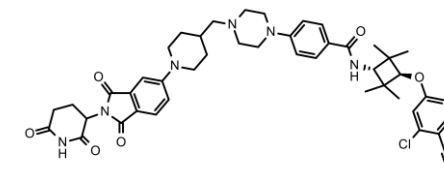
No correlation between EPSA and TPSA, particularly for bRo5 molecules



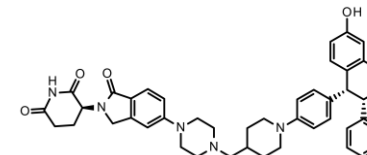
- Compliant with Lipinski's Ro5
- Breaking one rule
- Breaking two or more rules



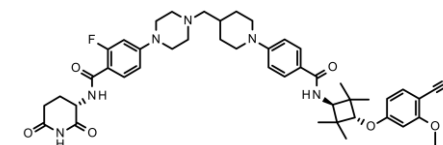
KT-474 (9)



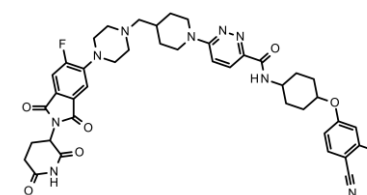
ARD-2128 (8)



ARV-471 (7)



ARV-766 (6)



ARV-110 (5)



Machine Learning Models

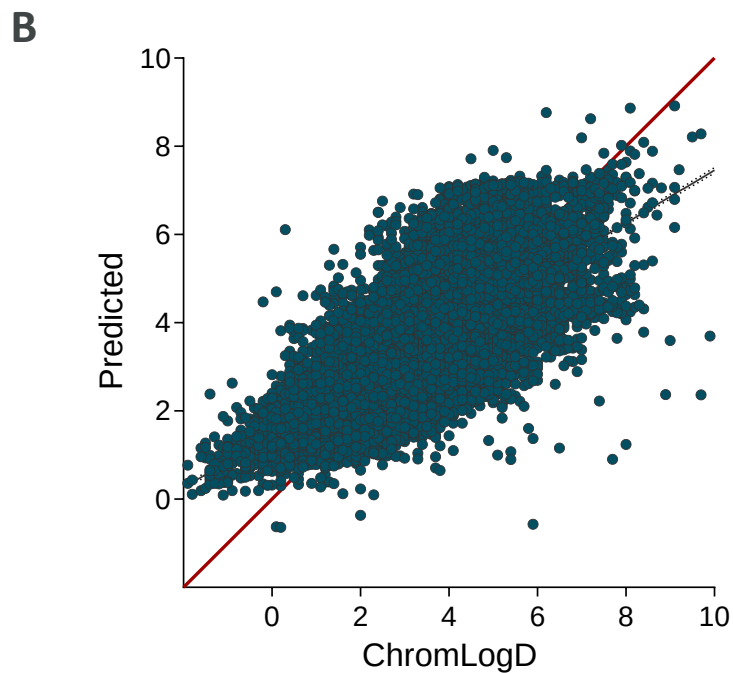
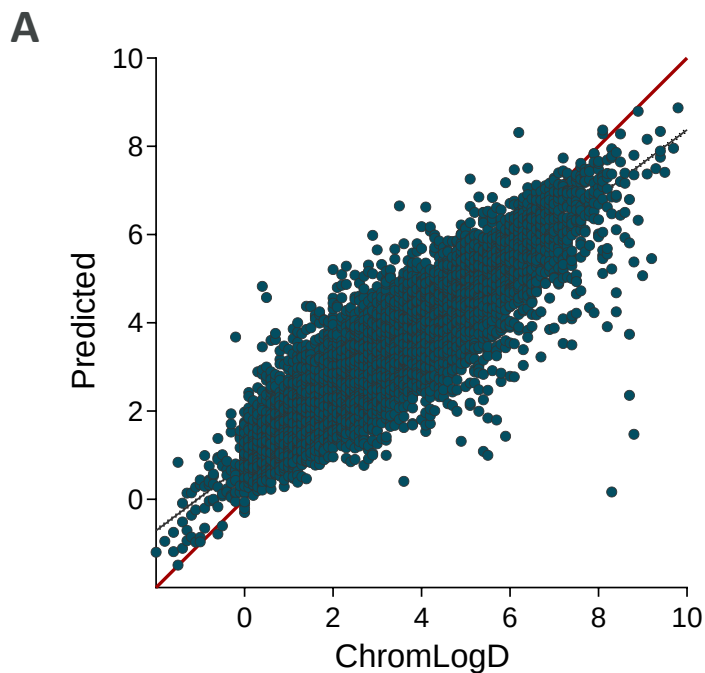
- Both ChromLogD and EPSA models were built using RF algorithm and a combination of ECFP4 and calculated 1D/2D molecular descriptors (e.g., number of hydrogen bond acceptors and donors, ring size, polarity, etc.)
- Two data splitting strategies were applied for model building and evaluation (training-test set ratio 80/20):
 - Stratified: Compound samples sorted in the ascending order of their averaged experimental values and every n -th sample selected into the test set (five trials)
 - Prospective: Compound samples ordered by the recorded experimental date, with the most recent experimental data used for model evaluation (realistic case scenario)
- The models were evaluated using R^2 and RMSE performance scores



ChromLogD Models

- In total, 77,792 compounds (from 801 medicinal chemistry projects and designated compound libraries) were available for ChromLogD model building and evaluation

	#Compounds	#Projects Test Set (Shared)	R ²	RMSE
ChromLogD stratified (A)	15,558	563 ± 8 (530 ± 9)	0.81 ± 0.01	0.72 ± 0.01
ChromLogD prospective (B)		402 (236)	0.51	1.20



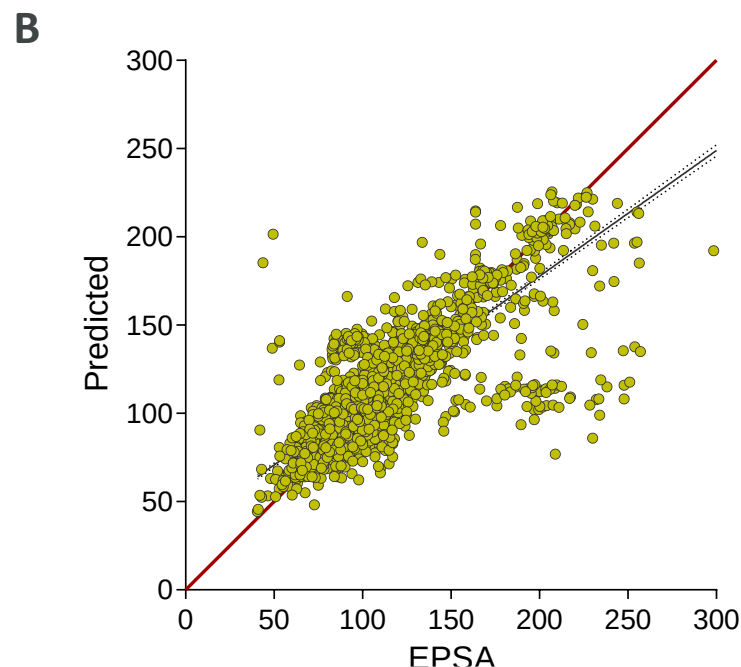
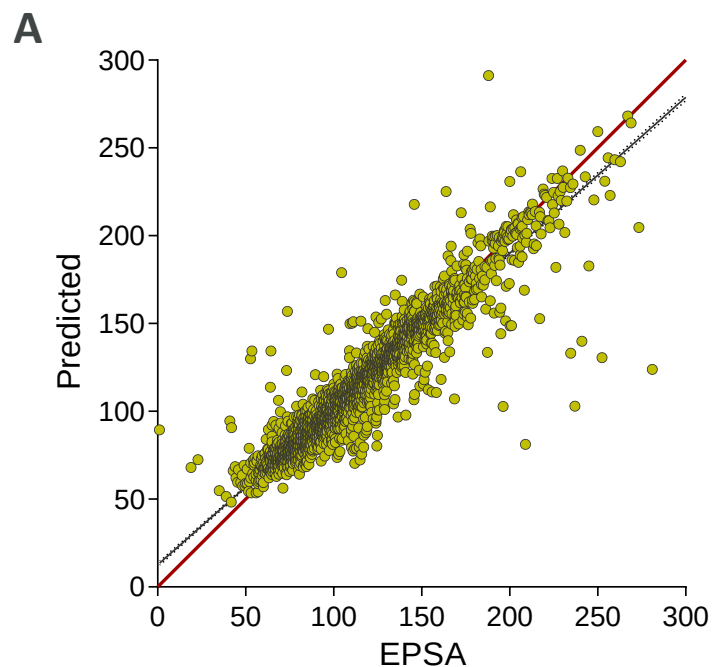
- Prospective test set corresponded to data generated in the past *four years* (*models built each month*)
- In-house model performance closer to the stratified models



EPSA Models

- In total, 18,064 compounds (from 384 projects), spanning a wide range of chemical modalities, were evaluated in the EPSA assay

	#Compounds	#Projects Test Set (Shared)	R ²	RMSE
EPSA stratified (A)	3612	210 ± 8 (183 ± 8)	0.89 ± 0.01	11.61 ± 0.61
EPSA prospective (B)		172 (106)	0.65	19.73

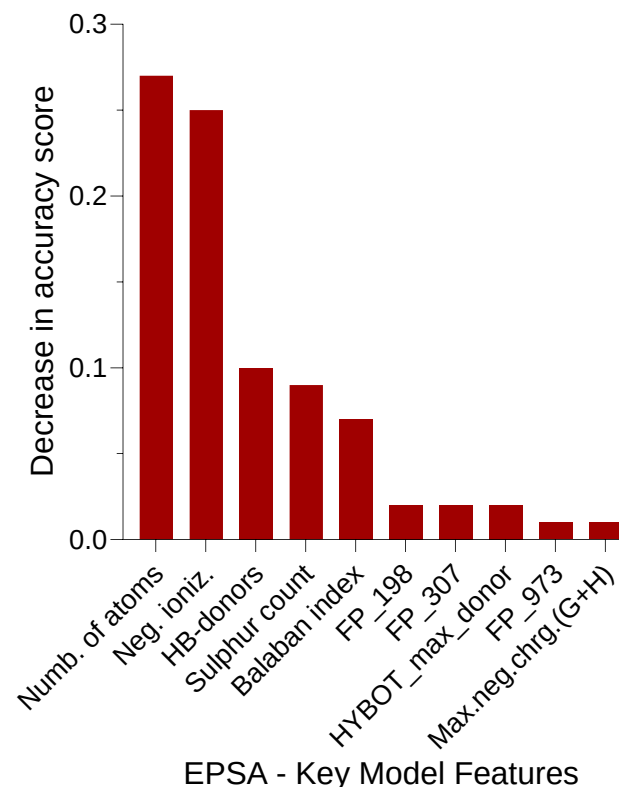
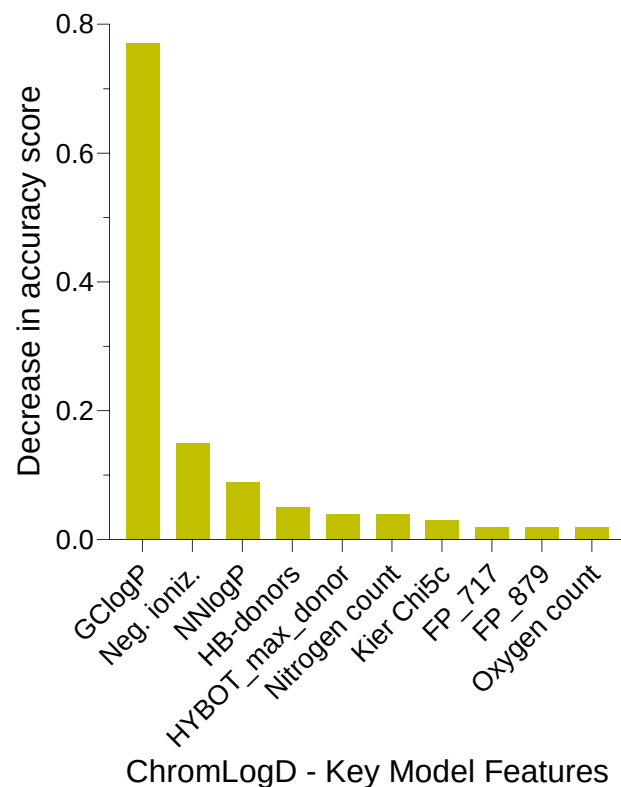


- Prospective test set corresponded to data generated in the past **seven months** (*models built each month*)
- In-house model performance closer to the stratified models



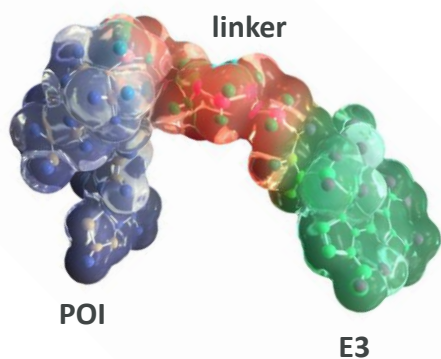
Model Feature Importance

- Permutation feature importance technique was applied to measure the contribution of each model feature to a fitted model's statistical performance
- A larger decrease in the accuracy score (R^2) indicates greater importance of that feature

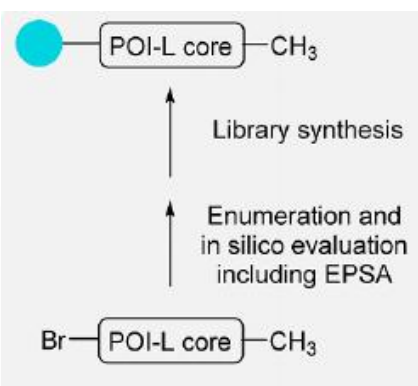


Project Example | EPSA Model

- A design goal of predicted EPSA < 150 for PROTACs was used to prioritize compounds, ultimately with the aim of facilitating oral bioavailability
- EPSA has proven to be a semi-additive parameter, i.e., within a particular chemical series, the predicted EPSA of the disconnected fragment is often proportional to the predicted EPSA of the full PROTAC
- EPSA model is predictive and transferable from a small-molecule warhead to its fully grown PROTAC

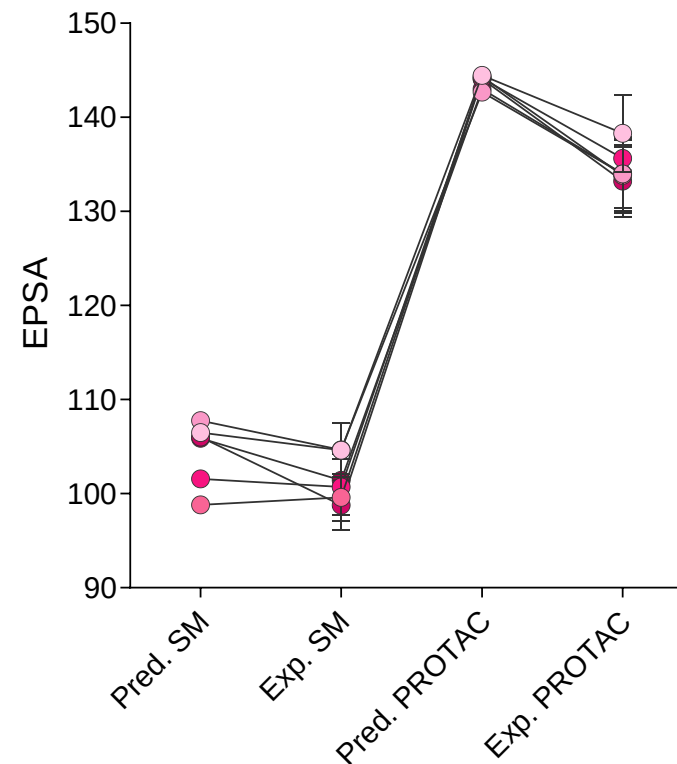
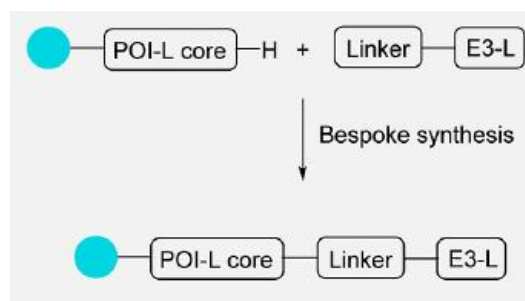


Step 1: Library design, evaluation, and synthesis



Approximately 400 building blocks were examined *in silico*

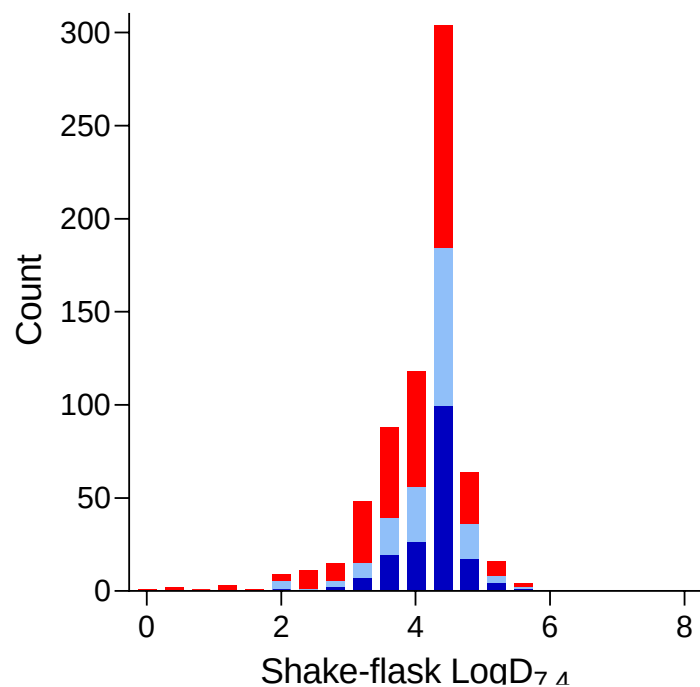
Step 2: PROTAC synthesis with best predicted substituents



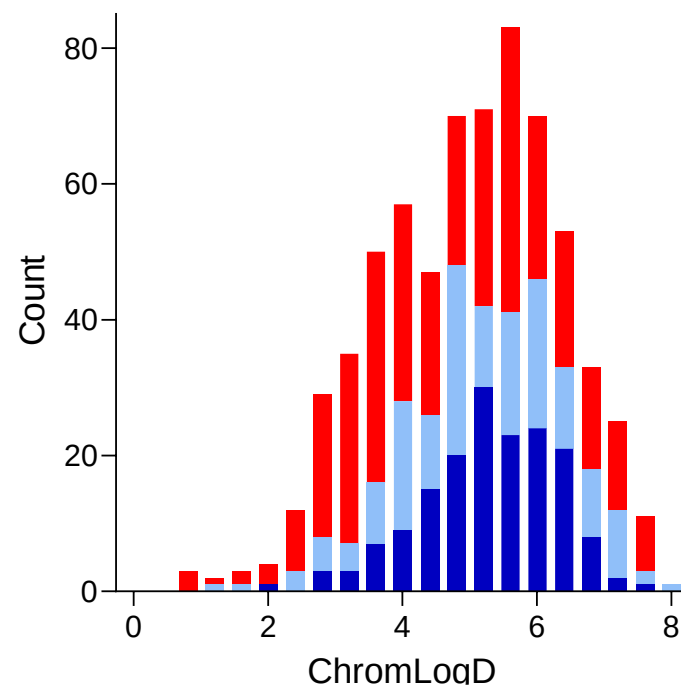
Oral Bioavailability for 691 PROTACs

- Next, we evaluated ChromLogD and EPSA values for 691 PROTACs with measured oral bioavailability in mice
- Majority of bioavailable compounds concentrate around shake-flask LogD values of 4.0-4.5 (**A**)
- A ChromLogD sweet spot for good bioavailability between 4.4 and 6.8 (**B**)
- The largest proportions of compounds with good bioavailability had EPSA between 110 and 150 (**C**)

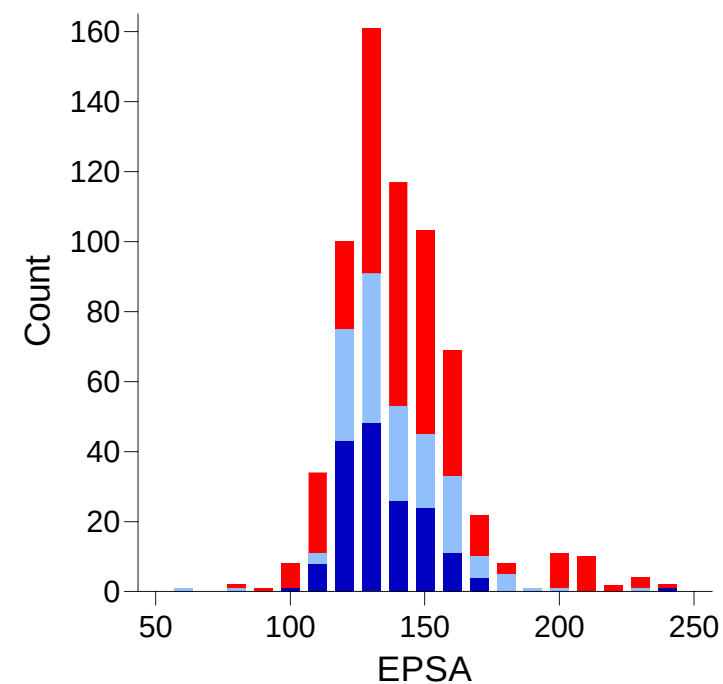
A



B



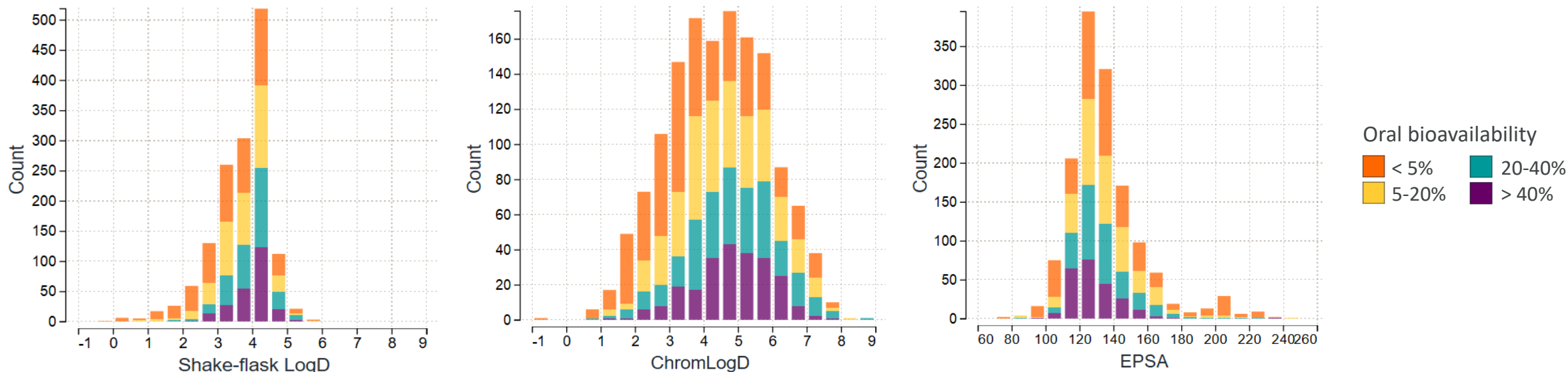
C



■ Poor bioavailability (<20%) ■ Moderate bioavailability (20-40%) ■ Good bioavailability (>40%)



Oral Bioavailability for ~1400 PROTACs (Sep 2025)



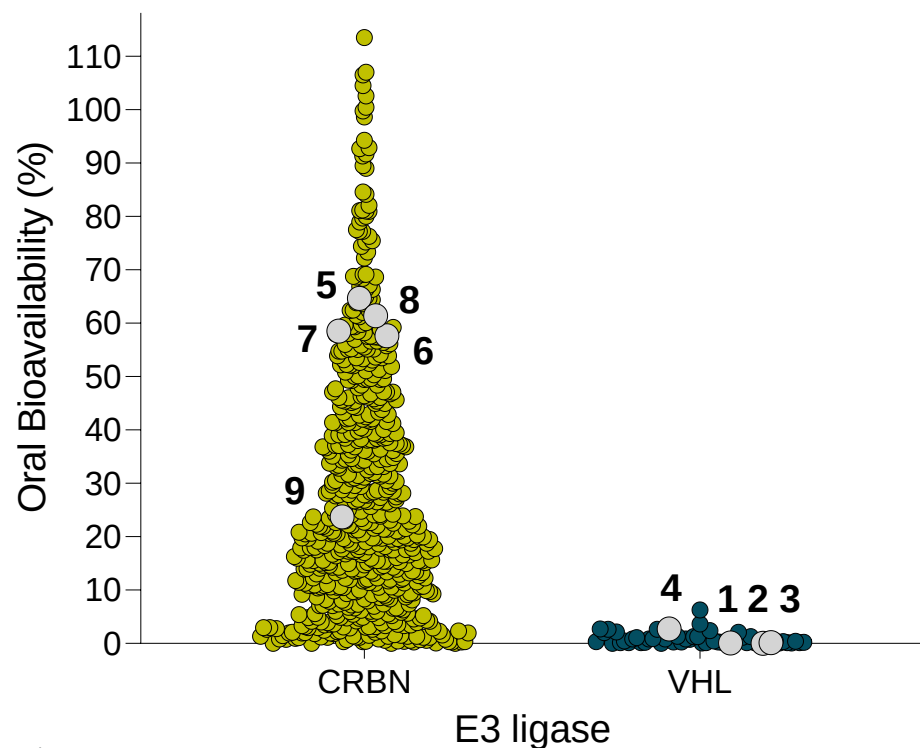
- The largest proportions of compounds with good bioavailability for:
 - Shake-flask LogD between 4.0 and 4.5
 - ChromLogD between 4.0 and 6.5
 - EPSA between 110 and 140

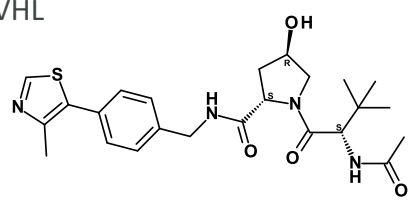
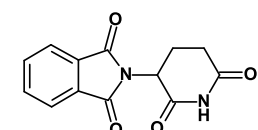
The majority of bioavailable compounds concentrate around shake-flask LogD where the assay plateaus



Oral Bioavailability for 691 PROTACs

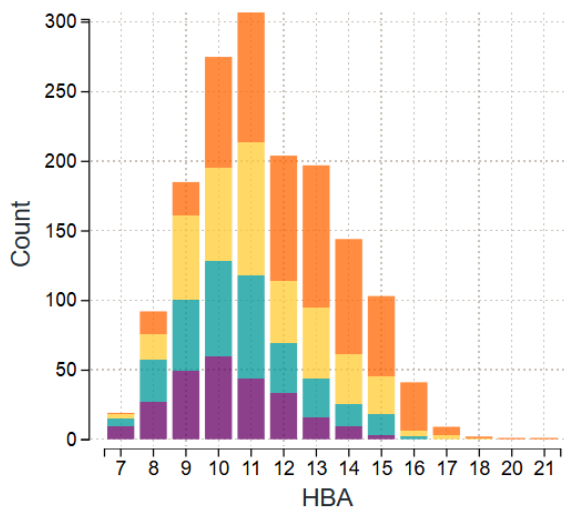
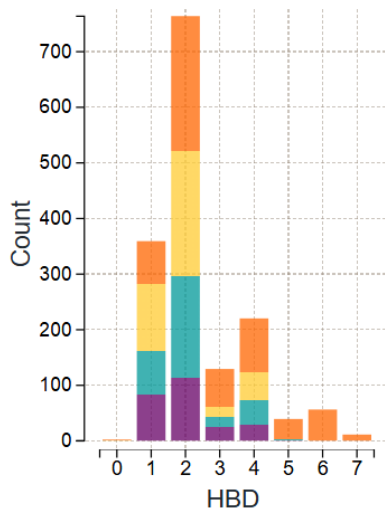
- Our in-house PROTAC data set consists of 641 CRBN- and 50 VHL-based PROTACs
- VHL PROTACs exhibit low bioavailability (< 5%) - attributed to the property of VHL warhead (high efflux and low permeability)
- CRBN PROTACs have better bioavailability – favorable properties of the thalidomide-based warhead, that is also smaller than the VHL warhead, allowing for better linker optimization



E3 warhead	Molecular Weight	Efflux Ratio	Caco2 intr. Perm. ($\times 10^{-6}$ cm/s)
VHL 	473	60	3
CRBN 	258	1	50

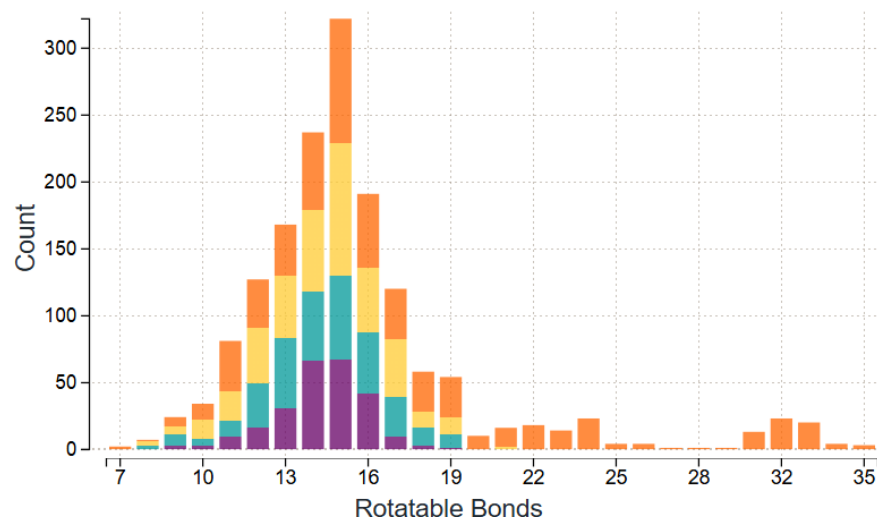
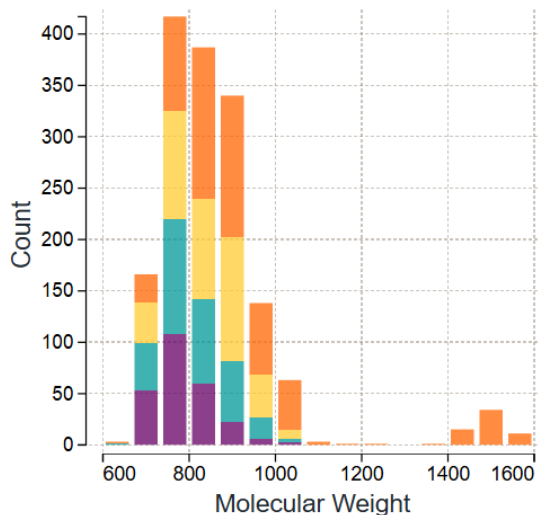


Oral Bioavailability for ~1400 PROTACs | Descriptors

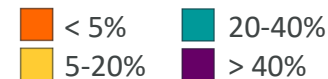


The largest proportions of compounds with good bioavailability for:

- $\text{HBD} \leq 2$
- $\text{HBA} < 13$
- $\text{MW} \leq 850$
- Rotatable bonds 13-16



Oral bioavailability



Final Remarks

- Delivery of high-quality PROTAC candidate drugs requires optimization of physicochemical and DMPK properties
- ChromLogD and EPSA experimental assays better accommodate bRo5 molecules and help determine their lipophilicity, permeability, and consequently bioavailability
- High-quality experimental assays are critical for building predictive and robust ML models – ChromLogD and EPSA models fit-for-purpose to drive PROTAC synthesis decisions
- In line with the collaborative nature of open science and artificial intelligence, we have publicly deposited ChromLogD and EPSA data that were used in the concordance testing



Acknowledgements

Zulma Santisteban Valencia

Jennifer Kingston

Hannah Rowbottom

Nadia Mann

Sophie Davies

Martin Ekblad

Silvio Di Castro

Karolina Kwapien

Erik Malmerberg

Stig D. Friis

Thomas Lundbäck

Tomas Leek

Johan Wernevik

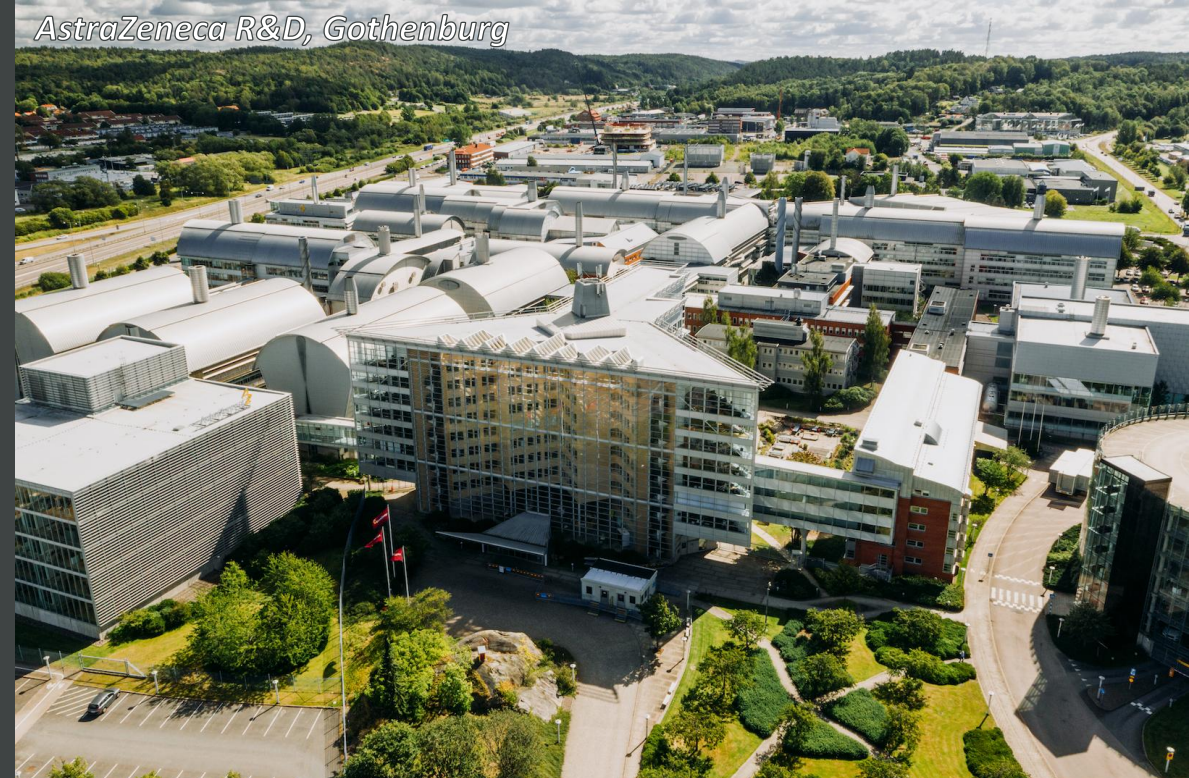
Link to paper



Link to data



AstraZeneca R&D, Gothenburg



AstraZeneca R&D, Cambridge



Confidentiality Notice

This file is private and may contain confidential and proprietary information. If you have received this file in error, please notify us and remove it from your system and note that you must not copy, distribute or take any action in reliance on it. Any unauthorized use or disclosure of the contents of this file is not permitted and may be unlawful.

AstraZeneca PLC, 1 Francis Crick Avenue, Cambridge Biomedical Campus, Cambridge, CB2 0AA, UK
+44(0)203 749 5000
www.astrazeneca.com

