Towards data-driven design of visible-light photoswitches as photopharmacological agents

Pavel SIDOROV
WPI-ICReDD, Hokkaido University







XXXII Symposium on Bioinformatics and
Computer-Aided Drug Discovery (BCADD-2026)

October 20, 2026

ICReDD philosophy and our experience

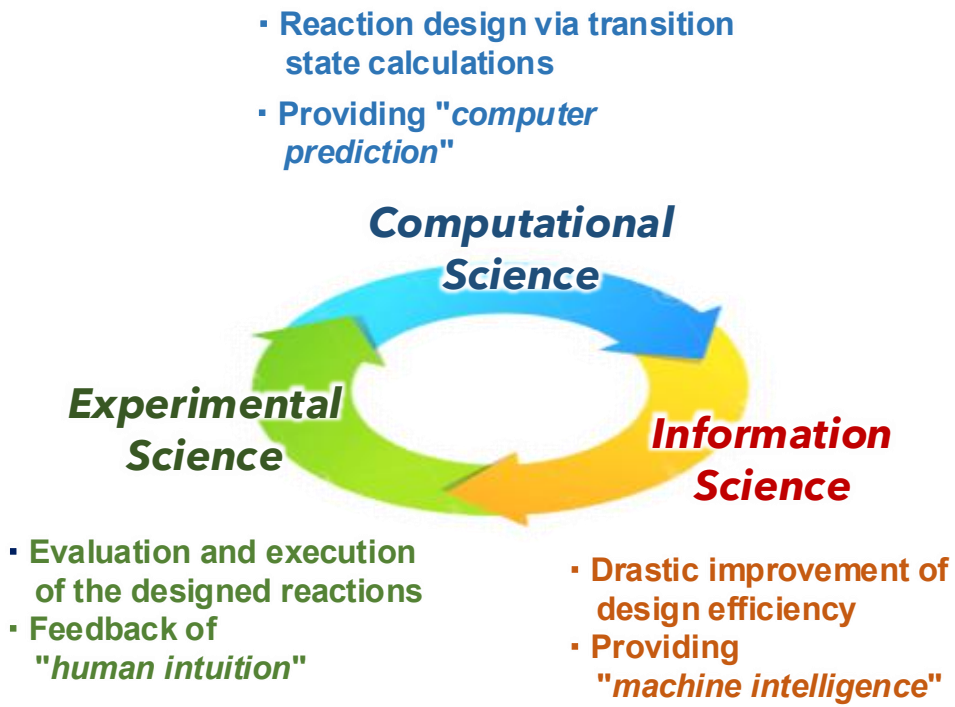
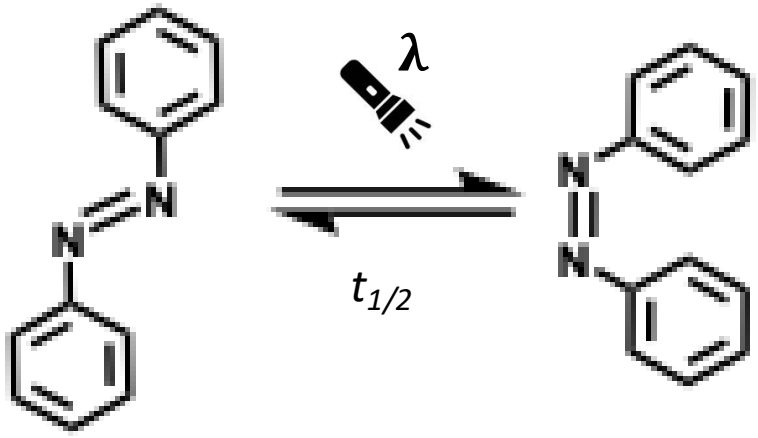


Image by Prof Maeda, WPI-ICReDD

- ICReDD strongly encourages the projects with emphasis on collaboration between fields of chemistry (experimental, computational, informatics).
- In our case – **experimental chemists**
- We are working on:
  - Predictive modeling for data from wet lab:
    - “small data”
  - Data management
  - Data analysis

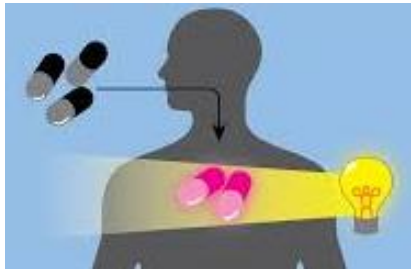
Photoswitches

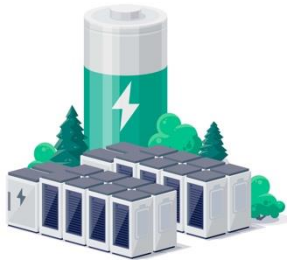
Molecular photoswitches are a class of compounds that can readily isomerize between distinct structures upon irradiation with light.



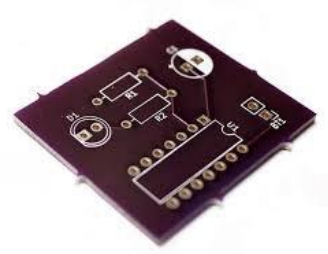
trans *cis*

photopharmacology





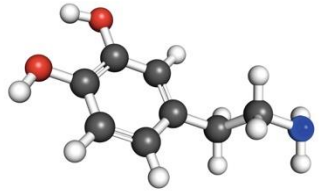
energy storage



electronics

Quantitative Structure-Property Relationship modeling

Quantitative Structure-Property Relationship (QSPR) modeling – establishing a relationship between the chemical structure (molecule, reaction, etc) and a property we want to predict.

$$Property = \mathbf{f}($$  $$)$$

$$Property = \mathbf{f}(\boldsymbol{Descriptors})$$

Mathematical relationship established with machine learning methods

Parameters directly derived from molecular structure

DOPtools: a library for automation of benchmarking and model building



SIDOROV



BYADI



GANTZER

`chem` module

- Calculation of fragments and fingerprints
  - (CircuS, ChyLine, RDkit FP)
- ColorAtom for fragment descriptors
- Solvent features

`cli` module

Command–line interface for all operations

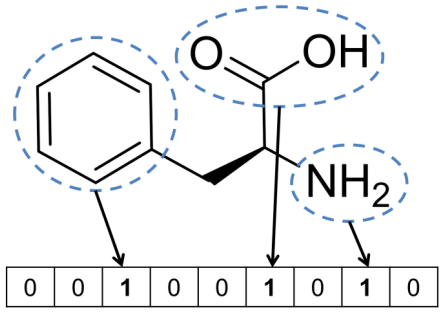
- **`launch_preparer`**
  - Input as Excel or .csv
  - Structure reading and standardization
  - Output as .svm or .csv
- **`launch_optimizer`**
  - Optimization of descriptor space and ML hyperparameters
- **`rebuilder`**
- **`plotter`**

DOPtools

`optimizer` module

- Batch calculation of descriptors
- Optimization of model hyperparameters
  - SVM, RandomForest, XGBoost
  - Regression and classification

chem module: available descriptors



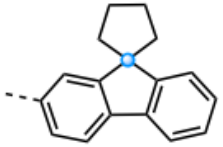


Fingerprints (RDkit)

- Morgan/features
- RDkit/layered
- Avalon
- Torsion
- AtomPairs

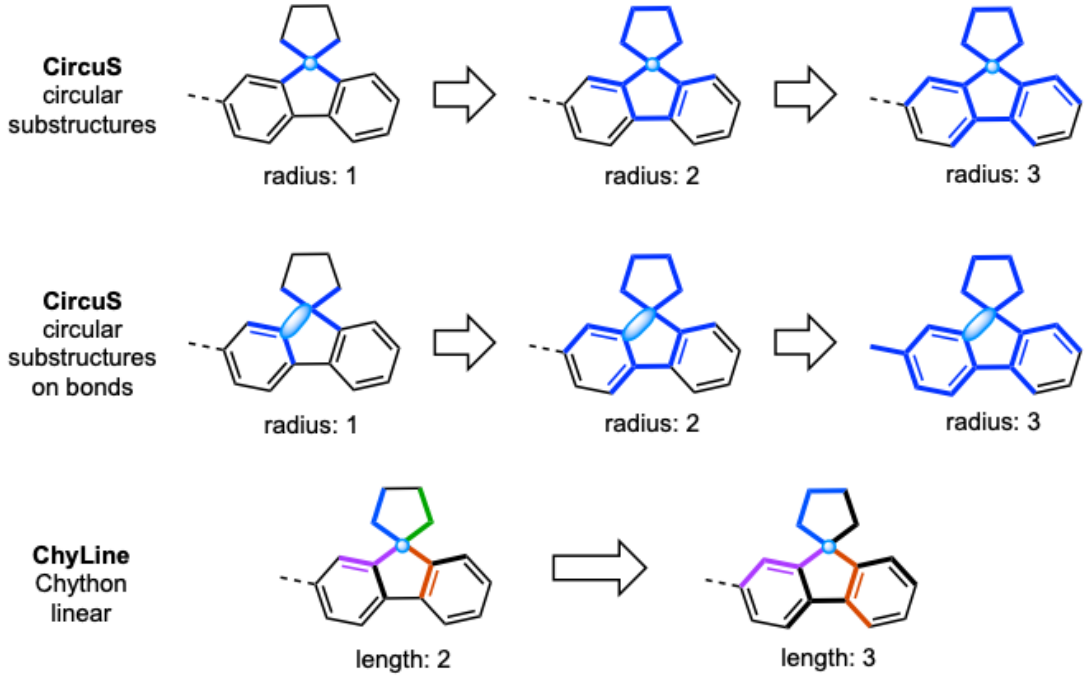
Solvents

- Acidity
- Basicity
- Polarizability
- Dipolarity



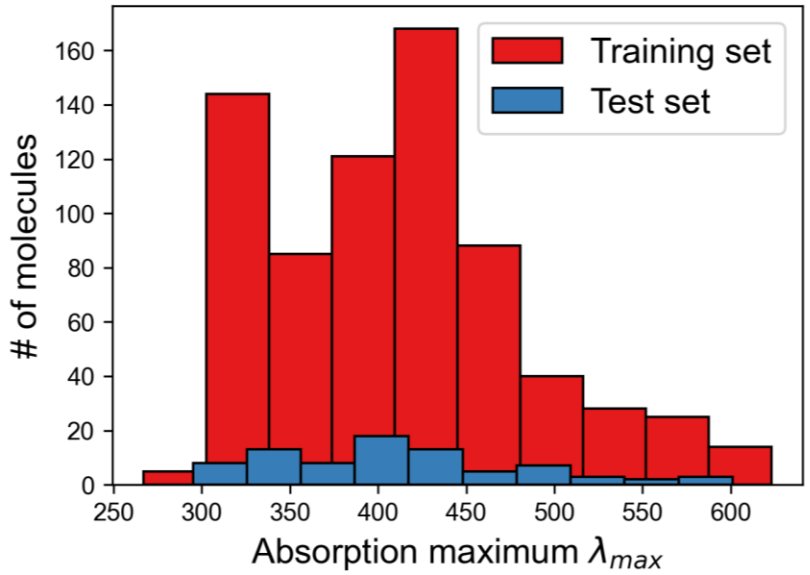
Fragments

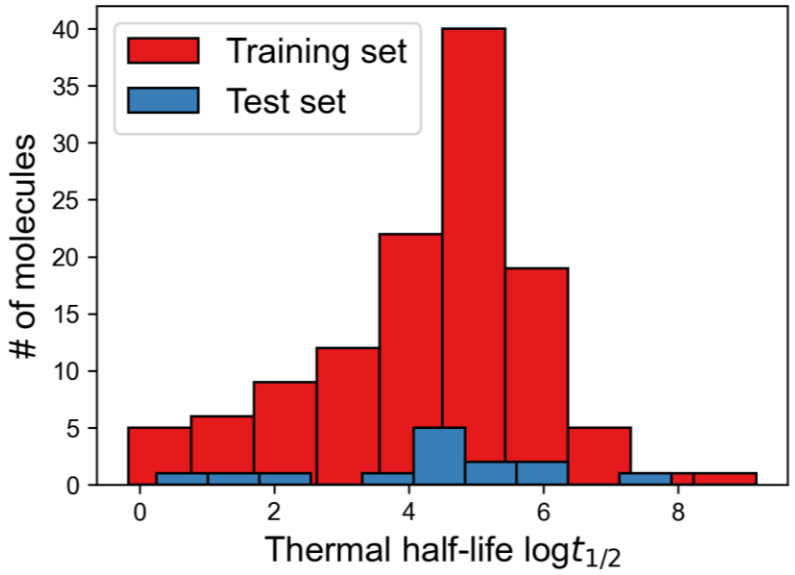
- Circular (CircuS)
- Linear (ChyLine)



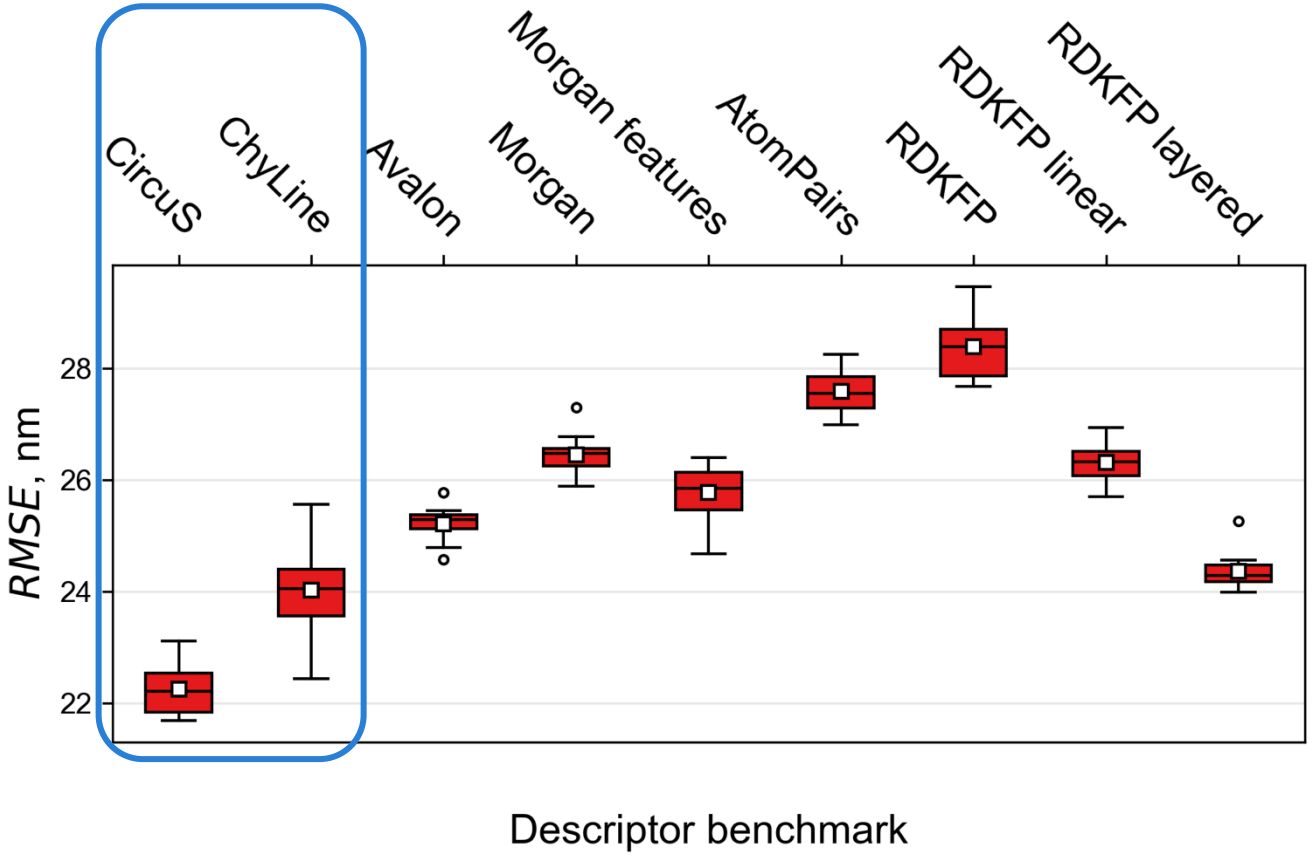
Initial studies: literature photoswitch data

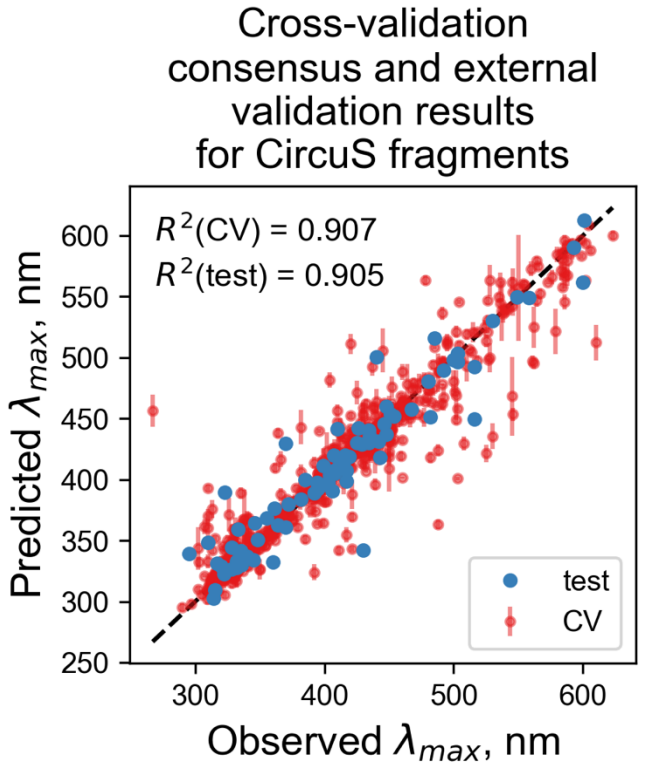
- **Literature data:**
  - ~800 molecules for λ_{max}
  - ~130 molecules for $t_{1/2}$
- **Training/test set separation**: 90% training / 10% test
- **Stratified** by property



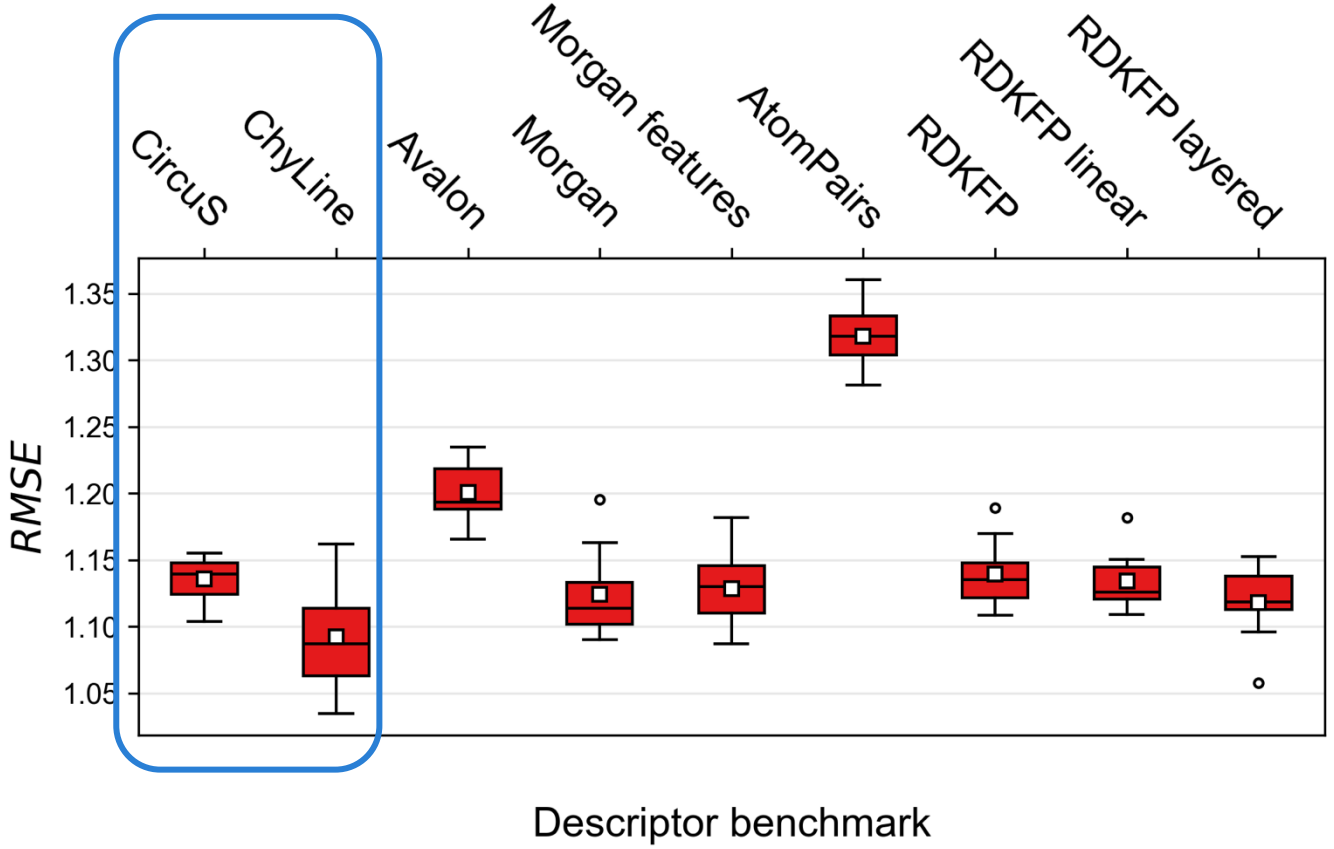


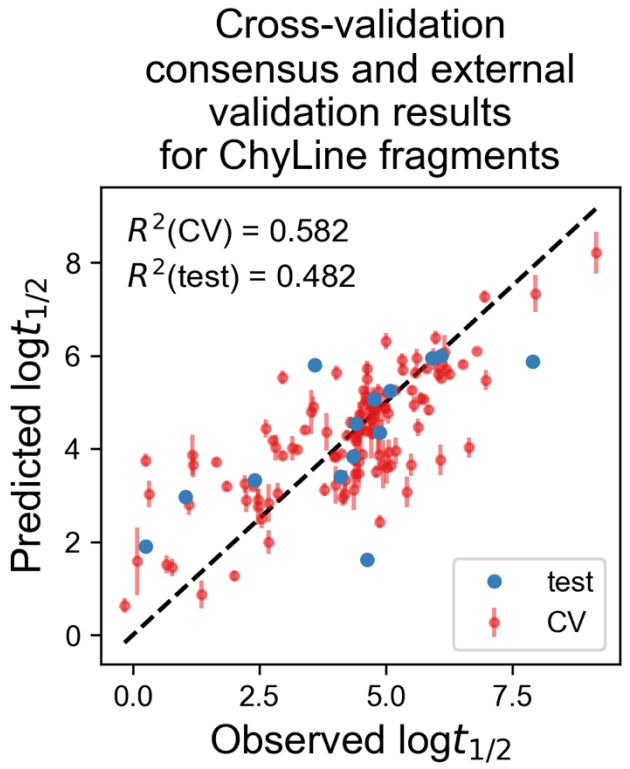
Benchmark results: Prediction of absorption





Benchmark results: Prediction of thermal half-life

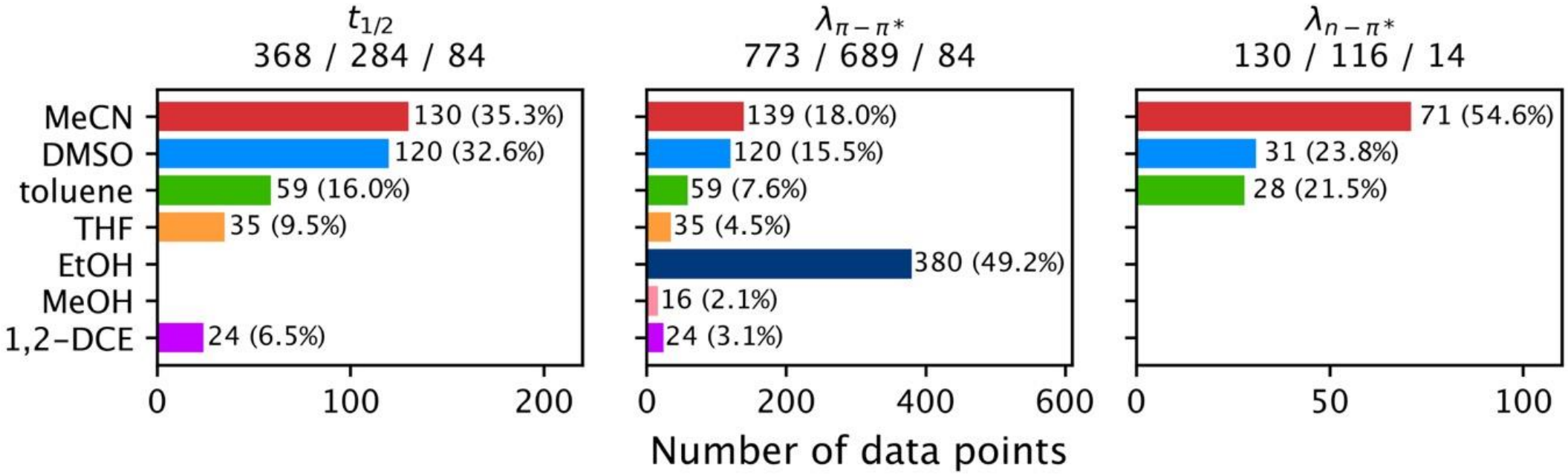


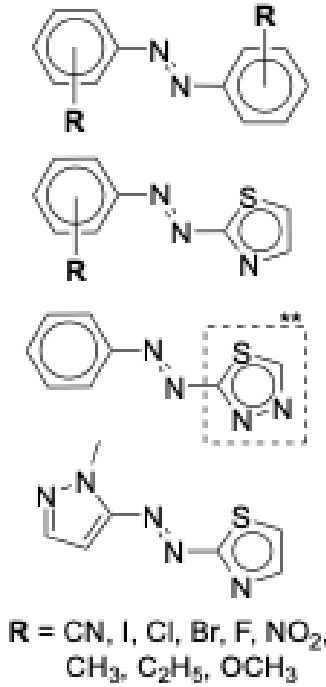


- Lack of data
- Model applicability
- Complexity of isomerization mechanism

Expanding the dataset

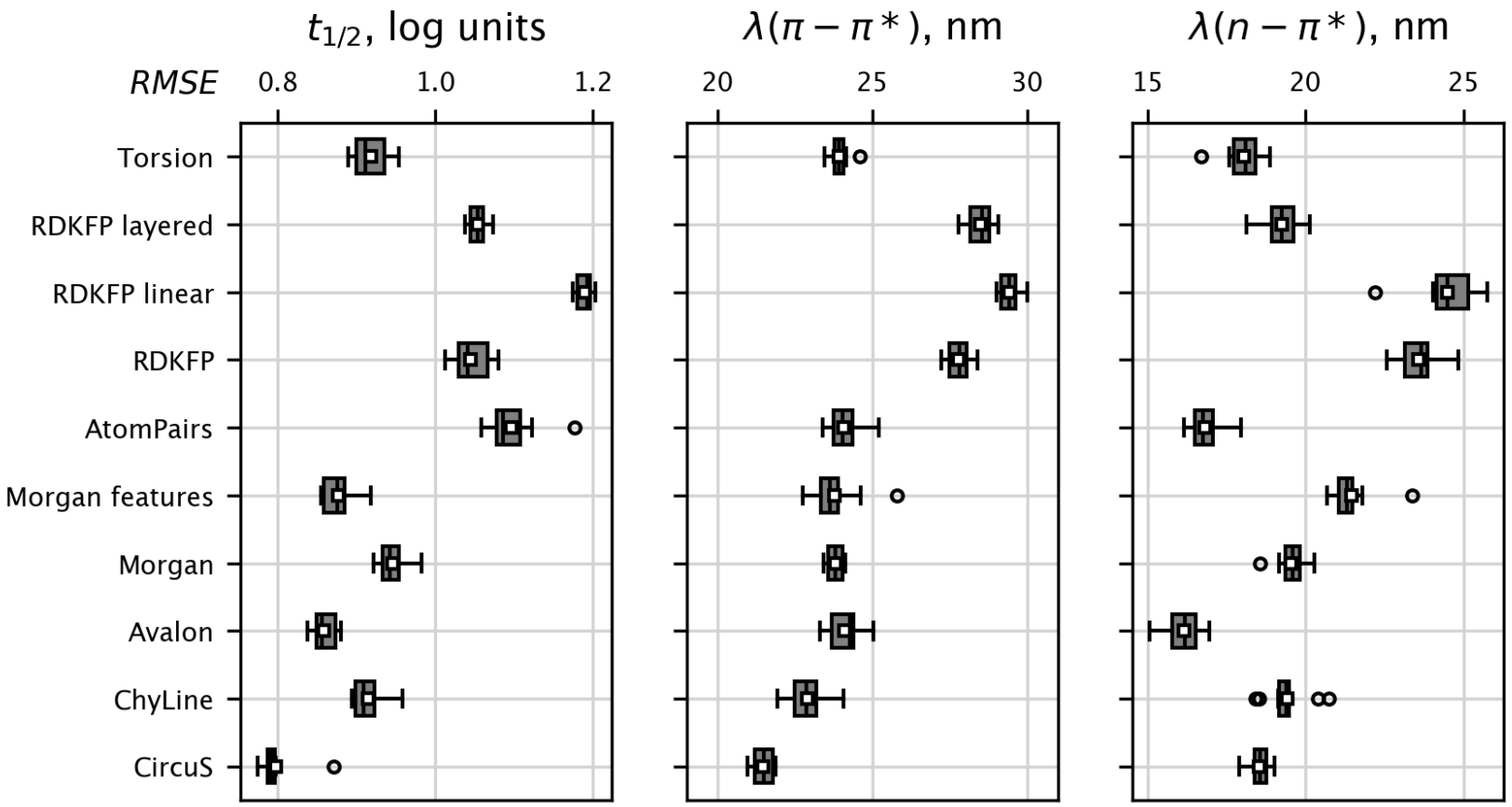
The data for all properties – $t_{1/2}$, $\lambda_{\pi\text{-}\pi^*}$, and $\lambda_{n\text{-}\pi^*}$ – was collected from literature, and further intra-laboratory measurements were performed in various solvents.



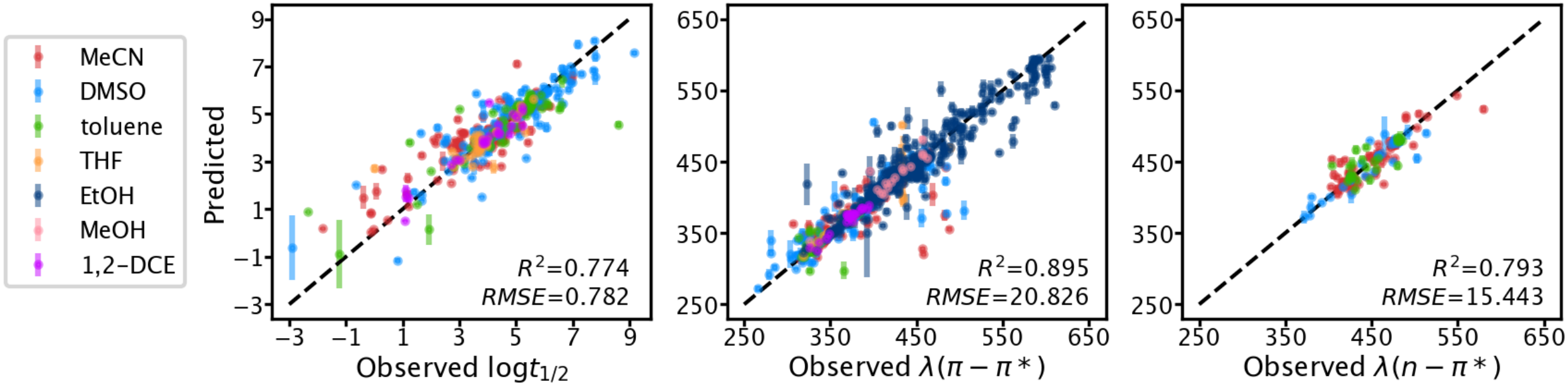


Benchmark results for the expanded dataset

- Benchmarks for half-life and $\lambda_{\pi\text{-}\pi^*}$ showed the superiority of the fragment descriptors.
- On the other hand, Avalon fingerprints were the best for $\lambda_{n\text{-}\pi^*}$.

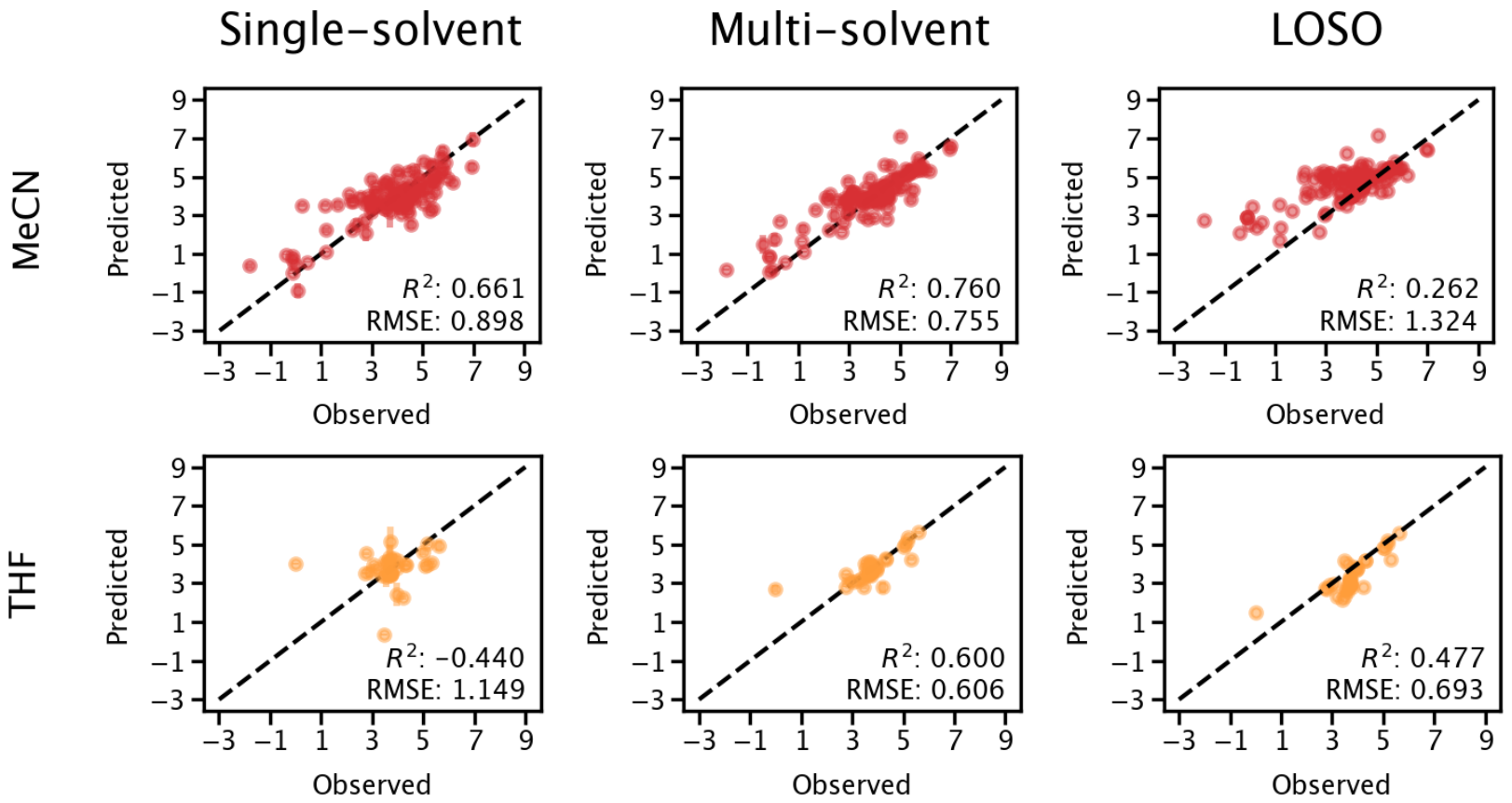


Benchmark results: the best models



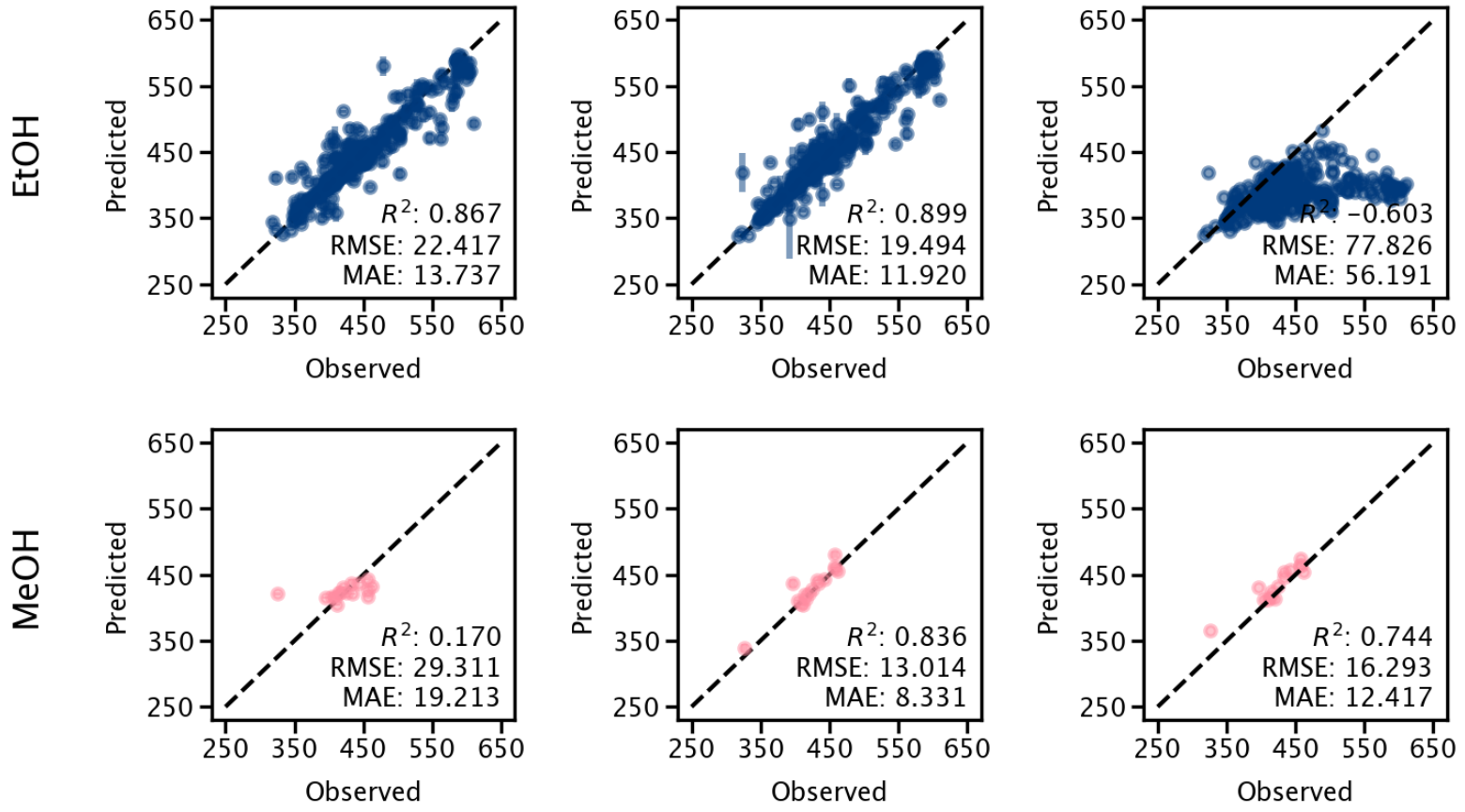
Model validation: single solvent vs multi-solvent vs LOSO

- **Multi-solvent** model achieved higher performance in all solvents in cross-validation compared to single solvent
- **Leave-one-solvent-out** models were predictive in datasets with lower data count, but struggled with solvents overrepresented in the dataset



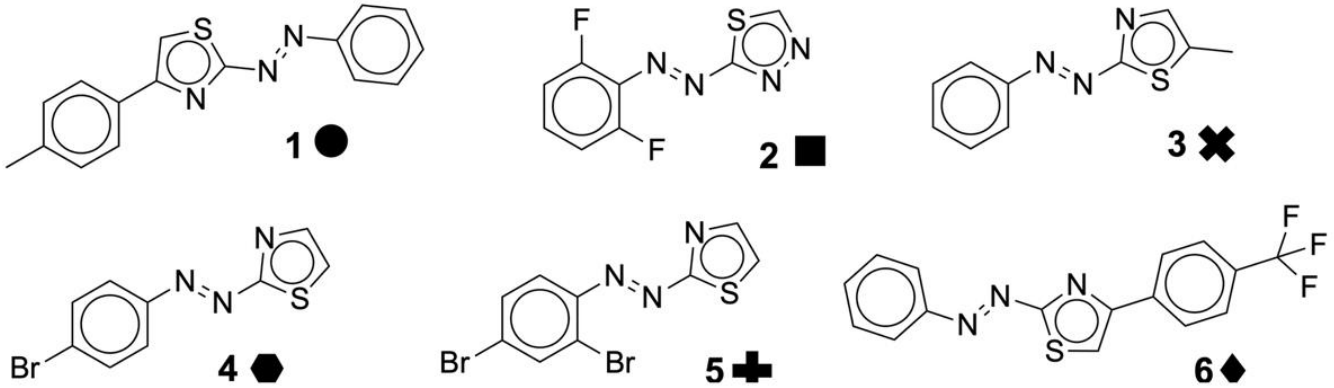
Model validation: single solvent vs multi-solvent vs LOSO

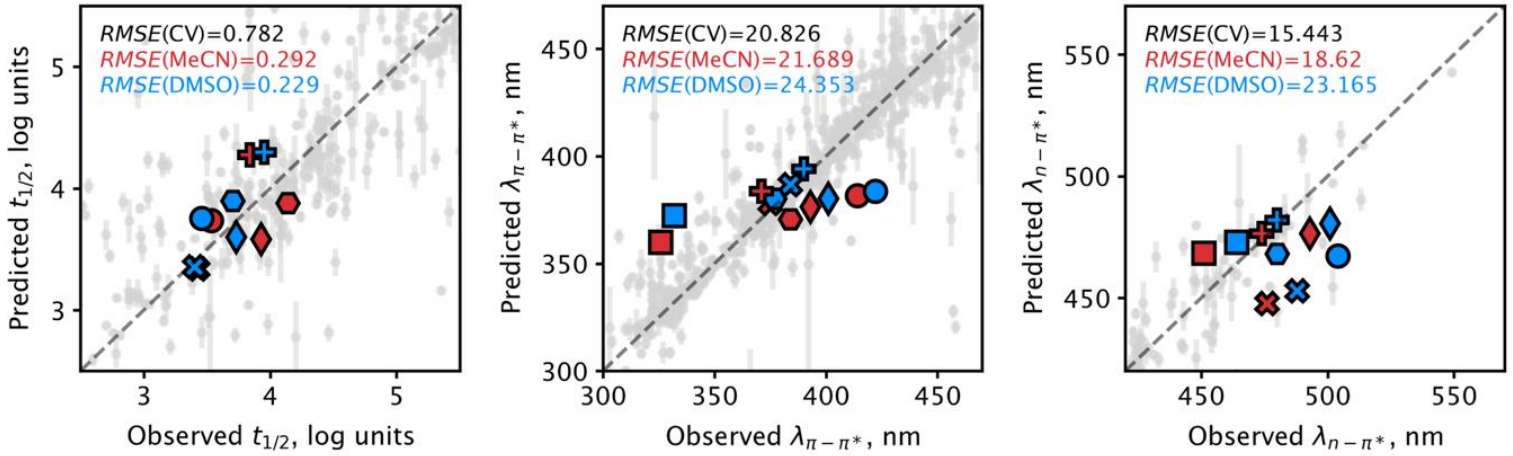
- The effect was even more pronounced for $\lambda_{\pi\text{-}\pi^*}$ models: ethanol that represented around 50% of all data was not predicted as well as other, smaller solvent sets.



External validation on newly synthesized compounds

- To assess the practical utility and generalizability of our models, we designed and synthesized six novel azo heteroaryl derivatives.
- Spectroscopic and thermal isomerization studies were conducted in acetonitrile and DMSO at 25°C, with all compounds exhibiting first-order kinetics.

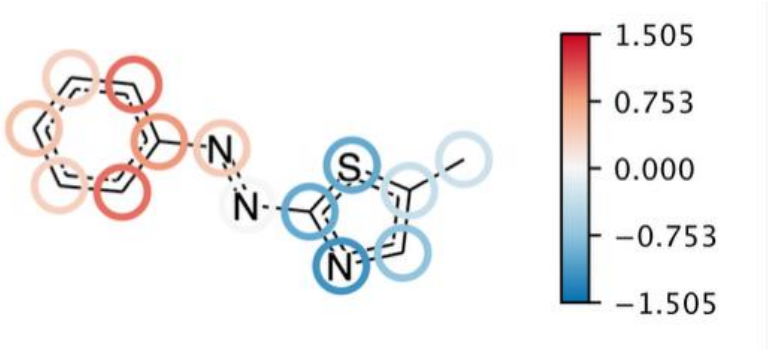




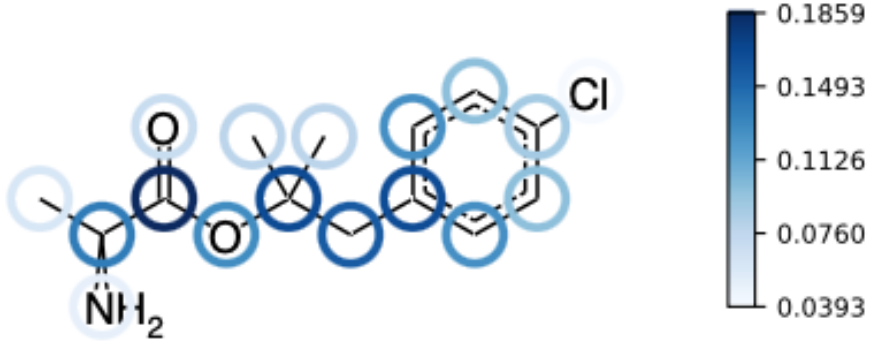
ColorAtom

- To interpret the predictions of the model, we use ColorAtom methodology. The contributions are calculated from partial derivatives of descriptors and are shown as colors around each atom.
- The colors are scaled to the maximum weight and are relative (a very bright color means the contribution is maximum, the less bright, the lower is the contribution).

Regression: **Red** is positive contribution, **Blue** is negative.

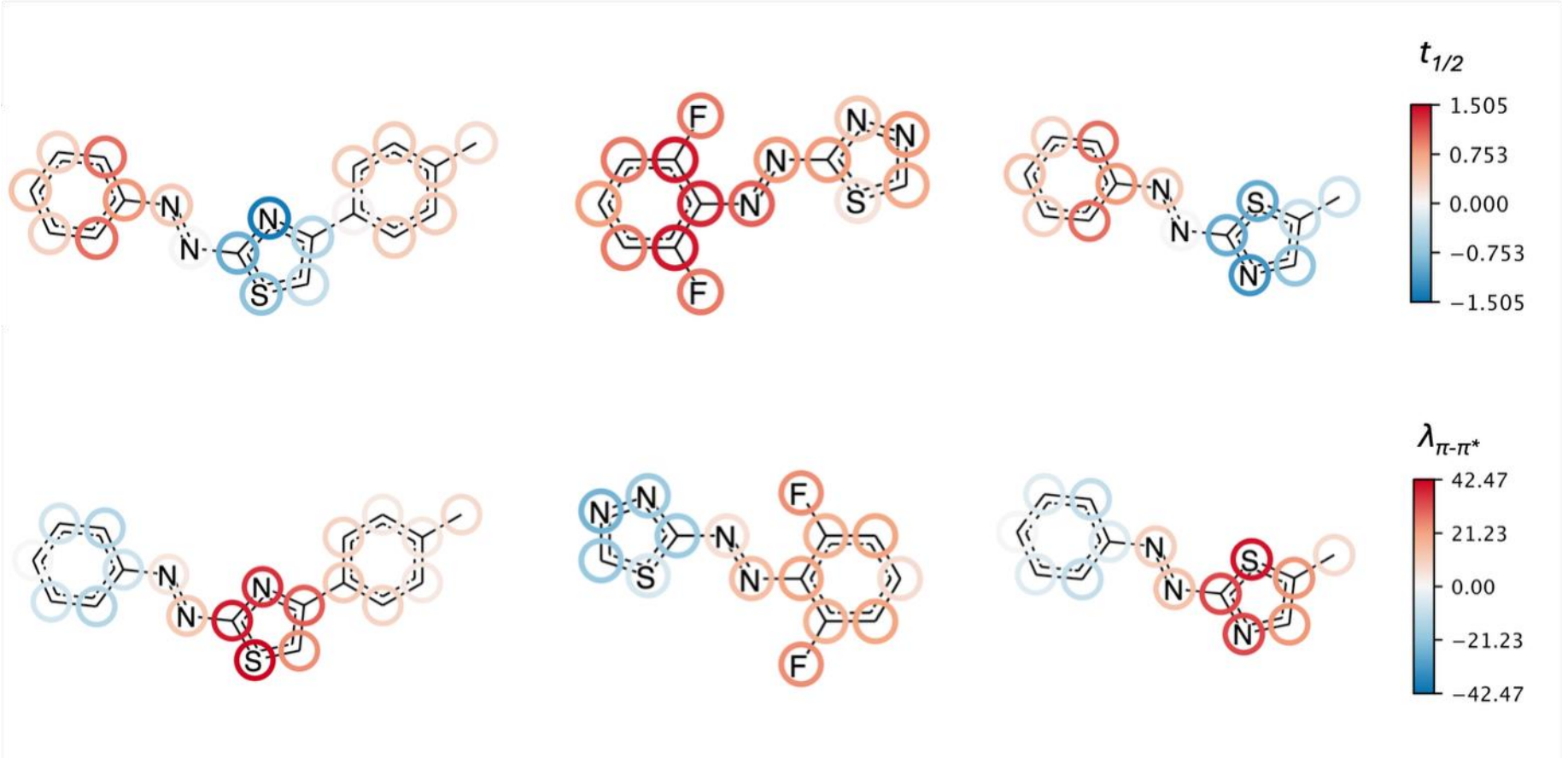


Classification: Color intensity shows the importance

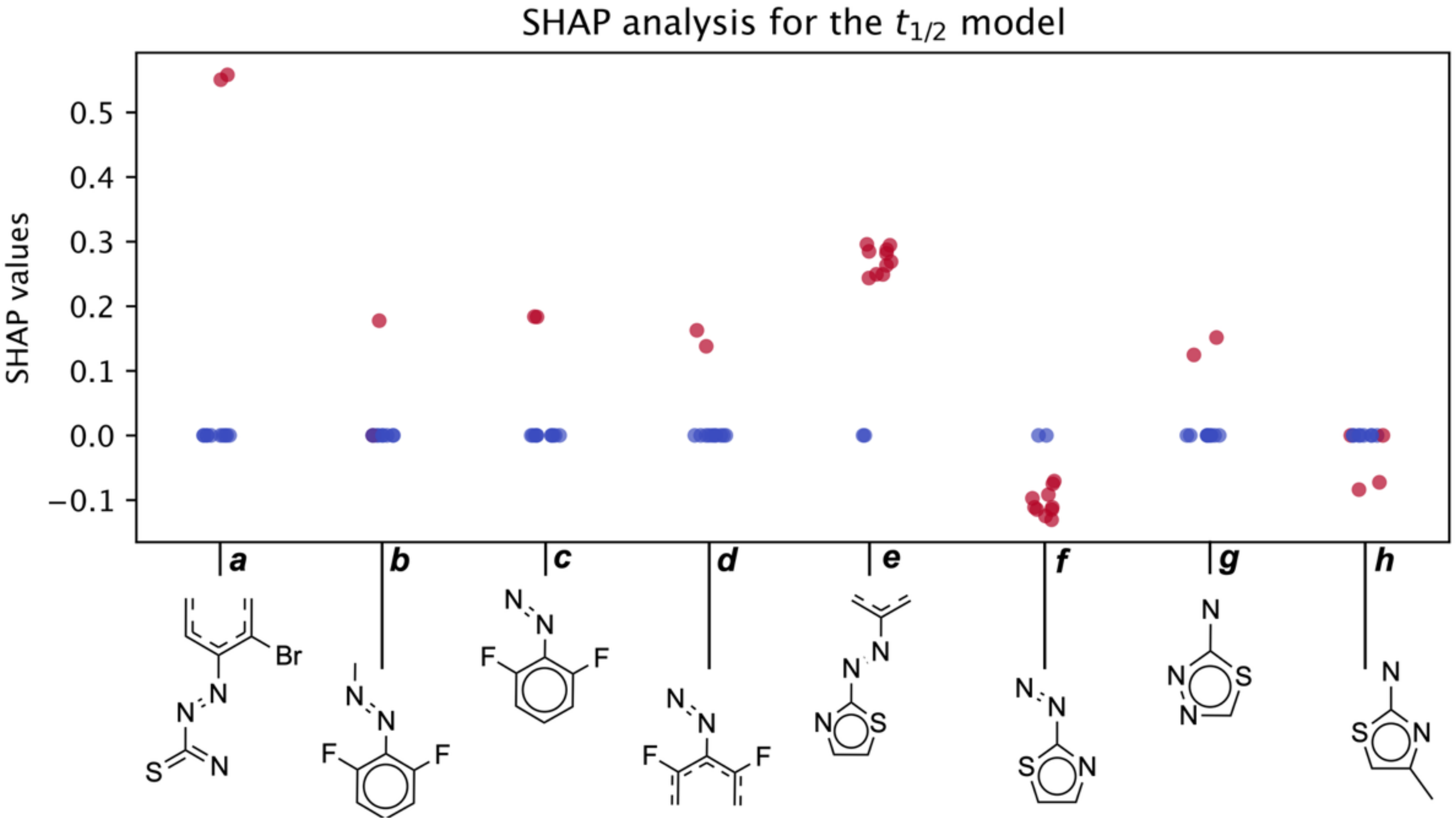


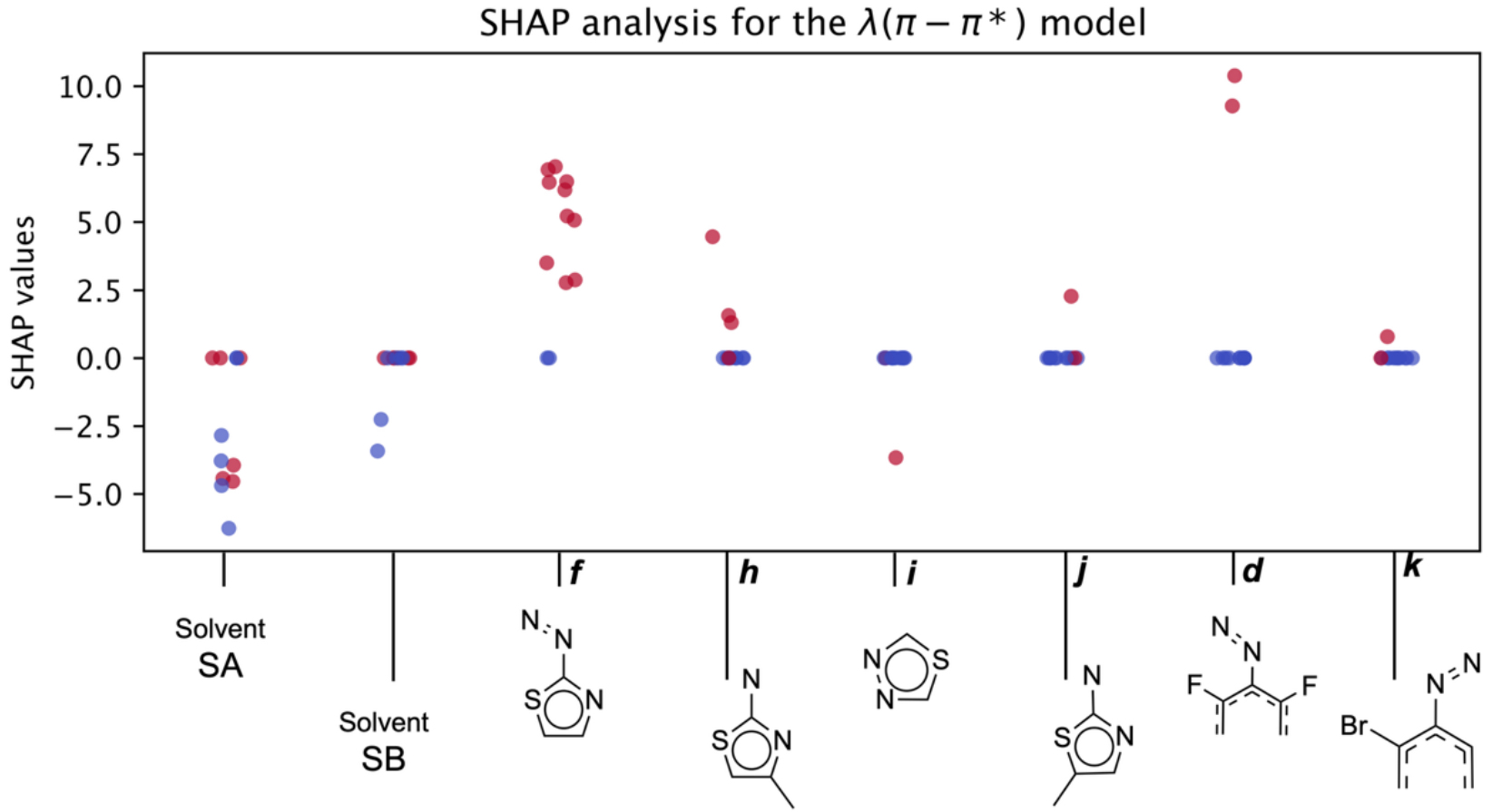
ColorAtom analysis

- ColorAtom highlighted the substructures in the molecules that correspond to increases and decreases of the property values.
- Importantly, it noted the often-observed inverse relationship between $t_{1/2}$ and $\lambda_{\pi\text{-}\pi^*}$.

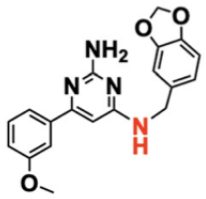


SHAP analysis

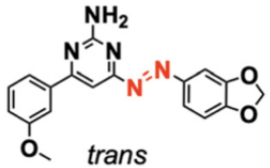



SHAP analysis for the λ(π − π*) model
SHAP values
10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
Solvent
SA
Solvent
SB
f
h
i
j
d
k
N
N
S
N
F
F
Br

Photoswitchable molecules as agonists of Wnt signaling pathway



BML-284



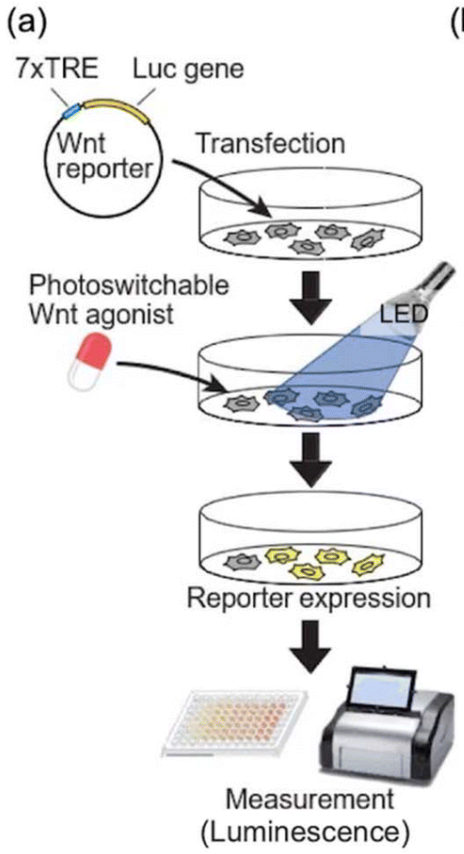
trans

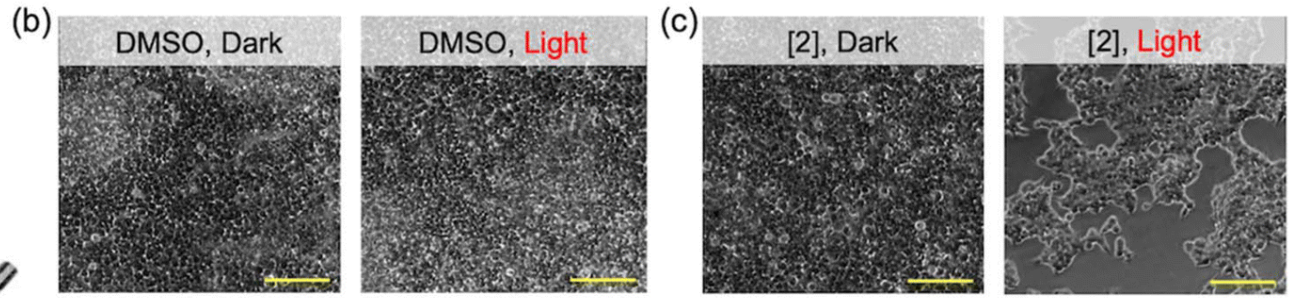
525 nm ⇅ 405 nm, 450 nm

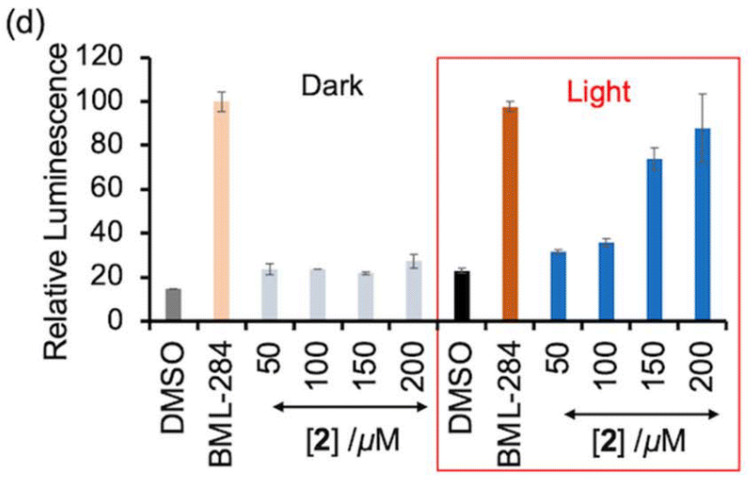


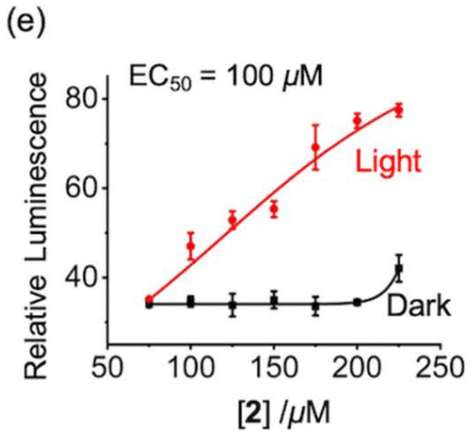
cis

1











IMAJO

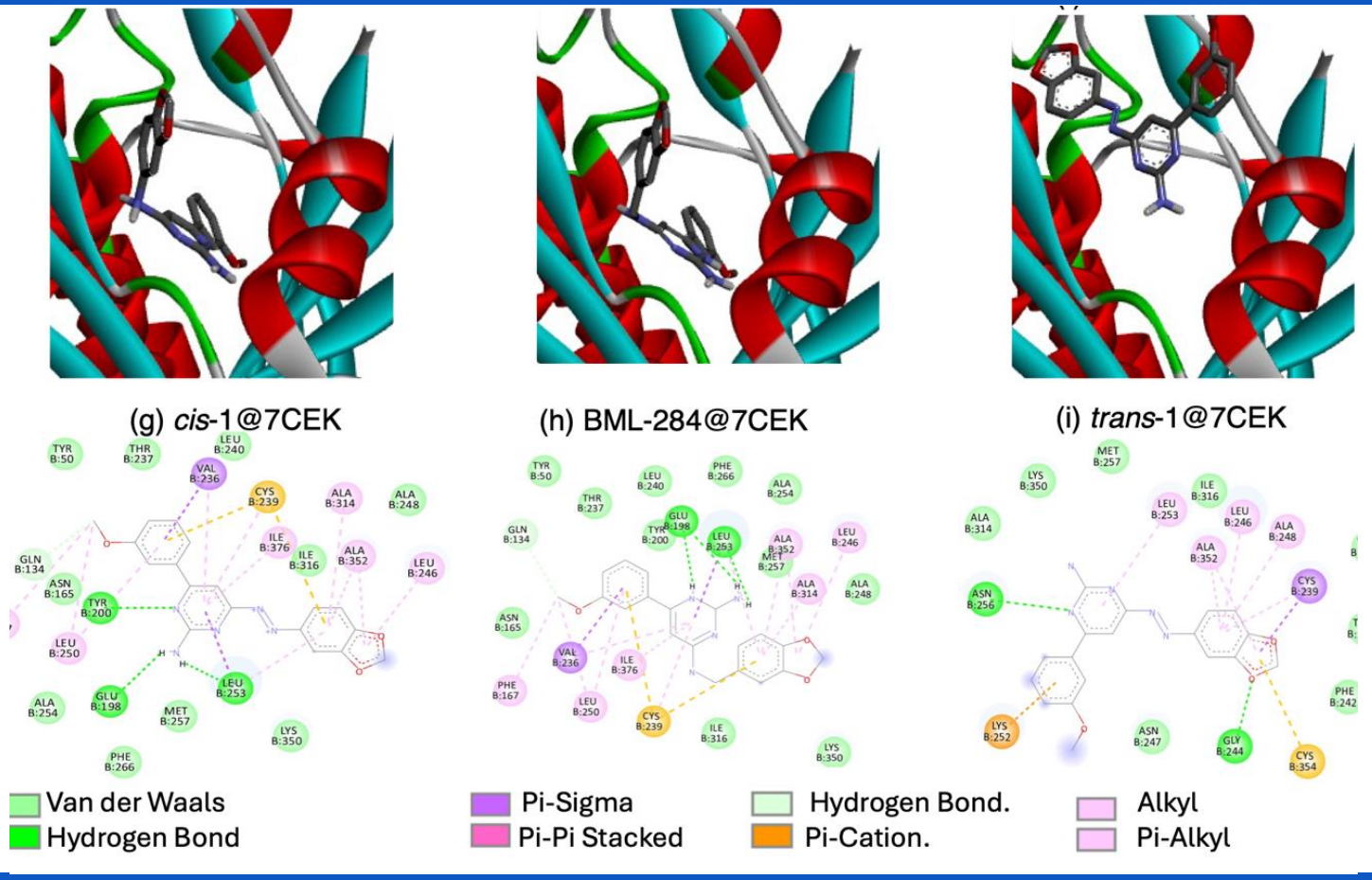


HASHIM

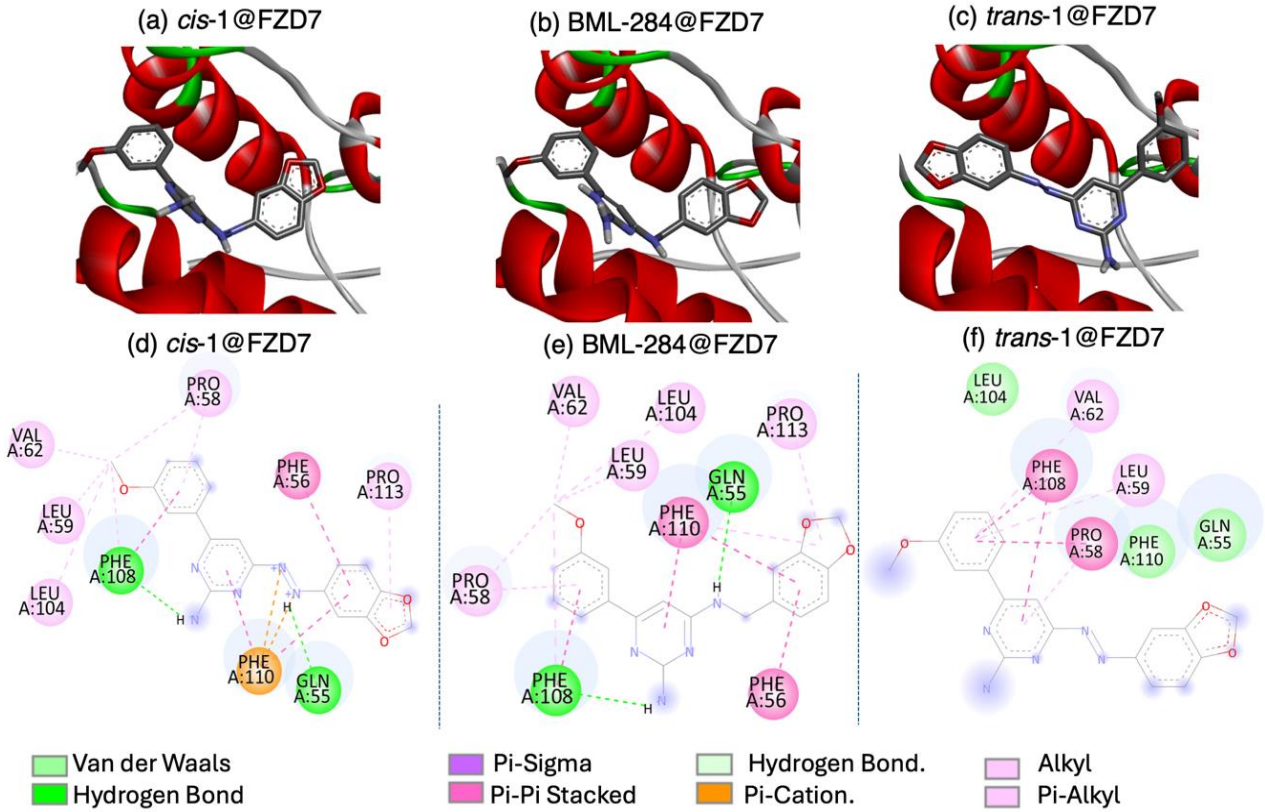
Tubulin docking studies

Ligand / pose	Docking score	RMSD
BML-284	−8.1	0.38 Å
cis-1	−7.7	0.67 Å
trans-1	−6.6	6.93 Å

- Tubulin is used as a benchmark, not as the proposed Wnt target.
- Cis-1 reproduced the BML-284 pocket position better than trans-1.
- Cis-1 interacts with the same amino acids as BML-284.



Wnt pathway target elucidation – FZD receptors



- FZD were selected as potential targets.
- Docking across FZD CRDs showed repeated preference towards cis-isomer.

Frizzled	BML	cis-1	trans-1	cis-1– trans-1 (Δ)
FZD2 (7X8P)	-8.7	-8.7	-7.9	-0.8
FZD4 (5BPB)	-8.1	-8.1	-8	-0.1
FZD5 (5URY)	-8.8	-8.8	-8.2	-0.6
FZD7 (5T44)	-8.5	-8.6	-7.9	-0.7
FZD8 (1IJY)	-8.2	-8.2	-7.2	-1
FZD10 (7X8T)	-6.7	-6.5	-6.4	-0.1

Ligands	Interacting amino acids
BML-284	GLN55 PHE56 PRO58 LEU59 VAL62 LEU104 PHE108 PHE110 GLN111 PRO113
cis-1	GLN55 PHE56 PRO58 LEU59 VAL62 LEU104 PHE108 PHE110 PRO113
trans-1	PRO58 LEU59 VAL62 PHE108

Conclusions

- Using machine learning on structural features (fragment counts of various topologies), we have successfully predicted photoswitching properties of various azo-photoswitches with high accuracy.
- Fragment descriptors allow to not only build accurate models, but also to apply the model interpretation techniques such as ColorAtom to guide subsequent design efforts and experiments.
- Docking studies allow to identify potential targets, further confirmed and evaluated by MD simulations.





IMAJO



HASHIM

ICReDD
Institute for Chemical Reaction Design and Discovery
HOKKAIDO UNIVERSITY

wpi
World Premier International Research Center Initiative

LIST PF

科研費
KAKENHI