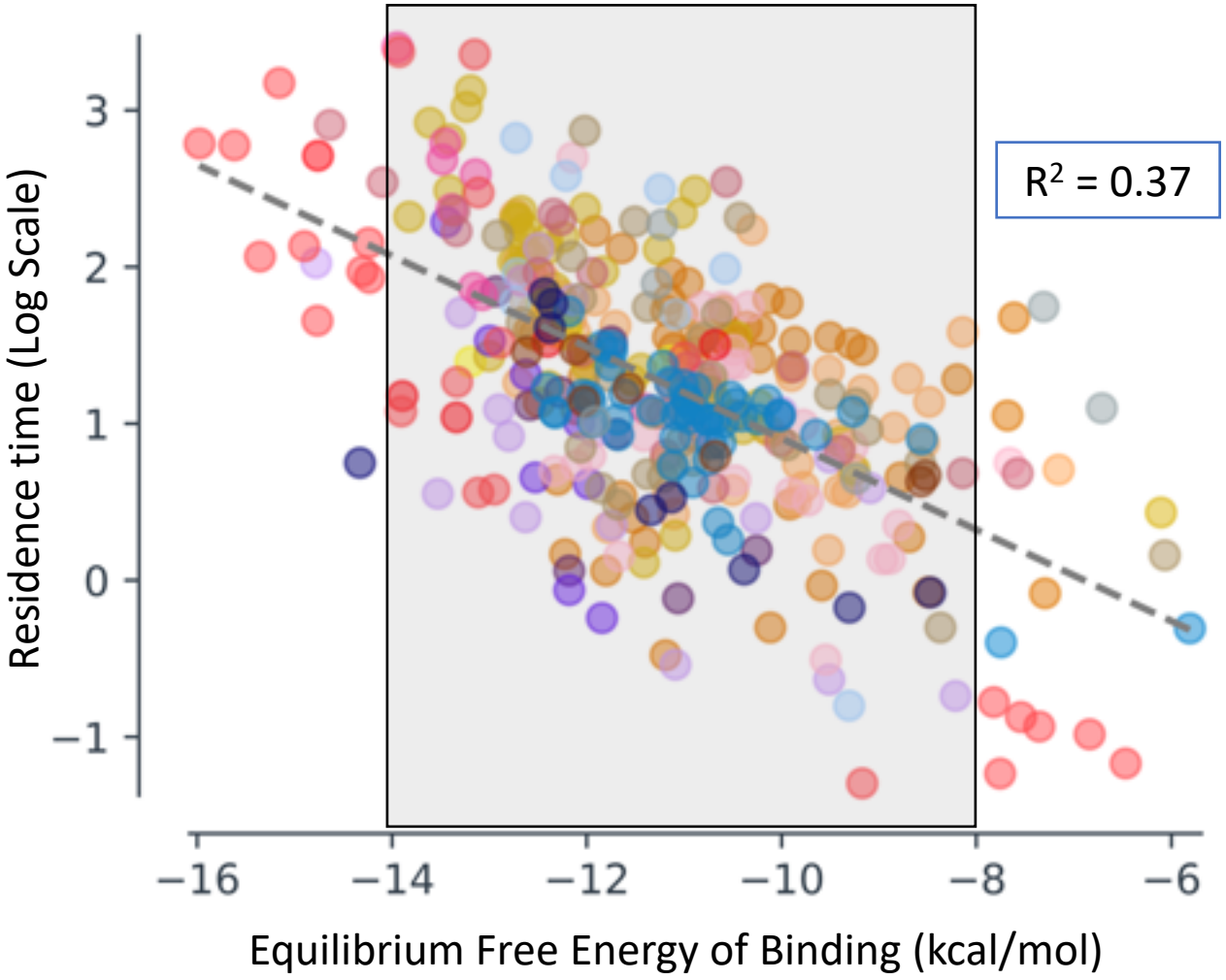
Using machine learning to understand the molecular properties that confer extended drug-target residence time

Andrew Potterton, Alexander Heifetz, Peter Coveney and Andrea Townsend-Nicholson

Introduction: Residence time is important in drug discovery

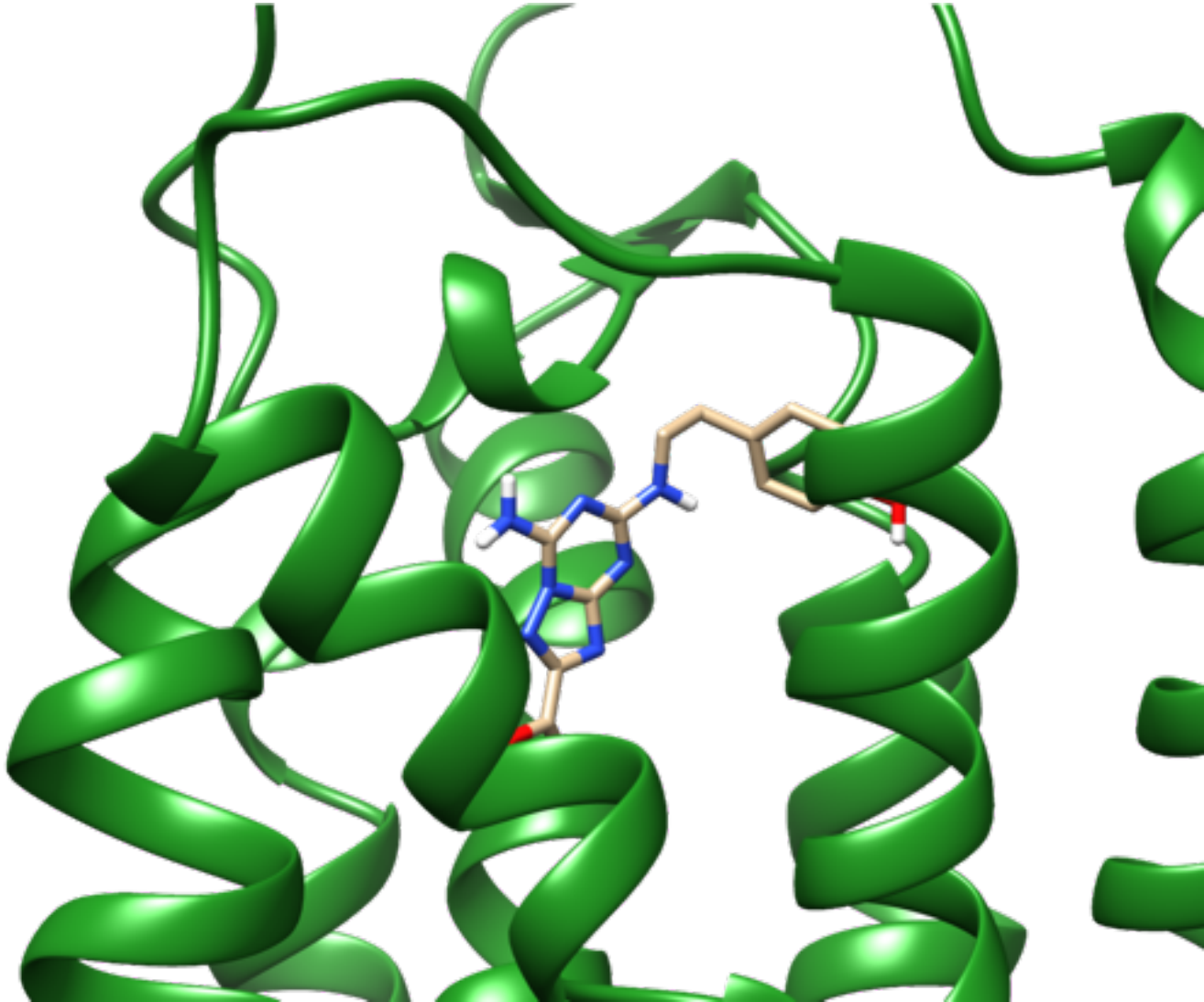


There is a weak correlation between GPCR kinetic and equilibrium binding values

- Great levels of attrition in drug discovery
- nM binders have great variability in residence time
- *In vivo* efficacy GPCR ligands has been proven to be linked to residence time
- Notable reviews suggest that residence time should be optimized in hit to lead and lead optimization phases of drug discovery

GPCRs

- 7TM helical structure
- 35% of drugs target GPCRs
- Cardiac, neurological and respiratory therapeutic targets



Kinetic Database

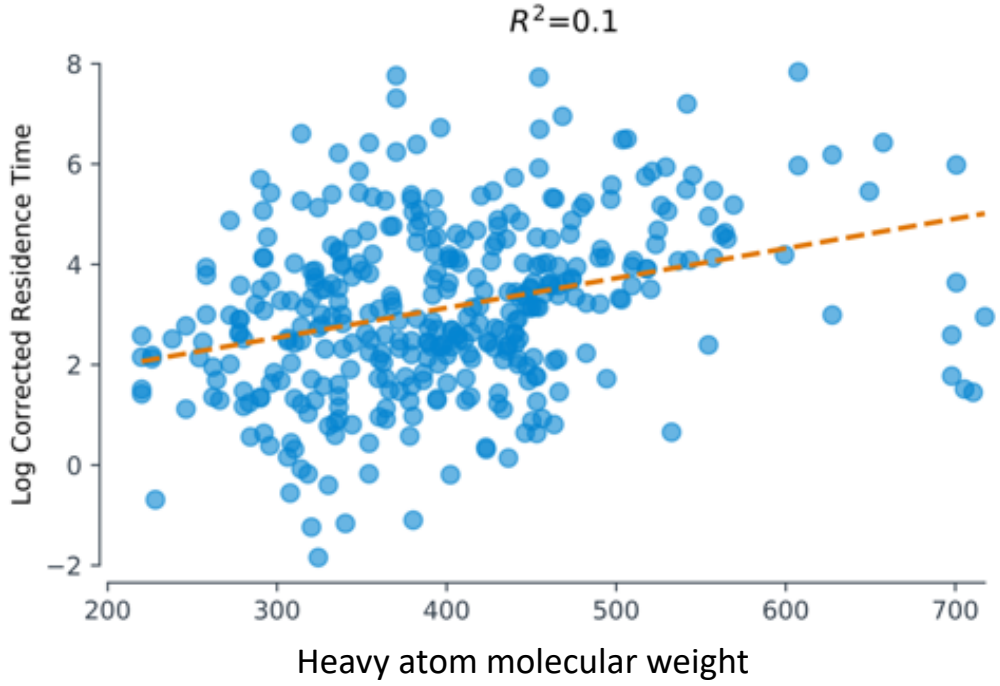
- Scraped publications for 550 entries
- Mainly radioligand kinetic binding assays
- Database contains:
  - 2D chemical structure
  - Kinetic binding data (temperature corrected)
  - Equilibrium binding data (K_D/K_i)
  - Clinical trials data

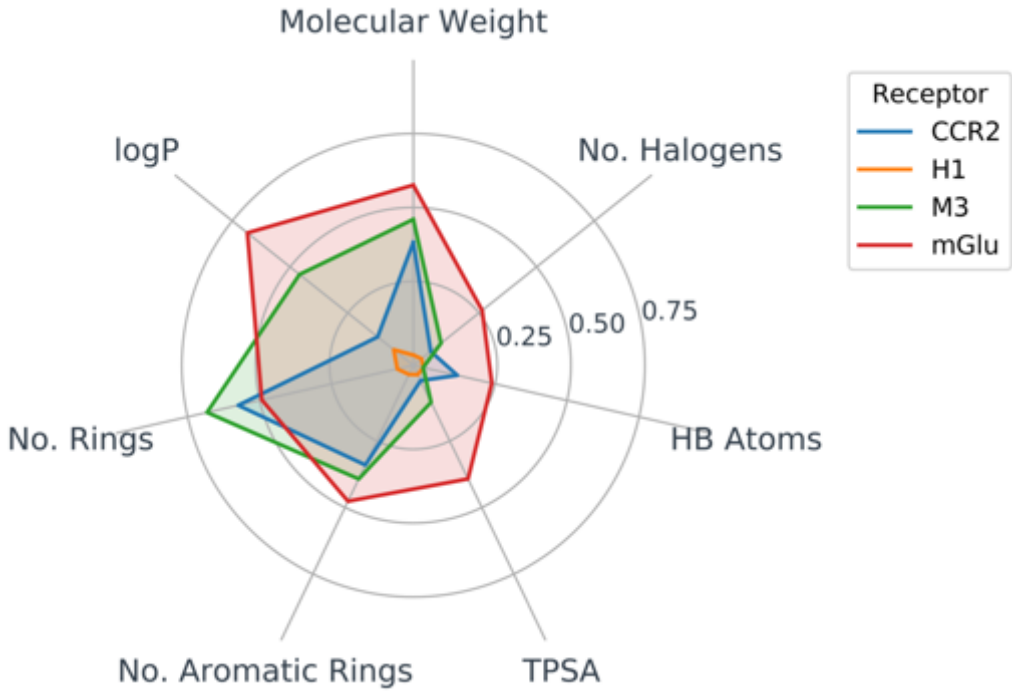
Receptor	GPCR Class	No. of entries	No. of unique ligands
5-HT_{2A}	A	1	1
5-HT_{2B}		7	7
Adenosine A_1		63	50
Adenosine A_{2A}		50	44
Adenosine A_{2B}		3	3
Adenosine A_3		65	56
α_{2A} Adrenergic		8	8
α_{2A} Adrenergic		8	8
β_1 Adrenergic		5	5
β_2 Adrenergic		32	20
Cannabinoid Type 1		12	10
Cannabinoid Type 2		24	23
C-C Chemokine Type 2		39	27
Histamine H_1		67	29
Muscarinic Acetylcholine M_2		17	8
Muscarinic Acetylcholine M_3		39	27
μ-Opioid		10	9
Tachykinin NK_1		7	7
Orexin Type 2		6	6
Corticotropin-Releasing Hormone 1	B	23	21
Metabotropic glutamate $mGluR_2$	C	51	50

Summary of ligand entries in database

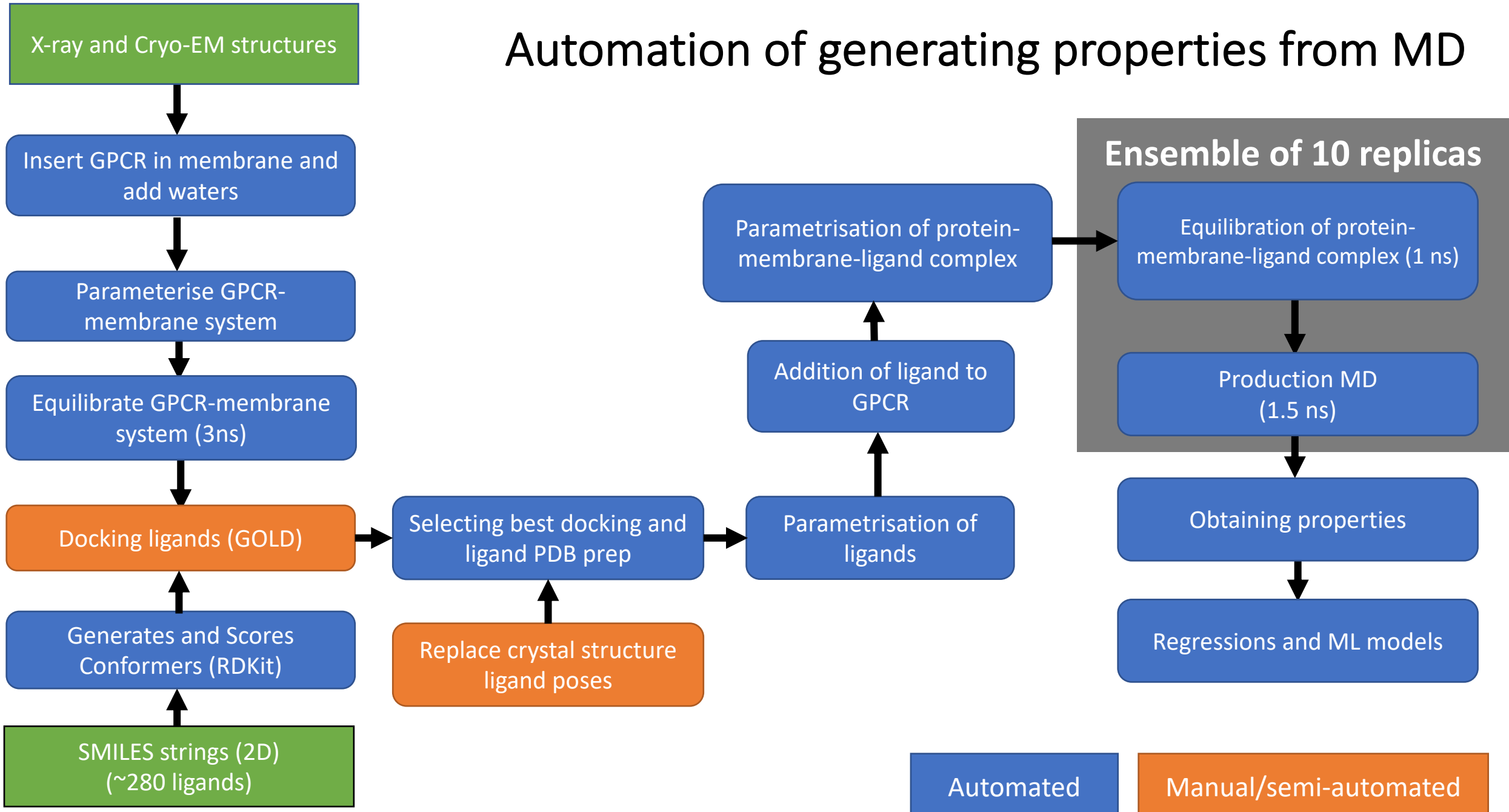
Correlations in data

- No correlation between logP, TPSA, number of hydrogen bond capable atoms and RTs of GPCR ligands
- The only two 2D properties that have any sort of weak correlation are: molecular weight and number of rings/aromatic rings

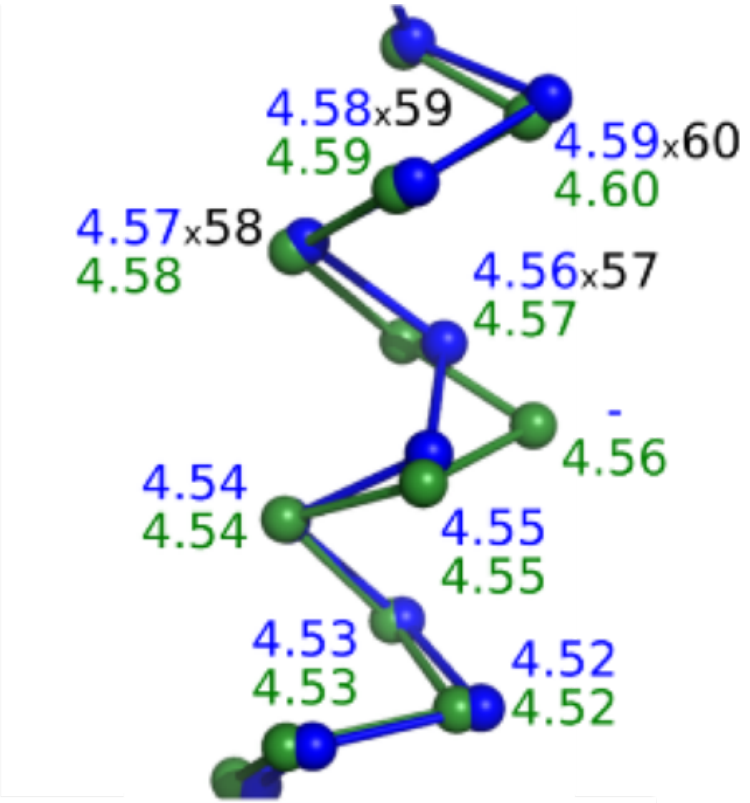




Automation of generating properties from MD



COMBINE analysis overfits to training data



MSE train: 0.36
MSE test: 0.75
R^2 train: 0.55
R^2 test: 0.01

Index	Elec_6.51	Elec_6.55	Elec_6.56	Elec_7.31	Elec_7.38
5HT2A_[3H…	-0.335378	-3.27144	0.0199144	0.532463	-0.586971
5HT2B_Cab…	-0.002424…	-1.64923	0.0683889	-0.137797	-0.95084
5HT2B_Dih…	0.143263	-1.90207	-0.0405811	-0.499482	0.468004
5HT2B_Erg…	-0.774223	-0.836446	0.118652	-6.1999	-0.203607
5HT2B_Per…	0.0374144	-1.45491	0.0412556	-3.70494	-0.891994
5HT2B_Ser…	0.0931056	-1.95418	-0.0554067	0.0308267	-0.469427
5HT2B_[12…	-0.211303	-1.99219	-0.0156267	-0.0710678	0.20127
5HT2B_[3H…	-0.634743	-2.30593	-0.006783…	-0.223672	0.24026
A1_13-Dua…	-1.14945	-8.50936	0.119008	-0.164228	0.0527778
A1_19-Dua…	-1.49955	-12.8792	0.298646	0.0953889	-0.200359
A1_22-Dua…	-1.56403	-12.6211	0.233458	-0.01549	-0.0924344
A1_33-Dua…	-1.43365	-12.5266	0.188422	-0.179567	0.203439
A1_35-Dua…	-1.5789	-12.3627	0.207316	-0.167307	0.13924

Combination of molecular properties predicts RT

- Gradient boosting regressor
- 20% test data
- 100 random training/test split

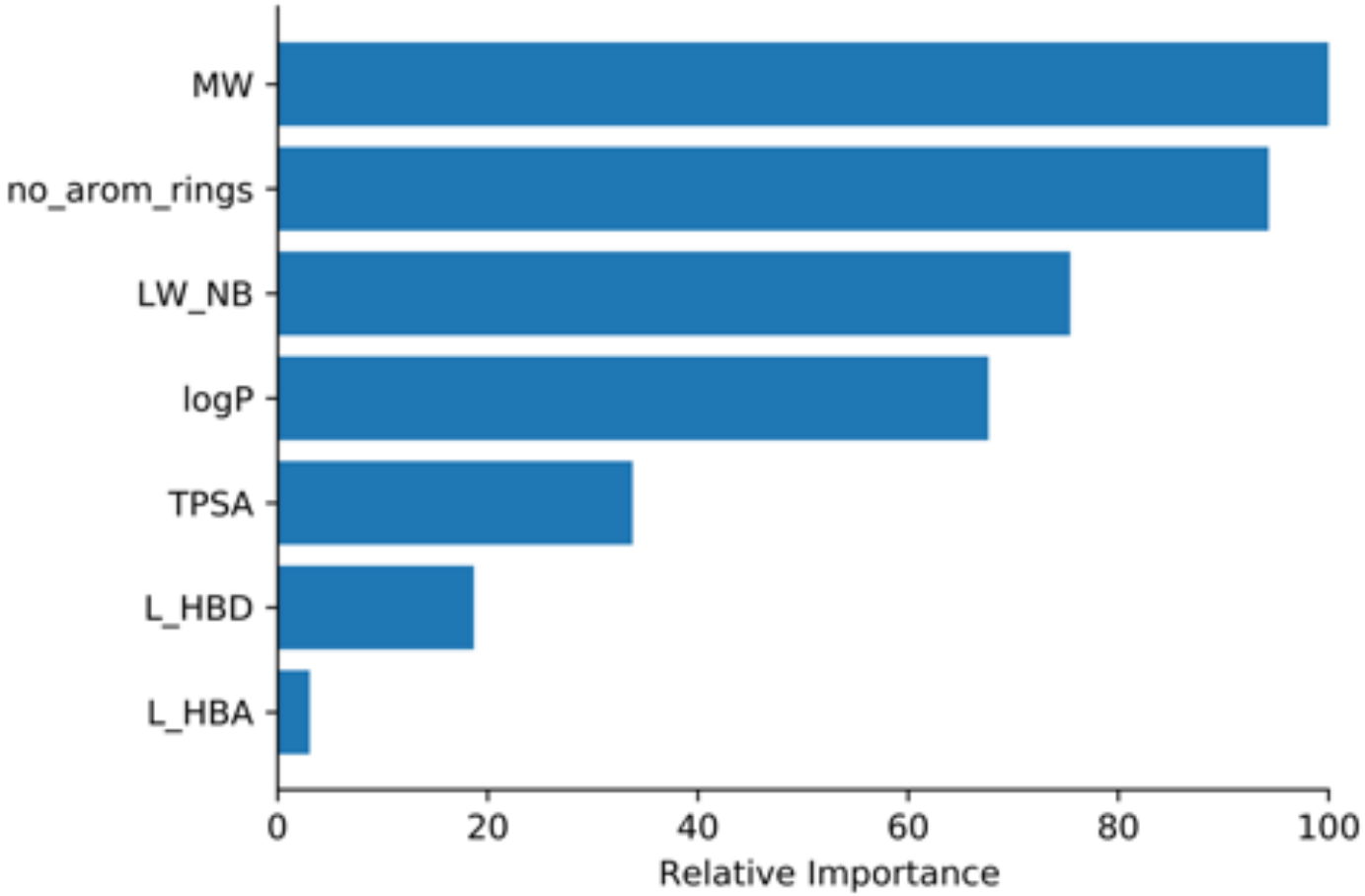
<u>Model Results:</u>

MSE train 0.28

MSE test 0.47

R^2 train 0.65

R^2 test 0.38



Next steps

- Adding more features:
  - ECL2 flexibility
  - Binding site flexibility

  - Adding spare matrices of experiment data to see if this improves prediction scores

Acknowledgments

Compute resources used:

SURF SARA

UCL

Professor Andrea Townsend-Nicholson

Dr Alexander Heifetz

Professor Peter Coveney

Email: andrew.potterton.13@ucl.ac.uk

LIDo London Interdisciplinary Doctoral Programme @ UCL, KCL, QMUL, Birkbeck, LSHTM, RVC

CompBioMed

BBSRC bioscience for the future

UCL

evotec