Guiding Anti-malarial Optimisation using Deep Learning Imputation and Compound Generation

ACS Fall National Meeting – August 24th, 2022

Benedict Irwin, Alexander Wade, Thomas Whitehead, Dr Matthew Segall

Overview

- Importance of antimalaria drug discovery
- Methods
  - Deep learning imputation
  - Med. chem. transformations for idea generation
- Application to anti-malaria optimisation – collaboration with Open Source Malaria
  - Chemistry and data
  - Model building and validation
  - Idea generation
  - Results
- Conclusions

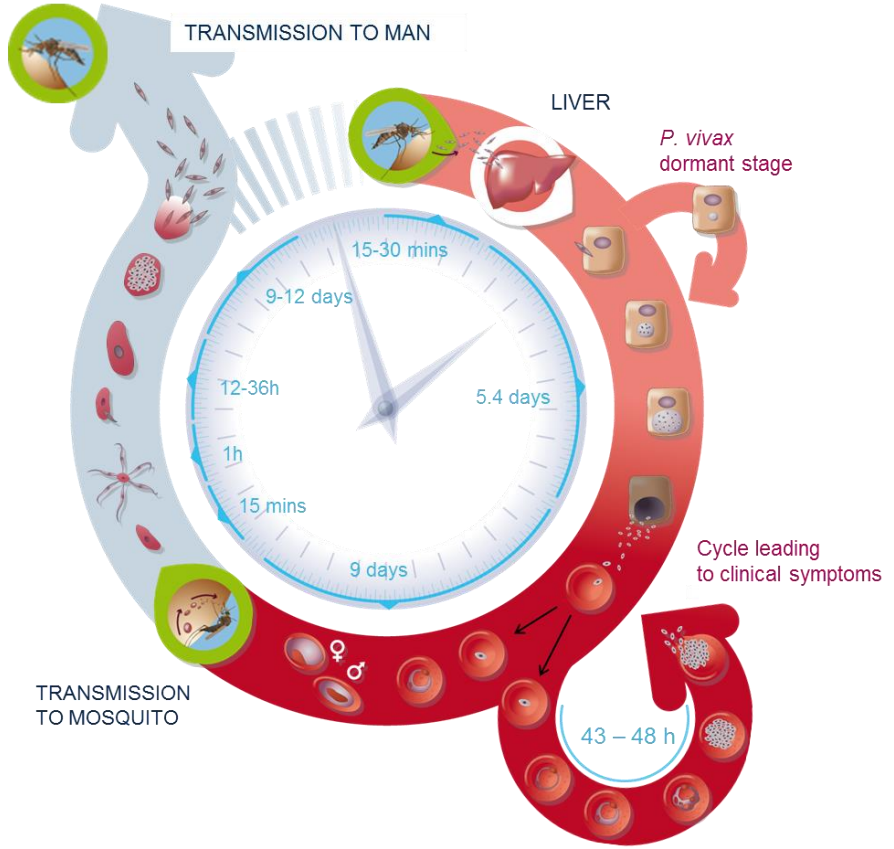
ACS Cent. Sci. **2016**, *2*, 687–701

optibrium™

Importance of Antimalaria Drug Discovery

- Infectious disease caused by *Plasmodium* parasites, mainly *P. falciparum* and *P. vivax*
- Transmitted through *Anopheles* mosquito
- ~212 million cases of malaria occurred worldwide in 2015, 429 000 mortalities*
- Global target: 90% reduction in incidence and mortality by 2030 compared to 2015
- Drug treatment threatened by emergence of resistance to artemisinin-based combination therapy in South-East Asia
- Novel drugs needed for treatment and transmission-blocking

MMV Medicines for Malaria Venture



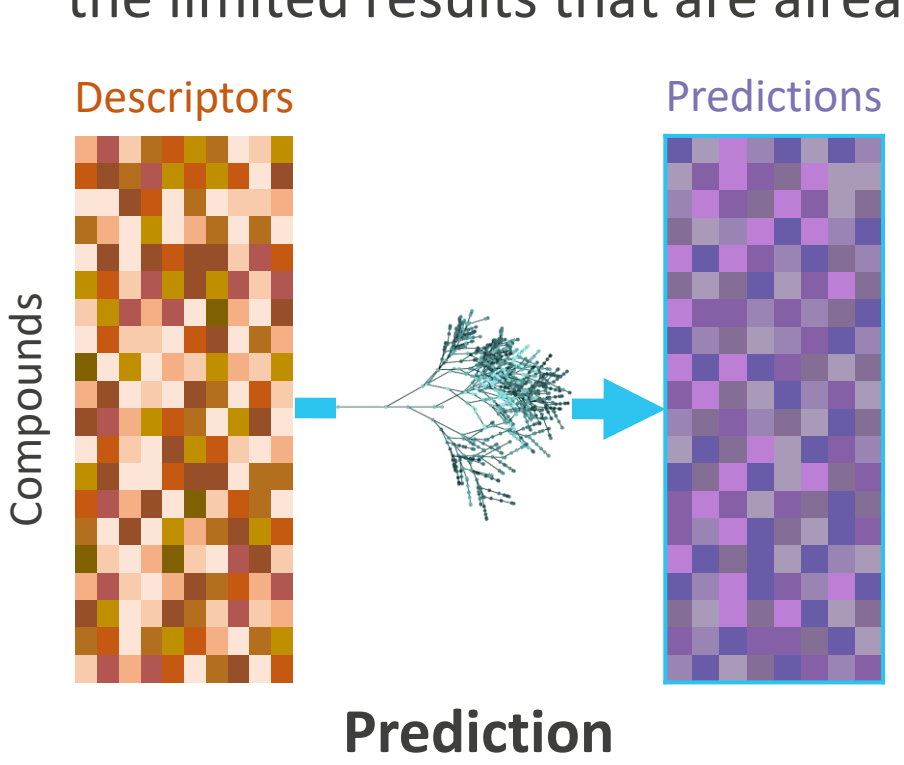
* World Health Organisation 2016 World Malaria Report

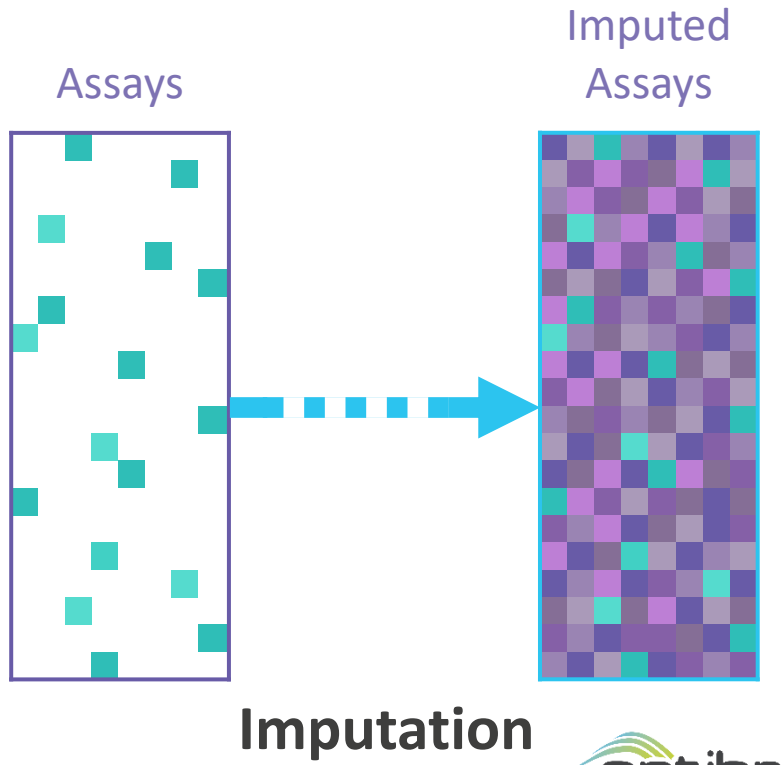
Introduction to Deep Learning Imputation

Prediction vs. Imputation

- Prediction uses input ‘features’ to predict one or more property values for a compound, e.g. QSAR models

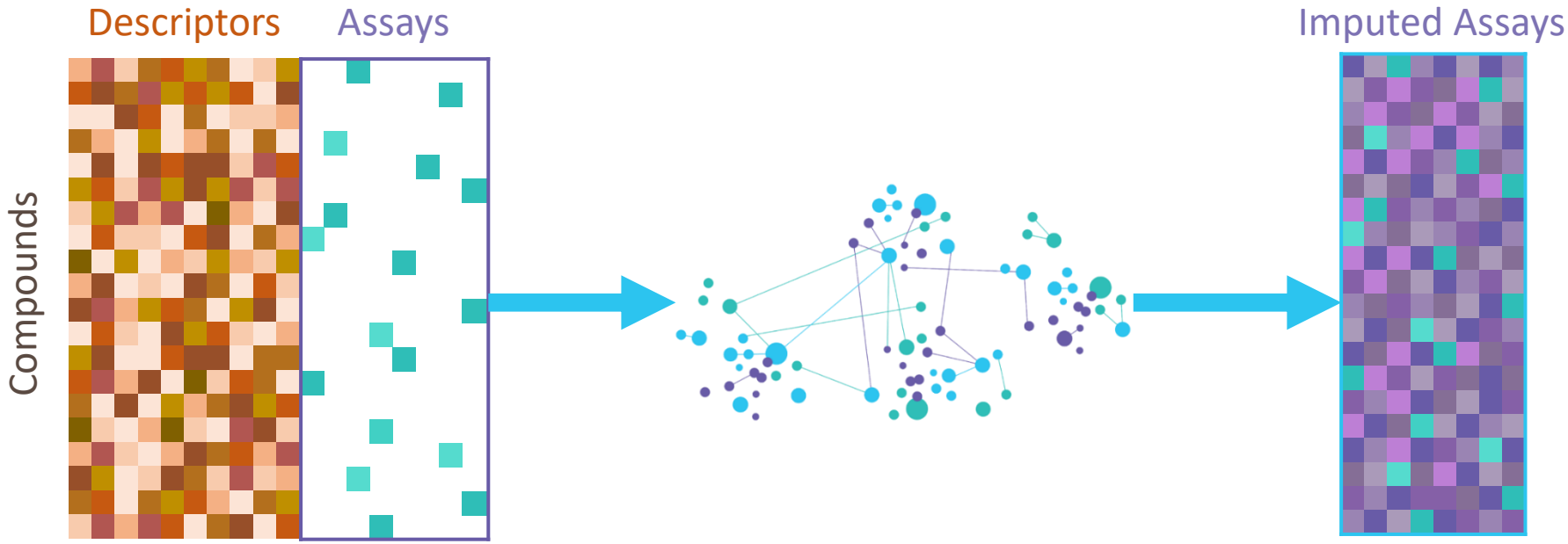
- Imputation is the process of filling in the gaps in sparse experimental data using the limited results that are already available





Alchemite™ Deep Learning Imputation
Optibrium's exclusive partnership with Intellegens

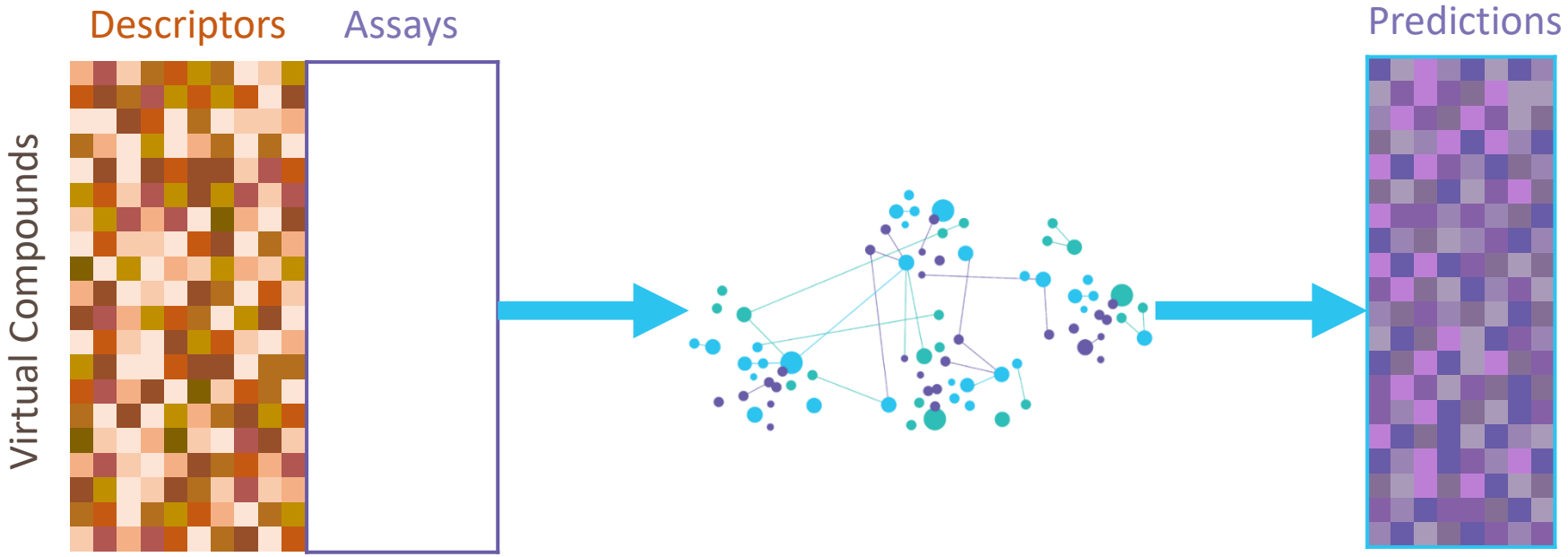
- Learns directly from relationships between experimental endpoints as well as SAR
  - Makes better use of sparse and noisy experimental data than conventional QSAR models
- 'Fills in' the gaps in your data and makes predictions for 'virtual' compounds
  - Generates more accurate predictions to target high-quality compounds



Whitehead *et al.* J. Chem Inf. Model. (2019) **59**(3) pp. 1197-1204, Irwin *et al.* J. Chem. Inf Model. (2020) **60**(6), pp. 2848–2857

Alchemite™ Deep Learning Imputation
Optibrium's exclusive partnership with Intellegens

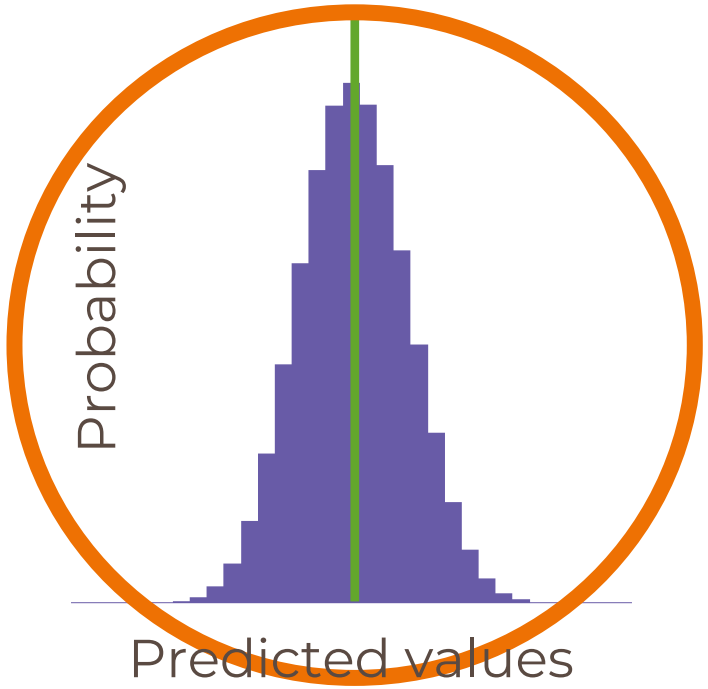
- Learns directly from relationships between experimental endpoints as well as SAR
  - Makes better use of sparse and noisy experimental data than conventional QSAR models
- 'Fills in' the gaps in your data and makes predictions for 'virtual' compounds
  - Generates more accurate predictions to target high-quality compounds

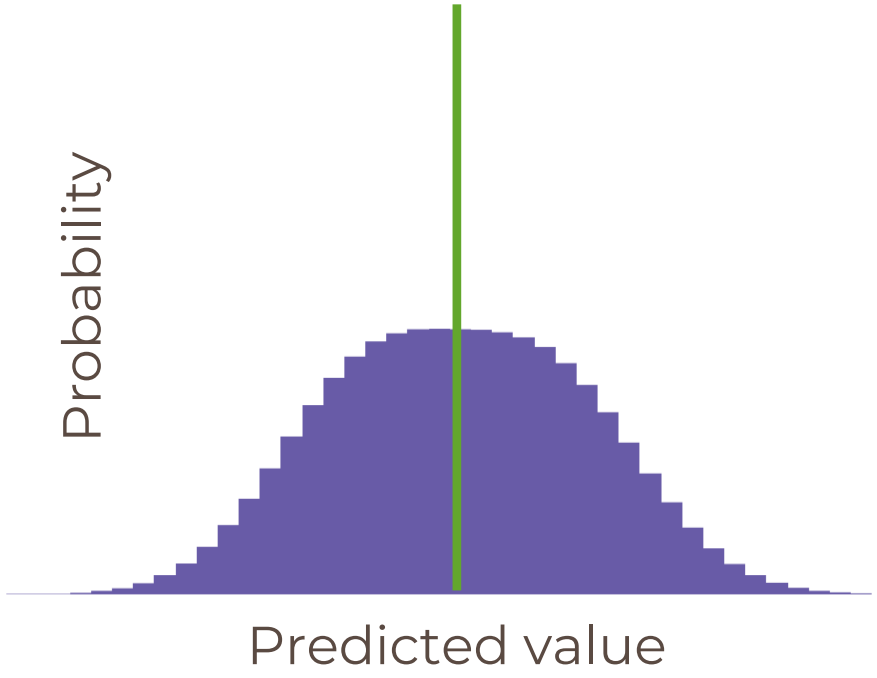


Whitehead *et al.* J. Chem Inf. Model. (2019) **59**(3) pp. 1197-1204, Irwin *et al.* J. Chem. Inf Model. (2020) **60**(6), pp. 2848–2857

Alchemite™ Deep Learning Imputation
Optibrium's exclusive partnership with Intellegens

- Estimates uncertainty in each individual prediction
  - Strong correlation between uncertainty estimates and observed accuracy on independent test sets
  - Highlights the most accurate predictions on which to base decisions
- Confidently targets high-quality compounds and prioritise experimental resources





Whitehead *et al.* J. Chem Inf. Model. (2019) **59**(3) pp. 1197-1204, Irwin *et al.* J. Chem. Inf Model. (2020) **60**(6), pp. 2848–2857

optibrium™

Introduction to Med. Chem. Transformations for Idea Generation

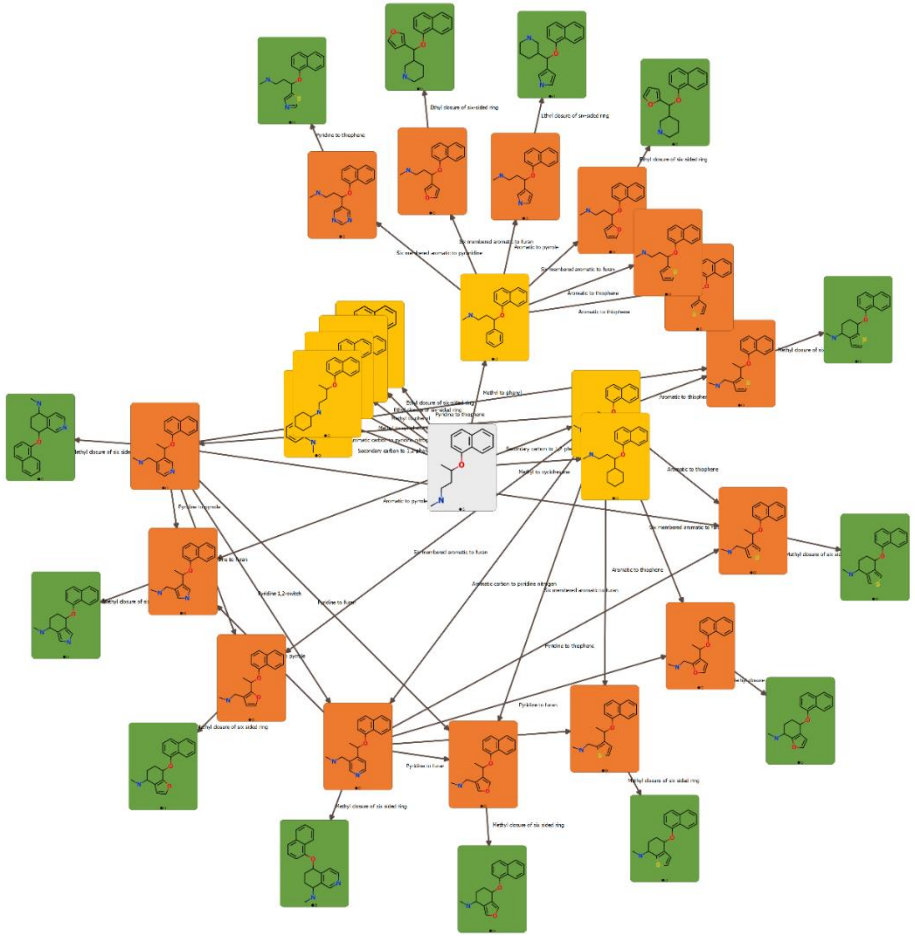
Chemistry Transformations

- Compounds must **make sense** from a medicinal chemistry perspective
- Apply ‘transformation rules’, derived from medicinal chemistry experience, to initial compound
  - Not only functional group replacement but also framework transformations
  - Library of >29,000 transformations – Including BIOSTER™ database
- Encode new rules to capture knowledge
  - Personal experience or in-house chemistries
  - Define groups of transformations tailored to specific objectives
- Apply iteratively to produce multiple ‘generations’

M.D. Segall *et al.* J. Chem. Inf. Model. (2011) **51** pp. 2967–2976. K.D. Stewart *et. al.* Bioorg. Med. Chem. **14** (2006) p. 7011

Expand Around a Compound or Chemical Series

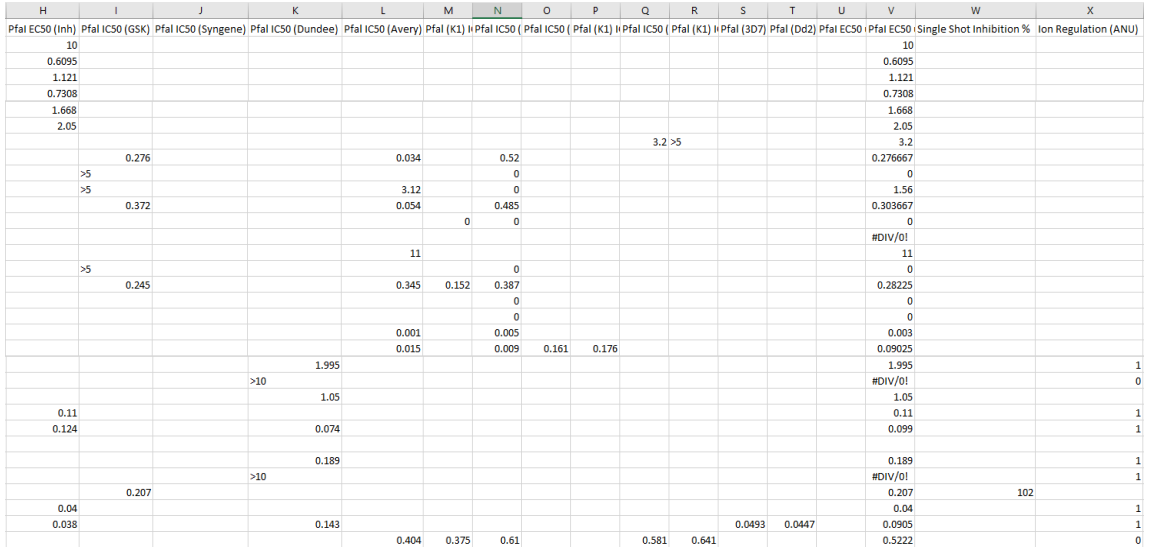
Potentially Exponential Growth



Application to Anti-Malarial Optimisation

Data and Modelling

- Data set of sparse antimalaria activity data
  - Different assays, different strains, different labs
- Competition run by OSM
  - Models from different organisations compared with blind test set
- Performed a stratified 80:20 split of the data into training and validation
- Descriptors
  - 330 sub-structural fragment and whole molecule properties
  - MW, logP, TPSA, V_x



H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
Pfal EC50 (Inh)	Pfal IC50 (GSK)	Pfal IC50 (Syngene)	Pfal IC50 (Dundee)	Pfal IC50 (Avery)	Pfal (K1) I	Pfal IC50 (	Pfal IC50 (	Pfal (K1) I	Pfal IC50 (	Pfal (K1) I	Pfal (3D7)	Pfal (Dd2)	Pfal EC50 (	Pfal EC50 (	Single Shot Inhibition %	Ion Regulation (ANU)
10														10		
0.6095														0.6095		
1.121														1.121		
0.7308														0.7308		
1.668														1.668		
2.05														2.05		
									3.2	>5				3.2		
	0.276			0.034		0.52								0.276667		
	>5					0								0		
	>5			3.12		0								1.56		
	0.372			0.054		0.485								0.303667		
					0	0								0		
														#DIV/0!		
				11										11		
	>5					0								0		
	0.245			0.345	0.152	0.387								0.28225		
						0								0		
						0								0		
				0.001		0.005								0.003		
				0.015		0.009	0.161	0.176						0.09025		
			1.995											1.995		1
			>10											#DIV/0!		0
			1.05											1.05		
0.11														0.11		1
0.124			0.074											0.099		1
			0.189											0.189		1
			>10											#DIV/0!		1
	0.207													0.207	102	
0.04														0.04		1
0.038			0.143								0.0493	0.0447		0.0905		1
				0.404	0.375	0.61			0.581	0.641				0.5222		0

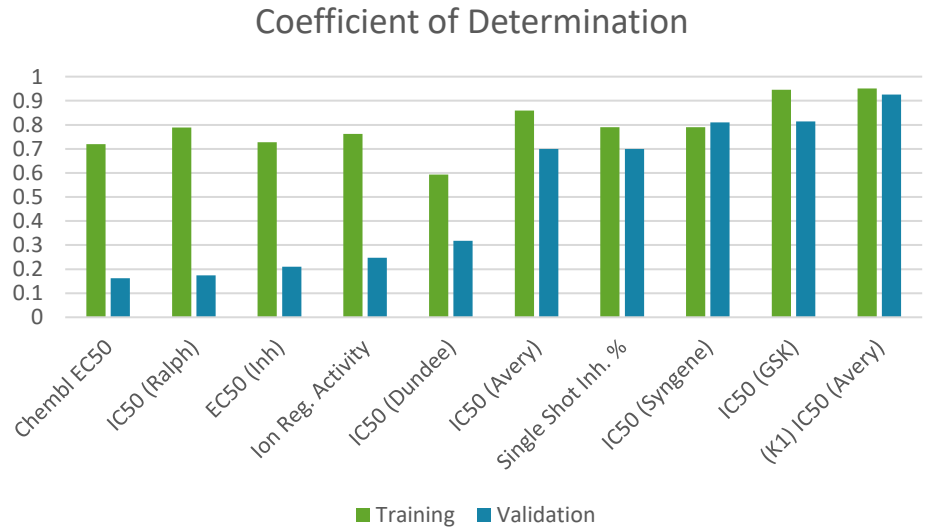
Assay	Min Value	Max Value
pEC50 ChEMBL	4.5	8.2
pEC50 (Inh)	4.7	7.8
(3D7) pIC50 (Broad)	4.7	7.3
(Dd2) pIC50 (Broad)	4.7	7.4
(K1) pIC50 (Avery)	4.5	7.2
(K1) pIC50 (Guy)	5.0	6.9
pIC50 (Avery)	4.1	9.0
pIC50 (Dundee)	4.0	7.3
pIC50 (GSK)	5.0	7.6
pIC50 (Guy)	5.5	7.0
pIC50 (Ralph)	5.7	8.3
pIC50 (Syngene)	4.0	7.2
Single Shot Inhibition %	0%	100%
Ion Regulation Activity	0	1

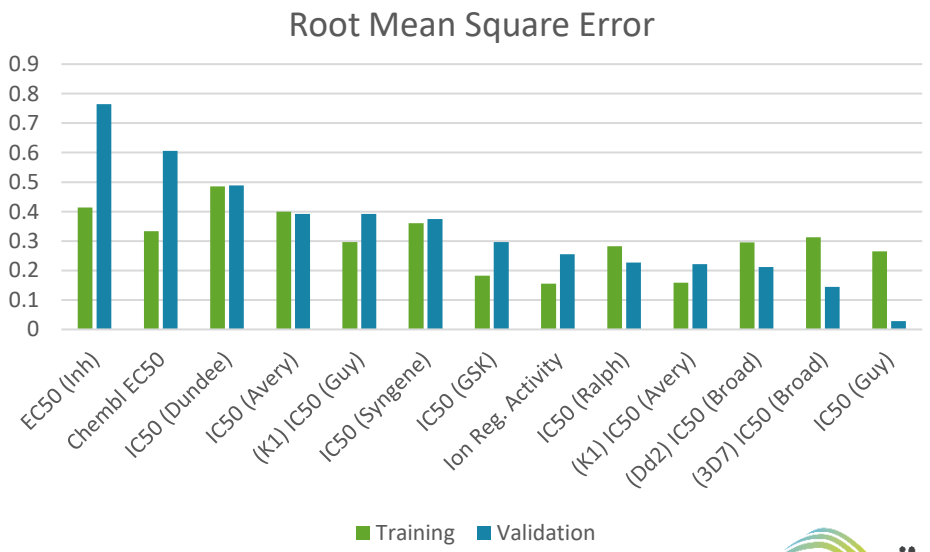
Watch our webinar http://bit.ly/ai_antimalarials

Tse *et al*. J. Med. Chem. (2021) **64**(22) pp 1645-16463

Validation Results – Imputation Model

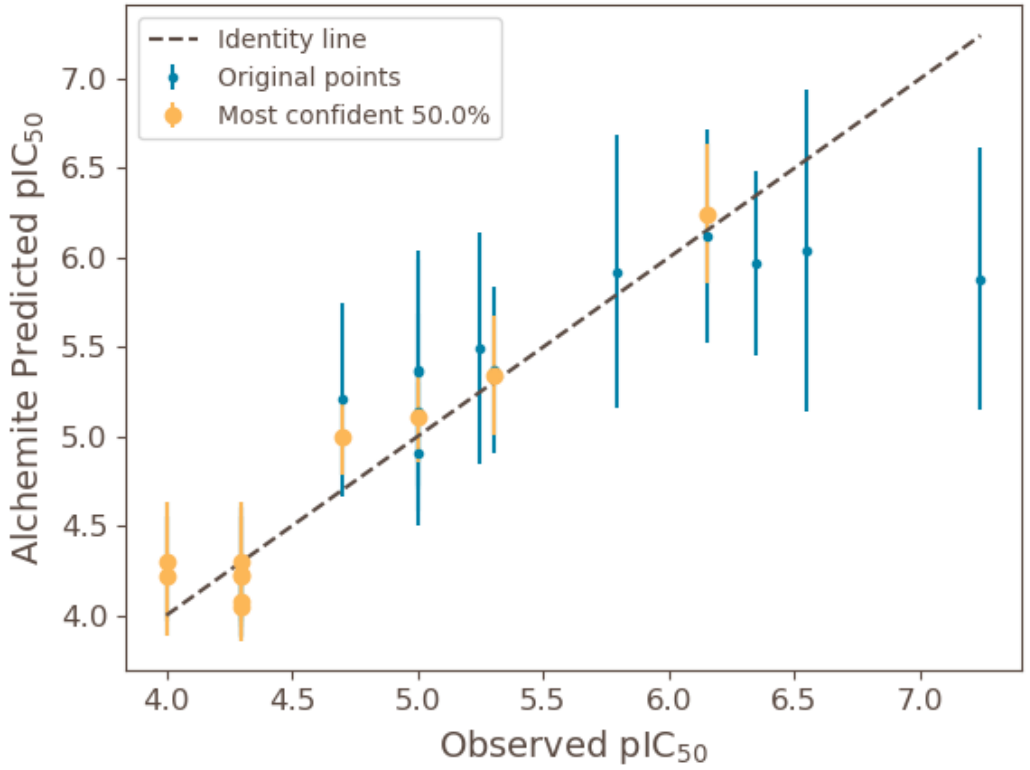
- Performance metrics
  - Coefficient of determination (R^2)
    - Higher is better
  - Root mean square error (RMSE)
    - Lower is better
- Performance on half of the endpoints is very good

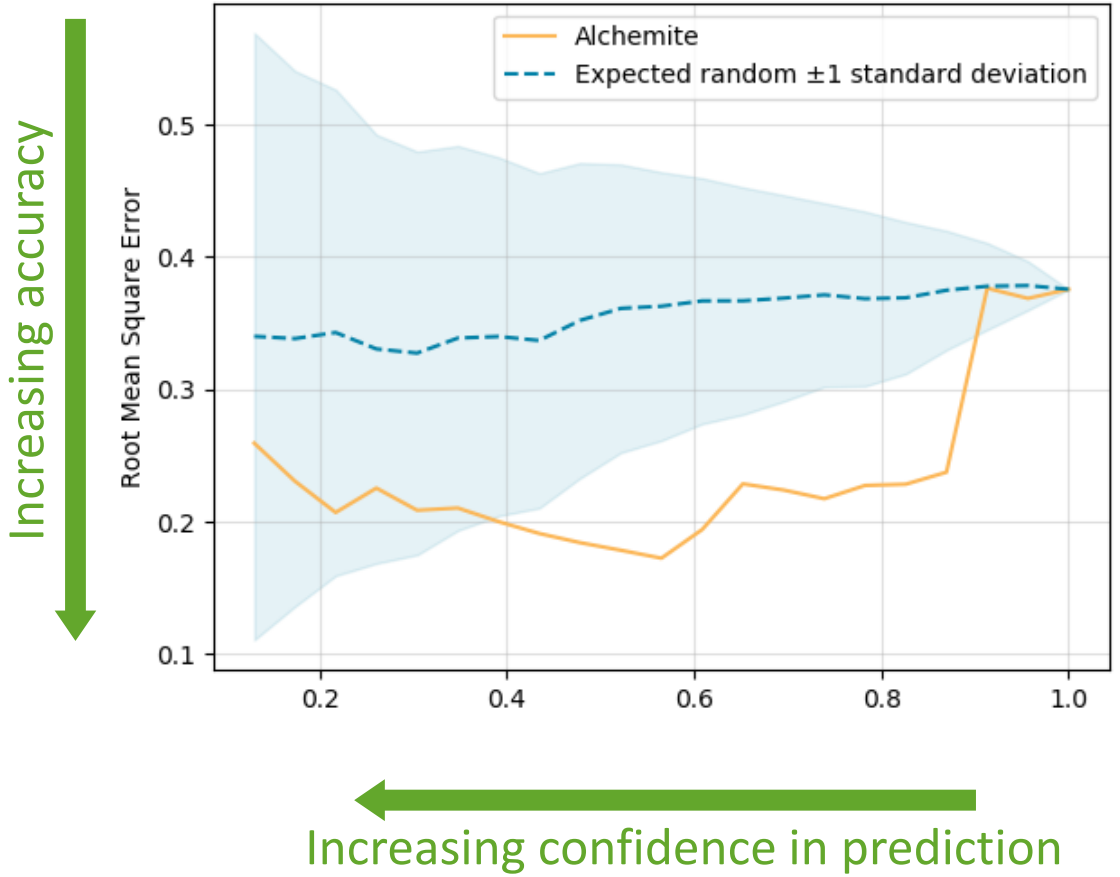




Watch our webinar http://bit.ly/ai_antimalarials
Tse *et al*. J. Med. Chem. (2021) **64**(22) pp 1645-16463

Uncertainty Analysis – e.g. Syngene Assay Validation





- Excellent correlation between model confidence (error bars) and observed accuracy

Watch our webinar http://bit.ly/ai_antimalarials Tse *et al*. J. Med. Chem. (2021) **64**(22) pp 1645-16463

Modelling Results – Comparison of Submissions on Blind Test Set

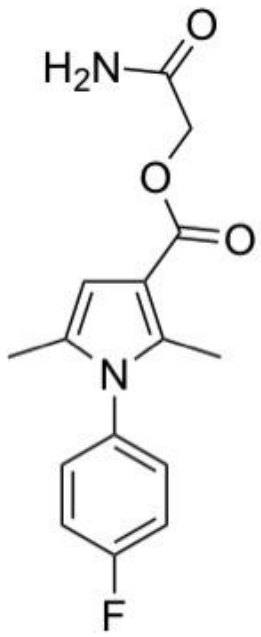
	Entrant (Affiliation)	Description of Model	Precision of Accurate Predictions (Active and Inactive)	Result
	Jonathan Cardoso-Silva (KCL)		36%	Runner-up
Private sector	Giovanni Cincilla (Molomics)	Logistic regression classifier model using a stochastic average gradient as solver, a uniform regularisation and a learning step size = 0.01.	91%	**Winner (company)**
Private sector	Mykola Galushka (Auromind)		58%	Runner-up
	Davy Guan (USyd)	Automated machine learning method to optimise QSAR models for the lowest Mean Absolute Error.	82%	**Winner (non-company)**
Private sector	Ben Irwin/Mario Öeren/Tom Whitehead (Optibrium/Intellegens)	Deep imputation [Whitehead2019] with quantum mechanical StarDrop6.6 Automodeller and pKa descriptors [Hunt2020].	81%	**Second place**
	Raymond Lui (USyd)		58%	Runner-up
	Slade Matthews (USyd)	Random forest model using 200 Mordred descriptors based on optimised 3D structures. Training RMSE = 0.805.		Runner-up
	Ho-Leung Ng (KSU)	QSAR model based on detailed homology modeling of PfATP4 and docking. 3D features are combined with 1D/2D QSAR features using XGBoost (gradient boosted trees) to make a regression model.	71%	Runner-up
Private sector	Vito Spadavecchio (Interlinked TX)		36%	Runner-up
Private sector	Laksh Aithani/Willem van Hoorn (Exscientia)	Ridge regression model with alpha = 1. ECFP4 fingerprints with (Morgan radius 2) were the input to the model.	81%	**Second place**

Progressed to compound idea generation

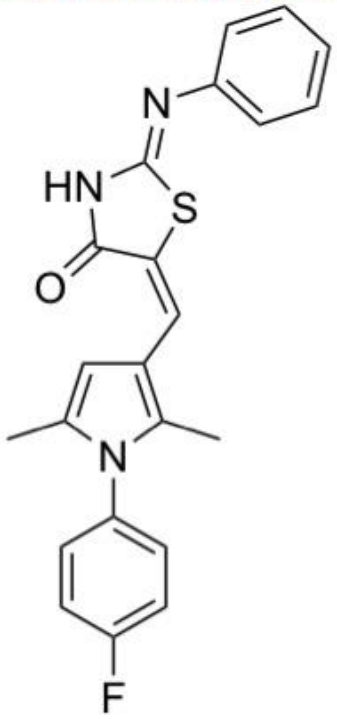
Watch our webinar http://bit.ly/ai_antimalarials Tse *et al*. J. Med. Chem. (2021) **64**(22) pp 1645-16463

Starting Point – Open Source Malaria ‘Series 4’

GSK TCAMS

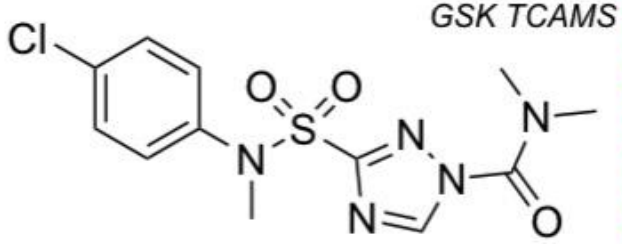


SERIES 1
Potent
Ester Problematic



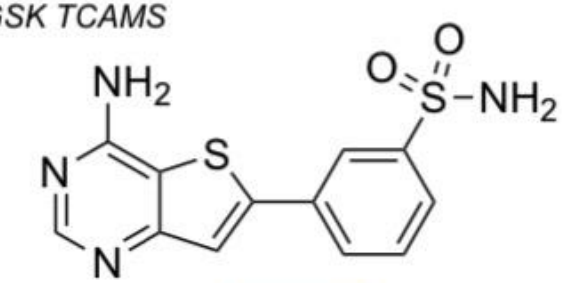
SERIES 1a
Potent
Gametocyte active
Low solubility

GSK TCAMS



SERIES 2
Potent
Duplication with Closed Group

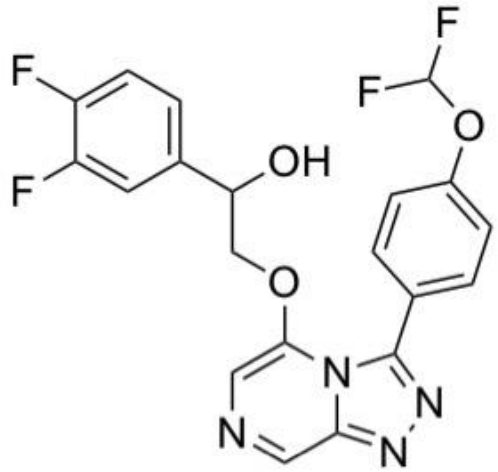
GSK TCAMS



SERIES 3
Potent
Excellent physicochem
Novel MoA: ASN Ligase or CA

Our focus

Pfizer



SERIES 4
Potent, including *in vivo*
Extensive SAR
Probable *Pf*ATP4 MoA
Requires improvement for Late Lead

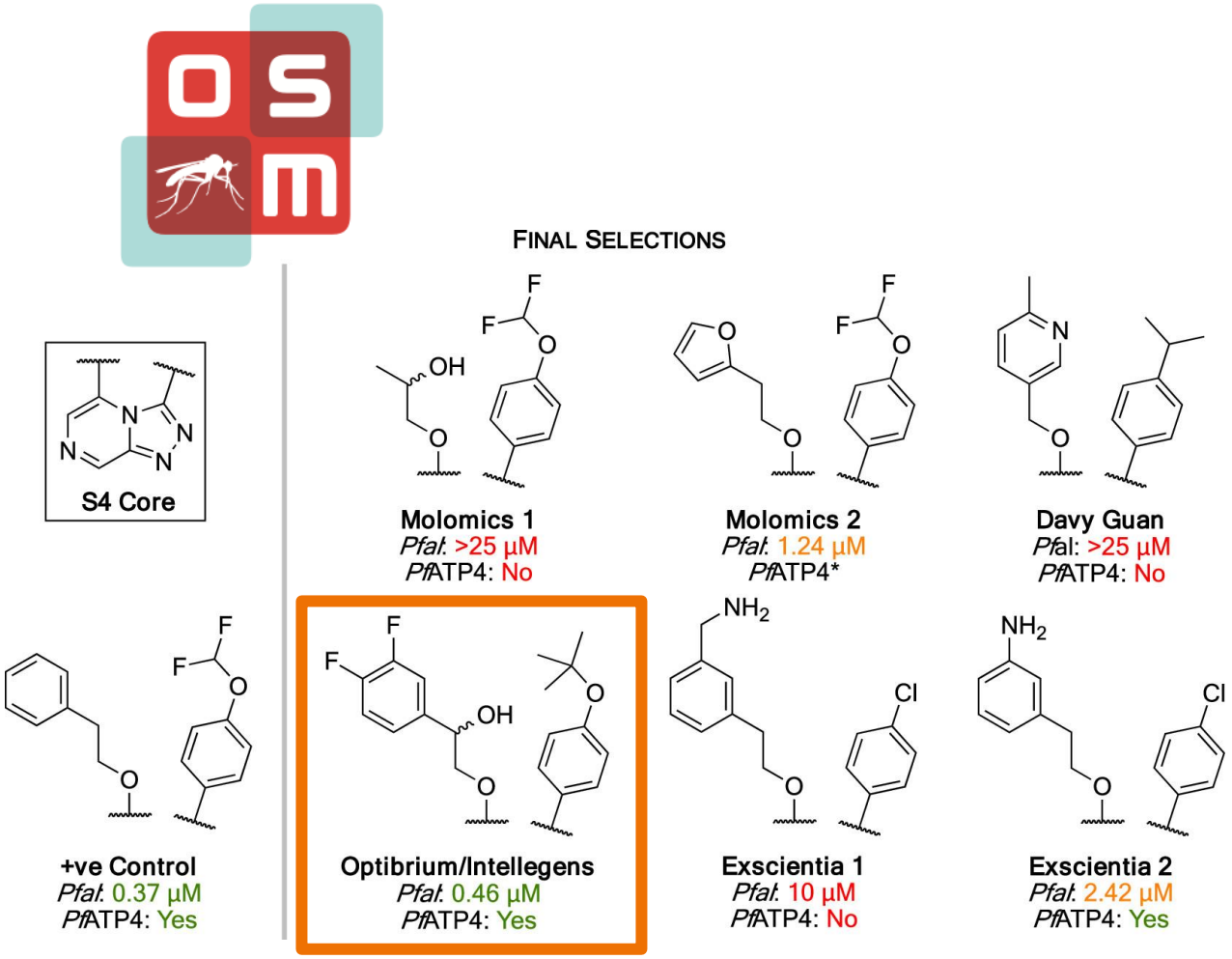
Series typically not exhausted - Decision points made to move to another.

Anyone free to employ OSM infrastructure to explore these, or other, series

www.opensourcemalaria.org, @O_S_M

New Compound Generation - Results

- A **confidently** predicted compound was synthesised and tested by OSM
  - **Only confirmed active** of those proposed
- "The Optibrium/Intellegens suggestion… was thought by the human team to be a certain inactive … yet this compound displayed good potency and is a particularly useful outcome (i.e., **the "Machine Overlords" class**)"



Watch our webinar http://bit.ly/ai_antimalarials

Tse *et al*. J. Med. Chem. (2021) **64**(22) pp 1645-16463

Conclusions

- We applied a combination of deep learning imputation and generative chemistry methods to explore optimisation strategies around an anti-malarial series
- By leveraging sparse data from multiple assays we built a model with good accuracy across several activity endpoints
  - The uncertainty estimates in each prediction correlated strongly with the observed accuracies
  - This enabled us to focus on novel compounds with the highest confidence in achieving activity
- The proposed compound was experimentally confirmed to be active and revealed a new optimisation approach for the chemical series
- For more information: matt@optibrium.com
  - Tse *et al*. J. Med. Chem. (2021) **64**(22) pp 1645-16463
  - Watch our webinar http://bit.ly/ai_antimalarials

optibrium™

Acknolwedgements

- Prof. Matthew Todd, Dr Edwin Tse and the rest of the Open Source Malaria team
- Optibrium
  - Benedict Irwin (now at MS Research)
  - Mario Öeren
  - Alexander Wade (University of Cambridge)
- Intellegens
  - Gareth Conduit
  - Tom Whitehead
- Everyone from the other organisations participating in the blind validation