A-WOL: Ligand Based Screening combined with HTS

J. Bibby[1], Neil G. Berry*[1], P. O'Neill[1], S. Ward[2], M. Taylor[2]
[1]Department of Chemistry, Crown Street, University of Liverpool, L69 7ZD, UK
Telephone: 44(0)151 794 3877, Fax: 44(0)151 794 3588, E-mail: ngberry@liv.ac.uk
[2]Liverpool School of Tropical Medicine, Pembroke Place, Liverpool L3 5QA, UK

Introduction

- Filariasis inflicts serious health problems throughout tropical communities.
- Lymphatic filariasis (Elephantiasis), and onchocerciasis (River blindness), together infect more than 150 million people[1].
- River Blindness is the second leading cause of blindness caused by infection.
  - 500,000 people blind due to River Blindness.
- 120 million people are currently infected by Lymphatic filariasis.
  - 40 million people disfigured, the second leading cause of global disability.
- Filarial species that infect people co-exist in mutualistic symbiosis with *Wolbachia* bacteria, which are essential for growth.
- Antibiotic anti-*Wolbachia* (A-WOL) therapy delivers safe macrofilaricidal activity with superior therapeutic outcomes compared to all standard anti-filarial treatments.
- The A·WOL targeted screening aims to identify novel hits with significant improvements in anti-*Wolbachia* activity.

Chemoinformatics

Similarity searching and machine learning methods have been employed in iterative cycles to select compound plates for screening from the MMV library. The results of each screening are fed back into the modelling to give improvement in the results of subsequent rounds of HTS.

Diversity and First Selection and Screen

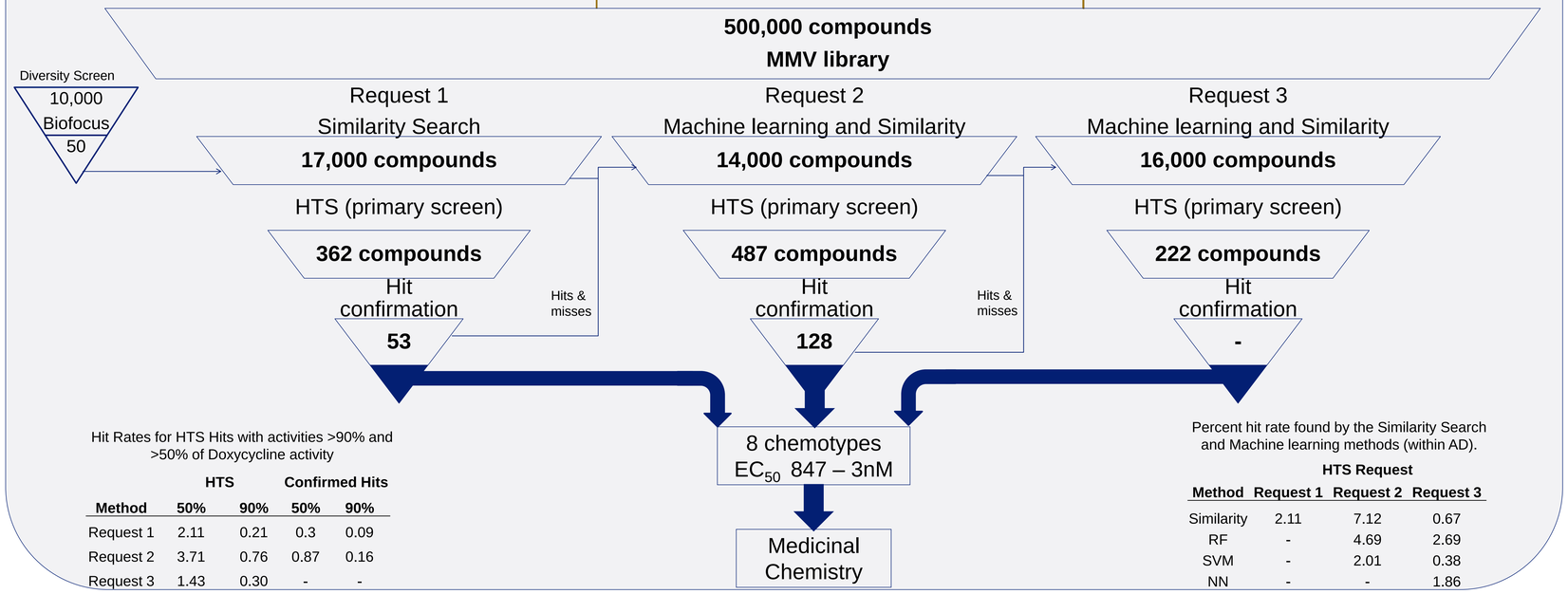
- Diverse set of 10,000 compounds selected from the Biofocus library.
- Screening identified 50 compounds with A-WOL activity.
- Similarity search based on the 50 actives (Request 1)
- Selected 17,000 compounds, 362 HTS hits.

Second Compound Selection and Screen

- Use hits and misses from Request 1.
- Range of computational modelling methods
  - Similarity Search
  - Random Forests
  - Support Vector Machines
- Trained on the data using:
  - 15 different physicochemical descriptors (e.g. LogP, number of rotable bonds, H bond donors.
  - fingerprint descriptors (ECFP4, FCFP4 and MDL MACCS)

Third Compound Selection and Screen

- The Results were again fed back into a third Request
- Neural Networks used in addition to RFs and SVMs
- Focus on hit diversity, finding novel chemotypes.
- Out of Application Domain predictions included.



Hit Rates for HTS Hits with activities >90% and >50% of Doxycycline activity

	HTS		Confirmed Hits	
Method	**50%**	**90%**	**50%**	**90%**
Request 1	2.11	0.21	0.3	0.09
Request 2	3.71	0.76	0.87	0.16
Request 3	1.43	0.30	-	-

Percent hit rate found by the Similarity Search and Machine learning methods (within AD).

	HTS Request		
Method	**Request 1**	**Request 2**	**Request 3**
Similarity	2.11	7.12	0.67
RF	-	4.69	2.69
SVM	-	2.01	0.38
NN	-	-	1.86

Conclusions and Future Work

The results demonstrate the effectiveness of ligand based methods in combination with HTS for the identification of novel antibacterial compounds.

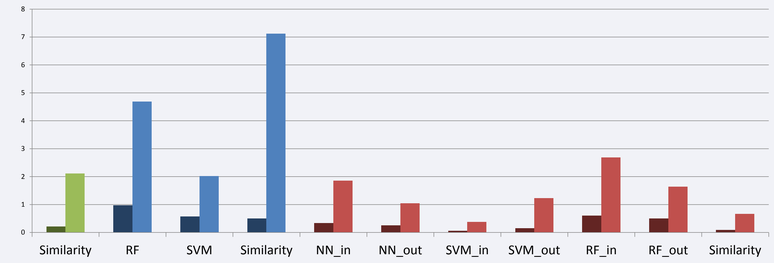
- 60,000 compounds have been screened so far.
- Iterative rounds of screening and modelling have shown an improvement at each stage.
  - either in terms of hit rates, or in the potency or diversity of hits found.
- 8 distinct tractable chemotypes identified.
- Activities in the nM range.
- Currently undergoing medicinal chemistry optimisation.
- Continuation of iterative screening.

This study successfully identified novel distinct chemotypes with potent A-WOL activity

Analysis of Methods

Machine Learning

- **Random Forest** models greatly increase the hit rate of potent hits.
  - This increase in hit rate was due to the expansion around existing chemotypes, with little detection of new chemotypes.
- **Support Vector Machines** models showed an overall poorer hit rate.
  - The hits detected were found in new areas of chemical space.
- **Neural Networks** combine hit rate with novel hit detection.



Hit Rates for potent hits (dark) and moderate hits (light) for the Similarity search and Machine Learning methods used in successive screens (Request 1: green; Request 2: blue; Request 3: red). With models applied within (in) or without (out) the AD.

The Applicability Domain (AD) is the chemical space on which the models were trained, and where the models can be reliably applied.

- In Request 2 only compounds which were within a models' AD were considered.
- In Request 3 a comparison was made between predictions within and without of the AD.
- As expected, there was a small reduction in total hit rate.
- The outside domain predictions selected an increased diversity of hits.

Data Fusion

- Each machine learning model produces a score for each compound, this is insufficient for the selection of plates for screening.
- Five methods of data fusion were used to combine scores of multiple models into a score that represents the likelihood of finding hits on that plate.
- Plates were ranked by this score with the highest scoring plates selected for screening in the second screen.

Hit Rates for HTS Hits with activities >90% and >50% of Doxycycline activity

		HTS Hits as a percentage	
	Method	**90%**	**50%**
Request 2	Reciprocal Rank	0.61	4.09
	Rank	0.46	10.27
	Sum	0.46	10.27
Request 3	Sum	0.26	0.93
	Parallel	0	0.26
	Z2	0.15	1.24

Acknowledgements
This work was supported by a grant from the Bill and Melinda Gates Foundation awarded to the Liverpool School of Tropical Medicine as part of the Anti-Wolbachia consortium.

References
1. Taylor, M. J.; Hoerauf, A.; Bockarie, M. Lymphatic filariasis and onchocerciasis. *Lancet***2010**. 376, 1175-1185.