Eve: Integration of machine learning with compound testing in a Robot Scientist

K. Williams[1], E. Bilsland[2,3], A. Sparkes[1], W. Aubrey[1], M. Young[4], L. N. Soldatova[5], K. De Grave[5], J. Ramon[6], M. de Clare[2], W. Sirawaraporn[7], K. M. Roper[8], S. G. Oliver[2] and R. D. King[8]

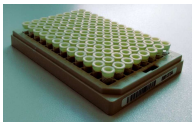
1. Department of Computer Science, Aberystwyth University
2. Cambridge Systems Biology Centre and Department of Biochemistry, University of Cambridge
3. Department of Structural and Functional Biology, UNICAMP
4. Institute of Biological, Environmental and Rural Sciences, Aberystwyth University
5. School of Computer Science, Brunel University
6. Department of Computer Science, KU Leuven
7. Department of Biochemistry, Mahidol University
8. Manchester Institute of Biotechnology and School of Computer Science, University of Manchester

Why Develop Robot Scientists?

- **Building a Robot Scientist forces us to make concrete engineering decisions involving the relation between abstract and physical objects, and between observed and theoretical phenomena, as well as the ways hypotheses are created.**
- **Some scientific problems are so complex they require a vast amount of research, and there are simply not enough human scientists to do it all; automation offers our best hope for solving those problems. We can address this problem using robot scientists to combine rapidly developing technologies for both laboratory automation and data analysis.**

We have so far built and validated two Robot Scientists: Adam – a high throughput system capable of analysing thousands of yeast strains – and more recently, Eve – a system designed to test drug or siRNA libraries against a small number of yeast strains [1].

Eve's Methodology

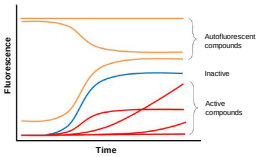


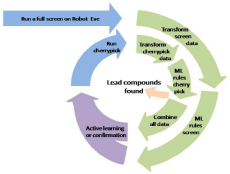
Phase 1: Mass-screening

Eve mass-screens a subset of its library to find "hit" compounds for the assay. This subset is currently chosen randomly, but could be selected using structural information about the target protein.

Phase 2: Hit-confirmation

Eve re-assays hit compounds using multiple repeats and titrations to eliminate false positives. Eve's integration of screening and hit-confirmation is similar to advanced screening systems that execute a high-throughput screen followed by high-content screening for selected compounds.





Phase 3: Automated QSAR analysis & intelligent library screening

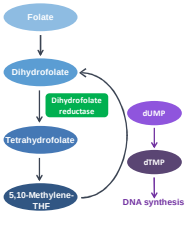
Using confirmed hits Eve executes cycles of machine learning that hypothesize QSARs, and tests them on new compounds. Eve currently lacks access to chemical synthesis automation, and therefore screens untested compounds from its library rather than synthesise new chemicals.

Case Study: Identifying TNP-760 as a Malaria Treatment

Validation of Eve focused on neglected tropical diseases including malaria, Chagas disease, leishmaniasis, schistosomiasis and African sleeping sickness [2]

- **These diseases infect hundreds of millions of people, and annually kill millions of people.**
- **The cause of these diseases is clear, and they can be treated simply by killing the parasites causing them. These criteria are not met for many diseases targeted by the pharmaceutical industry.**
- **There is little competition from the pharmaceutical industry.**

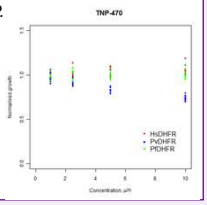
Plasmodium vivax is the most frequently occurring and widely distributed of the protozoal parasites that cause malaria, with at least 80 million infections occurring annually [3]. *P. vivax* cannot be continuously cultured in laboratory conditions, and the two main treatments, chloroquine and primaquine, have been in use for over 60 years.



Dihydrofolate reductase (DHFR) catalyses the reduction of dihydrofolate to tetrahydrofolate, an important step in the use of folates for purine synthesis. DHFR inhibitors are routinely used as prophylactics against malaria, but are no longer used as a standard treatment due to drug resistance.

TNP-470 is an angiogenesis inhibitor derived from the antimicrobial compound fumagillin, which has also been investigated as an anti-cancer drug. TNP-470 has binds to *P. falciparum* MetAP2 in vitro and inhibits growth of *P. falciparum* strains, as well as lowering parasitemia in mice [4].

Eve's results indicate that TNP-470 has high activity against *P. vivax* DHFR: the hit-confirmation curve displayed here shows an impact on *P. vivax* (blue), but not human (red) or *P. falciparum* (green) versions of the enzyme.



Enzyme inhibition assays demonstrated that *P. vivax* DHFR is 1000-fold more sensitive to TNP-470 than its human counterpart [5]. This is consistent with the results of Eve's assays and suggests that our approach identified a bona fide DHFR inhibitor with improved selectivity.

Benefits of Intelligent Screening

A Robot Scientist uses techniques from the field of artificial intelligence to carry out cycles of experimentation on a laboratory robotic system. It automatically generates hypotheses from background knowledge and models, designs physical experiments to test these hypotheses, carries out the experiments on a laboratory robotic system, and then analyses and interprets the results.

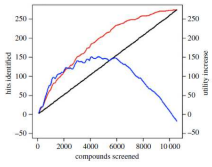
Improved data capture records:

- Goals, hypotheses, results, conclusions
- Direct experimental data
- Meta-data such as environmental conditions, detailed experiment content layout information, and instrument settings, protocols and runtime logs. These meta-data can be especially important when studying complex biological systems where the specifics of the environment can have such a large effect on results.

Intelligent screening yields increased hit rates:

In a simulation run of Eve's process of QSAR learning/testing from a compound library, we can see a clear increase in intelligent screening hit rates (red) compared to a standard brute-force approach (black).

This graph displays data simulating a screen of the Maybridge Hitfinder library against *Plasmodium vivax* DHFR. The differential utility of intelligent screening is shown in blue. It is cost-optimal to screen between a third and a half of Eve's small library, with a larger library the screened proportion would be expected to be smaller.



Eve's Equipment

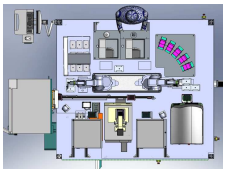
Consumables Handling:
RV-3SJB 5-axis robot
RV-3SJ 6-axis robot
Linear actuator
DC-96pro decapper
2x Orbit barcode readers
Xtr-96 barcode scanner

Liquid Transfer:
ECHO 550 acoustic transfer system
Bravo Velocity11 liquid handler
2x Multidrop 384 reagent dispensers

Storage:
2x Variomag teleshakers
6x consumable stacks
Cytomat automated drystore
Cytomat automated incubator
7x transfer stations

Measurement:
PHERAstar platereader
POLARStar Omega platereader
ImageXpress Micro cell imager

The modular nature of Eve's equipment allows use of a range of experimental formats.



Standard assay:

Compounds are added to assay plates via non-contact acoustic transfer. Yeast stocks are manually added to Eve's reagent dispensers and then equal volumes are added to each well. Eve's robot arms and linear actuators are used to transfer plates and other consumables within the system. Each plate is bar-coded, and all movements are recorded.

Assay plates are stored in a shaking incubator, and fluorescence or absorption is measured every 90 minutes using one of Eve's two microplate readers. Eve also has an automated microscope capable of taking both bright-field and fluorescence images across a broad range of wavelengths.

Scan this QR code to see footage of Eve in action on the BBC website!

Ongoing Projects:

DARPA Big Mechanism: Big mechanisms are large, explanatory models of complicated systems in which interactions have important causal effects. Using cancer pathways as a test case, the Big Mechanism program aims to develop technology to read research abstracts and papers to extract pieces of causal mechanisms, assemble these pieces into more complete causal models, and reason through and test these models to provide explanations. Eve's role is to test and further develop the hypotheses produced as part of this project.

CHIST-ERA An Adaptive Automated Scientific Laboratory (ADALAB): This project aims to develop a framework for semi-automated and automated knowledge discovery by teams of human and robot scientists. The framework will integrate and advance a number of ICT methodologies: knowledge representation, ontology engineering, semantic technologies, machine learning, bioinformatics, and automated experimentation (Robot Scientists). The AdaLab framework will be evaluated using the diauxic shift – a yeast pathway with biomedical relevance to cancer and ageing.

References

1. Sparkes, A., W *et al* (2010). "Towards Robot Scientists for autonomous scientific discovery." Autom Exp **2**: 1.
2. Williams, K., *et al* (2015). "Cheaper faster drug development validated by the repositioning of drugs against neglected tropical diseases." J R Soc Interface **12**(104): 20141289.
3. Baird, J. K. (2007). "Neglect of Plasmodium vivax malaria." Trends Parasitol **23**(11): 533-539
4. Zhang, P., *et al* (2002). "Angiogenesis inhibitors specific for methionine aminopeptidase 2 as drugs for malaria and leishmaniasis." J Biomed Sci **9**(1): 34-40.
5. Leartsakulpanich, U., *et al* (2002). "Molecular characterization of dihydrofolate reductase in relation to antifolate resistance in Plasmodium vivax." Mol Biochem Parasitol **119**(1): 63-73.