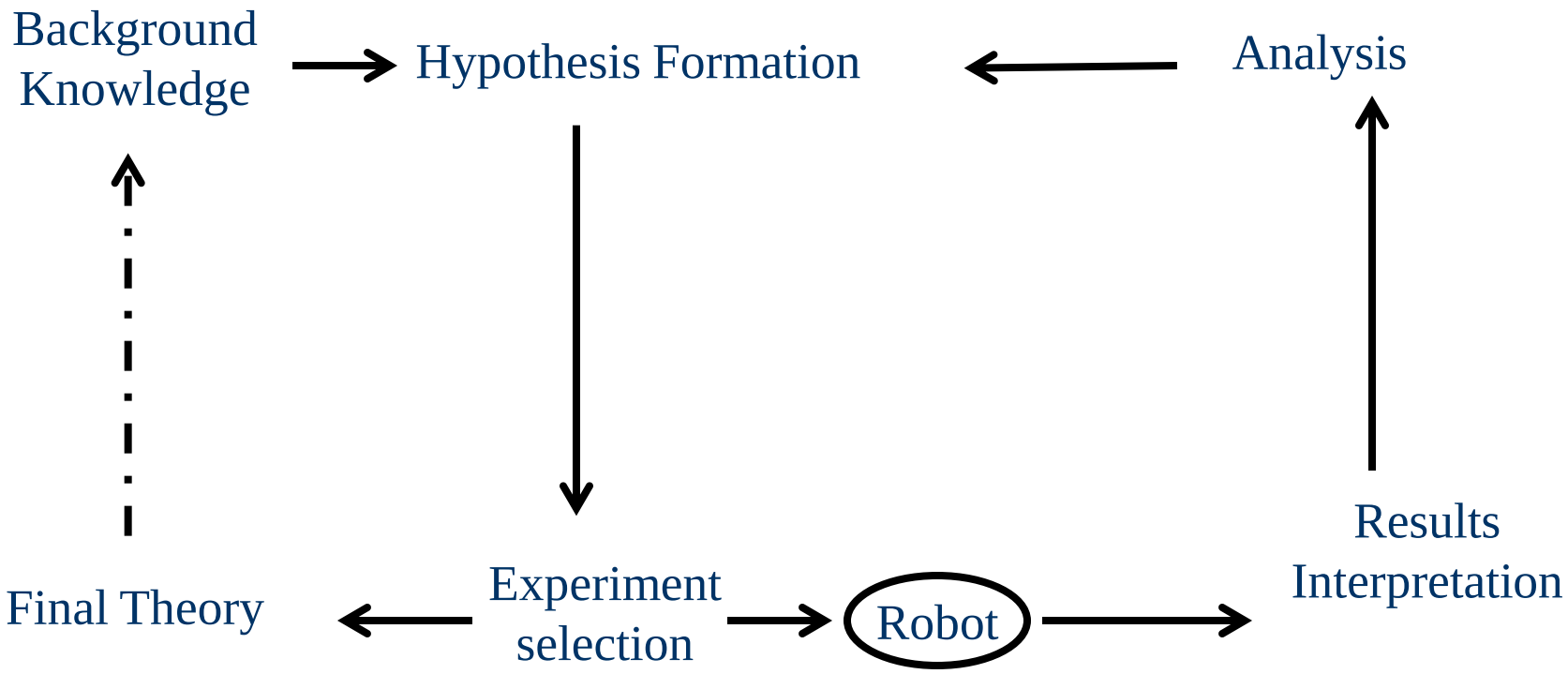
Automating Biology Using Robot Scientists

Ross D. King,
University of Manchester, ross.king@manchester.ac.uk

The Concept of a Robot Scientist

Computer systems capable of originating their own experiments, physically executing them, interpreting the results, and then repeating the cycle.



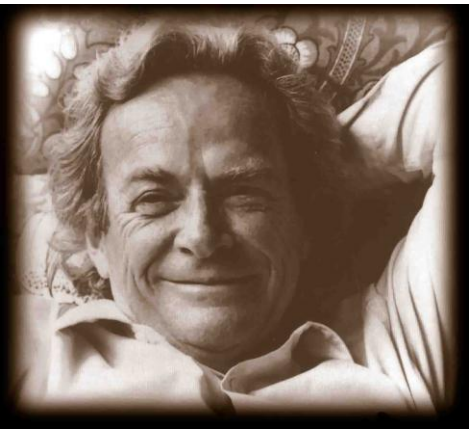
Talk Plan

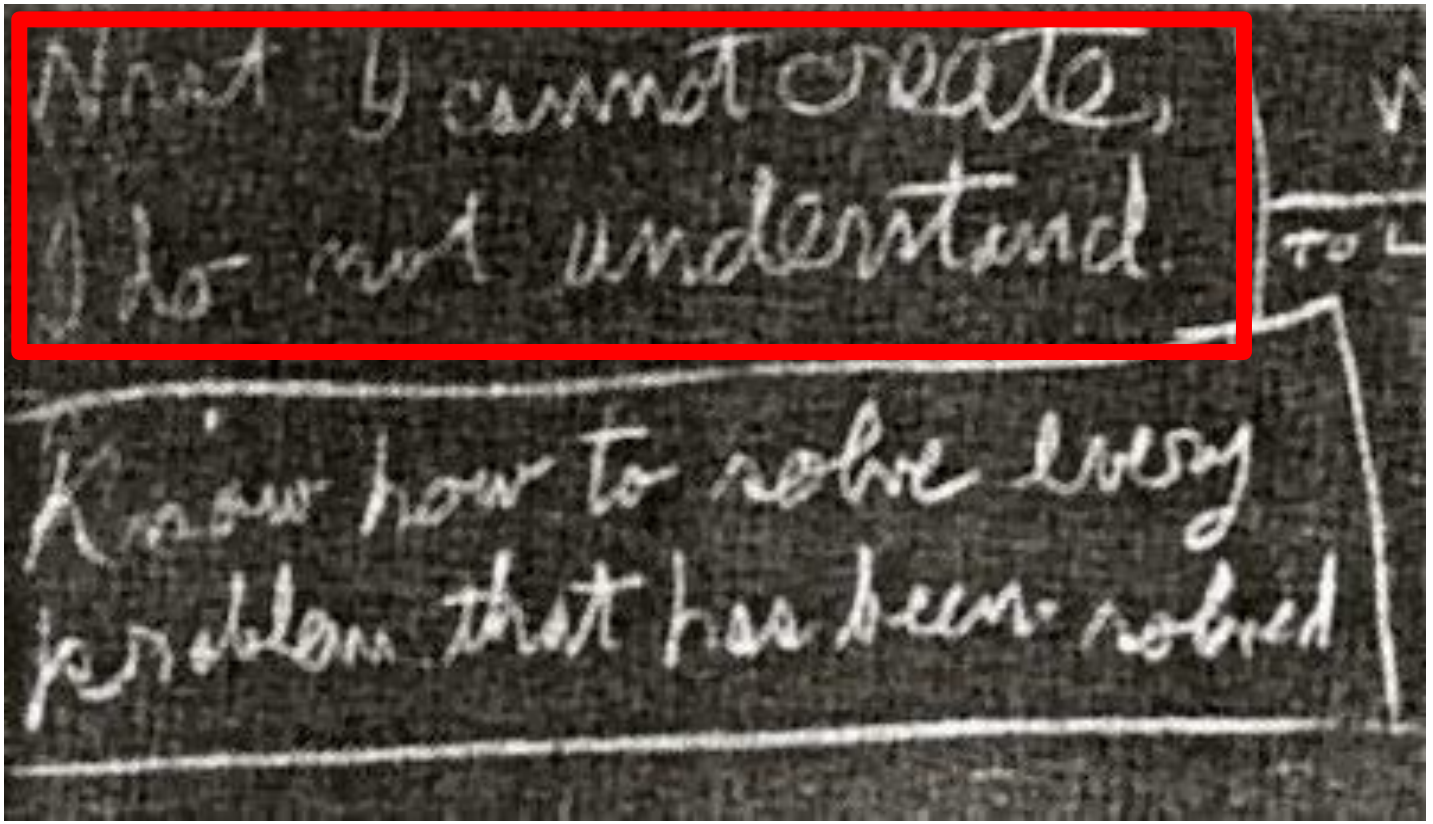
- Motivation: why science is a good application area for AI.
- Scientific Discovery: on the shoulder of giants.
- The Robot Scientist Adam: yeast functional genomics.
- The Robot Scientist Eve: drug design.
- Formalising Science: using logic to report science.
- Future Prospects?

Motivation: Philosophical

- What is Science?
- The question whether it is possible to automate scientific discovery seems to me central to understanding science.
- There is a strong philosophical position which holds that we do not fully understand a phenomenon unless we can make a machine which reproduces it.

Richard Feynman's Blackboard





"What I cannot create, I do not understand"

Motivation 2: Technological

- In many areas of science our ability to generate data is outstripping our ability to analyse it, e.g. genomics, drug screening..
- Data is being generated on an industrial scale.
- The analysis of scientific data needs to become as industrialised as its generation.

Motivation2: Technological

- Robot Scientists have the potential to increase the productivity of science. They can work cheaper, faster, more accurately, and longer than humans. They can also be easily multiplied.
  - *Enabling the high-throughput testing of hypotheses*.

- Robot Scientists have the potential to improve the quality of science.
  - *by enabling the description of experiments in greater detail and semantic clarity.*

The Complexity of Biological Systems

- Even simple “model” biological systems like that of *E. coli* and yeast are incredibly complicated.
- Thousands of genes, proteins, small-molecules, interacting together in complicated spatial temporal ways.
- Ockham's razor doesn’t work - system evolved.

- Not enough PhDs in the world to disentangle these systems.
- Need help - Robot Scientists.

Motivation 3: AI

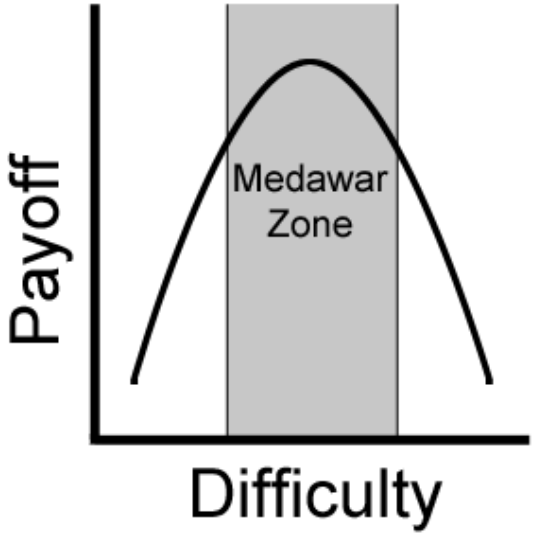
- Science is a wonderful test bed for AI.

- Compare/Contrast with Chess
  - Small abstract world: 64 squares, 36 pieces.
  - Computers now play chess much better than the best humans: ELO 2,800 v ELO 3,300.
  - Computers can now make strikingly beautiful moves.

  - No special “magic” for intelligence: increased quantity of search made a qualitative difference.

Motivation 3: AI

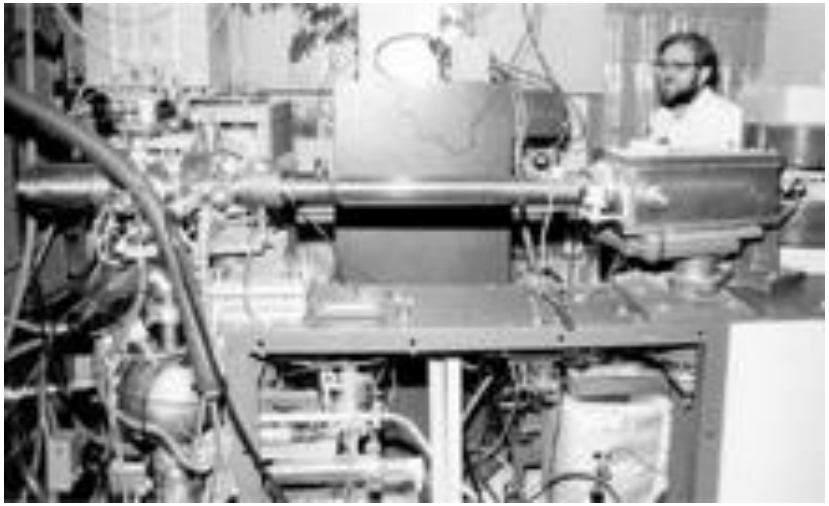
"The art of the soluble" Sir Peter Medawar

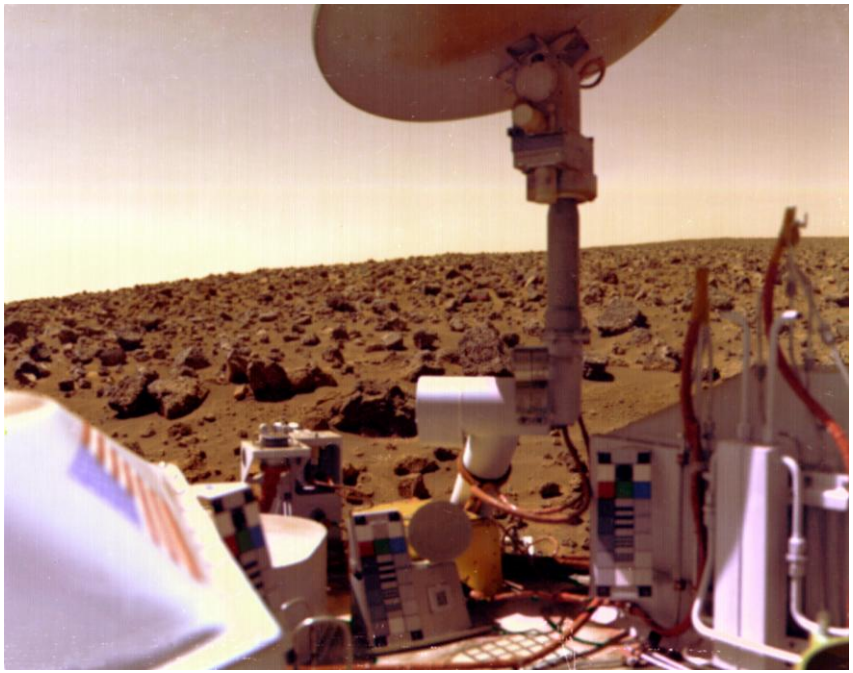


- Scientific problems are abstract, but involve the real-world.
- Scientific problems are restricted in scope – no need to know about "Cabbages and Kings".
- Nature is honest – no malicious agents.
- Nature is a worthy object of our study.
- The generation of scientific knowledge is a public good.

Scientific Discovery

Meta-Dendral

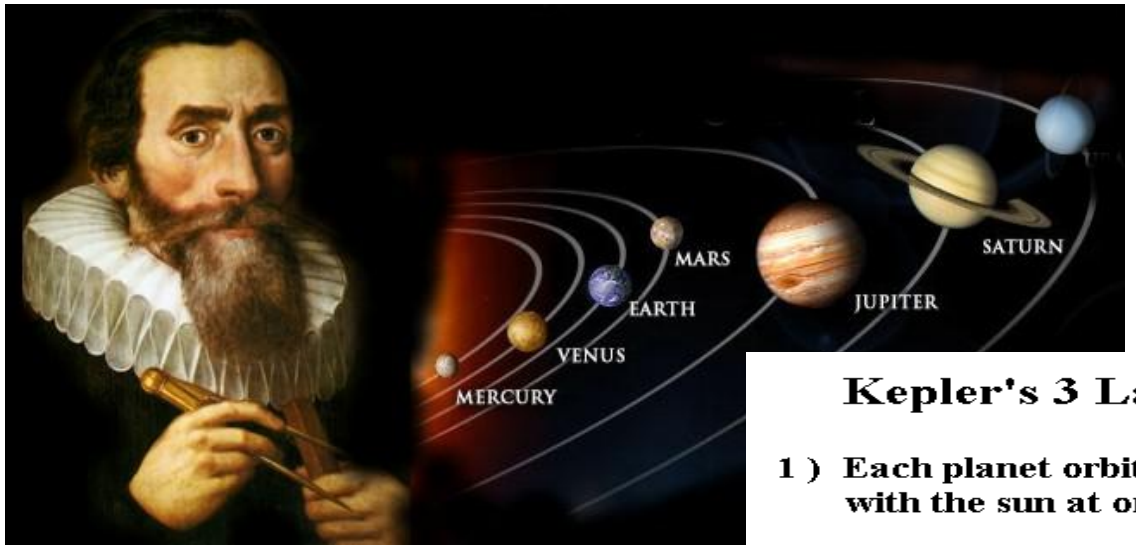




Analysis of mass-spectrometry data.

Joshua Lederburg, Ed. Feigenbaum, Bruce Buchanan, Karl Djerassi, *et al*. 1960-70s.

Bacon



Kepler's 3 Laws of Planetary Motion

1) **Each planet orbits the sun in an elliptical path with the sun at one focus**

2) **The radius vector (from sun to planet) sweeps out equal areas in equal time intervals**

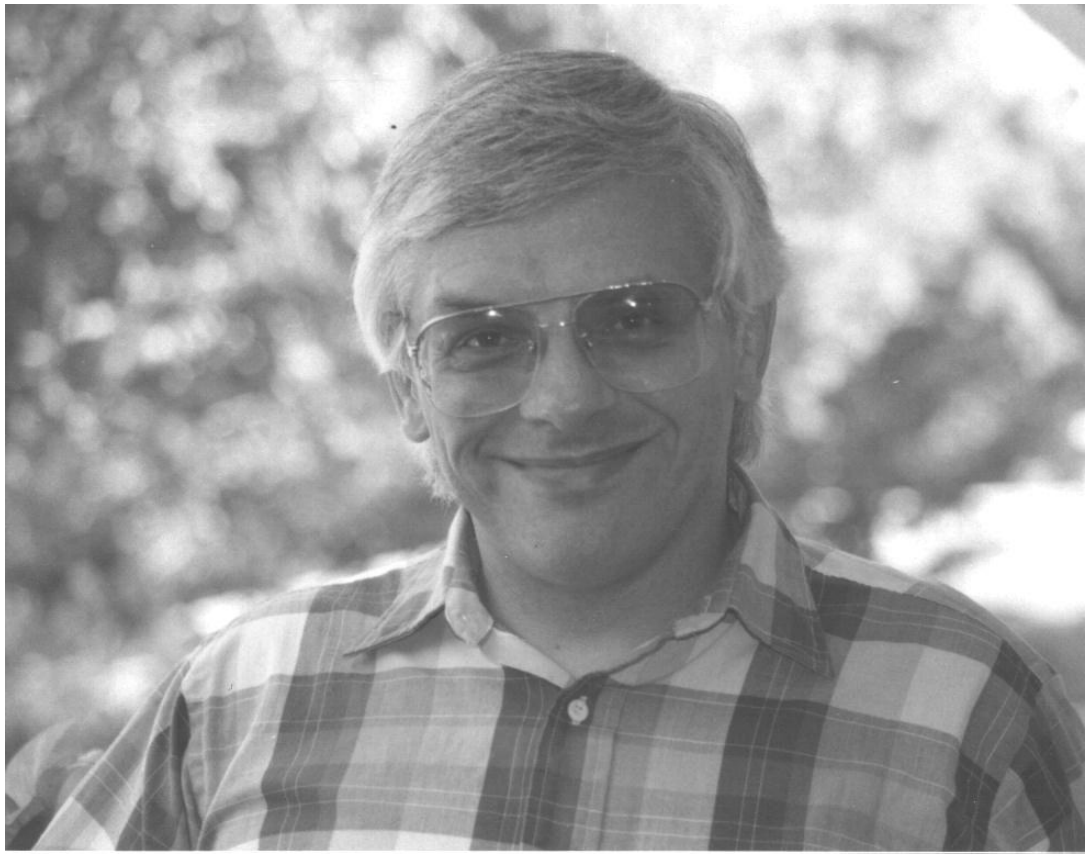
3) **The square of the period is proportional to the cube of the semi-major axis of the orbit**

i.e. $T^2 = k\ a^3$ **for some constant k**

Figure 11.1

Rediscovering physics and chemistry. Langley, Bradshaw, Simon (1979).

Into the Lab



Jan Zytkow (1944-2001)

- Automated discovery in a chemistry laboratory.

Zytkow, et al. (1990)

Robot Scientist Timeline

- 1999-2004 Initial Robot Scientist Project
  - Limited Hardware
  - Collaboration with Douglas Kell (Aber Biology), Steve Oliver (Manchester), Stephen Muggleton (Imperial)

  King *et al*. (2004) *Nature*, 427, 247-252

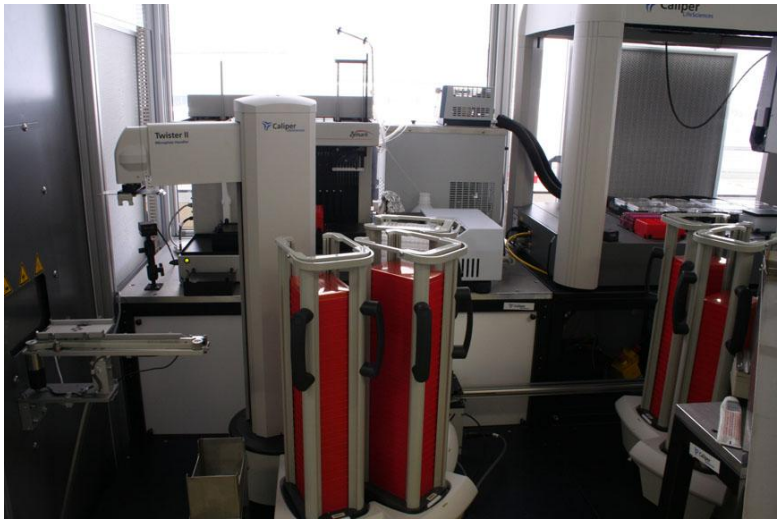
- 2004-2011 Adam Project – Yeast Functional Genomics
  - Sophisticated Laboratory Automation
  - Collaboration with Steve Oliver (Cambridge).

  King *et al*. (2009) *Science*, 324, 85-89

- 2008-2012 Eve Project – Drug Design for Tropical Diseases

Adam

Adam



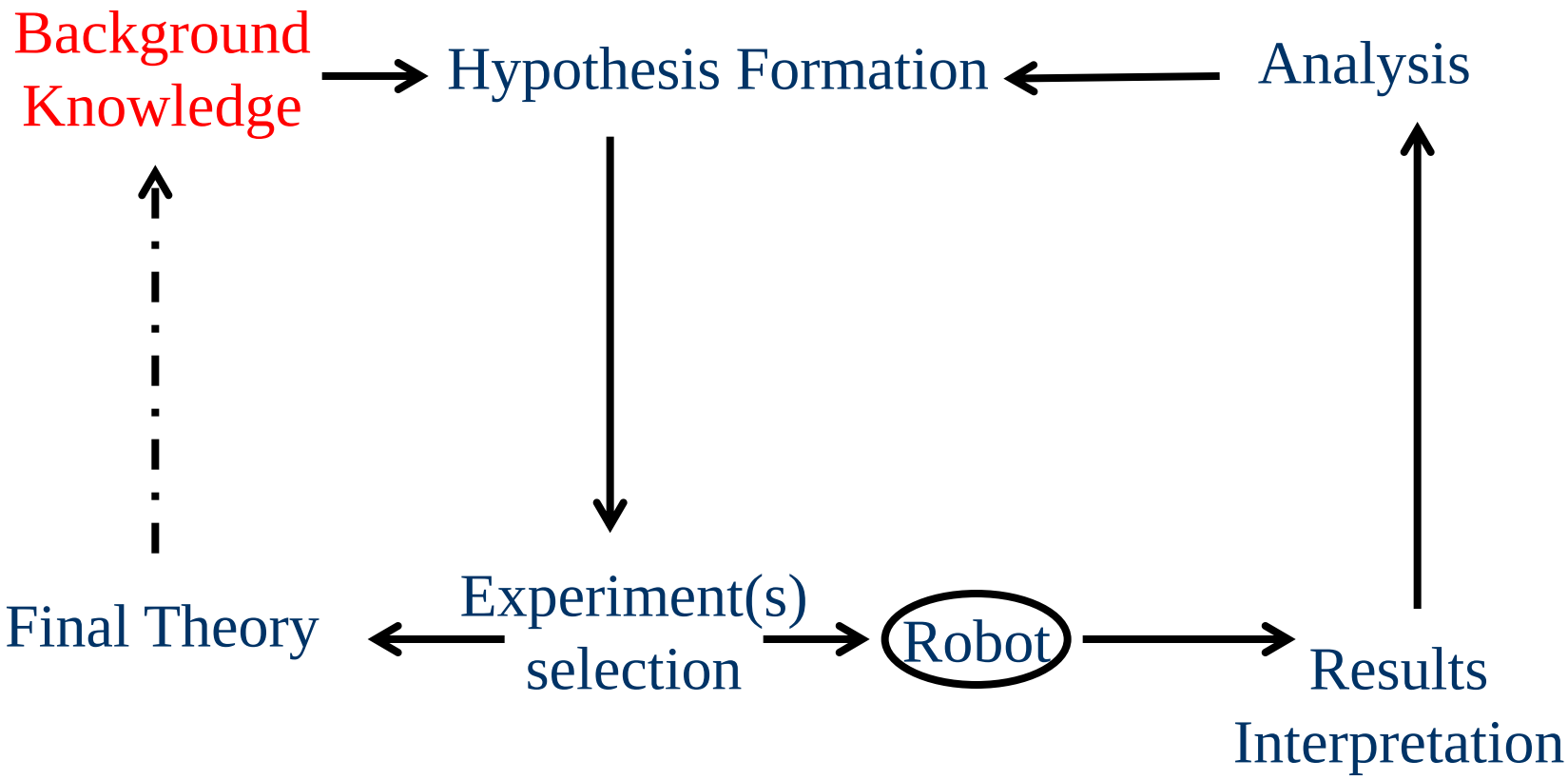
The Application Domain

- Functional genomics
- In yeast (*S. cerevisiae*) ~15% of the 6,000 genes still have no known function.
- EUROFAN 2 made all viable single deletant strains.
- Task to determine the “function” of a gene by growth experiments.

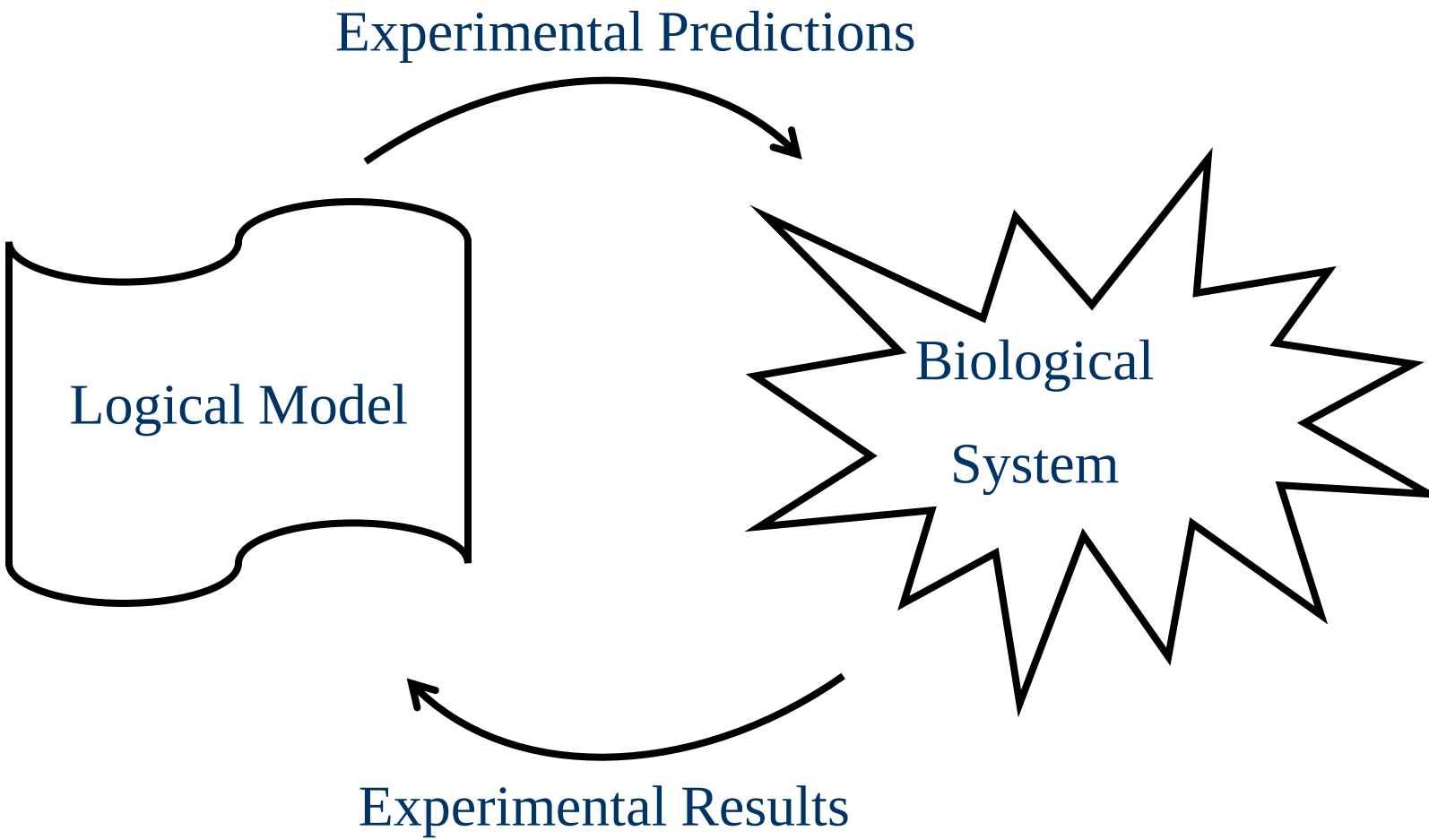
Formalising the Problem

- Use logic programming to represent background knowledge: metabolism modelled as a directed labeled hyper-graph.
- Use abduction to infer new hypotheses:
  - Abductive logic programming.
  - Techniques from Bioinformatics.
- Use active learning to decide efficient experiments: cost of compounds and time.
- Use machine learning to decide meaning of experimental results.

The Experimental Cycle



Model v Real-World

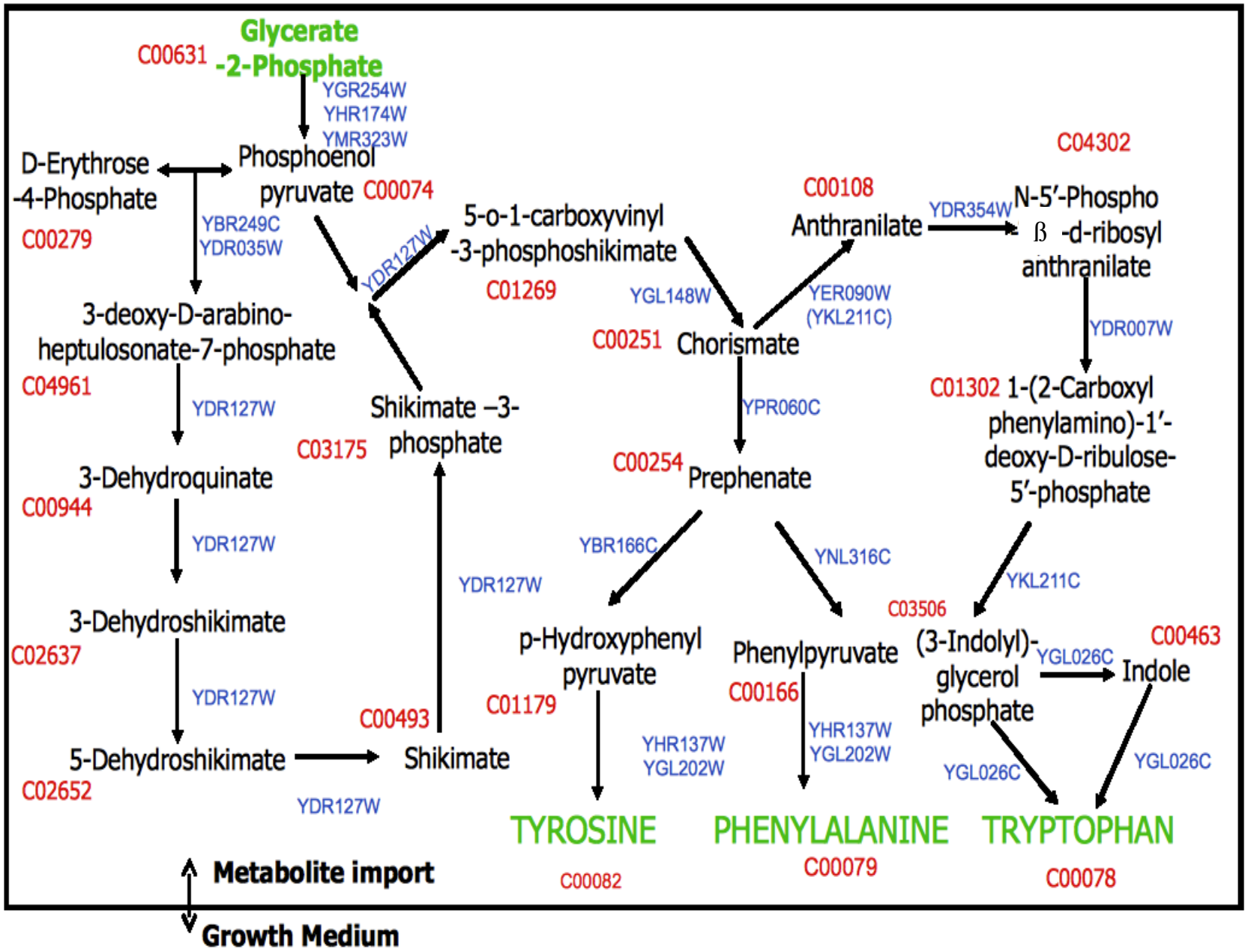


Logical Cell Model

- We have developed a logical formalism for modelling metabolic pathways (encoded in Prolog). This is essentially a directed labeled hyper-graph: with metabolites as nodes and enzymes as arcs.

- If a path can be found from cell inputs (metabolites in the growth medium) to all the cell outputs (essential compounds) then the cell can grow.

Phenyalanine, Tyrosine, and Tryptophan Pathways for *S. cerivisae*

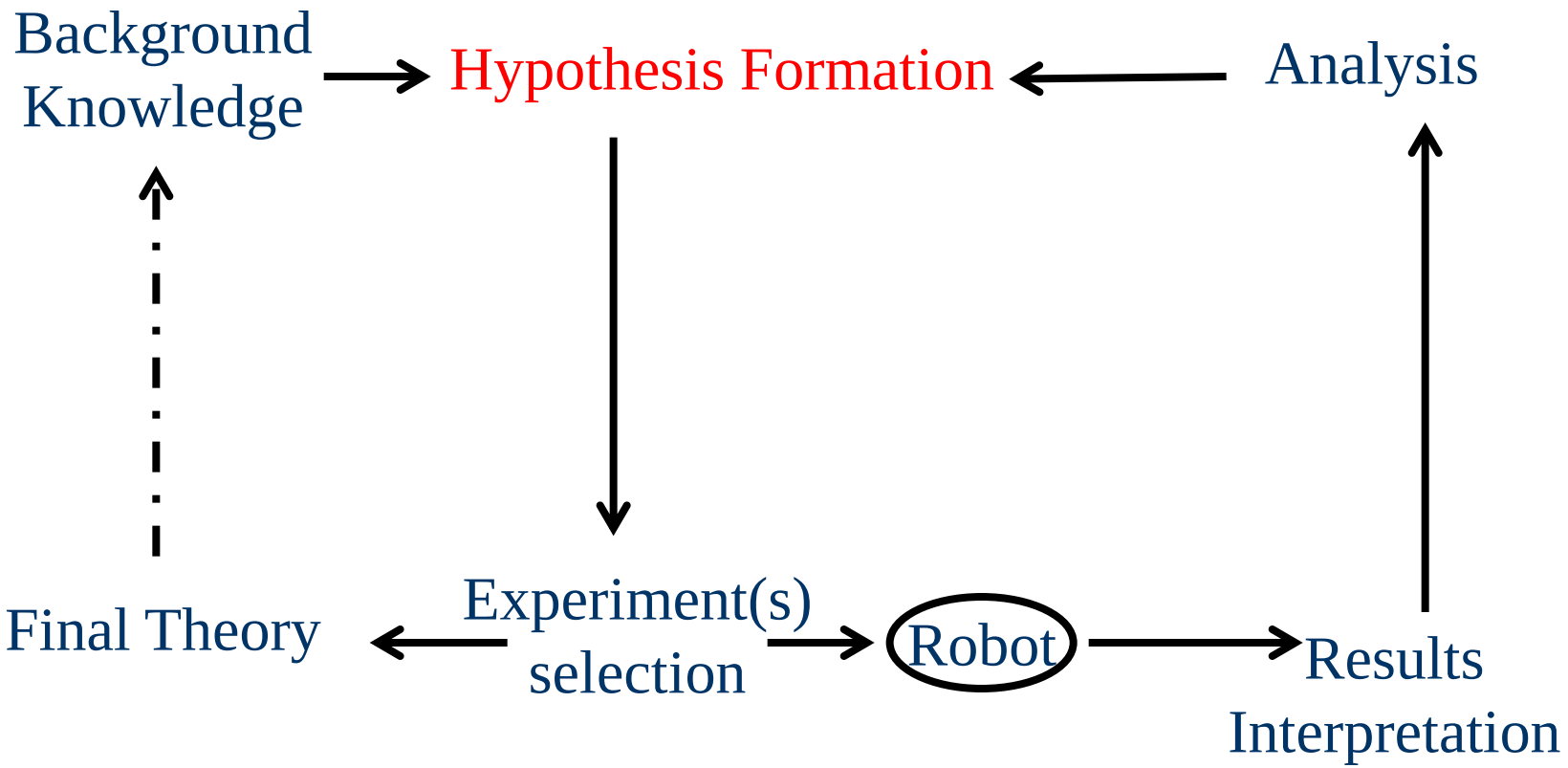


Genome Scale Model of Yeast Metabolism

- It covers most of what is known about yeast metabolism.
- Includes 1,166 ORFs (940 known, 226 inferred).
- Growth if path from growth medium to defined end-points.
- State-of-the-art accuracy in predicting cell viability.

- Now integrated with Yeast 4.0.

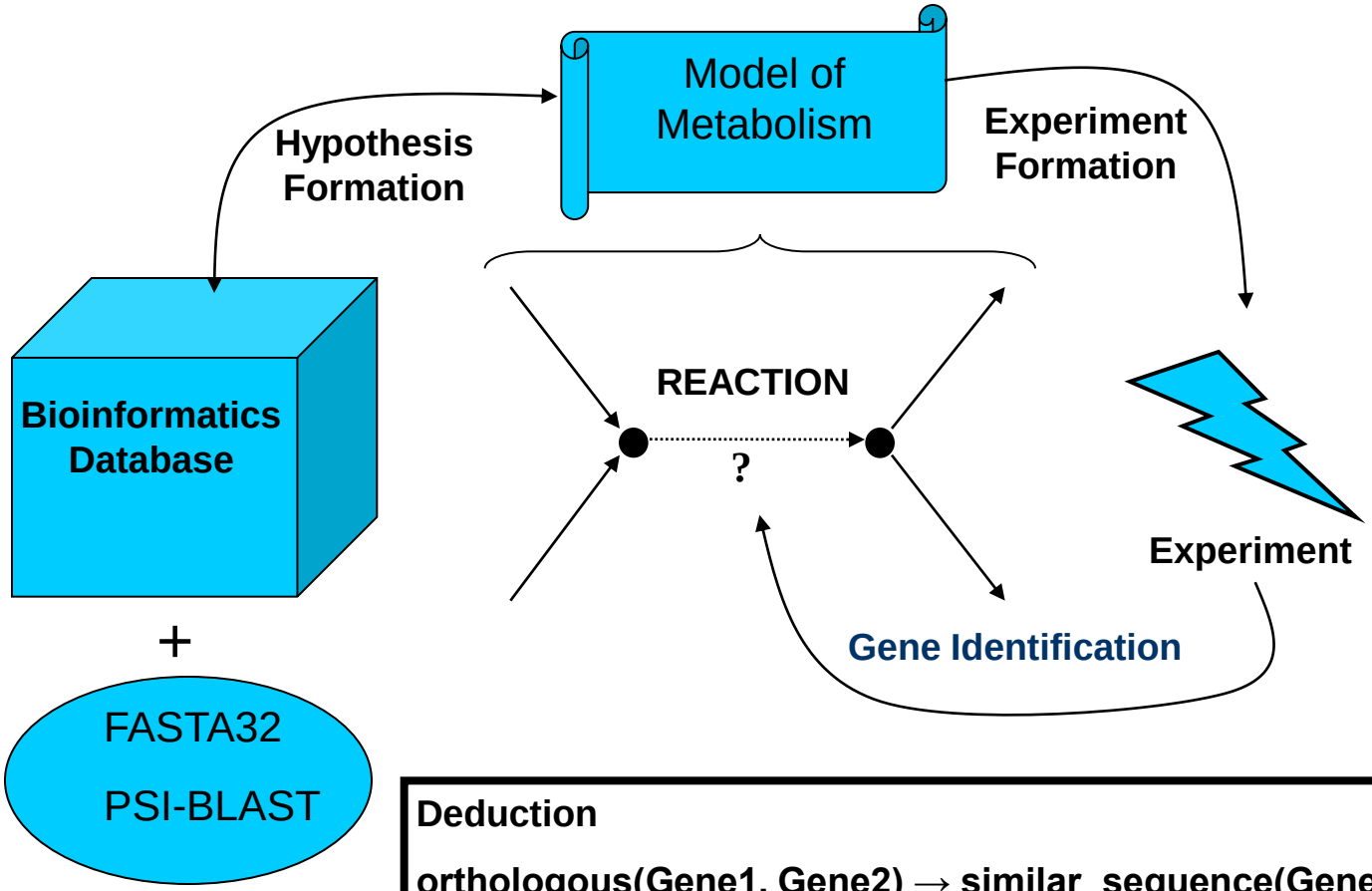
The Experimental Cycle



Inferring Hypotheses

- Science is based on the hypothetico-deductive method.
- In the philosophy of science. It has often been argued that only humans can make the "leaps of imagination" necessary to form hypotheses.
- In biology most hypothesis generation is abductive, not inductive.
- Adam used abductive inference to infer missing arcs/labels in its metabolic graph - hypotheses. With these missing nodes Adam could then deductively infer (explain) the observed experimental results.

Automated Model Completion



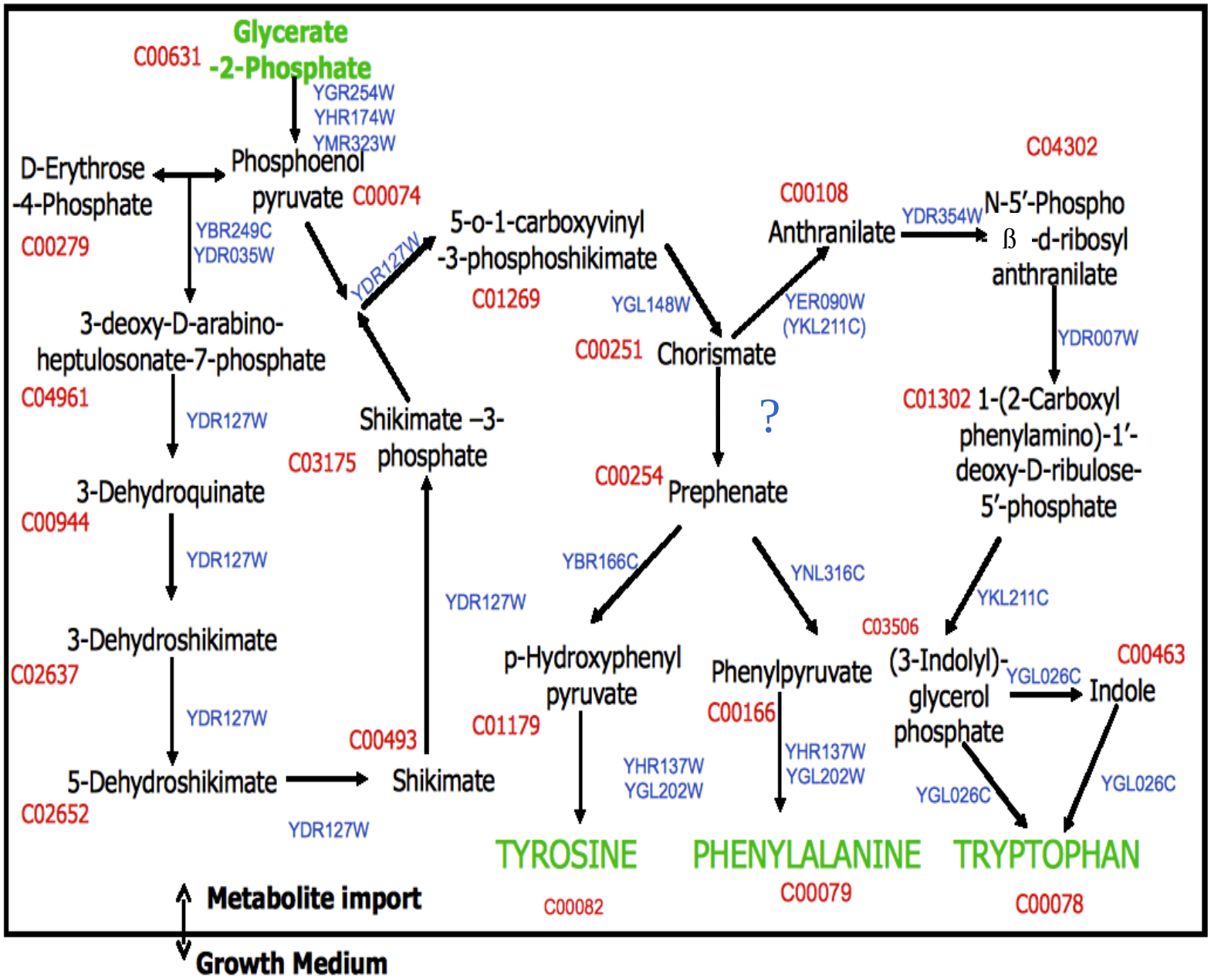
Deduction

orthologous(Gene1, Gene2) → similar_sequence(Gene1, Gene2).

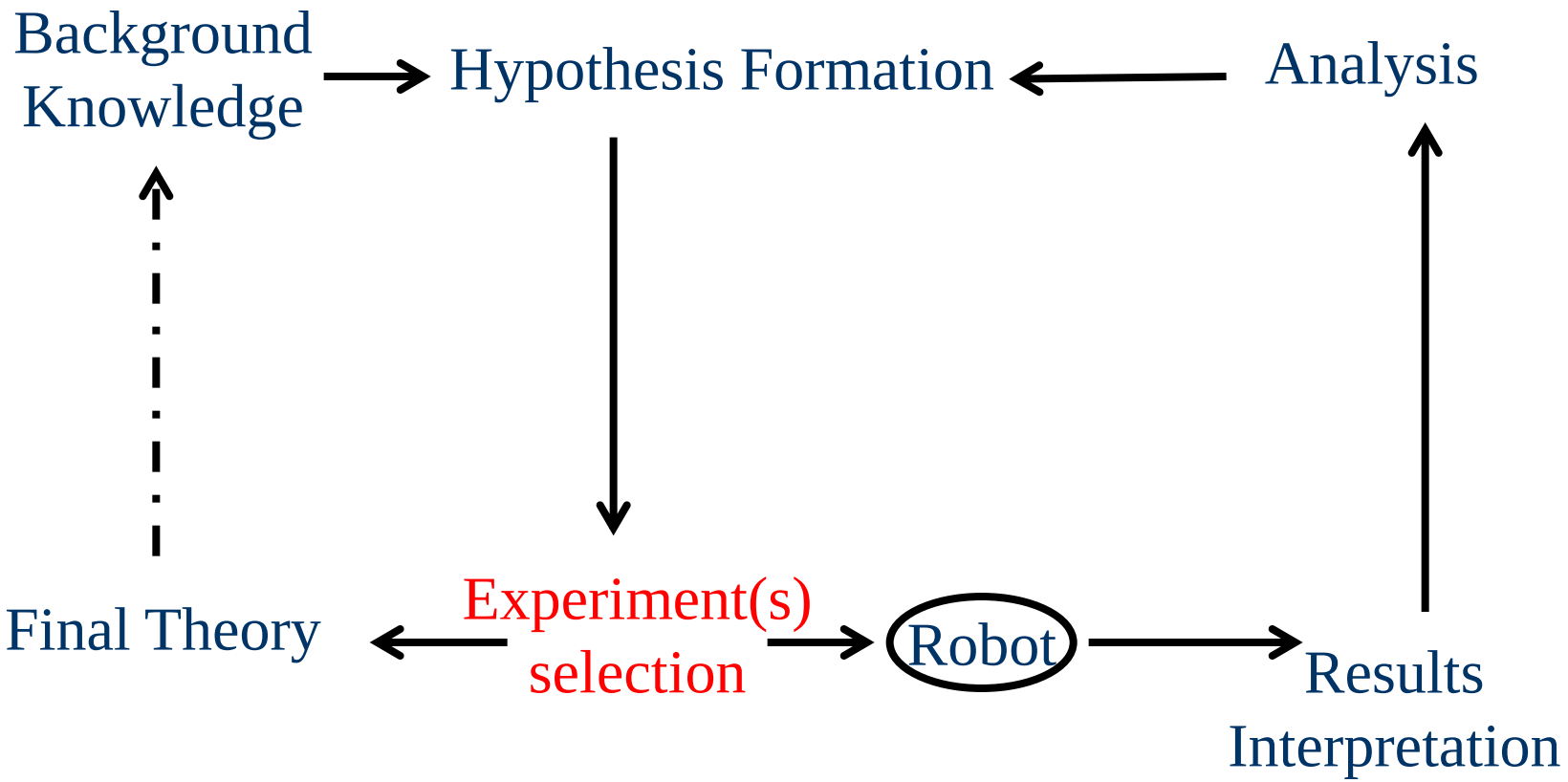
Abduction

similar_sequence(Gene1, Gene2) → orthologous(Gene1, Gene2).

Phenyalanine, Tyrosine, and Tryptophan Pathways for *S. cerivisae*



The Experimental Cycle

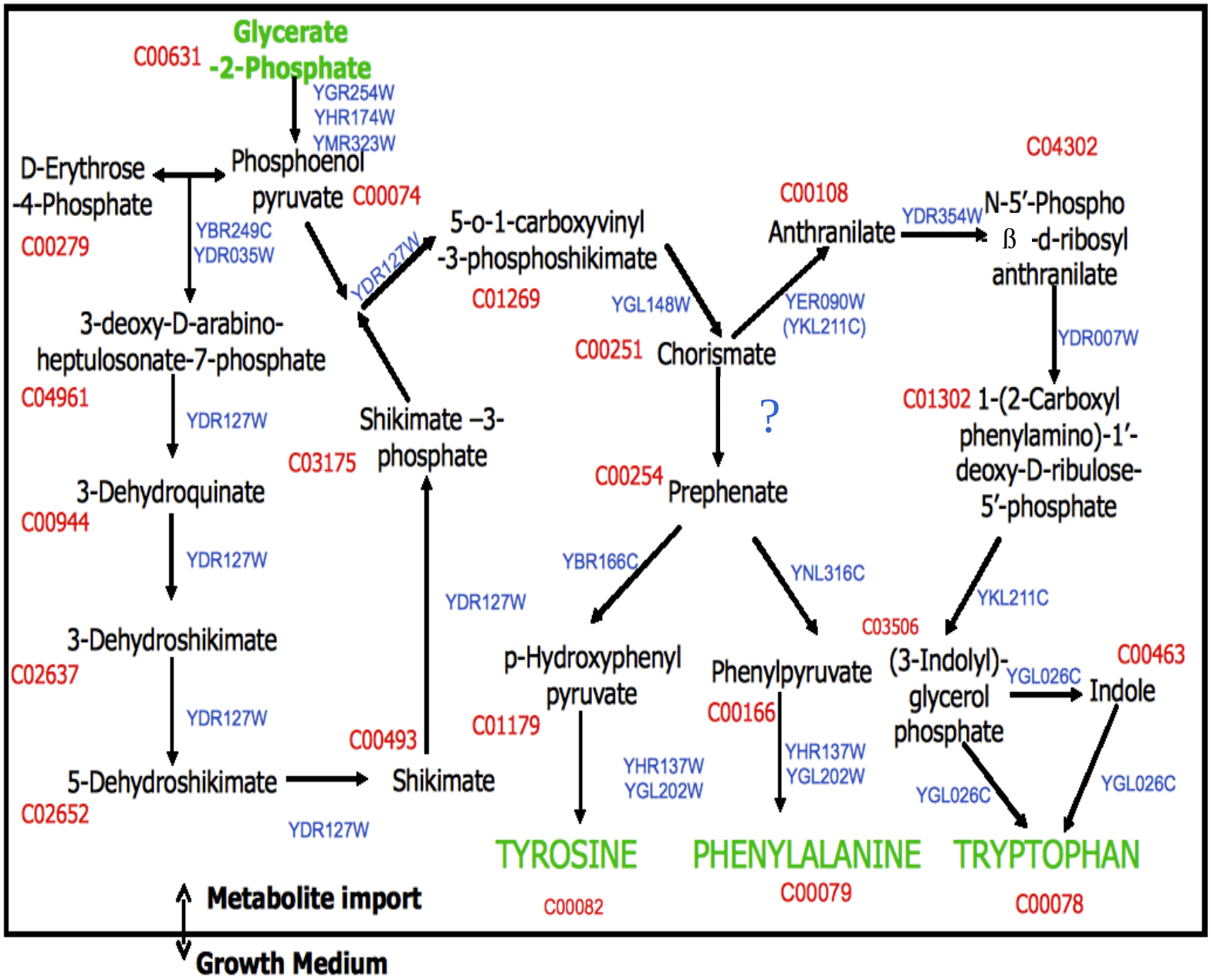


The Form of the Experiments

- Hypothesis 1: Gene YER152C codes for the enzyme the reaction: chorismate $\rightarrow$ prephenate.
- Hypothesis 2: Gene YGL060W codes for the enzyme the reaction: chorismate $\rightarrow$ prephenate.

- These can be tested by:
  - Growing YER152CΔ in environment +/- prephenate.
  - Growing YGL060WΔ in environment +/- prephenate.

Phenyalanine, Tyrosine, and Tryptophan Pathways for *S. cerivisae*



Inferring Experiments

Given a set of hypotheses we wish to infer an experiment that will efficiently discriminate between them

Assume:

- Every experiment has an associated cost.
- Each hypothesis has a probability of being correct.

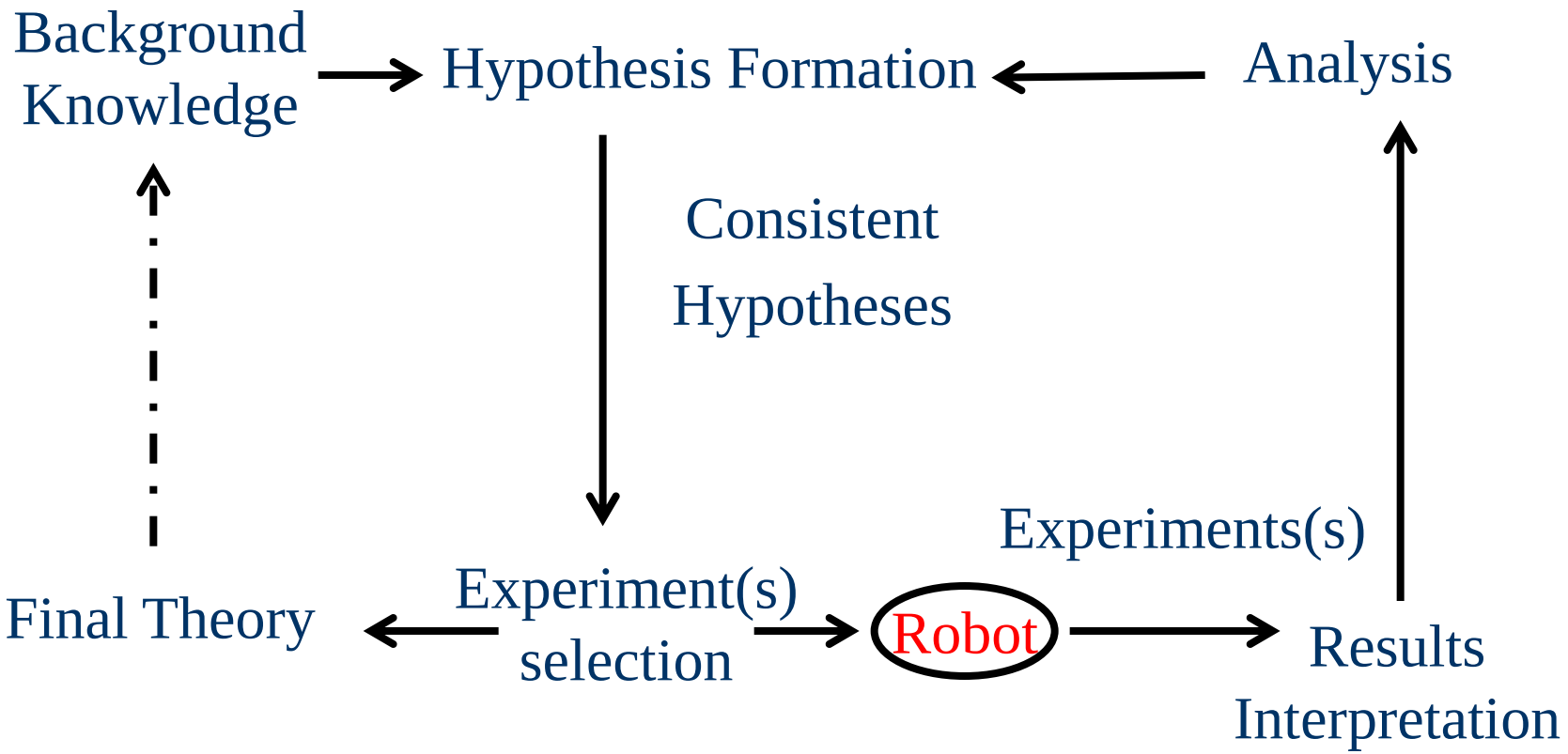
The task:

- To choose a series of experiments which minimise the expected cost of eliminating all but one hypothesis.

Active Learning

- In the 1972 Fedorov (Theory of optimal experiments) showed that this problem is in general intractable (NP complete).
- However, it can be shown that the problem is the same as finding an optimal decision tree; and it is known that this problem can be solved “nearly” optimally in polynomial time.
- We have shown that this strategy can outperform (get answer *faster* and *cheaper*) than simply choosing the cheapest experiment. Also better than humans on test problem.

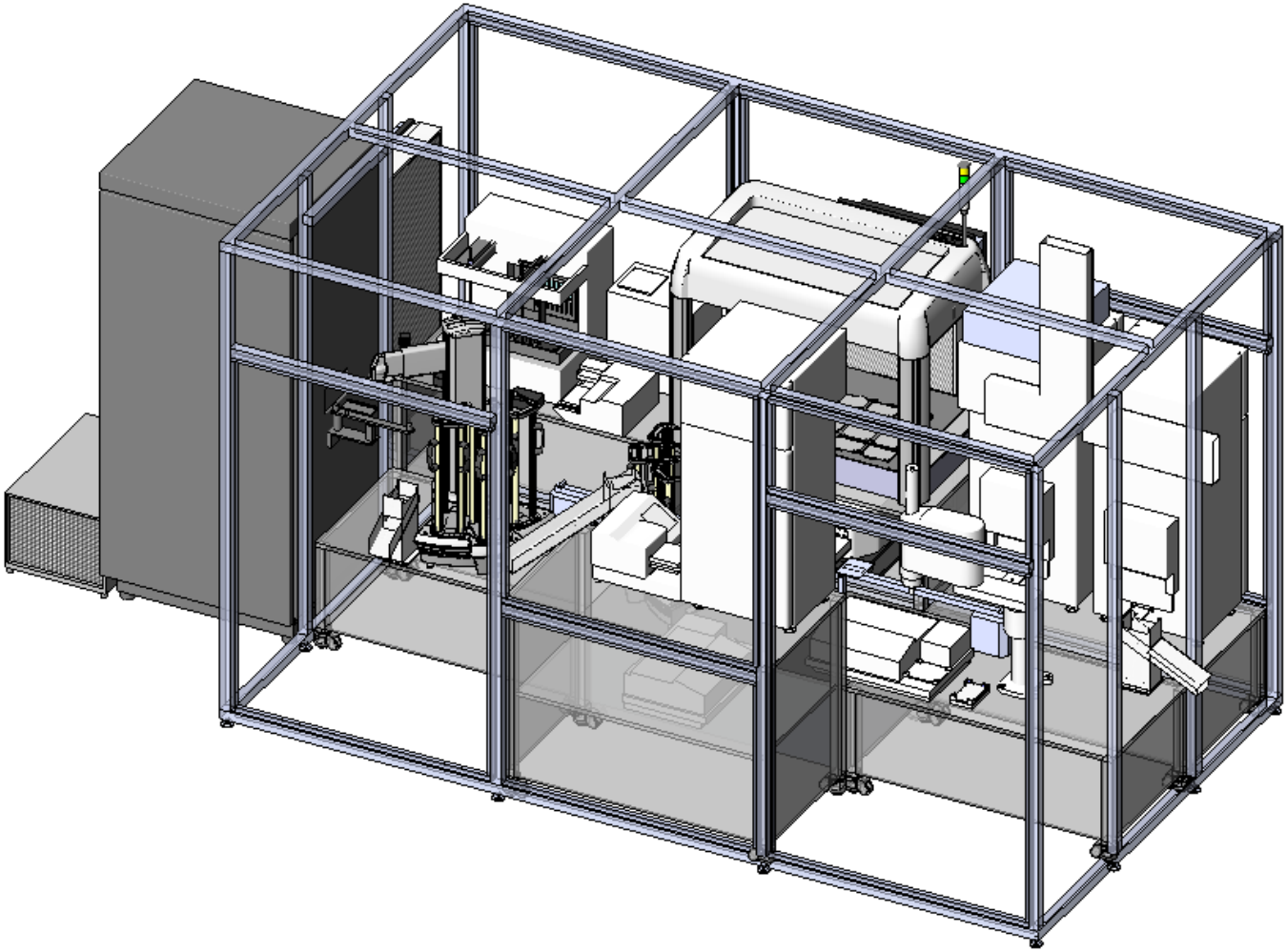
The Experimental Cycle



Adam

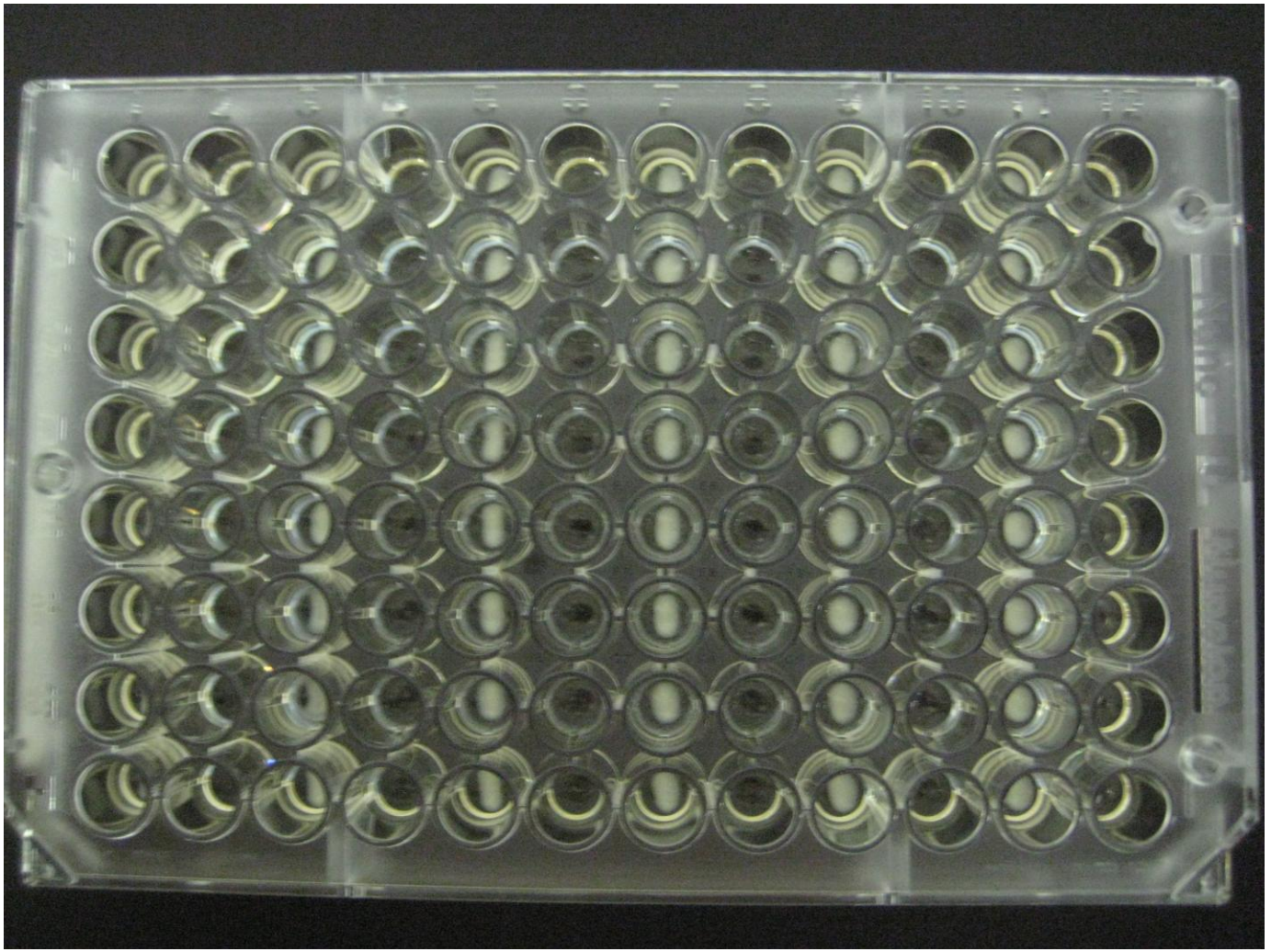
- One of the most sophisticated pieces of laboratory automation in the world.
- Designed to fully automate yeast growth experiments.
- Has a -20C freezer, 3 incubators, 2 readers, 3 liquid handlers, 3 robotic arms, 2 robot tracks, a centrifuge, a washer, an environmental control system, etc.
- Is capable of initiating ~1,000 new experiments and >200,000 observations per day in a continuous cycle.

Diagram of Adam

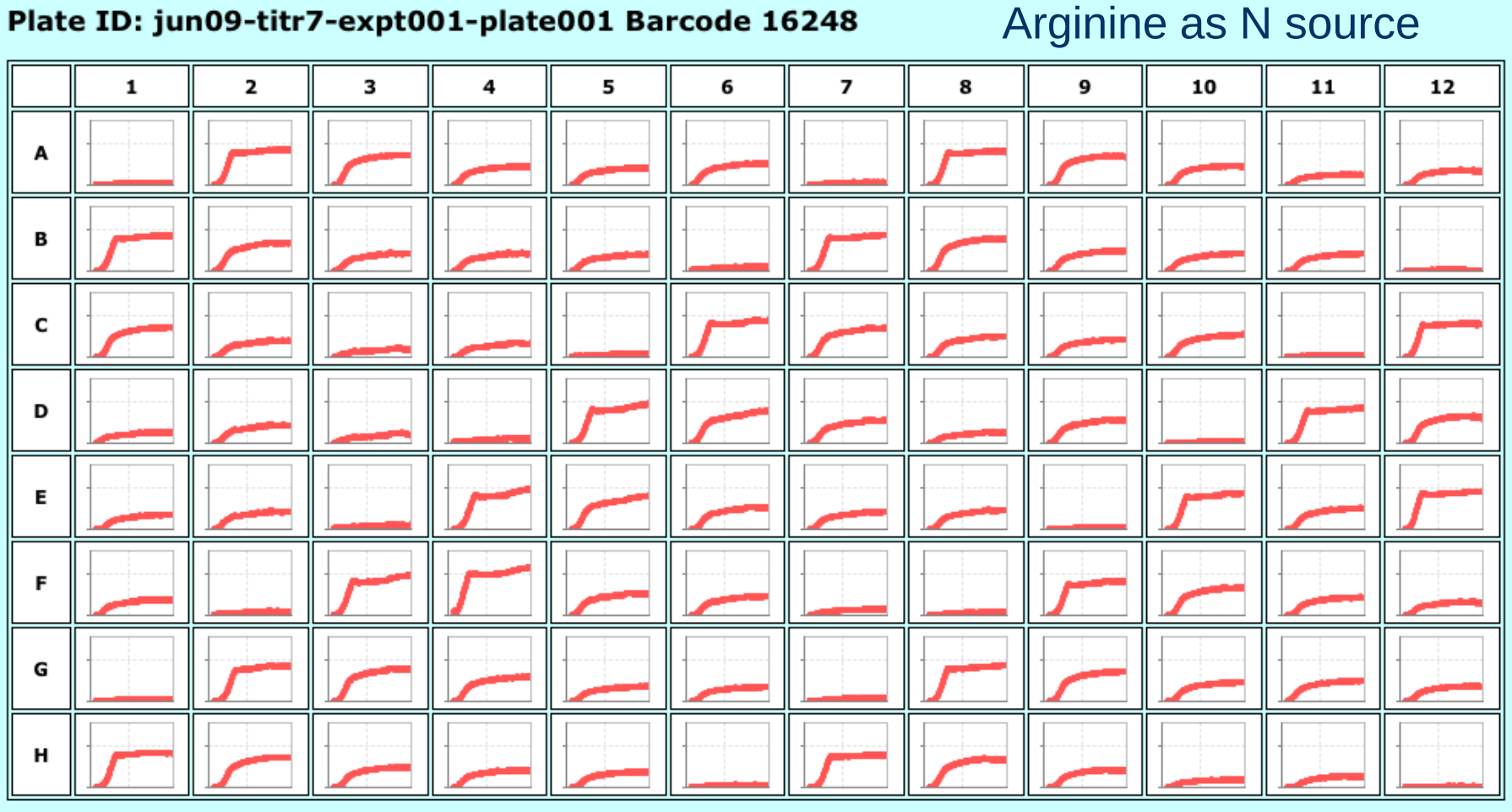


Adam in Action

Growth plates

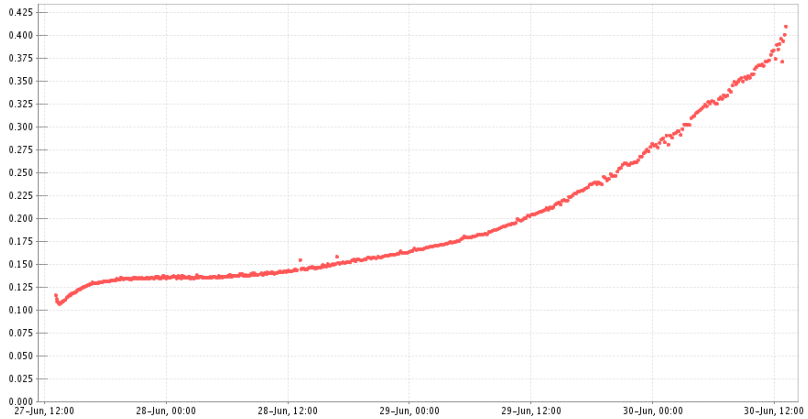


Example Growth Curves



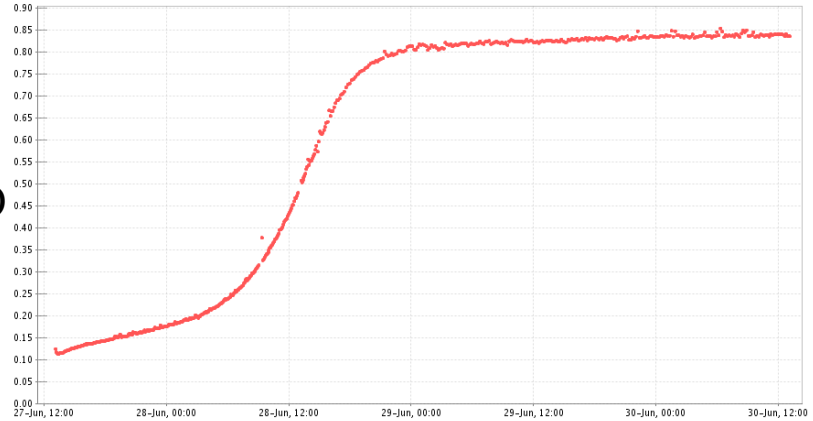
Growth curves

OD



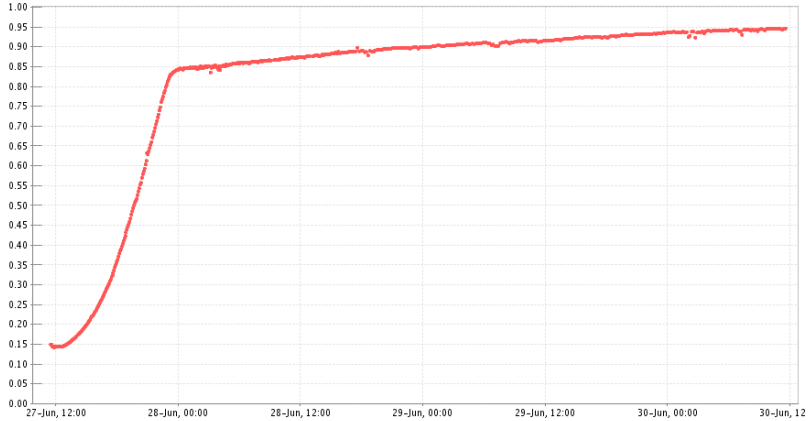
Time

OD



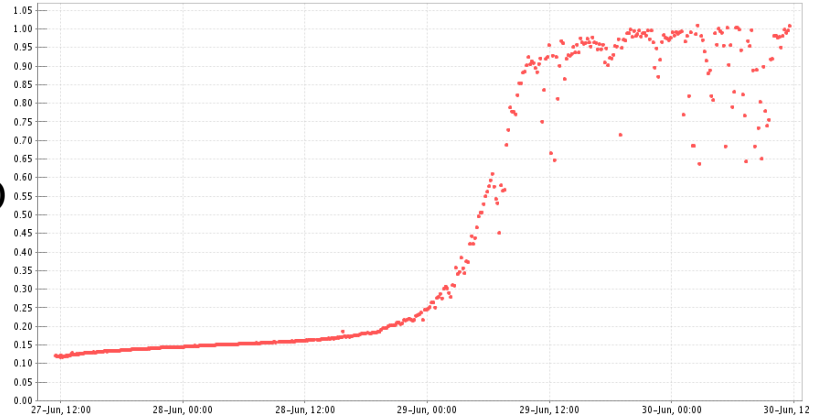
Time

OD



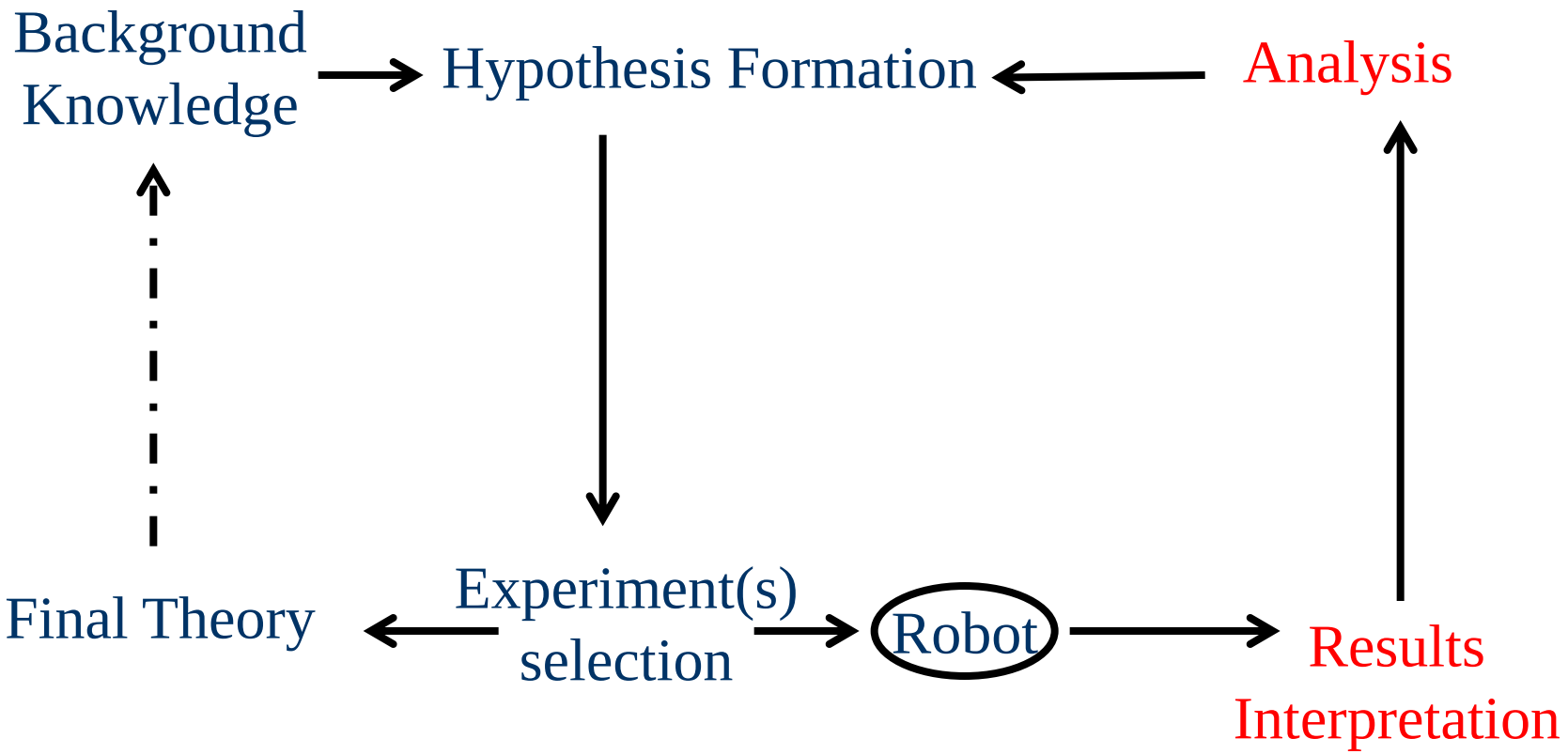
Time

OD



Time

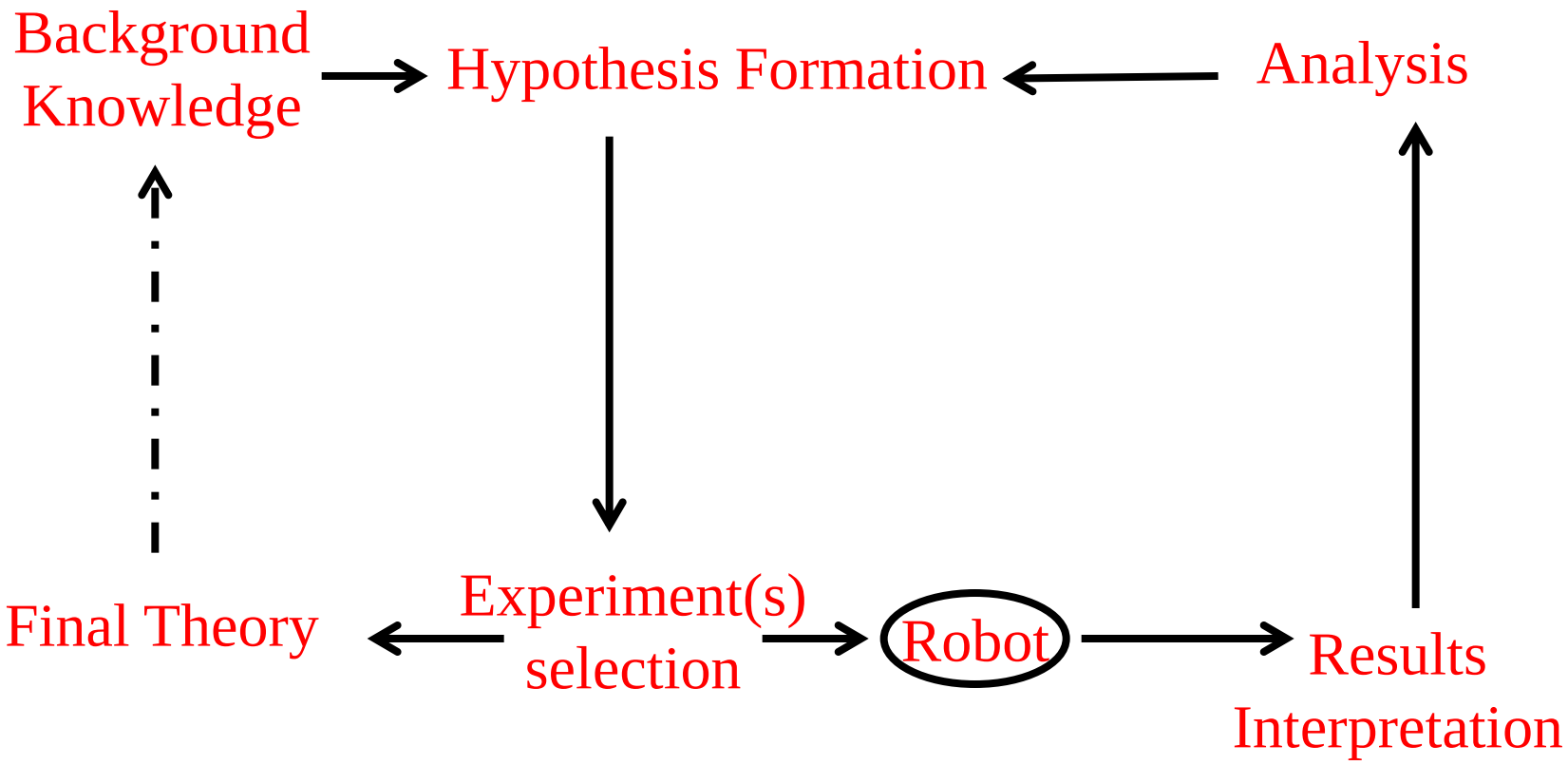
The Experimental Cycle



Qualitative to Quantitative

- The functions of most genes in *S. cerevisiae* that when deleted result in auxotrophy (no growth) have already been discovered.
- Most genes of unknown function only affect growth quantitatively.
- They may have slower growth (bradytrophs), faster growth, higher/lower biomass yield, etc.

The Experimental Cycle



Closing the Loop

- We have physically implemented all aspects of Adam.
- To the best of our knowledge Adam was the first AI system that can both explicitly form hypotheses and experiments, and physically do the experiments.

Discovery of Novel Science

Novel Science

- Adam generated and confirmed twelve novel functional-genomics hypotheses concerning the identify of genes encoding enzymes catalysing orphan reactions in the metabolic network of the yeast *S. cerevisiae.*

- Adam's conclusions have been manually verified using bioinformatic and biochemical evidence.

Novel Scientific Knowledge

	Orphan Enzyme	Hypothesised Gene	Prob.	Acc.	No.	Existing Annotation	Dry	Wet
1	glucosamine-6-phosphate deaminase (3.5.99.6)	YHR163W (SOL3)	$<10^{-4}$	97	8	'6-phosphogluconolactonase' ida	-	-
2	glutaminase (3.5.1.2)	YIL033C (BCY1)	$<10^{-4}$	92	11	'cAMP-dependent protein kinase inhibitor ' ida	✗ ?	-
3	L-threonine 3-dehydrogenase (1.1.1.103)	YDL168W (SFA1)	$<10^{-4}$	83	6	'alcohol dehydrogenase' ida	-	-
4	purine-nucleoside phosphorylase (2.4.2.1)	YLR209C (PNP1)	$<10^{-4}$	82	11	'purine-nucleoside phosphorylase' ida	✓	-
5	2-aminoadipate transaminase (2.6.1.39)	YGL202W (ARO8)	$<10^{-4}$	80	3	'aromatic-amino-acid transaminase' ida	✓	✓
6	5,10-methenyltetrahydrofolate synthetase (6.3.3.2)	YER183C (FAU1)	$<10^{-4}$	80	4	'5,10 formyltetrahydrofolate cyclo-ligase' ida	✓	-
7	glucosamine-6-phosphate deaminase (3.5.99.6)	YNR034W (SOL1)	$<10^{-4}$	79	2	'possible role in tRNA export'	-	-
8	pyridoxal kinase (2.7.1.35)	YPR121W (THI22)	$<10^{-4}$	78	1	'phosphomethylpyrimidine kinase' iss	-	-
9	mannitol-1-phosphate 5-dehydrogenase (1.1.1.17)	YNR073C	$<10^{-4}$	78	6	'putative mannitol dehydrogenase ' iss	-	-
10	1-acylglycerol-3-phosphate O-acyltransferase (2.3.1.51)	YDL052C (SLC1)	0.0001	80	6	'1-acylglycerol-3-phosphate O-acyltransferase' ida	✓	-
11	glucosamine-6-phosphate deaminase (3.5.99.6)	YGR248W (SOL4)	0.0002	78	2	'6-phosphogluconolactonase' ida	-	-
12	maleylacetoacetate isomerase (5.2.1.2)	YLL060C (GTT2)	0.0003	76	3	'glutathione S-transferase' ida	-	-
13	serine O-acetyltransferase (2.3.1.30)	YJL218W	0.0005	78	2	'unknown function'	-	-
14	L-threonine 3-dehydrogenase (1.1.1.103)	YLR070C (XYL2)	0.0052	75	6	'xylitol dehydrogenase' ida	-	-
15	2-aminoadipate transaminase (2.6.1.39)	YJL060W (BNA3)	0.0084	73	3	'kynurenine aminotransferase' ida	-	✓
16	pyridoxal kinase (2.7.1.35)	YNR027W	0.0259	76	2	'involved in bud-site selection' iss	-	-
17	polyamine oxidase (1.5.3.11)	YMR020W (FMS1)	0.0289	78	4	'polyamine oxidase' ida	✓	-
18	2-aminoadipate transaminase (2.6.1.39)	YER152C	0.0332	74	3	'uncharacterized'	-	✓
19	L-aspartate oxidase (1.4.3.16)	YJL045W	0.1300	72	1	'succinate dehydrogenase isozyme' iss	-	-
20	purine-nucleoside phosphorylase (2.4.2.1)	YLR017W (MEU1)	0.1421	72	6	'methylthioadenosine phosphorylase' ida	✓	-

Eve

Eve

Parasitic Diseases targeted

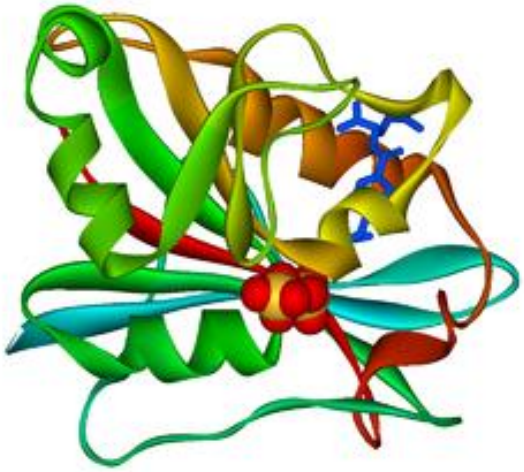
Malaria

Shistosomaisis

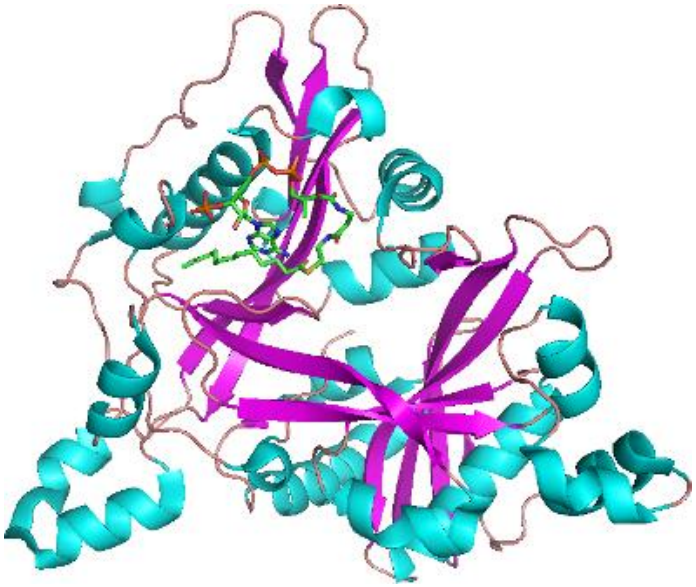
Leishmania

Chagas

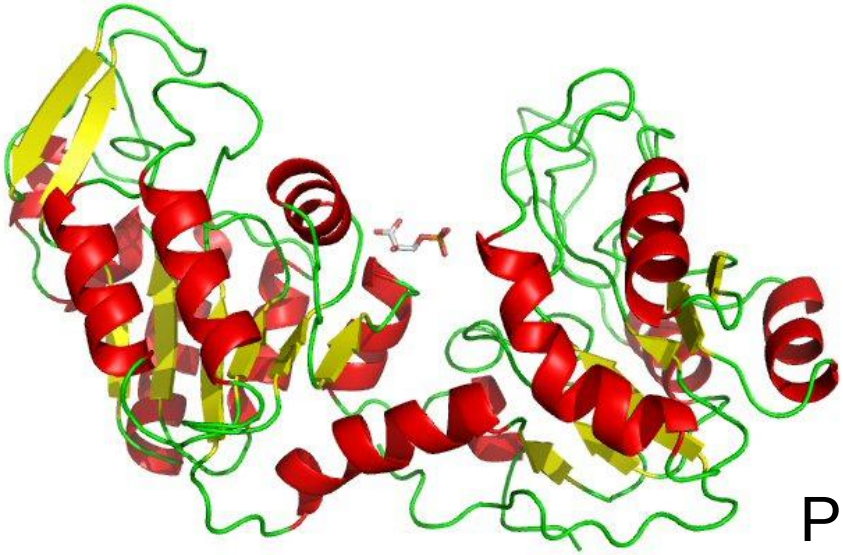
Enzymes Targeted



Dihydrofolate Reductase (DHFR)



N-myristoyl transferase



Phosphoglycerate kinase

Formalising the Problem

- Use graphs and standard chemoinformatic methods to represent background knowledge - the use of relations is planned.
- Uses induction (quantitative structure activity relationship – QSAR learning) to infer new hypotheses.
- Use active learning to decide efficient experiments, and econometric model to decide what compounds to test.

Design

- Novel design features: Combines: screening, hit confirmation, and QSAR learning.
  - Uses synthetic biology based yeast assays.
  - During the standard screening process Eve is able to decide to switch to QSAR mode.
  - Use cycles of active learning to improve QSARs.

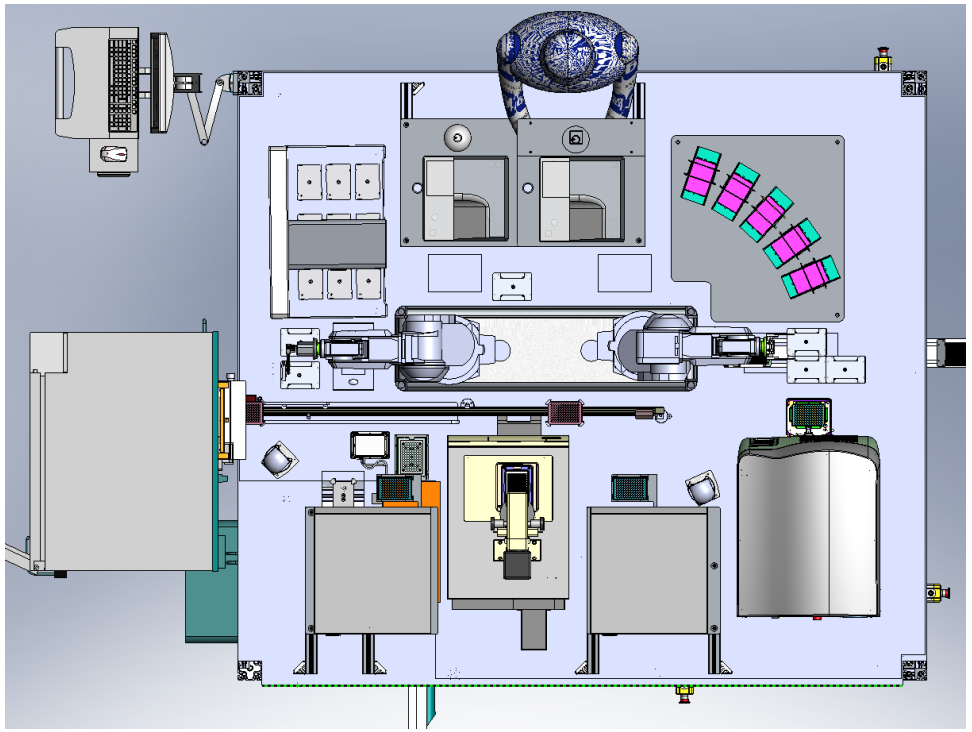
Synthetic Biology based Assays

- Our idea is to engineer cells to be analog computers.
- These computers will accurately estimate a biological function that corresponds to the set of desired assay properties.
- The function estimated is the utility of a compound against a disease.
- E.g. ((inhibit *P. vivax* DHFR) ∧ (¬ inhibit *H. sapiens* DHFR) ∧ (¬ cytotoxic)).

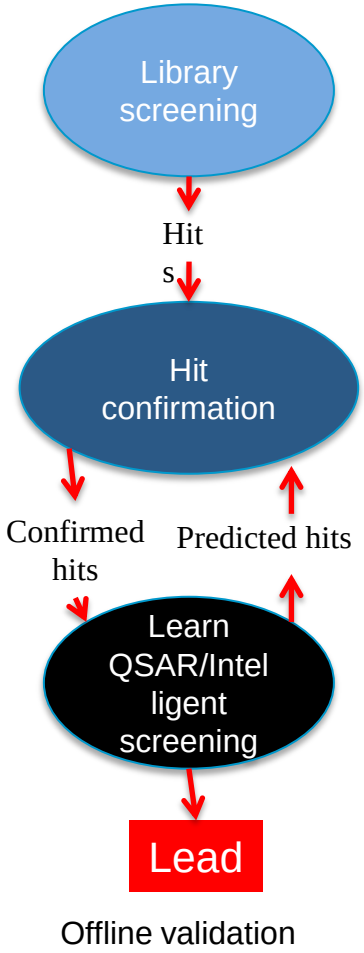
Eve's Hardware

Highlights of Eve's hardware:

- Acoustic liquid handling
- High throughput 384 well plates
- Two industrial robot arms
- Automated 60x microscope
- Liquid handlers, fluorescence readers, barcode scanners, dry store, incubator, tube decapper ...



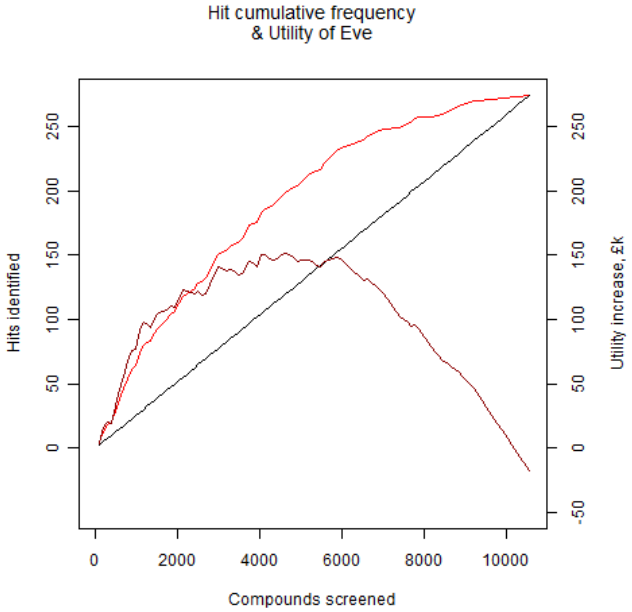
Eve's Intelligent Screening

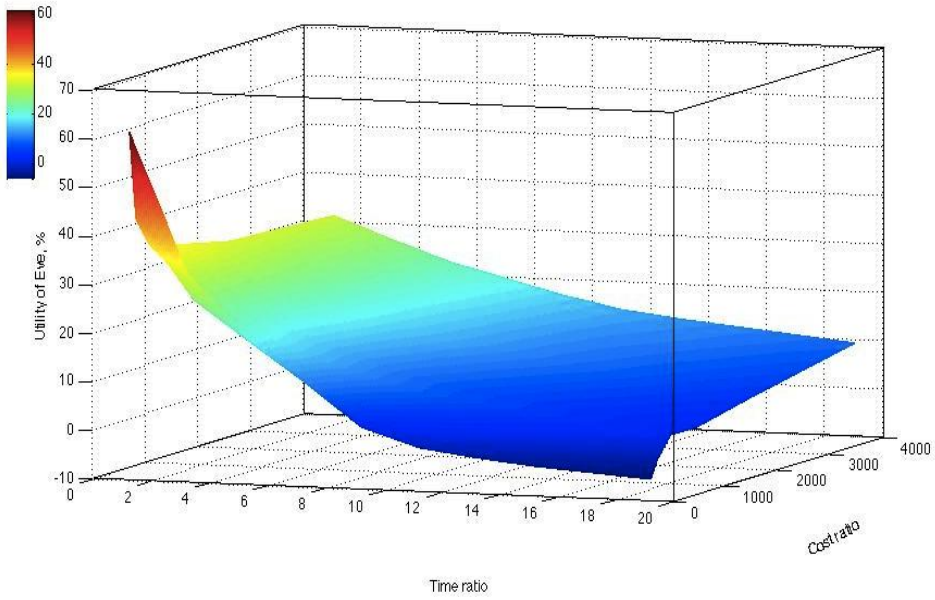


- Standard library screening is brute force: "begin at the beginning and go on till you come to the end: then stop".
- In the standard "pipeline" the 3 processes are not integrated.

- To learn its QSARs Eve uses Eve uses a Gaussian process model.
- Uses active learning to select predicted hits.

The Economics of Intelligent Screening





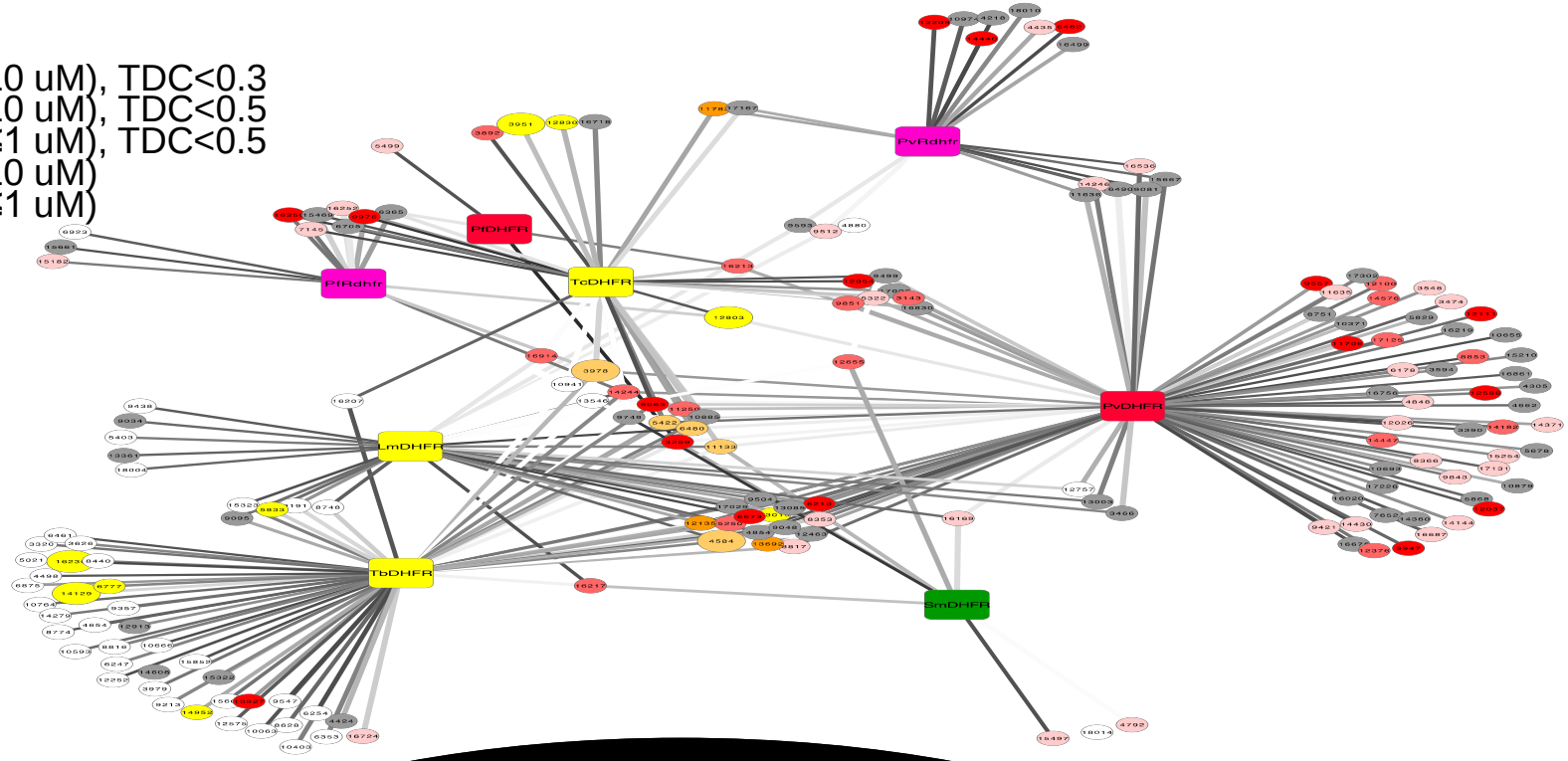
$$\Delta\, Utility\ of\ Eve = \sum_{1}^{Nm}(Tm + Cm) + \sum_{1}^{Nx}(Tc + Cc - Uh) + \sum_{1}^{Ne}(Tm - Tc + Cm - Cc)$$

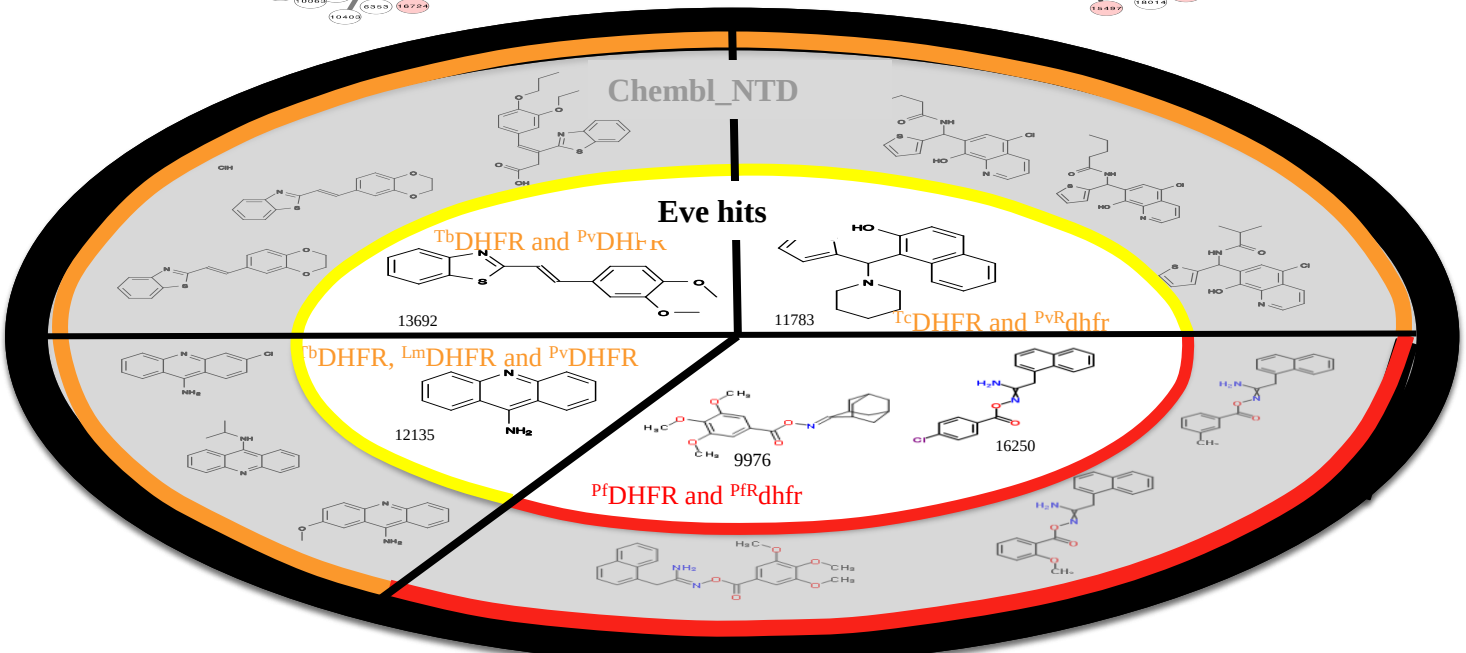
Nm	-	Number of compounds not assayed by Eve
Tm	-	Cost of the time to screen a compound using the mass screening assay
Cm	-	Cost of the loss of a compound in the mass screening assay
Nx	-	Number of hits missed by Eve
Tc	-	Cost of the time to screen a compound using a cherry-picking (confirmation or intelligent) assay
Cc	-	Cost of the loss of a compound in a cherry-picking assay
Uh	-	Utility of a hit
Ne	-	Number of compounds assayed by Eve

Table of Results

Disease	Species	Target	Number of confirmed hits	Repositioned drugs
Malaria	*Plasmodium falciparum*	DHFR	23	1
		DHFR res	1	
	Plasmodium vivax	DHFR	46	6
		DHFR res	24	
		PGK	*12*	
		NMT	*22*	1
African sleeping sickness	*Trypanosoma brucei*	DHFR	12	
		PGK	*4*	
		NMT	*23*	1
Chagas	*Trypanosoma cruzi*	DHFR	10	4
		PGK	*8*	
		NMT	*18*	1
Leishmaniasis	*Leishmania major*	DHFR	3	
Schistosomiasis	*Schistosoma mansoni*	DHFR	1	
		PGK	*5*	
		NMT	*5*	1
Bacterial infection	*Staphylococcus aureus*	DHFR	Awaiting analysis	

✔ T. brucei (10 uM), TDC<0.3
✔ T. brucei (10 uM), TDC<0.5
✔ T. brucei (≤1 uM), TDC<0.5
✔ T. brucei (10 uM)
✔ T. brucei (≤1 uM)
TDC <0.3
TDC <0.4
TDC <0.5
not validated
not tested





Planned Extensions

- Use automation to synthesise assays.
- Use automation to synthesise new compounds to test QSARs.
- These two steps would enable full automation of early stage drug design.

Formalising Science

Formalisation of Science

- The goal of science is to increase our knowledge of the natural world through the performance of experiments.
- This knowledge should be expressed in *formal logical languages*.
- Formal languages promote semantic clarity, which in turn supports the free exchange of scientific knowledge and simplifies scientific reasoning.

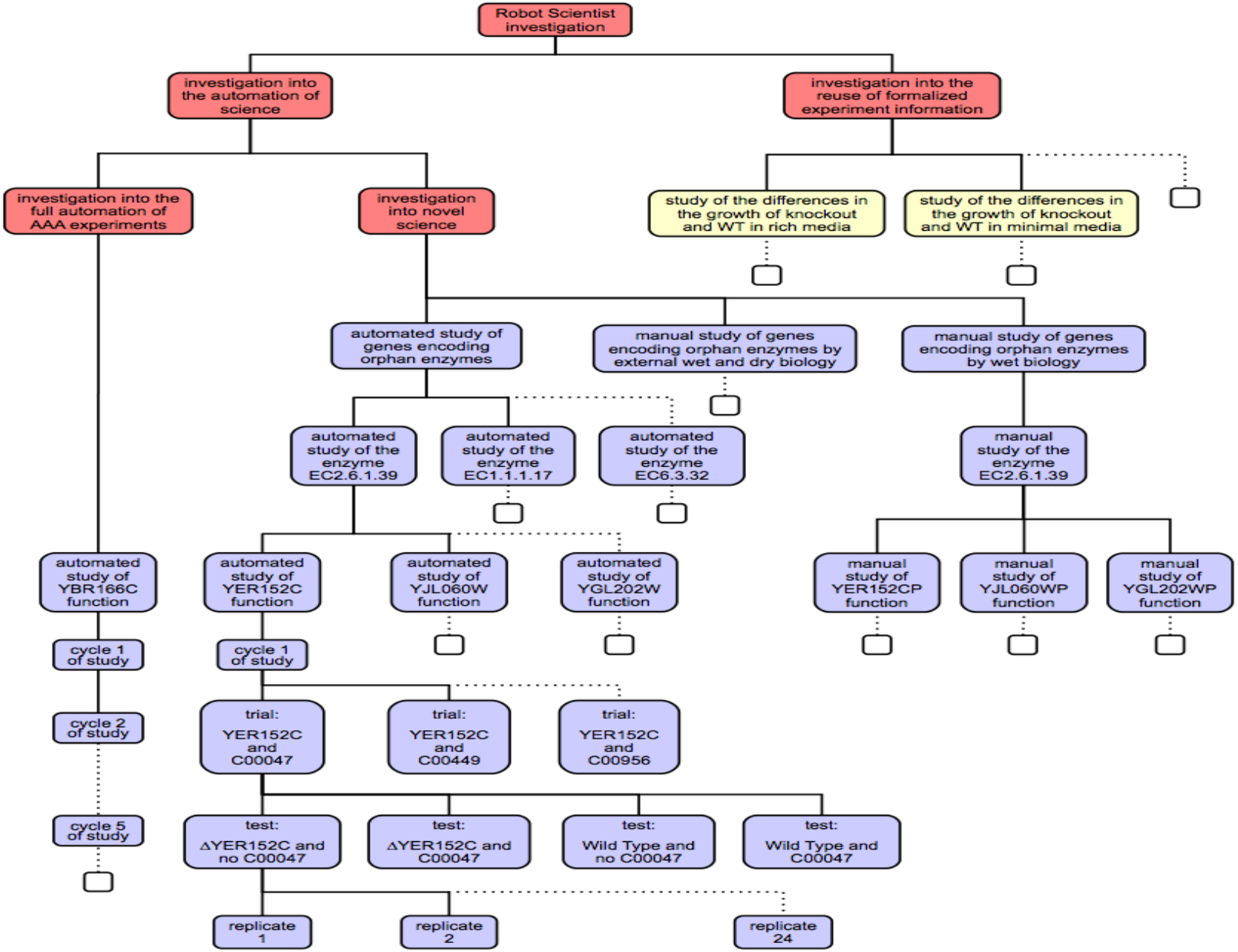
Robot Scientist & Formalisation

- Robot Scientists provide excellent test-beds for the development of methodologies for formalising science.
- Using them it is possible to completely capture and digitally curate all aspects of the scientific process.
- The ontology LABORS is designed to enable the open access of the Robot Scientist experimental data and metadata to the scientific community.

Soldatova *et al*. (2006) *Bioinformatics*

Adam's Investigations

- This formalisation involves >10,000 different research units in a nested tree-like structure 11 levels deep.
- It logically connects >6.6 million $OD600_{nm}$ measurements to hypotheses, experimental goals, results, etc.
- No previous large-scale experimental work has been so comprehensively described and recorded.


Robot Scientist investigation
investigation into the automation of science
investigation into the reuse of formalized experiment information
investigation into the full automation of AAA experiments
investigation into novel science
study of the differences in the growth of knockout and WT in rich media
study of the differences in the growth of knockout and WT in minimal media
automated study of genes encoding orphan enzymes
manual study of genes encoding orphan enzymes by external wet and dry biology
manual study of genes encoding orphan enzymes by wet biology
automated study of the enzyme EC2.6.1.39
automated study of the enzyme EC1.1.1.17
automated study of the enzyme EC6.3.32
manual study of the enzyme EC2.6.1.39
automated study of YBR166C function
automated study of YER152C function
automated study of YJL060W function
automated study of YGL202W function
manual study of YER152CP function
manual study of YJL060WP function
manual study of YGL202WP function
cycle 1 of study
cycle 2 of study
cycle 5 of study
cycle 1 of study
trial: YER152C and C00047
trial: YER152C and C00449
trial: YER152C and C00956
test: ΔYER152C and no C00047
test: ΔYER152C and C00047
test: Wild Type and no C00047
test: Wild Type and C00047
replicate 1
replicate 2
replicate 24

Levels in the Formalisation

Investigation into the automation of Science

Investigation into the automation of novel science

Investigation into the automated discovery of genes encoding orphan enzymes

Automated study of E.C.2.6.1.39 encoding

Cycle 1 of automated study of YER152C function

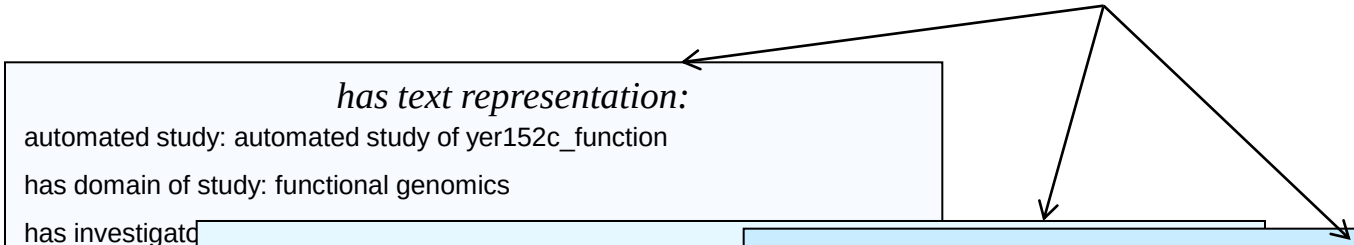
YER152C and Lysine automated trial

Experiment 1 (wild-type no metabolite)

Replicate 1 (well)

Observation 1

automated study of yer152c function



has text representation:

automated study: automated study of yer152c_function

has domain of study: functional genomics

has investigato

has goal: 'To te
with enzyme cl

has organism

has ncbi taxo

has hypothese

has researcl

has negative

has cycle 1 of

has study resu
encodes(yer15

highest

proporti

has study cond

has datalog representation:

a:automated study(X) :- a:automated_study_(
a:hypotheses-set(X) :- a:research_hypothesis
a:cycle_of_study(X) :- a:cycle_1_of_study_(X
a:hypotheses-set(X) :- a:negative_hypothesis
a:domain_of_study(Y) :- a: automated study(
a:investigator(Y) :- a: automated study(X), a:h
a:goal(Y) :- a: automated study(X), a:has_goa
a:organism_of_study (Y) :- a: automated stud
a:hypotheses-set(Y) :- a: automated study(X)
a:cycle_of_study(Y) :- a: automated study(X),
a:study_result(Y) :- a: automated study(X), a:
a:study_conclusion(Y) :- a: automated study(
a:domain_of_study(X) :- a:functional_genomi
a: investigator(X) :- a:adam.
a:goal(X) :- a: to_test_the_hypothesis_that_g
_encodes_an_enzyme_with_enzyme_class_
a:organism_of_study(X) :- a:saccharomyces_
a:study_result(X) :- a:the_strength_of_eviden
a:study_conclusion(X) :- a:hypothesis_1_con

has OWL representation:

```
<?xml version="1.0"?>
<rdf:RDF
   xmlns="http://www.owl-ontologies.com/Ontology1204198571.owl#"
<owl:Class rdf:ID="goal"/>
  <owl:Class rdf:ID="study_result"/>
  <owl:Class rdf:ID="ncbi_taxonomy_ID"/>
  <owl:Class rdf:ID="cycle_of_study"/>
  <owl:Class rdf:ID="negative_hypothesis">
    <rdfs:subClassOf>
      <owl:Class rdf:ID="hypotheses-set"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID="domain_of_study"/>
  <owl:Class rdf:ID="organism_of_study"/>
  <owl:Class rdf:ID="cycle_1_of_study_">
    <rdfs:subClassOf rdf:resource="#cycle_of_study"/>
  </owl:Class>
  <owl:Class rdf:ID="automated_study">
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:someValuesFrom rdf:resource="#goal"/>
        <owl:onProperty>
          <owl:ObjectProperty rdf:ID="has_goal"/>
        </owl:onProperty>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:someValuesFrom rdf:resource="#organism_of_study"/>
        <owl:onProperty>
          <owl:ObjectProperty rdf:ID="has_organism_of_study"/>
……..
```

Future Prospects

The Future?

- In chess there is a continuum of ability from novices up to Grandmasters.
- We argue that this is also true in science, from the simple research of Adam/Eve, through what most human scientists can achieve, up to the ability of a Newton or Einstein.
- If you accept this, then just as in chess, it is likely that advances in computer hardware and software will drive the development of ever smarter Robot Scientists.
- In favour of this argument are the ongoing development of AI and laboratory robotics.

Vision

- The collaboration between Human and Robot Scientists will produce better science than either can alone – human/computer teams still play better chess that either alone.
- Scientific knowledge will be primarily expressed in logic with associated probabilities and published using the Semantic Web.
- The improved productivity of science leads to societal benefits: better food security, better medicines, etc.

In a 100 years?

- The Physics Nobel Frank Wilczek is on record as saying that in 100 years' time the best physicist will be a machine.
- A key point about Robot Scientists is that their abilities can be objectively tested.
- So in a 100 years we will know is there are Robot Scientists doing world-class research or not.
- Time will tell.

Conclusions

- Science is a wonderful application area for AI.
- Automation is becoming increasingly important in scientific research e.g. DNA sequencing, drug design.
- The Robot Scientist concept represents the logical next step in scientific automation.
- The Robot Scientist Adam is the first machine to have discovered novel scientific knowledge.
- The Robot Scientist Eve is now finding lead compounds for neglected tropical diseases.
- Scientific knowledge should be expressed using logic.

Acknowledgments

ABERYSTWYTH / MANCHESTER

Wayne Aubrey

Amanda Clare

Douglas Kell

Maria Liakata

Chuan Lu

Magda Markham

Katherine Martin

Ronald Pateman

Jem Rowland

Andrew Sparkes

Larisa Soldatova

Mike Young

Ken Whelan

CAMBRIDGE

Steve Oliver

Elizabeth Bilsland

Pınar Pir

Harrt Moss

Michael de Clare

Mark Carrington

LEUVEN

Kurt De Grave

Luc De Raedt

Jan Ramon

Support from BB/F008228/1 from the UK Biotechnology & Biological Sciences Research Council and a contract from the European Commission under the FP7 Collaborative Programme, UNICELLSYS.

Parasites targeted

- Organism/disease:
  - *Plasmodium falciparum* (malaria),
  - *Plasmodium vivax* (malaria),
  - *Trypanosoma brucei* (sleeping sickness),
  - *Trypanosoma cruzi* (Chagas),
  - *Leishmania major* (leishmania),
  - *Schistosoma mansoni* (shistosomaisis),
  - Staphylococcus aureus
  - …

Abduction for Hypothesis Generation

Deduction

Rule: If a cell grows then it can synthesise tryptophan.

Fact: cell cannot synthesise tryptophan

$\therefore$ Cell cannot grow.

Given the rule $P \rightarrow Q$, and the fact $\neg Q$, infer the fact $\neg P$

(*modus tollens*)

Abduction

Rule: If a cell grows then it can synthesise tryptophan.

Fact: Cell cannot grow.

$\therefore$ Cell cannot synthesise tryptophan.

Given the rule $P \rightarrow Q$, and the fact $\neg P$, infer the fact $\neg Q$

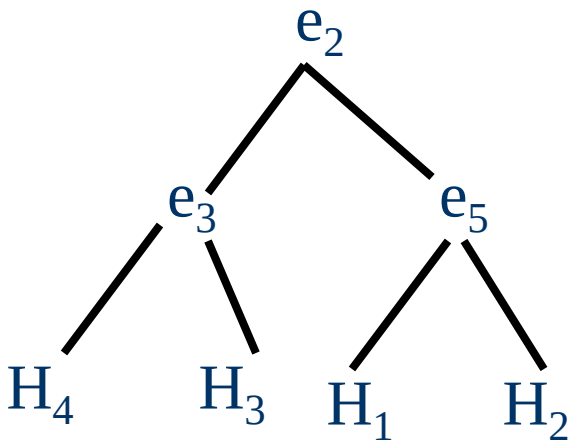
Locally Orphan Enzymes

- Adam's model of yeast metabolism has "locally orphan enzymes" these catalyse biochemical reactions known to be in yeast, but for which the coding genes are unknown.

- Adam uses bioinformatic methods to abduce genes which could encode these orphan enzymes - hypotheses.

How to choose the best experiment

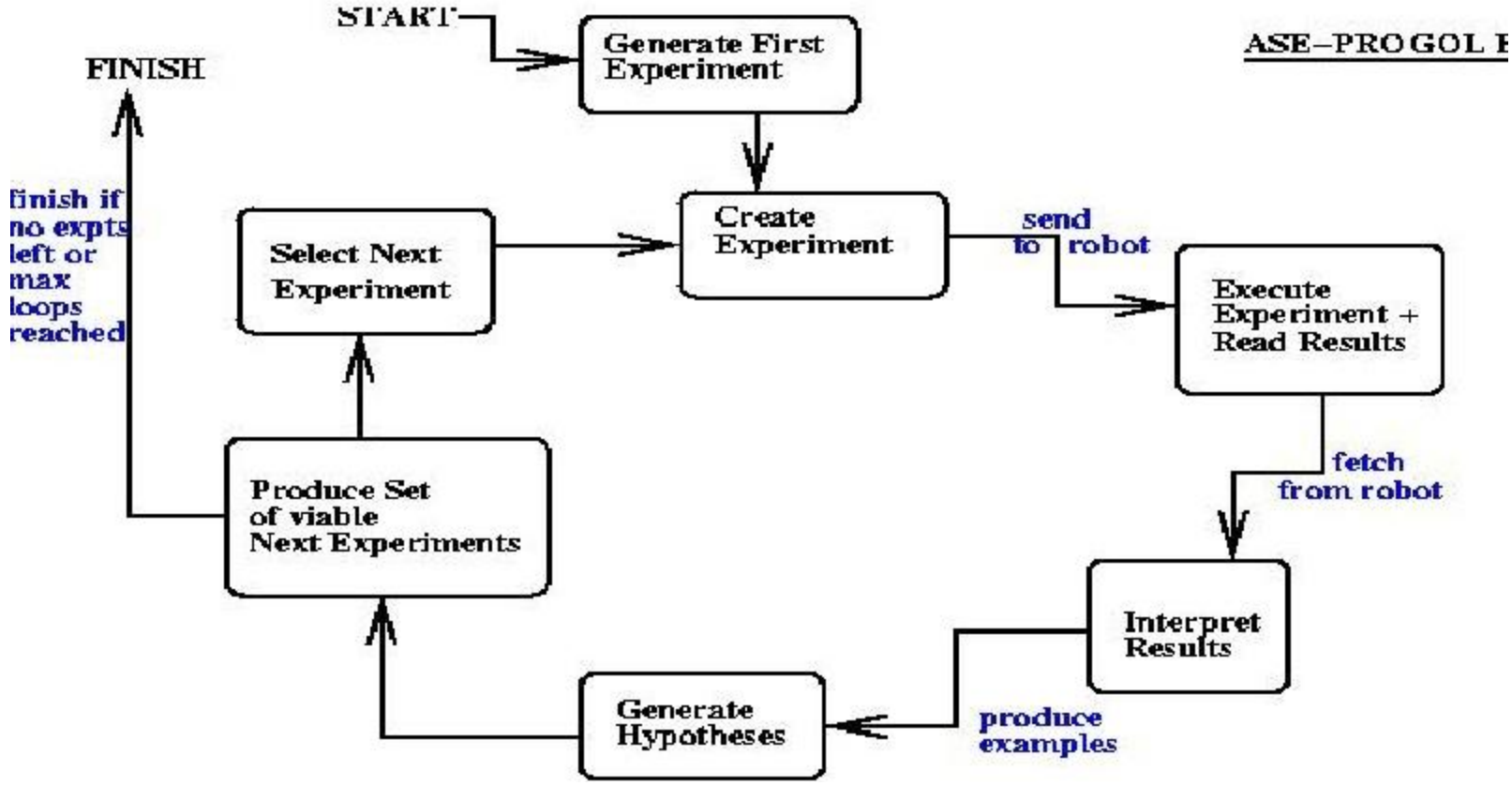
	e_1	e_2	…	e_m
H_1	T	F	…	T
…	F	T	…	F
H_n	F	T	…	T



Choosing the best experiment is equivalent to choosing the best node in a decision tree.

Bryant et al. (2001) ETAI 5, 1-36.

LIMS Setup


START
Generate First Experiment
ASE–PROGOL
FINISH
finish if no expts left or max loops reached
Select Next Experiment
Create Experiment
send to robot
Execute Experiment + Read Results
fetch from robot
Produce Set of viable Next Experiments
Interpret Results
Generate Hypotheses
produce examples

Eve in Action

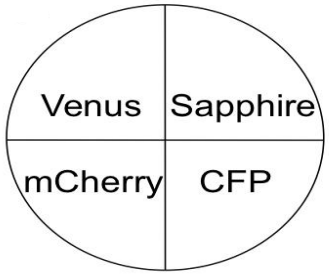
A 50 Year Old Puzzle

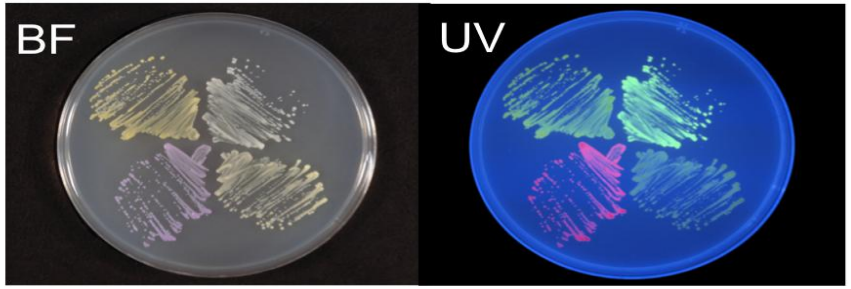
- The enzyme 2-aminoadipate: 2-oxoglutarate aminotransferase was a locally orphan enzyme.
- It is in the lysine biosynthesis pathway which has been studied for 50 years in fungi: target for antibiotics, and on path to penicillin.

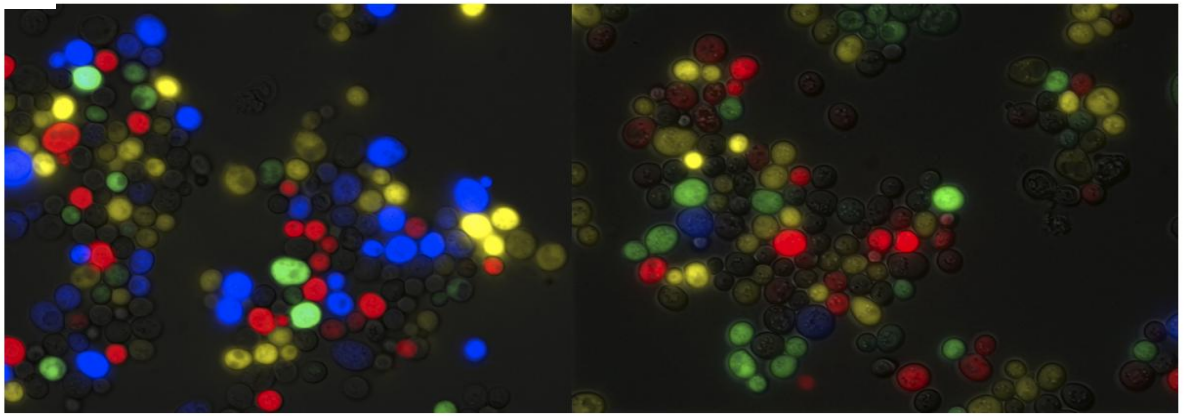
- Adam formed three hypotheses for the gene to encode this enzyme: YER152C, YJL060W, and YGL202W (in that order of probability).

- Currently KEGG states that YGL202W is the gene.
- Evidence from 1960's that at least 2 isoenzymes are involved.

Yeast Strains







No drug

Pyrimethamine

y^{Hs}DHFR+yEpmCherry

y^{Pf}DHFR+yEpSapphire

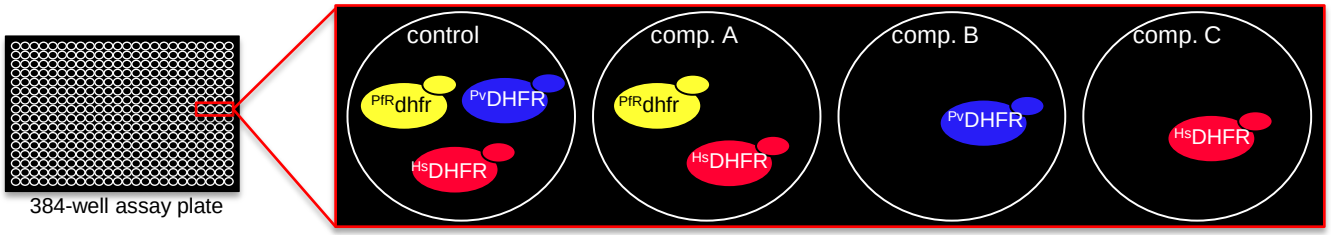
y^{PfR}dhfr+yEpVenus

y^{Sm}DHFR+yEpCFP

Confirmed New Knowledge

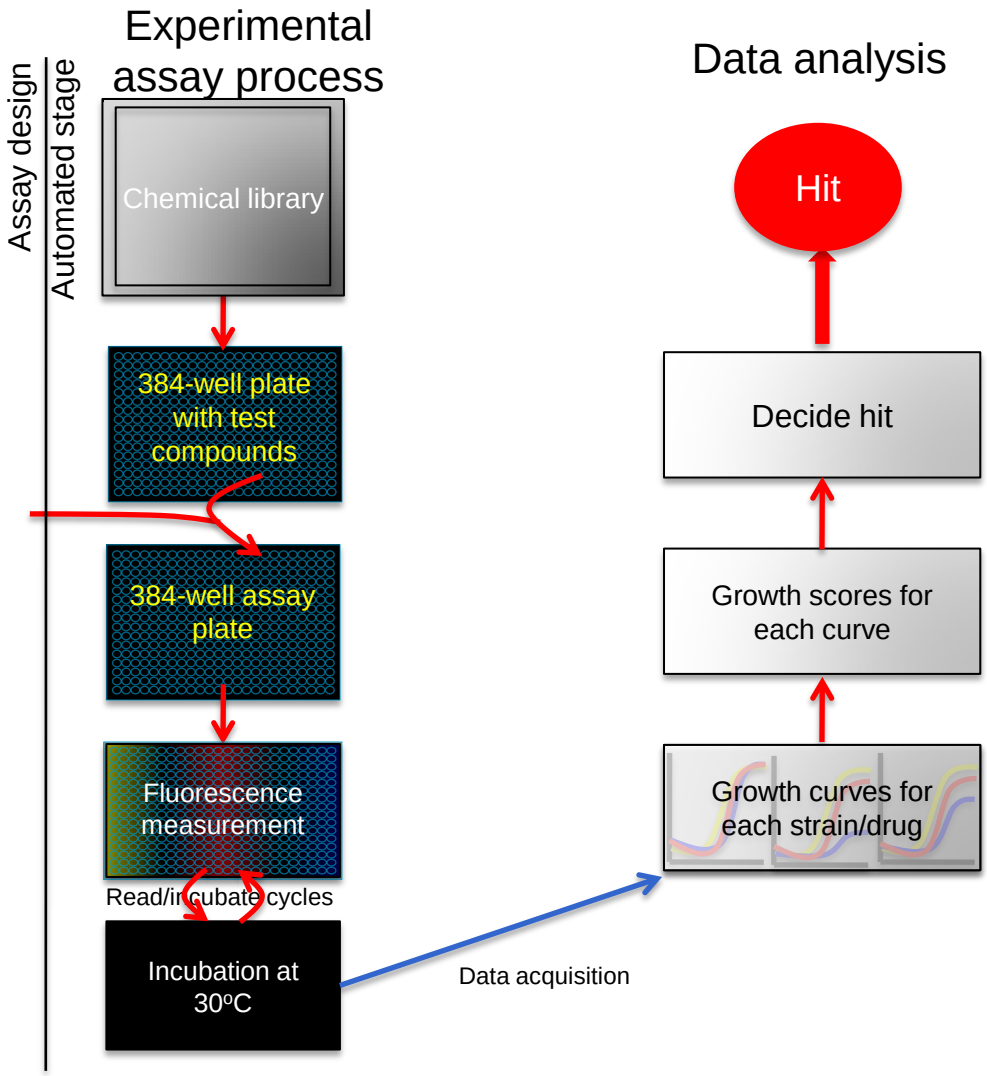
- Adam's differential growth experiments were consistent with all three genes encoding 2-oxoglutarate aminotransferase.
- Manual experiments: purified protein + enzyme assays are consistent.
  - YGL202W literature confirmed.
  - YJL060W (was annotated as an arylformamidase, new (08) annotation kynurenine aminotransferase)
  - YER152C (currently not annotated)
- YGL202W & YJL060W double knockout is lethal

A

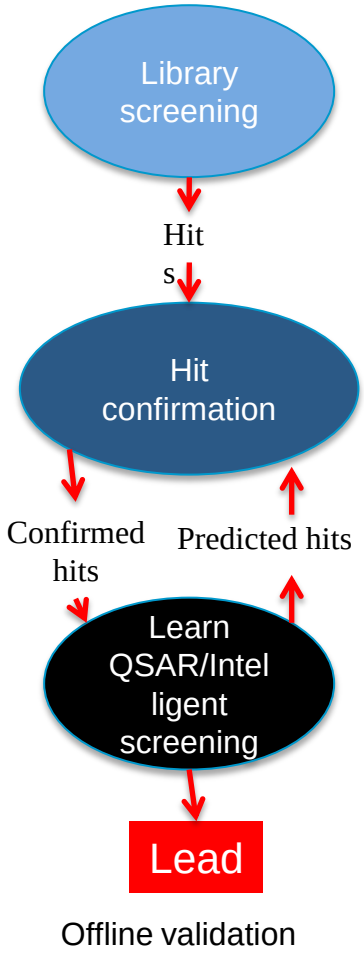


Synthetic Biology

B

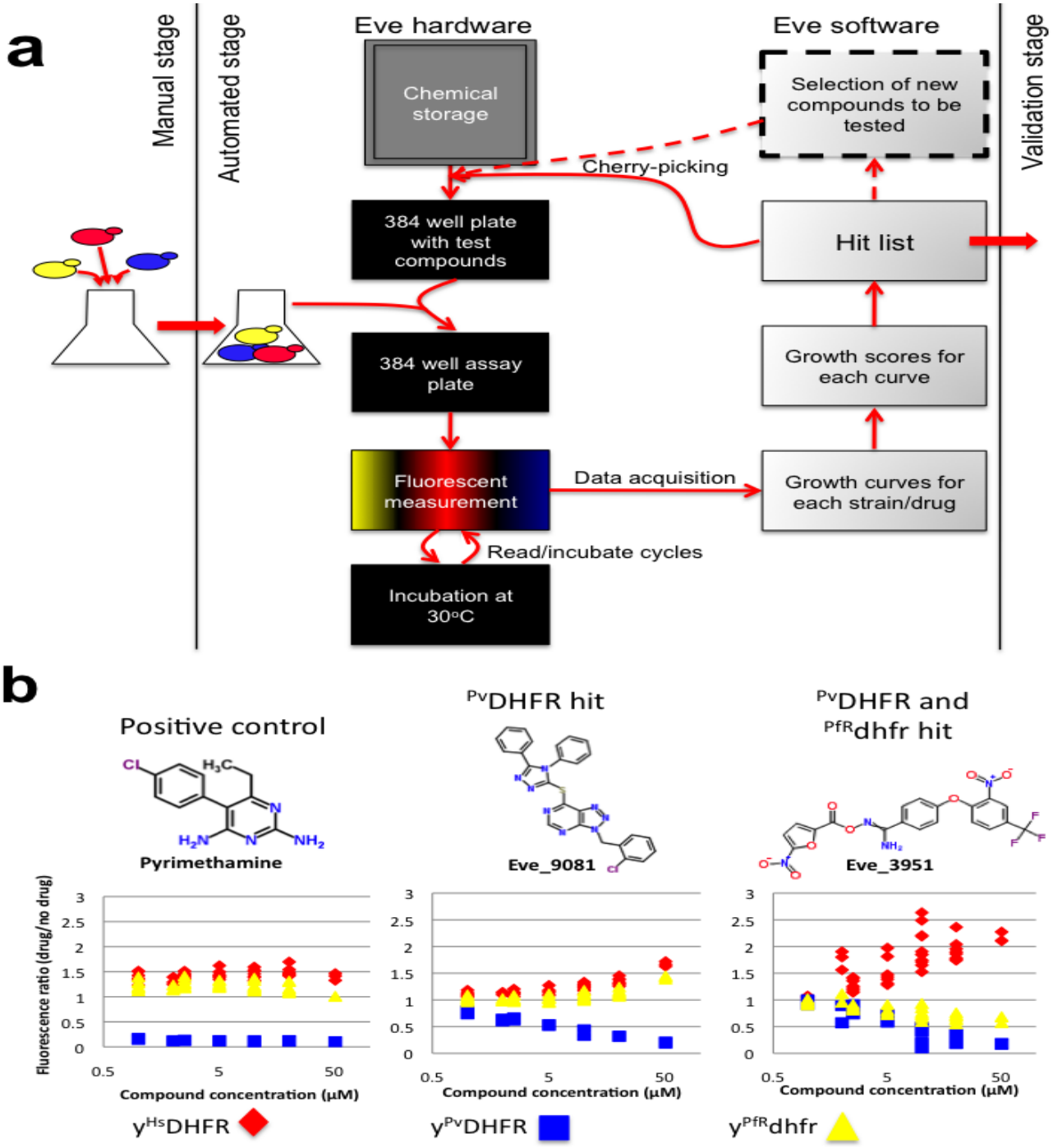


C Overview of Eve



Proces

s


a
Manual stage
Automated stage
Eve hardware
Eve software
Validation stage
Chemical storage
Selection of new compounds to be tested
Cherry-picking
384 well plate with test compounds
Hit list
384 well assay plate
Growth scores for each curve
Fluorescent measurement
Data acquisition
Growth curves for each strain/drug
Read/incubate cycles
Incubation at 30°C
b
Positive control
PvDHFR hit
PvDHFR and PfRdhfr hit
Pyrimethamine
Eve_9081
Eve_3951
Fluorescence ratio (drug/no drug)
Compound concentration (μM)
yHsDHFR
yPvDHFR
yPfRdhfr