

AAAI Spring Symposium 2023 on
Computational Approaches to Scientific Discovery

Imperial College
London



Human Comprehensible Active Learning of Genome-Scale Metabolic Networks

Lun Ai¹□, Shishun Liang², Wang-Zhou Dai³, Liam Hallett², Stephen H. Muggleton¹, Geoff S. Baldwin²

¹Department of Computing, Imperial College London, UK

²Department of Life Science, Imperial College London, UK

³School of Intelligence Science and Technology, Nanjing University, China

□lun.ai15@imperial.ac.uk

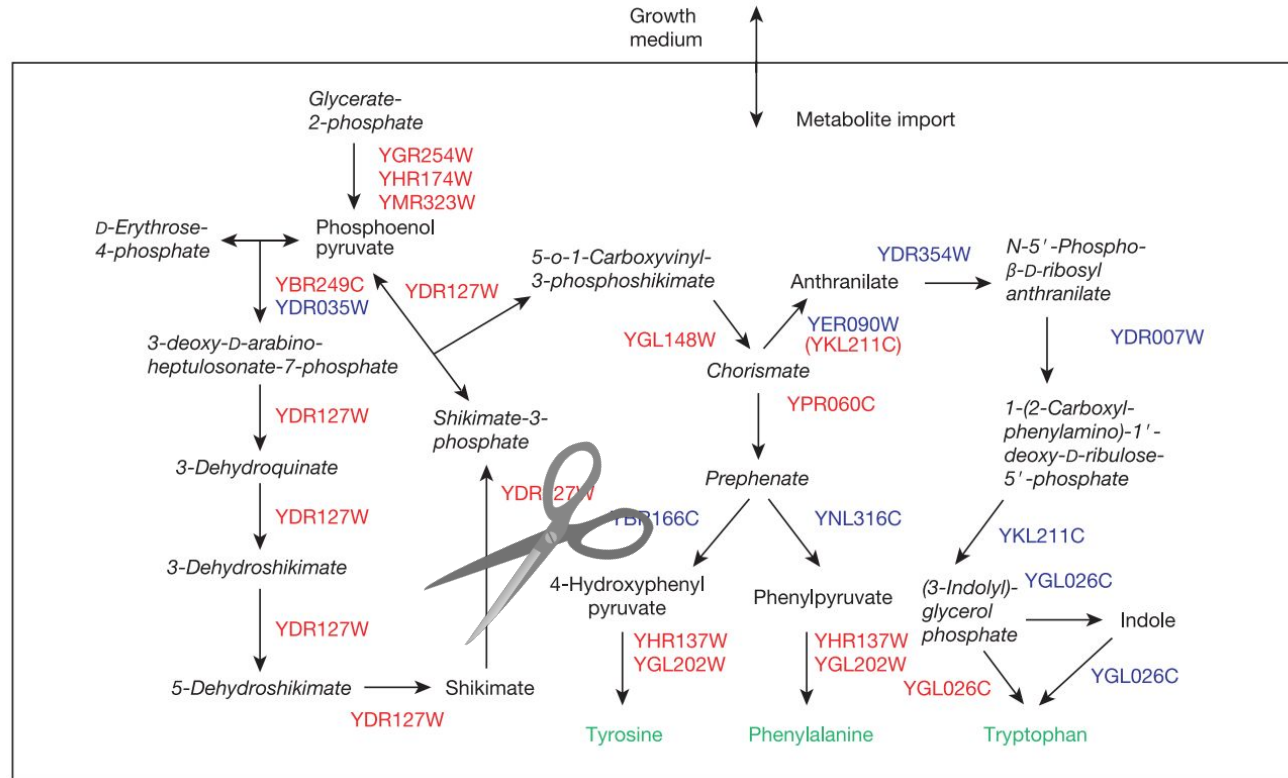
Outline

- **Scientific problem**
- Framework
- Novel matrix approach
- Results
- Summary

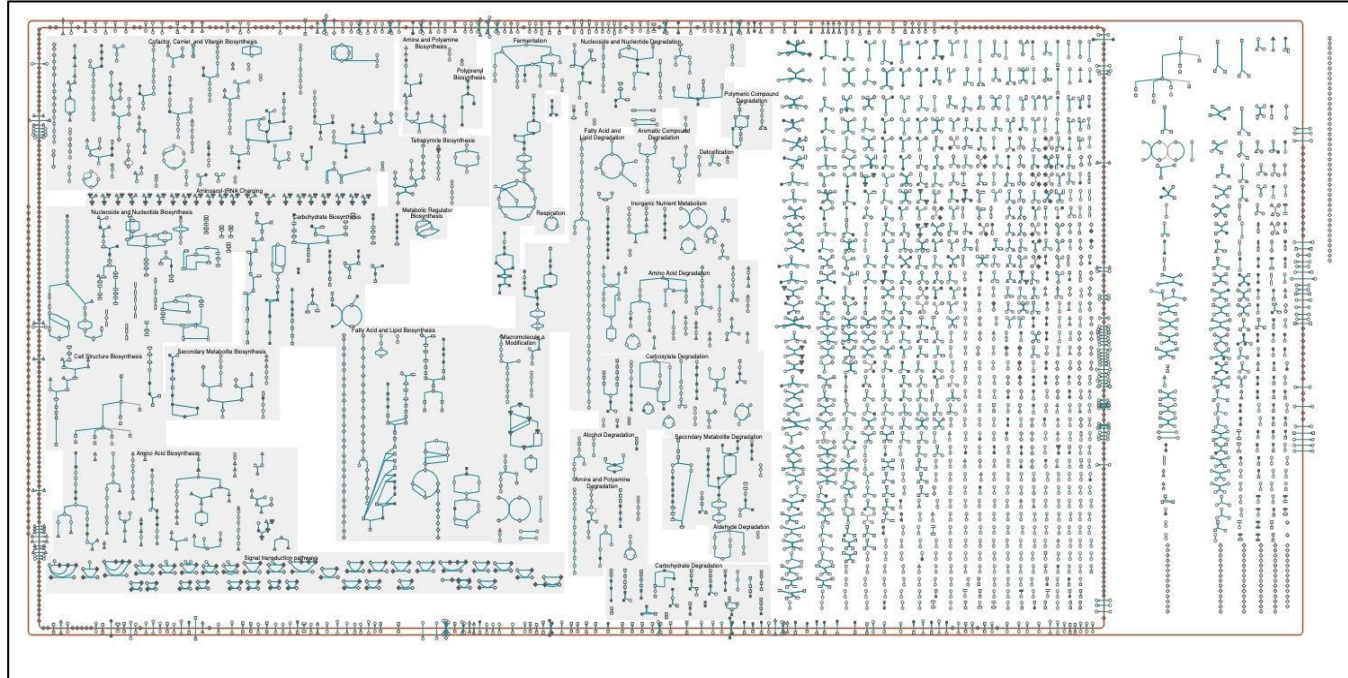
Importance to Synthetic Biology

- Aim:
 - Synthesise useful compounds
- Method:
 - Tuning for correct precursors
 - Engineering of metabolic networks
 - Examination of phenotypes

Exemplary metabolic network



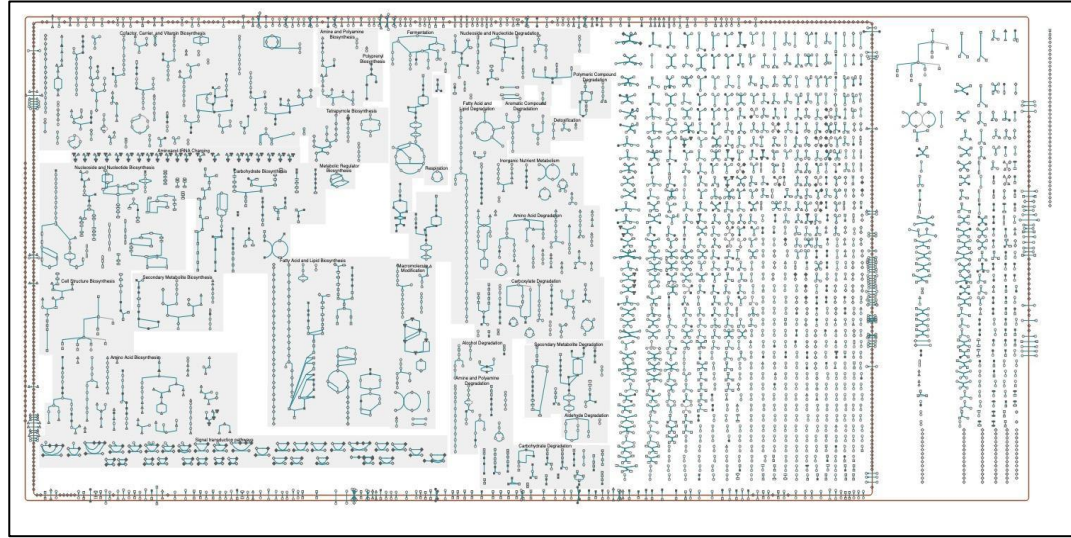
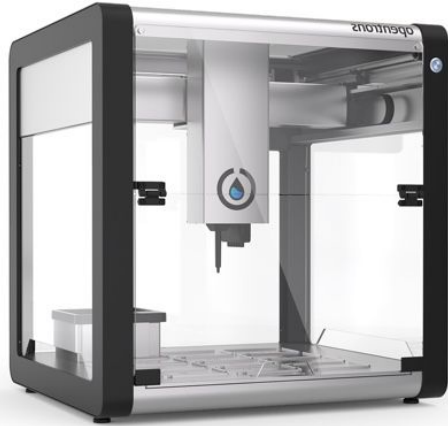
Efficiently learn and navigate genome-scale metabolic networks?



iML1515 (Monk et al. 2017), **100 times increase in model complexity**

Reduce cost and design space?

£££

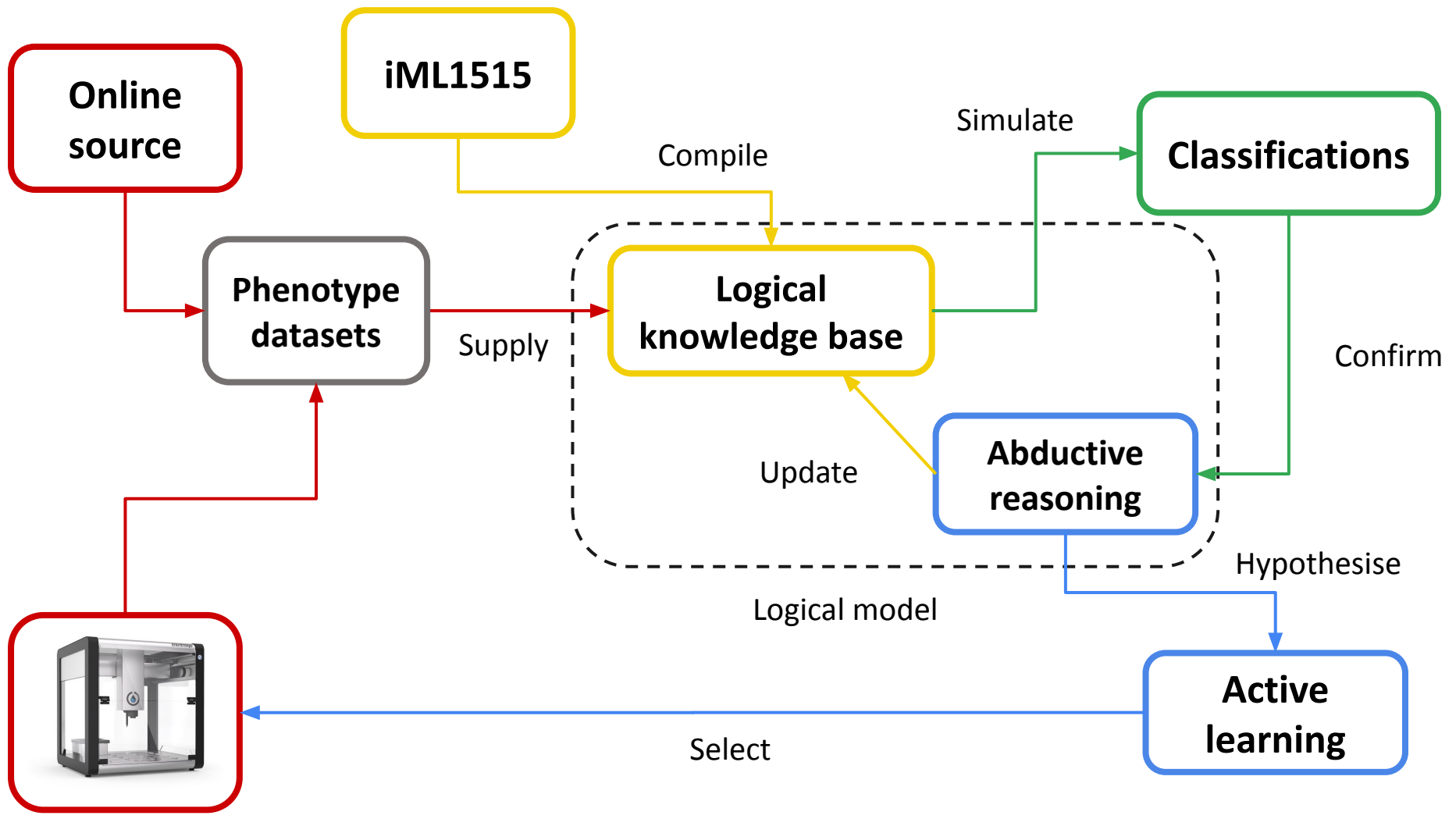


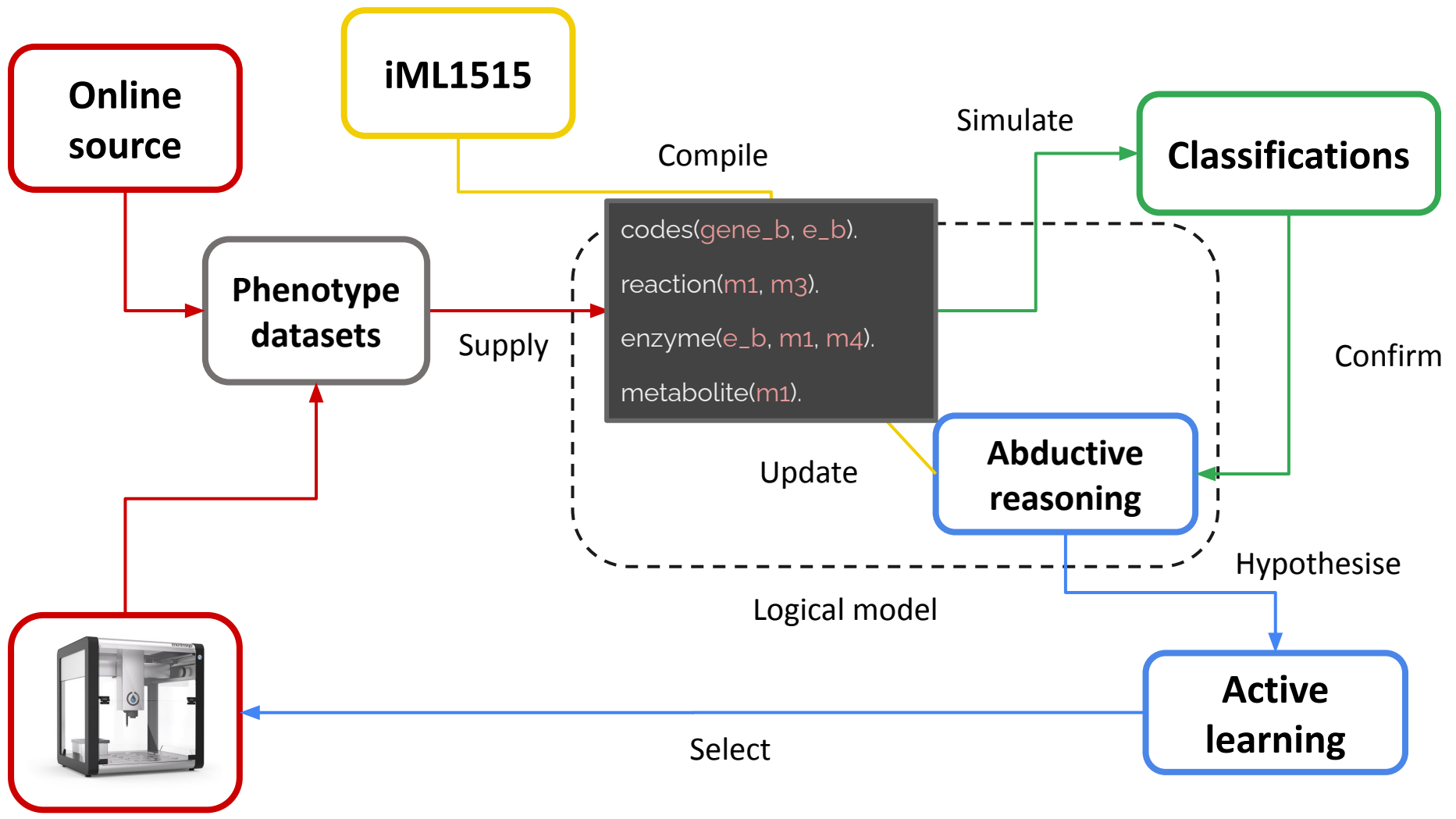
1515 genes + 2719 reactions

- Scientific problem
- **Framework**
 - **Model-Comprehend**
- Novel matrix approach
- Results
- Summary

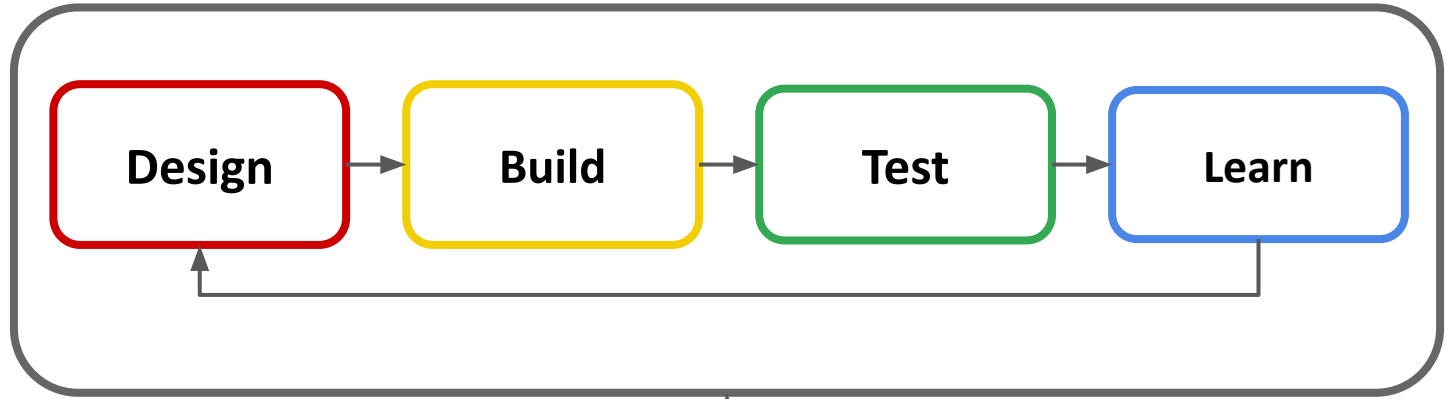
Model-Comprehend

- Automate **Design**, **Build**, **Test**, **Learn**
 - Rapid inferences
 - Hypothesis space reduction
 - Cost minimisation





Model



Comprehend

- Scientific problem
- Framework
- **Novel matrix approach**
 1. **Abduction**
 2. **Active learning**
- Results
- Summary

Abduction

$$p \leftarrow q \wedge r$$

- Phenotypes in various nutrient media
- Simulation based on metabolite saturation
- Hypothesise gene functions to explain data

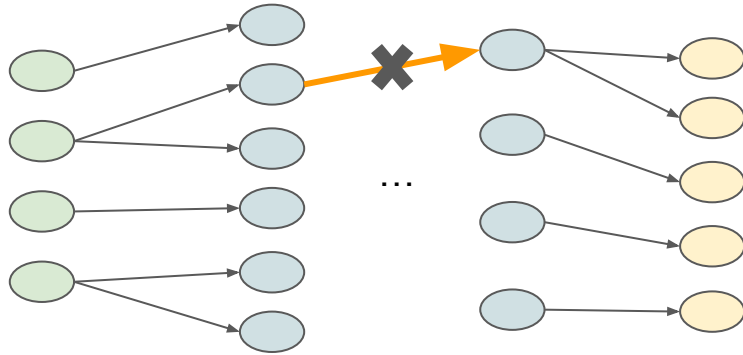
Experiment data:

`phenotypic_effect(Gene, Nutrients).`

Hypothesis:

`codes(Gene, Enzyme).`

Enzymes/reactions



Metabolite saturation

```
% description of effect using metabolic network
```

```
phenotypic_effect(Gene, Medium):-
```

```
  % abduced fact
```

```
  codes(Gene, Enzyme),
```

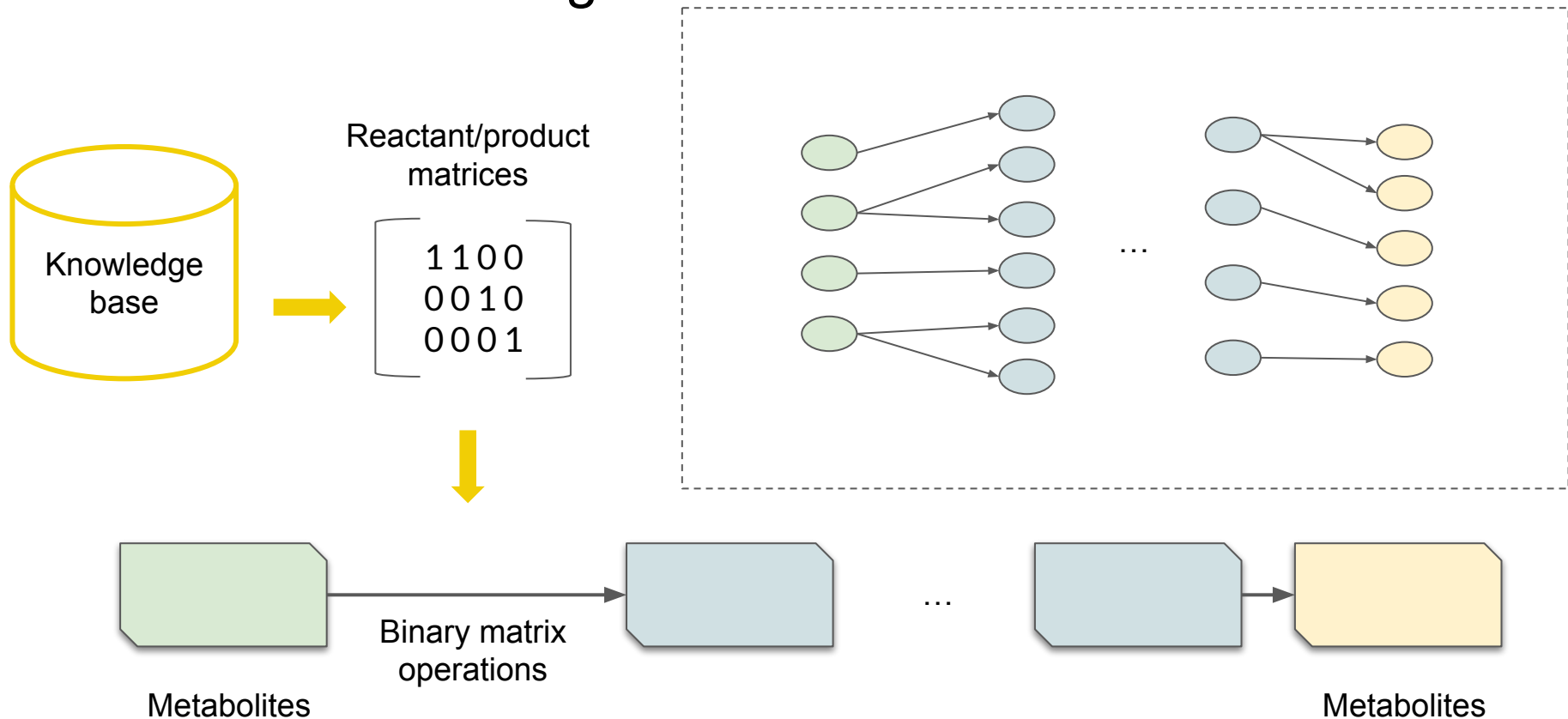
```
  cant_use_enzyme(Enzyme),
```

```
  %metabolic pathways
```

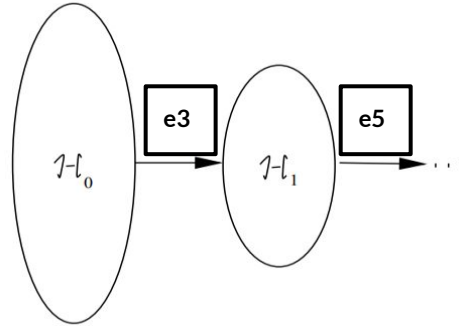
```
  metabolic_path(Medium, Metabolites),
```

```
  no_essential_molecule(Metabolites).
```

Novel matrix encoding



Active learning

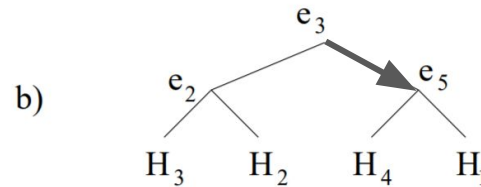


a) Binary discrimination of hypotheses

b) Approx. minimum cost binary decision tree

a)

	e_1	\dots	e_m
H_1	1	\dots	0
\dots	\dots	\dots	\dots
H_n	1	\dots	1

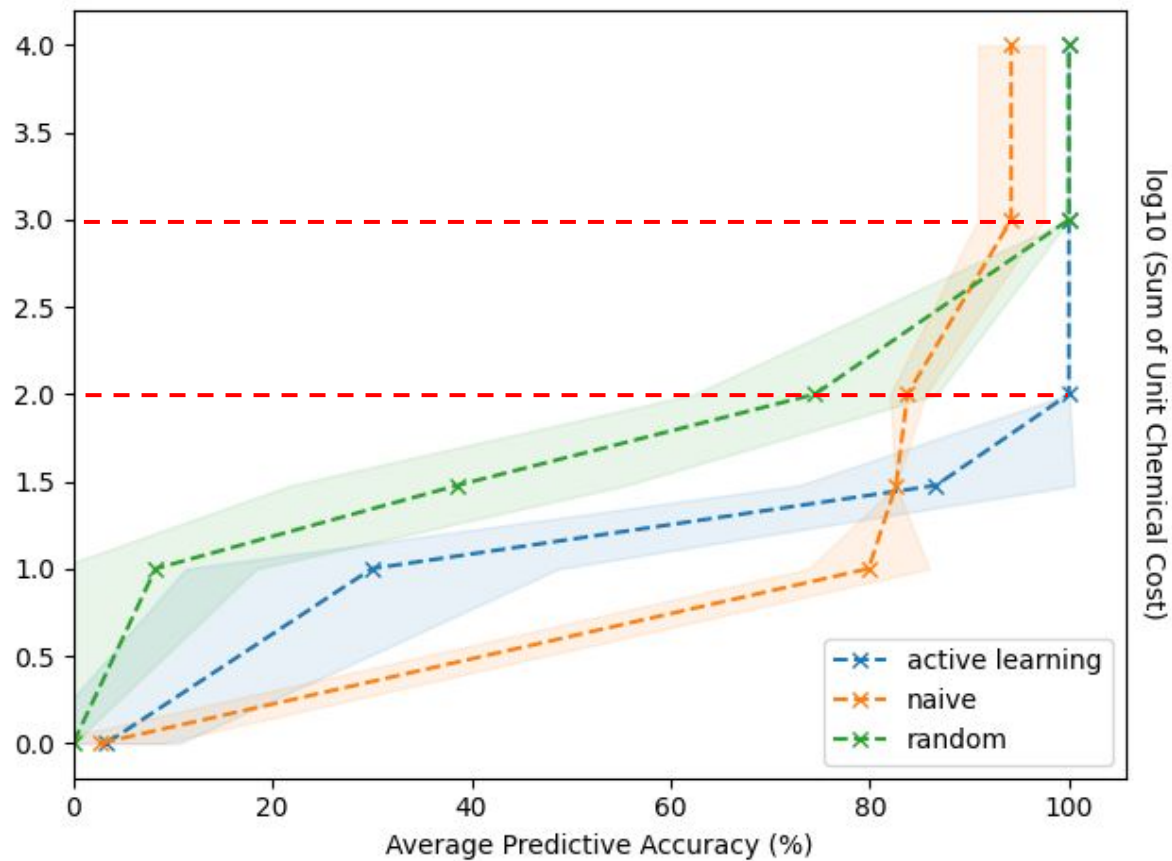


- Scientific problem
- Framework
- Novel matrix approach
- **Results**
 - **Runtime improvement**
 - **Cost reduction**
- Summary

> 4000 times better in runtime

Time per simulation	Robot Scientist (King et al. 2004)	MC
Without multi-threads	<u>≈250s</u>	≈0.6s
With multi-threads 20 cpus	≈27s	<u>≈0.06s</u>

10 times saving in cost



- Scientific problem
- Framework
- Novel matrix approach
- Results
- **Summary**

Summary

- To automate metabolic network engineering
- Overcome network complexity
 - 4000+ times better runtime
- Experimental design
 - 10 times lower cost
 - Most informative trials

Future work

- Generalisation of framework
 - Quantification
 - Multi-clause theories
- Optimisation of metabolic network
 - Validate hypotheses (CRISPRi)
 - Multiple gene loci
- Hypothesis comprehensibility (Ai et al. 2021)

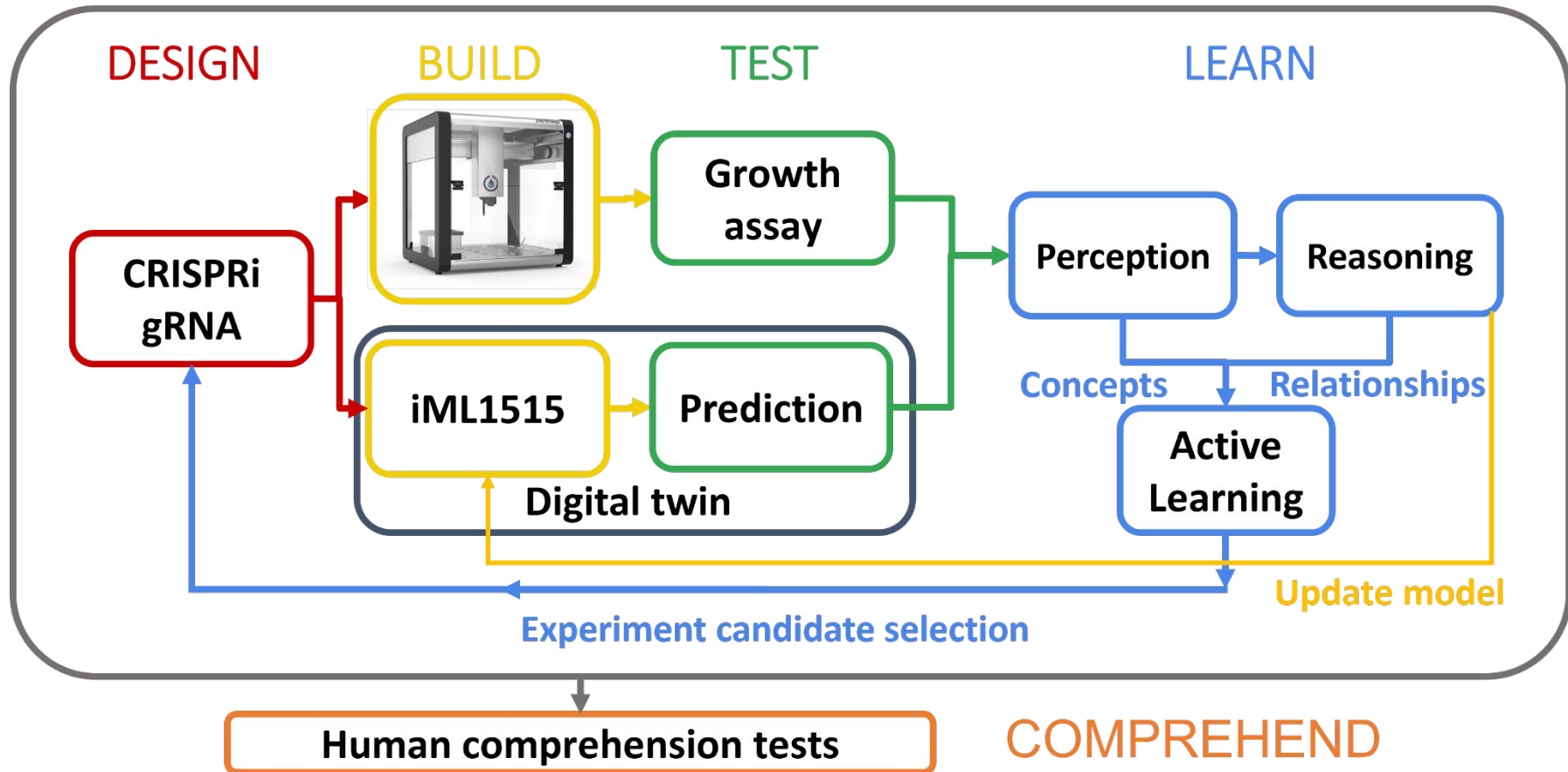
Lun Ai

Email: lun.ai15@imperial.ac.uk

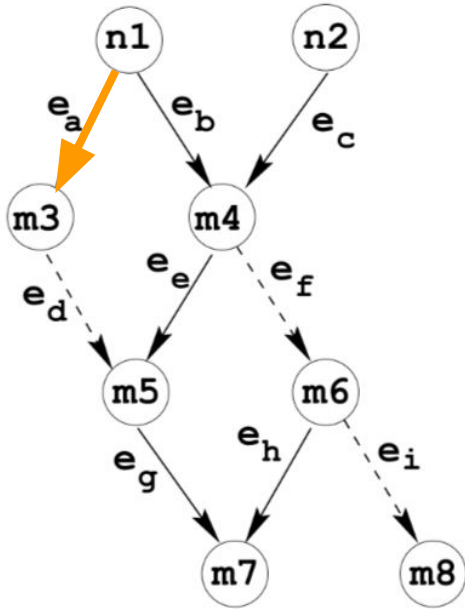
Website: <https://lai1997.github.io/>

Linkedin: <https://www.linkedin.com/in/lun-ai-46481a128/>

MODEL



Logical knowledge base



```
% description of effect using metabolic network
```

```
phenotypic_effect(Gene, Medium):-
```

```
    % abduced fact
```

```
    codes(Gene, Enzyme),
```

```
    cant_use_enzyme(Enzyme),
```

```
    %metabolic pathways
```

```
    metabolic_path(Medium, Metabolites),
```

```
    no_essential_molecule(Metabolites).
```

```
% static knowledge
```

```
codes(gene_b, e_b).
```

```
codes(gene_c, e_c).
```

```
codes(gene_e, e_e).
```

```
...
```

```
metabolic_step(n1, m3).
```

```
metabolic_step(n1, m4).
```

```
...
```

```
enzyme(e_a, n1, m3).
```

```
enzyme(e_b, n1, m4).
```

```
essential_molecule(m7).
```

```
essential_molecule(m8).
```

Four metabolites: m1, m2, m3, m4

Two orfs: g1, g2

Initial metabolic state: m1, m2, m3

Representation: [1, 1, 1, 0]

Logic encoding: mstate(0, 14)

Reactions:



Genes:

codes(g1, enz1)

codes(g2, enz2)

Reactant matrix (Rm)

enz1: 1,1,0,0

enz2_f: 0,0,1,0

enz2_b: 0,0,0,1

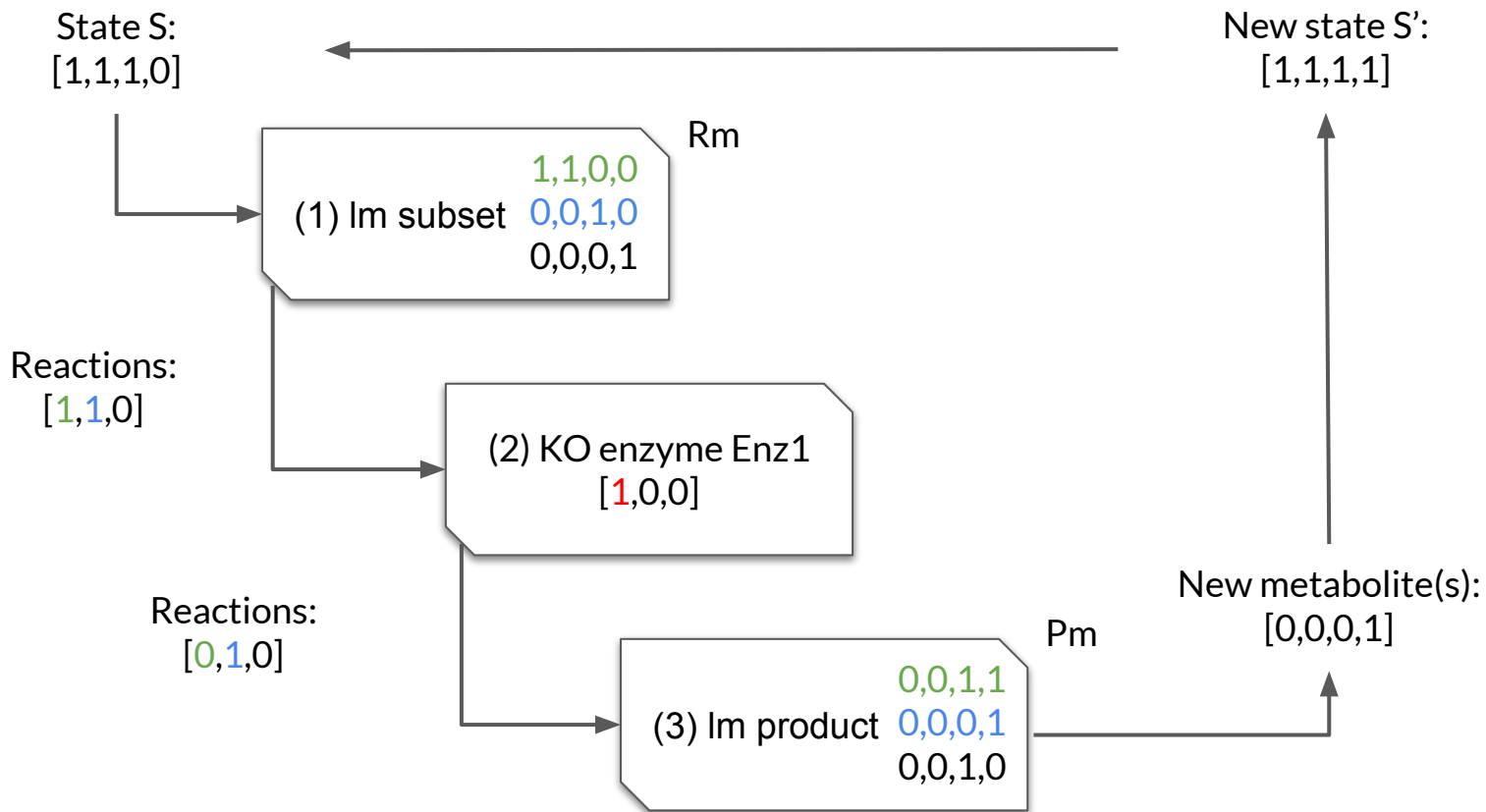
Product matrix (Pm)

enz1: 0,0,1,1

enz2_f: 0,0,0,1

enz2_b: 0,0,1,0

KO g1



KO g1 + g2

