



A large-scale exploration of cell-free compositions to maximize protein production using active learning

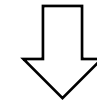
Olivier Borkowski^{*}, Mathilde Koch^{*}, Agnès Zettor, Amir Pandi, Angelo Cardoso Batista, Paul Soudier, Jean-Loup Faulon



Cell-free composition: A vast combinatorial space to explore

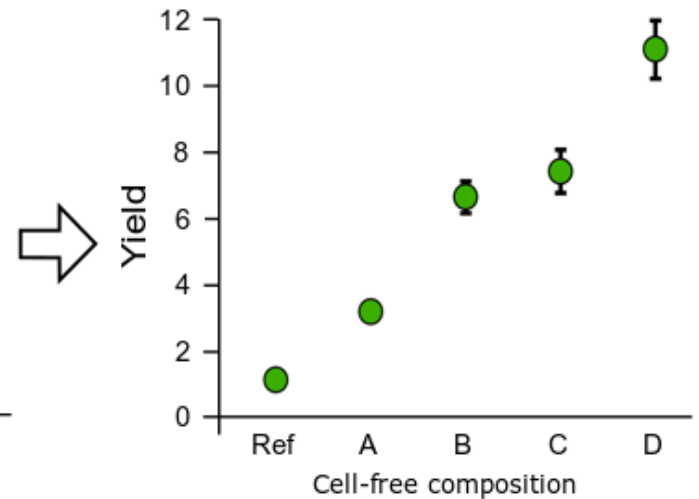
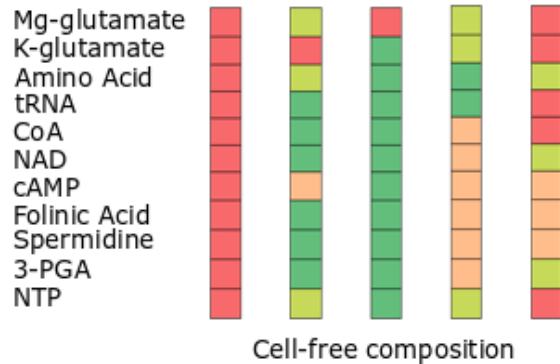
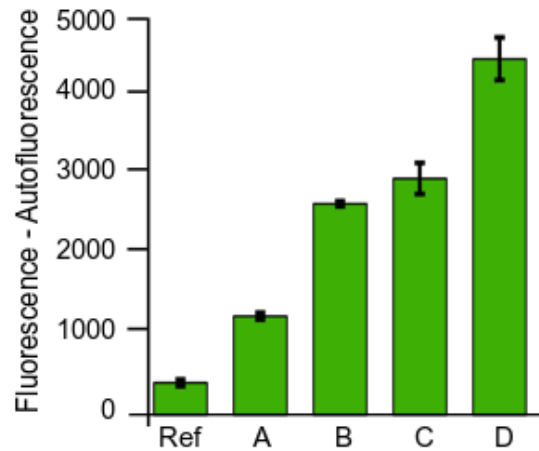
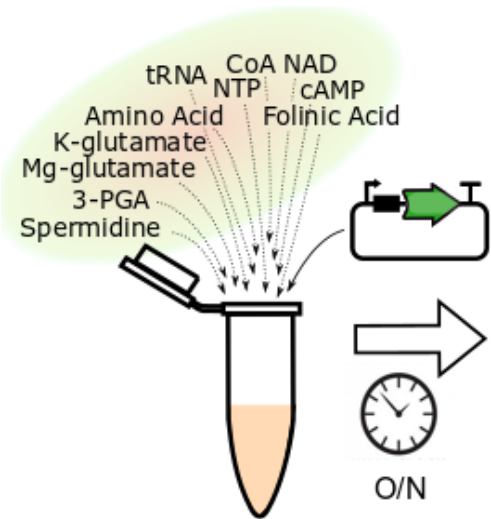
Component	Concentration			
Mg-glutamate (mM)	0.4	1.2	2	4
K-glutamate (mM)	8	24	40	80
Amino Acid (mM)	0.15	0.45	0.75	1.5
tRNA (mg/ml)	0.02	0.06	0.1	0.2
CoA (mM)	0.026	0.078	0.13	0.26
NAD (mM)	0.033	0.099	0.165	0.33
cAMP (mM)	0.075	0.225	0.375	0.75
Folinic Acid (mM)	0.0068	0.0204	0.034	0.068
Spermidine (mM)	0.1	0.3	0.5	1
3-PGA (mM)	3	9	15	30
NTP (mM)	0.15	0.45	0.75	1.5

Combinatorial space = $4^{11} = 4\,194\,304$ compositions



- Can we improve protein production for a specific lysate without increasing the price of cell-free reaction?
- Can we provide efficient predictions of protein production?
- Can we highlight the critical parameters involved in protein production?

A relative measure of efficiency: Yield

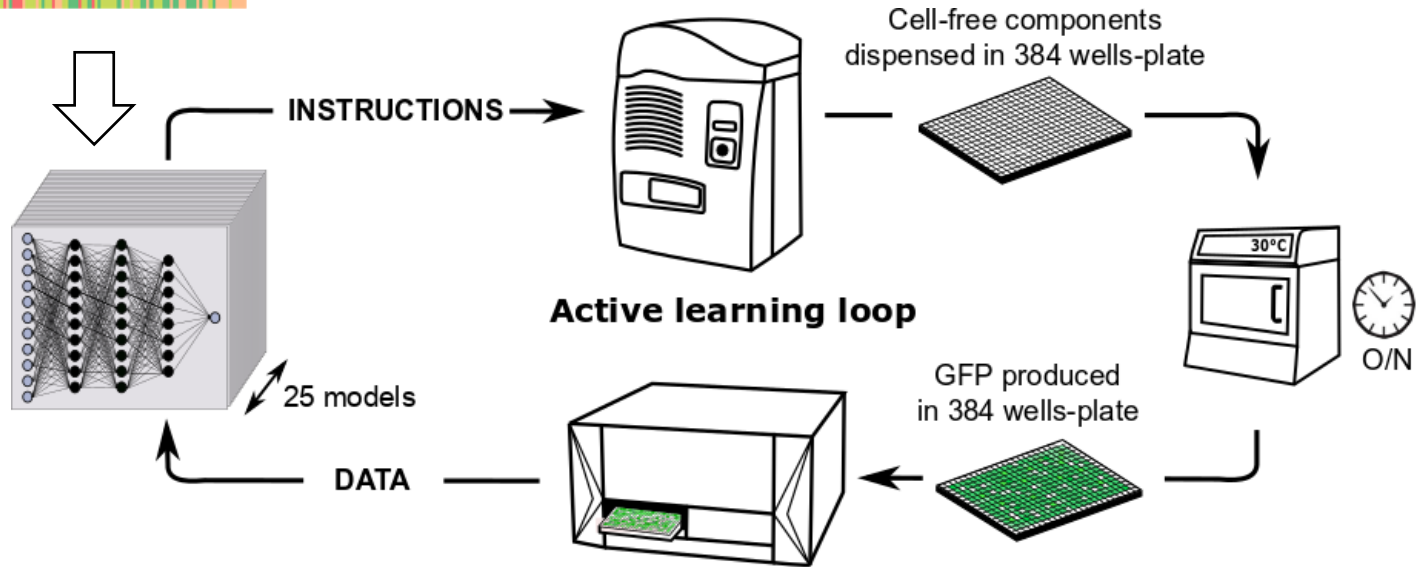
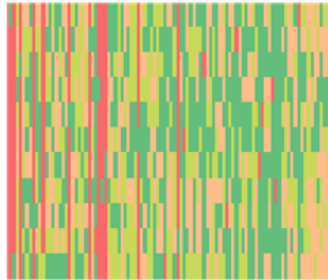


$$\text{Yield} = \frac{(\text{Fluorescence}_x - \text{Autofluorescence})}{(\text{Fluorescence}_{\text{ref}} - \text{Autofluorescence})}$$

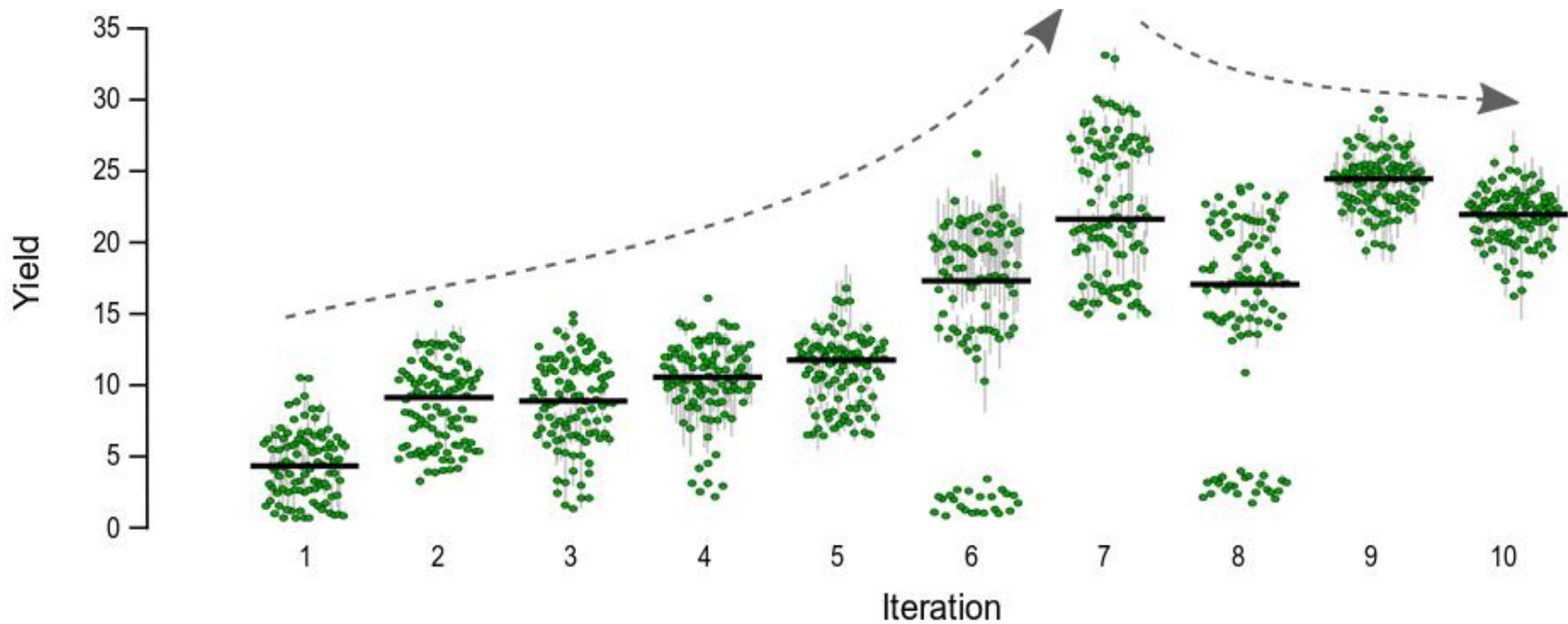
Exploration using active learning

Initial training set
102 Compositions

Mg-glutamate
K-glutamate
Amino Acid
tRNA
CoA
NAD
cAMP
Folic Acid
Spermidine
3-PGA
NTP

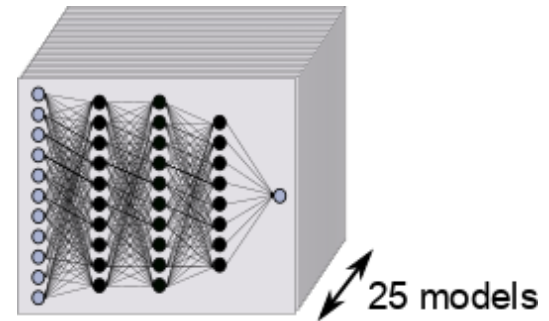
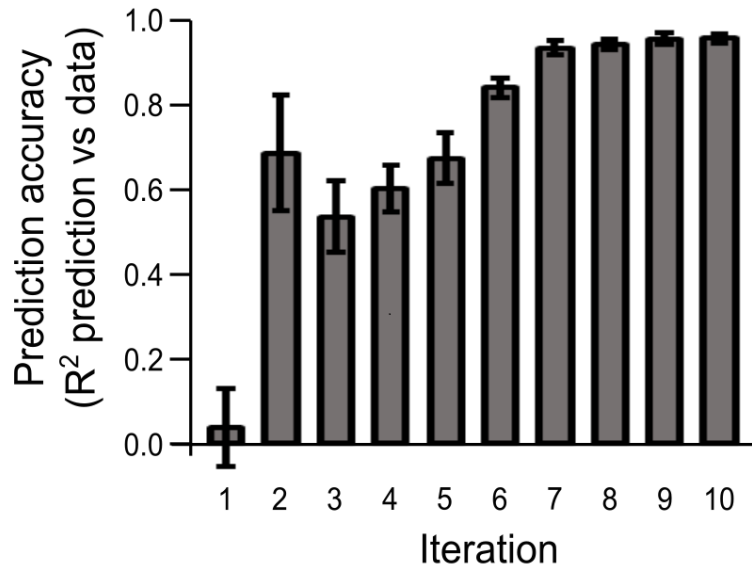


A 34 time Yield improvement using an home-made lysate

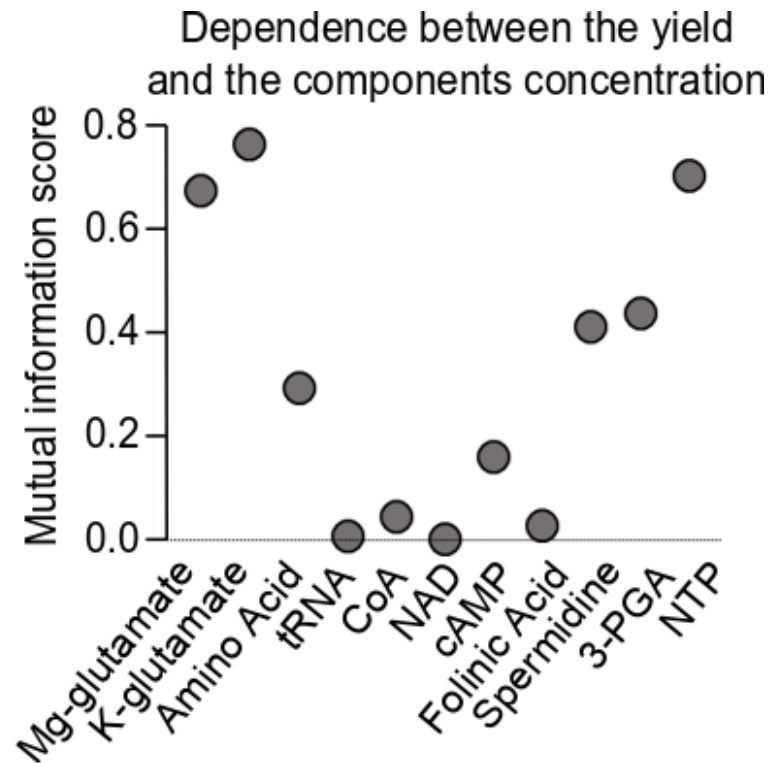


Prediction accuracy of the models

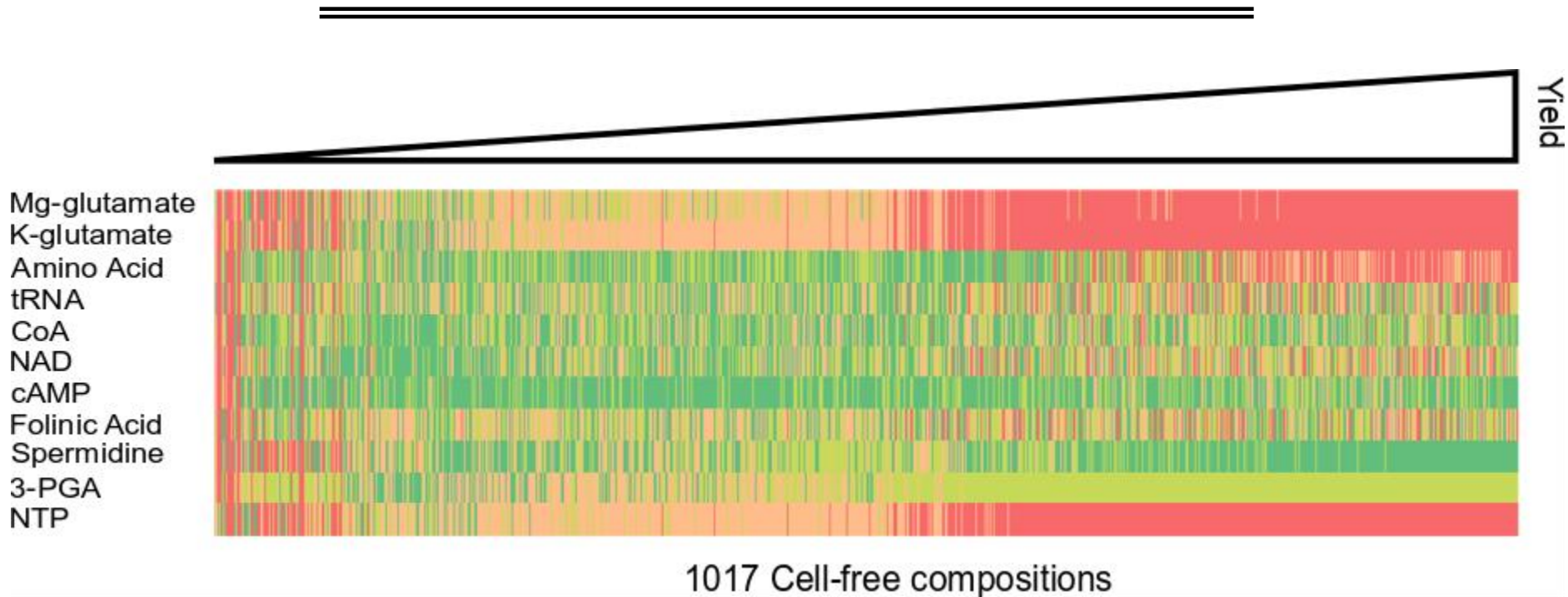
5-Fold cross validation



Impact of each compound on cell-free efficiency



Impact of each compound on cell-free efficiency



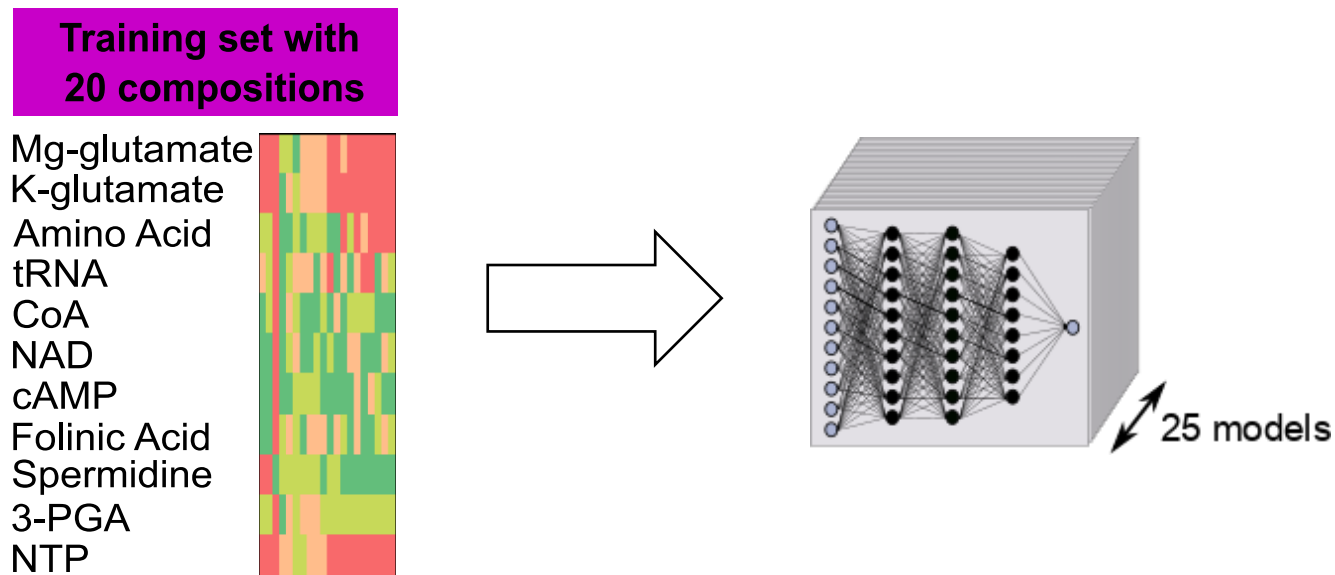
⇒ High concentration improve yield: Mg-glutamate, K-glutamate, Amino Acid and NTP

High concentration decrease yield: Spermidine, 3-PGA

A one-step method to maximize Yield in cell-free using home-made lysates

Lysate quality is user-dependent, strain-dependent...:

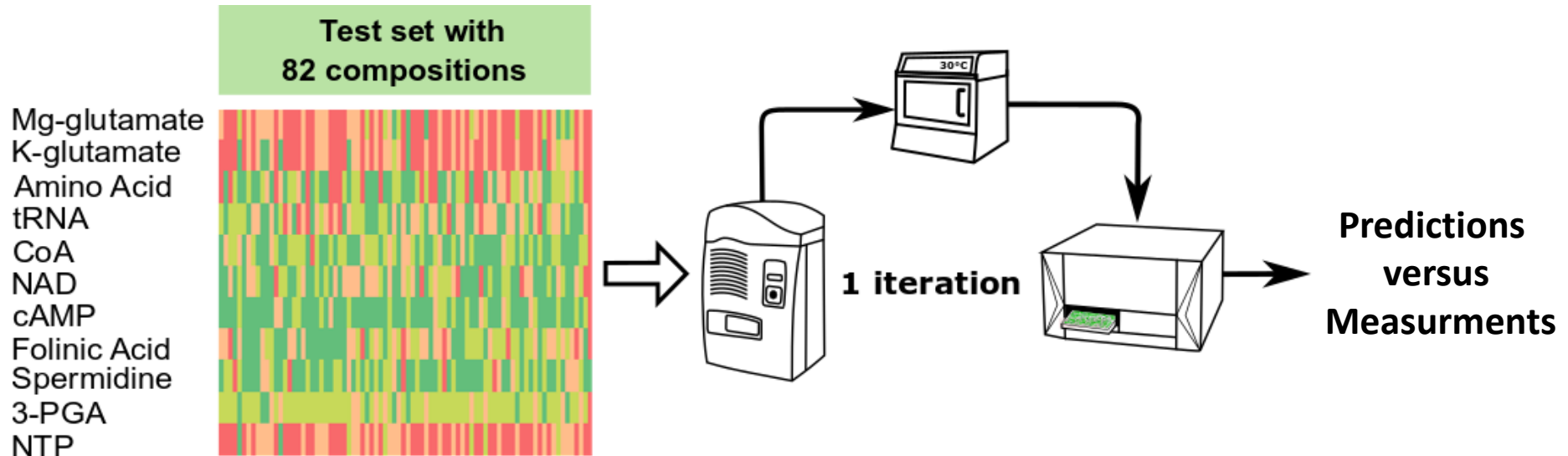
Can we provide a quick method to maximize protein production with any lysate?



Training a machine learning algorithm with the most informative compositions

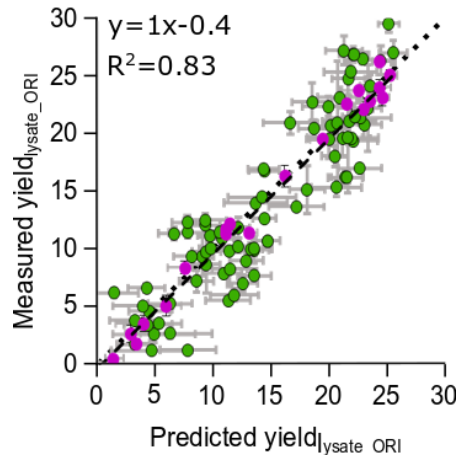
<https://github.com/brsynth>

Prediction of the Yields in cell-free based on home-made lysates

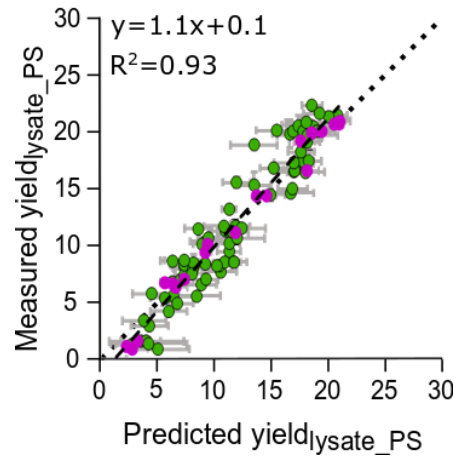


Predictions efficiencies (different experimentalists)

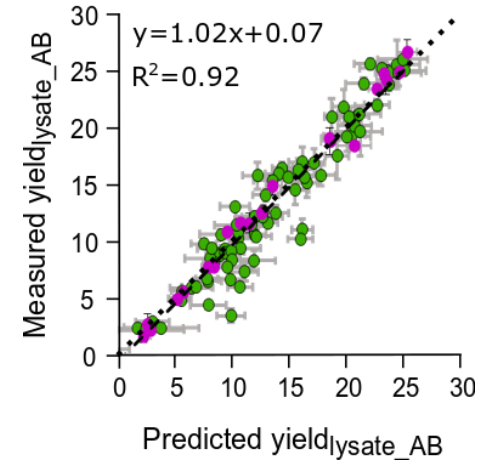
Lysate BL21 Ori



Lysate BL21 PS



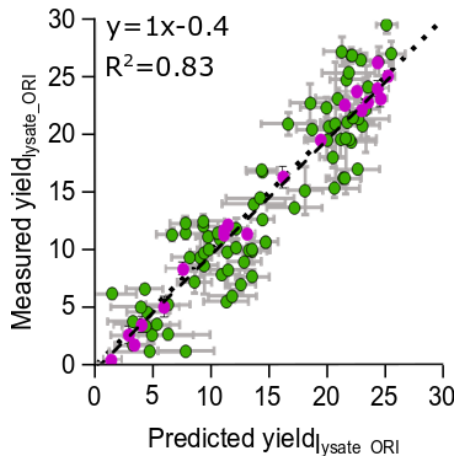
Lysate BL21 AB



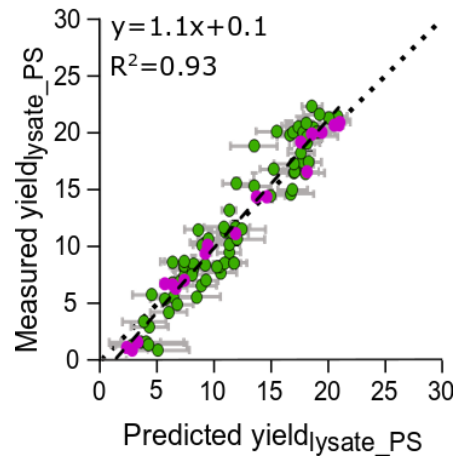
Predictions efficiencies

(different experimentalists, strains, conditions)

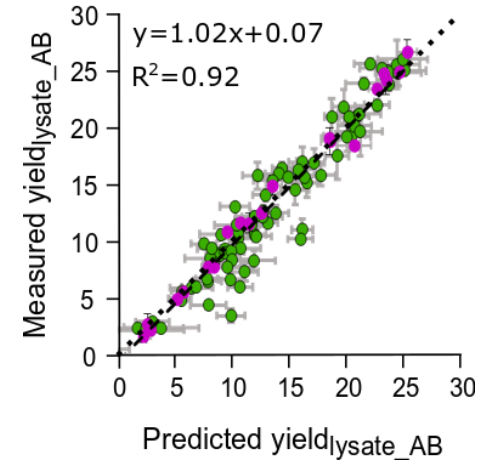
Lysate BL21 Ori



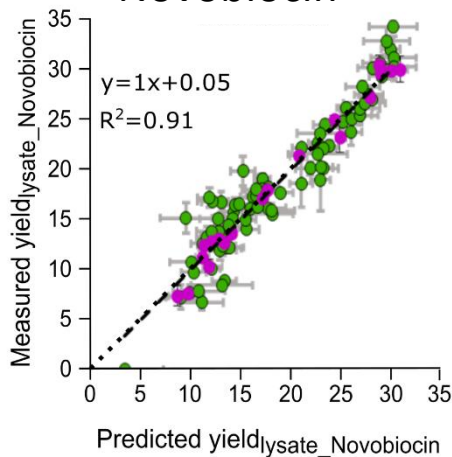
Lysate BL21 PS



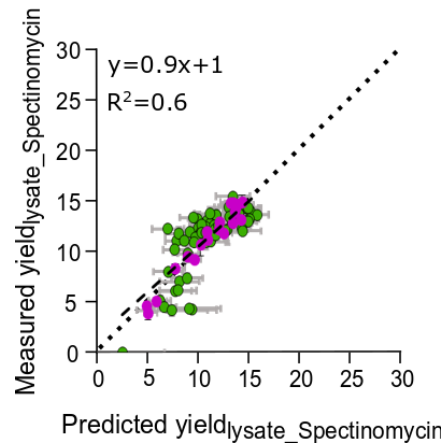
Lysate BL21 AB



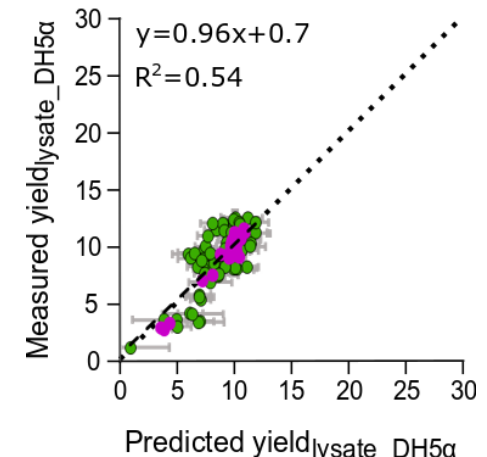
Novobiocin



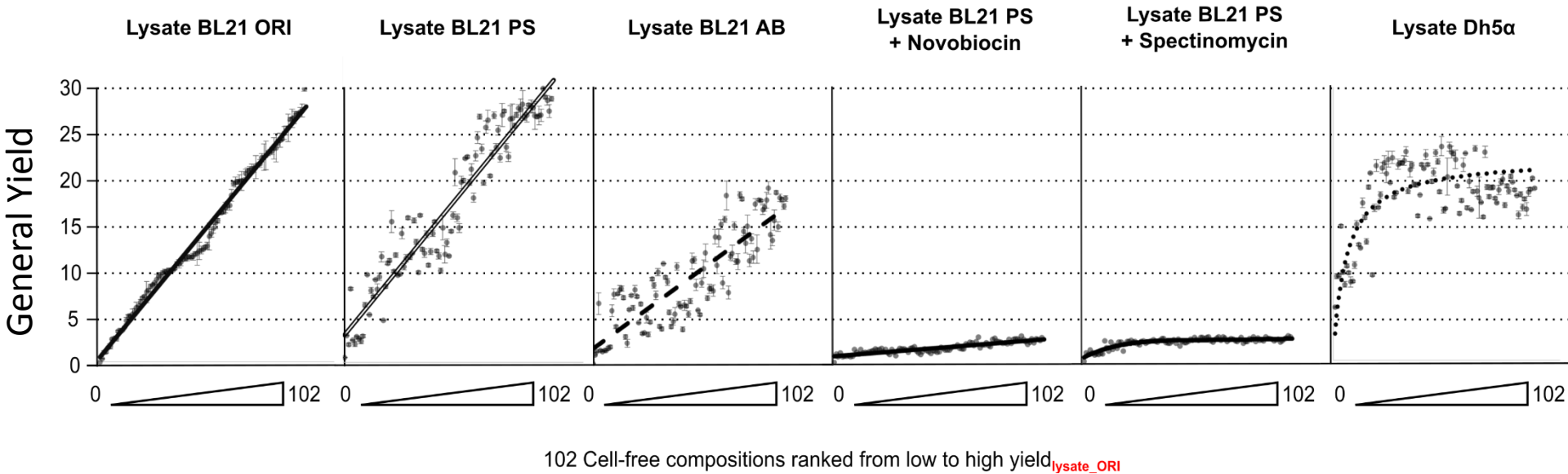
Spectinomycin



DH5a



“Absolute” comparison between lysates efficiencies



Conclusion

Cell-free composition:

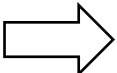
- We improved production of an home-made cell-free reactions without increasing the price
- We provided efficient predictions of protein production
- We highlighted the critical parameters involved in protein production

Lysate-independent maximization of portein production:

- We provide a one step method to maximize protein production for home-made cell-free

Transcription vs translation processes:

- Our data suggests that we mainly improved translation via composition optimization

 Our method can be easily extended to any other bacterial-based cell-free (*B. subtilis*, *V. natrigens*...)

The logo features a stylized network of blue and orange nodes connected by lines, resembling a molecular or biological structure.

Cell Free Systems Conference

Thank you

Bio-RetroSynth



Institut Pasteur



Mathilde Koch



Jean-Loup Faulon



Agnès Zettor



Fabrice Agou

Amir Pandi

Angelo Cardoso Batista

Paul Soudier

Manish Kushwaha

Melchior Du Lac

Ioana Grigoras-Popescu

Thomas Duigou