



HAL
open science

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

► To cite this version:

Olivier Borkowski, Mathilde Koch, Agnès Zettor, Amir Pandi, Angelo Cardoso Batista, et al.. A large-scale exploration of cell-free compositions to maximize protein production using active learning. Structural Biology and Chemistry Department days, Dec 2019, paris, France. hal-04292156

HAL Id: hal-04292156

<https://hal.science/hal-04292156>

Submitted on 21 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

Borkowski O*, Koch M*, Zettor A, Pandi A, Batista AC, Soudier P, Faulon J-L



Chemogenomic and Biological Screening Core Facility



Lysate-based cell-free systems have become a major platform to study gene expression¹⁻³ but batch-to-batch variation makes protein production difficult to predict⁴. Here we describe an active learning approach⁵ to explore a combinatorial space of ~4,000,000 cell-free compositions, maximizing protein production and identifying critical parameters involved in cell-free productivity. We also provide a one-step-method to achieve high quality predictions for protein production using minimal experimental effort regardless of the lysate quality.

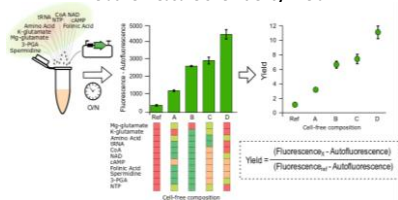
Cell-free composition:

A vast combinatorial space to explore

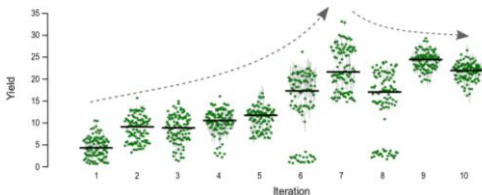
Component	Concentration
Mg glutamate (mM)	0.4 1.2 2 4
K-glutamate (mM)	8 24 40 80
Amino Acid (mM)	0.15 0.45 0.75 1.5
tRNA (mg/ml)	0.02 0.06 0.1 0.2
CoA (mM)	0.026 0.078 0.13 0.26
NAD (mM)	0.013 0.039 0.165 0.33
cAMP (mM)	0.075 0.225 0.375 0.75
Folic Acid (mM)	0.0048 0.0204 0.034 0.068
Spermidine (mM)	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

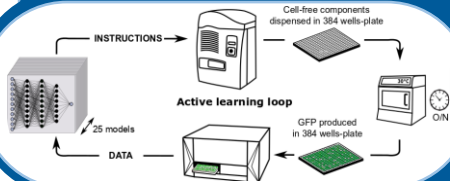
A relative measure of efficiency: Yield



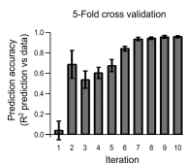
34 time Yield improvement after 7 iterations using a home-made lysate



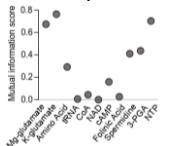
Active learning loop



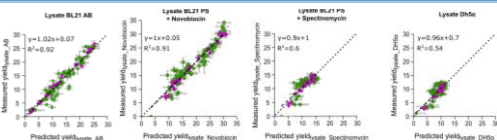
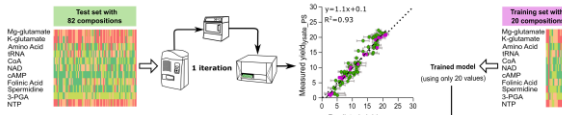
Prediction accuracy of the models



Mutual Information analysis



A one-step method to maximize Yield in cell-free platforms



References

- Orelle C. et al. Nature 524, 119–124 (2015)
- Moore S. J. et al. PNAS 115, 4340–4349 (2018)
- Noireaux V. et al. PNAS 100, 12672–77 (2003)
- Hogman CE C. et al. Metab Eng 261–69 (2012)
- Settles B. et al. (2009)