Models for Cell-Free Synthetic Biology: Make Prototyping Easier, Better, and Faster

Mathilde Koch [1], Jean-Loup Faulon [1,2,3] and Olivier Borkowski [1,2]*

[1] Micalis Institute, INRA, AgroParisTech, University of Paris-Saclay, Jouy-en-Josas, France, [2] Systems and Synthetic Biology Lab, CEA, CNRS, UMR 8030, Genomics Metabolics, University Paris-Saclay, Évry, France, [3] SYNBIOCHEM Center, School of Chemistry, Manchester Institute of Biotechnology, University of Manchester, Manchester, United Kingdom

Cell-free TX-TL is an increasingly mature and useful platform for prototyping, testing, and engineering biological parts and systems. However, to fully accomplish the promises of synthetic biology, mathematical models are required to facilitate the design and predict the behavior of biological components in cell-free extracts. We review here the latest models accounting for transcription, translation, competition, and depletion of resources as well as genome scale models for lysate-based cell-free TX-TL systems, including their current limitations. These models will have to find ways to account for batch-to-batch variability before being quantitatively predictive in cell-free lysate-based platforms.

Keywords: mathematical model, cell-free, prototyping, resource competition, transcription, translation, metabolism

INTRODUCTION

All the processes required to produce proteins in bacteria can be performed by adding DNA to a cell-free platform. After lysis of living cells, transcription, translation, degradation, and protein folding continue to operate as they do *in vivo* (Hodgman and Jewett, 2012; Sun et al., 2013; Takahashi et al., 2015a). Metabolic pathways like glycolysis or pentose phosphate pathway remain active and are used to regenerate ATP and maximize protein production over time (Kim and Swartz, 2001; Calhoun and Swartz, 2005). Protein production outside of the cell simplifies gene expression with well-defined parameters, easy to control inputs, faster time scale, and less numerous unknown interactions. As a result, many laboratories use cell-free as a prototyping platform to characterize expression of single proteins or complex metabolic pathways (Takahashi et al., 2015b; Wu et al., 2017; Borkowski et al., 2018). Mathematical models dedicated to cell-free emerged to predict protein production and understand the limits of this new platform. Cell-free properties are close to living organisms as the same processes take place in both systems, yet significant differences exist. For example, molecular crowding (Spruijt et al., 2014) and resources distribution (Sun et al., 2013) are significantly altered in cell-free and there is no resource competition with the host. Such differences oblige synthetic biologists to adapt the models already developed for living cells. This short review focuses on the recent deterministic models developed to understand lysate-based cell-free platforms and used to predict the behavior of simple or complex pathways (**Table 1**). Those models pave the way for efficient metabolic engineering in the emerging field of cell-free synthetic biology.

TABLE 1 | Deterministic models developed to understand cell-free.

Modeling strategy	Problems tackled	Level of detail	Description	Strength	Weakness	References
ODE	Protein production in cell-free	Simple*	Michaelis-Menten for the translation processes, as well as for degradation	Simple, quantitative	Parameters values are experiment dependent	Karzbrun et al., 2011
ODE	Protein production in cell-free	Simple*	Simple description of the transcription and translation processes. One parameter summarize each	Simple, qualitative	Parameters values are experiment dependent	Stögbauer et al., 2012; Siegal-Gaskins et al., 2014
ODE	Protein production in cell-free Resource competition in cell-free	Complex*	Detailed modeling: Particular focus on the translation process and the competition for the translation machinery	Generalization power quantitative	Parameters values are experiment dependent	Underwood et al., 2005; Borkowski et al., 2018
ODE	Protein production in cell-free Resource competition in cell-free	Simple*	Binding, unbinding and elongation bundled in one parameter. Accounts for number of RNAP, ribosomes, and promoter/RBS strengths	Easily adaptable to a new situation or phenomena	Necessitates parameter determination for new situation	Gyorgy and Murray, 2016; Halter et al., 2018; Voyvodic et al., 2018
ODE	Protein production in cell-free Resource depletion in cell-free	Complex*	Accounts for all known processes, including NTP consumption and degradation	Models biochemical phenonema precisely so generalisable	Important parameter identification/estimation needed	Siegal-Gaskins et al., 2013; Tuza et al., 2013; Nieß et al., 2017; Matsuura et al., 2018; Moore et al., 2018
Constraint based	Protein production with cell-free Metabolism in cell-free	Complex*	Modification of *E. Coli* metabolic model to account for cell-free constraints	Accounts for the full metabolism, can use constraints based methods such as FBA	Defining the objectives, require deeper knowledge of reactions in cell-free	Vilkhovoy et al., 2018

*Simple stands for one protein produced and limited amount of parameters (<10) Complex stands for more than one protein produced or/and large amount of parameters (more than 10).

TRANSLATION AND TRANSCRIPTION PROCESSES IN CELL-FREE

Lysate-based cell-free consists of a crude cell extract supplemented with buffer, amino acids, NTP, NAD, PEG, tRNA, and metabolic intermediates (Sun et al., 2013). A major advantage of cell-free is the absence of host regulations (Hodgman and Jewett, 2012), allowing circuits to function in isolation and an easy quantitative description of gene expression. A constitutively expressed gene in cell-free exhibits specific patterns at the translation and transcription levels. Protein production can be divided in 4 phases: in phase 1, the production rate increases over time, in phase 2, the production rate is constant during a 30 min/1 h, in phase 3 the production rate decreases slowly and eventually in phase 4 the production rate is null (Siegal-Gaskins et al., 2014; **Figure 1A**). A similar 4 phases pattern is observed with the mRNA concentration (Siegal-Gaskins et al., 2014; **Figure 1B**). ODE Models describing transcription, translation, mRNA, and protein degradation processes at various scales have been successfully used to predict mRNA and protein dynamics in lysate-based systems (Stögbauer et al., 2012; Tuza et al., 2013; Siegal-Gaskins et al., 2014; Nieß et al., 2017; Borkowski et al., 2018; Matsuura et al., 2018; Moore et al., 2018). DNA concentrations are usually considered constant: degradation is neglected as plasmid DNA or protected

linear template are used, and replication is considered not to happen since no dNTPs are added to the reaction mix.

Numerous models, with varying degrees of complexity, try to reproduce those production phases observed in cell-free reactions.

A simple model based on only 4 reactions and 10 parameters is sufficient to fit the full mRNA and protein dynamics during the first hours of reaction (Karzbrun et al., 2011). The transcription process is reduced to one step in which the RNA polymerase binds to the DNA; the rate of mRNA production depends only this binding rate and the DNA length. Similarly, the translation process is described as one binding step of the ribosome on the mRNA with the rate of protein production depending only on the binding rate and the mRNA length. This model is appropriate for the first hours before the consumption of resources and/or the waste accumulation (e.g., ATP degradation, toxic metabolites…) cause the reaction to stop (Siegal-Gaskins et al., 2014). A simple way to simulate the slow decay in synthesis is the consumption of the NTP over time. The transcription reaction slows down and eventually stops (Stögbauer et al., 2012; Tuza et al., 2013). The decrease in the NTP concentration (Kim and Swartz, 2001; Jewett et al., 2008; Moore et al., 2018) is an efficient method to obtain a decreasing transcription over time and simulate protein and mRNA production in cell-free but no experimental data either confirms or denies this approach. The accumulation of inactive
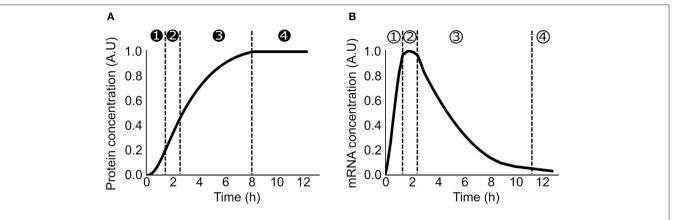
FIGURE 1 | Production of a constitutively expressed gene in cell-free **(A)** Protein concentration over time in a cell-free platform. ❶ the production rate increase over time ❷ the production rate is constant ❸ the production rate decreases ❹ the production rate is null **(B)** mRNA concentration over time in a cell-free platform ① the concentration increases over time ② the concentration is constant ③ the concentration decreases over time ④ the concentration is close to zero.

RNA polymerase/ribosome (Failmezger et al., 2016; Moore et al., 2018), accumulation of toxic metabolites (Kim and Swartz, 2000), or increase of the relative RNases concentration compare to the total amount of mRNA (Siegal-Gaskins et al., 2014) are possible other explanations of the arrest of protein production after 8 h. Voyvodic et al. (2018) added terms corresponding to decreasing resources for protein production and accumulation of toxic by-products as a reduction in production rates parametrized by a Michaelis-Menten like ratio, as an elegant way to account for the slowing production rate. As all cell-free models trying to account for the end of production after 8 h, the main issue is identifying the exact cause for decreasing production.

Models using Michaelis–Menten kinetics also succeed to capture protein production pattern in cell-free (Stögbauer et al., 2012). Those models precisely captured the observable mRNA and proteins dynamics in cell-free while remaining relatively coarse-grained. The model of Tuza et al. add extra steps in the transcription (and translation) process with a reversible binding of the RNA polymerase (ribosome) on the DNA (mRNA) followed by a reversible binding of the first NTP (amino acid) and eventually an irreversible elongation step. Moore et al. also developed a model accounting for reversible binding, unbinding and elongation steps, sharing the NTP energy source. Those more detailed descriptions of the transcription and translation processes lead to accurate predictions of the data obtained in cell-free and capture additional properties (Stögbauer et al., 2012; Tuza et al., 2013; Borkowski et al., 2018; Moore et al., 2018). For example, the non-additive cost of protein production when several genes are expressed requires higher level of complexity to be predicted (Borkowski et al., 2018).

While all models presented in this section described transcription and translation processes, the main challenge they faced is proper parameter identification, as biochemical parameters can vary widely from batch to batch and from *in vivo* to cell-free systems. Currently, models often used components concentration measured *in vivo* and estimated their concentration based on the dilution factor of the *E. coli*

cytoplasm after the lysate extraction protocol, which is not entirely satisfactory.

RESOURCE COMPETITION IN CELL-FREE

Resource competition is an important phenomenon that impacts circuit behavior in cell-free systems and should be accounted for in modeling approaches.

As a fixed amount of resources is present in the cell-free extract, competition has been measured between synthetic circuits (Siegal-Gaskins et al., 2014; Borkowski et al., 2018; Moore et al., 2018). Some of the previously described models take into account the limitation of each resource and include a fixed amount of transcription and translation machineries to predict the impact of resource competition in cell-free (**Figure 2A**; Siegal-Gaskins et al., 2013; Borkowski et al., 2018; Moore et al., 2018).

A maximal protein production is measured after a few hours before resources depletion and degradation (**Figure 1**). This upper limit on production rate is the result of one or several limited resources (RNA polymerase, NTP, ribosome, elongation factors, amino acids, chaperone, tRNA synthetase, or tRNA). DNA, NTP, amino acids, and T7 RNA polymerase are directly added to the mix so their impacts on the protein production can be easily measured. Increasing DNA concentration leads to an increase of protein production until a saturation point is reached (Siegal-Gaskins et al., 2014; Borkowski et al., 2018; Voyvodic et al., 2018), and toxicity can be observed with high DNA concentration (Borkowski et al., 2018). T7 polymerase (Siegal-Gaskins et al., 2014), amino acids, tRNA and nucleotides (Shin and Noireaux, 2010) are present in excess in the cell-free mix causing no noticeable competition for these resources. Eventually, High NTP concentration negatively affects the translation process (Nagaraj et al., 2017). Natural transcription and translation machineries are less controlled as they are added via the crude extract. Indirect measurements using competition for resources between two plasmids are
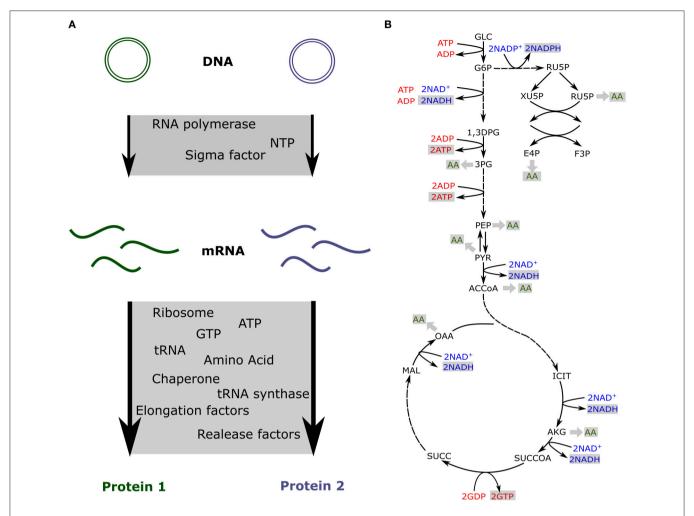
FIGURE 2 | Resource competition in cell-free **(A)** Competition for resources between two genes expressed in cell-free. The transcription and translation machineries amounts are fixed **(B)** Production of energy and amino acids via the core metabolic network describing glycolysis, pentose phosphate pathway, the TCA cycle and the Entner-Doudoroff pathway.

used to deduce competition for transcription and translation machineries (Underwood et al., 2005; Siegal-Gaskins et al., 2014; Gyorgy and Murray, 2016; Borkowski et al., 2018; Moore et al., 2018). The main source of competition can be at the transcription and/or translation level depending of the extract and the level of protein produced (Underwood et al., 2005; Li et al., 2014; Siegal-Gaskins et al., 2014; Gyorgy and Murray, 2016; Borkowski et al., 2018; Moore et al., 2018; Voyvodic et al., 2018). Parametrization of the models using the appropriate RNA polymerase and ribosomes concentrations and binding/unbinding rates allows an accurate description of the resource competition and to fit properly the production of several proteins expressed concurrently in cell-free (Siegal-Gaskins et al., 2014; Borkowski et al., 2018; Moore et al., 2018). Accounting for RNAP and ribosomes sharing between parts can also be leveraged to minimize the number of experiments required to fit parameters and obtain a predictive model (Halter et al., 2018). While not the main focus of this review, parameter estimation or identification is a major hurdle of

detailed models, and techniques from systems biology (e.g., Lillacci and Khammash, 2010) can be used to tackle this issue.

The models presented in this sections, while being able to account for transcription, translation and resource competition from a lack of generalizability due both to variability in experimental conditions as batches can differ greatly, and to scarcity of biochemical work measuring those parameters in cell-free setting as has been estimated from *in vivo* measurements.

METABOLISM IN CELL-FREE

The models presented in this section are constraint based, so as to take the whole metabolism into account and not the circuit in isolation as done in the previous sections.

In cell-free platforms, translation, and transcription are not the only active processes. Glycolysis, pentose phosphate pathway, TCA cycle, Entner-Doudoroff pathway, and amino

acid biosynthesis are still producing ATP, reducing equivalents, and amino acids (Kim and Swartz, 2001; Jewett and Swartz, 2004a,b). The previously described models account for the resources competition for a fixed amount of transcription and translation resources and usually do not include any metabolite production or consumption. Such an approach is quite limiting for metabolic pathway prototyping as those circuits also compete for metabolites (Wu et al., 2017; Borkowski et al., 2018). Constraint-based models have been used to simulate metabolites production and consumption when various proteins are produced at different levels (Vilkhovoy et al., 2018; **Figure 2B**). This model coupled transcription and translation processes with the availability of metabolic resources. Flux balance analysis was adapted to cell-free conditions with the objective function being the maximization of the protein translation rate. Growth associated reactions were removed and cell-free specific deletions were added from *E. coli* metabolic model, leading to 264 reactions and 146 species (Vilkhovoy et al., 2018). The stoichiometric network was adjusted to cell-free and fluxes were constrained by experimental measurements of glucose, nucleotides, amino and organic acid consumption and production rates. The transcription and translation were bound by Michaelis-Menten formula with a maximum transcription and translation rate depending on RNA polymerase concentration, RNA polymerase elongation rate, gene length, promoter strength, and the ribosome concentration, a polysome amplification constant, the ribosome elongation rate, the protein length, and the RBS strength, respectively. The energy efficiency was calculated using the ATP cost by transcription and translation processes. Transcription and translation rates are subject to resource constraints encoded by the metabolic network (**Figure 2B**). This model efficiently predicts proteins production and simulates optimal flux distribution in cell-free metabolic network. It makes predictions possible for metabolic engineering in cell-free as metabolites produced or consumed by a pathway will be accounted for via its energy efficiency.

Constraint based modeling for cell-free systems is an interesting field that would need further developments from the research community, both to include cell-free specific constraints and reactions, as well as to account for dynamic behavior such as metabolite exhaustion in cell-free systems.

CONCLUSION

Cell-free appeared as the ideal platform for circuits prototyping. It accelerates characterization and avoids the impact of the host on the circuit behavior. Models can be easily parametrized and predictions are easier and more accurate than *in vivo*, for qualitative behavior. Parametrization for quantitative behavior can be tackled using techniques from systems biology. Simple models succeed to accurately predict the simultaneous production levels of multiple proteins and the competition for the limited amount of resources in cell-free. A certain level of complexity is necessary to capture competition for metabolites but produces a powerful tool for metabolic engineering. The main limit for lysate-based cell-free in metabolic engineering and modeling remains extract preparation: extract efficiency can differ strongly depending on the experimentalist leading to variability of protein production and necessity of robust controls for each new batch, as well as uncertain parameters that vary with each batch for the modeler. Preliminary control of the extract quality and tuning of model parameters on each batch is required to obtain accurate predictions and precludes generalization. A way forward to both increase reproducibility and predictive modeling in cell-free systems would be a higher degree of automation in the extract production providing robust lysate preparation at affordable price.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

REFERENCES

Borkowski, O., Bricio, C., Murgiano, M., Rothschild-Mancinelli, B., Stan, G. B., and Ellis, T. (2018). Cell-free prediction of protein expression costs for growing cells. *Nat. Commun.* 9:1457. doi: 10.1038/s41467-018-03970-x

Calhoun, K. A., and Swartz, J. R. (2005). Energizing cell-free protein synthesis with glucose metabolism. *Biotechnol. Bioeng.* 90, 606–613 doi: 10.1002/bit. 20449

Failmezger, J., Nitschel, R., Sánchez-Kopper, A., Kraml, M., and Siemann-Herzberg, M. (2016). Site-specific cleavage of ribosomal RNA in *Escherichia coli*-based cell-free protein synthesis systems. *PLoS ONE* 11:e0168764. doi: 10.1371/journal.pone.0168764

Gyorgy, A., and Murray, R. M. (2016). "Quantifying resource competition and its effects in the TX-TL system," in *2016 IEEE 55th Conference on Decision and Control, CDC 2016* (Las Vegas, NV). doi: 10.1109/CDC.2016.7798775

Halter, W., Allgower, F., Murray, R. M., and Gyorgy, A. (2018). "Optimal experiment design and leveraging competition for shared resources in cell-free extracts," in *2018 Conference on Decision and Control* (Miami Beach, FL).

Hodgman, C. E., and Jewett, M. C. (2012). Cell-free synthetic biology: thinking outside the cell. *Metab. Eng.* 14, 261–269 doi: 10.1016/j.ymben.2011. 09.002

Jewett, M. C., Calhoun, K. A., Voloshin, A., Wuu, J. J., and Swartz, J. R. (2008). An integrated cell-free metabolic platform for protein production and synthetic biology. *Mol. Syst. Biol.* 4:220. doi: 10.1038/msb2008.57

Jewett, M. C., and Swartz, J. R. (2004a). Mimicking the *Escherichia coli* cytoplasmic environment activates long-lived and efficient cell-free protein synthesis. *Biotechnol. Bioeng.* 86, 19–26 doi: 10.1002/bit.20026

Jewett, M. C., and Swartz, J. R. (2004b). Substrate replenishment extends protein synthesis with an *in vitro* translation system designed to mimic the cytoplasm. *Biotechnol. Bioeng.* 87, 465–472 doi: 10.1002/bit.20139

Karzbrun, E., Shin, J., Bar-Ziv, R. H., and Noireaux, V. (2011). Coarse-grained dynamics of protein synthesis in a cell-free system. *Phys. Rev. Lett.* 106:048104. doi: 10.1103/PhysRevLett.106.048104

Kim, D. M., and Swartz, J. R. (2000). Prolonging cell-free protein synthesis by selective reagent additions. *Biotechnol. Prog.* 16, 385–390. doi: 10.1021/bp000031y

Kim, D. M., and Swartz, J. R. (2001). Regeneration of adenosine triphosphate from glycolytic intermediates for cell-free protein synthesis. *Biotechnol. Bioeng.* 74, 309–316. doi: 10.1002/bit.1121

Li, J., Gu, L., Aach, J., and Church, G. M. (2014). Improved cell-free RNA and protein synthesis system. *PLoS ONE* 9:e106232. doi: 10.1371/journal.pone.0106232

Lillacci, G., and Khammash, M. (2010). Parameter estimation and model selection in computational biology. *PLoS Comput. Biol.* 6:e1000696. doi: 10.1371/journal.pcbi.1000696

Matsuura, T., Hosoda, K., and Shimizu, Y. (2018). Robustness of a reconstituted *Escherichia coli* protein translation system analyzed by computational modeling. *ACS Synth. Biol.* 7, 1964–1972 doi: 10.1021/acssynbio.8b 00228

Moore, S. J., MacDonald, J. T., Wienecke, S., Ishwarbhai, A., Tsipa, A., Aw, R., et al. (2018). Rapid acquisition and model-based analysis of cell-free transcription–translation reactions from nonmodel bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 115, E4340–E4349. doi: 10.1073/pnas.1715806115

Nagaraj, V. J., Greene, J. M., Sengupta, A. M., and Sontag, E. D. (2017). Translation inhibition and resource balance in the TX-TL cell-free gene expression system. *Synthet. Biol.* 2:ysx005. doi: 10.1093/synbio/ysx005

Nieß, A., Failmezger, J., Kuschel, M., Siemann-Herzberg, M., and Takors, R. (2017). Experimentally validated model enables debottlenecking of *in vitro* protein synthesis and identifies a control shift under *in vivo* conditions. *ACS Synth. Biol.* 6, 1913–1921 doi: 10.1021/acssynbio.7b00117

Shin, J., and Noireaux, V. (2010). Efficient cell-free expression with the endogenous *E. coli* RNA polymerase and sigma factor 70. *J. Biol. Eng.* 4:8. doi: 10.1186/1754-1611-4-8

Siegal-Gaskins, D., Noireaux, V., and Murray, R. M. (2013). "Biomolecular resource utilization in elementary cell-free gene circuits," in *American Control Conference* (Washington, DC).

Siegal-Gaskins, D., Tuza, Z. A., Kim, J., Noireaux, V., and Murray, R. M. (2014). Resource usage and gene circuit performance characterization in a cell-free "breadboard." *ACS Synth. Biol.* 3, 416–425. doi: 10.1021/sb40 0203p

Spruijt, E., Sokolova, E., and Huck, W. T. (2014). Complexity of molecular crowding in cell-free enzymatic reaction networks. *Nat. Nanotechnol.* 9, 406–407. doi: 10.1038/nnano.2014.110

Stögbauer, T., Windhager, L., Zimmer, R., and Rädler, J. O. (2012). Experiment and mathematical modeling of gene expression dynamics in a cell-free system. *Integr. Biol.* 4, 494–501. doi: 10.1039/c2ib00102k

Sun, Z. Z., Hayes, C. A., Shin, J., Caschera, F., Murray, R. M., and Noireaux, V. (2013). Protocols for Implementing an *Escherichia coli* based TX-TL cell-free expression system for synthetic biology. *J. Vis. Exp.* 79:e50762. doi: 10.3791/50762

Takahashi, M. K., Chappell, J., Hayes, C. A., Sun, Z. Z., Kim, J., Singhal, V., et al. (2015a). Rapidly characterizing the fast dynamics of RNA genetic circuitry with cell-free transcription–translation (TX-TL) systems. *ACS Synth. Biol.* 4, 503–515. doi: 10.1021/sb400206c

Takahashi, M. K., Hayes, C. A., Chappell, J., Sun, Z. Z., Murray, R. M., Noireaux, V., et al. (2015b). Characterizing and prototyping genetic networks with cell-free transcription-translation reactions. *Methods* 86, 60–72. doi: 10.1016/j.ymeth.2015.05.020

Tuza, Z. A., Singhal, V., Kim, J., and Murray, R. M. (2013). An in silico modeling toolbox for rapid prototyping of circuits in a biomolecular "breadboard" system. in *Proceedings of the IEEE Conference on Decision and Control* (Florence).

Underwood, K. A., Swartz, J. R., and Puglisi, J. D. (2005). Quantitative polysome analysis identifies limitations in bacterial cell-free protein synthesis. *Biotechnol. Bioeng.* 91, 425–435. doi: 10.1002/bit.20529

Vilkhovoy, M., Horvath, N., Shih, C.-H., Wayman, J., Swartz, J., and Varner, J. D. (2018). Sequence specific modeling of *E. coli* cell-free protein synthesis. *ACS Synth Biol.* 7, 1844–1857 doi: 10.1021/acssynbio.7b00465

Voyvodic, P., Amir, P., Mathilde, K., Jean-Loup Faulon, J., and Bonnet, J. (2018). Plug-and-play metabolic transducers expand the chemical detection space of cell-free biosensors. *bioRxiv [Preprint]*. doi: 10.1101/397315

Wu, Y. Y., Sato, H., Huang, H., Culler, S. J., Khandurina, J., Harish, et al. (2017). System-level studies of a cell-free transcription-translation platform for metabolic engineering. *biorxiv [Preprint]*. doi: 10.1101/172007