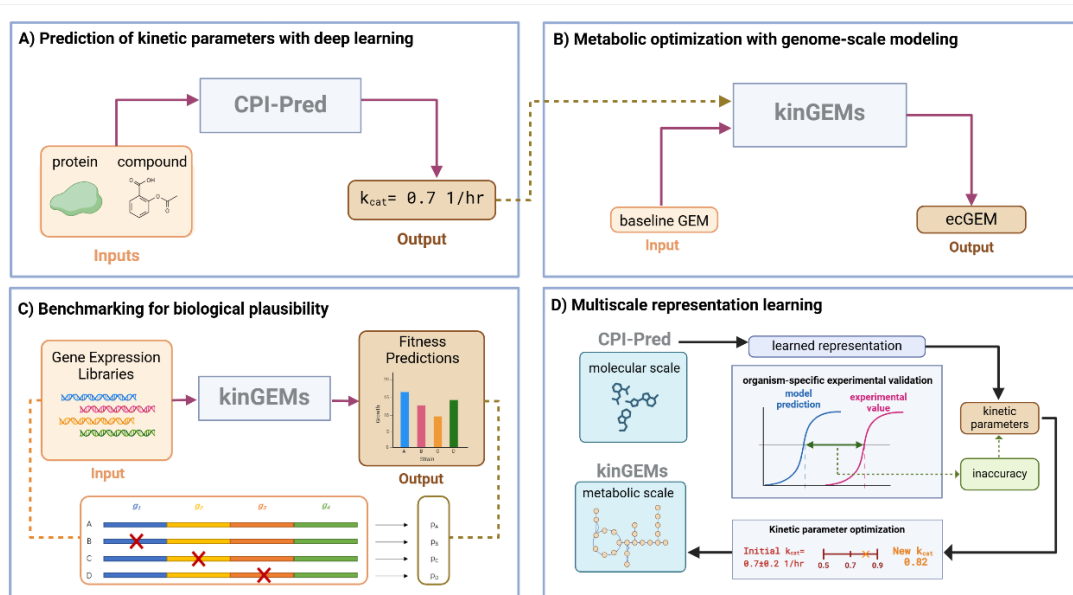


From Mechanisms to Models: Multi-Modal Machine Learning for Kinetically Constrained Genome-Scale Metabolic Models

Genome-scale metabolic models (GEMs) are a cornerstone for simulating cellular metabolism, but their predictive power is limited by the absence of enzyme kinetics. Incorporating enzyme constraints to these models (ecGEMs) narrows the solution space and links molecular parameters to organismal phenotypes. However, the lack of experimentally measured kinetic constants (e.g., k_{cat}) severely restricts scalability and transferability.



We introduce an integrated machine learning and modeling framework that balances precision and accuracy in ecGEMs. Our approach combines CPI-Pred, a deep learning model that predicts kinetic parameters from protein language model embeddings and compound representations, with kinGEMs, a pipeline that integrates these predictions into GEMs and refines the constraints through simulated annealing.

We evaluate the framework across three axes:

1. **Precision** (flux variability analysis): Incorporating CPI-Pred predictions reduces median flux variability by 3-fold compared to unconstrained GEMs, yielding more defined and interpretable solution spaces.
2. **Accuracy** (*E. coli* genetic and substrate perturbations): Using RB-TnSeq data, kinGEMs improves gene lethality prediction accuracy and ROC-AUC compared to baseline GEMs.
3. **Cross-organism generalization:** The same pipeline predicts gene knockout effects in *P. putida* and *S. elongatus* and recapitulates growth/no-growth outcomes in growth assays.

Together, these results demonstrate that ML-predicted kinetic parameters tuned in-vivo to match cellular contexts can systematically improve both the internal precision of metabolic models and their external accuracy in predicting phenotypes. Ongoing work scales the framework to the AGORA microbiome resource, enabling large-scale, interpretable simulations of microbial communities and perturbation studies. The integration of modern ML with mechanistic modeling offers a path toward more precise and accurate ecGEMs, broadening their impact in systems biology, metabolic engineering, and synthetic biology.