DOI 10.18699/SBB-2022-331

Genome-scale reconstructions of microbial biodynamics: great expectations, modest current success, and prospects for improvement

Panikov N.S.
Department of Chemistry and Chemical Biology, Northeastern University, Boston, MA, USA
n.panikov@northeastern.edu

Keywords: growth kinetics; conditional expression, cell composition; environmental factors; stress resistance; limitation

Motivation: *The importance of genome-scale models (GEM).* Today, NCBI lists more than 300,000 completed whole-genome projects, they cover almost the entire diversity of cultivable microorganisms. Further progress critically depends on how efficiently sequence data are converted into solid biological knowledge. Regular bioinformatics tools are not sufficient, and much higher expectations are coming from systems biology, a holistic approach aimed to discover emergent properties of biosystems using GEM as a research tool. The ultimate goal of GEM is the accurate prediction of *phenotype* from *genotype* under specified environmental conditions. If successful, it would be a remarkable scientific breakthrough, the next great step of the genomic revolution accelerating progress by substitution of real experiments with verified simulations. An especially high benefit is expected for studies of the hard-to-culture microorganisms, such as biotrophic species (symbionts or parasites poorly growing without the host) or not-yet-cultivable soil and aquatic bacteria with entirely unknown phenotypes.
Historical roots of contemporary GEMs: 1) biochemical modeling of metabolic cycles and 2) growth kinetics of microorganisms. The first approach uses equations of enzyme kinetics to analyze the regulatory metabolic circuits. The second approach simulates growth using the mass-balance ODEs for such complex variables as nutrients, cell mass, and products. The still-popular Monod model recognizes nutrients as the only controlling factor. More advanced models (structured, cybernetic, SCM) account for changeable cell constituents and attempt to mimic the regulatory mechanisms. The closest to GEMs is the synthetic chemostat model (SCM) [1], which links growth kinetics (e.g., specific growth rate, SGR) with conditionally expressed macromolecular cell composition (MMCC); it allows successful mechanistic simulation of a wide range of dynamic observation (Figure).
Contrary to enzyme kinetics, microbial growth is to be described in a dualistic way using rate and compositional variables. If SGR and MMCC remain constant in a stable environment, we call it *simple growth* (syn: balanced, steady-state). The real-life environment is not stable causing self-adjusting changes in both SGR and MMCC called the *complex biodynamics.* It prevails over simple patterns (Fig. 1, panels 1–3) but remains reproducible and allows precise kinetic studies.
Overview of GEM. Any microbial genome consists of three parts [2]: (i) the M-matrix (Metabolism) encoding enzymes of intermediary metabolism; (ii) the E-matrix (Expression) encoding enzymes and RNAs involved in transcription and translation; and (iii) the O-matrix (Operon), the transcriptional regulatory network (has not yet been incorporated into GEM). The first successful GEM known as FBA (Flux Balance Analyses), deals with the M-matrix. The FBA stems from a biochemical predecessor and

pragmatically ignores time, cell growth, and any kinetic and regulatory considerations. Instead, it uses several simplifying assumptions (steady-state, flux constraints, constant average MMCC) and the optimality principle (linear programming with a single objective function, e.g., the SGR maximization) to uniquely resolve the gigantic M-matrix (up to 30 % of the whole genome). The FBA turned out to be a very popular and efficient computational tool assisting metabolic engineering and displaying a reasonable agreement with key experimental data, such as gene essentiality, the effect of mutations, aerobic-to-anaerobic transition, etc. The next generation of GEM included dynamic FBA (added exchange module simulating nutrients uptake and release of products), ME-models (account of metabolism and expression), and finally the WC-models (the modular whole-cell reconstructions). The ME- and WC models have a larger genome coverage than FBA which became possible due to impressive progress in numeric integration of gigantic models, up to 10,500 non-linear ODE in the WC-model of *E. coli* [3], probably the largest sets of ODE in the history of science.

The limited success of the advanced GEMs. To reproduce the complex biodynamics, any GEM should satisfy the following three conditions:

1) to account for the effects of environmental factors (EF), as a minimum, the limiting substrate concentration in the M-matrix;
2) to simulate the adaptive MMCC changes including the conditional expression of proteins in response to variable EF;
3) to compute the SGR as a final model's output rather than a fixed input variable.

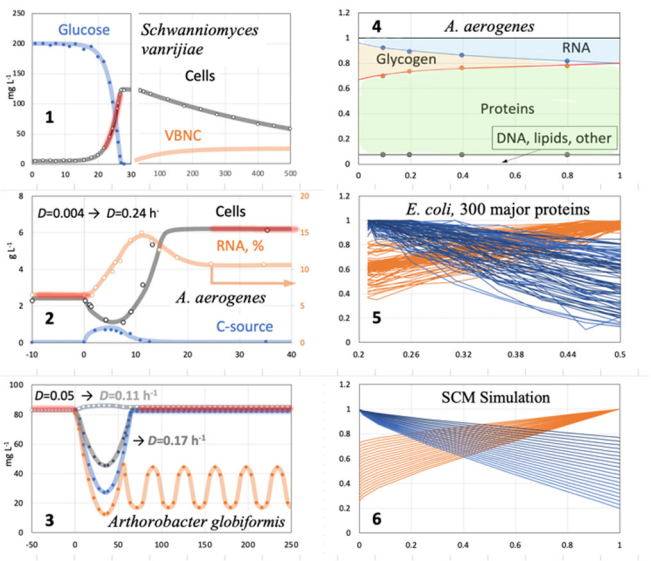

Fig. 1. Examples of experimentally observed and simulated simple (red-marked fragments of cells curves) and complex microbial dynamics: 1) batch culture from lag- to death phases including the formation of VBNC (viable-but-not-cultivable cells); 2) shift-up transient in a chemostat with overshoots and undershoots; 3) shift-up with sustained oscillations; 4) MMCC changes in chemostat; 5) proteomic profile changes in a chemostat, orange, and blue curves are the normalized content of the up- and downregulated proteins; 6) proteomic profile simulated by SCM. All experimental data were fitted to the SCM. From [4] with permission

The dFBA models contain the outdated Monod-style exchange module, so they do not meet the conditions 2) and 3) and can reproduce only the simple growth. The published ME-models [5] formally comply with all three requirements but in the wrong way. The most essential drawbacks were i) omitting the nutrient concentration as a variable, ii) incomplete list of changeable MMCC (see below), and iii) using SGR as an independent variable (widely spread misconception fully discussed in the talk). As a result, the ME models badly failed in an attempt to reproduce the proteomic profile. Ironically, the simple SCM turned out to be a better expression predictor (see Fig. 1, compare 5 and 6). The published WC models [3] assumed the constancy of SGR and MMCC therefore none of the three conditions are met and these models can reproduce

only the simple growth. The talk discusses the benefit of complex data in modeling and explains why the principle *learn to walk before you run*, does not work as expected.
What are the main deficiencies of existing GEMs and how they can be improved? Let us face it: *the first pancake is always lumpy*, and the recently constructed dynamic GEMs (dFBA, ME- and WC) have too many drawbacks. They are extremely bulky and not as useful in practice as the static FBA. Moreover, the accuracy of microbial dynamics simulation is not as good as with the best pre-genomic models. One reason is a generation gap: contemporary modelers overlooked the progress in microbial growth kinetics achieved before 2000. The talk specifies how the kinetic framework of existing GEM can be upgraded; the simplest solution is a merging of two models: the regular static FBA and the SCM.
Another discussion point is the strategy of GEM development. The current trend is to increase as much as possible the size of GEM (more genes and metabolites, supplementing the M- and E-matrices with kinetic and regulatory data, etc.), and further improve the computational algorithms. An expectation is that close to 100% coverage of the genome combined with constraints and objective functions like SGR-maximization is enough to provide an adequate simulation. We argue based on the history of mathematical biology that described strategy is counterproductive. Instead, we suggest trying more efforts to test the alternative hypotheses about the systems biology of bacteria. The objective functions should be related to the life strategies of the wild predecessors subjected to natural selection. The SGR-maximization as an objective function can be rejected as a wrong assumption based on the expression profile (see Fig. 1, panel 5). We conclude that the life strategy of *E coli* is to keep a balance between growth and stress resistance, which play the same role as the immune system in mammals.

References

1. Panikov N.S. Microbial Growth Kinetics. London: Chapman & Hall, 1995.
2. Palsson B.Ø. Systems biology: Constraint-based Reconstruction and Analysis. Cambridge: Cambridge Univ Press, 2015.
3. Macklin D.N., Ahn-Horst T.A., Choi H., Ruggero N.A., Carrera J., Mason J.C. et al. Simultaneous cross-evaluation of heterogeneous E. coli datasets via mechanistic simulation. *Science*. 2020;369(6502):eaav3751.
4. Panikov N.S. Genome-Scale Reconstruction of Microbial Dynamic Phenotype: Successes and Challenges. *Microorganisms.* 2021;9:2352-2400.
5. O'Brien E.J., Lerman J.A., Chang R.L., Hyduke D.R., Palsson B.O. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Mol Syst Biol.* 2013;9:1-13.