A modular minimal cell model: Purine and pyrimidine transport and metabolism

M. Castellanos¹, D. B. Wilson², and M. L. Shuler⁵

¹School of Chemical and Biomolecular Engineering and ²Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853-5201

Communicated by Edwin N. Lightfoot, University of Wisconsin, Madison, WI, February 10, 2004 (received for review October 7, 2003)

A more complete understanding of the relationship of cell physiology to genomic structure is desirable. Because of the intrinsic complexity of biological organisms, only the simplest cells will allow complete definition of all components and their interactions. The theoretical and experimental construction of a minimal cell has been suggested as a way to develop such an understanding. Our ultimate goal is to convert a “coarse-grain” lumped parameter computer model of Escherichia coli into a genetically and chemically detailed model of a “minimal cell.” The base E. coli model has been converted into a generalized model of a heterotrophic bacterium. This coarse-grain minimal cell model is functionally complete, with growth rate, composition, division, and changes in cell morphology as natural outputs from dynamic simulations where only the initial composition of the cell and of the medium are specified. A coarse-grain model uses pseudochemical species (or modules) that are aggregates of distinct chemical species that share similar chemistry and metabolic dynamics. This model provides a framework in which these modules can be “delumped” into chemical and genetic descriptions while maintaining connectivity to all other functional elements. Here we demonstrate that a detailed description of nucleotide precursors transport and metabolism is successfully integrated into the whole-cell model. This nucleotide submodel requires fewer (12) genes than other theoretical predictions in minimal cells. The demonstration of modularity suggests the possibility of developing modules in parallel and recombining them into a fully functional chemically and genetically detailed model of a prokaryote cell.

The basic design rules relating the regulation of cellular function to genomic structure is of broad interest. Bioinformatics emerged as an approach to convert static linear sequence genomic data into an understanding of the dynamic nonlinear function of living organisms. Initial efforts have focused on identifying the proteins encoded in the genome and, subsequently, identifying protein function and regulatory elements in the genome. These efforts are able to address specific questions but cannot translate genomic data broadly into an understanding of cell function. We propose a reverse approach. We ask how we would design a cell to achieve expected functions and, from that design, how we would write the genomic instructions. This approach follows the typical engineering design approach where desired performance dictates functional design, which is then translated into blueprints. To accomplish this goal, we are constructing a chemically and genetically minimal cell computer model. By modeling the essential regulatory structure and functions to maintain a living cell, we expect to better understand the relationship of genomic instructions to cell function and regulation.

A “minimal cell” is a hypothetical cell possessing the minimum functions required for sustained growth and reproduction in a maximally supportive culture environment. This organism is considered to live in a rich environment with preformed nutrients and relatively constant temperature and pH.

Recent developments in bioinformatics and comparative genomics have produced several candidate minimal gene sets. Mushegian and Koonin (1) compared the protein sequences of Mycoplasma genitalium (the smallest genome among self-replicating organisms, 580 kbp) and Haemophilus influenzae (1,703 kbp), Gram-positive and -negative bacteria, respectively, identifying 240 orthologous genes. They estimated a minimal genome composed of 262 genes by adding 22 genes to the 240 identified initially to fill missing links in essential pathways. Hutchison et al. (2) compared the sequences of M. genitalium and its closest relative Mycoplasma pneumoniae and used transposons to knock out specific genes identified by bioinformatics analysis as potentially essential; the number of genes needed for survival was estimated to be 265–350. Other observations were: (i) single gene knockout experiments are not definitive, because interacting gene products need multiple gene knockouts to really determine what is essential; (ii) the functions of many genes that appear to be essential are unknown. In 2001, Peterson and Fraser (3) reviewed some of the difficulties in a minimal cell project.

Burgard et al. (4) estimated the minimum required growth-sustaining core of metabolic reactions of Escherichia coli under different culture conditions. They found that the minimal reaction network set depends highly on the nutrient environment and the growth requirements imposed. E. coli grown on a medium containing only glucose requires at least 224 metabolic reactions, whereas the same cell cultured in a rich optimally engineered medium could theoretically support growth with 122 metabolic reactions. As expected, they observed that the required reaction set becomes larger with increasing growth rate.

Recently, Kobayashi et al. (5) estimated 271 genes as the minimal set by systematically inactivating single genes in Bacillus subtilis; 192 indispensable genes were found, and 79 others were predicted from the literature.

Pollack et al. (6) published an extensive review of the metabolic activities of Mycoplasmas, developing a list of the minimum enzymatic functions necessary for independent cellular life. It is important to recognize that the presence of a gene is not evidence that the gene is expressed or necessarily expressible. Important work has also been done on the proteome of M. genitalium (7) determining whether genes are being expressed and to what extent.

Other work has focused on finding essential genes; these results will inform discussion on the minimal gene set. An in silico approach using random mutagenesis with transposons estimated that the Mycobacterium tuberculosis genome contains 35% of essential genes (8). Additionally, the highly compacted genome of Nanoarchaeum equitans has been sequenced (9). A more direct approach to identifying essential genes is one with high-density transposon mutagenesis applied to the H. influenzae genome; this study estimated that 38% of the genes are critical for growth (10). Another study tested the deletion of two separate subsets of 53 unknown function genes in E. coli indicating 12 essential genes (11, 12). Judson and Mekalanos (13) describe a vector that allows the identification of essential genes in Gram-negative bacteria by replacing their natural promoter with Phad (arabinose-inducible promoter). They ap...
plied this approach to *Vibrio cholerae* to characterize 16 genes; genome-wide analysis has not been demonstrated. A combination of antisense technology and regulated expression has been used for identification of essential genes from the human pathogen, *Staphylococcus aureus* (14). Identification of essential genes in eukaryotes has also been accomplished. Analysis of chromosome V of *Saccharomyces cerevisiae* by genetic footprinting indicated that ~20% of the genes examined are essential (16).

Many investigators have made significant contributions to our understanding of bacterial metabolism. Some models are based primarily on stoichiometry, flux balance analysis, and metabolic control theory, (e.g., refs. 4, 17–20). Because these models are intrinsically static, they have limited ability to predict aspects of cell regulation and dynamic response (although by the addition of constraints, such as uptake rates of a nutrient, these models provide some insight into the dynamic state that can be achieved). Others have proposed methodology to incorporate more directly dynamic (kinetic) information into models of central metabolism (e.g., ref. 21). Others have attempted to model whole cells (22, 23), but those models, while attempting to be whole-cell models, neglect important nonmetabolic aspects of cell growth (e.g., control of chromosome replication or spatial issues associated with position of septa in cells). These studies and many others have contributed toward a systems biology perspective. However, so far all of these approaches are “incomplete” descriptions. Incomplete descriptions may lead to inaccurate conclusions, because there is an implicit assumption in such studies. The assumption is illustrated by considering the metabolic flux analysis of an isolated pathway. As shown by Schlosser and Bailey (24), such analysis is correct only if the output of the pathway cannot influence any input into the pathway. Any cell model that is incomplete assumes that no output of the model either directly or indirectly can influence any input or state within the model. A whole-cell model simulates the dynamics of intracellular processes, predicts cell responses, permits a detailed view of the dynamic changes inside the cell, and provides a more complete test of pathway analyses and metabolic manipulations. A whole-cell model facilitates understanding of the basic principles of cell function.

A minimal cell can be defined with experimental, computational, and theoretical approaches. However, no matter which approach is used to define the problem, there is agreement that one of the most important features of a true minimal genome is that it has the ability to replicate (25). The base model is the Cornell *E. coli* model (26) that accounts for control of chromosome replication, cell division, cell shape, and volume changes. The prototype version of the Cornell *E. coli* model was reported in 1979 (27) and modified in 1983 and 1984 (26, 28) to simulate the growth of an individual cell using 19 components. The general structure of the model is sketched in Fig. 1. The model’s cell pseudochemical components represent the aggregation of actual chemicals found within the cell. Each component lumps together species with similar chemistry and metabolic dynamics. The model includes: (i) pseudochemical reactions summarizing the stoichiometry relationships, (ii) kinetic relationships reflecting general dependencies of major metabolic pathways, (iii) metabolic control using the concentration of chemical components as signals, and (iv) evaluation of kinetic and stoichiometric parameters from independent experiments on cells growing exponentially in complex media. The Cornell *E. coli* model has been used to study glucose- (26) and ammonia-limited growth (28) and to simulate a population of cells (29, 30). Further work (31) expanded the model to include the uptake and incorporation of amino acids. More recently, the model was modified with a description of the control of the laci operon that provides the guidelines for general model modifications of transcription and translation mechanisms for future models (32, 33). The potential to use such single-cell models as tools in functional genomics has been discussed (34).

In 2001, Browning and Shuler (35) generalized the Cornell *E. coli* model into a coarse-grain minimal cell model with dimensionless parameters. This model tested the proposition that the ratio of rate constants of a reaction to other rate constants is the key determinant of cellular behavior, not the precise numerical value of the rate constants. This proposition was confirmed for over two orders of magnitude change in the kinetic ratio. Although this model was complete, it was not sufficiently detailed chemically to function as a minimal cell model that could be related directly to genomic structure. We refer to that model as a coarse-grain minimal cell model because of the lumping of chemical species into aggregated pseudochemical components that act as modules inside the coarse-grain model.

We hypothesize that the modular nature of the model will allow us to expand the description of a lumped compartment to identify all stable chemical species and the corresponding genes minimally necessary for a functionally complete cell. A chemically detailed delumped compartment will still maintain interconnections with other modules in a whole-cell model, allowing the detailed module to be embedded in a fully complete cell model. Showing that a modular unit is fully functional within a complete cell model is a critical step toward a totally genetically defined minimal cell model. The process of model building is facilitated if modules can be attacked in parallel and then combined into a fully complete cell model. We tested this hypothesis with simulations of a coarse-grain and partially genetically detailed minimal cell model (or mixed minimal cell model) that takes two of the modules (P_r, ribonucleotides and P_s, deoxyribonucleotides) and converts them into the minimum set of distinct chemical and genes that support life.

**Materials and Methods**

The *E. coli* single-cell model has been fully described (26–28, 31, 36). The coarse-grain minimal cell model (35) consists of nearly 40 nonlinear equations, which compute the interactions, synthesis, and degradation of the model components as well as algebraic expressions monitoring cycle parameters. The structure of the mixed minimal cell model allows predictions of...
physical quantities such as cell size and shape. In addition, the increased complexity enables the model to respond more realistically to changes in the nutritional environment.

Kinetic Equations. We incorporated into the coarse-grain minimal cell model 28 nonlinear differential equations that describe nucleotide metabolism. The equations were formulated by writing pseudochemical reactions that include the relationship between the nucleotides inside the cell, developing kinetic relationships that reflect the metabolic pathways and including metabolic control using the concentration of the chemical components as signals. These kinetic relationships are typically semiempirical in form and reflect the known factors modulating activity. Tables 2 and 3, which are published as supporting information on the PNAS web site, describe the nomenclature for the symbols representing the macromolecules in nucleotide metabolism and all of the equations currently used in the nucleotide submodel, respectively.

Kinetic Parameters. The 28 nonlinear differential equations include 136 new parameters, as equilibrium rate constants, stoichiometric values, maximum rates, and saturation/inhibition constants. The exact numerical value of the kinetic parameters is not critical to a hypothetical minimal cell; however, these numbers need to be physiologically plausible (to allow the cell to divide for essentially an infinite number of generations). More importantly, they must be in the correct relationship to one another (for details, refer to ref. 35). We require experimental evidence for the existence of the genes to be included in the mixed minimal cell model. M. genitalium is the smallest organism alive, but very few of its enzymes have been characterized kinetically. Enzyme studies are also very limited for other Mycoplasma species. The best-characterized organisms for parameter estimation are E. coli, B. subtilis, Bacillus cereus, and, for some parameters, H. influenzae. The enzyme characterizations for these organisms yielded at least three or more kinetic parameters for orthologous genes in the different species. The parameter values vary significantly from one species to another.

Evaluation of Parameters. Values for the kinetic parameters and sources or justification are given in Table 4, which is published as supporting information on the PNAS web site; 27 of the new parameters were obtained from mass balances and stoichiometry. Most of the maximum rate constants can be determined from basic measurements on appropriate enzymes or by using mass balances to determine the maximum net synthesis rate. Some saturation constants were determined from references to the literature, whereas others are estimated based on a prior heuristic rule (27) as 1/25 of the normal intracellular concentration. The stoichiometric values of the identified chemical reactions are used for the stoichiometric coefficients. Parameters were not adjusted to fit experimental data. In some cases, parameters in the literature were reported with several significant figures; in practice, the parameter values were rounded to three significant figures.

Results and Discussion

Transport. The utilization of any exogenous substrate depends on the activity of a membrane transport system that allows entry of the metabolite into the cell (36). A minimal cell would depend on the exogenous supply of many nutrients, because it would be missing most of the biosynthetic pathways that produce precursors of essential macromolecules. The percentage of genes devoted to transport in M. genitalium (10.7%) is not very different from H. influenzae (12.2%). A loss of specificity and broadened substrate selectivity so that a single carrier can transport several metabolites across the cell membrane explain the smaller number of transport systems in Mycoplasmas (37).

The transport mechanism in a mixed minimal cell model can be less specific than those of more evolved bacteria or multispecific (acting on several substrates) to compensate for the loss of specific transporters during genome compaction (37).

The current model already includes the transport of ammonia, glucose, and amino acids. The most important transporter to add to the model is a nucleotide precursor transporter. Experiments show that the incorporation of nucleosides, nucleotides, or free bases occurs in diverse prokaryotic organisms (38–45).

Nucleotide Precursor Transport System. Genomic comparison studies did not find any explicit free base, nucleoside, or nucleotide transporters in M. genitalium. Mushegian and Koonin (1) identified three polyamine transporters. A complete sequence analysis of M. pneumoniae did not identify transport systems for free bases that are essential components of Mycoplasmas growth media (46). One gene, nupG, is responsible for the transport of all nucleosides in E. coli (40), and comparable systems are found in B. cereus, B. subtilis, and H. influenzae. Based on the above observations, we include a single gene product for free base transport in the minimal cell using the following equation:

\[
R_{A3} = \frac{v_{A3}^{*}}{K_{A3}^{*} + C_{A3}^{*}} \frac{P_{21T}}{K_{P2} + \frac{P_{21T}^2}{V}} + \frac{P_{21T}}{K_{P2/21T} + \frac{P_{21T}^2}{V}}
\]

In Eq. 1, the first term on the right-hand side, \(v_{A3}^{*}\), is for the maximum rate of transport; the second term shows the dependency on the external concentration of free bases \(C_{A3}^{*}\); the third term reflects the saturable inhibition of uptake due to the intracellular free bases pool where \(P_2\) is the amount per cell of free bases and \(V\) is the cell volume; and the fourth term reflects the saturable dependence on cellular energy. \(P_{21T}\) is the ATP level; \(K_{A3}^{*}\), \(K_{P2}\), and \(K_{P2/21T}\) are saturation or equilibrium constants.

After incorporation of the free base transporter into the mixed minimal cell model, we have a complete cell model with an explicit linkage to extracellular free bases. (Fig. 6, which is published as supporting information on the PNAS web site, shows the mass of free bases inside the cell. The simulation starts at the initial value and equilibrates after three generations.)

Nucleotide Metabolism. The majority of purines and pyrimidines are present intracellularly as nucleotides. Most of these nucleotides are present in the macromolecules RNA and DNA. Razin and Avigan (47) reported RNA contents ranging from 8 to 17% and DNA from 4 to 7% for eight Mollicute strains. DNA, RNA, and ATP and GTP usage as energy sources generate the greatest net demand for nucleotides. The nutritional source of nucleoside or nucleotide precursors is the medium or host tissues. The minimal requirements for nucleotide precursors for most Mycoplasma species have not been determined.

From the genome sequence of M. genitalium, 21 genes were found in the category of nucleotide transport and metabolism (48). Hutchison’s minimal Mycoplasma genome gene set (2) includes 19 genes in this category. Mushegian and Koonin (1) included 23 genes for nucleotide metabolism in their minimal gene set; they also propose a pyrimidine nucleotide biosynthesis pathway. Kobayashi et al. (5) include 10 genes in the nucleotide category; they also point out that, due to single gene inactivation, the number of genes in their minimal gene set is likely to be underestimated. Missing steps to reach end pathway products are found in reviewing their diagrammatic pathways (see ref. 5). The list of essential genes by Kobayashi et al. (5) may not be complete,
Studied from the Gram-negative bacterium, the enzyme is believed to be essential for DNA and RNA synthesis. In another study, Lu and Inouye found a complementing gene in Hutchison’s knockout experiments did not lead to cell death (2). In another study, Lu and Inouye found a complementing gene in Hutchison’s knockout experiments did not lead to cell death (2). However, disruption of this gene did not affect cell viability except for a mutant phenotype (52). Furthermore, M. genitalium does not encode a homolog of this enzyme; Mushegian and Koonin (1) detected a novel M. genitalium Adenylate Kinase As Nucleoside Diphosphate Kinase (NDK). NDK is a ubiquitous enzyme that catalyzes the ATP-dependent synthesis of ribo- and deoxyribo-nucleoside triphosphates via a phosphoribosylamine intermediate. The structures of NDKs are highly conserved among genes of unknown function. However, disruption of this gene in Hutchison’s knockout experiments did not lead to cell death (2). In another study, Lu and Inouye found a complementing activity for the synthesis of nucleotides triphosphates (53). Assays for NDK activity in proteins in ndk-disrupted cells showed adenylate kinase was the complementing enzyme. Additionally, adenylate kinase is the only enzyme possessing complementary NDK activity among all of the nucleoside diphosphate kinases tested (53–55). The experimental evidence allows us to include adenylate kinase as the gene that catalyzes reactions labeled 4 in Figs. 2 and 3. Published literature from different bacteria (E. coli, B. subtilis, H. influenzae, and Mycoplasma species) and proposed minimal gene sets were carefully analyzed to produce the enzyme list (Table 1). This analysis provided a complete set of nucleotide metabolism enzymes. These studies provide us with some insights into the possible mechanisms for regulation of the nucleotide pools.

Fig. 2 shows the simplified pyrimidine pathway. The minimal cell lacks the capability for de novo synthesis of nucleotides from the simplest precursors but possesses a range of enzyme activities and transport functions to allow the use of free bases. Examination of Mycoplasma species (49) has provided an outline for pathways of purine nucleotide synthesis and salvage pathways. We propose a purine nucleotide metabolism pathway for a minimal cell. This pathway is shown in Fig. 3.

One of the enzymes in the nucleotide metabolism, ribonucleotide reductase that transforms ribonucleotides to deoxyribonucleotides, is already in the Cornell E. coli model (E1 in Fig. 1).

Adenylate Kinase As Nucleoside Diphosphate Kinase (NDK). NDK is a ubiquitous enzyme that catalyzes the ATP-dependent synthesis of ribo- and deoxyribo-nucleoside triphosphates via a phosphoribosylamine intermediate. The structures of NDKs are highly conserved from E. coli to human (43% identity; ref. 50). This enzyme is believed to be essential for DNA and RNA synthesis. Studies from the Gram-negative bacterium, Myxococcus xanthus have suggested that the enzyme may be essential for cell growth (51). However, an ndk gene disruption in E. coli did not affect cell viability except for a mutant phenotype (52). Furthermore, M. genitalium does not encode a homolog of this enzyme; Mushegian and Koonin (1) detected a novel ndk (MG268) among genes of unknown function. However, disruption of this gene in Hutchison’s knockout experiments did not lead to cell death (2). In another study, Lu and Inouye found a complementing activity for the synthesis of nucleotides triphosphates (53). Assays for NDK activity in proteins in ndk-disrupted cells showed adenylate kinase was the complementing enzyme. Additionally, adenylate kinase is the only enzyme possessing complementary NDK activity among all of the nucleoside diphosphate kinases tested (53–55). The experimental evidence allows us to include adenylate kinase as the gene that catalyzes reactions labeled 4 in Figs. 2 and 3. Published literature from different bacteria (E. coli, B. subtilis, H. influenzae, and Mycoplasma species) and proposed minimal gene sets were carefully analyzed to produce the enzyme list (Table 1). This analysis provided a complete set of nucleotide metabolism enzymes. These studies provide us with some insights into the possible mechanisms for regulation of the nucleotide pools.

Fig. 2 shows the simplified pyrimidine pathway. The minimal cell lacks the capability for de novo synthesis of nucleotides from the simplest precursors but possesses a range of enzyme activities and transport functions to allow the use of free bases. Examination of Mycoplasma species (49) has provided an outline for pathways of purine nucleotide synthesis and salvage pathways. We propose a purine nucleotide metabolism pathway for a minimal cell. This pathway is shown in Fig. 3.

One of the enzymes in the nucleotide metabolism, ribonucleotide reductase that transforms ribonucleotides to deoxyribonucleotides, is already in the Cornell E. coli model (E1 in Fig. 1).
Deoxyuridine triphosphatase catalyzes the hydrolysis of dUTP to dUMP and pyrophosphate (step 7, Fig. 2), which is an important feature to prevent uracil incorporation into DNA. This gene is essential for the viability of E. coli (65) and S. cerevisiae (66). dUMP is a weak competitive inhibitor (Table 3, second term, Eq. 12), and dUDP shows competitive inhibition (Table 3, fifth term, Eq. 12) (67).

Purine Metabolic Enzymes. Adenine phosphoribosyltransferase catalyzes the production of AMP (step 8, Fig. 3). This enzyme shows strong feedback inhibition by AMP, ADP, and ATP (Table 3, second, fifth, and sixth terms, respectively, Eq. 18) and particularly guanine nucleotides (49).

Guanine phosphoribosyltransferase, phosphorylates guanosine to GMP (step 9, Fig. 3). This enzyme shows tight feedback inhibition by GTP (Table 3, fifth term, Eq. 24), to prevent excessive salvage of guanosine from the external environment (68).

Adenylate kinase belongs to a family of small monomeric enzymes catalyzing the phosphoribosyl transfer reaction of AMP to ADP (step 4, Figs. 2 and 3). AMP is a strong inhibitor of the enzyme (69), although the mechanisms of inhibition are not known (70). As a NDK, adenylate kinase can use only ATP but not GTP as phosphate donor (Table 3, fourth term, Eqs. 5, 6, 11, 13, 16, 21, 22, 27, and 28). AMP is a potent inhibitor of the NDK function (Table 3, fifth term in equations mentioned above).

Guanylate kinase (GK) catalyzes the reversible phosphoryl transfer from ATP to GMP (step 10, Fig. 3). Characterization of this enzyme has been impaired due to its low abundance (71). The yeast GK kinetic data follow Michaelis–Menten kinetics, with GMP partially inhibiting the enzyme (Table 3, second term, Eq. 25) (72).

Model Response to Integration of Detailed Pyrimidine/Purine Pathways. There are no direct experimental data with which to test the mixed minimal cell model. Therefore, we compared available data for several bacterial organisms over a range of growth rates for glucose limited growth to determine whether the overall response is physiologic in nature. This exercise is an examination of the verisimilitude of the mechanisms imbedded in the model. In Fig. 4, the model’s predictions of cell size are compared to chemostat data, using transmission electron microscopy and gluteraldehyde fixed cells and measurements with a Coulter Counter. The model’s predictions are in close agreement with the trends in terms of the shape and size from the observed data.

Fig. 5 shows the model’s prediction of percent RNA and data from E. coli (73), Klebsiella aerogenes (74), Acinetobacter calcoaceticus (75), Aerobacter aerogenes (76), and Bacillus megaterium (76). The model predicts a trend similar to the experimental data, although the decrease in RNA with growth rate is less severe in the mixed minimal cell model than in the experimental measurements.

The model’s design allows the user to obtain the amount of any of the components at any point in time of the cell cycle or cell generation. A clear example is shown in Fig. 7, which is published as supporting information on the PNAS web site. Simulations for Figs. 5–7 were run with total equal extracellular concentration of free bases in the environment of 1×10⁻³ g/ml.

Conclusion

While maintaining “connectivity,” we have described a mixed model (coarse grain–partial genomically detailed) allowing the prediction of interactions between the environment and the
cellular physiological state with molecular level changes. Additionally, the results suggest that the problem of constructing a complete genetically and chemically detailed model using parallel delumping of modules is feasible because we have shown that we can evaluate the minimal cell responses from both the coarse-grain minimal cell model and mixed minimal cell model independent of each other.

In our mixed minimal cell model approach, we defined our pathways with essential functions and have found examples of specific genes that can accomplish these functions. We have constructed a pathway with 11 functions (12 genes), including secondary activity that permits growth from preformed nucleotides precursors and has complete nucleotide pathways. An example of secondary activity is the ensemble of functions that are performed by adenylyl kinase for several steps in both pyrimidine and purine metabolism. It is certainly possible that a smaller set of genes might be found, but we believe that the set of functions is minimal.

This observation reminds us of one of the important challenges for comparative genomics: nonorthologous gene displacement (same function being performed by unrelated or very distantly related nonorthologous proteins). After a cross-kingdom comparison of 21 bacterial, 4 archaea, and 1 eukaryote genome, only 80 genes (instead of 256) are universally present. Koonin (77) wrote: “It seems that a more general and hence more robust idea is a minimal set of functional niches, most of which can be filled by proteins that belong to two or more distinct families of orthologs. A conserved core of functions with a single, ubiquitous solution certainly exists.” We believe that the minimal cell model could reveal this last solution.

Our work shows that the modular nature of this model allows one to expand the description of a lumped compartment while still maintaining the interconnections and dynamic features with all other pathways in the whole-cell model. The ribonucleotide and deoxyribonucleotide modules contain the minimum number of essential functions and use fewer genes than previously proposed for other minimal gene sets that have verifiable complete nucleotide pathways.

We thank Kelvin Lec, Sam Browning, and Anjali Dhimane for helpful discussions. This work was supported in part by National Science Foundation Grant BES9909133 and the New York State Office of Science Technology and Academic Research.
Table 3. Kinetic equations

\[
\begin{align*}
\frac{dP_2}{dt} &= R_{p_2} \cdot S - \alpha_{p_2} \left( \frac{dP_{23M}}{dt} \right)_s - \left( \frac{dP_{23M}}{dt} \right)_p \\
\frac{dP_{23M}}{dt} &= k_1 \left( \frac{K_{p_{32M}}}{K_{P_{23M}} + \frac{P_{23M}}{V}} \right) \left( \frac{P_{32}}{V} \right) \left( \frac{P_{21T}}{V} \right) - \left( \frac{dP_{23M}}{dt} \right)_s - \left( \frac{dP_{23M}}{dt} \right)_p \\
\frac{dP_{23D}}{dt} &= k_2 \left( \frac{K_{p_{32D}}}{K_{P_{23D}} + \frac{P_{23D}}{V}} \right) \left( \frac{P_{32}}{V} \right) \left( \frac{P_{21D}}{V} \right) - \left( \frac{dP_{23D}}{dt} \right)_s - \left( \frac{dP_{23D}}{dt} \right)_p \\
\frac{dP_{23T}}{dt} &= k_3 \left( \frac{K_{p_{32T}}}{K_{P_{23T}} + \frac{P_{23T}}{V}} \right) \left( \frac{P_{32}}{V} \right) \left( \frac{P_{21T}}{V} \right) - \left( \frac{dP_{23T}}{dt} \right)_s - \left( \frac{dP_{23T}}{dt} \right)_p \\
\frac{dP_{3}}{dt} &= R_{p_3} \cdot S - \alpha_{p_3} \left( \frac{dP_{23M}}{dt} \right)_s - \left( \frac{dP_{23M}}{dt} \right)_p \\
\frac{dP_{25M}}{dt} &= k_4 \left( \frac{K_{p_{25M}}}{K_{P_{25M}} + \frac{P_{25M}}{V}} \right) \left( \frac{P_{25}}{V} \right) \left( \frac{P_{21T}}{V} \right) - \left( \frac{dP_{25M}}{dt} \right)_s - \left( \frac{dP_{25M}}{dt} \right)_p \\
\frac{dP_{25D}}{dt} &= k_5 \left( \frac{K_{p_{25D}}}{K_{P_{25D}} + \frac{P_{25D}}{V}} \right) \left( \frac{P_{25}}{V} \right) \left( \frac{P_{21D}}{V} \right) - \left( \frac{dP_{25D}}{dt} \right)_s - \left( \frac{dP_{25D}}{dt} \right)_p \\
\frac{dP_{25T}}{dt} &= k_6 \left( \frac{K_{p_{25T}}}{K_{P_{25T}} + \frac{P_{25T}}{V}} \right) \left( \frac{P_{25}}{V} \right) \left( \frac{P_{21T}}{V} \right) - \left( \frac{dP_{25T}}{dt} \right)_s - \left( \frac{dP_{25T}}{dt} \right)_p \\
\frac{dP_{3}}{dt} &= R_{p_3} \cdot S - \alpha_{p_3} \left( \frac{dP_{25M}}{dt} \right)_s - \left( \frac{dP_{25M}}{dt} \right)_p
\end{align*}
\]
\[
\frac{dP_25}{dt} = k_1 \left( \frac{K_{PM} + P_{25D}}{V} \right) \left( \frac{P_{25D}}{V + K_{PM25D}} \right) \left( \frac{P_{25T}}{V + K_{PM25T}} \right) \left( \frac{P_{25T}}{V + K_{PM25T}} \right) \left( \frac{P_{25T}}{V + K_{PM25T}} \right) \cdot V - \gamma \left( \frac{P_{25D}}{dt} \right)
\]

\[
\frac{dP_{25T}}{dt} = k_1 \left( \frac{K_{PM} + P_{25T}}{V} \right) \left( \frac{P_{25T}}{V + K_{PM25T}} \right) \left( \frac{P_{25T}}{V + K_{PM25T}} \right) \left( \frac{P_{25T}}{V + K_{PM25T}} \right) \left( \frac{P_{25T}}{V + K_{PM25T}} \right) \cdot V - \gamma \left( \frac{P_{25T}}{dt} \right)
\]

\[
\frac{dP_{25DM}}{dt} = k_1 \left( \frac{K_{PM} + P_{25DM}}{V} \right) \left( \frac{P_{25DM}}{V + K_{PM25DM}} \right) \left( \frac{P_{25DM}}{V + K_{PM25DM}} \right) \left( \frac{P_{25DM}}{V + K_{PM25DM}} \right) \left( \frac{P_{25DM}}{V + K_{PM25DM}} \right) \cdot V - \epsilon \left( \frac{P_{25DM}}{dt} \right)
\]

\[
\frac{dP_{24D}}{dt} = k_2 \left( \frac{K_{PM} + P_{24D}}{V} \right) \left( \frac{P_{24D}}{V + K_{PM24D}} \right) \left( \frac{P_{24D}}{V + K_{PM24D}} \right) \left( \frac{P_{24D}}{V + K_{PM24D}} \right) \left( \frac{P_{24D}}{V + K_{PM24D}} \right) \cdot V - \epsilon \left( \frac{P_{24D}}{dt} \right)
\]

\[
\frac{dP_{14D}}{dt} = k_2 \left( \frac{K_{PM} + P_{14D}}{V} \right) \left( \frac{P_{14D}}{V + K_{PM14D}} \right) \left( \frac{P_{14D}}{V + K_{PM14D}} \right) \left( \frac{P_{14D}}{V + K_{PM14D}} \right) \left( \frac{P_{14D}}{V + K_{PM14D}} \right) \cdot V - \epsilon \left( \frac{P_{14D}}{dt} \right)
\]

\[
\frac{dP_{25ID}}{dt} = k_3 \left( \frac{K_{PM} + P_{25ID}}{V} \right) \left( \frac{P_{25ID}}{V + K_{PM25ID}} \right) \left( \frac{P_{25ID}}{V + K_{PM25ID}} \right) \left( \frac{P_{25ID}}{V + K_{PM25ID}} \right) \left( \frac{P_{25ID}}{V + K_{PM25ID}} \right) \cdot V - \epsilon \left( \frac{P_{25ID}}{dt} \right)
\]

\[
\frac{dP_{25AD}}{dt} = k_2 \left( \frac{K_{PM} + P_{25AD}}{V} \right) \left( \frac{P_{25AD}}{V + K_{PM25AD}} \right) \left( \frac{P_{25AD}}{V + K_{PM25AD}} \right) \left( \frac{P_{25AD}}{V + K_{PM25AD}} \right) \left( \frac{P_{25AD}}{V + K_{PM25AD}} \right) \cdot V - \epsilon \left( \frac{P_{25AD}}{dt} \right)
\]

\[
\frac{dP_{25ID}}{dt} = k_3 \left( \frac{K_{PM} + P_{25ID}}{V} \right) \left( \frac{P_{25ID}}{V + K_{PM25ID}} \right) \left( \frac{P_{25ID}}{V + K_{PM25ID}} \right) \left( \frac{P_{25ID}}{V + K_{PM25ID}} \right) \left( \frac{P_{25ID}}{V + K_{PM25ID}} \right) \cdot V - \epsilon \left( \frac{P_{25ID}}{dt} \right)
\]

\[
\frac{dP_{21M}}{dt} = k_3 \left( \frac{K_{PM} + P_{21M}}{V} \right) \left( \frac{P_{21M}}{V + K_{PM21M}} \right) \left( \frac{P_{21M}}{V + K_{PM21M}} \right) \left( \frac{P_{21M}}{V + K_{PM21M}} \right) \left( \frac{P_{21M}}{V + K_{PM21M}} \right) \cdot V - \epsilon \left( \frac{P_{21M}}{dt} \right)
\]

\[
\frac{dP_{24D}}{dt} = k_2 \left( \frac{K_{PM} + P_{24D}}{V} \right) \left( \frac{P_{24D}}{V + K_{PM24D}} \right) \left( \frac{P_{24D}}{V + K_{PM24D}} \right) \left( \frac{P_{24D}}{V + K_{PM24D}} \right) \left( \frac{P_{24D}}{V + K_{PM24D}} \right) \cdot V - \epsilon \left( \frac{P_{24D}}{dt} \right)
\]
\[
\begin{align*}
\frac{dP_{21D}}{dt} &= k_{21} \left( \frac{K_{P_{10}}}{K_{P_{10}} + P_{21D}/P_{21M}} \right) \left( \frac{P_{31M}}{V} \right) \left( \frac{P_{21F}}{V} \right) \left( \frac{P_{21F}}{K_{P_{21D}/P_{21F}} + P_{21F}/V} \right) \cdot V - \gamma_{21} \left[ \frac{P_{21F} + P_{21D0}}{dt} \right] \\
\frac{dP_{21F}}{dt} &= k_{21} \left( \frac{K_{P_{10}}}{K_{P_{10}} + P_{21F}/P_{21D}} \right) \left( \frac{P_{21D}}{V} \right) \left( \frac{P_{21F}}{V} \right) \left( \frac{P_{21F}}{K_{P_{21D}/P_{21F}} + P_{21F}/V} \right) \cdot V - \gamma_{21} \left[ \frac{dM_{1}}{dt} \right] \\
\frac{dP_{21M}}{dt} &= k_{21} \left( \frac{K_{P_{10}}}{K_{P_{10}} + P_{21M}/P_{21D}} \right) \left( \frac{P_{21D}}{V} \right) \left( \frac{P_{21F}}{V} \right) \left( \frac{P_{21F}}{K_{P_{21D}/P_{21F}} + P_{21F}/V} \right) \cdot V - \gamma_{21} \left[ \frac{dM_{1}}{dt} \right] \\
\frac{dP_{21S}}{dt} &= R_{P_{10}} \cdot S - \alpha_{P_{10}} \left( \frac{dP_{21S}}{dt} \right) \left( \frac{dP_{21}}{dt} \right) \\
\frac{dP_{23M}}{dt} &= k_{23} \left( \frac{K_{P_{10}}}{K_{P_{10}} + P_{23M}/P_{23}} \right) \left( \frac{P_{23}}{V} \right) \left( \frac{P_{21F}}{V} \right) \left( \frac{P_{21F}}{K_{P_{23M}/P_{21F}} + P_{21F}/V} \right) \cdot V - \gamma_{23} \left[ \frac{P_{23M} + P_{21M}}{dt} \right] \\
\frac{dP_{23D}}{dt} &= k_{23} \left( \frac{K_{P_{10}}}{K_{P_{10}} + P_{23D}/P_{23}} \right) \left( \frac{P_{23}}{V} \right) \left( \frac{P_{21F}}{V} \right) \left( \frac{P_{21F}}{K_{P_{23D}/P_{21F}} + P_{21F}/V} \right) \cdot V - \gamma_{23} \left[ \frac{P_{23D} + P_{21D}}{dt} \right] \\
\frac{dP_{23F}}{dt} &= k_{23} \left( \frac{K_{P_{10}}}{K_{P_{10}} + P_{23F}/P_{23}} \right) \left( \frac{P_{23}}{V} \right) \left( \frac{P_{21F}}{V} \right) \left( \frac{P_{21F}}{K_{P_{23F}/P_{21F}} + P_{21F}/V} \right) \cdot V - \gamma_{23} \left[ \frac{P_{23F} + P_{21F}}{dt} \right] \\
\frac{dP_{23S}}{dt} &= k_{23} \left( \frac{K_{P_{10}}}{K_{P_{10}} + P_{23S}/P_{23}} \right) \left( \frac{P_{23}}{V} \right) \left( \frac{P_{21F}}{V} \right) \left( \frac{P_{21F}}{K_{P_{23S}/P_{21F}} + P_{21F}/V} \right) \cdot V - \gamma_{23} \left[ \frac{dM_{3}}{dt} \right] \\
\frac{dP_{24F}}{dt} &= k_{23} \left( \frac{K_{P_{10}}}{K_{P_{10}} + P_{24F}/P_{24}} \right) \left( \frac{P_{24}}{V} \right) \left( \frac{P_{21F}}{V} \right) \left( \frac{P_{21F}}{K_{P_{24F}P_{21F}} + P_{21F}/V} \right) \cdot V - \gamma_{24} \left[ \frac{P_{24F} + P_{21F}}{dt} \right]
\end{align*}
\]
Shown are the equations describing the rate of change of each component in the nucleotide metabolism pathway. S and D used as subscripts denote rate of synthesis and rate of degradation, respectively. Nomenclature is defined in Table 2. Additional symbols used are: $\alpha$- anabolic use of free bases, $\gamma$- conversion of precursor into macromolecule, $\varepsilon$- conversion of precursor into another, $S$ – cellular surface area, $V$-volume.