Computation, prediction, and experimental tests of fitness for bacteriophage T7 mutants with permuted genomes

Drew Endy*†, Lingchong You‡, John Yin‡, and Ian J. Molineux§

*The Molecular Sciences Institute, 2168 Shattuck Avenue, Berkeley, CA 94704; ‡Department of Chemical Engineering, University of Wisconsin, Madison, WI 53706; and §Department of Microbiology and Institute of Cellular and Molecular Biology, University of Texas, Austin, TX 78712

Communicated by Sydney Brenner, The Molecular Sciences Institute, Berkeley, CA, March 8, 2000 (received for review January 12, 2000)

We created a simulation based on experimental data from bacteriophage T7 that computes the developmental cycle of the wild-type phage and also of mutants that have an altered genome order. We used the simulation to compute the fitness of more than 10^5 mutants. We tested these computations by constructing and experimentally characterizing T7 mutants in which we repositioned gene *1*, coding for T7 RNA polymerase. Computed protein synthesis rates for ectopic gene *1* strains were in moderate agreement with observed rates. Computed phage-doubling rates were close to observations for two of four strains, but significantly overestimated those of the other two. Computations indicate that the genome organization of wild-type T7 is nearly optimal for growth: only 2.8% of random genome permutations were computed to grow faster, the highest 31% faster, than wild type. Specific discrepancies between computations and observations suggest that a better understanding of the translation efficiency of individual mRNAs and the functions of qualitatively ''nonessential'' genes will be needed to improve the T7 simulation. *In silico* representations of biological systems can serve to assess and advance our understanding of the underlying biology. Iteration between computation, prediction, and observation should increase the rate at which biological hypotheses are formulated and tested.

genetic networks | evolution | optimization | expression regulation

Research over the last century has produced a wealth of information on the mechanisms and rates for the processes that constitute biological systems. Integration of this information to create quantitative, high-resolution, system-scale models has begun more recently and reflects the necessity of having sufficient information before construction of constrained models. Examples include models for the growth of a single bacterial cell (1, 2), the regulation of genetic circuits (3, 4) and the cell cycle (5, 6), signal transduction (7–9), and metabolic pathways (10). A numerical model of a biological system is a complex hypothesis that can be used to compute the behavior of the system it represents (11). Such a model has heuristic value (12) when used to predict the effects of experimentally uncharacterized perturbations to the system, and the predicted perturbations then are compared with laboratory observations to refine the hypothesis instantiated by the model.

Here, we use a computer simulation for bacteriophage T7 development (13) in conjunction with laboratory experiments to compute, predict, and observe the effect of genome reorganization on T7 development. The simulation treats T7 at the logical scale of a self-replicating unit, from genome entry to production of progeny phage, and is resolved at the level of chemical species. Genome organization in T7 directly regulates the timing and level of gene expression during phage development. Reorganization of the T7 genome therefore should change the timing and level of gene expression and, presumably, the entire process of phage development. Comparison between the computed, predicted, and observed effects of genome reorganization allows for a rigorous test of the mathematical description of T7 development. Furthermore, as the development, function, and fitness of biological systems depend strongly on the regulation of gene expression (see, for example, ref. 14), the ability to compute and predict the consequences of gene expression variation is of general value.

Bacteriophage T7 is one of the best-described biological systems. It has been well studied by using both classical genetics and molecular genetics, and by biochemistry (15–17). The T7 genome (Fig. 1*B*), 39,937 bp of linear double-stranded DNA, contains five *Escherichia coli* promoters, 17 T7 promoters, three transcriptional terminators, and 10 RNase III recognition sites. Aside from the *E. coli* promoters, these genetic elements are named after the first gene that follows, thus R*1.3* is an RNase III recognition site immediately upstream of gene *1.3*. The 56 T7 genes are thought to encode 59 proteins that fall into three classes: class I, transition in metabolism from host to phage; class II, DNA replication; and class III, phage particle and DNA maturation and packaging.

After the initial 850 bp enters the infected cell, transcription causes internalization of the T7 genome (18–21). *E. coli* RNA polymerase (RNAP) initiates transcription from the A1, A2, and A3 promoters on the initial 850 bp and, at a rate of 40 bp per sec (30°C), pulls about 7 kb of the genome into the cell. This segment contains the gene for T7 RNAP (gene *1*). Once synthesized, T7 RNAP pulls the remainder of the genome into the cell at a rate of 250 bp per sec (20). The entire genome entry process takes about 8 min at 30°C, one-third of the normal growth cycle, and necessarily influences the timing of T7 gene expression. Furthermore, given 17 T7 promoters with different strengths and efficiencies distributed across the genome (22, 23), the slow kinetics of genome entry also regulates the levels of gene expression. Several natural T7-like phages are known to have an altered gene order, relative to T7 (I.J.M., unpublished data), but their developmental pathways have not been studied.

Materials and Methods

T7 Simulation. Previously, we used existing experimental data on T7 development to create a simulation for the infection of a single *E. coli* BL21 cell by a single wild-type T7 ($T7^+$) particle (13). The current simulation (T7v2.5) treats the genome as an ordered array of 74 functional genetic elements. Each element represents a section of the T7 genome encoding a gene, promoter, terminator, RNase III recognition site, or terminal repeat. Where sections of DNA encoding distinct functions overlap, multifunction elements are used. Fifty sections of

Abbreviations: RNAP, RNA polymerase; MOI, multiplicity of infection.

†To whom reprint requests should be addressed. E-mail: dendy@molsci.org.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Article published online before print: *Proc. Natl. Acad. Sci. USA*, 10.1073/pnas.090101397. Article and publication date are at www.pnas.org/cgi/doi/10.1073/pnas.090101397

"spacer" DNA that have no assigned function (average length ≈38 bp) are attached to the 3′ end of the upstream genetic element. The positions of the first and last elements, representing the terminal repeats, are fixed. T7v2.5 accounts for the mechanisms and rates of phage genome entry into the host cell, the logical definition of mRNA species based on the order of individual promoters, terminators, and RNase III recognition sites, synthesis of T7 mRNAs and proteins, replication of T7 DNA, assembly of procapsids, and the intracellular assembly of progeny phage. T7v2.5 resolves the above definitions and biochemical processes into a coupled set of differential equations that are integrated numerically. Output from T7v2.5 includes the computed state of genome entry and *in vivo* concentrations of T7 mRNA, proteins, DNA, procapsids, and progeny, each as a function of time after infection. Relative to earlier versions, the logical structure of T7v2.5 has been generalized such that all variables can be computed for T7 mutants whose genetic element order differs from that of wild type (ref. 24; D.E. and L.Y., unpublished data). We did not adjust any experimentally determined parameters used in T7v2.5 to fit observations. Although many parameters could be adjusted to provide a better match to the observed data, doing so would reduce the heuristic value of T7v2.5 and corrupt its value as an assay of our understanding of T7 biology. Source code, documentation, and an interactive version of the T7v2.5 are available at http://virus.molsci.org/t7.

We used T7v2.5 to compute the developmental cycle resulting from the infection of *E. coli* BL21 with all possible 72 gene *1* positional mutants and with 10^5 mutants created by assembling genomes from random permutations of the T7 genetic elements. The gene *1* element contains the T7 DNA from position 3,171 to 5,847 (16). Because host cell lysis is not understood mechanistically, it is not included in T7v2.5. Instead, we computed the intracellular development of each phage for 100 min of infection. As a result, T7v2.5 may overestimate the burst size of strains that would lyse before 100 min. Overestimates may result from the observation that replicated T7 DNA normally is packaged into procapsids with only 25–50% efficiency before cell lysis. Thus, to compare phages, we defined the computed fitness of each phage to be the maximum doubling rate, $\mu_m = \max_t\{\log2[Y(t)]/t\}$, where $Y(t)$ is the computed number of intracellular progeny as a function of time, t. μ_m defines the optimal time for phage induced lysis in environments containing infinite uninfected hosts.

Simulation Assumptions. Details of the assumptions used to construct T7v2.5 were presented elsewhere (13, 24). Notable is that the host cell is well mixed and has a constant volume throughout infection, that nucleoside triphosphates, amino acids, and ribosomes are not limiting, and that DNA packaging is rate limiting for phage particle formation. We also assume that the reactions comprising T7 development can be represented by deterministic kinetics. Other investigators (4, 25, 26) have shown how stochastic kinetics (27) can produce significant variations in gene expression independent of genetic and initial-condition heterogeneity, especially given low concentrations of reactants. However, we chose not to use stochastic kinetics here. The computational cost of a stochastic representation usually exceeds that resulting from a deterministic method, even with improvements in computational algorithms (28). Furthermore, the developmental pathway of T7 does not appear to bifurcate, unlike, for example, the lysis-lysogeny switch in phage lambda, and the use of deterministic kinetics in T7v2.5 does not appear to result in systematic deviations from experimental observation.

Bacteriophage and Bacteria. We constructed ectopic gene *1* phages by inserting the gene *1 Sau*3A I fragment from pAR1219 (29) into the *Bcl*I (*1.7*::*1*), *Bgl*II (*3.8*::*1*), and with an appropriate linker, *Pac*I (immediately downstream of gene *12*) sites of T7 Δ*1–3490* (30). Δ*1*–3490,*1.7*::*1*, Δ*1*–3490,*3.8*::*1*, and Δ*1*–3490, *Pac*I::*1* are referred to below as ecto*1.7*, ecto*3.8*, and ecto*12*, respectively. Recombinant phage genomes were transfected into IJ1127, a gene *1*-containing host, but the resulting phages subsequently were shown to grow in normal strains. We also inserted a noncoding 400-bp λ fragment (20) into the *Bcl*I (*1.7*::*c*) and *Bgl*II (*3.8*::*c*) sites of T7$^+$ DNA for use as control strains. Genes *1.7* and *3.8* are nonessential, but their functions have not been defined. 836::ø was constructed by inserting the 23-bp consensus late promoter at position 836 of T7$^+$. *E. coli* BL21 is *E. coli* B Gal$^-$ λ^s *hsdS*.

Characterization of Phage Development. We observed phage production in the laboratory by intracellular one-step growth experiments performed with *E. coli* BL21 grown aerobically at 30°C in LB media at a multiplicity of infection (MOI) of 0.01. Computed and observed intracellular one-step growth curves were compared by normalizing each data set to its maximum burst (see T7 simulation). We computed phage production directly from T7v2.5.

We observed protein synthesis in *E. coli* BL21 grown aerobically at 30°C in B2 glucose medium plus all amino acids, except cysteine and methionine, at a MOI of 10. The eclipse period and burst size of T7 in this medium are comparable to those in LB. Furthermore, the burst size at a MOI of 10 is comparable to that obtained at low MOI, and from this we assume that protein synthesis is comparable. At 2-min intervals, samples were labeled for 90 sec with [^{35}S]methionine plus [^{35}S]cysteine (30 μCi per ml), followed by a 45-sec chase using LB medium. Cells were lysed with a reducing, SDS-containing buffer, and proteins were separated by electrophoresis through a linear gradient of 7.5–17.5% polyacrylamide. Labeled proteins were visualized by using a PhosphorImager and IMAGEQUANT (Molecular Dynamics) and MACBAS (FujiFilm) software. Proteins that did not separate adequately to allow individual quantification were combined (gp3 and gp3.5; gp10A, gp6, and gp9; gp12 and gp15). Different experiments were normalized relative to each other using an average of observed synthesis rates for three host proteins (≈42 kDa, ≈38 kDa, and ≈30 kDa) and gp0.3, the first T7 protein to be synthesized. Within each strain, data were normalized on the number of cysteine and methionine residues in each protein. We compared the computed and observed data by calculating both absolute protein synthesis rates and the rates relative to wild type. The latter were obtained by dividing the rate of synthesis for each protein in an ectopic gene *1* strain infection by the rate of synthesis for the same protein during infection by wild type. We computed synthesis rates of all T7 proteins directly from T7v2.5.

Results

Phage Growth: Computation and Prediction. One simple genome perturbation is to reposition a single element in an otherwise constant genome. To explore the effects of such a perturbation, we used T7v2.5 to compute the course of T7 development resulting from the repositioning of the gene *1* element, encoding the T7 RNA polymerase, at each possible element position on an otherwise wild-type genome. Fig. 1*A* shows the computed maximum doubling rate, μ_m, as the gene *1* element is repositioned.

When positional mutants contain gene *1* close to the left end of the genome, where it cannot be expressed because it is upstream of all promoters, T7v2.5 computes that no progeny are produced and μ_m is zero, as expected on logical grounds. Several mutants that contain gene *1* downstream of the *E. coli* promoters, but upstream of its normal position have computed μ_m values greater than that of T7$^+$. Computations indicate that placing gene *1* upstream of its normal position allows it to enter the cell and be expressed earlier, relative to wild type. Because

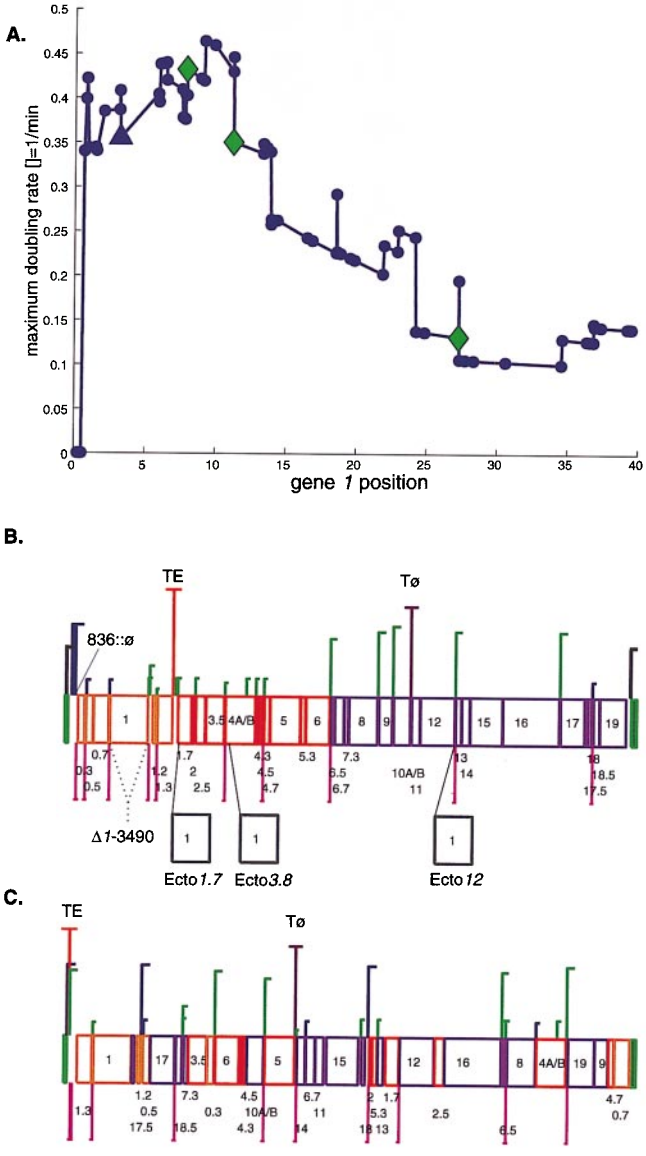

Fig. 1. (*A*) Computed maximum phage doubling rate, μ_m, as the gene *1* element is repositioned on the wild-type T7 genome (*B*). ●, μ_m at the 5′ end of the gene *1* element for 72 positional mutants. ◆, The computed μ_m for the ectopic gene *1* strains that we constructed and characterized experimentally. ▲, The μ_m computed for wild type. (*B*) The wild-type T7 ($T7^+$) genome. Boxes represent coding regions (genes are numbered as space permits), vertical lines with half bars represent *E. coli* (blue) and T7 (green) promoters; bar height is proportional to the observed or estimated strength of each promoter. The transcriptionally inactive T7 promoters, *øOL* and *øOR*, are shown in black. *E. coli* RNAP (TE) and T7 RNAP (Tø) terminators are shown as vertical lines with full bars. RNase III recognition sites are shown as vertical lines below the genome. As the genome is normally depicted, all transcription goes from left to right. The positions of the gene *1* element in the strains constructed and characterized in the laboratory are shown below the genome. (*C*) A random T7 permutation mutant that has a computed μ_m 21% above $T7^+$.

transcription by T7 RNAP is about six times faster than *E. coli* RNAP, earlier expression of gene *1* reduces the time required for genome internalization, increases gene expression rates, and results in a higher μ_m.

More unexpectedly, as gene *1* is positioned between the RNase III-processing sites R*1.1* and R*3.8*, the computed μ_m of the mutants is significantly greater than that of $T7^+$. These computations result from the establishment of an autocatalytic

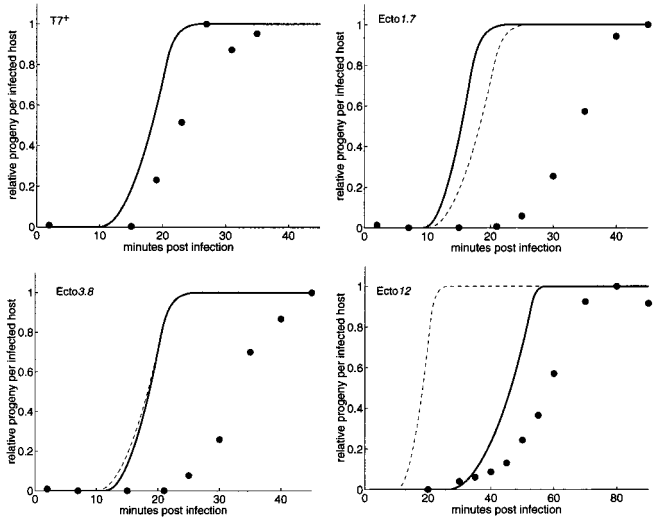

Fig. 2. Computed (solid lines) and observed (●) intracellular one-step growth curves for $T7^+$ and the ectopic gene *1* strains. Both computed and experimental data are normalized to the maximum burst for each strain (simulation: 181; experimental: $T7^+$, 103; ecto*1.7*, 75; ecto*3.8*, 102; ecto*12*, 107) to account for observed variations in DNA packaging efficiency (see *Materials and Methods*). Note the different time scale for ecto*12*. The computed $T7^+$ one-step growth curve (dashed lines) is provided for reference.

loop for gene *1* expression. Gene *1* positively regulates its own synthesis when positioned downstream of a plasmid-borne T7 promoter (31). Our computations indicate that increasing the concentration of active gp1 early in infection increases the expression of class I and class II genes and results in a higher μ_m. From this, we predicted that placing gene *1* downstream of an early T7 promoter, despite the delayed entry of gene *1* into the cell, would increase μ_m.

As gene *1* is positioned even further downstream, computed μ_m values decline, reaching a minimum when gene *1* lies between gene *16* and *ø17*. The computed general decline in μ_m results from the increased time of infection before gene *1* is expressed. However, significant drops in μ_m are computed to occur when gene *1* follows gene *3.5*, gene *4*, R*3.8*, R*13*, and the major transcription terminator Tø. We can easily explain only the last drop: when gene *1* is positioned immediately after Tø, gene *1* expression is reduced and T7 RNAP is limiting for phage development.

Phage Growth: Observation and Comparison to Computation. We constructed three phages in which gene *1* was repositioned within the nonessential genes *1.7* (ecto*1.7*) or *3.8* (ecto*3.8*), or in the intergenic region between gene *12* and promoter *ø13* (ecto*12*) (Fig. 1*B*). Control strains, two containing noncoding λ DNA within gene *1.7* (*1.7*::*c*) and *3.8* (*3.8*::*c*), and one containing the T7 late promoter inserted at position 836 (836::ø), also were made. All ectopic gene *1* strains were viable and eventually produced the same yield of progeny as wild type, although plaques of ecto*12* were small. The major difference, and the probable reason for the small plaque phenotype of ecto*12*, was the length of the eclipse period, which increased as gene *1* was placed even further downstream of its normal position. This repositioning of gene *1* increases the time required for genome entry; e.g., relative to $T7^+$, *E. coli* RNAP must transcribe an extra 21 kb at 40 bp per sec before ecto*12* T7 RNAP is synthesized.

The observed and computed one-step growth curves measuring the production of intracellular progeny over time for $T7^+$ and each ectopic strain are shown in Fig. 2. The observed growth

curves for T7$^+$ and ecto*12* are 4 and 10 min slower than their respective computed curves, but the maximum rates of phage production are comparable. Progeny ecto*1.7* and ecto*3.8* appear about 10 min later than computed and are produced more slowly than expected. We had computed that, relative to T7$^+$, a delay in gene *1* expression in ecto*1.7* and ecto*3.8* would be more than offset by overexpression of the ectopic gene *1* from upstream T7 promoters. We tried to understand these discrepancies by using *1.7*::c, *3.8*::c, and 836::ø. The eclipse period and time to maximum progeny of *1.7*::c were found to be longer than those of T7$^+$ (not shown), and the actual effect of disrupting gene *1.7*, which has no assigned biological function in T7v2.5, contributes about 40% of the difference between the computed and observed growth of ecto*1.7*. Interestingly, cells infected by ecto*1.7* accumulate more than twice the amount of T7 DNA as *1.7*::c (data available at http://virus.molsci.org/t7), an observation we cannot yet explain. Relative to T7$^+$, no significant differences from T7$^+$ have yet been found for *3.8*::c and 836::ø.

Protein Synthesis: Computation and Comparison to Observation. To highlight changes between wild type and the ectopic gene *1* mutants, we performed and quantified pulse–chase protein synthesis experiments. Fig. 3 shows the quantification and scaling process for T7 RNA polymerase. Data for all T7 proteins detected in our experiments were used to create the genomewide comparisons of computed and observed protein synthesis rates shown in Fig. 4 (gels, absolute protein synthesis rates, and cumulative synthesis data available at http://virus.molsci.org/t7).

The computed effect on protein synthesis initiation resulting from the repositioning of gene *1* is shown in Fig. 4 *A–C*. Relative to wild type, there is a genomewide decrease in the synthesis rates of individual proteins early in infection. As expected, the delay of synthesis initiation is most pronounced in ecto*12*. Ecto*1.7* gene *1* autocatalysis (Figs. 3 *C* and *E* and 4*A*) increased

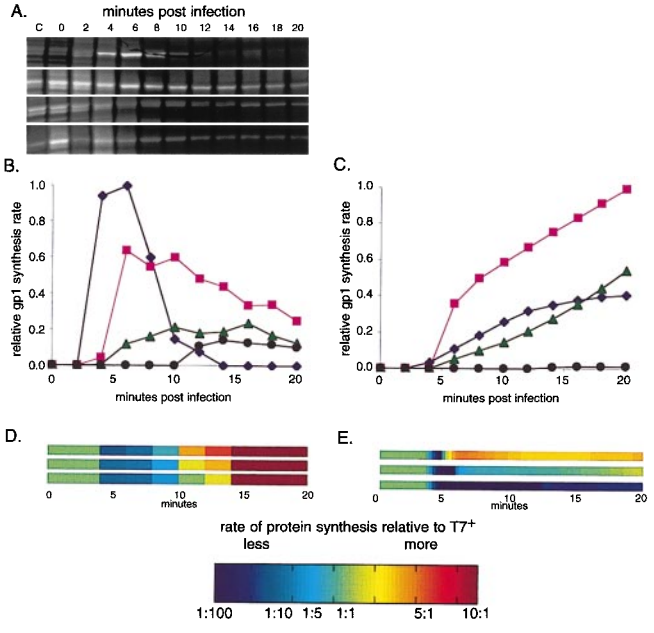

Fig. 3. (*A*) Observed, by pulse–chase labeling experiments, synthesis of T7 proteins. Synthesis of T7 RNA polymerase (gp1) is shown here; top to bottom, T7$^+$, ecto*1.7*, ecto*3.8*, ecto*12*. (*B*) Observed gp1 protein synthesis from *A* normalized to reflect differences between gels; T7$^+$ (◆), ecto*1.7* (■), ecto*3.8* (▲), and ecto*12* (●). (*C*) Computed gp1 protein synthesis (symbols as in *B*). (*D* and *E*) Observed (*D*) and computed (*E*) protein synthesis rates scaled relative to T7$^+$ (see *Materials and Methods*) to compare changes between the ectopic strains and T7$^+$; top to bottom, ecto*1.7*, ecto*3.8*, and ecto*12*.

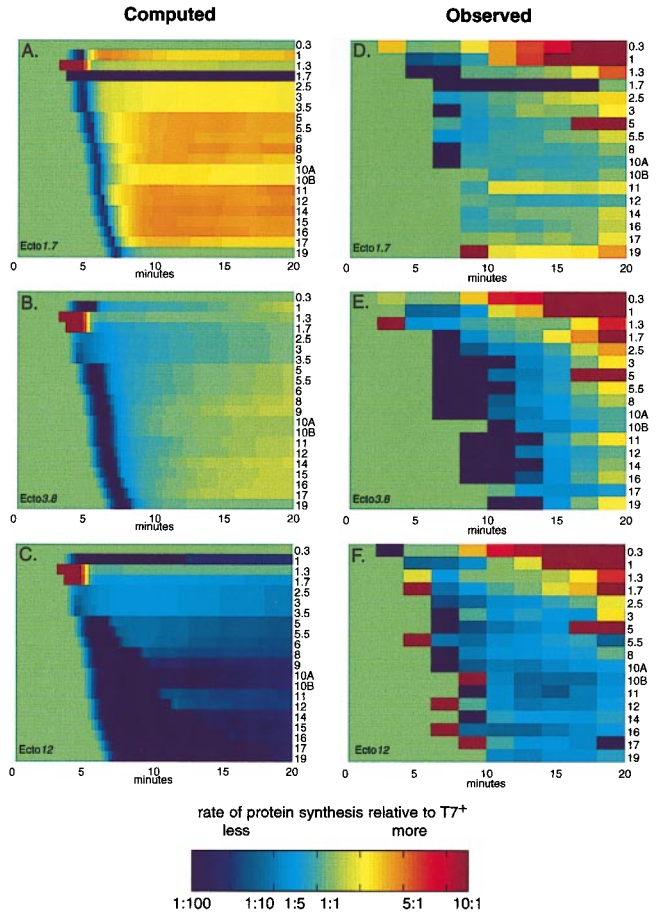

Fig. 4. Protein synthesis rates for the ectopic gene *1* strains relative to wild-type rates (see Fig. 3 legend and *Materials and Methods*). Computed and observed relative synthesis rates are given in the left and right columns, respectively. From top to bottom the plots show data for ecto*1.7*, ecto*3.8*, and ecto*12*, respectively. Proteins are identified by number on the right side of each row.

the computed rate of gp1 synthesis relative to T7$^+$ and, consequently, increased the computed synthesis rates for all later proteins. In turn, the computed increase in ecto*1.7* class II and III protein synthesis rates resulted in its computed higher μ_m.

The observed protein synthesis rates, relative to T7$^+$, demonstrate that changing the position of T7 gene *1* severely perturbs both the time of initiation and rates of synthesis of most T7 proteins (Fig.4 *D–F*). For example, ecto*1.7* gp1 synthesis started late and did not shut off. Furthermore, the observed initiation times of most ecto*1.7* proteins were delayed slightly, relative to T7$^+$. Additionally, neither class I nor class II proteins were shut off normally. Comparable changes were observed after infection by ecto*3.8* and ecto*12*; the largest changes were observed by using ecto*12*.

Comparing the computed and observed protein synthesis rates, relative to wild type, revealed that the computed increase in the synthesis rates of ecto*1.7* class II and class III proteins because of gene *1* autocatalysis were not observed (compare Fig. 4 *A* and *D*). T7$^+$ has the potential for gene *1* autocatalysis through the *øOL* promoter, but RNAs initiated from *øOL* have not been detected. *øOL* is, however, active because it can be used to catalyze genome entry (19). Furthermore, RNAs initiated from *øOL* are found in cells infected by T3, a close relative of T7 (32). Either gene *1* autocatalysis does not occur during ecto*1.7*

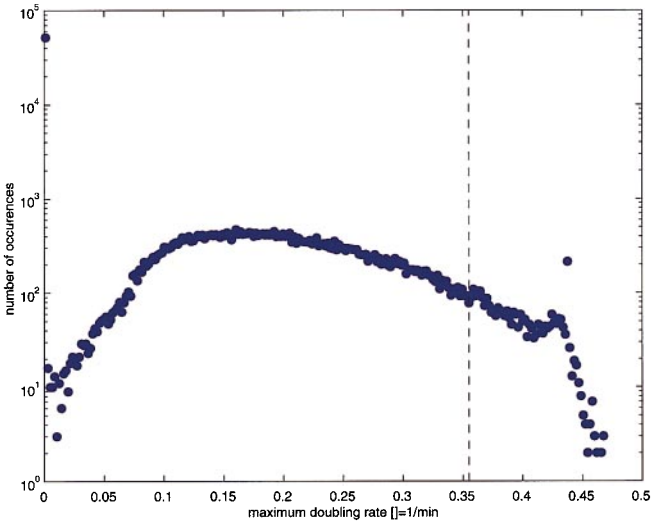

Fig. 5. The computed distribution of maximum doubling rates, μ_m, for a sample ($n = 10^5$) of mutant phage having genomes constructed from random permutations of T7 genetic elements. The dashed vertical line shows the wild-type μ_m (0.355 per min).

infection or its effects are more subtle than currently represented in the simulation.

The observed loss of temporal regulation for the shut-off of class I and class II protein synthesis in the ectopic gene *1* phages also was unanticipated (Fig. 4 *D–F*, and http://virus.molsci.org/t7). Shut-off of class I and II protein synthesis as a consequence of mRNA inactivation or by competition with class III mRNAs has been suggested (33–35), but how either mechanism could produce the effects observed here is unclear. Agreement between computation and observation would be significantly improved by an appreciation of the translation efficiencies of T7 mRNAs.

Random Genome Permutations. We have shown that our understanding of T7 development, as encoded in T7v2.5, is sufficient to successfully predict some of the changes that result from repositioning gene *1* in an otherwise wild-type genome. Our limited success encouraged us to use T7v2.5 to study genomes where the 72 internal genetic elements were randomly ordered. Given a genome that is an array of N genetic elements, there are N! possible permuted mutants (72! or $\approx 10^{104}$). We generated a set of T7 genome permutations at random and computed their μ_m values (Fig. 5). Of the 10^5 permutations studied, 51,643 were computed to not yield progeny: 41,112 because they failed to express all essential genes, 10,531 because they failed to express all essential genes to adequate levels. The remaining 48,357 permutations had μ_m values ranging from above 0 to 0.468 per min distributed about a mean μ_m of 0.204 per min. A total of 97,195 of the permuted genomes had μ_m values below that of the wild type (0.355 per min). A total of 654 permuted genomes had μ_m values significantly above that of $T7^+$, concentrated between 0.42 and 0.44 per min. The permuted genome of one mutant computed to have a μ_m equal to 0.43 per min is shown in Fig. 1*C*.

Discussion

We used a detailed computer simulation, entirely based on experimental data and faithful to current biological knowledge, to compute the development of bacteriophage T7 mutants that contain, relative to the wild type, altered genetic element orders. In doing so, we examined how variation in element order regulates the timing and level of gene expression and phage development. From these computations, we predicted that repositioning T7 gene *1*, encoding T7 RNA polymerase, would significantly affect phage development and, in one instance, increase the maximum phage doubling rate. We created T7 mutants in which gene *1* was repositioned and characterized these mutants experimentally. Ecto*1.7*, the phage predicted to grow faster than $T7^+$, did not achieve the anticipated growth rate in the laboratory. Observed changes in the timing and synthesis rates of individual T7 proteins, because of the repositioning of gene *1*, were in moderate agreement with computations, excepting gene *1* autocatalysis and translational regulation of certain T7 mRNAs. Disagreements between computation and observation suggest areas needing further investigation. Based only on current knowledge, we used T7v2.5 to compute the expected doubling rate for a set of random genome permutations. From these computations, we predicted that the permutation representing the wild-type genome lies at the 97th percentile of maximum phage doubling rate.

In constructing one ectopic gene *1* phage, we disrupted the coding region for the nonessential gene *1.7*. Subsequent experiments using a control phage, *1.7*::c, demonstrated that the expression of this gene, although qualitatively nonessential, has a quantitative effect on T7 development, most likely involving DNA replication. For example, we expect that phages containing an ectopic gene *1* near gene *1.7*, but remain 1.7^+, would make more phage DNA than $T7^+$. Interruption of another nonessential gene, *3.8*, has not been shown to affect phage development. At present, primary function has been assigned to 33 of the 59 expected T7 proteins. Thus, T7 is similar to more complex biological systems in which gene function has been assigned to only a subset of the gene complement. The characterized subset typically contains those genes that are either essential or have an otherwise easily scored phenotype. In the case of T7 at least, characterization of genes that, although nonessential, have quantitative effects on T7 development will be necessary to understand and compute its behavior.

Our experiments and computations support the idea that T7 development is qualitatively insensitive to genome organization. For example, gene *1* is normally expressed within 4 min of infection, and the remainder of T7 development is presumably adapted to that timing. Remarkably, despite a delay in the synthesis of T7 RNAP, corresponding to almost half the normal latent period, we found that ecto*12* produced an observed burst of progeny comparable to that of $T7^+$. Given this robustness of T7 development to repositioning of gene *1*, it was then somewhat less surprising that almost one-half of the 10^5 random genome permutations are computed to yield progeny, admittedly over a wide range of doubling rates. Qualitative system level robustness to change at the level of element order may be a general feature of viral genomes, including viruses that do not regulate expression by DNA entry. Neither ectopically positioning the N gene of vesicular stomatitis virus (36) nor rearranging the order of its P, M, and G genes (37) affected viability of the mutant viruses, although demonstrable changes in the rate of virus development were observed. Similarly, altering the positions of regulatory elements determining the lysis-lysogeny switch in phage λ had only modest effects on phage development (38).

Evolved systems in the wild are presumed to have high fitness, but the wild-type T7 studied today is merely the descendant of an arbitrary isolate taken 55 years ago (39). Although $T7^+$ may indeed have high fitness, the extent to which the genetic element order of $T7^+$ has been optimized to achieve that fitness is unknown. Our computations support the idea that the genetic element order of $T7^+$ is nearly optimal for growth; but the relationship between growth rate in any single environment and system fitness is not straightforward, even for a simple lytic phage. For example, it is possible mutants computed to grow faster than $T7^+$ would not grow well in restrictive environments. The host environment used for our computations provided

unlimited resources for T7 development (see *Materials and Methods*). Limiting resources (e.g., by limiting amino acid or nucleotide pools, or by forcing T7 mRNAs to compete for a finite number of ribosomes) usually reduces the computed maximum doubling rates of those mutants predicted to develop faster in unlimited-resource environments more severely than of $T7^+$ (D.E. and L.Y., unpublished data). Consistent with this idea, mutants that grow faster than wild type, but only under resource-rich conditions, have been selected experimentally (40, 41). Additionally, although 2.8% of all random permutation mutants were computed to have faster maximum doubling rates than $T7^+$, these mutants likely include many predicted to acquire a benefit from gene *1* autocatalysis, an effect that has not yet been demonstrated experimentally. We expect that the ability to compute, construct, and observe how changes in timing and levels of gene expression affect the development of a large number of T7 permutation-mutants in a range of environments will be useful both in studying the evolution of this phage and in quantifying its fitness.

In this initial study on T7 genome organization, repositioning of gene *1* produced major perturbations in T7 development. Computing and predicting the effects of an ectopic gene *1* is difficult because T7 RNAP is central to most aspects of phage development. The perturbations from the normal course of T7 development provided a rigorous test of our ability to predict the behavior of new mutant phage. The partial agreement between computation and observation that we obtained is sufficiently encouraging to continue the ongoing iteration of *in silico*-based computation and laboratory-based observation necessary to improve our understanding of the phage. This work suggests that we will, in time, be able to compute the reproductive properties of more complex organisms once they are sufficiently characterized to constrain computation. However, T7 has been intensively studied and has a relatively simple developmental pathway. Our limited ability to make accurate, quantitative predictions in T7 biology provides a cautionary example of the extent of characterization required for those attempting to model the behavior of more complex organisms. Still, as our experimental knowledge and computational abilities increase, simulations should improve as analysis tools for advancing our understanding of natural systems and could develop into design tools for creating new systems.

We thank Roger Brent and Jim Bull for their critical reading of the manuscript in various forms, and Priscilla Kemp for constructing *1.7*::c and *3.8*::c and for general laboratory assistance. This work was supported by grants from the Office of Naval Research (N00014-99-1-0556, to D.E., N00014-98-1-0226, to J.Y., N00014-97-1-0295, to I.J.M.) and the National Science Foundation (BES-9896067, to J.Y.).

1. Shuler, M. L., Leung, S. & Dick, C. C. (1979) *Ann. N.Y. Acad. Sci.* **326,** 35–55.
2. Domach, M. M., Leung, S. K., Cahn, R. E., Cocks, G. G. & Shuler, M. L. (1984) *Biotechnol. Bioeng.* **26,** 203–216.
3. Shea, M. A. & Ackers, G. K. (1985) *J. Mol. Biol.* **181,** 211–230.
4. Arkin, A. P., Ross, J. & McAdams, H. H. (1998) *Genetics* **149,** 1633–1648.
5. Novak, B., Csikasz-Nagy, A., Gyorffy, B., Chen, K. & Tyson, J. J. (1998) *Biophys. Chem.* **72,** 185–200.
6. Hatzimanikatis, V., Lee, K. H. & Bailey, J. E. (1999) *Biotechnol. Bioeng.* **65,** 631–637.
7. Bray, D., Bourret, R. B. & Simon, M. I. (1993) *Mol. Biol. Cell* **4,** 469–482.
8. Huang, C. Y. & Ferrell, J. E. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 10078–10083.
9. Alon, U., Surette, M. G., Barkai, N. & Leibler, S. (1999) *Nature (London)* **397,** 168–171.
10. Palsson, B. O. & Lee, I. D. (1993) *J. Theor. Biol.* **161,** 299–315.
11. Oreskes, N. (1997) in *10th International Congress of Logic, Methodology, and Philosophy of Science*, eds. Chiara, M. L. D., Doets, K., Mundici, D. & Benthem, J. (Kluwer, Dordrecht, The Netherlands), Vol. 2, p. 207.
12. Oreskes, N., Shrader-Frechette, K. S. & Belitz, K. (1994) *Science* **263,** 641–646.
13. Endy, D., Kong, D. & Yin, J. (1997) *Biotechnol. Bioeng.* **55,** 375–389.
14. Ferea, T. L., Botstein, D., Brown, P. O. & Rosenzweig, R. F. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 9721–9726.
15. Studier, F. W. & Dunn, J. J. (1983) *Cold Spring Harbor Symp. Quant. Biol.* **47,** 999–1007.
16. Dunn, J. J. & Studier, F. W. (1983) *J. Mol. Biol.* **166,** 477–535.
17. Molineux, I. J. (1999) in *Encyclopedia of Molecular Biology*, ed. Creighton, T. E. (Wiley, New York), p. 2495.
18. Zavriev, S. K. & Shemyakin, M. F. (1982) *Nucleic Acids Res.* **10,** 1635–1653.
19. Moffatt, B. A. & Studier, F. W. (1988) *J. Bacteriol.* **170,** 2095–2105.
20. Garcia, L. R. & Molineux, I. J. (1995) *J. Bacteriol.* **177,** 4066–4076.
21. Garcia, L. R. & Molineux, I. J. (1996) *J. Bacteriol.* **178,** 6921–6929.
22. McAllister, W. T. & Wu, H. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 804–808.
23. Ikeda, R. A. (1992) *J. Biol. Chem.* **267,** 11322–11328.
24. Endy, D. (1998) Dissertation (Dartmouth, Hanover, NH).
25. Spudich, J. L. & Koshland, D. E. (1976) *Nature (London)* **262,** 467–471.
26. McAdams, H. H. & Arkin, A. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 814–819.
27. Gillespie, D. T. (1976) *J. Comput. Phys.* **22,** 403–434.
28. Gibson, M. A. & Bruck, J. (2000) *J. Phys. Chem.*, **104,** 1876–1889.
29. Davanloo, P., Rosenberg, A. H., Dunn, J. J. & Studier, F. W. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 2035–2039.
30. Zhang, X. & Studier, F. W. (1995) *J. Mol. Biol.* **250,** 156–168.
31. Dubendorff, J. W. & Studier, F. W. (1991) *J. Mol. Biol.* **219,** 61–68.
32. Adhya, S., Basu, S., Sarkar, P. & Maitra, U. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 147–151.
33. Yamada, Y. & Nakada, D. (1976) *J. Mol. Biol.* **100,** 35–45.
34. Pfennig-Yeh, M. L., Ponta, H., Hirsch-Kauffmann, M., Rahmsdorf, H. J., Herrlich, P. & Schweiger, M. (1978) *Mol. Gen. Genet.* **166,** 127–140.
35. Strome, S. & Young, E. T. (1980) *J. Mol. Biol.* **136,** 417–432.
36. Wertz, G. W., Perepelitsa, V. P. & Ball, L. A. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 3501–3506.
37. Ball, L. A., Pringle, C. R., Flanagan, B., Perepelitsa, V. P. & Wertz, G. W. (1999) *J. Virol.* **73,** 4705–4712.
38. Little, J. W., Shepley, D. P. & Wert, D. W. (1999) *EMBO. J.* **18,** 4299–4307.
39. Demerec, M. & Fano, U. (1945) *Genetics* **30,** 119–136.
40. Studier, F. W. (1979) *Virology* **95,** 70–84.
41. Kong, D. & Yin, J. (1995) *Biotechnology* **13,** 583–586.