1. Historically, what two factors have held back scientists from constructing high-quality whole-cell models? (10 pts)

   Two critical factors in particular have hindered the construction of comprehensive, "whole- cell" computational models. **First, *until recently not enough has been known about the individual molecules and their interactions to completely model any one organism*.** The advent of genomics and other high-throughput measurement techniques have accelerated the characterization of some organisms to the extent that comprehensive modeling is now possible. For example, the mycoplasmas, a genus of bacteria with relatively small genomes that includes several pathogens, have recently been the subject of an exhaustive experimental effort by a European consortium to determine the transcriptome (Güell et al., 2009), proteome (Kuhner et al., 2009), and metabolome (Yus et al., 2009) of these organisms.

   The second limiting factor has been that no single computational method is sufficient to explain complex phenotypes in terms of molecular components and their interactions. The first approaches to modeling cellular physiology, based on ordinary differential equations (ODEs) (Atlas et al., 2008; Browning et al., 2004; Castellanos et al., 2004; Castellanos et al., 2007; Domach et al., 1984; Tomita et al., 1999), were limited by the difficulty in obtaining the necessary model parameters. Subsequently, alternative approaches were developed that require fewer parameters, including Boolean network modeling (Davidson et al., 2002) and constraint-based modeling (Orth et al., 2010; Thiele et al., 2009). However, the underlying assumptions of these methods do not apply to all cellular processes and conditions, and building a whole-cell model entirely based on either method is therefore impractical.

2. What general strategy was used to construct the model in this paper? (10 pts)
   First, we reconstructed the organization of the chromosome including the locations of each gene, transcription unit, promoter, and protein binding site. **Second**, we functionally annotated each gene beginning with the CMR annotation. Functional annotation was primarily based on homologs identified by bidirectional best BLAST. To fill gaps in the reconstructed organism, and to maximize the scope of the model, we expanded and refined each gene's annotation using primary research articles and reviews (see Data S1 and Table S3). **Third**, we curated the structure of each gene product, including the post-transcriptional and post-translational processing and modification of each RNA and protein and the subunit composition of each protein and ribonucleoprotein complex. After annotating each gene, we categorized the genes into 28 cellular processes. We curated the chemical reactions of each cellular process. The reconstruction was stored in a MySQL relational database. See Data S1 and Table S3 for further discussion of the reconstruction.

3. After it was constructed, how was the model validated experimentally? Is there enough data to validate all the parameters defined in the model? (10 pts)
   Next, we validated the model **against a broad range of independent datasets that were not used to construct the model and which encompass multiple biological functions** – metabolomics, transcriptomics, and proteomics – and scales, from single cells to populations. In

agreement with earlier reports (Yus et al., 2009), the model predicts that the flux through glycolysis is >100-fold more than that through the pentose phosphate and lipid biosynthesis pathways (Figure 2E). Furthermore, the predicted metabolite concentrations are within an order of magnitude of concentrations measured in Escherichia coli for 100% of the metabolites in one compilation of data (Sundararaj et al., 2004) and for 70% in a more recent high-throughput study (Bennett et al., 2009; Figure 2F). Our model also predicts "burst-like" protein synthesis due to the local effect of intermittent mRNA expression and the global effect of stochastic protein degradation on the availability of free amino acids for translation, comparable to recent reports by Yu et al., 2006 and So et al., 2011 (Figure 2G). The mRNA and protein level distributions predicted by our model are also consistent with recently reported single-cell measurements (Figure 2H, compare to Taniguichi et al., 2010). Taking all of these specific tests of the model's predictions together, we concluded that our model recapitulates experimental data across multiple biological functions and scales.

However, M. genitalium presents many challenges with regard to experimental tractability. **Resistance to most antibiotics, the lack of a chemically defined medium, and a cell size that requires advanced microscopy techniques for visualization, all greatly limit the range of experimental techniques available to study this organism. As a result, much of the data used to build and validate the model was obtained from other organisms.** Therefore, while the results we report suggest several new experiments that could yield important new insight with respect to M. genitalium function, comprehensive validation of our approach will require modeling more experimentally tractable organisms such as E. coli.

4. Give an example of how this model was used to for the purposes of predictive biology and explain and interpret the results. (10 pts)

    a. Models are often used to predict molecular interactions that are difficult or prohibitive to investigate experimentally.
    b. The model further predicts that the chromosome is explored very rapidly, with 50% of the chromosome having been bound by at least one protein within the first 6 min of the cell cycle, and 90% within the first 20 min.
    c. The model also predicts protein-protein collisions on the chromosome