Supplementary Information
A Biochemical Network Modeling of a Whole-Cell

Paulo E. P. Burke, Claudia B. L. Campos, Luciano da F. Costa, Marcos G. Quiles

1 The Modeling of Particular Interactions

1.1 Cell Division Reaction

The cell division is a biochemical and mechanical event involving several molecules and structures. In the *M. genitalium*’s whole-cell network it was modeled as a single reaction with all necessary molecules linked as modifiers. Figure S1 illustrates the reaction and Table S1.

Figure S1: Cell Division Reaction illustration.
### Table S1: Cell Division Reaction components.

| Illustration Name                          | WholeCellKB WID    | Network ID            |
|--------------------------------------------|--------------------|-----------------------|
| **Reactions**                              |                    |                       |
| Cell Division Reaction                     | -                  | cellular_division_reaction |
| **Molecules**                              |                    |                       |
| GTP                                         | GTP c              | c_GTP                |
| GDP                                         | GDP c              | c_GDP                |
| H2O                                        | H2O c              | c_H2O                |
| PI                                          | PI c               | c_PI                  |
| H                                           | H c                | c_H                  |
| Replication Terminus DNA Region            | -                  | c_Mgenitalium_Chr_1_region_2206_DNA |
| Chromosome Segregation Protein (MraZ)      | MG_470_MONOMER     | c_MG_470_MONOMER      |
| Chromosome Segregation Protein (Era)       | MG_384_MONOMER     | c_MG_384_MONOMER      |
| Chromosome Segregation Protein (CobQ)      | MG_221_OCTAMER     | c_MG_221_OCTAMER      |
| Chromosome Segregation Protein (Obg)       | MG_387_MONOMER     | c_MG_387_MONOMER      |
| Topoisomerase IV                           | MG_203_204_TETRAMER| c_MG_203_204_TETRAMER |
| Cell Division Protein ftsZ                 | MG_224_9MER_GDP    | c_MG_224_9MER_GDP     |
| MgPa Adhesin                               | MG_191_MONOMER     | tm_MG_191_MONOMER     |
| P110 Protein                               | MG_192_MONOMER     | tm_MG_192_MONOMER     |
| P200 Protein                               | MG_386_MONOMER     | tc_MG_386_MONOMER     |
| P32 Adhesin                                | MG_318_MONOMER     | tm_MG_318_MONOMER     |
| P65 Adhesin                                | MG_217_MONOMER     | tc_MG_217_MONOMER     |
| Cytadherence Accessory Protein 3           | MG_317_MONOMER     | tc_MG_317_MONOMER     |
| Cytadherence Accessory Protein 2           | MG_218_MONOMER     | tc_MG_218_MONOMER     |
| Cytadherence Accessory Protein 1           | MG_312_MONOMER     | tc_MG_312_MONOMER     |

### 1.2 Replication Reactions

The template for the replication reactions is described in the main document. Table S2 displays the information about the illustrated nodes.
### Table S2: Replication components.

| Illustration Name | WholeCellKB WID | Network ID |
|-------------------|----------------|------------|
| **Reactions**     |                |            |
| Replication Initiation |              | Mgenitalium_Chr_1_Replication_Initiation |
| Chromosome Region 0 Replication |              | Mgenitalium_Chr_1_region_0_DNA_Replication_Reaction |
| Chromosome Region 1 Replication |              | Mgenitalium_Chr_1_region_1_DNA_Replication_Reaction |
| Chromosome Region 2205 Replication |              | Mgenitalium_Chr_1_region_2205_DNA_Replication_Reaction |
| Chromosome Region 2207 Replication |              | Mgenitalium_Chr_1_region_2207_DNA_Replication_Reaction |
| Chromosome Region 4532 Replication |              | Mgenitalium_Chr_1_region_4532_DNA_Replication_Reaction |
| Chromosome Region 4533 Replication |              | Mgenitalium_Chr_1_region_4533_DNA_Replication_Reaction |
| Replication Terminus |              | Mgenitalium_Chr_1_region_2206_DNA_Replication_Reaction |
| **Molecules**     |                |            |
| dATP              | dATP          | c_dATP     |
| dTTP              | dTTP          | c_dTTP     |
| dCTP              | dCTP          | c_dCTP     |
| dGTP              | dGTP          | c_dGTP     |
| PPI               | PPI           | c_PPI      |
| PI                | PI            | c_PI       |
| Chromosome Region 0 |              | c_Mgenitalium_Chr_1_region_0_DNA |
| Chromosome Region 1 |              | c_Mgenitalium_Chr_1_region_1_DNA |
| Chromosome Region 2 |              | c_Mgenitalium_Chr_1_region_2_DNA |
| Chromosome Region 4534 |              | c_Mgenitalium_Chr_1_region_4534_DNA |
| Chromosome Region 4533 |              | c_Mgenitalium_Chr_1_region_4533_DNA |
| Chromosome Region 2206 |              | c_Mgenitalium_Chr_1_region_2206_DNA |
| DnaA ATP 7mer      |                |            |
| DnaABox Region     |                |            |
| DnaA ADP 7mer      | MG_469_7MER_ADPM | c_MG_469_7MER_ADPM |
| DnaABox Regions    |                |            |
| Replication Complex Region 0 |              | c_Mgenitalium_Chr_1_region_0_Replication_Complex |
| Replication Complex Region 1 |              | c_Mgenitalium_Chr_1_region_1_Replicating_Complex |
| Replication Complex Region 4534 |              | c_Mgenitalium_Chr_1_region_4534_Replication_Complex |
| Replication Complex Region 4533 |              | c_Mgenitalium_Chr_1_region_4533_Replicating_Complex |
| Replication Complex Region 2206 |              | c_Mgenitalium_Chr_1_region_2206_Replicating_Complex |
| DNA Topoisomerase I | MG_122_MONOMER | c_MG_122_MONOMER |
| DNA Topoisomerase IV | MG_203_204_TETRAMER | c_MG_203_204_TETRAMER |
| DNA Gyrase         | DNA_GYRASE    | c_DNA_GYRASE |
| DNA Primase        | MG_250_MONOMER | c_MG_250_MONOMER |
| DNA Helicase       | MG_094_HEXAMER | c_MG_094_HEXAMER |
| DNA Polymerase III Beta | MG_280_MONOMER | c_MG_280_MONOMER |
| DNA Polymerase Gamma Complex | DNA_POLYMERASE_GAMMA_COMPLEX | c_DNA_POLYMERASE_GAMMA_COMPLEX |
| DNA Polymerase Core | DNA_POLYMERASE_CORE | c_DNA_POLYMERASE_CORE |

### 1.3 Transcription Reactions

The template for the transcription reactions is described in the main document. Table S3 displays the information about the illustrated nodes. Once these re-
actions are templates, the exact name of molecules and reactions depends on the gene in the subject. Thus, we use the placeholder \textit{GENE} which can stand for the gene’s name or transcription units for single and polycistronic genes respectively. The placeholder \textit{CHRM,REG} stands for the chromosome region.

Table S3: Transcription components.

| Illustration Name | WholeCellKB WID | Network ID |
|-------------------|-----------------|------------|
| Reactions         |                 |            |
| Transcription Complex Formation - | \textit{GENE\_Transcription\_Complex\_Formation} |
| Transcription Elongation - | \textit{CHRM\_REG\_GENE\_Transcription\_Elongation} |
| Transcription End - | \textit{GENE\_Transcription\_End} |
| RNA Cleavage - | \textit{GENE\_Processing} |
| Maturation Reaction - | \textit{GENE\_Maturation} |
| Molecules         |                 |            |
| ATP               | ATP c\_ATP      |            |
| UTP               | UTP c\_UTP      |            |
| CTP               | CTP c\_CTP      |            |
| GTP               | GTP c\_GTP      |            |
| PPI               | PPI c\_PPI      |            |
| RNA Polymerase    | RNA\_POLYMERASE | c\_RNA\_POLYMERASE |
| RNA Polymerase Holoenzyme | RNA\_POLYMERASE\_HOLOENZYME | c\_RNA\_POLYMERASE\_HOLOENZYME |
| Sigma Factor      | MG\_249\_MONOMER | c\_MG\_249\_MONOMER |
| Transcription Factor - | Depends on the \textit{GENE} |
| DNA Region i with |                 |            |
| Transcription Factor - | Depends on the \textit{GENE} |
| Transcribing Complex - | c\_Mgenitalium\_Chr\_1\_region\_i\_DNA\_GENE\_Transcribing\_Complex |
| DNA Region i+1 - | c\_Mgenitalium\_Chr\_1\_region\_i+1\_DNA\_GENE\_Transcribing\_Complex |
| RNA Cleavage Factors | MG\_282\_MONOMER | c\_MG\_282\_MONOMER |
| RNA Elongation Factors | MG\_141\_MONOMER | c\_MG\_141\_MONOMER |
| RNA Release Factors | MG\_141\_MONOMER | c\_MG\_141\_MONOMER |
| RNA               | MG\_0003\_465   | c\_MG\_0003\_465 |
| RNAse             | MG\_110\_MONOMER | c\_MG\_110\_MONOMER |
| RNAse             | MG\_139\_DIMER  | c\_MG\_139\_DIMER |
| RNAse             | MG\_267\_DIMER  | c\_MG\_267\_DIMER |
| RNAse             | MG\_129\_DIMER  | c\_MG\_129\_DIMER |
| RNAse             | MG\_262\_DIMER  | c\_MG\_262\_DIMER |
| RNAse             | MG\_141\_DIMER  | c\_MG\_141\_DIMER |
| RNAse             | MG\_0003\_465   | c\_MG\_0003\_465 |
| RNAse             | MG\_110\_MONOMER | c\_MG\_110\_MONOMER |
| RNAse             | MG\_139\_DIMER  | c\_MG\_139\_DIMER |
| RNAse             | MG\_267\_DIMER  | c\_MG\_267\_DIMER |
| RNAse             | MG\_129\_DIMER  | c\_MG\_129\_DIMER |
| RNAse             | MG\_262\_DIMER  | c\_MG\_262\_DIMER |
| RNAse             | MG\_141\_DIMER  | c\_MG\_141\_DIMER |
| RNAse             | MG\_0003\_465   | c\_MG\_0003\_465 |
| RNAse             | MG\_110\_MONOMER | c\_MG\_110\_MONOMER |
| RNAse             | MG\_139\_DIMER  | c\_MG\_139\_DIMER |
| RNAse             | MG\_267\_DIMER  | c\_MG\_267\_DIMER |
| RNAse             | MG\_129\_DIMER  | c\_MG\_129\_DIMER |
| RNAse             | MG\_262\_DIMER  | c\_MG\_262\_DIMER |
| RNAse             | MG\_141\_DIMER  | c\_MG\_141\_DIMER |
| RNAse             | MG\_0003\_465   | c\_MG\_0003\_465 |
| RNAse             | MG\_110\_MONOMER | c\_MG\_110\_MONOMER |
| RNAse             | MG\_139\_DIMER  | c\_MG\_139\_DIMER |
| RNAse             | MG\_267\_DIMER  | c\_MG\_267\_DIMER |
| RNAse             | MG\_129\_DIMER  | c\_MG\_129\_DIMER |
| RNAse             | MG\_262\_DIMER  | c\_MG\_262\_DIMER |
| RNAse             | MG\_141\_DIMER  | c\_MG\_141\_DIMER |
| RNAse             | MG\_0003\_465   | c\_MG\_0003\_465 |
| RNAse             | MG\_110\_MONOMER | c\_MG\_110\_MONOMER |
| RNAse             | MG\_139\_DIMER  | c\_MG\_139\_DIMER |
| RNAse             | MG\_267\_DIMER  | c\_MG\_267\_DIMER |
| RNAse             | MG\_129\_DIMER  | c\_MG\_129\_DIMER |
| RNAse             | MG\_262\_DIMER  | c\_MG\_262\_DIMER |
| RNAse             | MG\_141\_DIMER  | c\_MG\_141\_DIMER |
| RNAse             | MG\_0003\_465   | c\_MG\_0003\_465 |
| RNAse             | MG\_110\_MONOMER | c\_MG\_110\_MONOMER |
| RNAse             | MG\_139\_DIMER  | c\_MG\_139\_DIMER |
| RNAse             | MG\_267\_DIMER  | c\_MG\_267\_DIMER |
| RNAse             | MG\_129\_DIMER  | c\_MG\_129\_DIMER |
| RNAse             | MG\_262\_DIMER  | c\_MG\_262\_DIMER |
| RNAse             | MG\_141\_DIMER  | c\_MG\_141\_DIMER |
| RNAse             | MG\_0003\_465   | c\_MG\_0003\_465 |

1.4 Transcription Stall Reactions

A transcription reaction can be interrupted for several reasons. One of them is the collision with other molecules in the same region of a DNA strand. Here we modeled the stall reaction for transcribing complexes when a replication complex is in the next chromosome region. Once the transcription reaction can be interrupted at many chromosome regions, one incomplete RNA molecule is created for each reaction. The name of the molecule carries its sequence.
### Table S4: Transcription Stall components.

| Illustration Name                  | WholeCellKB WID       | Network ID                                      |
|------------------------------------|-----------------------|-------------------------------------------------|
| **Reactions**                      |                       |                                                 |
| Transcription Stall Reaction       | -                     | GENE_Transcription_Complex_Stall                |
| **Molecules**                      |                       |                                                 |
| Transcribing Complex               | -                     | c_Mgenitalium_Chr_1_region_i+1_DNA_GENE_Transcribing_Complex |
| Chromosome Region i                | -                     | c_Mgenitalium_Chr_1_region_i+1_Replication_Complex |
| Replication Complex Region i+1     | -                     | c_Mgenitalium_Chr_1_region_i+1_DNA             |
| DNA Region i                       | -                     | c_Mgenitalium_Chr_1_region_i_DNA              |
| RNA Polymerase                     | RNA_POLYMERASE         | c_RNA_POLYMERASE                                |
| Transcription Elongation Factors   | MG_282_MONOMER        | c_MG_282_MONOMER                                |
| Incomplete RNA                     |                       | c_RNA_SEQUENCE                                  |

### 1.5 RNA Degradation Reactions

The RNA degradation reaction template is depicted in Figure S3. The Peptidyl-tRNA Hydrolase is needed only in the case of aminoacylated tRNAs. Modifications in RNAs were not taken into account due to inconsistencies in WholeCellKB. Table S5 shows the component’s names in WholeCellKB and the network model.
Figure S3: RNA Degradation Template.

Table S5: RNA Degradation components.

| Illustration Name      | WholeCellKB WID   | Network ID       |
|------------------------|-------------------|-----------------|
| Reactions              | RNA Degradation   | RNA_Degradation |
| Molecules              |                   |                 |
| ATP                    | ATP               | c_AMP           |
| UMP                    | UMP               | c_UMP           |
| CMP                    | CMP               | c_CMP           |
| GMP                    | GMP               | c_GMP           |
| PPI                    | PPI               | c_PPI           |
| H2O                    | H2O               | c_H2O           |
| H                      | H                 | c_H             |
| RNAs                  | MG_104_MONOMER   | c_MG_104_MONOMER|
| Peptidyl-tRNA Hydrolase| MG_083_MONOMER   | c_MG_083_MONOMER|
| RNA                    |                   | Depends on the RNA|
| Aminoacid              |                   | Depends on the RNA|

1.6 Translation Reactions

The template for the translation reactions is described in the main document. Table S6 displays the information about the illustrated nodes. Once these reactions are templates, the exact name of molecules and reactions depends on the gene in the subject. Thus, we use the placeholder GENE which can stand for the gene’s name. The placeholder LOC stands for location, which can be: cytosol (c), membrane (m), extracellular (e). The placeholder PROT is indi-
cated in the table as the Protein Monomer. Table S7 lists all the amino acids in the model. Table S8 shows all the tRNAs and their respective amino acid and codons.

Table S6: Translation components.

| Illustration Name       | WholeCellKB WID | Network ID                  |
|-------------------------|-----------------|-----------------------------|
| **Reactions**           |                 |                             |
| Translation Complex Formation | -          | PROT_Translation_Complex_Formation |
| Translation Elongation  | -          | PROT_Translation_Elongation |
| Maturation Reaction     | -          | PROT_Maturation              |
| **Molecules**           |                 |                             |
| GTP                     | GTP            | c_GTP                       |
| GDP                     | GDP            | c_GDP                       |
| H2O                     | H2O            | c_H2O                       |
| PI                      | PI             | c_PI                        |
| H                       | H              | c_H                         |
| Ribosome 70S            | RIBOSOME_70S   | c_RIBOSOME_70S              |
| IF-1                    | MG_173_MONOMER | c_MG_173_MONOMER            |
| IF-2                    | MG_142_MONOMER | c_MG_142_MONOMER            |
| IF-3                    | MG_196_MONOMER | c_MG_196_MONOMER            |
|                         | MG_026_MONOMER | c_MG_026_MONOMER            |
|                         | MG_089_DIMER   | c_MG_089_DIMER              |
| Elongation Auxiliaries  |                 |                             |
|                         | MG_258_MONOMER | c_MG_258_MONOMER            |
|                         | MG_433_DIMER   | c_MG_433_DIMER              |
|                         | MG_435_DIMER   | c_MG_435_DIMER              |
|                         | MG_451_DIMER   | c_MG_451_DIMER              |
| mRNA                    | -              | Depends on the PROT         |
| Chaperones              | -              | Depends on the PROT         |
| Translation Complex     | -              | c_PROT_Translation_Complex  |
| Imature Protein         | -              | c_Imature_PROT              |
| Protein Monomer         | -              | LOC_GENE_MONOMER (PROT)     |
| Modification Metabolites| -              | Depends on the PROT         |
| Modification Enzymes    | -              | Depends on the PROT         |
| Modification Side-Products | -        | Depends on the PROT         |
| Membrane Transporters   | -              | Depends on the PROT         |
Table S7: Aminoacids

| Name                     | WholeCellKB WID | Network ID |
|--------------------------|-----------------|------------|
| Alanine                  | ALA             | c(ALA)     |
| Arginine                 | ARG             | c(ARG)     |
| Asparagine               | ASN             | c(ASN)     |
| Aspartic Acid            | ASP             | c(ASP)     |
| Cysteine                 | CYS             | c(CYS)     |
| Formyl-Methionine        | FMET            | c(FMET)    |
| Glutamine                | GLN             | c(GLN)     |
| Glutamic Acid            | GLU             | c(GLU)     |
| Glycine                  | GLY             | c(GLY)     |
| Histidine                | HIS             | c(HIS)     |
| Isoleucine               | ILE             | c(ILE)     |
| Leucine                  | LEU             | c(LEU)     |
| Lysine                   | LYS             | c(LYS)     |
| Methionine               | MET             | c(MET)     |
| Phenylalanine            | PHE             | c(PHE)     |
| Proline                  | PRO             | c(PRO)     |
| Serine                   | SER             | c(SER)     |
| Threonine                | THR             | c(THR)     |
| Tryptophan               | TRP             | c(TRP)     |
| Tyrosine                 | TYR             | c(TYR)     |
| Valine                   | VAL             | c(VAL)     |
| tRNA | WCKB | WID | Aminoacid | tRNA Network ID | Aminoacylated tRNA Network ID |
|------|------|-----|-----------|----------------|-------------------------------|
| MG471 | ALA  | c_MG471 | MG471_ALA |
| MG472 | ILE  | c_MG472 | MG472_ILE |
| MG475 | SER  | c_MG475 | MG475_SER |
| MG479 | THR  | c_MG479 | MG479_THR |
| MG483 | CYS  | c_MG483 | MG483_CYS |
| MG484 | PRO  | c_MG484 | MG484_PRO |
| MG485 | MET  | c_MG485 | MG485_MET |
| MG486 | ILE  | c_MG486 | MG486_ILE |
| MG487 | SER  | c_MG487 | MG487_SER |
| MG488 | FMET | c_MG488 | MG488_FMET |
| MG489 | ASP  | c_MG489 | MG489 ASP |
| MG490 | PHE  | c_MG490 | MG490_PHE |
| MG492 | ARG  | c_MG492 | MG492_ARG |
| MG493 | GLY  | c_MG493 | MG493/GLY |
| MG495 | ARG  | c_MG495 | MG495_ARG |
| MG496 | TRP  | c_MG496 | MG496_TRP |
| MG497 | ARG  | c_MG497 | MG497_ARG |
| MG499 | GLY  | c_MG499 | MG499/GLY |
| MG500 | LEU  | c_MG500 | MG500_LEU |
| MG501 | LYS  | c_MG501 | MG501_LYS |
| MG502 | GLN  | c_MG502 | MG502/GLN |
| MG503 | TYR  | c_MG503 | MG503_TYR |
| MG504 | TRP  | c_MG504 | MG504_TRP |
| MG506 | SER  | c_MG506 | MG506_SER |
| MG507 | SER  | c_MG507 | MG507_SER |
| MG508 | LEU  | c_MG508 | MG508_LEU |
| MG509 | LYS  | c_MG509 | MG509_LYS |
| MG510 | THR  | c_MG510 | MG510_THR |
| MG511 | VAL  | c_MG511 | MG511_VAL |
| MG512 | THR  | c_MG512 | MG512_THR |
| MG513 | GLU  | c_MG513 | MG513/GLU |
| MG514 | ASN  | c_MG514 | MG514/ASN |
| MG518 | HIS  | c_MG518 | MG518_HIS |
| MG519 | LEU  | c_MG519 | MG519_LEU |
| MG520 | LEU  | c_MG520 | MG520_LEU |
| MG523 | ARG  | c_MG523 | MG523(ARG) |
1.7 Translation Stall Reactions

Just as transcription reactions, the translation process can be interrupted by several reasons too. However, when a transcription complex stalls, the incomplete protein needs to be tagged with a specific amino acid sequence in order to be rapidly degraded. Thus, this process is represented by two template reactions: the stall of the translation complex and the translation of the signal peptide. Once we do not represent intermediate molecules during the translation process, all stalled translation reactions will only produce the same incomplete peptide, which contains only the degradation signal sequence. The reactions’ templates are described in Figure S4 and Table S9.

Figure S4: Translation Stall Template.
Table S9: Translation Stall components.

| Illustration Name          | WholeCellKB WID | Network ID                                                                 |
|----------------------------|-----------------|-----------------------------------------------------------------------------|
| Reactions                  |                 |                                                                             |
| Translation Stall          | -               | PROT_Translation_Complex_Stall                                              |
| Stalled Translation Elongation | -           | Stalled_PROT_Translation_Complex_Translation_Elongation                      |
| Molecules                  |                 |                                                                             |
| GTP                        | GTP             | c,GTP                                                                      |
| GDP                        | GDP             | c,GDP                                                                      |
| H2O                        | H2O             | c,H2O                                                                      |
| PI                         | PI              | c_PI                                                                       |
| H                          | H               | c_H                                                                       |
| Ribosome 70S               | RIBOSOME_70S    | c,RIBOSOME_70S                                                            |
| IF-3                       | MG_196_MONOMER  | c,MG_196_MONOMER                                                           |
|                            | MG_026_MONOMER  | c,MG_026_MONOMER                                                           |
|                            | MG_089_DIMER    | c,MG_089_DIMER                                                            |
|                            | MG_258_MONOMER  | c,MG_258_MONOMER                                                           |
|                            | MG_433_DIMER    | c,MG_433_DIMER                                                            |
|                            | MG_435_DIMER    | c,MG_435_DIMER                                                            |
|                            | MG_451_DIMER    | c,MG_451_DIMER                                                            |
| Elongation Auxiliaries     |                 |                                                                             |
| mRNA                       | -               | Depends on the PROT                                                        |
| Translation Complex        | -               | c_PROT_Translation_Complex                                                  |
| Stalled Translation Complex | -           | c_Stalled_PROT_Translation_Complex                                          |
| tmRNA                      | MG_0004         | c,MG_0004                                                                  |
| Aminoacylated tmRNA        | -               | c,MG_0004_ALA                                                              |
| Proteolysis Tagged Peptide | -               | c_Peptide_ACKSKVNTCLLVNDIQYQHVFIVFV                                       |

1.8 Protein Degradation Reactions

The proteins produced by the cell can be degraded in order to recycle amino acids, control proteins’ concentration, remove defective proteins from the cytosol, and other reasons. Figure S5 and Table S10 shows the template for protein degradation reactions. According to the protein’s location (cytosol or membrane), different proteases can be recruited for its degradation. Proteins tagged with the Proteolysis Peptide are degraded by the membrane protease.
Figure S5: Protein Degradation Template.
Table S10: Protein Degradation components.

| Illustration Name | WholeCellKB WID | Network ID          |
|-------------------|-----------------|---------------------|
| Reactions         |                 |                     |
| Protein Degradation Reaction | - | PROT_Degradation     |
| Molecules         |                 |                     |
| ATP               | ATP             | c_ATP               |
| ADP               | ADP             | c_ADP               |
| H2O               | H2O             | c_H2O               |
| PI                | PI              | c_PI                |
| H                 | H               | c_H                 |
| Protein Monomer   | -               | Depends on the PROT |
| Aminoacids        | -               | Depends on the PROT composition |
| Prosthetic Groups | -               | Depends on the PROT composition |
| Ions              | -               | Depends on the PROT composition |
| Cytosol Protease  | MG_239_HEXAMER  | c_MG_239_HEXAMER    |
| Membrane Protease | MG_457_HEXAMER  | m_MG_457_HEXAMER    |
|                   | MG_020_MONOMER  | c_MG_020_MONOMER    |
|                   | MG_046_DIMER    | c_MG_046_DIMER      |
| Peptidases        | MG_183_MONOMER  | c_MG_183_MONOMER    |
|                   | MG_208_DIMER    | c_MG_208_DIMER      |
|                   | MG_324_MONOMER  | c_MG_324_MONOMER    |
|                   | MG_391_HEXAMER  | c_MG_391_HEXAMER    |

2 Software Structure and Implementation

The software called PiCell was developed to build the Whole-Cell Extended Biochemical Network of *Mycoplasma genitalium* but also being adaptable for other organisms. It is composed of three parts:

- Database Handler
- PiCell Core
- Network Constructor

that can be accessed by Python 3 scripts. The database handler is the interface between databases and the PiCell core. One handler should be implemented for each database to be used as a source of the model. The PiCell Core is responsible for organizing the data obtained from databases and create intermediate molecules and reactions in order to fulfill the central dogma of biology in the model. When all necessary information is gathered in the PiCell Core, it can be exported as a single network model, with linked molecule and reaction nodes, following the framework proposed in this work. This model is then
further submitted to validation and analyses. In Figure S6 the reader can find a schematic of the software implemented to build the *M. genitalium*'s network.

### 2.1 Database Handler

The necessary information for the model was acquired from the WholeCellKB through the WholeCellKB Handler, a piece of Python 3 code implemented specifically for this database. The data in the WholeCellKB database was available in several formats. The JSON format was chosen because of its easiness of access from Python. In addition to the JSON database file, the Handler can read two other files: one containing the database entries to be ignored, and another containing a name mapping to be applied in the database.

![Diagram showing the schematic implementation of the PiCell, a software to build Whole-Cell Extended Biochemical Networks.]

**Figure S6**: The schematic implementation of the PiCell, a software to build Whole-Cell Extended Biochemical Networks.

### 2.2 Model Builder

The control of the modeling is made through an IPython Notebook using the Jupyter interface. Before acquiring the database’s information, the model must be configured. Information about the canonical cellular processes must be provided in order to be constructed from the templates by the PiCell Core. The
The genetic information about the organism must also be provided. In the case of *M. genitalium*, it was also available in the WholeCellKB. The chromosome sequence, chromosome features, genes, and transcription units are necessary to construct the canonical processes.

Molecules and reactions to be added in the model can be retrieved from the database or inserted manually. An example of the latter is the cell division reaction and its structure and components can be found in Figure S1 and Table S1. Reactions, such as metabolic and aminoacylation, were retrieved from the database, as well as the participant molecules.

2.3 PiCell Core

The PiCell Core is responsible to structure the information acquired from databases and inserted manually in such a way that it can be more easily manipulated, checked for inconsistencies, and be further translated into an extended biochemical network.

**Chromosome Representation** The first function of the PiCell Core is to create a representation of the cell’s chromosomes based on the genetic information provided. Each chromosome is divided into regions according to annotated regions and respecting a maximum region length. In the case of *M. genitalium*, the maximum region length was set a very high value so that all the regions’ sizes are only constrained by the annotations in the genome. Transcription Units’ starts and ends were not considered in this process.

**Recursive Creation of Canonical Reactions** The second function of the PiCell Core is to generate missing canonical reactions for macromolecules in the model. This functionality is based on the premise that all macromolecules in the model must have at least one biosynthesis and one degradation reaction. Thus, this process can iterate from protein complexes needing their complexation reaction, up to the expression of their respective genes. For example, consider that a given metabolic reaction inserted in the model is catalyzed by a protein complex. The complex must be synthesized by a protein complexation reaction. The monomers required in this reaction must be synthesized by a translation reaction from their respective mRNA. The mRNA then needs to be synthesized by a transcription reaction from its respective DNA regions. Finally, DNA regions must be synthesized by their replication reactions. This cycle of reactions must be created for every macromolecule in the model. Similarly, the degradation reactions for each macromolecule is created. All reactions created by the PiCell Core are based on the templates described before. Particularities of each reaction created, such as specific chaperones in protein translation, are added in the reactions according to data availability in the database.
Consistency Checks  Additionally to the premise presented in the last paragraph, the PiCell Core performs a mass-balance check in order to probe for inconsistencies in the reactions. All metabolites must have their composition formula described in the model. From their atomic composition, their mass is estimated. Given that all macromolecules are combinations of basic metabolites, the mass of all molecules can be estimated upwards. Then, to check the mass-balance consistency of any reaction, we simply calculate the mass of reactants minus the mass of products. The absolute value obtained must be less than one, the mass of a hydrogen atom. It is important to notice that although this methodology adds an extra layer of confidence in the model, the correctness of all reactions still relies on the data sources.

Extended Biochemical Network Construction  After the model completion, it is ready to generate a working model following the extended biochemical network framework. For each reaction described in the PiCell core, a respective reaction is created in the network. The molecules are created respecting their location. If a given molecule can occur in more than one location, one molecule node is created for each location and linked to their respective reactions accordingly. Reversible reactions are represented by two reaction nodes, one for each direction. The final model can be exported in SBML, some network formats, and also as a networkx graph. The data formats are described in Section 3.

2.4 Software Dependencies

The PiCell is developed using Python 3 and it depends on some Python Packages. The packages are all open source and are listed in the following:

- json
- molmass
- networkx
- libsbml

For the scripts used in the analysis of the model, you will also need the following packages:

- numpy
- matplotlib
- openpyxl
- scipy

3 Network’s Data Formats

The M. genitalium’s whole-cell biochemical network is available in three formats, SBML, GML, and GraphML (Additional file 2).
Molecule nodes are stored as Species objects and reaction nodes as Reaction objects. The common reaction representation of SBML models is used but without kinetic laws associated. Catalytic molecules, such as enzymes, are connected to reactions as modifiers. The SBML file does not contain annotations for molecules or reactions. Also, SBML does not support setting the stoichiometry for the modifiers in reactions. Thus, we always recommend to use the GML or GraphML formats.

All the main nodes and edges attributes are described below. Particular nodes may contain more attributes with additional information.

**General node attributes:**
- annotations: Annotations from WholeCellKB and the authors
- crossreferences: From WholeCellKB
- degree: Number of connections
- indegree: number of inward connections
- instodegree: number of inward connections weighted by their stoichiometry
- name: Unique ID for each node
- originaldatabase: Whether it is from WholeCellKB or created by PiCell
- outdegree: Number of outward connections
- outstodegree: Number of outward connections weighted by their stoichiometry
- stodegree: Number connections weighted by their stoichiometry
- type: Molecule (m) or reaction (r)
- usualname: Human readable name

**Only molecule nodes’ attributes:**
- compartment: cytosol (c), membrane (m), extracellular (e), terminal organelle cytosol (tc), terminal organelle membrane (tm)
- moltype: Type of molecule (metabolite, protein, RNA, DNA, etc.)
- mw: Molecular weight

**Only reaction nodes’ attributes:**
- isreversible: If the reaction is reversible (True/False)
• process: Process where the reaction occurs
• reacttype: Type of the reaction (Metabolic, Protein Synthesis, etc.)

**Edge attributes:**
• type: type of connection: reactant (r), product (p), modifier (m)
• sto: stoichiometry

**Cascading failure algorithm**

**Algorithm 1** Cascading Failure

1: **procedure** recursiveNodeRemoval($N$, $TARGET\_REACTION$) \(\triangleright \) $N$ is the molecule node to remove
2: \hspace{1em} $R \leftarrow$ list of reactions where $N$ is reactant
3: \hspace{1em} $M \leftarrow$ empty list of molecules
4: \hspace{1em} **remove**($N$)
5: \hspace{1em} **while** \(\text{Length}(R) > 0\) **do**
6: \hspace{2em} **for all** $r$ in $R$ **do**
7: \hspace{3em} **for all** products of $r$ **do**
8: \hspace{4em} **if** product.indegree = 0 **then**
9: \hspace{4em} \hspace{1em} **append** product to $M$
10: \hspace{2em} \hspace{1em} **remove**(r)
11: \hspace{1em} **for all** $m$ in $M$ **do**
12: \hspace{2em} \hspace{1em} **append** reactions where $m$ is reactant to $R$
13: \hspace{2em} **remove**(m)
14: \hspace{1em} $M \leftarrow$ empty list of molecules