February 22, 2022 *doi: 10.2210/rcsb_pdb/goodsell-gallery-028*

Integrative Illustration of *Escherichia coli*

David S. Goodsell
Department of Integrative Structural and Computational Biology, The Scripps Research Institute
Research Collaboratory for Structural Bioinformatics Protein Data Bank, Rutgers, The State University of New Jersey

correspondence: goodsell@scripps.edu

***Escherichia coli* is a central model organism for research in molecular and cellular biology. Because of this, abundant information is available for all aspects of its structure and function, making it an attractive subject for whole cell structural modeling. For the past 30 years, I have integrated available information on *Escherichia coli* to create mesoscale illustrations that capture the current state of knowledge. These illustrations depict a cross section through the cell, showing all macromolecules to scale. They have been widely used in educational and outreach settings to depict the crowded nature of cells.**

Historical

In the early 1990s, I posed a hypothesis: "Is it possible to create an illustration of a living cell that shows all macromolecules, drawn to scale and integrating available structural and biophysical information?" I chose *Escherichia coli* as the subject, and at the time, the answer to this question was "Almost." The Science Citation Index was the primary resource for searching the scientific literature, requiring a laborious process of tracking down one piece of information at a time. Proteomic information was primarily available as lists of proteins identified one-by-one from 2D gel electrophoresis. The PDB archive contained about 300 structures and electron microscopy provided only the general contours of large assemblies. Much of my literature search focused on finding information on the abundance of molecular components, which ultimately relied heavily on the work of Neidhardt, Ingraham, and Schaechter [1]. Since the genome had not yet been determined, I also relied on studies of molecular weight distributions and presumed oligomerization states to fill in a representative collection of soluble enzymes. Three 100 nm square cross-sections (cell wall, cytoplasm and nucleoid) were presented [2], reflecting the state of knowledge at the time, tempered, of course, by my ability to find this information.

In the years since then, I have taken several opportunities to update these illustrations, both for integrating new information and for refining the design of the illustrations to improve clarity. The first edition of "The Machinery of Life" in 1993 included similar 100 nm squares using much of the same structural information, but with more artistic license in the orientation of fibrous molecules like DNA to avoid excessive clipping [3]. For an article in American Scientist in 1999, I chose a larger scene rendered in watercolor, developing the ultrastructure-based coloring scheme that I have used in most subsequent work (doi: 10.2210/rcsb_pdb/goodsell-gallery-001) [4]. For the second edition of "Machinery of Life" in 2009, most molecules in the illustration are based on known structures or nucleotide sequences, and the illustration was paired with a less-detailed illustration of the entire cell [5,6]. This whole-cell image was later updated in collaboration with Suckjoon Jun to explore the mechanisms of chromosome replication [7].

Today, the answer to this hypothesis is a nuanced "yes." Most of the central structural aspects of *Escherichia coli* have been studied, including genomes, proteomes with abundances, and atomic structures of many proteins and assemblies from this cell or homologs. Public databases such as GenBank [8], UniProt [9], and PDB [10] have made this information instantly available, and PubMed (pubmed.gov) has revolutionized the way that we access the scientific literature. Together, this availability and findability streamline the gathering and curation of information needed for this effort.

Design Choices

The primary goal of this series of work is to create a cross section of a portion of an *E. coli* cell, magnified to the level where individual molecules are visible (Fig. 1). Several design choices were necessary to make this possible and interpretable. From the beginning, I chose to include only macromolecules, since the illustrations become intractable when small molecules are included. I typically create illustrations at a consistent magnification of 2,000,000X, which is often reduced by about two-fold when printed or presented online. This magnification is sufficient to show a scene several hundred nanometers wide on a typical printed page while including enough detail that the molecular shapes are recognizable. In addition, everything is drawn in orthographic projection, so that the sizes and shapes of all molecules are directly comparable across the image, with no perspective distortion. The current illustration depicts a 225 nm square portion of the cell.

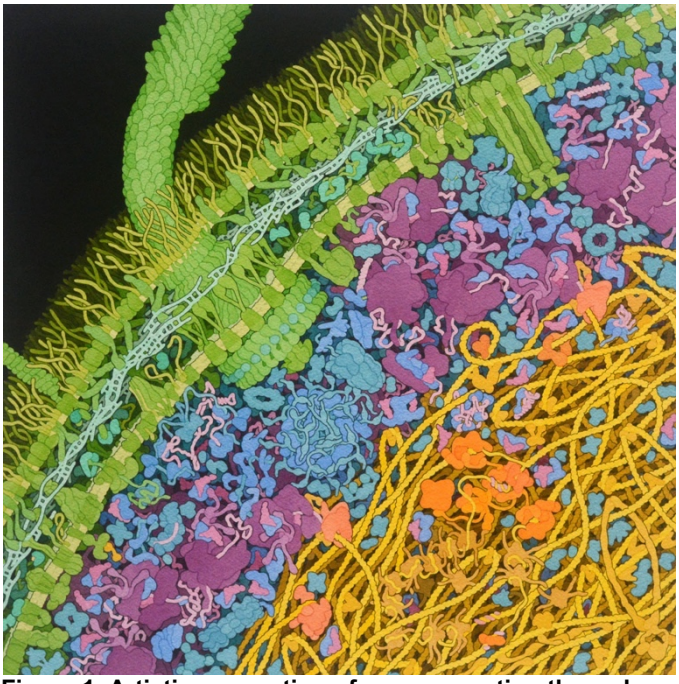

Figure 1. Artistic conception of a cross-section through an *Escherichia coli* cell

Since it is not possible to depict the entire proteome in an image this size, the molecules that are included in the scene are chosen using several steps. The most abundant molecules are added first: the top 20 biological assemblies include outer membrane lipoprotein lpp and several porins, EF-Tu and EF-Ts, ribosomal proteins and trigger factor, acyl-carrier protein, two cold shock proteins, glyceraldehyde-3-phosphate dehydrogenase and alkyl hydroperoxide reductase, ATP synthase, and nucleoid-associated proteins HU, H-NS, and fis [11]. I then added several specialized proteins that are related to the narrative of the illustration but have low copy numbers, including the flagellum and flagellar motor, proteins related to the fimbrium, and several cell wall proteins with interesting structure/function stories. Finally, I filled the remaining spaces with examples from top 200 entries in the proteome, excluding uncharacterized proteins and several proteins involved with processes not depicted in the scene, such as ribosomal assembly and cell division.

With each successive iteration of these *E. coli* illustrations, I have tried to highlight newly-discovered structural features along with a general presentation of major functional processes like protein synthesis and cell wall ultrastructure. In the current illustration, I drastically updated the flagellar motor and added a few new topics to the narrative: the role of intrinsically-disordered proteins in the degradosome, replisome, and ribosome stalk, a selection of non-coding RNA, current conceptions of the DNA replisome, and several remarkable aspects of lipopolysaccharide synthesis and transport.

I allow abundant artistic license to improve the interpretability of the image. This includes careful placement of DNA and peptidoglycan so that none are clipped by the cross-section, and placement of membranes to give appealing contours when clipped. Molecules are oriented with canonical views to allow easy recognition and showing interesting symmetry and interaction relationships. The coloring scheme is chosen to highlight the major ultrastructural compartments of the cell, with the cell wall in greens, nucleoid in yellows and oranges, and soluble cytoplasm in blues and purples. Within each compartment, analogous colors separate different functional classes. In the nucleoid, the DNA is in yellow, nucleoid-associated structural proteins and regulators in tan, and nucleoid-associated enzymes (polymerases and topoisomerases) in orange. In the cytoplasm, RNA-containing molecules (ribosomes, tRNA, mRNA, etc) are in magenta, protein synthesis enzymes and factors in cobalt blues and metabolic enzymes in turquoise. The cell wall uses yellower green to highlight lipopolysaccharides and bluer greens to separate the periplasm.

Finally, I also decided early in this project to be agnostic about the source of information in the image itself. I make no attempt to depict confidence levels or sources for each component, for example, by tuning the representation based on the experimental resolution of each structure. Rather, once I have the best information I can find for each component, I render the entire scene in a consistent style at a consistent level of detail. This choice is designed to simulate a snapshot of the cell, rather than present how each component has been studied. We are currently exploring 3-D models of cells as a mechanism for display of this type of information, where representations, coloring schemes, and annotations are more readily changeable [12].

Data Gathering and Curation

Proteomic information was taken from *E. coli* K-12 strain MG1655 with a MOPS complete growth condition [11]. As in previous work, we assumed a concentration of approximately 30% w/v of macromolecules in the soluble compartment [2]. The remarkably dense packing of outer membrane proteins and lipopolysaccharides was based on a comprehensive review [13], which reports 2-4 lipopolysaccharides in the outer leaflet and 4-10 phospholipids in the inner leaflet per outer membrane protein.

Gene names were used manually to identify proteins at UniProt, which then were used to identify appropriate entries at the PDB archive. If experimental structures were not available, modeled structures were obtained from SWISS-MODEL Repository [14] through links from UniProt. Annotations in UniProt were also used to identify interacting proteins in higher assemblies, and membrane-spanning regions and membrane orientation for cell wall proteins. Structures were viewed in Jmol [15] with a spacefilling representation colored by subunit and measured manually to define the size. Table 1 includes sources of structures depicted in the illustration (Fig. 2).

Identification of appropriate oligomeric states and interactions was by far the most laborious process. Sources in the scientific literature were found individually for each assembly. For typical oligomeric enzymes, the biological assembly included in the PDB archive was used. For more complex assemblies, often several PDB entries needed to be combined and the state depicted in the painting was based on models or schematic diagrams presented in the literature. A few of the challenging examples are described in the section below.

The process of creating this type of illustration provides an opportunity to identify existing gray areas in the structural biology of cells. Non-protein components remain a major challenge. The superhelical state of the DNA is based on modeling work [16,17]. The lipopolysaccharide core is based on PDB ID 6s8n and the polysaccharide chain includes 13-14 tetrasaccharide repeats [18], which are assumed to be in an extended conformation. Peptidoglycan is based on a hexamuropeptide from PDB ID 2mtz, with a wide range of chain lengths centered on 7-9 disaccharides [19]. Strands are depicted as highly crosslinked and mostly parallel to the membranes, although this is still a matter of conjecture and study [20].

Functional Complexes

With each year, added levels of complexity are being revealed within the inner structure of *E. coli* cells. This includes complex collections of molecules that interact to mediate construction and communication across the cell wall and molecular assemblies that manage large-scale tasks within the cytoplasm. Much of this structural understanding is being made possible by the recent revolution in structure determination by cryoelectron microscopy. In this illustration, I have updated or added several large functional complexes.

Flagellar motor. The flagellar motor is arguably the most impressive feature of the cell. Its complexity has been used both as an exemplar of molecular evolution [21], and conversely, to support conjectures of "irreducible complexity" in intelligent design. Abundant structural information is available for components of the motor and flagellum, but at the time of this illustration, they still needed to be integrated into a full conception. The overall structure is based on a 59 Å cryoelectron tomogram from EMD-5311 at the EMDataResource [22] and integrated models presented in several reviews [23,24]. Several atomic structures were also used, including the MS-ring from PDB ID 6sd3, the LP-ring from 7bgl, the rod/MS-ring/LP-ring/hook from 7cgo, the stator from 6ykm and 2zov, the exporter from 5b0o, and the flagellum and hook from 6jy0 and 6k3i.

Lipopolysaccharide synthesis and transport. The process of building and transporting lipopolysaccharides is filled with biomolecular wonders. Several of these are depicted in the illustration. The lipopolysaccharide core is built on the inner face of the cytoplasmic membrane, then flipped to the outer face for construction of the polysaccharide. There, the cone-shaped protein Wzz is one of the proteins thought to act as a ruler to determine the length of the polysaccharide chains [25]. The Lpt system, which spans the entire cell wall, then acts like an elevator to transport the completed lipopolysaccharides to the surface and insert them into the outer membrane. Structures for the components are included in Table 1, and several reviews were used to piece everything together [26–28].

Ribosomes. To me, the ever-expanding understanding of ribosome structure and function most exemplifies the steady march of structural biology. For my 1991 illustration, I relied on the shapes inferred from negative stain electron microscopy that laid the groundwork for early conceptions of ribosome structure and evolution [29]. In the following years, electron tomography continued to provide ever-finer views, and atomic structures were first determined in 2000. Today, detailed structures are available for ribosomes in many stages of initiation, elongation, termination, and inhibition. For the current illustration, I updated the representation of the ribosomes to include many functional features. The ribosomes include the L7/12 stalk that is shown interacting with elongation factors [30]. Trigger factor and other chaperones are shown interacting with the nascent polypeptide chain. Several states are shown, including an initiation complex from PDB ID 5me0, an expressome from 6x9q, and a ribosome bound to a secretory channel from 4v6m.

Degradosome. The role of intrinsically-disordered proteins is becoming increasingly clear in eukaryotes, but examples are also prevalent in bacterial systems, such as the L7/12 ribosome stalk and the degradosome. RNase E is a large protein that plays a catalytic and scaffolding role within the degradosome. The N-terminal domain acts as an endolytic ribonuclease, and the C-terminal half associates with other proteins included in RNA degradation, including the helicase RhlB, polynucleotide phosphorylase, and oddly, the glycolytic enzyme enolase. This scaffolding domain also includes RNA-binding domains and a membrane-targeting sequence. Structures of the components were integrated into the full assembly based on schematic diagrams [31].

Replisome. In previous illustrations, I depicted the replisome as a compact complex of two DNA polymerases and other machinery, based on widely-available schematic diagrams. The new illustration depicts the current conception of the replisome, which includes several DNA polymerases, a helicase complex, and a clamp loader complex, all tethered together by intrinsically-disordered linkers, based largely on a comprehensive review [32]. The transient single-strand portions of the lagging strand are shown interacting with single-stranded DNA-binding protein.

Building and Rendering the Illustration

The illustration was executed using previously reported techniques [33]. Briefly, style sheets for all molecules are created at a consistent magnification and used to develop a full sketch of the foreground of the scene. This is transferred to Arches 300 lb Rough Natural White paper using carbon paper. Color is added using a palette mixed largely from Windsor and Newton cadmium yellow, yellow ochre, Windsor red, viridian hue, cobalt blue, and Old Holland magenta. Background molecules are then added extemporaneously, depth-cued to darker shades using Vandyke brown for warm colors and ivory black for cool colors. Outlines are rendered in India ink with a 00 (0.3 mm) Rapidograph technical pen.

Applications

My own conception of the inner structure of cells has changed over the 30 years of creating these illustrations. In the early illustrations, I used the classic "bag of enzymes" concept of cellular structure: the cytoplasm was filled with soluble molecules and assemblies, and the membranes were studded with a collection of self-contained proteins. As a result, these early illustrations have been widely used to depict the crowded nature of cells, often being used as introductory figures in textbooks or professional presentations. The current illustration depicts a far more complex conception that includes extensive interaction between cytoplasmic, nucleoid, and cell wall components, and elaborate mechanisms for communication, transport, and infrastructure in the cell wall. My hope is that this will inspire additional uses for the illustration.

First, I hope that the illustration will help inspire wonder at the complexity of these so-called "simple" cells. My own sense of wonder continues to grow with each update, as I explore recently-revealed mechanisms. For example, in this current update, I was particularly struck by the Lpt system that spans the two membranes and acts as an elevator that transports lipopolysaccharides to the surface. To me, this is a truly remarkable example of evolutionary nanotechnology.

Second, as computational and bioinformatics hardware and software continue to improve, I am increasingly presenting these illustrations as a challenge to the field of 3D mesoscale modeling. In our own work in modeling bacterial cells, I often create these types of illustrations as a way to identify structural features that require development of new modeling techniques. For example, many aspects of *E. coli* structure are relatively easy to depict in an illustration, but still pose great challenges for modeling, such as the complex interplay of polymerases, topoisomerases, and supercoiling that define the state of DNA in the nucleoid, incorporation of intrinsically-disordered elements, and incorporation of insightful snapshots of multi-state processes such as protein synthesis or lipopolysaccharide synthesis and transport. Illustration provides a relatively nimble method to explore ideas for assembling and depicting these high-order structure/function features, to inform the more time-intensive structural modeling effort.

Availability

Acknowledgments

I gratefully acknowledge the RCSB Protein Data Bank (National Science Foundation DBI-1832184, National Institutes of Health GM133198, and US Department of Energy DE-SC0019749) for ongoing support of my education and outreach work, the Department of Integrative Structural and Computational Biology at Scripps Research, and NIH grant GM120604.

References

1. Neidhardt FC, Ingraham JL, Schaechter M. Physiology of the bacterial cell: a molecular approach. Sunderland, Mass: Sinauer Associates; 1990.
2. Goodsell DS. Inside a living cell. Trends Biochem Sci. 1991;16: 203–206.
3. Goodsell DS. The machinery of life. New York: Springer-Verlag; 1993.
4. Hoppert M, Mayer F. Prokaryotes: Even without membrane-bounded compartments, prokaryotes display a high degree of subcellular organization. Am Sci. 1999;87: 518–525.
5. Goodsell DS. The machinery of life. 2nd ed., corrected. New York: Copernicus Books; 2009.
6. Goodsell DS. Escherichia coli. Biochem Mol Biol Educ. 2009;37: 325–332.
7. Jun S. Chromosome, cell cycle, and entropy. Biophys J. 2015;108: 785–786.
8. Sayers EW, Cavanaugh M, Clark K, Ostell J, Pruitt KD, Karsch-Mizrachi I. GenBank. Nucleic Acids Res. 2020;48: D84–D86.
9. UniProt Consortium. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res. 2019;47: D506–D515.
10. Burley SK, Berman HM, Kleywegt GJ, Markley JL, Nakamura H, Velankar S. Protein Data Bank (PDB): The single global macromolecular structure archive. Methods Mol Biol Clifton NJ. 2017;1607: 627–641.
11. Li G-W, Burkhardt D, Gross C, Weissman JS. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. Cell. 2014;157: 624–635.

12. Maritan M, Autin L, Karr J, Covert MW, Olson AJ, Goodsell DS. Building structural models of a whole mycoplasma cell. J Mol Biol. 2022;434: 167351.
13. Horne JE, Brockwell DJ, Radford SE. Role of the lipid bilayer in outer membrane protein folding in Gram-negative bacteria. J Biol Chem. 2020;295: 10340–10367.
14. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 2018;46: W296–W303.
15. Hanson RM. *Jmol* – a paradigm shift in crystallographic visualization. J Appl Crystallogr. 2010;43: 1250–1260.
16. Hacker WC, Li S, Elcock AH. Features of genomic organization in a nucleotide-resolution molecular model of the Escherichia coli chromosome. Nucleic Acids Res. 2017;45: 7541–7554.
17. Goodsell DS, Autin L, Olson AJ. Lattice models of bacterial nucleoids. J Phys Chem B. 2018;122: 5441–5447.
18. King JD, Berry S, Clarke BR, Morris RJ, Whitfield C. Lipopolysaccharide O antigen size distribution is determined by a chain extension complex of variable stoichiometry in Escherichia coli O9a. Proc Natl Acad Sci. 2014;111: 6407–6412.
19. Koch AL. Length distribution of the peptidoglycan chains in the sacculus of Escherichia coli. J Theor Biol. 2000;204: 533–541. doi:10.1006/jtbi.2000.2039
20. Turner RD, Vollmer W, Foster SJ. Different walls for rods and balls: The diversity of peptidoglycan. Mol Microbiol. 2014;91: 862–874.
21. Liu R, Ochman H. Stepwise formation of the bacterial flagellar system. Proc Natl Acad Sci. 2007;104: 7116–7121.
22. Lawson CL, Patwardhan A, Baker ML, Hryc C, Garcia ES, Hudson BP, et al. EMDataBank unified data resource for 3DEM. Nucleic Acids Res. 2016;44: D396–D403.
23. Carroll BL, Liu J. Structural conservation and adaptation of the bacterial flagella motor. Biomolecules. 2020;10: 1492.
24. Khan S. The architectural dynamics of the bacterial flagellar motor switch. Biomolecules. 2020;10: 833.
25. Whitfield C, Williams DM, Kelly SD. Lipopolysaccharide O-antigens—bacterial glycans made to measure. J Biol Chem. 2020;295: 10593–10609.
26. Collins RF, Kargas V, Clarke BR, Siebert CA, Clare DK, Bond PJ, et al. Full-length, oligomeric structure of Wzz determined by cryoelectron microscopy reveals insights into membrane-bound states. Structure. 2017;25: 806-815.e3.
27. Li Y, Orlando BJ, Liao M. Structural basis of lipopolysaccharide extraction by the LptB2FGC complex. Nature. 2019;567: 486–490.
28. Rai AK, Mitchell AM. Enterobacterial common antigen: Synthesis and function of an enigmatic molecule. Garsin DA, editor. mBio. 2020;11.
29. Lake JA, Henderson E, Clark MW, Matheson AT. Mapping evolution with ribosome structure: intralineage constancy and interlineage variation. Proc Natl Acad Sci. 1982;79: 5948–5952.
30. Diaconu M, Kothe U, Schlünzen F, Fischer N, Harms JM, Tonevitsky AG, et al. Structural basis for the function of the ribosomal L7/12 stalk in factor binding and GTPase activation. Cell. 2005;121: 991–1004.
31. Bandyra KJ, Bouvier M, Carpousis AJ, Luisi BF. The social fabric of the RNA degradosome. Biochim Biophys Acta BBA - Gene Regul Mech. 2013;1829: 514–522.
32. Lewis JS, Jergic S, Dixon NE. The E. coli DNA replication fork. The Enzymes. Elsevier; 2016. pp. 31–88.
33. Goodsell D. Cellular landscapes in watercolor. J Biocommun. 2016;40: 22–26.

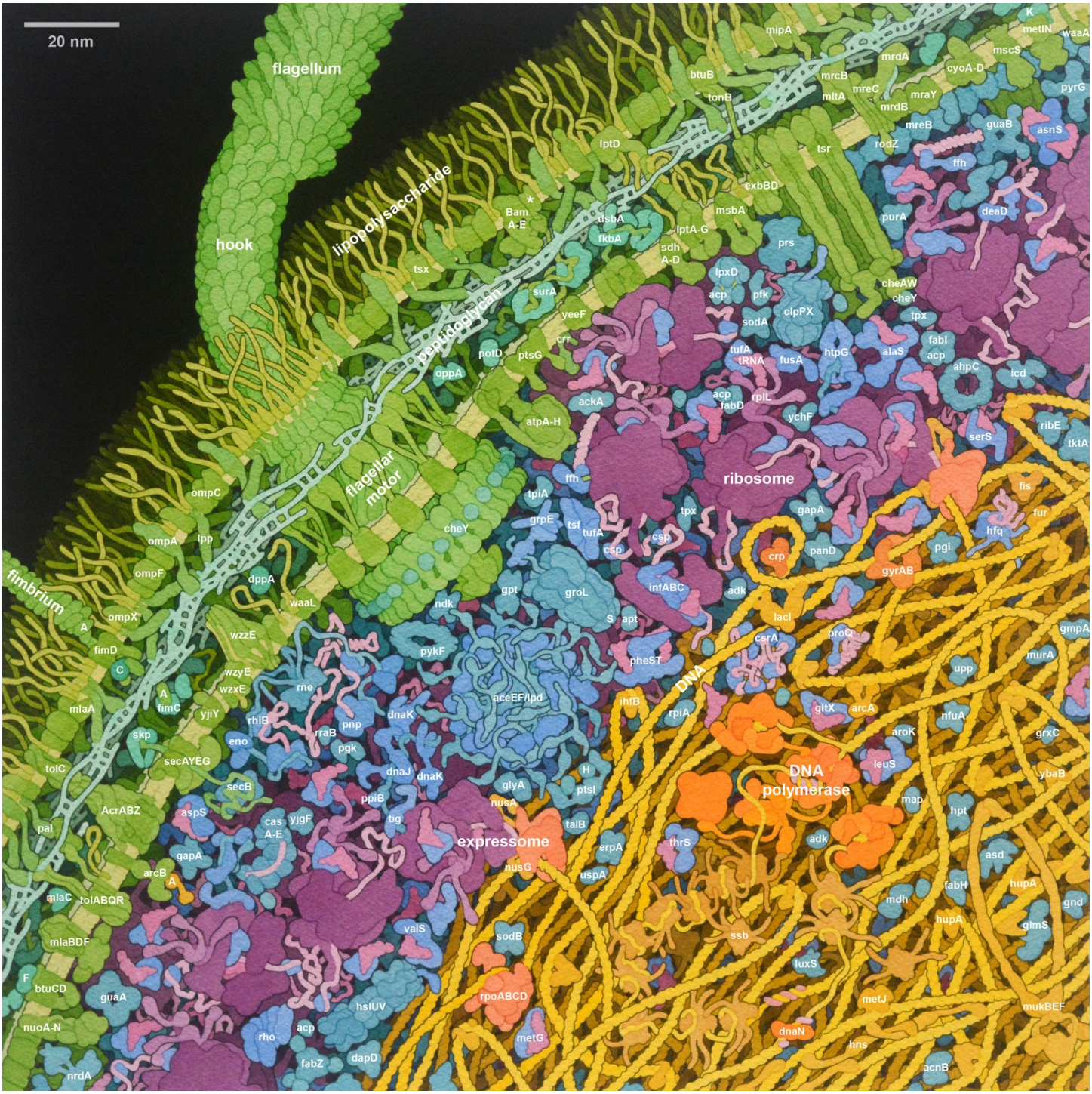

Figure 2. Key to the molecules depicted in the illustration
Biological assemblies are identified with gene names or descriptive names. Refer to Table 1 for full names and sources of structures.

Table 1. Proteins and Structural Sources

Structures from the PDB archive have four-character IDs, with homologs from other bacteria in parentheses. In the absence of a structure, six-character IDs are UniProt entries. Refer to the text for sources for the flagellar motor, ribosome, and replisome.

Gene	PDB/Uniprot	Name
Outer Membrane		
bamA-E	6v05,6lys	Outer membrane protein assembly complex (partially-folded protein denoted with *)
btuB	2gsk	B12 transporter
fimA	6c53	Type-1 fimbrial protein A
fimD	3rfz	Fimbrial usher
lpp	1eq7	Major outer membrane lipoprotein
mlaA	(5nup)	Maintenance of lipid asymmetry
mltA	2ae0	Membrane-bound lytic murein transglycosylase
ompA	1qjp,2mqe	Porin OmpA
ompC	2j1n	Porin OmpC
ompF	1hxt	Porin OmpF
ompX	1qj8	Porin OmpX
pal	2hqs,(2aiz)	Peptidoglycan-associated lipoprotein
tolC, AcrABZ	5ng5	Outer membrane protein TolC
tsx	1tly	Nucleoside-specific channel-forming protein Tsx
mipA	P0A908	MltA-interacting protein
Periplasm		
btuF	2qi9	Vitamin B12-binding protein
dppA	(6pu3)	Heme-binding protein
dsbAB	2zup	Thiol:disulfide interchange protein
fimC	3sqb	Chaperone protein FimC
fkpA	1q6u	FKBP-type peptidyl-prolyl cis-trans isomerase
metQ	6cvl	Methionine-binding protein
mlaC	5uwa	Phospholipid-binding protein
oppA	3tch	Periplasmic oligopeptide-binding protein
potD	1pot	Spermidine/putrescine-binding periplasmic protein
skp	1sg2	Periplasmic chaperone Skp
surA	1m5y	Chaperone SurA
Inner Membrane		
acrABZ, tolC	5ng5	Multidrug efflux pump
arcB	(4i5s)	Aerobic respiration control sensor protein ArcB
atpA-H	6oqr	ATP synthase
btuCD	2qi9	Vitamin B12 import channel
cheAW, tsr	6s1k	Chemotaxis signaling complex
cyoA-D	6wti	Cytochrome bo3
exbBD	5sv1,2pfu	Biopolymer transport protein
lptA-G	4q35,2r1a,6mhu	Lipopolysaccharide export system
metIN	6cvl	Methionine importer
mlaFEDB	6zy4	Maintenance of lipid asymmetry
mraY	(5ckr)	UDP-MurNAc-pentapeptide phosphotransferase
mrcB	3fwl	Penicillin-binding protein 1B
mrdA	6g9p	Penicillin-binding protein 2
mrdB	6pl5	Peptidoglycan glycosyltransferase
mreB	(7bvy)	Rod shape-determining protein
mreC	(2j5u)	Cell shape-determining protein
msbA	6uz2	ATP-dependent lipid A-core flippase
mscS	6pwp	Small conductance mechanosensitive channel
nuoA-N	(6zjy)	NADH-quinone oxidoreductase (Complex I)
ptsG	(5iw),1o2f	PTS system glucose-specific EIICB component
rfaL	P27243	O-antigen ligase
rodZ	(2wus)	Cytoskeleton protein
sdhABCD	1nek	Succinate dehydrogenase iron-sulfur subunit
secAYEG	(3din)	Protein-translocation channel
tolABRQ	(6ye4)	Tol-Pal system

tonB	2gsk	Protein TonB
tsr, cheAW	6s1k	Methyl-accepting chemotaxisis protein I
waaA	(2xci)	Kdo transferase
wzxE	P0AAA7	Lipid flippase
wzyE	P27835	ECA polymerase
wzzE	3b8o	ECA polysaccharide chain length modulation protein
yeeF	P0AA47	Low-affinity putrescine importer
yjiY	(3o7p)	Pyruvate/proton symporter

DNA-associated

arcA	(4kfc)	Aerobic respiration control protein ArcA
crp	1cgp	Catabolite gene activator
fis	3iv5	DNA-binding protein Fis
fusA	(2efg)	Elongation factor G
gyrAB	6rku	DNA gyrase
hns	1hnr,1ni8	DNA-binding protein H-NS
hupA/B	6o8q	DNA-binding protein HU
ihfAB	1owf	Integration host factor
lacI	1efa	Lac repressor
metJ	1cma	Met repressor
mukBEF	3ibp,3euh	Structural maintenance of chromosomes
rpoABCD	6x9q	RNA polymerase
ssb	1eqq	Single-stranded DNA-binding protein
ybaB	1pug	Nucleoid-associated protein YbaB

Cytoplasm

aceEF, lpd	1l8a,1e2o,4jdr	Pyruvate dehydrogenase complex
ackA	(3slc)	Acetate kinase
acnB	1l5j	Aconitate hydratase 2
acpP	1acp	Acyl-carrier protein
adk	1ake	Adenylate Kinase
ahpC	5b8a	Alkyl hydroperoxide reductase C
alaS	(3wqy)	Alanyl-tRNA synthetase
apt	2dy0	Adenine phosphoribosyltransferase
aroK	1kag	Shikimate kinase
asd	1brm	Aspartate-semialdehyde dehydrogenase
asnS	(4wj4)	Asparaginyl-tRNA synthetase
aspS	1c0a	Aspartyl-tRNA synthetase
casA-E	4qyz	CRISPR system Cascade
cheY	1ehc	Chemotaxis protein
clpPX	(6vfx)	ATP-dependent Clp protease
crr	2f3g	EIIA-Glc PTS system
cspA/C/E	1mjc	Cold shock protein
csrA	(2mf1)	Carbon storage regulator
dapD	(1kgt)	2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase
deaD	(5ivl)	Cold-shock DEAD box protein A
dnaJ	(4j80)	Chaperone DnaJ
dnaK	4b9q,2kho	Chaperone DnaK
eno	2fym	Enolase
erpA	(1x0g)	Iron-sulfur cluster insertion protein
fabD	6u0j	Malonyl CoA-ACP transacylase
fabH	1ebl	3-oxoacyl-ACP synthase 3
fabI	2fhs	Enoyl reductase
fabZ	6n3p	(3R)-hydroxymyristoyl-ACP dehydratase
ffh	2xxa	Signal recognition particle protein
gapA	1gad	Glyceraldehyde-3-phosphate dehydrogenase
glmS	1jxa	Glucosamine--fructose-6-phosphate aminotransferase
gltX	(1n77)	Glutamyl-tRNA synthetase
glyA	1dfo	Serine hydroxymethyltransferase
gnd	2zyd	6-phosphogluconate dehydrogenase
gpmA	1 e58	Phosphoglycerate mutase

gpt	1a95	Xanthine phosphoribosyltransferase
groLS	1aon	Chaperonin Gro-EL/ES
grpE	1dkg	Protein GrpE
grxC	1fov	Glutaredoxin-3
guaA	1gpm	GMP synthase
guaB	(4fxs)	IMP dehydrogenase
hfq	4v2s	RNA-binding protein Hfq
hpt	1g9s	Hypoxanthine phosphoribosyltransferase
hslUV	1yyf	ATP-dependent protease ATPase
htpG	2iop	Chaperone protein HtpG
icd	1ai2	Isocitrate dehydrogenase
infABC	5lmv	Translation initiation factor IF1-3
iscUS	3lvl	NifU-like protein
leuS	5omw	Leucyl-tRNA synthetase
lpxD	4ihg	UDP-3-O-(hydroxytetradecanoyl)glucosamine N-acyltransferase
luxS	(5v2w)	S-ribosylhomocysteine lyase
map	1c21	Methionine aminopeptidase
mdh	1emd	Malate dehydrogenase
metG	(2csx)	Methionyl-tRNA synthetase
murA	3kqj	UDP-N-acetylglucosamine 1-carboxyvinyltransferase
ndk	2hur	Nucleoside diphosphate kinase
nfuA	(2z51)	Fe/S biogenesis protein
nrdA	1mrr,3r1r	Ribonucleoside-diphosphate reductase 1
nusA	6x9q	Transcription antitermination protein
nusG	6x9q	Transcription antitermination protein
panD	1aw8	Aspartate 1-decarboxylase
pgi	3nbu	Glucose-6-phosphate isomerase
pgk	1zmr	Phosphoglycerate kinase
pheST	(1eiy)	Phenylalanyl-tRNA synthetase
pnp	3gcm	Polyribonucleotide nucleotidyltransferase
ppiB	1lop	Cyclophilin A
proQ	5nbb, 5nb9	RNA chaperone
prs	4s2u	Ribose-phosphate pyrophosphokinase
ptsH	1cm3	Phosphocarrier HPr
ptsI	2xdf	Phosphoenolpyruvate-protein phosphotransferase
purA	1ade	Adenylosuccinate synthetase
pykF	1e0t	Pyruvate kinase I
pyrG	1s1m	CTP synthase
rhlB	(2db3)	ATP-dependent RNA helicase
rho	1pv4	Transcription termination factor Rho
ribE	1i8d	Riboflavin synthase
rne	2bx2	ribonuclease E
rpiA	1ks2	Ribose-5-phosphate isomerase A
rraB	(1nxi)	Regulator of ribonuclease activity B
secB	1qyn	Protein-export protein SecB
serS	6r1m,(1ser)	Seryl-tRNA synthetase
sodA	1i0h	Superoxide dismutase (Mn)
sodB	1isa	Superoxide dismutase (Fe)
talB	1onr	Transaldolase B
thrS	1qf6	Threonyl-tRNA synthetase
tig	1w26	Trigger factor
tktA	1qgd	Transketolase 1
tpiA	1tre	Triosephosphate isomerase
tpx	3hvs	Thiol peroxidase
tsf	1efu	Elongation factor Ts
tufA/B	1ob2	Elongation factor Tu
upp	1jp3	Undecaprenyl pyrophosphate synthase
uspA	(1jmv)	Universal stress protein A
valS	(1gax)	Valyl-tRNA synthetase
ychF	(1jal)	Ribosome-binding ATPase
yjgF	1qu9	Enamine/imine deaminase