

# Molecular Dynamics Simulations of Periplasmic Proteins Interacting with the Peptidoglycan Layer of *Escherichia coli*

Pranav Lalgudi<sup>1</sup> and Adrian H. Elcock<sup>2</sup>

<sup>1</sup>Lynbrook High School

<sup>2</sup>The University of Iowa

## Summary

**There has long been interest in using *in silico* methods to model complex biological systems, such as cells, for self-evident implications in molecular biology and medicine. Molecular dynamics (MD) simulations are ideal for the construction of such models due to the atomic resolution they provide. We use MD in this work to construct and simulate a preliminary model of the periplasmic space, the peptidoglycan layer and its associated proteins, in an *E. coli* cell. During an initial 23 nanoseconds (ns) MD simulation including 13 of the most abundant periplasmic proteins, the proteins and the peptidoglycan layer were characterized by several interactions. This work provides one of the first steps along the path to creating a complete and comprehensive model of a biological cell.**

**Received:** August 26, 2015; **Accepted:** February 7, 2016; **Published:** April 6, 2016

**Copyright:** © 2016 Lalgudi and Elcock. All JEI articles are distributed under the attribution non-commercial, no derivative license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>). This means that anyone is free to share, copy and distribute an unaltered article for non-commercial purposes provided the original author and source is credited.

## Introduction

There has long been interest in modeling complex biological systems in order to advance biology, medicine, and science. The construction of a complete computational model of a biological cell, for example, could revolutionize the manner in which research is conducted, allowing for testing of hypotheses via *in silico* computer-based approaches in conjunction with experiments performed *in vitro* and *in vivo*, or as a stand-alone approach.

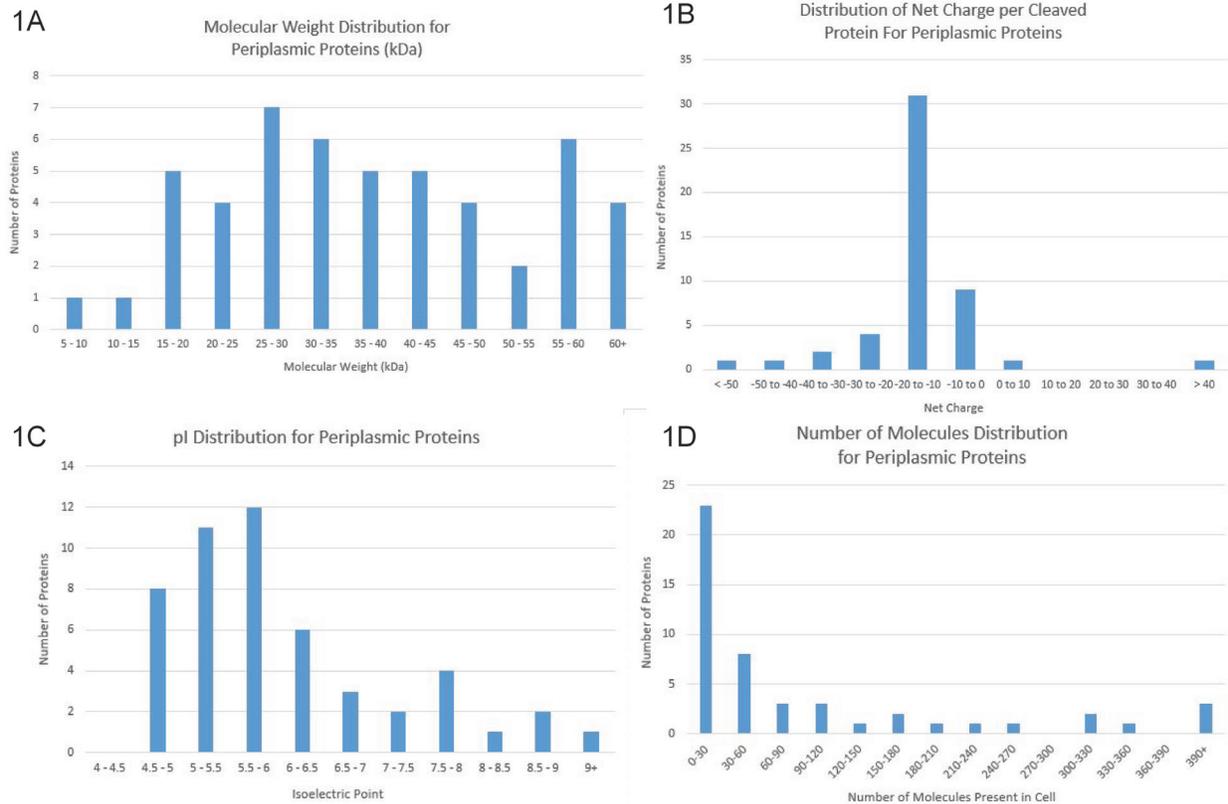
*Escherichia coli* (*E. coli*) is a common model organism used in molecular biology research, so a model of this single-celled Gram-negative bacterium would be immensely useful. An *E. coli* cell has a cytoplasm containing the chromosomal DNA, ribosomes, and many other proteins bounded by a complex cell wall. The latter is composed of inner and outer phospholipid bilayer membranes that together bound an aqueous

space called the periplasm. This space is filled with ions, proteins, and other solutes. It also contains the peptidoglycan layer, a mesh of sugars and amino acids assembled into interlocking strands in a complex network. The periplasm is an important site for protein localization (1) and can constitute up to 40% of the volume of a Gram-negative bacterium (2). The periplasm adopts the pH of the extracellular environment, usually around 7.5 to 7.6 (3), and is a crucial component of an *E. coli* cell. Constructing a model of the periplasm is therefore an important step toward building a model of an entire bacterial cell.

The work described here involves the construction of an atomically accurate preliminary model of the periplasm, the first of its kind, following a similar approach used by McGuffee and Elcock to build a structurally detailed model of the bacterial cytoplasm (4). The aim of this work was to construct a model of the periplasmic space that can eventually be used in conjunction with that of the cytoplasm and with future models of the inner and outer membrane to create a complete, accurate, and comprehensive model of an *E. coli* cell.

Molecular dynamics (MD) and Brownian dynamics (BD) simulations are powerful tools that enable the modeling and visualization of molecular interactions at the atomic level. They take into consideration the important forces that impact atoms, including bond stretching, angle bending, dihedral torsions, electrostatic interactions, and van der Waals forces. They create force vectors to predict trajectories for the atoms using classical Newtonian mechanics that are updated on very small timescales. This allows one to visualize the conformational behavior of biomolecules in aqueous solutions, and thus, provides a means to 'animate' otherwise static structural models. Recent advancements in quantitative proteomics (elucidating and quantifying specific proteins in cells) and structural proteomics (solving structural coordinates of proteins, making available the 3-dimensional structures for use) provide the raw data necessary to set up and perform MD and BD simulations of large-scale biomolecular systems.

We therefore attempted to create a stable preliminary model of the *E. coli* periplasm and subject it to MD simulation. To this end, 13 of the most abundant periplasmic proteins, as determined by a quantitative



**Figure 1: Physicochemical features of periplasmic proteins.** (A) Molecular weight distribution, (B) net charge distribution, (C) pI distribution, and (D) the predicted numbers of proteins in a typical cell.

proteomics study (5), as well as corresponding structural data, were used to create a preliminary model of the periplasm. Throughout an initial MD simulation of 23-nanosecond duration, several intriguing interactions between the proteins in the model and the peptidoglycan layer were observed.

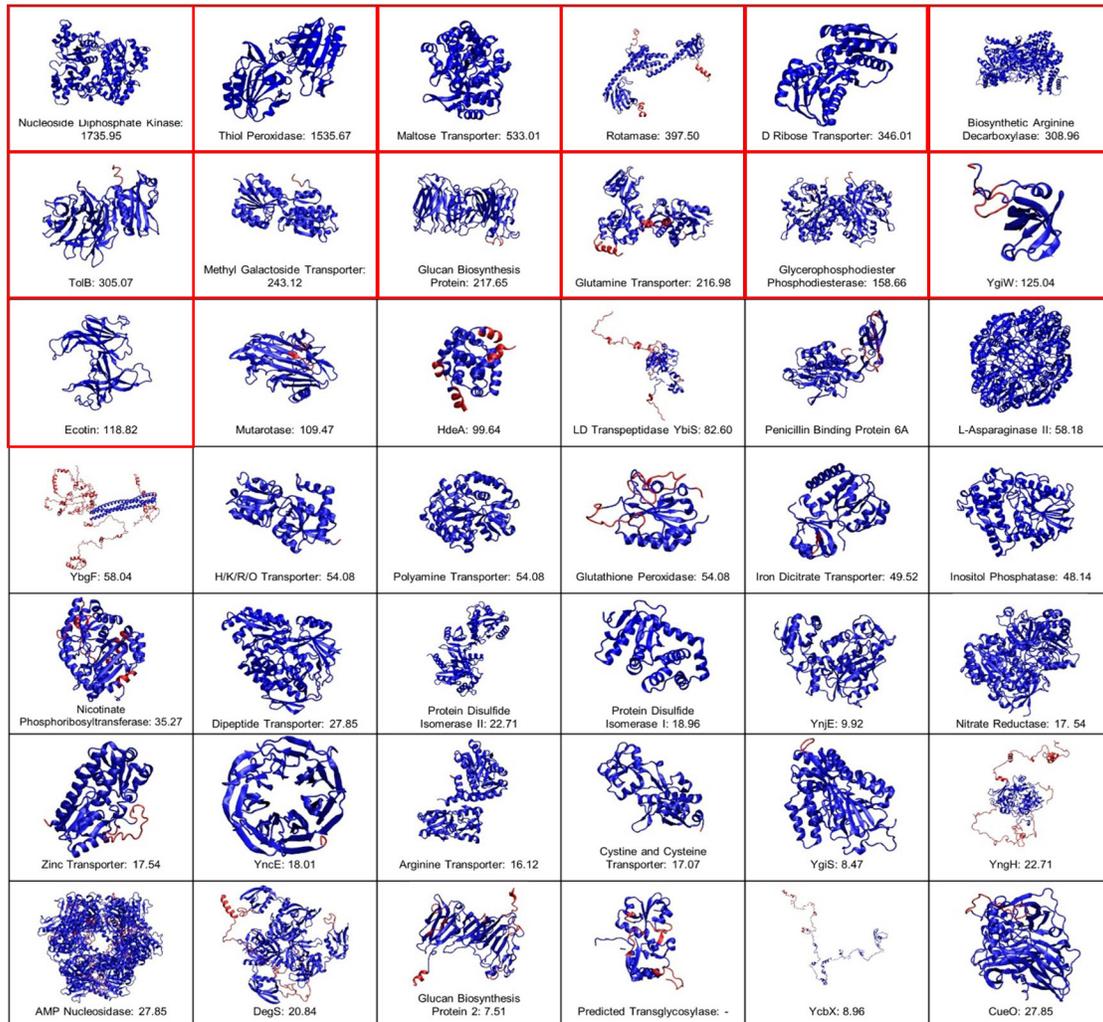
## Results

Figures 1A–D show histograms depicting several properties of the periplasmic proteins researched in this work. Figure 1A shows that the molecular weights of the proteins vary from less than 10 kDa to beyond 60 kDa, with an average of ~30 kDa. The overwhelming majority of these proteins are negatively charged with one exception: membrane-associated serine endoprotease, which has an estimated net charge of  $> +40e$  (Figure 1B). Consistent with this, the isoelectric points for most of the periplasmic proteins are slightly acidic (i.e. below pH 7), which may affect their interactions with each other and with the peptidoglycan layer (Figure 1C). Finally, the distribution of estimated copy numbers for periplasmic proteins indicates that a substantial number of proteins are present in very low copy numbers (i.e.,  $< 30$  per cell), while a few are present in very high copy numbers (Figure 1D). Therefore, a few highly abundant protein

types dominate the protein composition of the periplasm.

The structures of all 42 proteins for which structures were modeled, together with their estimated copy numbers per cell, are shown in Figure 2. It is interesting to note that some of these models, such as those for LD-transpeptidase and transglycosylase, contain regions with little or no secondary structure. These regions were predicted to be intrinsically disordered and have no stable configuration, based on the protein's FASTA sequences entered into IUPred, a tool that predicts intrinsically disordered protein regions (26). Thus, it is impossible for the homology modeling program to create the protein's structure in these regions, since a stable structure does not exist. This explains the largely unstructured regions of these few proteins. Other proteins shown in Figure 2 are large, structured oligomers whose components dictate their function. Several of the transporter proteins, such as those that transport methyl galactoside and the charged amino acids histidine, lysine, arginine, and ornithine, have structures composed of transmembrane alpha helix domains. Therefore, these proteins should not be considered as freely diffusing in the periplasm.

The movements of the 13 proteins selected for inclusion in the periplasm model and their internal flexibilities during the simulation were largely dictated by



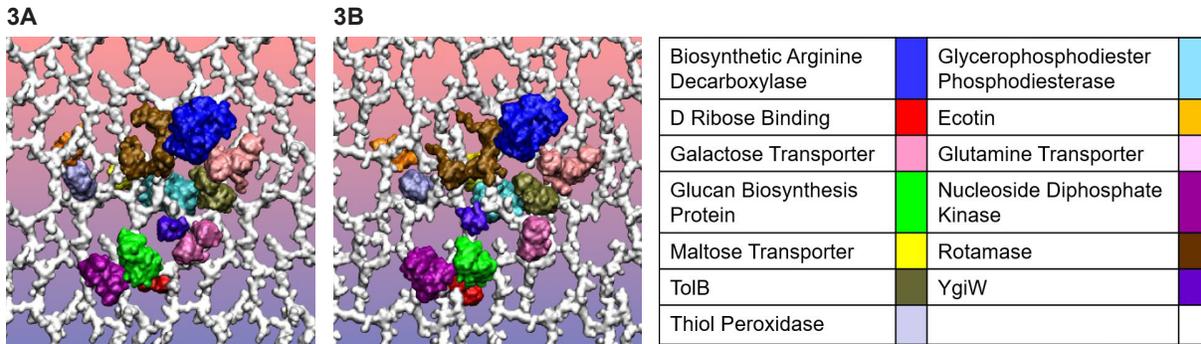
**Figure 2: Gallery of 42 periplasmic proteins whose structures were modeled.** Blue indicates amino acids whose structure was taken from the homology template; red indicates amino acids that were model-built. 13 proteins selected for periplasm modeling are outlined in red.

their structures. Large oligomers that had a well-defined structure, such as L-asparaginase II, did not change very much throughout the duration of the simulation. On the contrary, others, such as D-ribose binding protein and ecotin, had an intrinsically flexible structure that may have contributed to their behavior during the simulation.

**Figures 3A and 3B** compare the beginning configuration of the simulation with that produced at the end of the simulation. Most proteins moved very little, due to the relatively short timescale of the simulation, but some did behave interestingly. TolB (dark sage color) interacted with the peptidoglycan layer and became stuck in the meshwork of amino acids and sugars. Rotamase (brown) also latched on to the peptidoglycan layer and was stuck for a large part of the simulation. Biosynthetic arginine decarboxylase (blue), which has a net charge of -66.6, was repelled from the peptidoglycan model, which had a charge of -276, and was observed to

be at a distance of approximately 84 angstroms, around 1/3 of the width of the periplasmic space (27). In addition, a varied spectrum of pore sizes was observed in the peptidoglycan, some large enough for proteins to pass through. TolB was close to passing through a hole, but got caught on the peptidoglycan.

Root mean square deviation (RMSD) values from the original structure were calculated for five randomly selected proteins (biosynthetic arginine decarboxylase, ecotin, nucleoside diphosphate kinase, D-ribose binding protein, and glucan biosynthesis protein) and were plotted against simulation time, as shown in **Figure 4**. Examination of the graphs indicates structural stability for three of the five proteins as the slope of the best fit line goes to zero as the time increases, indicating that the structure does not change significantly. For the two other proteins, ecotin and D-ribose binding protein, however, greater fluctuations were observed. These two proteins'



**Figure 3: Simulated movement of protein structure.** Pictures of the system at the beginning (A) and end (B) of the simulation. A key is provided that matches proteins with their respective colors.

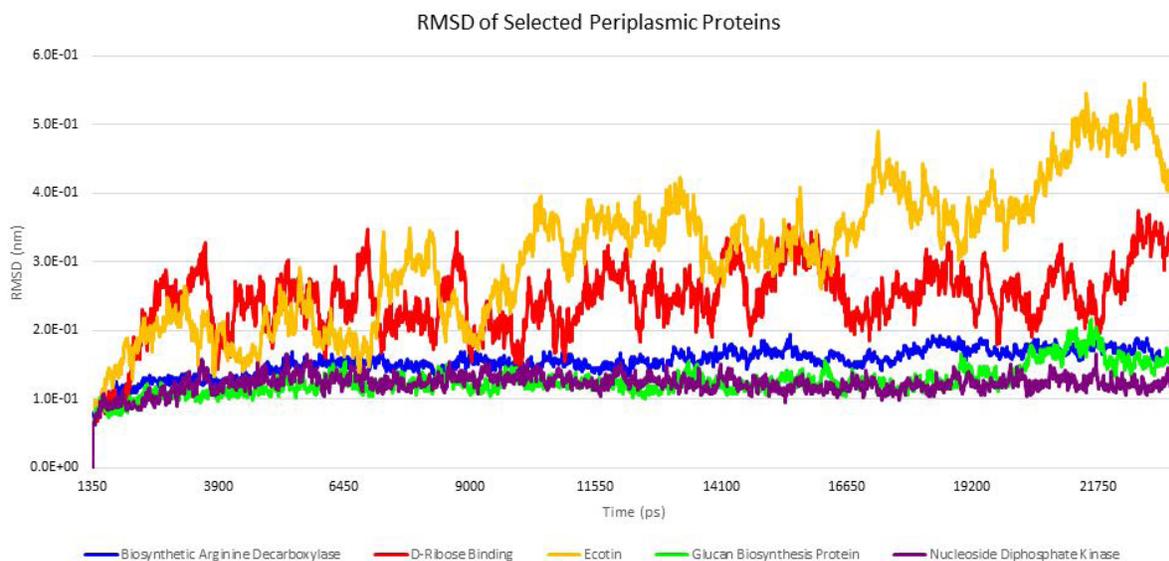
conformations varied significantly, as both graphs have large periods of oscillation between structures similar to the starting point and those that are more different. This greater flexibility can be explained when one looks in more detail at the structures of the proteins.

The ecotin dimer, shown in **Figure 5**, takes on a distinctive Z-shaped structure that is likely to be inherently flexible; the two “arms” of the Z have relatively large freedom to flex along the structural support of the middle bar. The D-ribose binding protein shown in **Figure 6**, on the other hand, has potential flexibility between its two globular-shaped chains as the interface of these two chains is very small. While these features account for the observed fluctuating RMSD plots, it is important to note that the fluctuations are not so great as to indicate fundamentally unstable structures. Obviously, longer simulations need to be conducted to obtain a better picture of protein behavior on a longer timescale, but

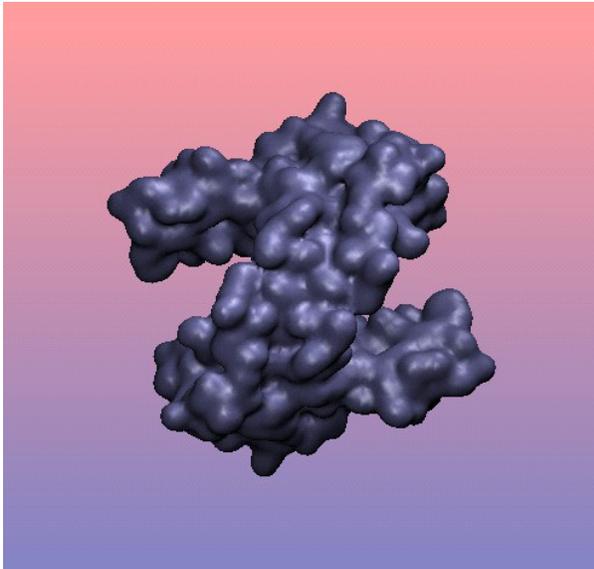
these preliminary simulations show that this area holds promise and interesting prospects for the future.

### Discussion

The construction of atomically accurate models of intracellular environments provides us with the building blocks needed for the creation of a whole-cell model. Here, the first dynamic model of the periplasmic space at the atomic scale has been created, and the interactions of the proteins in this space with the peptidoglycan mesh have been visualized. It is interesting to note that this model was made possible largely by recent advancements in the burgeoning fields of biomolecular and biophysical visualization, as well as those in quantitative proteomics and systems biology. As large-scale experimental studies continue to be conducted and more data is generated, it will be possible to constantly update and improve models, as well as create new ones



**Figure 4: Plots of RMSD of 5 simulated periplasmic proteins.** Ecotin (Orange), D-ribose binding protein (Red), biosynthetic arginine decarboxylase (Blue), glucan biosynthesis protein (Green), nucleoside diphosphate kinase (Purple). The five were randomly selected.

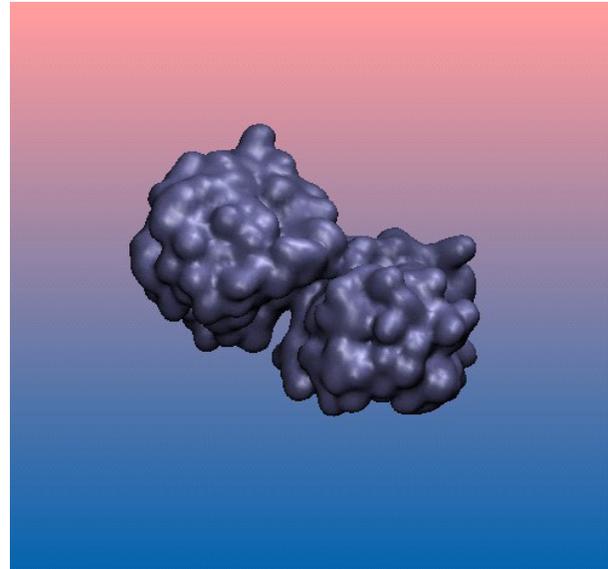


**Figure 5: The Z-shaped structure of ecotin.**

based on new incoming data, to create useful models of biological systems.

We think that the work described here shows that using current molecular dynamics simulations, protein modeling, and structural visualization software, in conjunction with high-throughput quantitative proteomics studies, can lead to the creation of a meaningful model of the *E. coli* periplasmic space. Moving forward, however, it is important to be aware of the potential limitations of the approach adopted here. For example, while the protein structures produced here through the use of homology modelling procedures are likely to be fairly reliable due to the high sequence conservation between the template and the target sequences, it is important to note that for other proteins it is possible that no sufficiently good template upon which to construct a homology model exists. In such cases, homology modelling is likely to yield inaccurate models and so-called *ab initio* prediction algorithms are likely required (28).

It is also important to acknowledge that, while MD simulations can provide invaluable insights into the behavior of biomolecules, they are also limited by a number of factors. For instance, if the force field used does not accurately model all the interactions between atoms, the simulation will, in turn, be inaccurate. In addition, MD simulations are also limited by the available computational power, which directly affects the size and timescale of the simulations. The simulation reported here should, therefore, be considered preliminary, as the system's behavior has only been observed over a relatively short time scale, and even then has only been simulated once. Additional longer simulations – ideally on a microsecond timescale – will be needed



**Figure 6: The dimeric conformation of D-ribose binding protein.**

to obtain a better picture of the nature of interactions between proteins and the peptidoglycan layer. It is also important to note that the present work used a relatively small 6-stranded peptidoglycan sheet, which limited the number of proteins that could be included in the simulations. However, in the future, the peptidoglycan model could be expanded to a larger size and more proteins can be added to obtain a more complete picture.

Despite the limitations of the current model, it is expected that further advancements in the fields of structural biology, quantitative proteomics, and computational modeling will enable increasingly complete and accurate models of intracellular conditions. Such models could prove to be immensely useful in all areas of the biological sciences and would advance and revolutionize research, allowing for the testing of biological hypotheses and experiments done *in silico*, as well as allowing the verification of *in vitro* results. A unicellular model would eventually lead the way to one of a more complex biological system, such as a multicellular organism with specialized cells. This would greatly facilitate work done with subjects, such as humans and live organisms, by offering a cheaper, organism-free method to preliminarily test hypotheses. This would be a powerful quantitative biological visualization tool when used in conjunction with *in vitro* and *in vivo* experiments that could drastically improve the quality of research in the future. Future directions and the next steps for the near future are to perform simulations with the entire periplasmic protein fraction created here and to add the inner and outer membranes, transmembrane proteins, and eventually, the cytoplasm to build a complete, accurate, and comprehensive model of the *E. coli* cell.

## Methods

In this work, the selected proteins came from a comprehensive and unbiased quantitative proteomics study in *E. coli* performed by Masuda *et al.* (5). The supporting information of their publication provides a list of 1,270 proteins that were identified using their novel protease and bile salt protocol to rapidly digest the membrane and extract the cell's proteins, which eliminates a bias for soluble proteins. These proteins were listed with features important to the simulations, such as copy number per cell and intracellular location. Of the 38 proteins that were originally listed by the authors as being periplasmic, 9 were also present in another cellular compartment. There were 601 proteins identified by the authors whose location was listed as a "0", which the authors used to indicate that the location was unknown. Thus, it was necessary that all 1,270 of the proteins were, for the sake of consistency, checked for location. Using this data sheet, all of the 1,270 accession numbers were individually examined in the Biocyc database (6), which links to the UniProt protein sequence database (7). The proteins that were periplasmic were identified and compared to those in the original list. All 38 proteins identified by Masuda *et al.* were featured in the new list. However, there were 15 new proteins that existed in the periplasm and another cellular compartment, while 11 novel proteins were found exclusively in the periplasm. As only one copy number is given for each protein in the Masuda *et al.* dataset, it is not clear how many copies of each protein are located in each cellular compartment for those that are found in multiple compartments. Since it is not known, therefore, how many copies to add to the periplasm in this simulation, these proteins were omitted from the initial model described here.

The names and accession codes of the periplasmic proteins were then recorded in an Excel spreadsheet, and the amino acid (FASTA) sequences for each protein were obtained from the Uniprot database. The ExPASy pI/MW tool was used to compute the pI and the molecular weight of the proteins using the FASTA sequences as input (8, 9, 10), and PROTEIN CALCULATOR v3.4, which uses pKa values from Stryer's Biochemistry was used to compute electric charges at pH 7.0 (11). The molecular masses were then used to determine the percentages of the entire mass of the cell that each protein comprised. Finally, the number of molecules present per protein was calculated by dividing the copy number reported by Masuda *et al.* by the number of polypeptide chains in each protein. After computing the percentage of molecules comprised by each protein, the INT function was used to calculate how many copies of each protein should be used in the simulation given a particular total number of molecules.

## Homology Modeling

To create the structures of the proteins, homology modeling was used. A BLAST search was carried out on the SEQATOMS webserver using the FASTA sequences as input (RCSB; 12). If the reported e-values were less than  $1 \times 10^{-4}$ , then homology could be inferred (13), and the four-character alphanumerical codes of the corresponding structure were used to obtain the oligomeric states and ligands from the RCSB entry. However, if the e-values were not less than the above cutoff, no homology could be inferred, indicating that no good templates existed upon which to build these proteins. Examples for which this was the case include YcbX and YngH, which are thought to contain intrinsically disordered regions. Although it is possible that threading and *ab initio* modeling algorithms could have been used to create putative structures for these proteins, it would be very difficult to predict the tertiary and quaternary structures with high accuracy; to avoid the possibility that poorly modeled structures could lead to unrealistic behavior, these proteins were not included.

When there were sufficiently good templates, the PDB files were downloaded and edited manually to remove heteroatoms (non-protein atoms) and change formatting. Using SwissPDBViewer, the sequence alignment obtained from the SEQATOMS search was used to align the sequences of the protein of interest with that of its homolog for which a structure was available (14). This initial, aligned structure was formatted using the in-house Perl CLEAN.SWISS program, missing loops were created using LOOPY (15), and missing N and C termini were created using an in house program. After this, SwissPDBViewer was used to fill in any missing sidechains, and the resulting PDB structure file was verified to have been completed using VMD Beta view, which rendered constant regions blue and those that were changed red. The final structure was then considered ready for use in a subsequent MD simulation. This procedure was successful for most modeled proteins, but failed for three: YdeN, beta-D-glucoside glucohydrolase, and a predicted mechanosensitive channel, even though they were found to have existing templates in the protein database. It is unclear as to why these proteins could not be modeled successfully, but it is likely that internal steric clashes in the models prevented the CLEAN.SWISS program from completing the addition of missing loops.

## Protein Concentrations in the Periplasm

To determine how many protein molecules in total were to be used in the simulation based on a given volume, the combined concentration of proteins in the periplasm needed to be calculated. The volume of an *E. coli* cell is estimated to be  $1.1 \mu\text{m}^3$  (16), and the periplasm

makes up, on average, ~16% of the cell's total volume (17, 18). Therefore, the total periplasmic volume is  $0.16 \times 1.1 \times 10^{-18} = 1.76 \times 10^{-19} \text{ m}^3$ . There are estimated to be  $2.35 \times 10^6$  proteins in an entire *E. coli* cell (17), of which 4% are thought to be periplasmic (19). Thus there are approximately  $2.35 \times 10^6 \times 0.04 = 94,000$  periplasmic proteins. This means that the number of moles of protein in the periplasm is  $94,000 / 6.022 \times 10^{23} = 1.56 \times 10^{-19} \text{ mol}$ . Therefore, the concentration of proteins in the periplasm is  $1.56 \times 10^{-19} \text{ mol} / 1.76 \times 10^{-19} \text{ m}^3 = 0.89 \text{ mM}$ . Given that the periplasmic volume ranges from 10% to 40% of the total cell volume, we expect the combined protein concentration to range from 2.22 to 0.55 mM assuming that the protein copy numbers remain approximately constant.

### Molecular Dynamics Simulations

All MD simulations were performed using the freely available GROMACS software (20). The `pdb2gmx` command was used to create topology files from each protein's structure (pdb) file that were to be used in the GROMACS simulation. The AMBER99SBnmr1-ILDN force field (21) was used to describe protein atoms, and the TIP4P-Ew (22) model was used for water molecules. A 6-stranded model of the peptidoglycan layer constructed by Chengxuan Guo, a member of the group, was also included in the simulation; parameters for the sugar atoms of the peptidoglycan were taken from the GLYCAM force field (23). To reduce computational expense, we limited the size of our initial simulation system to the following dimensions:  $246 \times 243 \times 180 \text{ \AA}$ , with the peptidoglycan model oriented in the x-y plane. The volume of this system is  $1.08 \times 10^{-23} \text{ m}^3$ . For such a system size, we expect that the total number of periplasmic proteins would range from 3.6 to 14.4. A whole number on the higher side, 13, was chosen from this range, as choosing a larger number of proteins to be placed in the simulation would facilitate the interactions between them and the peptidoglycan layer, allowing for easier visualization of this behavior.

The proteins chosen for the simulation were the 13 most abundant proteins in the periplasmic space, as determined by their estimated experimental copy number (5). These proteins are listed in **Table 1**, along with their copy numbers per cell and other important details including their oligomeric states and PDB and Uniprot IDs. Each of these proteins was added to the system using an in-house script that places molecules randomly while avoiding steric clashes with atoms that have already been placed. The system was then solvated in a box of water with periodic boundary conditions.  $\text{Na}^+$  and  $\text{Cl}^-$  ions were added accordingly at a concentration of 150 mM, and 9 additional sodium ions were then added to neutralize the charge of the system. The temperature

was then incrementally raised from 50 Kelvin, to room temperature, 298 K, and Newtonian dynamics was computed for all atoms in the system using GROMACS version 4.5.1 (20). The simulation was performed on 6 compute-nodes and 288 cores of a PC computer cluster, 96 of which were used to perform calculations of long-range electrostatic interactions using the Particle Mesh Ewald (PME) method (24). This setup was determined to provide the best performance by running a series of experimental 5 picoseconds (ps) simulations. The entire system was simulated for 23.368 ns, over the course of 10 days, with a timestep of 2.5 ps and then stopped for viewing and analysis. An `xtc` file without water molecules displaying 1 frame every 10 picoseconds was generated, and the resulting movie was then viewed with VMD version 1.9.1 (25). As periodic boundary conditions took effect, separating proteins that crossed the edge of the simulation cell, an in-house program was used to create whole versions of the proteins for the purpose of facilitating visualization. The simulation was then viewed, and the resulting interactions were observed.

### Acknowledgements

P.L. expresses his sincere gratitude to Gabriel Velez, Wesley Lay, Casey Andrews, and Shuxiang Li for their guidance throughout this project and to Chengxuan Guo, who provided the peptidoglycan model. Special thanks also to Dr. Kate Degner for her helpful advice and resources throughout this project.

### References

- Schlegel, S., Rujas, E., Ytteberg, A. J., Zubarev, R. A., Luirink, J., & de Gier J. W. "Optimizing heterologous protein production in the periplasm of *E. coli* by regulating gene expression levels." *Microbial Cell Factories* 12 (2013): 24.
- Graham, L. L., Beveridge, T. J., & Nanninga, N. "Periplasmic space and the concept of the periplasm." *Trends Biochem Sci.* 16 (1991): 328-9.
- Wilks, J. C., & Slonczewski, J. H. "pH of the Cytoplasm and Periplasm of *Escherichia coli*: Rapid Measurement by Green Fluorescent Protein Fluorimetry." *J. Bacteriol.* 189 (2007) 5601-5607.
- McGuffee S. R., & Elcock A. H. "Diffusion, crowding & protein stability in a dynamic molecular model of the bacterial cytoplasm." *PLoS Comput. Biol.* 6 (2010): e1000694.
- Masuda, T., Saito, N., Tomita, M., & Ishihama, M. "Unbiased Quantitation of *Escherichia coli* Membrane Proteome Using Phase Transfer Surfactants." *J. Mol. Cell. Proteom.* 8 (2009): 2770-2777.
- Caspi, R., Altman, T., Dale, J. M., Dreher, K., Fulcher, C. A., Gilham, F., Kaipa, P., Karthikeyan, A. S., Kothari, A., Krummenacker, M., Latendresse, M., Muel-

- ler, L. A., Paley, S., Popescu, L., Pujar, A., Shearer, A.G., Zhang, P., & Karp, P. D. "The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases." *Nucleic Acids Res.* 38 (2010): D473-9.
7. The UniProt Consortium. "Activities at the Universal Protein Resource (UniProt)." *Nucleic Acids Res.* 42 (2014): D191-D198.
  8. Bjellqvist, B., Hughes, G. J., Pasquali, Ch., Paquet, N., Ravier, F., Sanchez, J.-Ch., Frutiger, S. & Hochstrasser, D. F. "The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences." *Electrophoresis* 14 (1993): 1023-1031.
  9. Bjellqvist, B., Basse, B., Olsen, E. & Celis, J. E. "Reference points for comparisons of two-dimensional maps of proteins from different human cell types defined in a pH scale where isoelectric points correlate with polypeptide compositions." *Electrophoresis* 15 (1994): 529-539.
  10. Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M. R., Appel R. D., & Bairoch A. "Protein Identification and Analysis Tools on the ExPASy Server." (in) John M. Walker (ed). *The Proteomics Protocols Handbook*. Humana Press (2005).
  11. Stryer, Lubert. *Biochemistry*. 3<sup>rd</sup> ed. W.H. Freeman & Company, 3rd Edition 1988. Print.
  12. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., & Bourne, P. E. "The Protein Data Bank." *Nucleic Acids Res.* 28 (2010): 235-242.
  13. Pevsner, Jonathan. *Bioinformatics and Functional Genomics*. John Wiley and Sons, Inc. 2009, Print.
  14. Brandt, B. W., Heringa, J., & Leunissen, J. A. M. "SEQATOMS: a web tool for identifying missing regions in PDB in sequence context." *Nucleic Acids Res.* 36 (2008): W255-W259.
  15. Xiang Z., Soto C. S., Honig B. "Evaluating conformational free energies: The colony energy and its application to the problem of loop prediction." *Proc. Natl. Acad. Sci. USA* 99 (2002): 7432-7437
  16. Kubitschek, H. E., & Friske, J. A. "Determination of bacterial cell volume with the Coulter Counter." *J. Bacteriol.* 168 (1986): 1466-7.
  17. Neidhardt F. C. *Escherichia coli and Salmonella: Cellular and Molecular Biology*. Vol 1. ASM Press. 1996.
  18. Pilizota T, & Shaevitz, J. W. "Fast, multiphase volume adaptation to hyperosmotic shock by *Escherichia coli*." *PLoS One.* 7 (2012): e35205.
  19. Goodsell D. S. "Inside a living cell." *Trends Biochem. Sci.* 16 (1991): 203-6.
  20. Hess, B., Kutzner, C., van der Spoel, D., & Lindahl, E. "GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation." *J. Chem. Theory Comput.* 4 (2008): 435-447.
  21. Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., *et al.* "Improved side-chain torsion potentials for the Amber ff99SB protein force field." *Proteins* 78 (2010): 1950-8.
  22. Horn, H. W., Swope, W. C., Pitner, J. W., Madura, J. D., Dick, T. J., Hura, G. L., & Head-Gordon, T. "Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew." *J. Chem. Phys.* 120 (2004): 9665-9678.
  23. Kirschner, K., Yongye, A. B., Tschampel, S. M., Gonzalez-Outeirino, J., Daniels, C. R., Lachele-Foley, B., & Woods, R. J. "GLYCAM06: A generalizable biomolecular force field. Carbohydrates." *J. Comput. Chem.* 29 (2008): 622-655.
  24. Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., & Pedersen, L. G. "A smooth particle mesh Ewald method." *J. Chem. Phys.* 18 (1997): 1463-1472.
  25. Humphrey, W., Dalke, A. & Schulten, K. "VMD - Visual Molecular Dynamics" *J. Mol. Graph.* 14 (1996): 33-38.
  26. Dosztányi, Z., Csizmók, V., Tompa, P., & Simon, I. "IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content." *Bioinformatics* 21 (2005): 3433-3434.
  27. Vollmer, W., & Seligman, S. J. "Architecture of peptidoglycan: more data and more models." *Trends Microbiol.* 18 (2010): 59-66.
  28. Baker, D., & Sali, A. "Protein Structure Prediction and Structural Genomics." *Science.* 5540 (2001): 93-96.

Protein	Uniprot	PDB ID Used	Copy	Alignment State	Ligands	Confidence of Alignment	E value	Chains	Molecular Weight per Cleaved Chain	Total Weight	Mass Percentage of 13 Proteins	Cumulative Percentage	pI	Net Charge per Cleaved Chain	Net Charge per Molecule	Net Charge per Cleaved Molecule	Number of Molecules	Molecule Percentage of 13 Proteins	Number of Molecules Not Rounded	Number of Molecules/Total	Disulfide Bridges
multifunctional nucleoside diphosphate kinase -1, cytosolic, endonuclease -1, 5-phosphoribosyltransferase	P12202	2HUR	175.95	4mer, 2HUR	Magnesium, Nicotinamide	334	2.00E-77	4.00	1233.25	2655023.39	15.94	15.94	5.35	-1.1	-12.4	-11.4	423.89	11.15	2,155	3.00	100 and 534 on each other
ribonucleoside diphosphate kinase II, cytosolic	P12217	3A83	1578.97	2mer, 3A83	Adenosine, dATP, dTTP, dCTP, dGTP	334	2.00E-89	2.00	17704.12	27217985.86	16.28	32.22	4.75	-4.9	-9.8	-9.8	787.84	21.49	3,158	5.00	
ribonucleoside diphosphate kinase II, cytosolic	P12204	3B74	333.01	Monomer	Adenosine	759	0	1.00	40707.31	2697408.65	12.69	44.91	5.21	-4.3	-7.3	-7.3	333.01	24.62	3,561	4.00	
ribonucleoside diphosphate kinase II, cytosolic	P12205	1D6U	397.50	2mer, 1D6U	Adenosine, dATP, dTTP, dCTP, dGTP	479	1.00E-415	1.00	2823.85	1043392.83	6.24	51.16	6.75	2.2	6.0	6.0	198.73	5.36	1,336	1.00	
ribonucleoside diphosphate kinase II, cytosolic	P12206	1O89	348.01	Monomer	Ribose	511	1.00E-548	1.00	28974.47	8803451.36	5.90	57.06	5.89	0.7	-1.3	-1.3	246.01	6.98	2,322	2.00	
ribonucleoside diphosphate kinase II, cytosolic	P12207	3A2Q	104.84	2mer, 3A2Q	dATP, dTTP, dCTP, dGTP	1311	0	1.00	71999.45	2181465.11	11.67	62.73	4.82	-0.11	-0.13	-0.13	154.48	4.32	1,038	1.00	
ribonucleoside diphosphate kinase II, cytosolic	P12208	3A2X	350.07	Monomer	dATP, dTTP, dCTP, dGTP	884	0	1.00	43951.64	13551552.31	7.97	70.70	6.24	1.2	-1.1	-1.1	350.07	8.54	2,050	2.00	
ribonucleoside diphosphate kinase II, cytosolic	P12213	2W9H	243.12	Monomer	dATP, dTTP, dCTP, dGTP	612	1.00E-575	1.00	33397.7	8111335.22	4.86	83.56	5.25	-3.1	-3.1	-3.1	243.12	6.81	1,631	2.00	
ribonucleoside diphosphate kinase II, cytosolic	P12225	1TXX	217.83	Monomer	dATP, dTTP, dCTP, dGTP	493	0	1.00	35855.18	12551482.03	7.22	75.78	6.28	0.7	-1.3	-1.3	217.83	6.15	1,648	1.00	
ribonucleoside diphosphate kinase II, cytosolic	P12235	1A66	118.88	2mer	dATP, dTTP, dCTP, dGTP	450	1.00E-427	1.00	34883.41	5448860.70	3.24	64.03	6.87	2.4	6.3	6.3	108.49	3.04	670	1.00	
ribonucleoside diphosphate kinase II, cytosolic	P12252	1N0X	123.04	2mer, 1Y07	dATP, dTTP, dCTP, dGTP	731	0	1.00	38200	6069313.00	3.85	67.88	5.21	-6.6	-16.2	-16.2	79.33	2.22	533	1.00	
ribonucleoside diphosphate kinase II, cytosolic	P12257	1E7Y	118.81	2mer, 1E7Y	dATP, dTTP, dCTP, dGTP	290	3.00E-579	2.00	15976.15	1691195.30	0.90	88.85	4.75	-2.8	-4.8	-4.8	123.04	3.50	834	1.00	
ribonucleoside diphosphate kinase II, cytosolic	P12257	1E7Y	118.81	2mer, 1E7Y	dATP, dTTP, dCTP, dGTP	290	3.00E-579	2.00	16099.52	4912844.87	1.15	200.00	5.84	0.1	-0.9	-1.8	99.41	1.66	639	0.00	

Table 1: Detailed information on each of the 13 simulated proteins.