

EcoProDB: the protein knowledge base for *Escherichia coli*

Hongseok Yun^{1,2}, Jeong Wook Lee¹, Joonwoo Jeong², Jaesung Chung³, Jong Myoung Park¹,
Han Na Myoung¹, Sang Yup Lee^{1,2,3,*}

1. Department of Chemical and Biomolecular Engineering, KAIST, Daejeon, Korea

2. Bioinformatics Research Center, KAIST, Daejeon, Korea

3. Department of BioSystems, KAIST, Daejeon, Korea

*email: leesy@kaist.ac.kr

Recent advances in x-omics technologies have facilitated understanding of global behaviors of microorganisms. Among them, proteomics has become an essential tool to provide information on physiological functions, translocation, and subcellular localization of the whole set of proteins in the cell. Significant impact it brings out essentially have led to the increasing volume of proteome data; protein interactions, expression profiles, 2-D gel and non-gel based data. Also, various databases for protein information have been constructed such as UniProt (<http://www.expasy.org>) and NCBI (<http://www.ncbi.nlm.nih.gov>), and used in fundamental and applied research. Meanwhile, the most extensively studied organism, *Escherichia coli*, is receiving renewed interest as an excellent model for various systems-level research approaches including proteomics. Consequently, the number and quality of resolved *E. coli* proteins are exponentially increasing. In the previous studies, the most frequently undertaken approach on *E. coli* proteome is the gel-based method, which is currently the most widely used as a core of proteomic technology.

To date, we have collected *E. coli* proteins which have been identified on 2-D gels from *E. coli* SWISS 2DPAGE database (<http://www.expasy.org/ch2d/>) and the literatures. As a result, we constructed a protein database for *E. coli* called EcoProDB. The objective of constructing EcoProDB is to integrate widely dispersed information of proteins that are identified on 2-D gels along with other resources and to provide comprehensive platform for the expression levels of many heterogeneous proteins under different genetic and environmental conditions, using the interactive interface and search mechanism. EcoProDB services subcellular localization, theoretical and experimental 2-D maps, and integrated protein information. In addition, the on-line tools are able to upload user's data for the comparative analysis and database expansion. Thus, using EcoProDB, users can efficiently grasp the core information associated with the proteins and 2-D gel results obtained from several different experimental sets for more convenient and enhanced analysis. EcoProDB is publicly available at <http://eecoli.kaist.ac.kr>. [This work was supported by the Korea Science and Engineering Foundation (KOSEF) grant funded by the Korea government (MOST) (No. M10309020000-03B5002-00000). Further supports by LG Chem Chair Professorship, Microsoft and IBM SUR program are appreciated.]

Figure 1: the screenshots of EcoProDB.

The screenshot displays the EcoProDB website interface, which is part of the KAIST Institute for the BioCentury. The page features a navigation menu with tabs for Protein, Reference, Localization, Theoretical Map, 2D Page Maps, Contact, Submission, and Online Tool. A search bar is located at the top, with a 'Go' button and a search scope selector (Name, Accession Num, Gene name, PI, Mw, Function/Expression).

The main content area is divided into several sections:

- Entry Information:** Displays details for protein P08443 (Substrin domainase [Mn]). It includes the protein description, pI/Mw (pI: 4.42, Mw: 50), protein function (involved in cell cycle regulation), and a list of references.
- 2D gel information:** Shows the protein's position on a 2D gel with pI (4.5-7.0) on the x-axis and Mw (10-70 kD) on the y-axis. The protein is highlighted in red.
- 2D gel images:** Three 2D gel images are shown, labeled (A), (B), and (C), illustrating the protein's position and changes in expression under different conditions.
- Protein list:** A list of proteins is displayed, including AcpB, AcpF, AcpM, and AcpN, with their theoretical Mw and pI values.

The interface is designed for easy navigation and data visualization, providing comprehensive information about the proteins in the database.