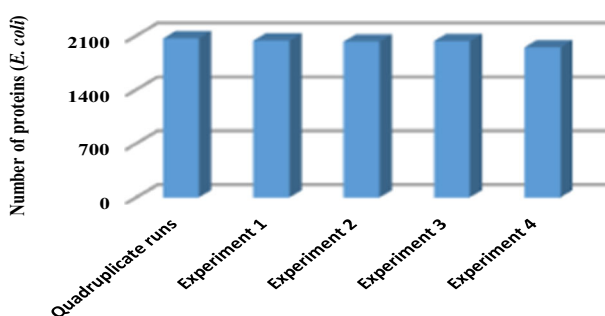


RESEARCH ARTICLE

Single-Run Mass Spectrometry Analysis Provides Deep Insight into *E. coli* Proteome

Bhaswati Chatterjee,¹ Suman S. Thakur² ¹National Institute of Pharmaceutical Education and Research (NIPER), NIPER-Hyderabad, (Dept. of Pharmaceuticals, Ministry of Chemicals and Fertilizers, Govt. of India), Balanagar, Hyderabad, Telangana 500 037, India²Centre for Cellular and Molecular Biology, Proteomics and Cell Signaling, Lab E409, Uppal Road, Hyderabad, 500007, India

Abstract. Single-run mass spectrometry has enabled the detection and quantifications of *E. coli* proteins. A total of 2068 proteins quantified by intensity-based absolute quantification (iBAQ) Schwanhäusser et al.: (Nature. **473**, 337–342, 2011) procedure were obtained with single enzyme-trypsin, without pre-fractionation, by quadruplicate long liquid chromatography runs coupled with high-resolution linear trap quadrupole (LTQ)-Orbitrap Velos mass spectrometry.

The single-run of 12 h has ability to cover almost 98% of the quadruplicate LC-MS/MS runs of *E. coli* proteome and is therefore almost equivalent to quadruplicate LC-MS/MS runs. These quantified proteins are about 52% of the total proteins present in *E. coli* genome according to Uniprot database. The quantified proteins covered almost all of the proteins in folate biosynthesis. Remarkably greater part of Gene Ontology (GO) Barrell et al.: (Nucleic Acids Res. **37**, D396–D403, 2009), Ashburner et al.: (Nat. Genet. **25**, 25–29, 2000) annotations, signaling pathways along with protein-protein interactions were covered. Some of the important biological processes-cell cycle, DNA repair, ion transport, ubiquinone biosynthetic process, pseudouridine synthesis, peptidoglycan biosynthetic process, RNA processing, and translation-revealed protein-protein interaction network generated by Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) Jensen, et al.: (Nucleic Acids Res 37, D412–D126, 2009) database. Therefore, to achieve the saturation point of detection of maximum number of proteins in single LC-MS/MS run, 12-h liquid chromatography gradient is appropriate.

Keywords: Single-run, Mass spectrometry, *E.coli*, Proteome, iBAQ

Received: 16 March 2018/Revised: 2 September 2018/Accepted: 6 September 2018/Published Online: 26 September 2018

Introduction

Single-run proteomic analysis using liquid chromatography coupled with high resolution mass spectrometry has evolved as an important technology for studying large number of proteins in an unbiased manner using less amount of starting material, no requirement for pre-fractionation or biochemical

fractionation techniques and involving less measuring time in mass spectrometry. It has been used to study deep proteome of yeast and mammalian cell line [1]. Mass spectrometry-based proteomics technology have been invariably used to study protein-protein interactions [2–4], cell signaling [3–6], subcellular organelles [2, 3, 5], and complexes involving proteins [2, 3, 5].

The liquid chromatography has a great power for separating analytes [7], and it works in a wonderful way when combined with high-resolution mass spectrometry leading to their identification. Different types of columns of various lengths 66-cm [8], 80-cm [9], 350-cm [10], 70-cm × 20- μ m inner diameter silica-based monolithic columns [11], 60-cm long triphasic capillary column [12], and bead sizes 1.5- μ m nonporous

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s13361-018-2066-z>) contains supplementary material, which is available to authorized users.

Correspondence to: Bhaswati Chatterjee; e-mail: bhaswati@niperhyd.ac.in, Suman Thakur; e-mail: sst@ccmb.res.in

octadecylsilane-modified silica particles [8], 1.7 μm C_{18} particles [13], 3- μm C_{18} -bonded particles [9, 13], and 100- μm inner diameter monolithic silica- C_{18} [10] were used to improve chromatography separation and resolution. For deep proteome analysis of complex mixtures, generally protein or peptide fractionation is involved before liquid chromatography-tandem mass spectrometry runs. However, this involves large amount of starting materials and measurement time. The in-depth coverage of the *E. coli* proteome was tried by several groups including Wiśniewski and Rakus [14], Macek and coworkers [15] and Heinemann and coworkers [16]. Moreover, the use of expensive filtration devices for sample preparation technique, Filter-Aided Sample Preparation (FASP) [14], different protein or peptide fractionation techniques requiring more starting materials, more measuring time, and also higher version of linear trap quadrupole (LTQ)-Orbitrap Velos mass spectrometry accompanied by various quantification techniques such as label-free quantification [16], selected reaction monitoring (SRM) [16], stable isotope dilution (SID) [16], Super-Stable Isotope Labeling by/with Amino acids in Cell culture (SILAC) approach [15], and total protein approach [14] were reported by them [14–16].

Earlier, we had used liquid chromatography runs of 8-h duration, using in-house made column oven, column length of 50 cm packed with 1.8- μm C_{18} beads, coupled with high-resolution LTQ-Orbitrap Velos mass spectrometry to identify 2990 yeast proteins and 5376 mammalian cell line proteins in triplicate analysis without any pre-fractionation steps [1].

Here, we have increased the duration of liquid chromatography runs to 12 h instead of 8 h [1]. Now, by using LTQ-Orbitrap Velos mass spectrometry coupled with our earlier described method for preparation of column [1], column oven [1], along with intensity-based absolute quantification (iBAQ) [17] approach, we set to investigate the in-depth proteome of *E. coli* using single enzyme-trypsin and without any pre-fractionation techniques. We will demonstrate the coverage of *E. coli* cellular pathways, biological process, signaling pathways, and protein-protein interaction network in an unbiased manner. Further, we will investigate how the single-run of 12 h has almost the same coverage ability as the quadruplicate runs. Here, in our study, we have employed iBAQ [17] approach to study dynamic range of the proteins identified in *E. coli*. The iBAQ algorithm [17] is used to determine each protein's approximate abundance by normalizing the added intensities of the peptides by the theoretical number of the peptides observed for the proteins.

Material and Methods

Preparation of E. coli Cell Lysates

The *Escherichia coli* (strain K12) was grown in 10 ml LB media at 37 °C to an $\text{OD}_{600} = 0.7$. Cells were harvested by centrifugation and re-suspended in 500 μL 8 M Urea, 100 mM Tris-HCl pH 7.4. Five hundred microliters of Zirconia beads were added and cells were lysed in a FastPrep-24 (MP

Biomedical) three times for 45 s. Lysates were cleared by centrifugation at $17000\times g$ for 5 min and protein concentration was measured. Fifty micrograms of total proteins were reduced with 1 mM DTT and alkylated with iodacetamide.

In-Solution Digestion

E. coli proteins were digested overnight with Trypsin at a concentration of 1:50 (*w/w*) at room temperature. After digestion sample were acidified with TFA and loaded on Stage Tips (5 μg /Stage Tip) as described previously [18]. From Stage Tips, the peptides were eluted using buffer B (80% acetonitrile, 0.5% acetic acid).

LC-MS Analysis

The peptides were loaded on in-house made 50-cm column length (75- μm inner diameter, 1.8- μm C_{18} beads (Reprosil-AQ Pur, Dr.Maisch) using nano-HPLC with a flow rate of 75 nl/min. This was coupled with linear trap quadrupole (LTQ)-Orbitrap Velos mass spectrometer (Thermo Fisher Scientific). Peptides were loaded onto the column with buffer A (0.5% acetic acid) and eluted with a 12-h linear gradient. Data was recorded in data-dependent mode. The acquirement of survey scans was carried out in the Orbitrap mass analyzer using 25 most intense peaks as described earlier [1]. The fragmentation of the peptides was done using HCD method [19].

Data Analysis

The acquired raw data was processed using MaxQuant [15] version 1.2.2.5 and database used was (ncbi.Ecoli_K12_substr-DH10B.25-Jan-2010.fasta) with Andromeda search engine [20]. The search included cysteine carbamidomethylation as a fixed modification, and N-acetylation of protein and oxidation of methionine as variable modifications. Up to two missed cleavages were allowed for protease digestion. Only peptides with minimum six amino acid length were considered for identification. The false discovery rate (FDR) for the identification of peptides and proteins was 1%. The iBAQ algorithm was used for quantification of proteins [17]. The identified proteins were filtered to remove the reverse hits and contaminants. The annotations were provided with Gene Ontology (GO) biological process, molecular function, cellular component [21, 22], and Kyoto Encyclopedia of Genes and Genome (KEGG) [23] pathway. We used STRING database version 9.1 (<http://string-db.org/>) for protein-protein interactions. We considered only those protein-protein interactions, for our dataset, whose confidence score was ≥ 0.7 (high confidence). We also visualized these interactions using Cytoscape [24].

Results and Discussion

iBAQ Quantification of E.coli Proteome

We chose to investigate *E. coli* K-12 because it is genetically best understood prokaryotic organism and biologically safe vehicle to conduct biological experiments [25]. To investigate

the number of proteins identifiable in single-run analysis in *E. coli* without pre-fractionation, we used long column of 50 cm with 1.8- μ m C₁₈ particle sizes with 12-h gradient in LTQ-Orbitrap-Velos that has high scanning speed. We performed quadruplicate analysis of *E. coli* peptides in order to have uniformity in the analysis and detected 2068 quantified proteins with about 2 days measurement time (Supplementary Table S1). Here, we refer quadruplicate runs as 4 single-runs \times 12 h, while single-run is 1 single-run \times 12 h.

Each single-run of 12 h identifies and quantifies 2038, 2025, 2031, and 1950 proteins respectively. The Pearson's correlation coefficient of iBAQ protein quantification between quadruplicate runs and each single-runs as Experiment 1 containing 2038 proteins is 0.984, Experiment 2 containing 2025 proteins is 0.985, Experiment 3 containing 2031 proteins is 0.985 and Experiment 4 with 1950 proteins is 0.956 (Supplementary Figure S1). Notably, it covered 98.5%, 97.9%, 98.2%, and 94.3% respectively of 2068 quantified proteins in quadruplicate run analysis. It is clear; each single-run is almost equivalent to quadruplicate runs (Table 1). Interestingly, it has identified and quantified about 52% of the total proteins present in *E. coli* genome according to Uniprot database. Notably, during 8-h single LC-MS/MS run, we were able to achieve 87% coverage of triplicate LC-MS/MS runs of HEK 293 cell line proteome [1], while in this 12-h single-LC-MS/MS run, we are able to achieve 98% coverage of quadruplicate LC-MS/MS runs of *E. coli* proteome [Table 1]. This experiment suggests each single-run of 12 h is sufficient to identify enough proteins of *E. coli*, so as to decipher the signaling and proteome information. The confidence of single LC-MS/MS run data was increased by running replicates such as duplicate, triplicate, and quadruplicate runs and identifying the same proteins again. In order to improve the dynamic range of the single-run analysis, "match between runs" was performed in MaxQuant environment [26] by matching peptide intensities of each single-runs of quadruplicate analysis with each other. This shows each single-run of 12 h has achieved enough saturation point and dynamic range for the detection and identification of maximum number of proteins that are achieved by quadruplicate run analysis.

To roughly calculate the abundance of each protein detected in a proteome, we used intensity-based absolute quantification (iBAQ) algorithm where the intensity of the peptide identifying the protein is summed-up followed by the normalization of these total peptide intensities by the theoretical number of peptides observed for the

proteins. These protein intensities normalized are converted to copy number of the proteins based on the total amount of the proteins in the sample analyzed [17]. The values of iBAQ tell about the absolute amount of each protein [27]. We are not claiming to get absolute quantification without spiked standards. However, we do not expect any global deviation from the reported values. To obtain global view of dynamic range of *E. coli* proteome, the graph was plotted to show the cumulative benefaction of each proteins to the total proteins. The 22 most abundant proteins correspond to 25% of the proteome (Figure 1) with outer membrane protein A, DNA-binding protein HU1, braun lipoprotein, and glyceraldehyde-3-phosphate dehydrogenase helping to constitute more than 1% each of the proteome.

We also calculated the median iBAQ values of the *E. coli* proteome and plotted the absolute abundance of quantified 2068 *E. coli* proteins. The iBAQ values vary by over five orders of magnitude in *E. coli* proteome (Figure 2A). The most highly expressed 25 proteins include glycy radical cofactor, outer membrane protein, DNA-binding protein, braun lipoprotein, glyceraldehyde-3-phosphate dehydrogenase, ribosomal proteins, chaperonin (Figure 2B), and 25 most low abundant proteins with their low intensity are ATP-binding protein, glycerate kinase, ABC transporter periplasmic-binding protein, uncharacterized proteins, acriflavine resistance protein, cell filamentation protein, heat-shock regulatory protein, lipoprotein, and other proteins (Figure 2C). The above-stated values are the approximate estimation of abundance of the expressed proteins, and there might be an underestimation for low expressed proteins. Here, the proteome recorded by LC-MS/MS is not complete probably due to low expression levels of some of *E. coli* proteins, and also no pre-fractionation of *E. coli* peptides and proteins were performed. In spite of the limitations, the single-run were able to cover considerable depth of high abundant and substantial number of low abundant proteins.

Coverage of Cellular Components, Biological Processes and Pathways

To check the coverage of our dataset with the known cellular pathways, cellular compartments, biological processes, we overlapped our quadruplicate dataset with the database of the Kyoto Encyclopedia of Genes and Genome (KEGG) [23], and Gene Ontology (GO) annotations [21, 22] (Figure 3).

Table 1. Identification of proteins in single- and quadruplicate runs

	Proteins in a single-run	Proteins in quadruplicate runs	Percentage coverage in a single-run	Difference in proteins between quadruplicate and single-run	Difference % in proteins between quadruplicate and single-run
Experiment 1	2038	2068	98.5	30	1.45
Experiment 2	2025		97.9	43	2.07
Experiment 3	2031		98.2	37	1.78
Experiment 4	1950		94.3	118	5.7

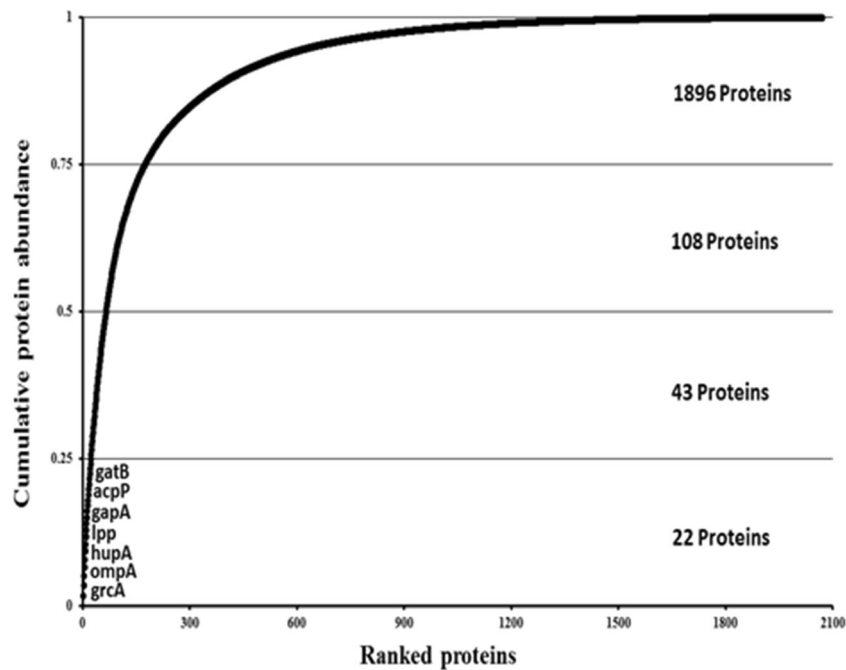


Figure 1. Quantitation of identified *E. coli* K-12 proteins by iBAQ. Cumulative proteins abundance distribution from high abundant proteins to low abundant proteins. Some of the high abundant proteins are labeled in the graph

Interestingly, 93 of 158 KEGG pathways were presented with 80% of their members while 139 of 158 with 50% members (Figure 3C, Supplementary Table S2). Some examples of

signaling pathways include citrate cycle–TCA cycle (26 of 27 proteins) and mismatch repair (19 of 22 proteins) as well. Moreover, a comparison of some of the important KEGG

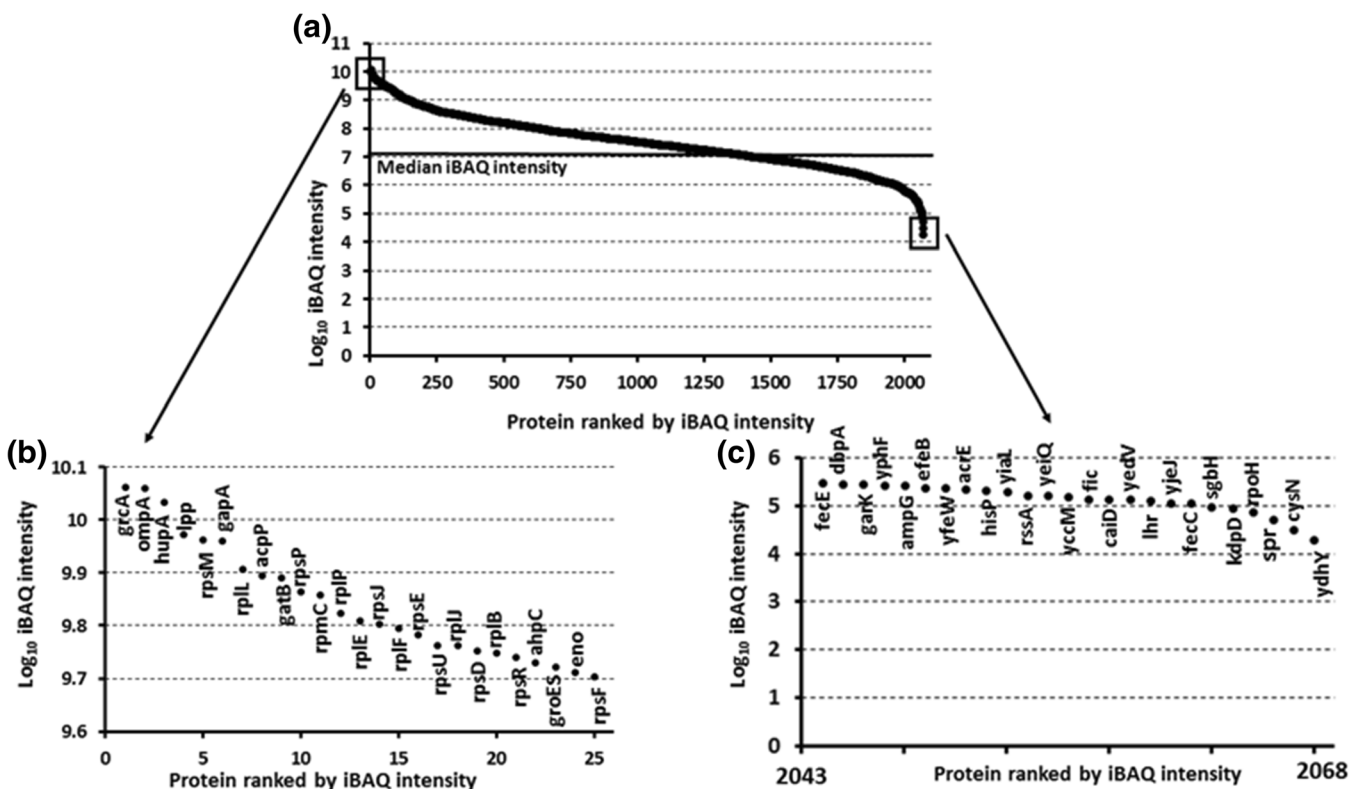


Figure 2. (a) Dynamic distribution of *E. coli* proteins based on their iBAQ intensity. (b) Distribution of 25 most high abundant proteins. (c) 25 most low abundant protein distribution

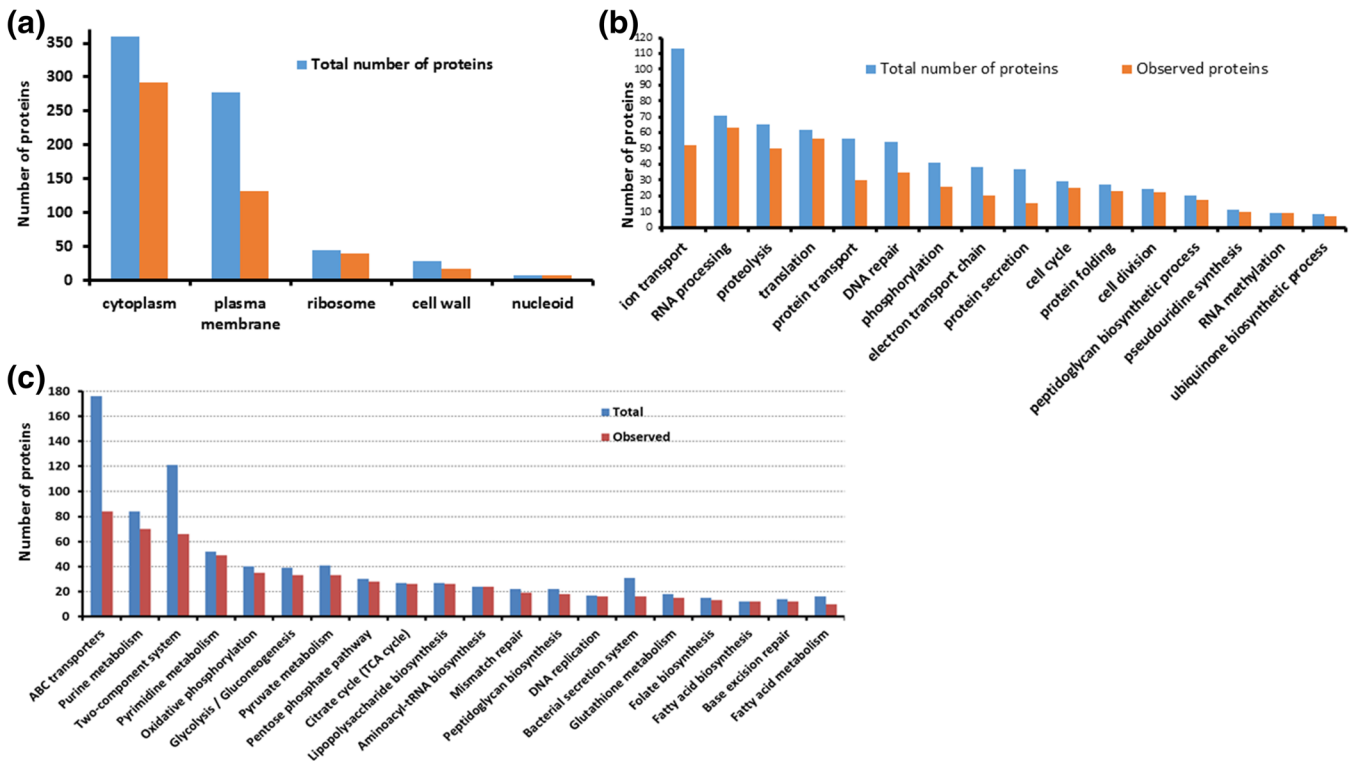


Figure 3. Distribution of different GO annotations and KEGG pathways for the observed proteins in *E. coli* in relation to the total number of proteins. (a) Cellular compartments (b) biological processes (c) KEGG pathways

pathways covered by quadruplicate runs with each single-run such as Experiment 1, Experiment 2, Experiment 3 and Experiment 4 shows that single-run is almost as good as quadruplicate runs (Supplementary Figure S2).

We looked at folate biosynthesis (Figure 4) and found that our dataset covered almost all proteins except two proteins. The proteins (*folE*, *moaC*, *moaD*, *moaE*, *sscR*, *ycgF*, *queC*, *queF*, *phoA*, *folB*, *folK*, *folP*, *folC*, *folA*, *pabC*, *pabA*) are detected and

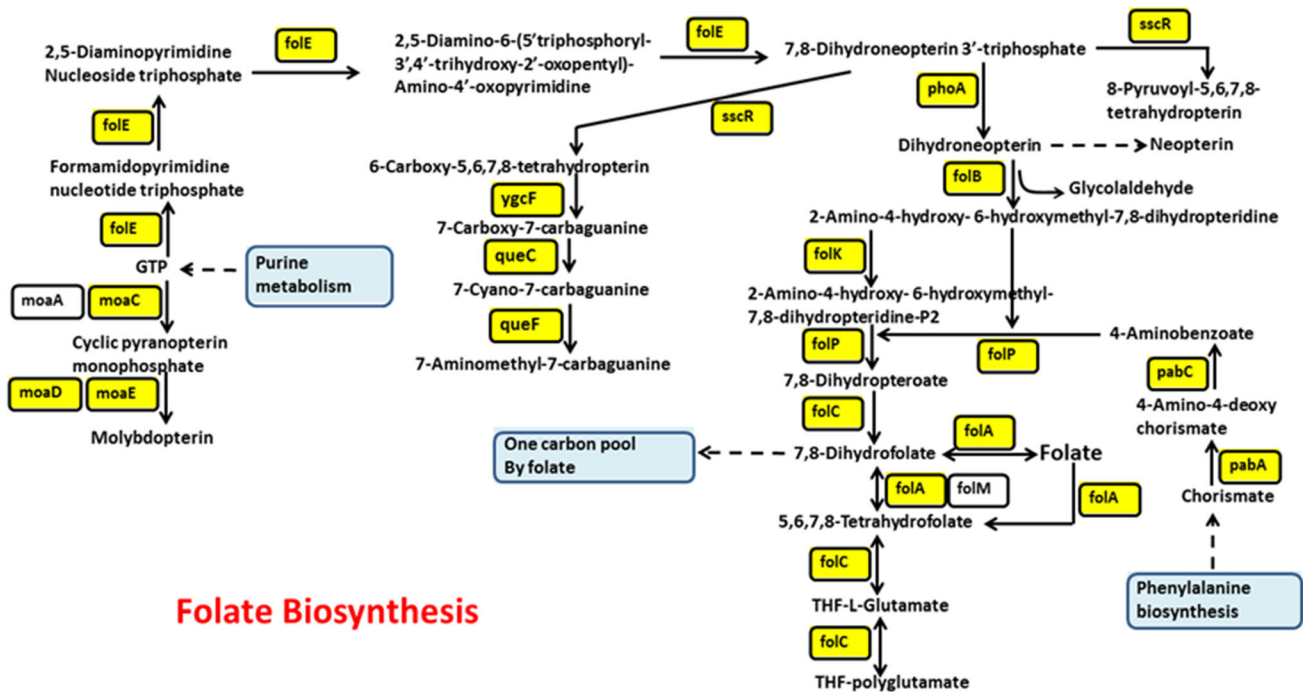


Figure 4. Schematic representation of proteins involved in folate biosynthesis. Proteins that were identified in our analysis are in yellow color while the proteins not found in our analysis are white in color

identified by LC-MS/MS run analysis with 9, 7, 3, 4, 4, 7, 7, 11, 2, 4, 2, 10, 13, 6, 3, and 2 unique peptides respectively (Supplementary Table S3). Interestingly, each single-run covered almost all of the proteins in folate biosynthesis as quadruplicate runs (Supplementary Figure S3).

For the GO annotations, quadruplicate run dataset covered almost all cellular compartments and processes. Remarkably, 675 of 1262 biological processes and 64 of 119 cellular components were represented with 80% of their members (Supplementary Tables S4 and S5). Some of the major coverage of cellular compartments and biological processes are shown in Figure 3A, B. This proves that almost all the KEGG pathways, cellular compartments, and biological processes are covered by this dynamic range of quadruplicate LC-MS/MS runs. Furthermore, each single-run has ability to cover mostly all cellular compartment and biological processes as quadruplicate runs (Supplementary Figures S4 and S5). Note that all the members of signaling pathways, biological process are not expressed in *E. coli* grown in laboratory standard conditions.

Protein-Protein Interactions

We wanted to investigate known protein-protein interactions network in some of the important biological processes in *E. coli* by our quadruplicate run LC-MS/MS dataset. Therefore, we took the proteins of some of the biological process from Figure 3B and used Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database that provides physical and functional interactions [28]. The STRING generated protein-protein interaction network with tight connected clusters. This is seen in Figures 5 and 6 with some of the important biological

processes such as cell cycle (Figure 5A), DNA repair (Figure 5B), ion transport (Figure 5C), ubiquinone biosynthetic process (Figure 5D), pseudouridine synthesis (Figure 6A), peptidoglycan biosynthetic process (Figure 6B), RNA processing (Figure 6C), and translation (Figure 6D). These protein-protein interactions are detected and identified by different number of unique peptides (Supplementary Table S6). Interestingly, we also compared quadruplicate runs with each of the single-run of some of the biological process from Figure 3b to demonstrate that single-run is almost as good as quadruplicate run (Supplementary Figures S6 and S7).

To investigate the common proteins and protein-protein interaction between different biological processes, we combined the proteins of cell cycle and peptidoglycan biosynthetic process in STRING and found that there is dense interaction network between these two processes with presence of common proteins *mraY*, *murF*, *murE*, *murG*, *murD*, and *murC* that are identified by 2, 6, 16, 12, 12, and 15 number of unique peptides respectively (Supplementary Figure S8, Supplementary Table S6). The proteins from pseudouridine synthesis, RNA processing and translation process also forms protein interaction network among them when combined together in STRING (Supplementary Figure S9). The proteins common to both pseudouridine synthesis and RNA processing are *truA*, *truD*, and *truB* that are identified by 6, 13, and 13 number of unique peptides respectively (Supplementary Figure S9, Supplementary Table S6). The *rplA* protein is common between RNA processing and translation process and is identified by 21 unique peptides. Similarly, transcription and translation processes also form protein-protein interaction network when combined together in STRING, and these interaction networks are identified by different unique

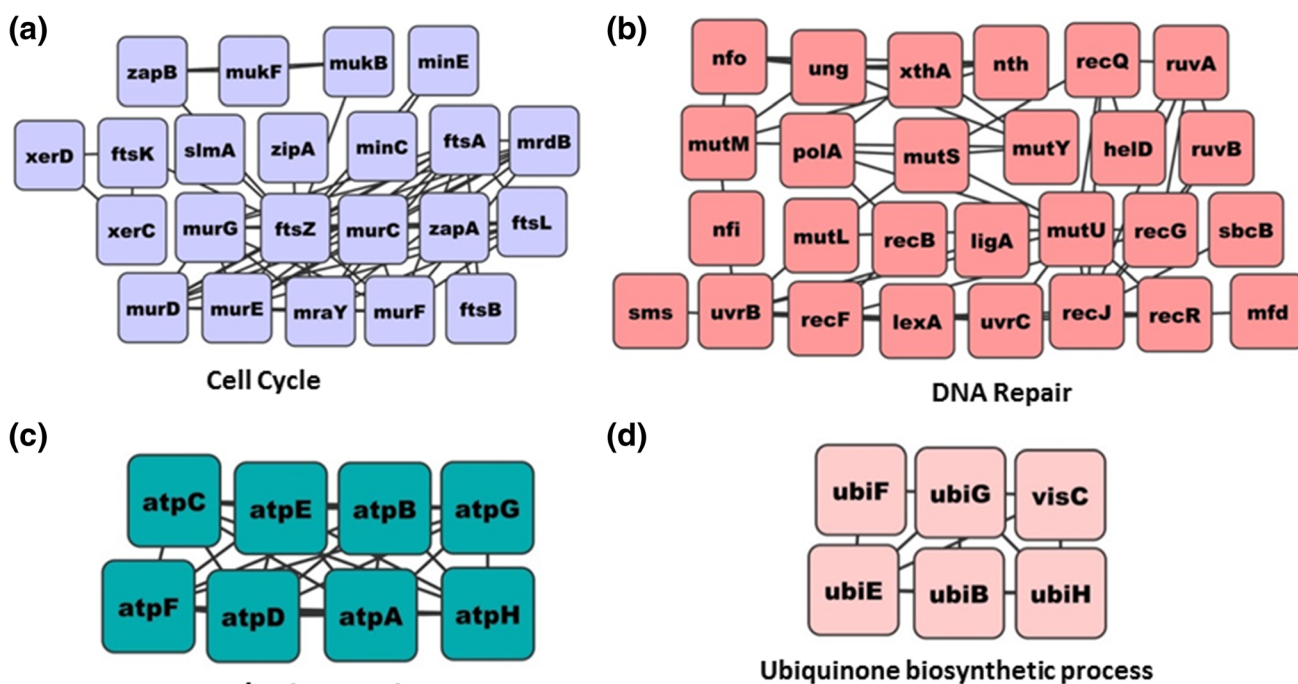


Figure 5. Protein-protein interactions network in (a) cell cycle (b) DNA repair (c) ion transport (d) ubiquinone biosynthetic process identified by our LC-MS/MS dataset of *E. coli*. These were analyzed by STRING and visualized by Cytoscape

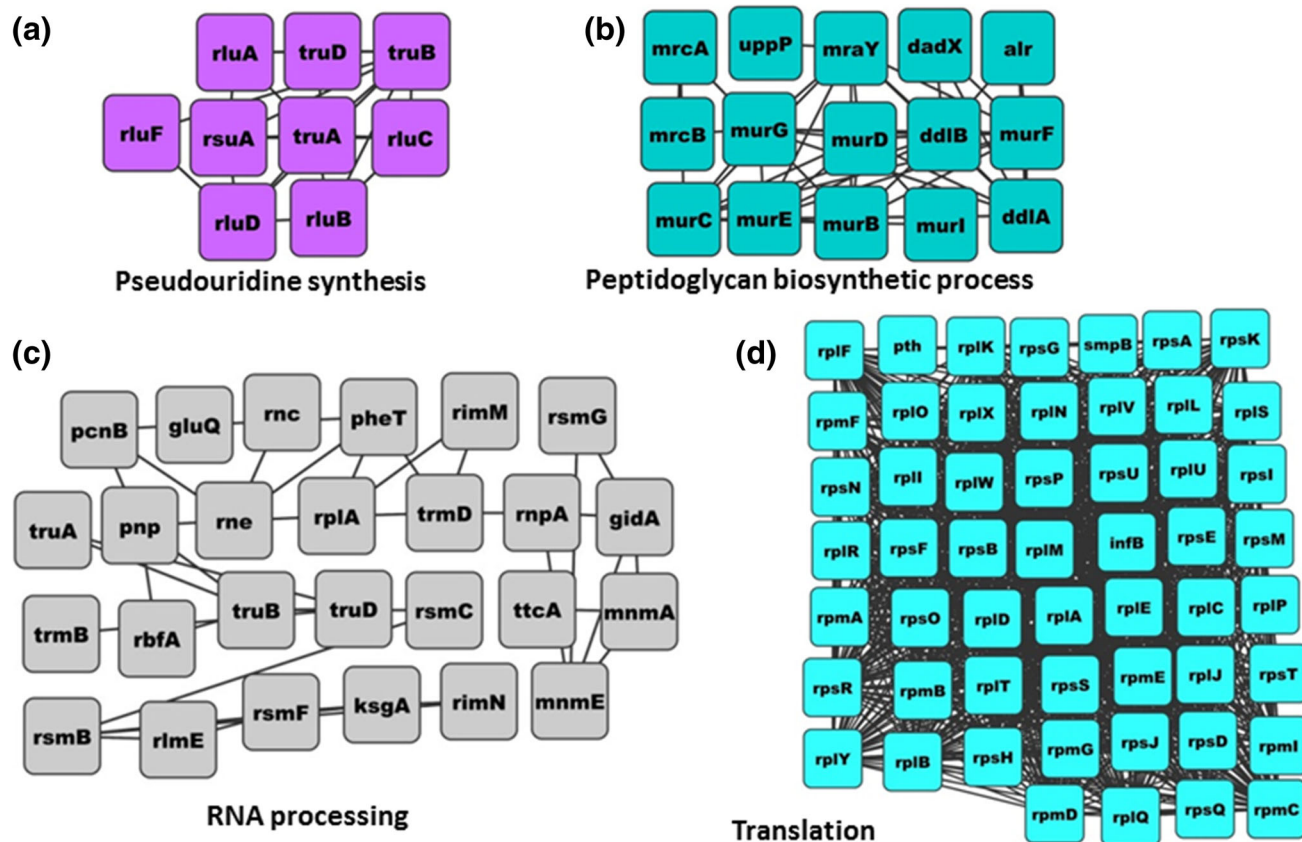


Figure 6. Protein-protein interaction network in (a) pseudouridine synthesis, (b) peptidoglycan biosynthetic process, (c) RNA processing, and (d) translation identified by our LC-MS/MS dataset of *E. coli*. These were analyzed by STRING and visualized by Cytoscape

peptides (Supplementary Figure S10, Supplementary Table S6). This shows that different protein-protein interaction network of *E. coli* are also covered by our quadruplicate run dataset and identified with different unique peptides.

Conclusions and Outlook

Here, we see that without pre-fractionation, *E. coli* proteome quantified by iBAQ was covered to a considerable depth in quadruplicate LC-MS/MS run analysis using single enzyme-trypsin. These quantified proteins are about 52% of the total proteins present in *E. coli* genome according to Uniprot database. Our dataset captured a large part of GO annotations and signaling pathways along with protein-protein interactions. Most of the proteins of folate biosynthesis have been observed in single-run. Here, the protein-protein interaction networks that are covered in this quadruplicate run analysis are derived from various active interaction sources using STRING. We saw biological processes sharing common proteins among themselves along with protein-protein interaction network among different biological processes. Our data shows great sensitivity and coverage of quantified *E. coli* proteome using LTQ-Orbitrap Velos mass spectrometry and Proxeon-Nano ESI liquid chromatography pumps where

there was no requirement of expensive UPLC with very high pressure.

This single-run LC-MS/MS analysis involves considerable reduced sample preparation, a few micro gram *E. coli* proteins, economical in-house made long chromatography column, column oven and reduced measurement time in mass spectrometry compared to extensive fractionation. The assurance of single LC-MS/MS run is increased by running replicates such as duplicate, triplicate, and quadruplicate runs and identifying the same proteins again. Our single-run method using liquid chromatography runs of 12 h was able to achieve 98% coverage of quadruplicate runs of *E. coli* proteome. This method is very useful for quick understanding of complex cellular system of *E. coli*. The protein-protein interaction network shown here in an unbiased way is an important step in the destination of interactome mapping by capturing important properties. It can serve as a platform for many hypothesis-driven experiments by cumulative effect of interplay of different signaling pathways, protein interaction network, and quantified proteins. Moreover, mutations can also be carried out without affecting the functions of networking proteins. Maybe in the future, advancement in technology will help to identify whole of *E. coli* proteome in single-run without any requirement for replicates and pre-fractionation experiments.

Accession Informations

All described nano-LC-MS/MS data (Raw files) may be downloaded from ProteomeXchange (<http://proteomecentral.proteomexchange.org>). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier Project accession: PXD010139. Login for reviewers is provided under Username: reviewer22855@ebi.ac.uk and Password: 8fraiZ3Z.

Acknowledgements

We are thankful to Prof. Matthias Mann for his great help. Mass spectrometry experiment has done in Prof. Matthias Mann lab at Department of Proteomics and Signal Transduction, Max-Planck-Institute of Biochemistry, Munich, Germany. We are thankful to Dr. Florian Fröhlich for providing *E. coli*.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

References

- Thakur, S.S., Geiger, T., Chatterjee, B., Bandilla, P., Fröhlich, F., Cox, J., Mann, M.: Deep and highly sensitive proteome coverage by LC-MS/MS without pre-fractionation. *Mol. Cell. Proteomics*. **10**, M110.003699 (2011)
- Aebersold, R., Mann, M.: Mass spectrometry-based proteomics. *Nature*. **422**, 198–207 (2003)
- Cravatt, B.F., Simon, G.M., Yates 3rd, J.R.: The biological impact of mass spectrometry-based proteomics. *Nature*. **450**, 991–1000 (2007)
- Choudhary, C., Mann, M.: Decoding signalling networks by mass spectrometry-based proteomics. *Nat. Rev. Mol. Cell Biol.* **11**, 427–439 (2010)
- Yates 3rd, J.R., Gilchrist, A., Howell, K.E., Bergeron, J.J.: Proteomics of organelles and large cellular structures. *Nat. Rev. Mol. Cell Biol.* **6**, 702–714 (2005)
- Bantscheff, M., Eberhard, D., Abraham, Y., Bastuck, S., Boesche, M., Hobson, S., Mathieson, T., Perrin, J., Raida, M., Rau, C., Reader, V., Sweetman, G., Bauer, A., Bouwmeester, T., Hopf, C., Kruse, U., Neubauer, G., Ramsden, N., Rick, J., Kuster, B., Drewes, G.: Quantitative chemical proteomics reveals mechanisms of action of clinical ABL kinase inhibitors. *Nat. Biotechnol.* **25**, 1035–1044 (2007)
- Guiochon, G.: The limits of the separation power of unidimensional column liquid chromatography. *J. Chromatogr. A*. **1126**, 6–49 (2006)
- MacNair, J.E., Lewis, K.C., Jorgenson, J.W.: Ultrahigh-pressure reversed-phase liquid chromatography in packed capillary columns. *Anal. Chem.* **69**, 983–989 (1997)
- Shen, Y., Zhao, R., Belov, M.E., Conrads, T.P., Anderson, G.A., Tang, K., Pasa-Tolić, L., Veenstra, T.D., Lipton, M.S., Udseth, H.R., Smith, R.D.: Packed capillary reversed-phase liquid chromatography with high-performance electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry for proteomics. *Anal. Chem.* **73**, 1766–1775 (2001)
- Iwasaki, M., Miwa, S., Ikegami, T., Tomita, M., Tanaka, N., Ishihama, Y.: One-dimensional capillary liquid chromatographic separation coupled with tandem mass spectrometry unveils the *Escherichia coli* proteome on a microarray scale. *Anal. Chem.* **82**, 2616–2620 (2010)
- Luo, Q., Shen, Y., Hixson, K.K., Zhao, R., Yang, F., Moore, R.J., Mottaz, H.M., Smith, R.D.: Preparation of 20-microm-i.d. silica-based monolithic columns and their performance for proteomics analyses. *Anal. Chem.* **77**, 5028–5035 (2005)
- Motoyama, A., Venable, J.D., Ruse, C.I., Yates 3rd, J.R.: Automated ultra-high-pressure multidimensional protein identification technology (UHP-MudPIT) for improved peptide identification of proteomic samples. *Anal. Chem.* **78**, 5109–5118 (2006)
- Liu, H., Finch, J.W., Lavallee, M.J., Collamati, R.A., Benevides, C.C., Gebler, J.C.: Effects of column length, particle size, gradient, length and flow rate on peak capacity of nano-scale liquid chromatography for peptide separations. *J. Chromatogr. A*. **1147**, 30–36 (2014)
- Wiśniewskia, J.R., Rakus, D.: Quantitative analysis of the *Escherichia coli* proteome. *Data Brief*. **1**, 7–11 (2014)
- Soufi, B., Krug, K., Harst, A., Macek, B.: Characterization of the *E. coli* proteome and its modifications during growth and ethanol stress. *Front. Microbiol.* **6**, 103 (2015)
- Schmidt, A., Kochanowski, K., Vedelaar, S., Ahmé, E., Volkmer, B., Callipo, L., Knoops, K., Bauer, M., Aebersold, R., Heinemann, M.: The quantitative and condition-dependent *Escherichia coli* proteome. *Nat. Biotechnol.* **34**, 104–110 (2016)
- Schwahnhauser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., Selbach, M.: Global quantification of mammalian gene expression control. *Nature*. **473**, 337–342 (2011)
- Rappsilber, J., Ishihama, Y., Mann, M.: Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75**, 663–670 (2003)
- Olsen, J.V., Macek, B., Lange, O., Makarov, A., Homing, S., Mann, M.: Higher-energy C-trap dissociation for peptide modification analysis. *Nat. Methods*. **4**, 709–712 (2007)
- Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A., Olsen, J.V., Mann, M.: Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10**, 1794–1805 (2011)
- Barrell, D., Dimmer, E., Huntley, R.P., Binns, D., O'Donovan, C., Apweiler, R.: The GOA database in 2009—an integrated gene ontology annotation resource. *Nucleic Acids Res.* **37**, D396–D403 (2009)
- Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G.: Gene ontology: tool for the unification of biology. The gene ontology consortium. *Nat. Genet.* **25**, 25–29 (2000)
- Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., Hattori, M.: The KEGG resource for deciphering the genome. *Nucleic Acids Res.* **32**, D277–D280 (2004)
- Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T.: Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003)
- Kuhnert, P., Nicolet, J., Frey, J.: Rapid and accurate identification of *Escherichia coli* K-12 strains. *Appl. Environ. Microbiol.* **61**, 4135–4139 (1995)
- Cox, J., Mann, M.: MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008)
- Nagaraj, N., Wisniewski, J.R., Geiger, T., Cox, J., Kircher, M., Kelso, J., Pääbo, S., Mann, M.: Deep proteome and transcriptome mapping of a human cancer cell line. *Mol. Syst. Biol.* **7**, 548 (2011)
- Jensen, L.J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller, J., Doerks, T., Julien, P., Roth, A., Simonovic, M., Bork, P., von Mering, C.: STRING 8—a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res.* **37**, D412–D416 (2009)