Deciphering the Cell Envelope Proteome of *Escherichia coli*

M Papanastasiou1, G Orfanoudaki1,2, M Frantzèskos-Sardis1,2, P Mavroudis2, M Aivaliotis1, S Karamanou1, A Economou1,2

1Institute of Molecular Biology & Biotechnology, FORTH, Iraklio-Crete, Greece
2Department of Biology, UoC, Iraklio-Crete, Greece

**Introduction**

- The E. coli cell envelope is a complex assembly that comprises the inner and outer membrane, separated by the periplasm and a number of proteins associated peripherally to the membranes.
- Cell envelope proteins are crucial for cell viability and pathogenicity, vaccine development and recombinant protein biotechnology. Characterizing their function is crucial therefore for understanding of how a cell operates.
- Most proteomic studies carried out to date have dealt mainly with the soluble counterparts of E. coli. Membrane proteomes are under-represented due to the hydrophobic nature of their proteins [1].
- In this study, we combine traditional biochemical techniques with mass spectrometry analysis and transcriptomics data available in the literature, in an effort to elucidate the cell envelope proteome (CEP) of E. coli.
- In addition, using manual and multi-combinatorial bioinformatics approaches of the available E. coli proteome databases, we perform extensive curation and topological re-evaluation of strain K12 proteins. The topology curation is extrapolated also for the BL21(DE3) strain.
- Overall, this work will serve as a basis for the analysis of the cell envelope interactome of laboratory and enteropathogenic E. coli strains.

**A. Bioinformatics Approach**

**Figure 1**: Strategy for the systematic curation of the E. coli K12 cell envelope proteome.

For the sub-cellular annotation of CEP proteins, Uniprot, Bernal & Daley [1] EchoLOCATION [2] and data from the literature were used as a starting point. The proteins were cross-checked with a number of prediction tools (LipoP, SignalP, TatP, TMHMM, Proteins, PSORTB). BLAST and proteomics data reported in the literature. By combining these results, a final curated E. coli CEP proteome emerged.

**Figure 2**: (A) Protein categorization and sub-cellular annotation of E. coli strains (B) K12 and (C) BL21(DE3). The K12 and BL21(DE3) proteomes are composed of 4406 and 4156 proteins respectively. It is shown that the cell envelope comprises ~45% of each proteome, a number far exceeding current proteome databases, we perform extensive curation and topological re-evaluation of strain K12 proteins. The topology curation is extrapolated also for the BL21(DE3) strain.

**Figure 3** (A) Effect of sample preparation treatment on the nature of proteins detected by nanoLC-MS/MS. Cytoplasmic proteins tend to give intense signals (44%) even after the first sucrose gradient treatment of inverted membrane vesicles (IMVs). Extensive washing of the membranes and flotation of IMVs in a sucrose gradient (IMVs treated) results in an enriched CEP proteome fraction (70%).

**Results & Discussion**

**Figure 4**: Breakdown scheme of *E. coli* undetected basic CEP proteome. Out of the 809 proteins that have not been identified, and taking into account MS parameters used in this study as well as transcriptomics data available in the literature, there is still 468 proteins to be detected. In-solution trypsin digestion of the sample is believed to aid in the identification of those proteins and result in a complete CEP proteome coverage by nanoLC-MS/MS.

**Materials & Methods**

**References**

1) Weiner & Li, 2008, Biochimica Biophysica Acta
2) Bernal & Daley, 2009, Trends in Microbiology
3) Horler et al, 2009, Bioinformatics

**Acknowledgements**: The research leading to these results has received funding from the European Commission 6th Framework Programme agreement n° LSHC-CT-2006-037834 ‘Streptomics’. M.F.S. is supported by a Greek Ministry of Science and Technology fellowship and G.O. is Onassis Foundation pre-doctoral fellow. M.F.S. received funding from the EC Sixth Framework Programme agreement 229823 Capacities-FP7-REGPOT-2008-1/ project «ProFI».

**Figure 2**: (A) Protein categorization and sub-cellular annotation of E. coli strains (B) K12 and (C) BL21(DE3). The K12 and BL21(DE3) proteomes are composed of 4406 and 4156 proteins respectively. It is shown that the cell envelope comprises ~45% of each proteome, a number far exceeding current proteome databases, we perform extensive curation and topological re-evaluation of strain K12 proteins. The topology curation is extrapolated also for the BL21(DE3) strain.

**Figure 3** (A) Effect of sample preparation treatment on the nature of proteins detected by nanoLC-MS/MS. Cytoplasmic proteins tend to give intense signals (44%) even after the first sucrose gradient treatment of inverted membrane vesicles (IMVs). Extensive washing of the membranes and flotation of IMVs in a sucrose gradient (IMVs treated) results in an enriched CEP proteome fraction (70%).

**Figure 4**: Breakdown scheme of *E. coli* undetected basic CEP proteome. Out of the 809 proteins that have not been identified, and taking into account MS parameters used in this study as well as transcriptomics data available in the literature, there is still 468 proteins to be detected. In-solution trypsin digestion of the sample is believed to aid in the identification of those proteins and result in a complete CEP proteome coverage by nanoLC-MS/MS.