World Academy of Science, Engineering and Technology
International Journal of Biotechnology and Bioengineering
Vol:10, No:09, 2016

Identification and Characterization of Small Peptides Encoded by Small Open Reading Frames using Mass Spectrometry and Bioinformatics

Authors : Su Mon Saw, Joe Rothnagel

Abstract : Short open reading frames (sORFs) located in 5'UTR of mRNAs are known as uORFs. Characterization of uORF-encoded peptides (uPEPs) i.e., a subset of short open reading frame encoded peptides (sPEPs) and their translation regulation lead to understanding of causes of genetic disease, proteome complexity and development of treatments. Existence of uORFs within cellular proteome could be detected by LC-MS/MS. The ability of uORF to be translated into uPEP and achievement of uPEP identification will allow uPEP's characterization, structures, functions, subcellular localization, evolutionary maintenance (conservation in human and other species) and abundance in cells. It is hypothesized that a subset of sORFs are translatable and that their encoded sPEPs are functional and are endogenously expressed contributing to the eukaryotic cellular proteome complexity. This project aimed to investigate whether sORFs encode functional peptides. Liquid chromatography-mass spectrometry (LC-MS) and bioinformatics were thus employed. Due to probable low abundance of sPEPs and small in sizes, the need for efficient peptide enrichment strategies for enriching small proteins and depleting the sub-proteome of large and abundant proteins is crucial for identifying sPEPs. Low molecular weight proteins were extracted using SDS-PAGE from Human Embryonic Kidney (HEK293) cells and Strong Cation Exchange Chromatography (SCX) from secreted HEK293 cells. Extracted proteins were digested by trypsin to peptides, which were detected by LC-MS/MS. The MS/MS data obtained was searched against Swiss-Prot using MASCOT version 2.4 to filter out known proteins, and all unmatched spectra were re-searched against human RefSeq database. ProteinPilot v5.0.1 was used to identify sPEPs by searching against human RefSeq, Vanderperre and Human Alternative Open Reading Frame (HaltORF) databases. Potential sPEPs were analyzed by bioinformatics. Since SDS PAGE electrophoresis could not separate proteins <20kDa, this could not identify sPEPs. All MASCOT-identified peptide fragments were parts of main open reading frame (mORF) by ORF Finder search and blastp search. No sPEP was detected and existence of sPEPs could not be identified in this study. 13 translated sORFs in HEK293 cells by mass spectrometry in previous studies were characterized by bioinformatics. Identified sPEPs from previous studies were <100 amino acids and <15 kDa. Bioinformatics results showed that sORFs are translated to sPEPs and contribute to proteome complexity. uPEP translated from uORF of SLC35A4 was strongly conserved in human and mouse while uPEP translated from uORF of MKKS was strongly conserved in human and Rhesus monkey. Cross-species conserved uORFs in association with protein translation strongly suggest evolutionary maintenance of coding sequence and indicate probable functional expression of peptides encoded within these uORFs. Translation of sORFs was confirmed by mass spectrometry and sPEPs were characterized with bioinformatics.

Keywords : bioinformatics, HEK293 cells, liquid chromatography-mass spectrometry, ProteinPilot, Strong Cation Exchange Chromatography, SDS-PAGE, sPEPs
Conference Title : ICABB 2016 : International Conference on Advanced Bioengineering and Biotechnology
Conference Location : Singapore, Singapore
Conference Dates : September 08-09, 2016