OPEN ACCESS Freely available online

ISSN: 2329-6682
Gene Technology

Commentary

Novel Gene Discovery through N-Terminal Proteomics in *Mycobacterium Tuberculosis*

Shu Mai*

Department of Biotechnology, Zhejiang University, Hangzhou, China

DESCRIPTION

Severe Tuberculosis (TB) is caused by *Mycobacterium Tuberculosis* (MTB). The complete genome of *M. tuberculosis* type strain, H37Rv, was first sequenced and annotated with 4010 sequences. Protein N-termini are crucial structural and functional components of annotated genes. Important regulatory information and disease-causing gene processes are provided by the border of a certain ORF's 5' Deoxyribonucleic Acid (DNA) sequence. Post-Translational Modification (PTM), which substantially influences the fate of the protein itself, is likewise determined by the N-terminal structure.

PTM includes the removal of the N-terminal initiator methionine, acetylation, and Proteolytic maturation by signal- and pro-peptide removal. The prevalence of gene prediction errors, despite the fact that genome sequencing and numerous gene prediction methods have been widely used in genome annotation. Because of the information gap in gene structure and regulation model, the inaccurate prediction of Translation Initiation Sites (TISs) in prokaryotic genomes is one of these inaccuracies that are particularly pervasive. The bacterial genome's start codon error rates ranged from 10% to 44%.

Several enrichment techniques using various derivatization chemicals and liquid-phase separation have been developed to increase the analytical power of N-terminal in complicated samples. They were able to detect 278 N-CLAP peptides using selective affinity purification and Tandem Mass Spectrometry (TMS) analysis an N-terminal enrichment technique called N-CLAP (N-Terminalomics by Chemical Labelling of the -Amine of Proteins). TNBS (2,4,6-Trinitrobenzenesulfonic Acid) and the amino group of internal peptides react to produce N-terminal peptides using the COFRADIC method, which was developed.

The H37Rv genome of *M. tuberculosis* has been fully sequenced and annotated for more than 20 years. Several tactics, including the use of MS-based technologies, have been used to increase the annotation accuracy of H37Rv. Since 2004, proteogenomics has emerged as one of the most effective tools for improving genome annotation in numerous species, including Eukarya, Bacteria, and Archaea. In *M. tuberculosis* H37Rv, a sizable portion of annotated genes were confirmed and novel genes were reannotated with gene-specific new peptides. The identification of a gene's N-terminal is one of the most significant stages in gene reannotation because of its crucial features and the stability of proteins in cells. Due to the lack of coupled technology for proteogenomics and N-terminal proteomics, the specific TISs for the majority of the annotated genes in H37Rv have not yet been determined. Prior to protease digestion TMPP-labeled the N-terminal of the whole cell lysate collected from H37Rv, which improved the N-terminal peptide identification due to greater ionisation efficiency from TMPP-labeled peptides.

The TMPP labelled samples even further to discover more N-terminal peptides by minimizing their complexity. Using a proteogenomics technique and a six-frame database search, we were able to identify 2123 N-terminal peptides for 1141 ORFs, including 43 incorrectly annotated genes and 1 new gene. In comparison to the N-terminal identification following positive or negative choices for the tagged N-terminal peptides, the N-terminal peptide made up 6.38% of all identified peptides.

Together with correcting 43 existing genes with new N-termini, the newly discovered peptides. With the aid of these outcomes, we were able to create the first N-terminomics landscape for *M. tuberculosis* H37Rv. We used a comparative genomic technique to check the genome annotations in the other 129 sequenced type strains of *Mycobacteriaceae*, including *Mycobacterium*, *Mycolicibacter*, *Mycobacteroides*, and *Mycoicibacillus*, using the corrected N-termini of 43 annotated genes. We used TMPP labelling and an enhanced StageTip separation technique to conduct a deep coverage N-terminomics study on *M. tuberculosis* H37Rv. By identifying 2234 naturally occurring N-terminal peptides, re-correcting 43 genes' N-terminal boundaries, and discovering the novel gene Rv1078A for H37Rv, we were able to create an N-terminal proteome landscape using the optimum parameters. Also, we discovered that GTG and TTG were both two of the three key start codons in *M. tuberculosis* H37Rv, the third being ATG, which even took up 46% of all the updated N-

Correspondence to: Shu Mai, Department of Biotechnology, Zhejiang University, Hangzhou, China, Email: shumai@gmail.cn

Received: 02-Feb-2023, Manuscript No. RDT-23-20651; **Editor assigned:** 06-Feb-2023, PreQC No. RDT-23-20651 (PQ); **Reviewed:** 20-Feb-2023, QC No. RDT-23-20651; **Revised:** 27-Feb-2023, Manuscript No. RDT-23-20651 (R); **Published:** 06-Mar-2023, DOI: 10.35248/2329-6682.23.12.216

Citation: Mai S (2023) Novel Gene Discovery through N-Terminal Proteomics in *Mycobacterium Tuberculosis*. Gene Technol. 12:216.

Copyright: © 2023 Mai S. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

termini of 43 genes. One of the main causes of incorrect genome annotation is over-prediction of the initiation codon with ATG. The technique and the data presented here are anticipated to be very helpful for carrying out effective N-terminal proteomics analysis, even though some of the parameters may need to be adjusted when used to other bacteria. In a large-scale proteomics investigation, TMPP labelling enrichment was an effective method for identifying N-terminal peptides, which helped to confirm or correct the TISs of annotated encoding genes in the genome.