

Global landscape of lysine succinylome in *Saccharopolyspora erythraea*

Xiang Ke¹, Xing Jiang¹, Xiwei Tian¹, and Ju Chu¹

¹East China University of Science and Technology

June 13, 2023

Abstract

As a novel protein post-translational modification, lysine succinylation is widely involved in metabolism regulation. In this study, we focused on the distribution of lysine succinylation sites and their physiological functions in *Saccharopolyspora erythraea*. Using high-resolution 4D label free mass spectrometry, a large and global protein succinylome was identified in a hypersuccinylated strain E3ΔsucC. The results showed that succinylated proteins are predominantly involved in protein synthesis-related pathways (e.g., ribosomes, tRNA) and metabolic pathways, such as the TCA cycle. Proteins in these pathways generally have a higher lysine content, suggesting that lysine succinylation may have a greater regulatory role in biochemical reactions involving acidic substrates. Motif analysis revealed that charged amino acids (D, E, K, R and W) display a more regular distribution around acylation sites, implying that the polar effect between residues may be the key factor influencing lysine succinylation. Based on predicted protein structures, we highlighted the potential impact of lysine succinylation on enzyme activity in the TCA cycle. In conclusion, this study offers valuable insights into the regulation of lysine succinylation and contributes to a comprehensive understanding of its physiological functions in actinomycetes.

Global landscape of lysine succinylome in *Saccharopolyspora erythraea*

Xiang Ke, Xing Jiang, Xiwei Tian* and Ju Chu*

State Key Laboratory of Bioreactor Engineering, East China University of Science and Technology, Shanghai 200237, China

*Correspondence:

Xiwei Tian: xiweitian@ecust.edu.cn; Tel. (+86) 021-64257019; 130 Meilong Road, Shanghai 200237, China.

Ju Chu: juchu@ecust.edu.cn; Tel. (+86) 021-64253021; 130 Meilong Road, Shanghai 200237, China.

Abbreviations:

PTM, post-translational modification; TCA, tricarboxylic acid; CoA, coenzyme A; FDR, false detection rate; KEGG, Kyoto Encyclopedia of Genes and Genomes.

Key words: *Saccharopolyspora erythraea*, Mass spectrometry, Succinylome

ABSTRACT

As a novel protein post-translational modification, lysine succinylation is widely involved in metabolism regulation. In this study, we focused on the distribution of lysine succinylation sites and their physiological functions in *Saccharopolyspora erythraea*. Using high-resolution 4D label free mass spectrometry, a large and global protein succinylome was identified in a hypersuccinylated strain E3ΔsucC. The results showed that succinylated proteins are predominantly involved in protein synthesis-related pathways (e.g., ribosomes, tRNA) and metabolic pathways, such as the TCA cycle. Proteins in these pathways generally have a higher

lysine content, suggesting that lysine succinylation may have a greater regulatory role in biochemical reactions involving acidic substrates. Motif analysis revealed that charged amino acids (D, E, K, R and W) display a more regular distribution around acylation sites, implying that the polar effect between residues may be the key factor influencing lysine succinylation. Based on predicted protein structures, we highlighted the potential impact of lysine succinylation on enzyme activity in the TCA cycle. In conclusion, this study offers valuable insights into the regulation of lysine succinylation and contributes to a comprehensive understanding of its physiological functions in actinomycetes.

Statement of significance

This study provides a comprehensive protein succinylome of *Saccharopolyspora erythraea*, aiming to explore the integrated impact of succinylation on cellular primary and secondary metabolism. The corresponding analysis not only reveals the functional distribution of succinylated proteins in *S. erythraea*, but also explores the effects of lysine proportion, amino acid residue polarity, protein structure, and other factors on the probability of lysine succinylation. These findings would not only enhance understanding of the physiological functions of protein acylation modifications, but also help to uncover the potential regulatory mechanisms of lysine succinylation.

Adapting to external environmental challenges and maintaining internal metabolic stability are crucial for the survival of cells and this requires a dynamic diversity in its proteome. While Although transcription and translation processes can achieve this, they come at a costly expense of time and energy. However, post-translational modifications (PTMs) can effectively enhance the heterogeneity of protein function and are considered an elegant mechanism for cells to dynamically regulate metabolism [1]. To date, over 600 PTMs have been identified, with approximately half of all proteinogenic amino acids capable of undergoing modification [2]. Notably, lysine acylation modification can rapidly respond to intracellular metabolism, making it a prominent modification conserved across phylogeny [3].

Nε-lysine acetylation is a critical process whereby an acyl group is donated to the epsilon amino group of lysine, and acyl-CoAs are the main intracellular donors of acyl group [4]. Depending on the acyl groups, lysine acylation can be classified into different types, such as acetylation, malonylation, succinylation, and others. Mass spectrometry technology has enabled the detection of thousands of acylation sites in various organisms, which highlights the prominent role that lysine acylation plays in regulating cell growth and metabolic functions [5]. Nevertheless, the comprehension of the intracellular distribution patterns of lysine acylation sites remains limited. Identifying the key factors that impact lysine acylation could provide valuable insights into its physiological functions.

Actinomycetes are the rich source of natural products, including polyketides and terpenes, which are synthesized from acyl-coenzyme A (acyl-CoA) precursors [6]. *Saccharopolyspora erythraea* is one such organism that produces erythromycin, a polyketide compound synthesized from propionyl-CoA and methylmalonyl-CoA [7]. Succinyl-CoA plays a critical role in the tricarboxylic acid (TCA) cycle and is a crucial source of these two acyl-CoAs. Thus, it is considered to be a key node linking primary and secondary metabolic pathways in *S. erythraea*. Notably, succinyl-CoA also functions as an acyl donor in protein succinylation, studying on it might be an efficacious way to reveal the regulatory mechanisms of primary and secondary metabolism in *S. erythraea*.

We conducted a high-resolution mass spectrometry-based proteomic analysis in the engineered strain E3ΔsucC, which has a higher level of intracellular protein succinylation, to investigate the systematic succinylome profiling of *S. erythraea* [8]. Our results identified a total of 5,531 succinylation sites from 1,654 protein sequences, all at a 1% false detection rate (FDR) (supplementary Figure S1, Table S1). On average, each modified protein had 3.34 succinylation sites, and 23.12% (1,654/7,154) of all proteins were found to have the detected succinylation sites. Among the identified modified proteins, 658 proteins had only one succinylation site, accounting for 39.8%, while 90 proteins had more than ten succinylation sites, accounting for 5.4% (Figure 1A, B). In particular, five proteins had over 30 succinylation sites: 2-oxoglutarate dehydrogenase E1 component (36), DNA-directed RNA polymerase subunit beta rpoC (35), chaperone protein dnaK (32),

elongation factor G fusA (32), and 60 kDa chaperonin groL (30). Subcellular localization analysis showed that 63.5% of succinylated proteins were located in the cytoplasm, while 14% were membrane-associated (Figure 1C). Previous studies in other microorganisms such as *Escherichia coli*, *Streptomyces coelicolor*, and *Bacillus subtilis*, resulting in the identification of 2580/670, 673/427, and 2150/634 succinylation sites and succinylated protein, respectively [9-11]. In comparison, our study identified a greater quantity succinylome with broader protein coverage. A high density of succinylation sites within a sample can be crucial for investigating the distribution patterns of these sites and delving into their physiological functions. For instance, machine learning methods were recently employed to predict acylation sites, a more comprehensive acylation site library might sustain the training and validation processes for these methods [12].

To shed light on the function of succinylation in cellular processes, we conducted a GO classification analysis. The classification results (supplementary Figure S2) relating to molecular function, biological process, and cellular component categories showed that the largest protein group of succinyl proteins are associated with catalytic activity, organonitrogen compound biosynthetic processes, and cytoplasm, which accounts for 24, 33, and 34% of the total succinyl proteins, respectively. Protein domains are the structural basis of their physiological functions. Therefore, to further identify the functions associated with succinylation, we annotated the domains of the identified succinylated proteins, and then carried out enrichment analysis (supplementary Table S2). The enriched succinylated proteins, as depicted in supplementary Figure S3, displayed functional domains such as DEAD/DEAH box helicase, histidine phosphatase superfamily (branch 1), cold-shock DNA-binding domain, and helix-turn-helix, which were predominantly associated with nucleotide binding. The KEGG enrichment results (Figure 2) highlighted that the succinylated proteins were mainly engaged in pathways such as ribosome, RNA degradation, citrate cycle (TCA cycle), and aminoacyl-tRNA biosynthesis. Previous researches suggested proteins in these pathways usually contain a higher proportion of basic amino acids (K, R), aiding in acidic substrate binding [13]. Thus, we continually calculated the proportion of lysine residues in proteins within the abovementioned KEGG pathways (supplementary Figure S4). The results revealed that most of them had a higher proportion of lysine residues, with ribosomal proteins exhibiting the highest lysine content at 8.38%, marking a 4.76-fold increase compared to the average level of 1.76%. It implied that lysine succinylation might play a more important regulatory role in biochemical reactions involving acidic substrates.

The biosynthesis of 12-, 14-, and 16-membered macrolides pathway and the polyketide sugar unit biosynthesis pathway, which are components of erythromycin synthesis, were enriched for succinylated proteins (supplementary Figure S5). These results suggested that succinylated proteins may play a direct role in the regulation of erythromycin synthesis. However, 12-, 14- and 16-membered macrolides and polyketide sugar unit biosynthesis pathways exhibited relatively lower lysine contents. Taking into account that protein expression levels may serve as a potential factor affecting acylation probability, we employed transcriptome data from *S. erythraea* E3 under similar growth conditions to gauge the average expression level of proteins within each pathway. Data indicated that these two pathways exhibited a higher degree of expression in *S. erythraea*, potentially bolstering the probability of lysine succinylation for these proteins (supplementary Figure S4, Table S3). In conclusion, we speculated that the proportion of lysine residues in proteins and the intracellular protein concentration may be important factors that affect the distribution of protein acylation.

Furthermore, we analyzed flanking protein sequences (10 amino acids upstream and downstream of succinylated lysine sites) using Motif-X to explore the amino acid preferences adjacent to modified lysine residues (supplementary Table S4). The heat map displayed the significance of amino acids present around modification sites, with a noticeable degree of symmetry in their distribution (Figure 3). Acidic amino acids, Aspartic acid (D) and Glutamic acid (E), appeared more frequently around the modification site, while basic amino acids had a lower frequency nearby but significantly increased further away. Nonpolar amino acids were distributed less prevalently around the modification site, possibly due to differing levels of residue hydrophobicity and distinct positions within the protein.

Previous acylome analysis have identified the common presence of polar amino acids E, K, and R in acylation site motifs, but the underlying reasons have not been clearly described [11; 14]. In this study, motif analysis

showed that charged amino acids (D, E, K, R and W) exhibit a more regular distribution around acylation sites. Based on this result, we speculated that the polar effect between residues may be the key factor influencing lysine succinylation. The negative charge carried by acidic amino acids D and E could interact with the lysine side chain, enhancing its electrophilicity and reactivity, thereby increasing its susceptibility to react with succinyl-CoA. Notably, the longer side chain of E might extend its polar effect on lysine residues, explaining its higher occurrence frequency even at more distant positions. Similarly, the lower occurrence frequency of basic amino acids adjacent to acylation sites could result from a depolarization effect. However, the frequency of basic amino acids appearing a few residues away from the acylation site increases significantly, possibly due to the weakening of the depolarization effect and the more prominent recruitment of succinyl-CoA by basic amino acid residues. The different distribution between K and R might also be attributed to the broader range of depolarization effect caused by the longer side chain of R. Given the significant impact of neighboring amino acids on lysine succinylation, modifying the amino acids surrounding the lysine residue could be a novel strategy for regulating the level of lysine succinylation and enzyme activity.

Otherwise, we explored the lysine succinylation distribution pattern on the TCA cycle protein structures. With NetSurfP 3.0 algorithm, we predicted the secondary structure of succinylated proteins in the TCA cycle, and the results were shown in the supplementary Table S5 and Figure S6 [15]. In terms of α -helix and β -sheet, the probability distribution of modified and non-modified lysine residues appears to be inherently similar. By using the Wilcox test, no significant difference in solvent accessibility was found between modified and unmodified lysine (supplement Figure S7). According to the 3D structure predicted by alphafold, most of the succinylated lysine residues were found to be located on the surface of proteins, corresponding to the high solvent accessibility of lysine residues. For multi-subunit protein complexes, these modifications could affect the binding between subunits and subsequently influence their function. Then, we used POCASO to predict possible active pocket of five single-chain proteins, citrate synthase, isocitrate dehydrogenase, aconitase, fumarate dehydrogenase, and malate dehydrogenase, in TCA cycle [16]. Several succinylated lysine residues were found to be located around the active pockets or active sites, indicating the potential impact on protein activity (supplement Figure S8). This analysis suggested that the function of multiple proteins in the TCA cycle might be regulated by succinylation.

In summary, by employing high-resolution 4D label free mass spectrometry analysis and a highly succinylated strain, we identified a relatively large protein succinylome in *S. erythraea*. Our analysis revealed that succinylation probably tends to occur in biochemical pathways that bind to acidic substrates, as associated proteins have a higher lysine content. Additionally, polar amino acids surrounding acylation sites display a more regular distribution pattern, potentially due to the polar effect between amino acid residues. Lastly, we investigated the distribution of acylation sites in TCA cycle proteins based on predicted structures, highlighting the potential impact of acylation modifications on enzyme activity. These findings may contribute to a comprehensive understanding of lysine succinylation regulation and provide valuable omics data for further analysis and research advancement.

ACKNOWLEDGMENTS

This work was financially supported by the National Key Research Development Program of China (No.2022YFC2105403), the National Natural Science Foundation of China (No. 21276081), the Scientific Research Think Tank of Biological Manufacturing Industry in Qingdao (QDSWZK202004), the Fundamental Research Funds for the Central Universities (JKF01211116, JKVF1211036, JKF01211708).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The MS data have been deposited to the ProteomeXchange Consortium with the dataset identifier PXD042164 (<https://www.iprox.cn>).

REFERENCES

- [1] Macek, B., Forchhammer, K., Hardouin, J., Weber-Ban, E., Grangeasse, C., & Mijakovic, I. (2019). Protein post-translational modifications in bacteria. *Nat Rev Microbiol*, *17* (11), 651-664. doi:10.1038/s41579-019-0243-0
- [2] Xu, H., Wang, Y., Lin, S., Deng, W., Peng, D., Cui, Q., & Xue, Y. (2018). PTMD: A Database of Human Disease-associated Post-translational Modifications. *Genom Proteom Bioinf*, *16* (4), 244-251. doi:10.1016/j.gpb.2018.06.004
- [3] Hu, L. I., Lima, B. P., & Wolfe, A. J. (2010). Bacterial protein acetylation: the dawning of a new age. *Mol Microbiol*, *77* (1), 15-21. doi:10.1111/j.1365-2958.2010.07204.x
- [4] Christensen, D. G., Xie, X., Basisty, N., Byrnes, J., McSweeney, S., Schilling, B., & Wolfe, A. J. (2019). Post-translational Protein Acetylation: An Elegant Mechanism for Bacteria to Dynamically Regulate Metabolic Functions. *Front Microbiol*, *10*, 1604. doi:10.3389/fmicb.2019.01604
- [5] Bradley, D. (2022). The evolution of post-translational modifications. *Curr Opin Genet Dev*, *76*, 101956. doi:10.1016/j.gde.2022.101956
- [6] Barka, E. A., Vatsa, P., Sanchez, L., Gaveau-Vaillant, N., Jacquard, C., Meier-Kolthoff, J. P., . . . van Wezel, G. P. (2016). Taxonomy, Physiology, and Natural Products of Actinobacteria. *Microbiol Mol Biol Rev*, *80* (1), 1-43. doi:10.1128/MMBR.00019-15
- [7] Park, J. W., & Yoon, Y. J. (2019). Recent advances in the discovery and combinatorial biosynthesis of microbial 14-membered macrolides and macrolactones. *J Ind Microbiol Biotechnol*, *46* (3-4), 445-458. doi:10.1007/s10295-018-2095-4
- [8] Ke, X., Jiang, X., Huang, M., Tian, X., & Chu, J. (2022). Engineering of succinyl-CoA metabolism in view of succinylation regulation to improve the erythromycin production. *Appl Microbiol Biotechnol*, *106* (13-16), 5153-5165. doi:10.1007/s00253-022-12060-4
- [9] Colak, G., Xie, Z., Zhu, A. Y., Dai, L., Lu, Z., Zhang, Y., . . . Tan, M. (2013). Identification of lysine succinylation substrates and the succinylation regulatory enzyme CobB in Escherichia coli. *Mol Cell Proteomics*, *12* (12), 3509-3520. doi:10.1074/mcp.M113.031567
- [10] Zhang, H., Li, P., Ren, S., Cheng, Z., Zhao, G., & Zhao, W. (2019). ScCobB2-mediated Lysine Desuccinylation Regulates Protein Biosynthesis and Carbon Metabolism in *Streptomyces coelicolor*. *Mol Cell Proteomics*, *18* (10), 2003-2017. doi:10.1074/mcp.RA118.001298
- [11] Zhang, M., Liu, T., Wang, L., Huang, Y., Fan, R., Ma, K., . . . Xu, J. Y. (2023). Global landscape of lysine acylomes in *Bacillus subtilis*. *J Proteomics*, *271*, 104767. doi:10.1016/j.jprot.2022.104767
- [12] Hasan, M. M., & Kurata, H. (2018). GPSuc: Global Prediction of Generic and Species-specific Succinylation Sites by aggregating multiple sequence features. *PLoS One*, *13* (10), e0200283. doi:10.1371/journal.pone.0200283
- [13] An, H., Ordureau, A., Korner, M., Paulo, J. A., & Harper, J. W. (2020). Systematic quantitative analysis of ribosome inventory during nutrient stress. *Nature*, *583* (7815), 303-309. doi:10.1038/s41586-020-2446-y
- [14] Zhao, Y., Han, Y., Sun, Y., Wei, Z., Chen, J., Niu, X., . . . Gao, X. (2020). Comprehensive Succinylome Profiling Reveals the Pivotal Role of Lysine Succinylation in Energy Metabolism and Quorum Sensing of *Staphylococcus epidermidis*. *Front Microbiol*, *11*, 632367. doi:10.3389/fmicb.2020.632367
- [15] Hoie, M. H., Kiehl, E. N., Petersen, B., Nielsen, M., Winther, O., Nielsen, H., . . . Marcatili, P. (2022). NetSurfP-3.0: accurate and fast prediction of protein structural features by protein language models and deep learning. *Nucleic Acids Res*, *50* (W1), W510-W515. doi:10.1093/nar/gkac439

[16] Yu, J., Zhou, Y., Tanaka, I., & Yao, M. (2010). Roll: a new algorithm for the detection of protein pockets and cavities with a rolling probe sphere. *Bioinformatics*, 26 (1), 46-52. doi:10.1093/bioinformatics/btp599

Figure legends

Figure 1 Profile of *S. erythraea* lysine succinylation proteome. (A) The number of proteins with the various numbers of succinylated sites. (B) Pie charts show the distribution of the number of lysine acylated sites per protein. (C) Classification of succinylated proteins on the basis of subcellular location.

Figure 2 KEGG enrichment analysis of the acetylated proteins in *S. erythraea*. A two-tailed Fisher's exact test was used, p-value < 0.05 were considered significant.

Figure 3 A heat map of amino acid compositions surrounding the succinylated lysine. The motifs consisting of 20 residues (10 amino acids upstream and downstream of the site) surrounding the succinylated lysine are shown.



