

Review

Not peer-reviewed version

Advances in Decoding Bacterial N-Terminal Proteoforms: Technologies, Challenges, and Functional Insights

Valdes Shauwaert and [Petra Van Damme](#)*

Posted Date: 5 June 2026

doi: 10.20944/preprints202606.0506.v1

Keywords: alternative translation initiation; genome engineering; multiplexed recombineering; N-terminal proteoforms; N-terminomics; riboproteogenomics



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC, OpenAlex.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Review

Advances in Decoding Bacterial N-Terminal Proteoforms: Technologies, Challenges, and Functional Insights

Valdes Snauwaert and Petra Van Damme *

iRIP Unit, Laboratory of Microbiology, Department of Biochemistry and Microbiology, Ghent University, Gent, Belgium

* Correspondence: petra.vandamme@ugent.be

Abstract

The bacterial proteome is a highly dynamic landscape rather than a static reflection of the genome. Recent research revealed that proteome complexity extends far beyond canonical gene annotation, with N-terminal (Nt-)proteoforms emerging as an important underexplored additional regulatory layer. These molecular variants originate from a single genetic locus through alternative translation initiation at internal or external in-frame start sites, thereby generating N-terminal heterogeneity that can influence protein stability, subcellular localization, interaction networks, and the stoichiometric assembly of multiprotein complexes. While recent advances in riboproteogenomics, N-terminomics, and computational annotation strategies have enabled proteoform mapping at single-amino acid resolution, rapid high-throughput discovery currently outpaces downstream functional characterization. This review discusses the technological advances driving Nt-proteoform discovery, including emerging ribosome profiling and proteogenomic approaches, and further evaluates strategies for the functional characterization of Nt-proteoform. Particular emphasis is placed on the transition from conventional plasmid-based heterologous expression systems toward precise genome-engineering approaches that enable selective manipulation of alternative translation initiation events within their native genomic context. Such targeted strategies are essential to bridge the gap between Nt-proteoform identification and functional understanding, ultimately uncovering how individual bacterial genomic loci can encode proteoforms with distinct and potentially polarized roles in bacterial physiology and pathogenesis.

Keywords: alternative translation initiation; genome engineering; multiplexed recombineering; N-terminal proteoforms; N-terminomics; riboproteogenomics

1. N-Terminal Proteoforms Expand Bacterial Proteome Complexity

The central dogma of molecular biology, originally proposed by Francis Crick, describes a unidirectional flow of genetic information from DNA through RNA to functional proteins [1]. However, the discovery of alternative splicing by Sharp and Roberts demonstrated that eukaryotic gene sequences are often non-contiguous, thereby challenging the presumed linear relationship between genes and their protein products. More recently, the identification of defense-associated reverse transcriptases (DRTs) capable of synthesizing alternating poly(GT/AC) double-stranded DNA (dsDNA) revealed a protein-templated mechanism for sequence-specific DNA synthesis, further expanding the conceptual boundaries of information flow in biology [2].

In bacteria, proteome complexity is substantially amplified by alternative translation initiation events and a broad repertoire of co- and post-translational modifications, collectively generating a diverse array of proteoforms (Figure 1). Proteoforms encompass all distinct protein variants derived from a single genetic locus [3].

Among bacterial protein modifications, one of the most prevalent is the co-translational removal of the N-terminal formyl group from the initiator methionine (iMet) by peptide deformylase (PDF), which associates with the ribosomal exit tunnel during translation elongation [4–6]. This processing event affects more than 94% of proteins in *Escherichia coli* (*E. coli*) [7] and exposes a free N-terminal amine group that permits further N-terminal maturation. Subsequently, methionine aminopeptidase (MetAP) may remove the iMet depending on the identity of the penultimate acid residue, a process estimated to affect more than half of mature bacterial N-termini [7–9].

In addition, protein N-termini may undergo (partial) Nt-acetylation mediated by N-terminal acetyltransferases (NATs), a process that occurs post-translationally in bacteria [10]. NATs can irreversibly transfer an acetyl group from acetyl-coenzyme A (acetyl-CoA) to the α -amino group of either the deformylated initiator methionine or the newly exposed N-terminal residue following iMet excision. Although bacterial N-terminal acetylation is generally observed at substoichiometric levels, up to 10–15% of bacterial proteins are estimated to contain detectable N-terminal acetylation [11–15]. Other protein modifications, including phosphorylation, lipidation, glycosylation, and pupylation, further contribute to bacterial proteome diversification (Figure 1A). Proteoform diversity is additionally expanded through coordinated proteolytic processing events, most notably the cleavage of signal peptides by dedicated signal peptidases [16]. Prior to removal, these N-terminal sequences serve essential regulatory functions by directing subcellular localization and mediating interactions with molecular chaperones that maintain proteins in a translocation-competent state. Subsequent site-specific proteolytic processing ultimately yields the mature and localized proteoform (Figure 1B) [17].

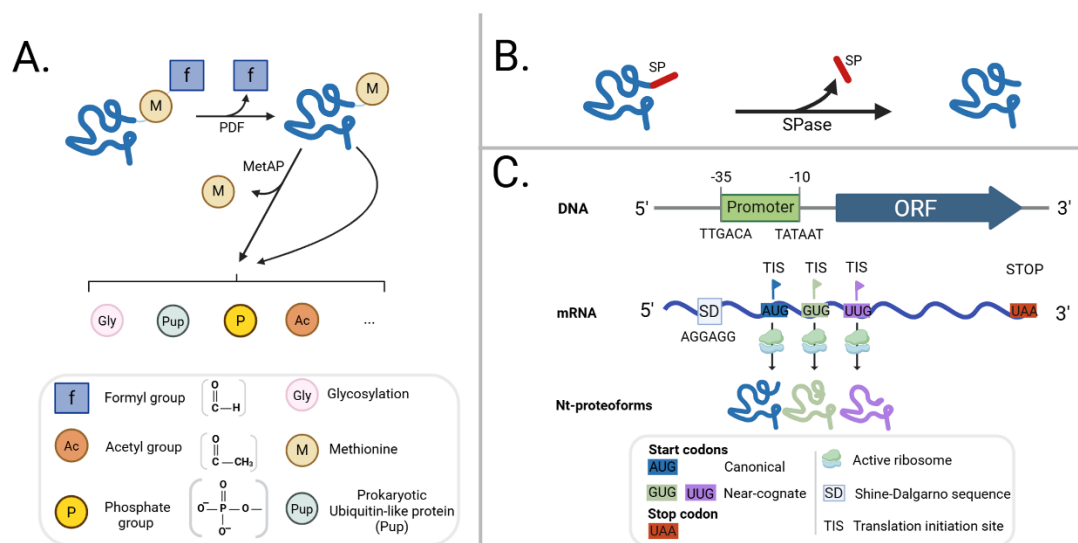


Figure 1. Mechanisms contributing to bacterial N-terminal (Nt-) proteoform diversity. Distinct Nt-proteoforms can arise through multiple co- and post-translational mechanisms that collectively expand bacterial proteome complexity. **(A) N-terminal processing and protein modifications:** Nascent bacterial proteins initially carry a formylmethionine residue. The Nt-formyl group (f) can be removed co-translationally by peptide deformylase (PDF), after which the initiator methionine (M) may either be retained or cleaved by methionine aminopeptidase (MetAP). Mature proteins may subsequently undergo additional post-translational modifications (PTMs), including phosphorylation, pupylation (addition of a prokaryotic ubiquitin-like protein), glycosylation, and other modifications, thereby further contributing to N-terminal heterogeneity. **(B) Proteolytic processing:** Cleavage of N-terminal signal peptides (SPs) by dedicated signal peptidases (SPases) generates mature processed proteoforms with altered N-termini following protein translocation. **(C) Alternative translation initiation:** Multiple proteoforms can originate from a single gene through the use of alternative translation initiation sites (aTISs) within the same transcript. Translation may initiate from the canonical AUG start codon or from near-cognate initiation codons such as GUG and UUG. Promoter architecture, including the -10 and -35 promoter elements, as well as the Shine–Dalgarno (SD) sequence involved in ribosome recruitment, is indicated.

Utilization of distinct TISs results in Nt-proteoforms differing in N-terminal composition and length. Figure created using BioRender.

Among these protein variants, distinct molecular forms originating from the same gene locus represent a major source of bacterial protein diversity. Beyond the previously discussed N-terminal maturation and processing events, Nt-proteoforms may additionally arise through the selection of alternative translation initiation sites (aTISs) distinct from the database-annotated or canonical translation initiation site (dbTISs) (Figure 1C). Translation from internal or external in-frame and out-of-frame start sites significantly expands the bacterial proteome and can yield Nt-proteoforms that differ in their N-terminal composition and, consequently, their biological properties [18,19]. These alternative initiation events generate functionally distinct proteins that contribute to a more complex and versatile 'alternative' proteome than previously recognized.

Different protein variants originating from a single gene locus can arise through a combination of transcriptional and translational mechanisms. At the transcriptional level, the use of alternative transcription start sites (TSSs) can directly influence the selection and usage of aTISs. For instance, TSSs located within annotated coding regions may generate shorter transcripts that alter the set of translation initiation sites accessible to the ribosome. Accordingly, the observation that some experimentally mapped TSS reside precisely between dbTISs and aTISs further illustrates how transcriptional boundaries can shape alternative translation initiation patterns [19].

Alongside transcription, multiple translational features further influence TIS selection. Due to bacterial transcription-translation coupling, local mRNA secondary structures or riboswitches may promote or inhibit translation initiation at specific TISs by masking or exposing ribosome binding sites (RBSs) [20–26]. In addition, the presence and strength of the Shine-Dalgarno (SD) sequences, together with their spacing relative to the initiating codon, strongly affect translation initiation efficiency. However, the presence of an SD sequence followed by an AUG triplet does not necessarily guarantee translation initiation [27], and because some leaderless mRNAs can be efficiently translated in the absence of an SD sequence [28], TIS selection remains highly dynamic and context-dependent.

Although AUG is the predominant start codon in *E. coli*, accounting for approximately 83% of annotated genes, near-cognate initiating codons such as GUG (14%) and UUG (3%) are also frequently utilized [29,30]. Similar patterns of start codon frequency have been observed across the bacterial domain [31]. Beyond these three major initiation codons, several additional non-AUG codons have been reported to support translation initiation [32]. Importantly, our previous work demonstrated that near-cognate start codons account for approximately one-quarter of newly identified Nt-proteoforms in *Salmonella enterica* serovar Typhimurium (*S. Typhimurium*) [33].

Collectively, these mechanisms demonstrate that the bacterial translome is not merely a static reflection of genome annotation, but rather a highly dynamic and adaptable landscape. In particular, the widespread occurrence of alternative translation initiation suggests that bacterial coding capacity extends substantially beyond canonical gene models. By generating multiple Nt-proteoforms from single genetic loci, bacteria can expand their functional repertoire without increasing genome size. Consequently, N-terminal heterogeneity constitutes an important and still underexplored regulatory layer that can influence protein stability, subcellular localization, interaction networks, enzymatic activity, and multiprotein complex assembly, among other biological processes.

2. Biological Relevance of Nt-Proteoforms

Heterogeneity at the N-terminus of bacterial proteins represents an important regulatory layer that may influence subcellular localization, multiprotein complex assembly, protein stability, and ultimately protein function (Figure 2).

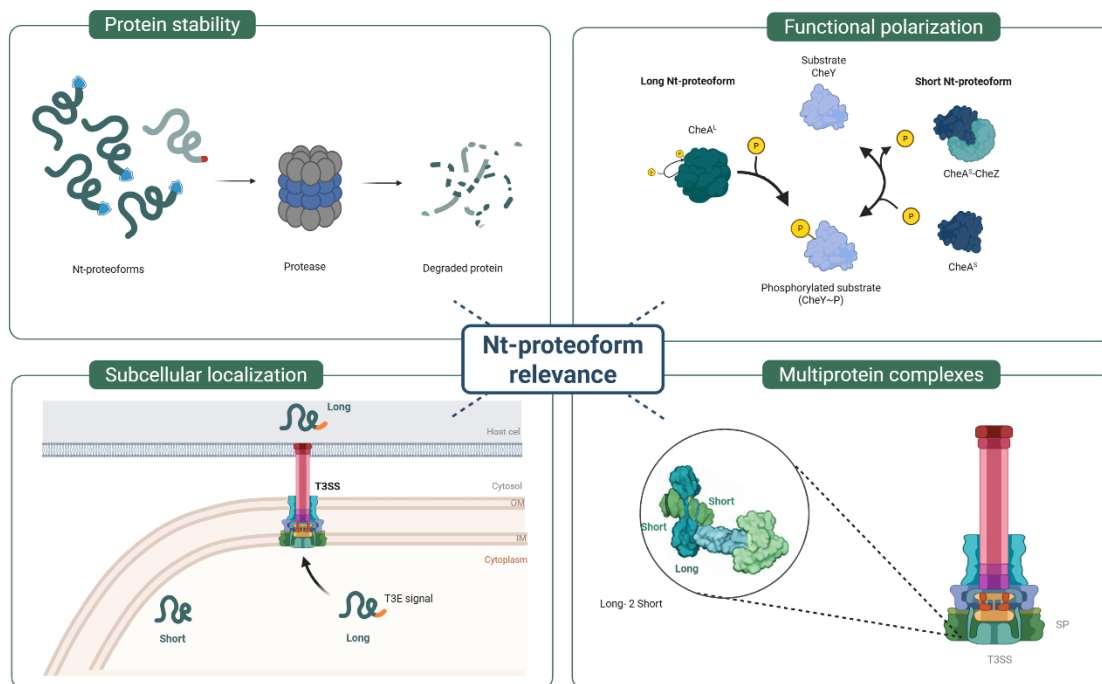


Figure 2. Biological relevance and functional implications of bacterial N-terminal (Nt-)proteoforms. The bacterial transcriptome constitutes a highly dynamic landscape in which N-terminal heterogeneity introduces multiple regulatory layers that influence protein fate and function. **Protein stability:** Variations at the N-terminus can modulate protein half-life through the bacterial N-degron pathway, in which specific N-terminal residues determine recognition by the ClpS-ClpAP proteolytic machinery. **Functional polarization:** Distinct N-termini can alter biochemical activity and *signaling* behaviour. The long CheA Nt-proteiform (CheA^L) contains an N-terminal autophosphorylation site that is absent in the shorter CheA^S proteoform. Consequently, CheA^L primarily functions as a kinase that phosphorylates CheY, whereas CheA^S interacts with CheZ to promote CheY dephosphorylation. **Subcellular localization:** N-terminal extensions can encode transport or secretion signals, such as N-terminal type III effector (T3E) secretion signals, thereby enabling selective translocation of specific proteoforms (e.g., SseL^L) into the host cell, while truncated proteoforms lacking these signals (e.g., SseL^S) remain confined to the bacterial cytoplasm. **Multiprotein complex assembly:** Nt-proteiform diversity can further regulate the stoichiometric assembly and functionality of higher-order molecular complexes, exemplified by the T3SS sorting platform (SP), where defined ratios between long and short isoforms (e.g., one SpaO^L and two SpaO^S subunits) are required for proper injectisome assembly and activity. Figure created using BioRender.

N-terminal variation fundamentally dictates spatial organization, as illustrated by the architecture of cyanobacterial β -carboxysomes. Within these proteinaceous microcompartments, ribulose-1,5-biphosphate carboxylase/oxygenase (Rubisco), the key enzyme of the Calvin cycle, is sequestered to optimize CO₂ fixation efficiency [34]. In *Synechococcus* sp. strain PCC 7942, an internal GUG₂₁₆ initiation codon within *cmM* generates two distinct isoforms, forming an N-terminal proteoform pair consisting of the full-length CcmM58 (58 kDa) and the N-terminally truncated CcmM35 (35 kDa). This 23 kDa difference fundamentally dictates their cellular functions. CcmM58, retaining its complete N-terminal region, is localized to the inner shell of the carboxysome, where it recruits the carboxysomal carbonic anhydrase CcaA and interconnects outer shell components [35–37]. Conversely, the truncated CcmM35 proteoform is confined to the carboxysome lumen and functions exclusively in organizing Rubisco into a paracrystalline matrix. This example highlights how N-terminal diversity can physically segregate structural and scaffolding functions within a single microcompartment to ensure efficient CO₂ fixation.

Beyond mediating spatial organization within intracellular bacterial microcompartments, Nt-proteoforms can additionally control protein targeting and secretion in bacterial virulence systems.

In bacterial pathogens such as *S. Typhimurium*, more than 40 type III effector (T3E) proteins are secreted through the type III secretion system (T3SS) to manipulate host cell biology and promote bacterial survival, replication, and dissemination [38–40]. The effector SseL exists as two proteoforms: a long isoform (SseL^L, 38 kDa) and a shorter isoform lacking 23 N-terminal residues (SseL^S, 29 kDa) [33]. The N-terminal extension unique to SseL^L contains a T3E secretion signal, thereby enabling translocation into the host cell environment. In contrast, SseL^S remains confined to the bacterial cytosol and may therefore fulfill distinct intracellular functions. In this context, it is important to note that T3Es were long assumed to remain inactive within the bacterium and only adopt their functional conformations following host-cell translocation, owing to chaperone-mediated maintenance of a secretion-competent unfolded state. However, recent studies have demonstrated that certain T3Es can exert important intrabacterial activities involved in the regulation of bacterial physiology [41,42]. This heteromorphic phenomenon is further corroborated by CyaA' (calmodulin-dependent adenylate cyclase) translocation assays demonstrating that the SseL leader sequence constitutes the minimal determinant required for host-cell entry [43]. Accordingly, several additional candidate Nt-proteoform pairs, including SsaQ, YdgA, YfhG, MgtC, OmpX, and PagC, are predicted to exhibit distinct subcellular localization among their constituting Nt-proteoform members in *Salmonella* [33].

Beyond influencing effector localization and host-cell targeting, Nt-proteoforms can additionally modulate the assembly and stoichiometry of larger multiprotein machineries such as the T3SS sorting platform. The T3SS injectisome itself is composed of a cytosolic sorting complex, multiple membrane-embedded structures, a hollow needle spanning the bacterial membranes, and a translocon complex located at the tip (extensively reviewed in [44] and [45]). Central to this machinery is the cytosolic sorting platform, a multiprotein complex positioned at the base of the injectisome that coordinates effector sorting and unfolding prior to secretion. Previous studies demonstrated that SpaO^S (11 kDa) forms homodimers that subsequently associate with a single SpaO^L (34 kDa) subunit to generate a heterotrimeric complex with a strict 2:1 stoichiometry (2 SpaO^S: 1 SpaO^L). Within this assembly, SpaO^L acts as a central interaction hub that promotes SP pod formation [46,47]. Although deletion of SpaO^S partially impairs T3SS functionality, secretion activity is not completely abolished, suggesting an ancillary yet functionally important role in T3SS-mediated protein translocation [47,48]. Consistent with this observation, deletion of the homologues Spa33^S (12 kDa) (proteoform in *Shigella*) resulted in residual, albeit detectable, T3SS activity. Conversely, conventional *in vitro* studies in both *Yersinia* [49,50] and *Shigella* [51] indicate that the short SpaO^S isoform is indispensable for proper T3SS function.

In addition to controlling protein localization and multiprotein complex assembly, Nt-proteoforms may also directly modulate biochemical activity and signalling behaviour, as exemplified by the chemotaxis regulator CheA. This histidine kinase functions as a central regulator of bacterial chemotaxis. Upon binding of chemical stimuli to chemoreceptors, CheA undergoes autophosphorylation, thereby initiating a signal transduction cascade that governs flagellar rotation, receptor adaptation, and methylation [52,53]. In *E. coli*, an internal AUG₉₈ initiation site generates the shorter CheA^S (60 kDa) isoform alongside the full-length CheA^L (71 kDa) isoform. The N-terminal difference between these variants dictates both their complex assembly and biochemical activities. CheA^L forms both homodimers and heterodimers with CheA^S and functions exclusively as a kinase responsible for phosphorylating the response regulator CheY [54]. In contrast, CheA^S lacks the critical His₄₈ residue and instead fulfills distinct regulatory functions. Specifically, CheA^S localizes as small foci at the cell poles and participates in both phosphorylating and dephosphorylating complexes [55,56]. Furthermore, CheA^S interacts with and enhances the activity of CheZ-mediated dephosphorylation of phospho-CheY, while CheA^L lacks this capability [56,57]. This example illustrates how N-terminal heterogeneity enables a single genetic locus to generate proteins with functionally polarized roles.

Besides influencing localization, multiprotein complex assembly, and biochemical specialization, Nt-proteoforms may also profoundly affect protein stability. Importantly, not all Nt-proteoforms differ by extensive N-terminal truncations or extensions, as many identified proteoform

architectures vary by only a few amino acid residues [19,33]. Nevertheless, even these subtle N-terminal differences may drastically influence protein stability through the N-degron pathway (formerly the N-end rule pathway), in which the identity of the N-terminal residue acts as a primary determinant of proteolytic fate [58–60]. In bacteria, this pathway is primarily mediated by the ATP-dependent ClpAP and its adaptor ClpS, a 12-kDa Leu/N-recogin that captures substrates and delivers them to the protease machinery [61]. ClpS preferentially recognizes primary destabilizing residues such as Leu, Phe, Trp, and Tyr, whereas proteins carrying stabilizing N-terminal residues (Met, Ser, Thr, Ala, Val, and Gly) generally exhibit prolonged half-lives. Although bacterial N-degron pathways firmly establish the importance of N-terminal identity in determining protein stability, endogenous Nt-proteoform architectures differing only minimally at their N-termini have rarely been directly linked to divergent degradation kinetics. However, analogous functional consequences of subtle N-terminal variation are well established in eukaryotes, where human Nt-proteoforms differing by only a few residues can display markedly distinct stability profiles [62].

3. Mapping the Bacterial Translatome: Genome Annotation, Ribosome Profiling, and N-Terminomics

The increasing prevalence of Nt-proteoforms within the bacterial translatome reveals a regulatory layer that is frequently overlooked by traditional genome annotation approaches. Accurate delineation of this proteomic landscape increasingly relies on integrated frameworks combining computational prediction with high-resolution experimental data (Figure 3). Three complementary pillars currently underpin this strategy: computational modeling (Figure 3C), ranging from classical *ab initio* gene prediction algorithms to modern genomic language models (gLMs); ribosome profiling (Figure 3A), which identifies actively translated regions and TISs; and N-terminomics (Figure 3B), which provides direct mass spectrometry (MS)-based evidence of protein start sites. Together, these complementary approaches are shifting bacterial genome annotation from static gene predictions toward dynamic translatome- and proteome-aware annotation strategies [63].

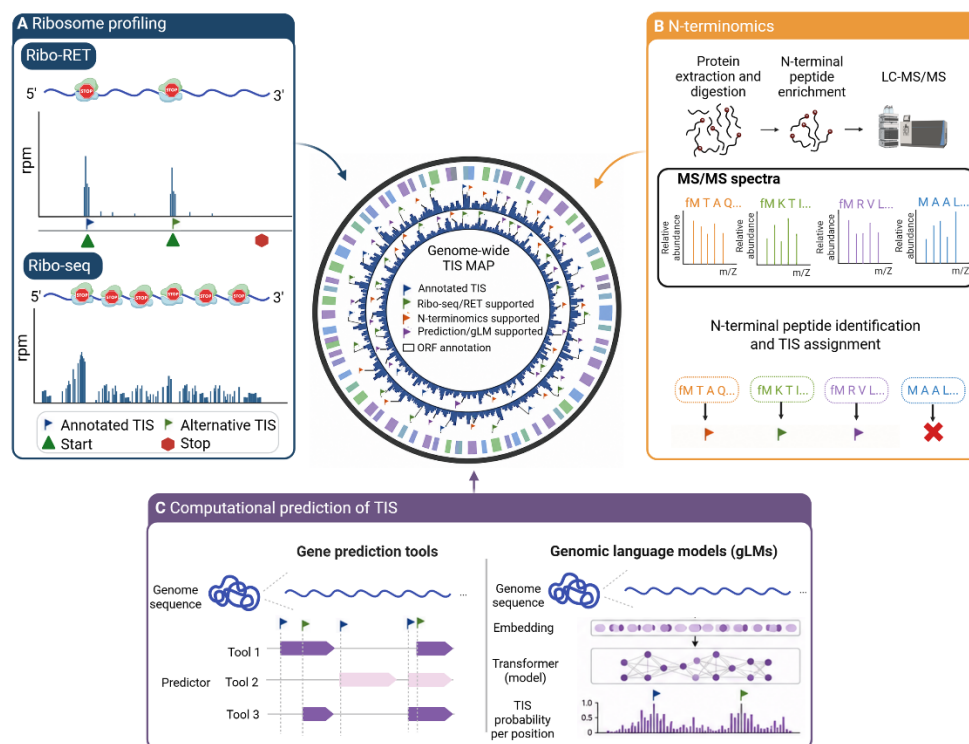


Figure 3. Integrated riboproteogenomic framework for high-resolution translation initiation site (TIS) mapping and bacterial genome annotation. Complementary experimental and computational approaches collectively

enable the identification and validation of annotated and alternative translation initiation sites (aTISs) within bacterial genomes. The central circular overview integrates genome-wide TIS mapping results supported by ribosome profiling, N-terminomics, and computational prediction approaches. Annotated TISs, experimentally supported aTISs, and computationally predicted TISs are represented alongside the corresponding open reading frame (ORF) architecture. (A) Ribosome profiling. Conventional Ribo-seq captures actively translated genomic elements through ribosome-protected mRNA fragments. Specialized approaches such as retapamulin-assisted ribosome (Ribo-RET) enrich initiating ribosomes through antibiotic-mediated stalling at start codons, thereby improving the detection of canonical and aTISs. (B) N-terminomics. N-terminal proteomics workflows combine protein extraction, proteolytic digestion, selective enrichment of N-terminal peptides, and LC-MS/MS analysis to identify proteoform-specific N-termini. The presence of formylated methionine (fM) serves as a direct proxy for an actual initiator methionine. Consequently, detection of fM-containing N-terminal peptides provides direct molecular evidence for translation initiation events and enables TIS delineation at single-amino acid resolution. (C) Computational TIS prediction. Computational approaches ranging from classical ab initio gene finders (e.g., GeneMark, Prodigal, and GLIMMER) to modern genomic language models (gLMs) are used to predict coding sequences (CDSs) and TISs. Figure created using BioRender.

3.1. Classical Genome Annotation Approaches and Their Limitations

Historically, the existence of Nt-proteoforms became apparent through observations that single gene loci could produce multiple protein variants differing at their N-termini. However, accurately defining the genomic boundaries and translation initiation events underlying this heterogeneity remained technically challenging. At the turn of the millennium, the first generation of microbial genome annotation tools emerged, most notably GLIMMER [64] and GeneMark [65]. These algorithms used probabilistic sequence models to identify coding regions within microbial genomes and represented major advances in automated gene prediction. However, early annotation tools were often computationally intensive and displayed limited accuracy in delineating complex genomic architectures, particularly TISs.

Methodological development subsequently focused on improving the precision of TIS identification, most notably through the development of the PROkaryotic Dynamic programming Gene-finding Algorithm (Prodigal) [66]. Despite their widespread use, substantial discrepancies persist among annotation platforms. Comparative analysis of GLIMMER, GeneMark, and Prodigal revealed only 70% consensus between predictions, highlighting the inherent algorithmic biases and limitations within individual tools [67]. To meet increasing high-throughput demands, automated annotation pipelines such as the TIGR Annotation Engine [68] and the Bacterial Annotation System (BASys) [69] were subsequently developed to streamline these workflows. This evolution ultimately culminated in fully automated annotation services for archaeal and bacterial genomes, including Rapid Annotations using Subsystems Technology (RAST) [70], which was later integrated into Bacterial and Viral Bioinformatics Resource Center (BV-BRC) [71]. Despite the comparatively simple gene architecture of prokaryotes relative to eukaryotes, modern annotation tools still struggle to resolve internal genomic complexity. Most annotation pipelines fail to detect multiple internal TISs (iTISs) within a single coding sequence (CDS). Widely used annotation frameworks such as the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) rely heavily on homology-based methodologies, integrating Hidden Markov Models (HMMs), BlastRules, and curated Conserved Domain Database (CDD) architectures [72,73]. Currently, approximately 79% of the RefSeq proteins are annotated based on matches to a curated protein family model (PFM) [74]. Importantly, these tools frequently prioritize the longest possible ORF to maximize coding information, although this does not necessarily reflect the true bacterial translome [63].

Although current ab initio gene prediction algorithms robustly identify canonical ORFs, their sensitivity declines substantially for non-canonical genomic elements. First, many annotation tools exclude sequences shorter than 150-300 nucleotides, thereby systematically overlooking small open reading frames (sORFs) [75–77]. Increasing evidence demonstrates that these small proteins participate in critical biological processes, including protein recruitment [78,79], protein stability

[80,81], and multiprotein complex assembly [82–84], among others. Second, current annotation frameworks generally fail to identify aTISs that give rise to Nt-proteoform. Homology-driven pipelines such as PGAP frequently enforce previously annotated start sites across related genomes, thereby masking biologically relevant variation in translation initiation [73]. Finally, annotation tools prioritizing computational speed, including RAST and Prokka, often perform poorly when annotating specialized or highly divergent genes lacking close homologs.

3.2. Machine Learning and Genomics Language Models

Recent advances in machine learning have substantially reshaped genome annotation strategies. In particular, gLMs such as DNABERT and Evo 2 interpret DNA as structured sequential information, enabling the extraction of both local sequence motifs and long-range contextual relationships [85–87]. Compared with traditional alignment- or homology-based approaches, these models offer improved adaptability for resolving complex bacterial gene architectures, offering a more adaptive alternative to traditional genome annotation.

Various deep learning-based frameworks have been proposed for gene prediction, including conventional neural networks [88]. More recently, GeneLM, a DNABERT-based architecture, was applied to prokaryotic gene prediction using a two-stage approach involving initial CDS identification followed by precise TIS delineation across diverse bacterial species [89]. Benchmarking analyses demonstrated that GeneLM consistently outperformed classical tools like Prodigal, GeneMark, and Glimmer, as well as specialized deep learning approaches including TITER [90] and DeepGSR [91], both of which exhibit comparatively lower TIS prediction performance.

Despite these advances, the performance of gLMs remains highly dependent on the quality and diversity of their training datasets. Nevertheless, gLMs are expected to provide increasingly accurate frameworks for identifying novel gene structures and refining gene boundaries in both well-characterized and previously unannotated genomes.

3.3. Ribosome Profiling-Based Translation Initiation Site Mapping

The limitations of purely computational annotation strategies underscore the importance of experimental validation and evidence-based annotation approaches. Proteogenomic (re-)annotation efforts have therefore been widely applied to identify novel genes and to correct missing or inaccurate annotations [92–95]. In particular, the rapid evolution of riboproteogenomics has largely been driven by the advent of ribosome profiling (Ribo-seq). General Ribo-seq approaches infer translation by capturing ribosome footprints across the transcriptome, thereby substantially improving ORF detection and Nt-proteoform discovery. One of the first bacterial genome annotation approaches integrating Ribo-seq-derived translational evidence was RibosomE Profiling Assisted (Re-)AnnotATIOn (REPARATIOn), which enabled *de novo* ORF delineation in prokaryotic genomes [96]. Consequently, extensive re-annotation efforts have been performed in so-called well-annotated bacteria, including *E. coli*, *S. Typhimurium*, and *Bacillus subtilis* [75,96,97].

To move beyond general ORF delineation and improve TIS resolution, several specialized Ribo-seq approaches have been developed to selectively enrich initiating ribosomes. One of the earliest approaches, tetracycline-inhibited ribosome profiling (TetRP), was introduced as a relatively simple and comprehensive strategy for bacterial TIS annotation [98]. Tetracycline blocks aminoacyl-tRNA entry into the ribosomal A-site and is therefore classically regarded as an elongation inhibitor. To further improve TIS resolution, Meydan et al. subsequently developed retapamulin-assisted ribosome profiling (Ribo-RET), which selectively enriches initiating ribosomes through retapamulin-mediated stalling. Using this approach, internal start codons were identified in more than one hundred *E. coli* genes [19]. Application of Ribo-RET in *S. Typhimurium* further revealed 26 candidate Nt-proteoform pairs following manual curation [33], demonstrating that in both *E. coli* and *S. Typhimurium*, the majority of identified aTISs cluster near the beginning of the CDSs.

Despite these methodological advances, precise TIS delineation remains highly challenging due to several technical and biological limitations. One major limitation stems from incomplete

elongation inhibition during antibiotic treatment. Because tetracycline can interact with ribosomes throughout multiple elongation cycles, TetRP cannot reliably discriminate between elongating and initiating ribosomes, thereby limiting its sensitivity for detecting aTISs [98]. Retapamulin-based approaches face a related constraint, as signals originating from elongating ribosomes are not completely eliminated. This residual background complicates confident identification of internal start codons, particularly within highly expressed ORFs [99]. Furthermore, Ribo-RET peaks at putative initiation sites do not inherently demonstrate productive, full-length translation in the absence of antibiotic treatment, meaning that some detected events may represent abortive or non-functional initiation [100].

In addition to antibiotic-related artifacts, ribosome profiling approaches also suffer from intrinsic resolution constraints because translation initiation is inferred indirectly through ribosome footprints. Since translating ribosomes physically protect RNA fragments of approximately 22-35 nucleotides, overlapping footprints complicate precise discrimination between closely positioned TISs [101].

Moreover, the reliance of ribosome profiling on living cells can complicate the distinction between direct and indirect translational effects. Translation-targeting antibiotics and their associated uptake mechanisms may inadvertently induce generalized stress responses and widespread off-target transcriptional and translational changes that obscure native cellular physiology [102,103].

To minimize such confounding *in vivo* artifacts, *in vitro* Ribo-seq (INRI-seq) was developed as a cell-free alternative that enables translation analysis on customizable synthetic transcriptomes [104]. Although INRI-seq successfully confirmed the annotated TIS for approximately 70% of *E. coli* genes and validated 51 out of the 64 aTISs previously identified by *in vivo* Ribo-RET [19], the method remains constrained by its current library design, which captures only the first 50 codons of a CDS. Consequently, alternative TIS detection is inherently restricted to the first ~150 nucleotides of a gene, thereby limiting the identification of more downstream internal initiation events.

3.4. N-terminomics Approaches for Nt-Proteoform Discovery

Together with ribosome profiling, N-terminomics has become an important tool for high-resolution identification of prokaryotic translation initiation sites (TISs). In conventional bottom-up proteomics workflows, alternative Nt-proteoforms are often difficult to distinguish from their canonical counterparts because most internal tryptic peptides are shared. Proteoform-specific information is mainly confined to the altered N-terminal peptide and, in some cases, to the absence of peptides from the truncated N-terminal region, making dedicated N-terminal proteomics (N-terminomics) strategies essential for direct TIS identification [105,106].

Techniques such as Combined FRActional Diagonal Chromatography (COFRADIC) [107,108], Terminal Amine Isotopic Labeling (TAILS) [109], and LysN Amino Terminal Enrichment (LATE) [110] employ selective chemical derivatization and chromatographic enrichment to isolate N-terminal peptides prior to LC-MS/MS analysis. By reducing sample complexity and enriching proteoform-specific N-terminal peptides, these approaches facilitate direct detection of both canonical and alternative TISs. Comprehensive overviews of bacterial N-terminomics strategies have previously been provided by Berry et al. [106]. Recently, a deformylation-assisted N-terminomics workflow termed TRANslation Initiation SPOTTER (TRAINSPOTTER) was developed [111]. By exploiting the presence of the nascent N-formyl group—optionally stabilized through peptide deformylase (PDF) inhibition—TRAINSPOTTER enables proteome-wide detection of nascent N-termini and thereby provides direct molecular evidence for translation initiation at single-amino acid resolution. Nevertheless, the efficacy of MS-based TIS delineation remains highly dependent on both the completeness of the underlying protein sequence databases and the applied database search strategy. The composition of the proteomic search space fundamentally defines the boundaries of novel proteoform discovery. Historically, unannotated TISs, including aTISs and initiation events originating from non-canonical start codons, frequently escaped MS-based detection because they were absent from standard reference databases. However, this limitation is not absolute and becomes

particularly problematic when stringent enzymatic cleavage constraints are imposed during database searching or when non-AUG initiation events are excluded from the databases.

To circumvent these limitations, the TRAINSPOTTER workflow interrogates MS datasets against a customized six-frame translation (6-FT) database combined with semi-specific search parameters that account for iMet processing rules. This relatively unconstrained search strategy enables identification of internal and alternative N-termini, thereby providing direct physical evidence for productive translation initiation from both canonical and non-canonical start codons.

Despite these advances, current N-terminomics workflows remain limited in their ability to capture TIS-indicative N-termini of proteins undergoing signal peptide cleavage. In such cases, the original N-terminus is removed during protein translocation, effectively erasing the molecular evidence of the initial TIS and complicating comprehensive mapping of the bacterial N-terminal proteome landscape.

4. Methodological Strategies for Decoupling Bacterial N-Terminal Proteoforms

While the growing number of identified aTISs reveals a highly modular bacterial proteome, the capacity for high-throughput discovery currently exceeds downstream functional validation. Consequently, substantial gaps remain in our understanding of the precise physiological roles of individual Nt-proteoforms.

4.1. Classical Heterologous Complementation Approaches

Initial functional characterization and decoupling of bacterial Nt-proteoforms—defined as the selective expression and analysis of individual proteoforms—primarily relied on deletion of the endogenous gene followed by heterologous complementation. Within this framework, individual proteoforms are typically expressed from (inducible) plasmids to evaluate their ability to complement a mutant phenotype. This strategy provided some of the earliest functional insights into the isoforms of Initiation Factor 2 (IF2), encoded by *infB*. Expression of IF2 α and IF2 β from pEV1 and pBAD24 vectors enabled characterization of their distinct ribosomal binding properties and respective contributions to replication restart [112,113]. Similarly, plasmid-based complementation experiments demonstrated that a 46-amino acid shorter proteoform of the major penicillin-binding protein PBP-1B, encoded by *mrcB*, could restore thermosensitive growth defects in *E. coli* [114]. Another classical example is represented by the ClpB proteoform pair, originally identified due to pronounced molecular weight differences and subsequently decoupled using pBS or pGB2-based expression systems [115–117].

Despite their widespread utility in gain-of-function studies, heterologous expression systems frequently fail to recapitulate the native biological context. Plasmid-based protein overexpression often drives protein concentrations far beyond physiological levels, potentially inducing molecular crowding, aberrant subcellular localization, inclusion body (IB) formation, or artificial protein-protein interactions (PPIs). For example, overexpression of periplasmic proteins may saturate Sec-translocon capacity, thereby impairing the secretion of endogenous proteins and promoting cytoplasmic protein aggregation [118]. Similarly, excessive production of aggregation-prone proteins can promote artificial non-physiological PPIs or sequester essential factors into toxic aggregates [119]. Such artifacts are particularly disruptive for multiprotein complexes and signaling pathways, where strict stoichiometric balance is essential for proper function. Furthermore, proteoform expression is often tightly coupled to specific growth phases or environmental stimuli [115,120–124], whereas inducible plasmid systems frequently bypass the native transcriptional and post-transcriptional regulatory architecture. Finally, episomal maintenance often requires continuous antibiotic selection, imposing an additional metabolic burden that may fundamentally alter cellular physiology [125].

4.2. Endogenous Chromosomal Manipulation Strategies

To overcome the limitations associated with heterologous overexpression systems, increasing emphasis has been placed on endogenous genome engineering strategies. By manipulating genes directly within their native chromosomal context, proteoform expression remains under endogenous promoter and regulatory control. This approach can be combined with small high-sensitivity peptide tags, such as HiBiT or 3 x FLAG, to facilitate proteoform detection at physiological expression levels while avoiding disruptive overexpression [126,127].

The *S. Typhimurium* protein SpaO represents a landmark example of Nt-proteoform decoupling via endogenous genome manipulation. The *spaO* gene contains an internal GTG₂₀₃ initiation codon to yield two distinct proteoforms: the full-length SpaO^L and the truncated SpaO^S proteoforms. To dissect their individual functions, Lara-Tejero et al. employed R6K-based allelic exchange to steer selective Nt-proteoform expression [47,128]. This suicide plasmid-based allelic exchange strategy enables precise chromosomal mutagenesis through homologous recombination. Following counterselection using markers such as *sacB*, a second recombination event removes the plasmid backbone and leaves behind the desired scarless mutation. Specifically, this approach was used to mutate the internal GTG₂₀₃ start codon to the non-initiating GCG codon, successfully eliminating SpaO^S production and enabling selective functional interrogation of SpaO^L. Additionally, loss-of-function ($\Delta spaO$ mutant) phenotypes were complemented by expressing a wild-type *spaO* copy in trans from the pSB4545 plasmid.

4.3. (Multiplex) Recombineering Approaches

More recent developments have implemented multiplexed recombineering approaches to precisely manipulate Nt-proteoform expression while simultaneously enabling high-sensitivity detection through integration of peptide tags such as the luminescent HiBiT tag [33]. This 11-amino acid peptide (VSGWRLFKKIS) associates with high affinity to the complementary Large BiT (LgBiT) subunit to form an active luciferase, thereby enabling highly sensitive bioluminescent quantification. A cornerstone of modern bacterial genome engineering is the λ -Red recombineering system, which mediates homologous recombination through the phage-derived Gam, Bet, and Exo proteins. This platform enables precise genomic engineering using either single-stranded DNA (ssDNA) oligonucleotides or dsDNA repair templates [129].

The use of ssDNA oligonucleotides, commonly referred to as oligo-mediated allelic replacement (OMAR) [130], is particularly well-suited for Nt-proteoform decoupling because it permits seamless conversion of individual TIS codons into non-initiating codons without introducing selection markers that could perturb operon polarity. Maintaining operon integrity is biologically important in prokaryotes, where genes are often tightly clustered and co-transcribed. Residual recombination scars or selection markers may disrupt important transcriptional and translational regulatory sequences, including promoters, terminators, and RBSs [131–133]. Consequently, OMAR is particularly well suited for TIS mutagenesis while preserving the native physiological context of the targeted genetic locus.

Nevertheless, careful oligonucleotide design remains essential when modifying start codons. Near-cognate codons such as ATA should generally be avoided because they may still permit low-level “leaky” translation initiation [32,134]. Instead, codons lacking initiation potential are preferred, ideally while preserving the encoded amino acid or introducing only conservative amino acid substitutions where possible. For example, CTT can serve as a reliable non-initiating alternative in certain leucine-compatible contexts [32]. Furthermore, because closely spaced TISs may overlap with SD-like sequences or other cis-regulatory elements, even single nucleotide substitutions may inadvertently disrupt translational coupling or generate unintended polar effects across polycistronic operons [135–137]. Predictive tools such as the Genetic Systems Calculator [138] therefore provide valuable support for evaluating the translational consequences of engineered mutations.

Beyond design constraints, execution of precise tag-free TIS modifications poses substantial screening challenges. Because OMAR typically generates scarless and non-selectable mutations, the identification of correctly edited clones becomes experimentally challenging. Reported baseline

efficiencies for OMAR vary considerably, typically ranging from 6% to 20% depending on the genetic background, mutation type, and oligonucleotide modifications [130]. To overcome these low baseline frequencies in the absence of direct selection, co-selection MAGE (Cos-MAGE) can be implemented. By co-targeting a nearby selectable chromosomal marker with an additional ssDNA oligo, CoS-MAGE enriches for subpopulations containing both mutations, thereby substantially increasing the allelic replacement frequency (ARF) of the unselectable TIS mutation [139].

To identify successful mutants within these enriched populations, alternative screening approaches such as allele-specific colony PCR (ASC-PCR) are required [140]. While OMAR efficiencies have historically been estimated using selectable chromosomal markers that generate easily screenable phenotypes on selective media—thereby enabling ARF calculation as a percentage of the surviving population[141]—our previous work demonstrated that ASC-PCR enables determination of actual TIS conversion frequencies [142]. Broader optimization strategies for recombineering efficiency have been comprehensively reviewed elsewhere [130].

Overall, multiplexed recombineering approaches have previously been shown to effectively steer Nt-proteoform expression at the genomic level while simultaneously allowing sensitive detection of endogenous expression [33]. Decoupling Nt-proteoform expression is particularly relevant for validation purposes when multiple TISs are located in close proximity within the bacterial genome. Furthermore, this approach enables characterization of conditionally expressed proteoforms, as exemplified by the *S. Typhimurium* FruK and *E. coli* RNA polymerase Sigma S (σ^S) proteoform pairs [33,143].

4.4. CRISPR-Based Nt-Proteoform Engineering

Since the development of clustered regularly interspaced short palindromic repeats (CRISPR)-based genome engineering, CRISPR/Cas systems have become widely implemented for bacterial genome manipulation [144–147]. These approaches rely on guide RNA (gRNA)-directed Cas-mediated cleavage at target protospacer sequences adjacent to protospacer-adjacent motifs (PAMs), thereby enabling counterselection of nonedited wild-type cells. Importantly, CRISPR-based strategies can eliminate the need for selectable markers, classical counterselection systems, or dedicated recombineering platforms. Many early CRISPR/Cas-based bacterial editing strategies required the introduction of additional mutations within the protospacer or PAM sequence to prevent repeated Cas-mediated cleavage following successful genome editing. While initial CRISPR-Cas9-mediated oligonucleotide-directed mutagenesis approaches achieved relatively high editing efficiencies for two- to three-base mutations, introduction of single-nucleotide mutations proved substantially more challenging [148,149]. Such single-base edits were often obtained at frequencies below 3%, largely due to the mismatch tolerance of the CRISPR/Cas system [150]. However, subsequent optimization of sgRNA design strategies dramatically improved single-nucleotide editing efficiencies to between 36% and 95% [150,151].

More recently, the RECKLEEN platform combines λ -Red recombineering with CRISPR-Cas9-mediated counterselection, achieving editing efficiencies approaching 100% in *Klebsiella* [152]. Nevertheless, CRISPR-mediated genome engineering still faces several important limitations, including off-target mutagenesis and the occurrence of ‘escaper’ colonies containing wild-type clones or unintended genomic alterations [153]. In addition, excessive Cas9 expression may exert cytotoxic effects, thereby limiting the broader applicability of certain CRISPR-based editing systems [153]. Despite these limitations, increasingly precise CRISPR-based genome engineering technologies hold potential for Nt-proteoform research by enabling selective manipulation of endogenous proteoform expression within native chromosomal contexts.

5. Functional Characterization of Nt-Proteoforms

The rapid expansion of computational TIS prediction, ribosome profiling, and N-terminomics datasets has substantially increased the number of candidate bacterial Nt-proteoforms. Nevertheless, functional validation and mechanistic characterization of these proteoforms remain comparatively

limited. Recent advances in endogenous genome engineering now enable the selective decoupling of individual proteoforms within their native chromosomal context, thereby facilitating direct investigation of their physiological relevance. Such strategies provide opportunities to study proteoform-specific effects of cellular physiology, subcellular localization, interaction networks, stability, and conditional regulation.

Recently, multiplexed recombinering toolkits have been extended toward genomic integration of promiscuous biotin ligases (PBLs), thereby enabling proximity-dependent biotinylation approaches such as BioID [142]. These systems exploit engineered ligases, including BioID/BirA* [154], BioID2 [155], TurboID, and miniTurboID [156] to map local protein interaction environments, although their application in prokaryotes remains limited [157,158]. Resolving the specific proxoemes of decoupled Nt-proteoforms may provide important insights into how N-terminal variation influences protein interactions, complex assembly, and functional specialization (Figure 4) [159,160].

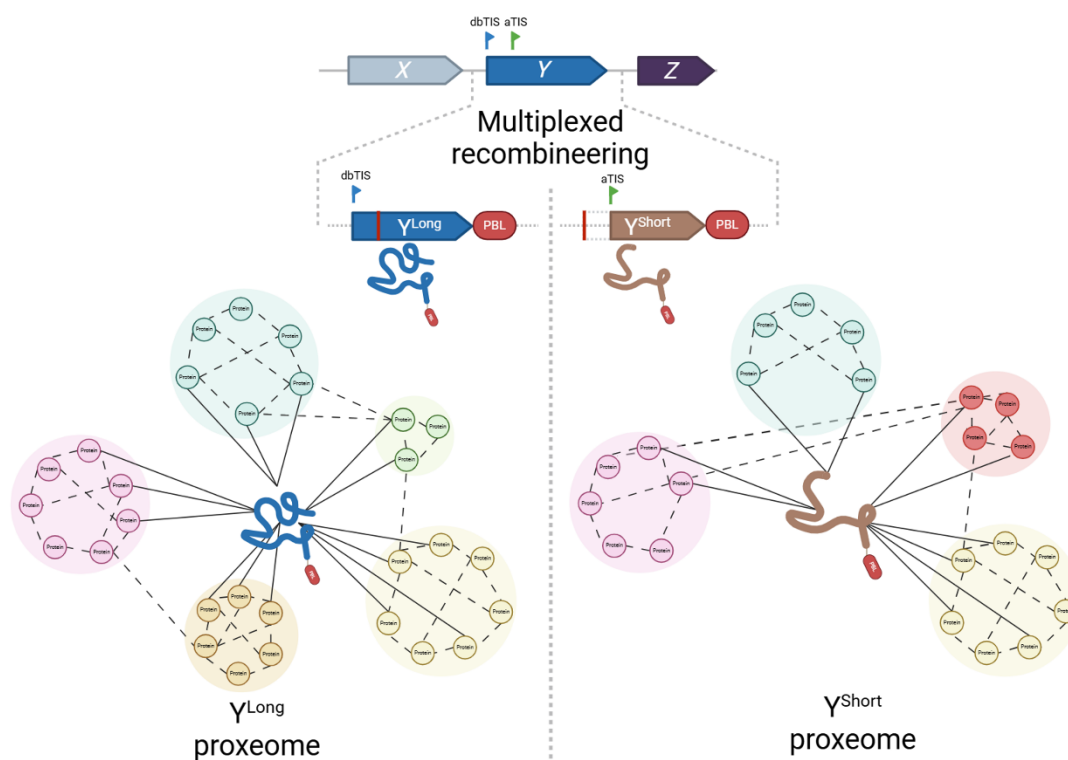


Figure 4. Multiplexed recombinering enables proteoform-specific proximity interactome (“proxoeme”) mapping. Schematic representation of a bacterial genomic locus comprising genes X, Y and Z. Gene Y encodes two Nt-proteoforms generated through alternative translation initiation. The long proteoform (Y^{Long}) is translated from the database-annotated translation initiation site (dbTIS; blue flag), whereas the short proteoform (Y^{Short}) originates from an alternative translation initiation site (aTIS; green flag). Multiplexed recombinering is used to selectively inactivate either TIS while simultaneously integrating a C-terminal promiscuous biotin ligase (PBL) fusion. This genome engineering strategy generates strains selectively expressing either Y^{Long} or Y^{Short} -PBL under endogenous regulatory control. Subsequent proximity-dependent biotin identification (BioID) approaches enable mapping of the specific proxoemes associated with each Nt-proteoform. Solid lines denote direct or stable protein interactions, whereas dashed lines represent indirect, transient, or shared interaction networks. Figure created using BioRender.

Nt-proteoform variation may additionally influence protein stability. In eukaryotes, altered N-termini have been shown to substantially affect proteoform half-life [62]. Protein stability is commonly investigated using pulse-chase approaches such as pulse SILAC (pSILAC) combined with COFRADIC, which enables temporal monitoring of proteoform degradation dynamics [108,111,161–164]. Although initially developed for eukaryotic cells, recent methodological adaptations have

enabled highly efficient SILAC labeling in bacteria, thereby expanding the applicability of these approaches to prokaryotic proteome dynamics [165]. Interestingly, a recent *E. coli* study suggested that N-terminal identity may not always represent the dominant determinant of protein stability in vivo [166]. However, unstable proteoforms may evade detection entirely due to rapid turnover, complicating the interpretation of such datasets.

Collectively, the combination of precise endogenous genome engineering with complementary proteomic and interaction-mapping technologies provides an emerging framework for systematic functional characterization of bacterial Nt-proteoforms.

6. Discussion

Although Nt-proteoforms have now been identified across phylogenetically diverse bacterial species, their overall prevalence and evolutionary distribution across bacterial phyla remain largely unresolved. Current evidence is still largely limited to a relatively small number of experimentally investigated lineages, spanning taxonomically distant groups such as *Streptomyces* and *Cyanobacteria* [35,167]. However, recent riboproteogenomic advances are increasingly enabling more systematic and large-scale identification of bacterial Nt-proteoforms. By revealing a far more dynamic and heterogeneous proteomic landscape than previously appreciated, these methods have fundamentally transformed our understanding of the bacterial translome. In particular, the widespread occurrence of aTIS events demonstrates that single bacterial loci can generate multiple Nt-proteoforms with potentially distinct biological properties. However, the rapid expansion of high-throughput discovery datasets is currently outpacing downstream functional characterization, leaving the physiological relevance of many candidate Nt-proteoforms unresolved.

A central challenge in the field is distinguishing functional Nt-proteoforms from pervasive or non-productive translational noise. Ribosome profiling approaches such as Ribo-RET provide highly sensitive snapshots of ribosome occupancy and initiation-site selection, but fundamentally measure ribosome-protected mRNA fragments, rather than stable protein products. Consequently, not every detected initiation event necessarily yields a biologically relevant proteoform. Nevertheless, our previous work demonstrated that aggregate Ribo-seq and Ribo-RET signals correlate strongly with experimentally determined protein abundances. This correlation indicates that ribosome occupancy metrics can serve as valuable quantitative proxies for translation output [33], and that signal intensity, reproducibility, positional conservation, sequence context, and integration with complementary proteomic evidence serve as critical parameters for prioritizing candidate functional Nt-proteoforms.

Accordingly, orthogonal peptide-level validation remains essential. N-terminomics approaches such as TRAINSPOTTER provide direct molecular evidence for productive translation initiation by experimentally identifying proteoform-specific N-termini [111]. The integration of quantitative ribosome profiling with high-confidence N-terminal proteomics, therefore represents a particularly powerful framework for distinguishing genuine proteoforms from pervasive translational noise and for delineating bacterial translation initiation events at single-amino-acid resolution.

From an evolutionary perspective, alternative translation initiation likely represents an efficient strategy for expanding bacterial functional capacity without increasing genome size. Bacterial genomes are highly compact and coding dense, with approximately 90% of genomic DNA dedicated to coding regions [168]. Within this constrained genomic architecture, the generation of multiple Nt-proteoforms from a single transcript provides a mechanism to diversify protein functionality while minimizing the genetic burden associated with gene duplication. In this regard, Nt-proteoform generation may represent an efficient strategy for expanding proteomic and regulatory complexity in streamlined prokaryotic systems.

Beyond coding efficiency, alternative translation initiation may additionally provide important regulatory and kinetic advantages. Because transcription and translation are tightly coupled in bacteria, modulation of translation initiation can enable highly rapid proteomic adaptation without requiring de novo transcriptional programs [169]. Changes in local mRNA structure, ribosome binding site accessibility, or environmental cues may rapidly shift TIS usage and thereby alter

proteoform production under changing growth conditions. Such mechanisms could facilitate fast post-transcriptional remodelling of the bacterial proteome during environmental, thermal, nutritional, or host-associated stress responses. Consistent with this concept, several characterized proteoform-encoding genes, including *if2*, *clpB*, *safA*, and σ^S , display condition-dependent expression or specialized stress-associated functionality [122,170,171].

Alternative translation initiation may also contribute to stoichiometric regulation of multi-protein complexes. In the T3SS sorting platform of *S. Typhimurium*, defined ratios between long and short SsaQ and SpaO proteoforms are required for proper injectisome assembly and functionality [47,48,172]. Encoding multiple structural variants within a single transcriptional unit may therefore provide a robust mechanism to coordinate relative subunit abundance while minimizing the need for additional transcriptional regulatory layers.

In addition, Nt-proteoform diversity may influence protein stability and conditional protein turnover. Although bacterial N-degron pathways clearly establish the importance of N-terminal identity in proteolytic targeting, endogenous Nt-proteoforms differing only minimally at their N-termini have thus far rarely been systematically linked to distinct degradation kinetics. It nevertheless remains highly plausible that certain alternative proteoforms evolved as transient or condition-specific molecular variants with specialized accessory functions and tightly regulated half-lives.

Ultimately, systematic functional characterization of bacterial Nt-proteoforms will require a transition away from artifact-prone plasmid overexpression systems toward precise endogenous genome engineering approaches. The use of advanced genome engineering technologies now enables selective manipulation of individual proteoforms at endogenous levels. Combined with complementary approaches, including proximity labeling, quantitative proteomics, and proteoform stability profiling, these tools may facilitate bridging the gap between high-throughput Nt-proteoform discovery and deep mechanistic biological understanding. Collectively, continued integration of riboproteogenomics, N-terminomics, and endogenous genome engineering is expected to substantially refine our understanding of bacterial proteome organization and regulatory complexity. Resolving how Nt-proteoforms contribute to bacterial physiology, adaptation, and pathogenesis will likely represent an important frontier in future bacterial proteome biology.

Author Contributions: Conceptualization, V.S. and P.V.D.; writing-original draft preparation, V.S.; writing—review and editing V.S. and P.V.D.; visualization, V.S.; supervision, P.V.D.; project administration, P.V.D.; funding acquisition, P.V.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Research Foundation – Flanders (FWO-Vlaanderen; project no. G088726N), awarded to P.V.D., and the Special Research Fund (BOF) of Ghent University (reference no. BOF25/CDV/006).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

6-FT	Six-frame translation
ARF	Allelic replacement frequency
ASC-PCR	Allele-specific colony PCR
BASys	Bacterial Annotation system
BV-BRC	Bacterial and Viral Bioinformatics Resource Center
CDD	Conserved Domain Database
CDS	Coding sequence
COFRADIC	Combined FRActional Diagonal Chromatography

Cos-MAGE	Co-selection MAGE
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CyaA	Calmodulin-dependent adenylate cyclase
dbTIS	Database-annotated translation initiation site
DRTs	Defense-associated reverse transcriptases
dsDNA	Double stranded DNA
<i>E. coli</i>	<i>Escherichia coli</i>
gLM	Genomic language model
gRNA	guide RNA
HMM	Hidden Markov Model
IB	Inclusion Body
IF2	Initiation Factor 2
iMet	Initiator methionine
INRI-seq	in vitro Ribo-seq
iTIS	Internal translation initiation site
kDa	Kilodalton
LATE	LysN Amino Terminal Enrichment
LC-MS/MS	Liquid Chromatography-Tandem Mass Spectrometry
LgBiT	Large BiT
MAGE	Multiplex Automated Genome Engineering
MetAP	Methionine aminopeptidase
mRNA	Messenger RNA
MS	Mass spectrometry
NAT	N-terminal acetyl transferase
N-terminomics	N-terminal proteomics
Nt-proteoforms	N-terminal proteoforms
OMAR	Oligo-mediated allelic replacement
ORF	Open reading frame
PAM	Protospacer-adjacent motifs
PBL	Promiscuous biotin ligase
PDF	Peptide deformylase
PFM	Protein family model
PGAP	Prokaryotic Genome Annotation Pipeline
PPIs	Protein-protein interactions
Prodigal	Prokaryotic Dynamic programming Gene-finding Algorithm
pSILAC	Pulse Stable isotope labeling by amino acids in cell culture
RAST	Rapid Annotations using subsystems Technology
RECKLEEN	Recombineering/CRISPR-based <i>KLebsiella</i> Engineering for Efficient Nucleotide editing
REPARATION	Ribosome Profiling Assisted (Re-)AnnotATIION
Ribo-RET	Retapamulin-assisted ribosome profiling
Ribo-seq	Ribosome profiling
Rubisco	Ribulose-1,5-biphosphate carboxylase/oxygenase
<i>S. Typhimurium</i>	<i>Salmonella enterica</i> serovar Typhimurium
SD	Shine-Dalgarno
sgRNA	Single guide RNA
sORF	Small open reading frame
SP	Signal peptidase
SP	Sorting platform
T3E	Type III secretion system effector
T3SS	Type III secretion system
TAILS	Terminal Amine Isotopic Labeling
TetRP	Tetracycline-inhibited ribosome profiling
TRANSPOTTER	TRANslation Initiation SPOTTER
TSS	Transcription start site

References

1. Crick, F. Central Dogma of Molecular Biology. *Nature* **1970**, *227*, 561–563, doi:10.1038/227561a0.
2. Deng, P.; Lee, H.; Armijo, C.; Wang, H.; Gao, A. Protein-templated synthesis of dinucleotide repeat DNA by an antiphage reverse transcriptase. *Science* **2026**, *0*, eaed1656, doi:doi:10.1126/science.aed1656.
3. Smith, L.M.; Kelleher, N.L.; Linial, M.; Goodlett, D.; Langridge-Smith, P.; Ah Goo, Y.; Safford, G.; Bonilla*, L.; Kruppa, G.; Zubarev, R.; et al. Proteoform: a single term describing protein complexity. *Nature Methods* **2013**, *10*, 186–187, doi:10.1038/nmeth.2369.
4. Bingel-Erlenmeyer, R.; Kohler, R.; Kramer, G.; Sandikci, A.; Antolić, S.; Maier, T.; Schaffitzel, C.; Wiedmann, B.; Bukau, B.; Ban, N. A peptide deformylase–ribosome complex reveals mechanism of nascent chain processing. *Nature* **2008**, *452*, 108–111, doi:10.1038/nature06683.
5. Giglione, C.; Fieulaine, S.; Meinnel, T. N-terminal protein modifications: Bringing back into play the ribosome. *Biochimie* **2015**, *114*, 134–146, doi:<https://doi.org/10.1016/j.biochi.2014.11.008>.
6. Meinnel, T.; Mechulam, Y.; Blanquet, S. Methionine as translation start signal: A review of the enzymes of the pathway in *Escherichia coli*. *Biochimie* **1993**, *75*, 1061–1075, doi:[https://doi.org/10.1016/0300-9084\(93\)90005-D](https://doi.org/10.1016/0300-9084(93)90005-D).
7. Bienvenut, W.V.; Giglione, C.; Meinnel, T. Proteome-wide analysis of the amino terminal status of *Escherichia coli* proteins at the steady-state and upon deformylation inhibition. *PROTEOMICS* **2015**, *15*, 2503–2518, doi:<https://doi.org/10.1002/pmic.201500027>.
8. Solbiati, J.; Chapman-Smith, A.; Miller, J.L.; Miller, C.G.; Cronan, J.E. Processing of the N termini of nascent polypeptide chains requires deformylation prior to methionine removal Edited by M. Gottesman. *Journal of Molecular Biology* **1999**, *290*, 607–614, doi:<https://doi.org/10.1006/jmbi.1999.2913>.
9. Frottin, F.; Martinez, A.; Peynot, P.; Mitra, S.; Holz, R.C.; Giglione, C.; Meinnel, T. The Proteomics of N-terminal Methionine Cleavage. *Molecular & Cellular Proteomics* **2006**, *5*, 2336–2349, doi:<https://doi.org/10.1074/mcp.M600225-MCP200>.
10. VanDrisse, C.M.; Escalante-Semerena, J.C. Protein Acetylation in Bacteria. *Annual Review of Microbiology* **2019**, *73*, 111–132, doi:<https://doi.org/10.1146/annurev-micro-020518-115526>.
11. Thompson, C.R.; Champion, M.M.; Champion, P.A. Quantitative N-Terminal Footprinting of Pathogenic *Mycobacteria* Reveals Differential Protein Acetylation. *Journal of Proteome Research* **2018**, *17*, 3246–3258, doi:10.1021/acs.jproteome.8b00373.
12. Christensen, D.G.; Baumgartner, J.T.; Xie, X.; Jew, K.M.; Basisty, N.; Schilling, B.; Kuhn, M.L.; Wolfe, A.J. Mechanisms, Detection, and Relevance of Protein Acetylation in Prokaryotes. *mBio* **2019**, *10*, 10.1128/mbio.02708–02718, doi:doi:10.1128/mbio.02708-18.
13. Bonissone, S.; Gupta, N.; Romine, M.; Bradshaw, R.A.; Pevzner, P.A. N-terminal Protein Processing: A Comparative Proteogenomic Analysis*. *Molecular & Cellular Proteomics* **2013**, *12*, 14–28, doi:<https://doi.org/10.1074/mcp.M112.019075>.
14. Ouidir, T.; Jarnier, F.; Cosette, P.; Jouenne, T.; Hardouin, J. Characterization of N-terminal protein modifications in *Pseudomonas aeruginosa* PA14. *Journal of Proteomics* **2015**, *114*, 214–225, doi:<https://doi.org/10.1016/j.jprot.2014.11.006>.
15. Jones, J.D.; O'Connor, C.D. Protein acetylation in prokaryotes. *PROTEOMICS* **2011**, *11*, 3012–3022, doi:<https://doi.org/10.1002/pmic.201000812>.
16. Hegde, R.S.; Bernstein, H.D. The surprising complexity of signal sequences. *Trends in Biochemical Sciences* **2006**, *31*, 563–571, doi:<https://doi.org/10.1016/j.tibs.2006.08.004>.
17. Kaushik, S.; He, H.; Dalbey, R.E. Bacterial Signal Peptides- Navigating the Journey of Proteins. *Frontiers in Physiology* **2022**, *Volume 13 - 2022*, doi:10.3389/fphys.2022.933153.
18. Huvet, M.; Stumpf, M.P.H. Overlapping genes: a window on gene evolvability. *BMC Genomics* **2014**, *15*, 721, doi:10.1186/1471-2164-15-721.
19. Meydan, S.; Marks, J.; Klepacki, D.; Sharma, V.; Baranov, P.V.; Firth, A.E.; Margus, T.; Kefi, A.; Vázquez-Laslop, N.; Mankin, A.S. Retapamulin-Assisted Ribosome Profiling Reveals the Alternative Bacterial Proteome. *Molecular Cell* **2019**, *74*, 481–493.e486, doi:<https://doi.org/10.1016/j.molcel.2019.02.017>.

20. Altuvia, S.; Zhang, A.; Argaman, L.; Tiwari, A.; Storz, G. The *Escherichia coli* OxyS regulatory RNA represses *flhA* translation by blocking ribosome binding. *The EMBO Journal* **1998**, *17*, 6069–6075, doi:10.1093/emboj/17.20.6069.
21. Majdalani, N.; Cuning, C.; Sledjeski, D.; Elliott, T.; Gottesman, S. DsrA RNA regulates translation of RpoS message by an anti-antisense mechanism, independent of its action as an antisilencer of transcription. *Proceedings of the National Academy of Sciences* **1998**, *95*, 12462–12467, doi:doi:10.1073/pnas.95.21.12462.
22. Storz, G.; Opdyke, J.A.; Zhang, A. Controlling mRNA stability and translation with small, noncoding RNAs. *Current Opinion in Microbiology* **2004**, *7*, 140–144, doi:<https://doi.org/10.1016/j.mib.2004.02.015>.
23. Winkler, W.; Nahvi, A.; Breaker, R.R. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* **2002**, *419*, 952–956, doi:10.1038/nature01145.
24. Mandal, M.; Breaker, R.R. Gene regulation by riboswitches. *Nature Reviews Molecular Cell Biology* **2004**, *5*, 451–463, doi:10.1038/nrm1403.
25. Moine, H.; Romby, P.; Springer, M.; Grunberg-Manago, M.; Ebel, J.-P.; Ehresmann, B.; Ehresmann, C. *Escherichia coli* threonyl-tRNA synthetase and tRNA^{Thr} modulate the binding of the ribosome to the translational initiation site of the *ThrS* mRNA. *Journal of Molecular Biology* **1990**, *216*, 299–310, doi:[https://doi.org/10.1016/S0022-2836\(05\)80321-3](https://doi.org/10.1016/S0022-2836(05)80321-3).
26. Babitzke, P.; Baker, C.S.; Romeo, T. Regulation of Translation Initiation by RNA Binding Proteins. *Annual Review of Microbiology* **2009**, *63*, 27–44, doi:<https://doi.org/10.1146/annurev.micro.091208.073514>.
27. Perham, R.N. Structural aspects of biomolecular recognition and self-assembly. *Biosensors and Bioelectronics* **1994**, *9*, 753–760, doi:[https://doi.org/10.1016/0956-5663\(94\)80074-X](https://doi.org/10.1016/0956-5663(94)80074-X).
28. Scharff, L.B.; Childs, L.; Walther, D.; Bock, R. Local Absence of Secondary Structure Permits Translation of mRNAs that Lack Ribosome-Binding Sites. *PLOS Genetics* **2011**, *7*, e1002155, doi:10.1371/journal.pgen.1002155.
29. Nirenberg, M.; Leder, P. RNA Codewords and Protein Synthesis. *Science* **1964**, *145*, 1399–1407, doi:doi:10.1126/science.145.3639.1399.
30. Blattner, F.R.; Plunkett, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; et al. The Complete Genome Sequence of *Escherichia coli* K-12. *Science* **1997**, *277*, 1453–1462, doi:doi:10.1126/science.277.5331.1453.
31. Villegas, A.; Kropinski, A.M. An analysis of initiation codon utilization in the Domain Bacteria – concerns about the quality of bacterial genome annotation. *Microbiology* **2008**, *154*, 2559–2661, doi:<https://doi.org/10.1099/mic.0.2008/021360-0>.
32. Hecht, A.; Glasgow, J.; Jaschke, P.R.; Bawazer, L.A.; Munson, M.S.; Cochran, J.R.; Endy, D.; Salit, M. Measurements of translation initiation from all 64 codons in *E. coli*. *Nucleic Acids Research* **2017**, *45*, 3615–3626, doi:10.1093/nar/gkx070.
33. Fijalkowski, I.; Snauwaert, V.; Damme, P.V. Proteins à la carte: riboproteogenomic exploration of bacterial N-terminal proteoform expression. *mBio* **2024**, *15*, e00333–00324, doi:doi:10.1128/mbio.00333-24.
34. Cannon, G.C.; Bradburne, C.E.; Aldrich, H.C.; Baker, S.H.; Heinhorst, S.; Shively, J.M. Microcompartments in Prokaryotes: Carboxysomes and Related Polyhedra. *Applied and Environmental Microbiology* **2001**, *67*, 5351–5361, doi:doi:10.1128/AEM.67.12.5351-5361.2001.
35. Long, B.M.; Tucker, L.; Badger, M.R.; Price, G.D. Functional Cyanobacterial β -Carboxysomes Have an Absolute Requirement for Both Long and Short Forms of the CcmM Protein *Plant Physiology* **2010**, *153*, 285–293, doi:10.1104/pp.110.154948.
36. Long, B.M.; Badger, M.R.; Whitney, S.M.; Price, G.D. Analysis of Carboxysomes from *Synechococcus* PCC7942 Reveals Multiple Rubisco Complexes with Carboxysomal Proteins CcmM and CcaA. *Journal of Biological Chemistry* **2007**, *282*, 29323–29335, doi:<https://doi.org/10.1074/jbc.M703896200>.
37. Cot, S.S.-W.; So, A.K.-C.; Espie, G.S. A Multiprotein Bicarbonate Dehydration Complex Essential to Carboxysome Function in Cyanobacteria. *Journal of Bacteriology* **2008**, *190*, 936–945, doi:doi:10.1128/jb.01283-07.
38. Van Damme, P.; Jonckheere, V.; Simoens, L. *Systematic real-time profiling of Salmonella type III effector translocation provides quantitative resolution of the T3SS-1/T3SS-2 secretion dichotomy*; 2026.

39. Worley, M.J. *Salmonella* Type III Secretion System Effectors. *International Journal of Molecular Sciences* **2025**, *26*, 2611.
40. Pillay, T.D.; Hettiarachchi, S.U.; Gan, J.; Diaz-Del-Olmo, I.; Yu, X.-J.; Muench, J.H.; Thurston, T.L.M.; Pearson, J.S. Speaking the host language: how *Salmonella* effector proteins manipulate the host. *Microbiology* **2023**, *169*, doi:<https://doi.org/10.1099/mic.0.001342>.
41. El Qaidi, S.; Scott, N.E.; Hays, M.P.; Geisbrecht, B.V.; Watkins, S.; Hardwidge, P.R. An intra-bacterial activity for a T3SS effector. *Scientific Reports* **2020**, *10*, 1073, doi:10.1038/s41598-020-58062-y.
42. Hasan, M.K.; El Qaidi, S.; Hardwidge, P.R. The T3SS Effector Protease NleC Is Active within *Citrobacter rodentium*. *Pathogens* **2021**, *10*, 589.
43. Niemann, G.S.; Brown, R.N.; Mushamiri, I.T.; Nguyen, N.T.; Taiwo, R.; Stufkens, A.; Smith, R.D.; Adkins, J.N.; McDermott, J.E.; Heffron, F. RNA Type III Secretion Signals That Require Hfq. *Journal of Bacteriology* **2013**, *195*, 2119–2125, doi:10.1128/jb.00024-13.
44. Büttner, D. Protein export according to schedule: architecture, assembly, and regulation of type III secretion systems from plant- and animal-pathogenic bacteria. *Microbiol Mol Biol Rev* **2012**, *76*, 262–310, doi:10.1128/mmmbr.05017-11.
45. Abrusci, P.; McDowell, M.A.; Lea, S.M.; Johnson, S. Building a secreting nanomachine: a structural overview of the T3SS. *Current Opinion in Structural Biology* **2014**, *25*, 111–117, doi:<https://doi.org/10.1016/j.sbi.2013.11.001>.
46. Morita-Ishihara, T.; Ogawa, M.; Sagara, H.; Yoshida, M.; Katayama, E.; Sasakawa, C. *Shigella* Spa33 Is an Essential C-ring Component of Type III Secretion Machinery. *Journal of Biological Chemistry* **2006**, *281*, 599–607, doi:10.1074/jbc.M509644200.
47. Lara-Tejero, M.; Qin, Z.; Hu, B.; Butan, C.; Liu, J.; Galán, J.E. Role of SpaO in the assembly of the sorting platform of a *Salmonella* type III secretion system. *PLOS Pathogens* **2019**, *15*, e1007565, doi:10.1371/journal.ppat.1007565.
48. Soto, J.E.; Wang, T.; Galán, J.E.; Lara-Tejero, M. Interplay between SpaO variants shapes the architecture of the *Salmonella* type III secretion sorting platform. *mBio* **2026**, *17*, e00155–00126, doi:10.1128/mbio.00155-26.
49. Bzymek, K.P.; Hamaoka, B.Y.; Ghosh, P. Two Translation Products of *Yersinia* yscQ Assemble To Form a Complex Essential to Type III Secretion. *Biochemistry* **2012**, *51*, 1669–1677, doi:10.1021/bi201792p.
50. Diepold, A.; Kudryashev, M.; Delalez, N.J.; Berry, R.M.; Armitage, J.P. Composition, Formation, and Regulation of the Cytosolic C-ring, a Dynamic Component of the Type III Secretion Injectisome. *PLOS Biology* **2015**, *13*, e1002039, doi:10.1371/journal.pbio.1002039.
51. McDowell, M.A.; Marcoux, J.; McVicker, G.; Johnson, S.; Fong, Y.H.; Stevens, R.; Bowman, L.A.H.; Degiacomi, M.T.; Yan, J.; Wise, A.; et al. Characterisation of *Shigella* Spa33 and *Thermotoga* FliM/N reveals a new model for C-ring assembly in T3SS. *Molecular Microbiology* **2016**, *99*, 749–766, doi:<https://doi.org/10.1111/mmi.13267>.
52. Wadhams, G.H.; Armitage, J.P. Making sense of it all: bacterial chemotaxis. *Nature Reviews Molecular Cell Biology* **2004**, *5*, 1024–1037, doi:10.1038/nrm1524.
53. Levit, M.N.; Stock, J.B. Receptor Methylation Controls the Magnitude of Stimulus-Response Coupling in Bacterial Chemotaxis. *Journal of Biological Chemistry* **2002**, *277*, 36760–36765, doi:10.1074/jbc.M204325200.
54. Hess, J.F.; Oosawa, K.; Kaplan, N.; Simon, M.I. Phosphorylation of three proteins in the signaling pathway of bacterial chemotaxis. *Cell* **1988**, *53*, 79–87, doi:[https://doi.org/10.1016/0092-8674\(88\)90489-8](https://doi.org/10.1016/0092-8674(88)90489-8).
55. Wolfe, A.J.; Stewart, R.C. The short form of the CheA protein restores kinase activity and chemotactic ability to kinase-deficient mutants. *Proceedings of the National Academy of Sciences* **1993**, *90*, 1518–1522, doi:10.1073/pnas.90.4.1518.
56. Wang, H.; Matsumura, P. Phosphorylating and dephosphorylating protein complexes in bacterial chemotaxis. *Journal of Bacteriology* **1997**, *179*, 287–289, doi:10.1128/jb.179.1.287-289.1997.
57. O’Conno, C.; Matsumura, P. The Accessibility of Cys-120 in CheAS Is Important for the Binding of CheZ and Enhancement of CheZ Phosphatase Activity. *Biochemistry* **2004**, *43*, 6909–6916, doi:10.1021/bi035592n.
58. Tobias, J.W.; Shrader, T.E.; Rocap, G.; Varshavsky, A. The N-End Rule in Bacteria. *Science* **1991**, *254*, 1374–1377, doi:10.1126/science.1962196.

59. Dougan, D.A.; Truscott, K.N.; Zeth, K. The bacterial N-end rule pathway: expect the unexpected. *Molecular Microbiology* **2010**, *76*, 545–558, doi:<https://doi.org/10.1111/j.1365-2958.2010.07120.x>.
60. Heo, A.J.; Kim, S.B.; Kwon, Y.T.; Ji, C.H. The N-degron pathway: From basic science to therapeutic applications. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **2023**, *1866*, 194934, doi:<https://doi.org/10.1016/j.bbagrm.2023.194934>.
61. Varshavsky, A. N-degron and C-degron pathways of protein degradation. *Proceedings of the National Academy of Sciences* **2019**, *116*, 358–366, doi:10.1073/pnas.1816596116.
62. Gawron, D.; Ndah, E.; Gevaert, K.; Van Damme, P. Positional proteomics reveals differences in N-terminal proteoform stability. *Molecular Systems Biology* **2016**, *12*, MSB156662, doi:10.15252/msb.20156662.
63. Fijalkowska, D.; Fijalkowski, I.; Willems, P.; Van Damme, P. Bacterial riboproteogenomics: the era of N-terminal proteoform existence revealed. *FEMS Microbiology Reviews* **2020**, *44*, 418–431, doi:10.1093/femsre/fuaa013.
64. Salzberg, S.L.; Delcher, A.L.; Kasif, S.; White, O. Microbial gene identification using interpolated Markov models. *Nucleic Acids Research* **1998**, *26*, 544–548, doi:10.1093/nar/26.2.544.
65. Borodovsky, M.; Rudd, K.E.; Koonin, E.V. Intrinsic and extrinsic approaches for detecting genes in a bacterial genome. *Nucleic Acids Research* **1994**, *22*, 4756–4767, doi:10.1093/nar/22.22.4756.
66. Hyatt, D.; Chen, G.-L.; LoCascio, P.F.; Land, M.L.; Larimer, F.W.; Hauser, L.J. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **2010**, *11*, 119, doi:10.1186/1471-2105-11-119.
67. Tripp, H.J.; Sutton, G.; White, O.; Wortman, J.; Pati, A.; Mikhailova, N.; Ovchinnikova, G.; Payne, S.H.; Kyrpides, N.C.; Ivanova, N. Toward a standard in structural genome annotation for prokaryotes. *Standards in Genomic Sciences* **2015**, *10*, 45, doi:10.1186/s40793-015-0034-9.
68. Poole, F.L.; Gerwe, B.A.; Hopkins, R.C.; Schut, G.J.; Weinberg, M.V.; Jenney, F.E.; Adams, M.W.W. Defining Genes in the Genome of the Hyperthermophilic Archaeon *Pyrococcus furiosus*: Implications for All Microbial Genomes. *Journal of Bacteriology* **2005**, *187*, 7325–7332, doi:10.1128/jb.187.21.7325-7332.2005.
69. Van Domselaar, G.H.; Stothard, P.; Shrivastava, S.; Cruz, J.A.; Guo, A.; Dong, X.; Lu, P.; Szafron, D.; Greiner, R.; Wishart, D.S. BASys: a web server for automated bacterial genome annotation. *Nucleic Acids Research* **2005**, *33*, W455–W459, doi:10.1093/nar/gki593.
70. Aziz, R.K.; Bartels, D.; Best, A.A.; DeJongh, M.; Disz, T.; Edwards, R.A.; Formsma, K.; Gerdes, S.; Glass, E.M.; Kubal, M.; et al. The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics* **2008**, *9*, 75, doi:10.1186/1471-2164-9-75.
71. Olson, R.D.; Assaf, R.; Brettin, T.; Conrad, N.; Cucinell, C.; Davis, James J.; Dempsey, Donald M.; Dickerman, A.; Dietrich, Emily M.; Kenyon, Ronald W.; et al. Introducing the Bacterial and Viral Bioinformatics Resource Center (BV-BRC): a resource combining PATRIC, IRD and ViPR. *Nucleic Acids Research* **2022**, *51*, D678–D689, doi:10.1093/nar/gkac1003.
72. Haft, D.H.; Selengut, J.D.; Richter, R.A.; Harkins, D.; Basu, M.K.; Beck, E. TIGRFAMs and Genome Properties in 2013. *Nucleic Acids Research* **2012**, *41*, D387–D395, doi:10.1093/nar/gks1234.
73. Tatusova, T.; DiCuccio, M.; Badretdin, A.; Chetvernin, V.; Nawrocki, E.P.; Zaslavsky, L.; Lomsadze, A.; Pruitt, K.D.; Borodovsky, M.; Ostell, J. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Research* **2016**, *44*, 6614–6624, doi:10.1093/nar/gkw569.
74. Li, W.; O'Neill, K.R.; Haft, D.H.; DiCuccio, M.; Chetvernin, V.; Badretdin, A.; Coulouris, G.; Chitsaz, F.; Derbyshire, Myra K.; Durkin, A.S.; et al. RefSeq: expanding the Prokaryotic Genome Annotation Pipeline reach with protein family model curation. *Nucleic Acids Research* **2020**, *49*, D1020–D1028, doi:10.1093/nar/gkaa1105.
75. Baek, J.; Lee, J.; Yoon, K.; Lee, H. Identification of Unannotated Small Genes in *Salmonella*. *G3 Genes | Genomes | Genetics* **2017**, *7*, 983–989, doi:10.1534/g3.116.036939.
76. Fijalkowski, I.; Willems, P.; Jonckheere, V.; Simoens, L.; Van Damme, P. Hidden in plain sight: challenges in proteomics detection of small ORF-encoded polypeptides. *microLife* **2022**, *3*, doi:10.1093/femsml/uqac005.
77. Simoens, L.; Fijalkowski, I.; Van Damme, P. Exposing the small protein load of bacterial life. *FEMS Microbiology Reviews* **2023**, *47*, doi:10.1093/femsre/fuad063.

78. Kim, E.Y.; Tyndall, E.R.; Huang, K.C.; Tian, F.; Ramamurthi, K.S. Dash-and-Recruit Mechanism Drives Membrane Curvature Recognition by the Small Bacterial Protein SpoVM. *Cell Systems* **2017**, *5*, 518–526.e513, doi:<https://doi.org/10.1016/j.cels.2017.10.004>.
79. Raina, M.; Aoyama, J.J.; Bhatt, S.; Paul, B.J.; Zhang, A.; Updegrove, T.B.; Miranda-Ríos, J.; Storz, G. Dual-function AzuCR RNA modulates carbon metabolism. *Proceedings of the National Academy of Sciences* **2022**, *119*, e2117930119, doi:doi:10.1073/pnas.2117930119.
80. Yeom, J.; Shao, Y.; Groisman, E.A. Small proteins regulate *Salmonella* survival inside macrophages by controlling degradation of a magnesium transporter. *Proceedings of the National Academy of Sciences* **2020**, *117*, 20235–20243, doi:doi:10.1073/pnas.2006116117.
81. Ebmeier, S.E.; Tan, I.S.; Clapham, K.R.; Ramamurthi, K.S. Small proteins link coat and cortex assembly during sporulation in *Bacillus subtilis*. *Molecular Microbiology* **2012**, *84*, 682–696, doi:<https://doi.org/10.1111/j.1365-2958.2012.08052.x>.
82. Gaßel, M.; Möllenkamp, T.; Puppe, W.; Altendorf, K. The KdpF Subunit Is Part of the K⁺-translocating Kdp Complex of *Escherichia coli* and Is Responsible for Stabilization of the Complex in Vitro. *Journal of Biological Chemistry* **1999**, *274*, 37901–37907, doi:<https://doi.org/10.1074/jbc.274.53.37901>.
83. Chen, H.; Luo, Q.; Yin, J.; Gao, T.; Gao, H. Evidence for the requirement of CydX in function but not assembly of the cytochrome bd oxidase in *Shewanella oneidensis*. *Biochimica et Biophysica Acta (BBA) - General Subjects* **2015**, *1850*, 318–328, doi:<https://doi.org/10.1016/j.bbagen.2014.10.005>.
84. Sun, Y.-H.; de Jong, M.F.; den Hartigh, A.B.; Roux, C.M.; Rolan, H.G.; Tsolis, R.M. The small protein CydX is required for function of cytochrome bd oxidase in *Brucella abortus*. *Frontiers in Cellular and Infection Microbiology* **2012**, *Volume 2 - 2012*, doi:10.3389/fcimb.2012.00047.
85. Benegas, G.; Ye, C.; Albors, C.; Li, J.C.; Song, Y.S. Genomic language models: opportunities and challenges. *Trends in Genetics* **2025**, *41*, 286–302, doi:10.1016/j.tig.2024.11.013.
86. Gorenstein, L.; Konen, E.; Green, M.; Klang, E. Bidirectional Encoder Representations from Transformers in Radiology: A Systematic Review of Natural Language Processing Applications. *Journal of the American College of Radiology* **2024**, *21*, 914–941, doi:10.1016/j.jacr.2024.01.012.
87. Brixi, G.; Durrant, M.G.; Ku, J.; Naghipourfar, M.; Poli, M.; Sun, G.; Brockman, G.; Chang, D.; Fanton, A.; Gonzalez, G.A.; et al. Genome modelling and design across all domains of life with Evo 2. *Nature* **2026**, *652*, 1349–1361, doi:10.1038/s41586-026-10176-5.
88. Al-Ajlan, A.; El Allali, A. CNN-MGP: Convolutional Neural Networks for Metagenomics Gene Prediction. *Interdisciplinary Sciences: Computational Life Sciences* **2019**, *11*, 628–635, doi:10.1007/s12539-018-0313-4.
89. Akotenou, G.; El Allali, A. Genomic language models (gLMs) decode bacterial genomes for improved gene prediction and translation initiation site identification. *Briefings in Bioinformatics* **2025**, *26*, doi:10.1093/bib/bbaf311.
90. Zhang, S.; Hu, H.; Jiang, T.; Zhang, L.; Zeng, J. TITER: predicting translation initiation sites by deep learning. *Bioinformatics* **2017**, *33*, i234–i242, doi:10.1093/bioinformatics/btx247.
91. Kalkatawi, M.; Magana-Mora, A.; Jankovic, B.; Bajic, V.B. DeepGSR: an optimized deep-learning structure for the recognition of genomic signals and regions. *Bioinformatics* **2019**, *35*, 1125–1132, doi:10.1093/bioinformatics/bty752.
92. Kumar, D.; Yadav, A.K.; Kadimi, P.K.; Nagaraj, S.H.; Grimmond, S.M.; Dash, D. Proteogenomic Analysis of *Bradyrhizobium japonicum* USDA110 Using Genosuite, an Automated Multi-algorithmic Pipeline*. *Molecular & Cellular Proteomics* **2013**, *12*, 3388–3397, doi:<https://doi.org/10.1074/mcp.M112.027169>.
93. Potgieter, M.G.; Nakedi, K.C.; Ambler, J.M.; Nel, A.J.M.; Garnett, S.; Soares, N.C.; Mulder, N.; Blackburn, J.M. Proteogenomic Analysis of *Mycobacterium smegmatis* Using High Resolution Mass Spectrometry. *Frontiers in Microbiology* **2016**, *Volume 7 - 2016*, doi:10.3389/fmicb.2016.00427.
94. Čuklina, J.; Hahn, J.; Imakaev, M.; Omasits, U.; Förstner, K.U.; Ljubimov, N.; Goebel, M.; Pessi, G.; Fischer, H.-M.; Ahrens, C.H.; et al. Genome-wide transcription start site mapping of *Bradyrhizobium japonicum* grown free-living or in symbiosis – a rich resource to identify new transcripts, proteins and to study gene regulation. *BMC Genomics* **2016**, *17*, 302, doi:10.1186/s12864-016-2602-9.

95. Abendroth, U.; Adlung, N.; Otto, A.; Grüneisen, B.; Becher, D.; Bonas, U. Identification of new protein-coding genes with a potential role in the virulence of the plant pathogen *Xanthomonas euvesicatoria*. *BMC Genomics* **2017**, *18*, 625, doi:10.1186/s12864-017-4041-7.
96. Ndah, E.; Jonckheere, V.; Giess, A.; Valen, E.; Menschaert, G.; Van Damme, P. REPARATION: ribosome profiling assisted (re-)annotation of bacterial genomes. *Nucleic Acids Research* **2017**, *45*, e168–e168, doi:10.1093/nar/gkx758.
97. Giess, A.; Jonckheere, V.; Ndah, E.; Chyżyńska, K.; Van Damme, P.; Valen, E. Ribosome signatures aid bacterial translation initiation site identification. *BMC Biology* **2017**, *15*, 76, doi:10.1186/s12915-017-0416-0.
98. Nakahigashi, K.; Takai, Y.; Kimura, M.; Abe, N.; Nakayashiki, T.; Shiwa, Y.; Yoshikawa, H.; Wanner, B.L.; Ishihama, Y.; Mori, H. Comprehensive identification of translation start sites by tetracycline-inhibited ribosome profiling. *DNA Research* **2016**, *23*, 193–201, doi:10.1093/dnares/dsw008.
99. Eisenberg, A.R.; Higdon, A.L.; Hollerer, I.; Fields, A.P.; Jungreis, I.; Diamond, P.D.; Kellis, M.; Jovanovic, M.; Brar, G.A. Translation Initiation Site Profiling Reveals Widespread Synthesis of Non-AUG-Initiated Protein Isoforms in Yeast. *Cell Systems* **2020**, *11*, 145–160.e145, doi:10.1016/j.cels.2020.06.011.
100. Stringer, A.; Smith, C.; Mangano, K.; Wade, J.T. Identification of Novel Translated Small Open Reading Frames in *Escherichia coli* Using Complementary Ribosome Profiling Approaches. *Journal of Bacteriology* **2022**, *204*, e00352–00321, doi:doi:10.1128/JB.00352-21.
101. Limbu, M.S.; Xiong, T.; Wang, S. A review of Ribosome profiling and tools used in Ribo-seq data analysis. *Computational and Structural Biotechnology Journal* **2024**, *23*, 1912–1918, doi:10.1016/j.csbj.2024.04.051.
102. Kole, R.; Krainer, A.R.; Altman, S. RNA therapeutics: beyond RNA interference and antisense oligonucleotides. *Nature Reviews Drug Discovery* **2012**, *11*, 125–140, doi:10.1038/nrd3625.
103. Pifer, R.; Greenberg, D.E. Antisense antibacterial compounds. *Translational Research* **2020**, *223*, 89–106, doi:<https://doi.org/10.1016/j.trsl.2020.06.001>.
104. Hör, J.; Jung, J.; Đurica-Mitić, S.; Barquist, L.; Vogel, J. INRI-seq enables global cell-free analysis of translation initiation and off-target effects of antisense inhibitors. *Nucleic Acids Research* **2022**, *50*, e128–e128, doi:10.1093/nar/gkac838.
105. Koch, A.; Gawron, D.; Steyaert, S.; Ndah, E.; Crappé, J.; De Keulenaer, S.; De Meester, E.; Ma, M.; Shen, B.; Gevaert, K.; et al. A proteogenomics approach integrating proteomics and ribosome profiling increases the efficiency of protein identification and enables the discovery of alternative translation start sites. *PROTEOMICS* **2014**, *14*, 2688–2698, doi:<https://doi.org/10.1002/pmic.201400180>.
106. Berry, I.J.; Steele, J.R.; Padula, M.P.; Djordjevic, S.P. The application of terminomics for the identification of protein start sites and proteoforms in bacteria. *PROTEOMICS* **2016**, *16*, 257–272, doi:<https://doi.org/10.1002/pmic.201500319>.
107. Gevaert, K.; Goethals, M.; Martens, L.; Van Damme, J.; Staes, A.; Thomas, G.R.; Vandekerckhove, J. Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides. *Nature Biotechnology* **2003**, *21*, 566–569, doi:10.1038/nbt810.
108. Staes, A.; Van Damme, P.; Helsens, K.; Demol, H.; Vandekerckhove, J.; Gevaert, K. Improved recovery of proteome-informative, protein N-terminal peptides by combined fractional diagonal chromatography (COFRADIC). *PROTEOMICS* **2008**, *8*, 1362–1370, doi:<https://doi.org/10.1002/pmic.200700950>.
109. Kleifeld, O.; Doucet, A.; auf dem Keller, U.; Prudova, A.; Schilling, O.; Kainthan, R.K.; Starr, A.E.; Foster, L.J.; Kizhakkedathu, J.N.; Overall, C.M. Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products. *Nature Biotechnology* **2010**, *28*, 281–288, doi:10.1038/nbt.1611.
110. Hanna, R.; Rozenberg, A.; Saied, L.; Ben-Yosef, D.; Lavy, T.; Kleifeld, O. In-Depth Characterization of Apoptosis N-Terminome Reveals a Link Between Caspase-3 Cleavage and Posttranslational N-Terminal Acetylation. *Molecular & Cellular Proteomics* **2023**, *22*, 100584, doi:<https://doi.org/10.1016/j.mcpro.2023.100584>.
111. Van Damme, P. TRAINSPOTTER: Profiling Nascent Protein N-Termini Indicative of Bacterial Translation Initiation via Deformylation-Assisted N-Terminomics *Nucleic Acids Research* **2026**, *Accepted Author Manuscript*.

112. Caserta, E.; Tomšic, J.; Spurio, R.; La Teana, A.; Pon, C.L.; Gualerzi, C.O. Translation Initiation Factor IF2 Interacts with the 30 S Ribosomal Subunit via Two Separate Binding Sites. *Journal of Molecular Biology* **2006**, *362*, 787–799, doi:<https://doi.org/10.1016/j.jmb.2006.07.043>.
113. North, S.H.; Kirtland, S.E.; Nakai, H. Translation factor IF2 at the interface of transposition and replication by the PriA-PriC pathway. *Molecular Microbiology* **2007**, *66*, 1566–1578, doi:<https://doi.org/10.1111/j.1365-2958.2007.06022.x>.
114. Kato, J.-i.; Suzuki, H.; Hirota, Y. Overlapping of the coding regions for α and γ components of penicillin-binding protein 1 b in *Escherichia coli*. *Molecular and General Genetics MGG* **1984**, *196*, 449–457, doi:10.1007/BF00436192.
115. Park, S.K.; Kim, K.I.; Woo, K.M.; Seol, J.H.; Tanaka, K.; Ichihara, A.; Ha, D.B.; Chung, C.H. Site-directed mutagenesis of the dual translational initiation sites of the clpB gene of *Escherichia coli* and characterization of its gene products. *J Biol Chem* **1993**, *268*, 20170–20174.
116. Beinker, P.; Schlee, S.; Groemping, Y.; Seidel, R.; Reinstein, J. The N Terminus of ClpB from *Thermus thermophilus* Is Not Essential for the Chaperone Activity. *Journal of Biological Chemistry* **2002**, *277*, 47160–47166, doi:10.1074/jbc.M207853200.
117. Nagy, M.; Guenther, I.; Akoyev, V.; Barnett, M.E.; Zavodszky, M.I.; Kedzierska-Mieszkowska, S.; Zolkiewski, M. Synergistic Cooperation between Two ClpB Isoforms in Aggregate Reactivation. *Journal of Molecular Biology* **2010**, *396*, 697–707, doi:<https://doi.org/10.1016/j.jmb.2009.11.059>.
118. Schlegel, S.; Rujas, E.; Ytterberg, A.J.; Zubarev, R.A.; Luirink, J.; de Gier, J.-W. Optimizing heterologous protein production in the periplasm of *E. coli* by regulating gene expression levels. *Microbial Cell Factories* **2013**, *12*, 24, doi:10.1186/1475-2859-12-24.
119. Bhattacharyya, S.; Bershtein, S.; Yan, J.; Argun, T.; Gilson, A.I.; Trauger, S.A.; Shakhnovich, E.I. Transient protein-protein interactions perturb *E. coli* metabolome and cause gene dosage toxicity. *eLife* **2016**, *5*, e20309, doi:10.7554/eLife.20309.
120. Omairi-Nasser, A.; de Gracia, A.G.; Ajlani, G. A larger transcript is required for the synthesis of the smaller isoform of ferredoxin:NADP oxidoreductase. *Molecular Microbiology* **2011**, *81*, 1178–1189, doi:<https://doi.org/10.1111/j.1365-2958.2011.07739.x>.
121. Thomas, J.-C.; Ughy, B.; Lagoutte, B.; Ajlani, G. A second isoform of the ferredoxin:NADP oxidoreductase generated by an in-frame initiation of translation. *Proceedings of the National Academy of Sciences* **2006**, *103*, 18368–18373, doi:10.1073/pnas.0607718103.
122. Giuliodori, A.M.; Brandi, A.; Gualerzi, C.O.; Pon, C.L. Preferential translation of cold-shock mRNAs during cold adaptation. *Rna* **2004**, *10*, 265–276, doi:10.1261/rna.5164904.
123. Squires, C.L.; Pedersen, S.; Ross, B.M.; Squires, C. ClpB is the *Escherichia coli* heat shock protein F84.1. *Journal of Bacteriology* **1991**, *173*, 4254–4262, doi:10.1128/jb.173.14.4254-4262.1991.
124. Yoshida, A.; Tomita, T.; Kuzuyama, T.; Nishiyama, M. Mechanism of Concerted Inhibition of $\alpha\beta_2$ -type Hetero-oligomeric Aspartate Kinase from *Corynebacterium glutamicum*. *Journal of Biological Chemistry* **2010**, *285*, 27477–27486, doi:10.1074/jbc.M110.111153.
125. Wein, T.; Hülter, N.F.; Mizrahi, I.; Dagan, T. Emergence of plasmid stability under non-selective conditions maintains antibiotic resistance. *Nature Communications* **2019**, *10*, 2595, doi:10.1038/s41467-019-10600-7.
126. Schwinn, M.K.; Machleidt, T.; Zimmerman, K.; Eggers, C.T.; Dixon, A.S.; Hurst, R.; Hall, M.P.; Encell, L.P.; Binkowski, B.F.; Wood, K.V. CRISPR-Mediated Tagging of Endogenous Proteins with a Luminescent Peptide. *ACS Chemical Biology* **2018**, *13*, 467–474, doi:10.1021/acscchembio.7b00549.
127. Einhauer, A.; Jungbauer, A. The FLAG™ peptide, a versatile fusion tag for the purification of recombinant proteins. *Journal of Biochemical and Biophysical Methods* **2001**, *49*, 455–465, doi:[https://doi.org/10.1016/S0165-022X\(01\)00213-5](https://doi.org/10.1016/S0165-022X(01)00213-5).
128. Penfold, R.J.; Pemberton, J.M. An improved suicide vector for construction of chromosomal insertion mutations in bacteria. *Gene* **1992**, *118*, 145–146, doi:[https://doi.org/10.1016/0378-1119\(92\)90263-O](https://doi.org/10.1016/0378-1119(92)90263-O).
129. Fels, U.; Gevaert, K.; Van Damme, P. Bacterial Genetic Engineering by Means of Recombineering for Reverse Genetics. *Frontiers in Microbiology* **2020**, *Volume 11 - 2020*, doi:10.3389/fmicb.2020.548410.

130. Wang, H.H.; Xu, G.; Vonner, A.J.; Church, G. Modified bases enable high-efficiency oligonucleotide-mediated allelic replacement via mismatch repair evasion. *Nucleic Acids Research* **2011**, *39*, 7336–7347, doi:10.1093/nar/gkr183.
131. Brandis, G.; Cao, S.; Hughes, D. Operon Concatenation Is an Ancient Feature That Restricts the Potential to Rearrange Bacterial Chromosomes. *Molecular Biology and Evolution* **2019**, *36*, 1990–2000, doi:10.1093/molbev/msz129.
132. Gao, G.; Le, D.; Huang, L.; Lu, H.; Narumi, I.; Hua, Y. Internal promoter characterization and expression of the *Deinococcus radiodurans* ppri-folP gene cluster. *FEMS Microbiology Letters* **2006**, *257*, 195–201, doi:10.1111/j.1574-6968.2006.00169.x.
133. Knöppel, A.; Näsval, J.; Andersson, D.I. Compensating the Fitness Costs of Synonymous Mutations. *Molecular Biology and Evolution* **2016**, *33*, 1461–1477, doi:10.1093/molbev/msw028.
134. K. E. Köpke, A.; A. Leggatt, P. Initiation of translation at an AUA codon for an archaeobacterial protein gene expressed in *E.coli*. *Nucleic Acids Research* **1991**, *19*, 5169–5172, doi:10.1093/nar/19.19.5169.
135. Johnson, Z.I.; Chisholm, S.W. Properties of overlapping genes are conserved across microbial genomes. *Genome Research* **2004**, *14*, 2268–2272, doi:10.1101/gr.2433104.
136. Huber, M.; Vogel, N.; Borst, A.; Pfeiffer, F.; Karamycheva, S.; Wolf, Y.I.; Koonin, E.V.; Soppa, J. Unidirectional gene pairs in archaea and bacteria require overlaps or very short intergenic distances for translational coupling via termination-reinitiation and often encode subunits of heteromeric complexes. *Frontiers in Microbiology* **2023**, Volume 14 - 2023, doi:10.3389/fmicb.2023.1291523.
137. Brown, K.M.; Wade, J.T. Translational coupling of neighboring genes in prokaryotes. *Journal of Bacteriology* **2025**, *207*, e00255–00225, doi:doi:10.1128/jb.00255-25.
138. Salis, H.M. Chapter two - The Ribosome Binding Site Calculator. In *Methods in Enzymology*, Voigt, C., Ed.; Academic Press: 2011; Volume 498, pp. 19–42.
139. Wang, H.H.; Kim, H.; Cong, L.; Jeong, J.; Bang, D.; Church, G.M. Genome-scale promoter engineering by coselection MAGE. *Nature Methods* **2012**, *9*, 591–593, doi:10.1038/nmeth.1971.
140. Kwok, P.-Y. Methods for Genotyping Single Nucleotide Polymorphisms. *Annual Review of Genomics and Human Genetics* **2001**, *2*, 235–258, doi:<https://doi.org/10.1146/annurev.genom.2.1.235>.
141. Ellis, H.M.; Yu, D.; DiTizio, T.; Court, D.L. High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proceedings of the National Academy of Sciences* **2001**, *98*, 6742–6746, doi:doi:10.1073/pnas.121164898.
142. Snauwaert, V.; Jonckheere, V.; Van Damme, P. Unraveling N-Terminal Proteoform Interactomes via Multiplexed Recombineering in *Salmonella*. In *Proximity-Dependent Protein Biotinylation: Methods and Protocols*, Van Damme, P., Ed.; Springer US: New York, NY, 2025; pp. 81–102.
143. Subbarayan, P.R.; Sarkar, M. A stop codon-dependent internal secondary translation initiation region in *Escherichia coli* rpoS. *Rna* **2004**, *10*, 1359–1365, doi:10.1261/rna.7500604.
144. Jiang, W.; Bikard, D.; Cox, D.; Zhang, F.; Marraffini, L.A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature Biotechnology* **2013**, *31*, 233–239, doi:10.1038/nbt.2508.
145. Li, Y.; Lin, Z.; Huang, C.; Zhang, Y.; Wang, Z.; Tang, Y.-j.; Chen, T.; Zhao, X. Metabolic engineering of *Escherichia coli* using CRISPR-Cas9 mediated genome editing. *Metabolic Engineering* **2015**, *31*, 13–21, doi:<https://doi.org/10.1016/j.ymben.2015.06.006>.
146. Huang, C.; Ding, T.; Wang, J.; Wang, X.; Guo, L.; Wang, J.; Zhu, L.; Bi, C.; Zhang, X.; Ma, X.; et al. CRISPR-Cas9-assisted native end-joining editing offers a simple strategy for efficient genetic engineering in *Escherichia coli*. *Applied Microbiology and Biotechnology* **2019**, *103*, 8497–8509, doi:10.1007/s00253-019-10104-w.
147. Huang, C.; Guo, L.; Wang, J.; Wang, N.; Huo, Y.-X. Efficient long fragment editing technique enables large-scale and scarless bacterial genome engineering. *Applied Microbiology and Biotechnology* **2020**, *104*, 7943–7956, doi:10.1007/s00253-020-10819-1.
148. Reisch, C.R.; Prather, K.L.J. The no-SCAR (Scarless Cas9 Assisted Recombineering) system for genome editing in *Escherichia coli*. *Scientific Reports* **2015**, *5*, 15096, doi:10.1038/srep15096.
149. Ronda, C.; Pedersen, L.E.; Sommer, M.O.A.; Nielsen, A.T. CRMAGE: CRISPR Optimized MAGE Recombineering. *Scientific Reports* **2016**, *6*, 19452, doi:10.1038/srep19452.

150. Lee, H.J.; Kim, H.J.; Lee, S.J. CRISPR-Cas9-mediated pinpoint microbial genome editing aided by target-mismatched sgRNAs. *Genome Research* **2020**, *30*, 768–775, doi:10.1101/gr.257493.119.
151. Lim, S.R.; Lee, H.J.; Kim, H.J.; Lee, S.J. Multiplex Single-Nucleotide Microbial Genome Editing Achieved by CRISPR-Cas9 Using 5'-End-Truncated sgRNAs. *ACS Synthetic Biology* **2023**, *12*, 2203–2207, doi:10.1021/acssynbio.3c00323.
152. Elsayed, E.M.; Stukenberg, D.; Meier, D.; Schmeck, B.; Becker, A. RECKLEEN is a lambda Red/CRISPR-Cas9 based single plasmid platform for enhanced genome editing in *Klebsiella pneumoniae*. *Communications Biology* **2025**, *8*, 1509, doi:10.1038/s42003-025-08934-8.
153. Vento, J.M.; Crook, N.; Beisel, C.L. Barriers to genome editing with CRISPR in bacteria. *Journal of Industrial Microbiology and Biotechnology* **2019**, *46*, 1327–1341, doi:10.1007/s10295-019-02195-1.
154. Roux, K.J.; Kim, D.I.; Raida, M.; Burke, B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *Journal of Cell Biology* **2012**, *196*, 801–810, doi:10.1083/jcb.201112098.
155. Kim, D.I.; Jensen, S.C.; Noble, K.A.; KC, B.; Roux, K.H.; Motamedchaboki, K.; Roux, K.J.; Zheng, Y. An improved smaller biotin ligase for BioID proximity labeling. *Molecular Biology of the Cell* **2016**, *27*, 1188–1196, doi:doi:10.1091/mbc.E15-12-0844.
156. Branon, T.C.; Bosch, J.A.; Sanchez, A.D.; Udeshi, N.D.; Svinkina, T.; Carr, S.A.; Feldman, J.L.; Perrimon, N.; Ting, A.Y. Efficient proximity labeling in living cells and organisms with TurboID. *Nature Biotechnology* **2018**, *36*, 880–887, doi:10.1038/nbt.4201.
157. Herfurth, M.; Müller, F.; Søgaard-Andersen, L.; Glatter, T. A miniTurbo-based proximity labeling protocol to identify conditional protein interactomes in vivo in *Myxococcus xanthus*. *STAR Protocols* **2023**, *4*, 102657, doi:<https://doi.org/10.1016/j.xpro.2023.102657>.
158. Remy, O.; Santin, Y.G.; Jonckheere, V.; Tesseur, C.; Kaljević, J.; Damme, P.V.; Laloux, G. Distinct dynamics and proximity networks of hub proteins at the prey-invading cell pole in a predatory bacterium. *Journal of Bacteriology* **2024**, *206*, e00014–00024, doi:doi:10.1128/jb.00014-24.
159. Jonckheere, V.; Van Damme, P. N-Terminal Acetyltransferase Naa40p Whereabouts Put into N-Terminal Proteoform Perspective. *International Journal of Molecular Sciences* **2021**, *22*, 3690.
160. Bogaert, A.; Fijalkowska, D.; Staes, A.; Van de Steene, T.; Vuylsteke, M.; Stadler, C.; Eyckerman, S.; Spirohn, K.; Hao, T.; Calderwood, M.A.; et al. N-terminal proteoforms may engage in different protein complexes. *Life Science Alliance* **2023**, *6*, e202301972, doi:10.26508/lsa.202301972.
161. Schwanhäusser, B.; Busse, D.; Li, N.; Dittmar, G.; Schuchhardt, J.; Wolf, J.; Chen, W.; Selbach, M. Global quantification of mammalian gene expression control. *Nature* **2011**, *473*, 337–342, doi:10.1038/nature10098.
162. Jayapal, K.P.; Sui, S.; Philp, R.J.; Kok, Y.-J.; Yap, M.G.S.; Griffin, T.J.; Hu, W.-S. Multitagging Proteomic Strategy to Estimate Protein Turnover Rates in Dynamic Systems. *Journal of Proteome Research* **2010**, *9*, 2087–2097, doi:10.1021/pr9007738.
163. Fierro-Monti, I.; Racle, J.; Hernandez, C.; Waridel, P.; Hatzimanikatis, V.; Quadroni, M. A Novel Pulse-Chase SILAC Strategy Measures Changes in Protein Decay and Synthesis Rates Induced by Perturbation of Proteostasis with an Hsp90 Inhibitor. *PLOS ONE* **2013**, *8*, e80423, doi:10.1371/journal.pone.0080423.
164. Boisvert, F.-M.; Ahmad, Y.; Gierliński, M.; Charrière, F.; Lamont, D.; Scott, M.; Barton, G.; Lamond, A.I. A Quantitative Spatial Proteomics Analysis of Proteome Turnover in Human Cells. *Molecular & Cellular Proteomics* **2012**, *11*, doi:10.1074/mcp.M111.011429.
165. Han, J.; Yi, S.; Zhao, X.; Zheng, Y.; Yang, D.; Du, G.; Yang, X.-Y.; He, Q.-Y.; Sun, X. Improved SILAC method for double labeling of bacterial proteome. *Journal of Proteomics* **2019**, *194*, 89–98, doi:<https://doi.org/10.1016/j.jprot.2018.12.011>.
166. Gupta, M.; Johnson, A.N.T.; Cruz, E.R.; Costa, E.J.; Guest, R.L.; Li, S.H.-J.; Hart, E.M.; Nguyen, T.; Stadlmeier, M.; Bratton, B.P.; et al. Global protein turnover quantification in *Escherichia coli* reveals cytoplasmic recycling under nitrogen limitation. *Nature Communications* **2024**, *15*, 5890, doi:10.1038/s41467-024-49920-8.
167. Xue, Y.; Sherman, D.H. Alternative modular polyketide synthase expression controls macrolactone structure. *Nature* **2000**, *403*, 571–575, doi:10.1038/35000624.

168. Bohlin, J.; Pettersson, J.H.-O. Evolution of Genomic Base Composition: From Single Cell Microbes to Multicellular Animals. *Computational and Structural Biotechnology Journal* **2019**, *17*, 362–370, doi:doi:10.1016/j.csbj.2019.03.001.
169. Pan, T.; Sosnick, T. RNA FOLDING DURING TRANSCRIPTION. *Annual Review of Biophysics* **2006**, *35*, 161–175, doi:<https://doi.org/10.1146/annurev.biophys.35.040405.102053>.
170. Chow, I.T.; Baneyx, F. Coordinated synthesis of the two ClpB isoforms improves the ability of *Escherichia coli* to survive thermal stress. *FEBS Letters* **2005**, *579*, 4235–4241, doi:<https://doi.org/10.1016/j.febslet.2005.06.054>.
171. Ozin, A.J.; Costa, T.; Henriques, A.O.; Moran, C.P. Alternative Translation Initiation Produces a Short Form of a Spore Coat Protein in *Bacillus subtilis*. *Journal of Bacteriology* **2001**, *183*, 2032–2040, doi:doi:10.1128/jb.183.6.2032-2040.2001.
172. Yu, X.-J.; Liu, M.; Matthews, S.; Holden, D.W. Tandem Translation Generates a Chaperone for the *Salmonella* Type III Secretion System Protein SsaQ*. *Journal of Biological Chemistry* **2011**, *286*, 36098–36107, doi:<https://doi.org/10.1074/jbc.M111.278663>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.