



Zakład Biologii Obliczeniowej  
Instytut Biologii Molekularnej i Biotechnologii  
Wydział Biologii  
Uniwersytet im. Adama Mickiewicza w Poznaniu

**Jan Grzegorz Kosiński**

**Identyfikacja i charakterystyka rho-niezależnych terminatorów transkrypcji u bakterii  
oraz ich udziału w odpowiedzi na stres antybiotykowy**

Identification and characteristics of the intrinsic transcription terminators in bacteria and their  
contribution to the antibiotic stress response

Praca doktorska przygotowana pod kierunkiem  
**prof. UAM dr hab. Marka Żywickiego**

Poznań, 2024



*Dedykowane mojej żonie, Paulinie,  
za nieocenione wsparcie i nieustanną wiarę we mnie.*

## Spis treści

1.	Streszczenie	5
2.	Abstract	8
3.	Źródła finansowania	10
4.	Lista publikacji wchodzących w skład rozprawy	11
5.	Dorobek publikacyjny autora	12
6.	Wykaz stosowanych skrótów	14
7.	Wprowadzenie	17
7.1.	Centralny dogmat biologii molekularnej	19
7.2.	Synteza cząsteczek RNA u bakterii	22
7.3.	Rola cząsteczek RNA w komórkach bakteryjnych	27
7.4.	Wpływ struktur drugorzędowych rna na regulację ekspresji genów	29
7.5.	Terminacja transkrypcji a problem lekooporności bakterii	30
7.6.	Molekularne mechanizmy lekooporności gronkowca złocistego	38
7.7.	Metody identyfikacji miejsc terminacji transkrypcji	46
8.	Cel pracy	51
9.	Metody	52
9.1.	Wykorzystane zestawy danych	52
9.2.	Identyfikacja rho-niezależnych terminatorów transkrypcji na podstawie danych Term-seq	54
9.3.	Przewidywanie rho-niezależnych terminatorów transkrypcji na podstawie danych RNA-seq u gronkowca złocistego	55
9.4.	Utworzenie referencyjnego atlasu rho-niezależnych terminatorów transkrypcji u gronkowca złocistego	57
9.5.	Identyfikacja miejsc warunkowej terminacji transkrypcji w odpowiedzi na stres antybiotykowy u gronkowca złocistego	58
10.	Wyniki i dyskusja	60
10.1.	<i>TERMITE</i> – metoda identyfikacji i adnotacji stabilnych 3' końców RNA na podstawie wyników sekwencjonowania z zastosowaniem protokołu Term-seq	60
10.2.	<i>termiRNator</i> – metoda przewidywania miejsc terminacji transkrypcji na podstawie danych RNA-seq wykorzystująca algorytmy uczenia maszynowego	62
10.3.	<i>AGouTI</i> – metoda adnotacji interwałów genomowych oraz transkryptomowych z wykorzystaniem informacji zawartych w plikach gtf lub gff	70
10.4.	Atlas rho-niezależnych terminatorów transkrypcji u metycylooopornych szczepów gronkowca złocistego	71
10.5.	Charakterystyka rho-niezależnych terminatorów transkrypcji u gronkowca złocistego	73
10.6.	Warunkowa terminacja transkrypcji w odpowiedzi na stres antybiotykowy	90
11.	Podsumowanie	99
12.	Spis literatury	102
13.	Załączniki	109
14.	Oświadczenia współautorów	130

## 1. Streszczenie

Transkrypcja jest procesem, który umożliwia propagację informacji genetycznej z poziomu DNA do RNA. Mimo że pełne spektrum funkcji pełnionych przez cząsteczki RNA nie zostało jeszcze w pełni poznane, wiadomo, że różnorodność mechanizmów i procesów, w których są one bezpośrednio zaangażowane, jest nadzwyczajna. Pomimo pełnienia roli matrycy niezbędnej do syntezy białek, cząsteczki RNA pełnią także szeroki zakres funkcji regulacyjnych wewnątrz komórki. Dlatego też kluczowe jest pełne zrozumienie sposobów regulacji tych procesów oraz praktyczne wykorzystanie tej wiedzy, na przykład w leczeniu chorób zakaźnych.

W bakteriach cząsteczki RNA mogą uczestniczyć w regulacji odpowiedzi komórkowych na różne bodźce środowiskowe lub fizjologiczne przy pomocy kilku dobrze poznanych mechanizmów. RNA może związać się w złożone i aktywnie zmieniające się struktury drugorzędowe, które tworzą się poprzez parowanie zasad z innymi przyległymi lub odległymi częściami tej samej cząsteczki. Zmiany struktur drugorzędowych RNA mogą modulować szereg procesów związanych z ekspresją genów, zazwyczaj poprzez wpływanie na aktywność polimeraz RNA, rybosomów i innych kluczowych enzymów lub kompleksów. Indukowane zmiany w strukturze drugorzędowej RNA mogą również w konsekwencji prowadzić do przedwczesnej terminacji transkrypcji.

Niniejsza rozprawa prezentuje szereg innowacyjnych narzędzi bioinformatycznych, opracowanych w celu wykrywania i adnotacji rho-niezależnych terminatorów transkrypcji u bakterii. Wykorzystują one dane pochodzące z wysokoprzepustowych technik sekwencjonowania 3' końców RNA (Term-seq) oraz całkowitego sekwencjonowania RNA (RNA-seq).

Pierwsze z prezentowanych w rozprawie narzędzi ma na celu identyfikację stabilnych 3' końców RNA w ramach sygnału reprezentowanego jako pokrycie genomu przez odczyty uzyskiwane metodą Term-seq. Opracowany algorytm, wykorzystując metodę IDR (ang. *Irreproducible Discovery Rate*), pozwala skutecznie różnicować sygnały zachowawcze pomiędzy replikatami biologicznymi, od pozostałych, w znacznej mierze reprezentujących pośrednie produkty przetwarzania RNA lub jego degradacji. Opracowane oprogramowanie umożliwia także precyzyjną adnotację szeregu cech związanych z 3' końcami cząsteczek RNA, w szczególności identyfikując te będące wynikiem rho-niezależnej terminacji transkrypcji.

Drugi z opracowanych algorytmów służy do przewidywania miejsc rho-niezależnej terminacji transkrypcji na podstawie danych z sekwencjonowania RNA (RNA-seq) oraz metod uczenia maszynowego. Do utworzenia modelu predykcyjnego wykorzystane zostały publicznie dostępne zbiory odczytów, uzyskanych przy pomocy protokołów RNA-seq oraz Term-seq. Zastosowanie wspomnianych narzędzi bioinformatycznych do analizy szerokiego zestawu publicznie dostępnych danych umożliwiło stworzenie niezwykle rozbudowanego atlasu rho-niezależnych terminatorów bakteryjnych, zidentyfikowanych przy pomocy technik wysokoprzepustowego sekwencjonowania.

Rolą trzeciego z przygotowanych programów jest adnotacja interwałów genomowych lub transkryptomowych przy pomocy informacji zawartych w plikach GTF lub GFF. Wymienione formaty plików służą do przechowywania adnotacji genów, transkryptów oraz innych cech genomowych poznanych w obrębie danego genomu. Zastosowanie niniejszego narzędzia pozwoliło uzupełnić zestaw adnotacji przedstawionych w ramach opracowanego atlasu rho-niezależnych terminatorów bakteryjnych.

Prezentowana praca przedstawia kompleksowe porównanie terminatorów, występujących u wybranych gatunków bakterii, ujawniając ich znaczącą zmienność między różnymi taksonami. Obserwacja ta poddaje w wątpliwość dominujące przeświadczenie mówiące o znacznym podobieństwie sekwencji i struktur terminatorów występujących u bakterii. Przeświadczenie to jest głównie wynikiem badań przeprowadzonych na organizmach modelowych, takich jak *Escherichia coli* i *Bacillus subtilis*. Przedstawiona praca szczegółowo omawia zauważalne różnice w strukturze drugorzędowej, kompozycji nukleotydowej, a także lokalizacji genomowej zidentyfikowanych rho-niezależnych terminatorów transkrypcji.

Niezwykle istotnym elementem prezentowanej rozprawy jest starannie przygotowany atlas miejsc rho-niezależnej terminacji transkrypcji u *Staphylococcus aureus*, bakterii o wysokim znaczeniu klinicznym, charakteryzującej się szczepami wykazującymi oporność na wiele znanych leków. Atlas ten został opracowany na podstawie analizy 140 publicznie dostępnych zbiorów danych, stanowiąc tym samym według najlepszej wiedzy autora, najbardziej wszechstronne zestawienie terminatorów opracowane do dziś dla *S. aureus*.

Ponadto, niniejsza praca dyplomowa prezentuje wyniki, które ujawniają kilka kontekstowo specyficznych zdarzeń terminacji transkrypcji, prowadzących do regulacji ekspresji genów w odpowiedzi na stres wywołany stosowaniem antybiotyków u *S. aureus*. Wyniki te mają potencjalne znaczenie terapeutyczne, pozwalając szczegółowo zrozumieć wpływ stosowanych antybiotyków na bakterie stanowiące wysokie zagrożenie kliniczne.

**Słowa kluczowe:** rho-niezależna terminacja transkrypcji, warunkowa terminacja transkrypcji, Term-seq, *Staphylococcus aureus*, stres antybiotykowy

## 2. Abstract

Transcription is an essential process allowing the propagation of genetic information from DNA to RNA. The full spectrum of RNA roles in cells is yet to be uncovered. However, it is already known that the variability of mechanisms and processes in which RNA molecules are involved is extraordinary. Despite acting as a template for the synthesis of proteins, RNAs perform a wide array of regulatory functions within the cell. Therefore, it is crucial to fully understand how RNA molecules regulate gene expression in bacteria and how people can use this knowledge to their advantage, e.g., in treating infectious diseases.

In bacteria, RNA molecules can modulate cellular responses to various environmental or physiological stimuli via several well-understood mechanisms. RNAs can fold into sophisticated and actively changing secondary structures forming through base-pairing with other adjacent or distant parts of the same molecule. RNA secondary structures can modulate a range of processes involved in gene expression, usually by affecting the work of RNA polymerases, ribosomes, and other crucial enzymes or complexes. One of the most common mechanisms of regulating gene expression is via changes in the RNA secondary structure leading to premature transcription termination.

This thesis introduces innovative bioinformatic tools developed to detect and annotate intrinsic transcription terminators in bacterial organisms, utilizing 3'-end RNA sequencing (Term-seq) and total RNA sequencing (RNA-seq) data.

The first tool aims to identify stable 3' RNA ends within the signal represented by the genomic coverage of 5' Term-seq read ends. A novel algorithm, based on Irreproducible Discovery Rates, is employed to effectively distinguish signals being reproducible between biological replicates from others, likely attributed to intermediate products of RNA processing or degradation. Moreover, the provided software enables the annotation of various features associated with the 3' ends of RNA molecules, specifically identifying those resulting from intrinsic transcription termination.

The second algorithm introduced as a part of this thesis is a machine-learning method for accurate prediction of intrinsic termination events based on RNA-seq data, leveraging publicly available datasets from RNA-seq and Term-seq protocols. The development of this toolkit facilitated the creation of a comprehensive atlas of intrinsic bacterial terminators, validated through high-throughput sequencing techniques.



The role of the third presented software is to annotate genomic or transcriptomic intervals using information from GTF or GFF files. These file formats are used to store annotations of genes, transcripts, and other genomic features identified within a given genome. The application of this tool allowed for the completion of the annotations presented in the atlas of intrinsic bacterial terminators.

This study's author has comprehensively compared terminators across the bacterial kingdom through computational analysis, revealing substantial variability between different taxa. This finding challenges the prevailing assumption of their sequence and structural similarity that has been primarily derived from research conducted on model organisms like *Escherichia coli* and *Bacillus subtilis*. The thesis extensively discusses the notable differences observed in intrinsic terminators' secondary structure, nucleotide composition, and genomic location.

Significantly, the author presents a meticulous atlas of intrinsic termination events in *Staphylococcus aureus*, a highly clinically relevant bacterium known for its multidrug resistance. Based on the analysis of 140 publicly available datasets, this atlas represents, to the best of the author's knowledge, the most comprehensive compilation of *S. aureus* terminators compiled to date.

Furthermore, the thesis introduces additional results unveiling several context-specific transcription termination events governing gene expression in response to antibiotic-induced stress in *S. aureus*. These findings allow us to better understand the therapeutic implications of specific antibiotic treatments for bacteria of high clinical importance.

**Keywords:** intrinsic transcription termination, conditional transcription termination, Term-seq, *Staphylococcus aureus*, antibiotic stress

### **3. Źródła finansowania**

Niniejsza praca powstała przy wsparciu finansowym Narodowego Centrum Nauki [2017/25/B/NZ6/00642 dla Marka Żywickiego], Wydziału Biologii Uniwersytetu im. Adama Mickiewicza w Poznaniu [GDWB-11/2020, POWR.0302.00-00-I006/17 dla Jana Grzegorza Kosińskiego] oraz programu Uniwersytet Jutra II realizowanego na Uniwersytecie im. Adama Mickiewicza w Poznaniu [POWR.03.05.00-00-Z303/18 dla Jana Grzegorza Kosińskiego].

#### 4. Lista publikacji wchodzących w skład rozprawy

- **Kosiński JG**, Ranaweera S, Chełkowska-Pauszek A, Kashlev M, Babitzke P, Żywicki M. *Characterization of bacterial intrinsic transcription terminators identified with TERMITE – a novel method for comprehensive analysis of Term-seq data*. bioRxiv. 2024 May. doi: 10.1101/2024.05.16.594410
- **Kosiński JG**, Żywicki M. *AGouTI - Flexible Annotation of Genomic and Transcriptomic Intervals*. PLoS Comput Biol. 2023 Oct 18;19(10):e1011527. doi: 10.1371/journal.pcbi.1011527. PMID: 37851672; PMCID: PMC10621952
- Chełkowska-Pauszek A \*, **Kosiński JG** \*, Marciniak K \*, Wysocka M \*, Bąkowska-Żywicka K, Żywicki M. *The Role of RNA Secondary Structure in Regulation of Gene Expression in Bacteria*. Int J Mol Sci. 2021 Jul 22;22(15):7845. doi: 10.3390/ijms22157845. PMID: 34360611; PMCID: PMC8346122

## 5. Dorobek publikacyjny autora

- **Kosiński JG**, Ranaweera S, Chełkowska-Pauszek A, Kashlev M, Babitzke P, Żywicki M. *Characterization of bacterial intrinsic transcription terminators identified with TERMITE – a novel method for comprehensive analysis of Term-seq data*. bioRxiv. 2024 May. doi: 10.1101/2024.05.16.594410
- **Kosiński JG**, Żywicki M. *AGouTI - Flexible Annotation of Genomic and Transcriptomic Intervals*. PLoS Comput Biol. 2023 Oct 18;19(10):e1011527. doi: 10.1371/journal.pcbi.1011527. PMID: 37851672; PMCID: PMC10621952
- Pierański MK, **Kosiński JG**, Szymczak K, Sadowski P, Grinholc M. *Antimicrobial Photodynamic Inactivation: An Alternative for Group B Streptococcus Vaginal Colonization in a Murine Experimental Model*. Antioxidants (Basel). 2023 Apr 1;12(4):847. doi: 10.3390/antiox12040847. PMID: 37107222; PMCID: PMC10135335
- Rucińska A, Świerszcz S, Nobis M, Zubek S, Boczkowska M, Olszak M, **Kosiński JG**, Nowak S, Nowak A. *Is it possible to understand a book missing a quarter of the letters? Unveiling the belowground species richness of grasslands*. Agriculture, Ecosystems & Environment. 2022 Feb 1;324:107683. doi:10.1016/j.agee.2021.107683
- Chełkowska-Pauszek A \*, **Kosiński JG** \*, Marciniak K \*, Wysocka M \*, Bąkowska-Żywicka K, Żywicki M. *The Role of RNA Secondary Structure in Regulation of Gene Expression in Bacteria*. Int J Mol Sci. 2021 Jul 22;22(15):7845. doi: 10.3390/ijms22157845. PMID: 34360611; PMCID: PMC8346122
- Glazinska P, Kulasek M, Glinkowski W, Wysocka M, **Kosiński JG**. *LuluDB-The Database Created Based on Small RNA, Transcriptome, and Degradome Sequencing Shows the Wide Landscape of Non-coding and Coding RNA in Yellow Lupine (Lupinus luteus L.) Flowers and Pods*. Front Genet. 2020 May 15;11:455. doi: 10.3389/fgene.2020.00455. PMID: 32499815; PMCID: PMC7242762

- Glazińska P, Kulasek M, Glinkowski W, Wojciechowski W, **Kosiński J**. *Integrated Analysis of Small RNA, Transcriptome and Degradome Sequencing Provides New Insights into Floral Development and Abscission in Yellow Lupine (*Lupinus luteus* L.)*. *Int J Mol Sci*. 2019 Oct 16;20(20):5122. doi: 10.3390/ijms20205122. PMID: 31623090; PMCID: PMC6854478
- Rosikiewicz W, Kabza M, **Kosinski JG**, Ciomborowska-Basheer J, Kubiak MR, Makalowska I. *RetrogeneDB-a database of plant and animal retrocopies*. *Database (Oxford)*. 2017 Jan 1;2017:bax038. doi: 10.1093/database/bax038. PMID: 29220443; PMCID: PMC5509963

## 6. Wykaz stosowanych skrótów

**DNA** – ang. *deoxyribonucleic acid*; kwas deoksyrybonukleinowy

**RNA** – ang. *ribonucleic acid*; kwas rybonukleinowy

**cDNA** – ang. *complementary DNA*, komplementarne DNA; dwuniciowe DNA powstające w procesie odwrotnej transkrypcji na matrycy jednoniciowego RNA

**tRNA** – ang. *transfer RNA*; transportujący RNA

**rRNA** – ang. *ribosomal RNA*; rybosomalny RNA

**sRNA** – ang. *small RNA*; krótkie, niekodujące cząsteczki RNA pełniące w komórce rolę regulatorowe

**UP** – ang. *Upstream Promoter*; charakterystyczny region wchodzący w skład promotorów bakteryjnych zlokalizowany w kierunku 5' względem elementu -35

**CRE** – ang. *Core-Recognition Element*; charakterystyczny region wchodzący w skład promotorów bakteryjnych zlokalizowany na pozycjach od -4 do +2 (gdzie +1 oznacza miejsce startu transkrypcji)

**transportery ABC** – ang. *ATP-binding cassette*; jedna z rodzin białek transporterowych odpowiedzialnych za ATP-zależne transportowanie szkodliwych substancji poza obszar komórki bakteryjnej

**ncRNA** – ang. *non-coding RNA*; niekodujące cząsteczki RNA

**RBS** – ang. *ribosome-binding site*; miejsce wiązania rybosomu do RNA

**SEnd-seq** - ang. *simultaneous 5' and 3' end sequencing*; metoda równoczesnego sekwencjonowania 5' i 3' końców cząsteczek RNA u bakterii

**RNA-seq** – ang. *RNA sequencing*; metoda wysokoprzepustowego sekwencjonowania RNA

**Term-seq** - metoda wysokoprzepustowego sekwencjonowania 3' końców RNA u bakterii

**PCR** – ang. *polymerase chain reaction*; reakcja łańcuchowa polimerazy służąca do amplifikacji fragmentów DNA

**MRSA** – ang. *methicilin-resistant S. aureus*; szczep Gronkowca złocistego odporny na metycylinę

**SCCmec** - ang. *staphylococcal cassette chromosome mec*; mobilny element genomowy warunkujący oporność na penicyliny u *S. aureus*

**PBP2** – ang. *penicillin-binding protein*; białko wiążące penicylinę, odpowiedzialne za syntezę peptydoglikanu, będącego budulcem bakteryjnej ściany komórkowej; antybiotyki z rodziny penicylin blokują jego aktywność enzymatyczną

**VISA** – ang. *vancomycin-intermediate-resistant S. aureus*; szczepy *S. aureus* o zwiększonej oporności na wankomycynę

**VRSA** – ang. *vancomycin-resistant S. aureus*; szczepy *S. aureus* o wysokiej oporności na wankomycynę

**MIC** – ang. *minimum inhibitory concentration*; minimalne stężenie antybiotyku hamujące wzrost bakterii

**CPM** – ang. *counts per million mapped reads*; metoda normalizacji ekspresji genów. CPM reprezentuje liczbę fragmentów sekwencji RNA przypisanych do danego genu przypadających na jeden milion fragmentów obecnych w danej próbce.

**PP** – liczba poprawnie przewidzianych pozytywnych przypadków wśród wszystkich pozytywnych przypadków obecnych w zbiorze danych

**PN** – liczba poprawnie przewidzianych negatywnych przypadków wśród wszystkich negatywnych przypadków obecnych w zbiorze danych

**FP** – liczba przypadków negatywnych błędnie sklasyfikowanych jako pozytywne

**FN** – liczba przypadków pozytywnych błędnie sklasyfikowanych jako negatywne

**POT** – ang. *point of termination*; referencyjny punkt terminacji transkrypcji oznaczany przez program *TERMIte*

**BHI/TS** – pożywka stosowana do hodowli bakteryjnych *Brain Heart Infusion/Tryptone Soy*

**THB** - pożywka stosowana do hodowli bakteryjnych *Todd Hewitt Broth*

**CAMHB** - pożywka stosowana do hodowli bakteryjnych *Cation-Adjust Mueller Hinton*

**MHB** - pożywka stosowana do hodowli bakteryjnych *Mueller Hinton*

**RPLB** - pożywka stosowana do hodowli bakteryjnych *RPMI-1640* suplementowana 10% LB

W języku angielskim istnieją dwa sformułowania opisujące ‘rho-niezależne’ terminatory transkrypcji u bakterii. Są to odpowiednio określenia ‘rho-independent transcription terminators’ oraz ‘intrinsic transcription terminators’. Oba wyrażenia dotyczą tego samego mechanizmu terminacji, pierwsze z nich wyraźnie jednak podkreśla brak udziału białka *Rho* w tym procesie. Obecnie w debacie naukowej pojawia się coraz więcej doniesień, że białko *Rho* może także odgrywać rolę w rodzaju terminacji, który do tej pory był uznawany za niezależny od jego udziału. W języku polskim nie funkcjonuje natomiast odpowiednik określenia ‘intrinsic transcription terminator’, które wydaje się obecnie bardziej odpowiednie do zastosowania w niniejszej rozprawie. Mimo tego w niniejszej pracy używana jest obowiązująca nomenklatura (tj. sformułowanie ‘rho-niezależne terminatory transkrypcji’), która powinna zostać niezwłocznie dostosowana, jeżeli tylko opisywane doniesienia zostaną potwierdzone.



## 7. Wprowadzenie

W obrębie pojedynczej cząsteczki RNA mogą formować się niestałe struktury drugorzędowe, powstające na drodze wzajemnego wiązania się komplementarnych fragmentów. Struktura RNA może z kolei znacząco wpływać na proces ekspresji genów, regulując tym samym wiele szlaków komórkowych odpowiedzialnych za syntezę produktów niezbędnych do prawidłowego funkcjonowania organizmu (Chełkowska-Pauszek et al., 2021). Fałdowanie RNA może wpływać na szereg interakcji pomiędzy RNA a maszyną komórkową, warunkując sposoby i możliwości oddziaływania RNA, m. in. z polimerazami RNA, rybosomami oraz niektórymi białkami. Powszechnie wiadomym jest, że opisywane struktury drugorzędowe mogą regulować różnorodne etapy przepływu informacji genetycznej w systemach biologicznych, w tym procesy transkrypcji czy translacji (Chełkowska-Pauszek et al., 2021).

Proces terminacji transkrypcji u bakterii jest jednym z procesów ściśle warunkowanych obecnością określonych struktur drugorzędowych w obrębie cząsteczek RNA. W wielu przypadkach bakterie wykształciły mechanizmy, oparte o alternatywną (przedwczesną) terminację transkrypcji, regulujące wiele szlaków metabolicznych, w tym również pełniące kluczowe role w uzyskiwaniu przez bakterie oporności na działanie antybiotyków (Dersch et al., 2017). Elementy struktury drugorzędowej RNA oraz różnorodne procesy regulacyjne oparte o dynamikę RNA, mogą nie tylko służyć do „obrony” przed działaniem leków, ale także jako cel dla potencjalnych nowo opracowywanych klas antybiotyków. Zaburzenie procesów regulacji ekspresji genów, opartych o struktury drugorzędowe u bakterii, może prowadzić do destabilizacji komórki, a w konsekwencji także jej śmierci (Machtel et al., 2016).

Badanie mechanizmów regulujących ekspresję genów u bakterii w skali genomowej, w tym terminacji transkrypcji, nie byłoby możliwe, gdyby nie niezwykle dynamiczny rozwój metod sekwencjonowania nowej generacji, który możemy obserwować na przestrzeni ostatniej dekady. Sekwencjonowanie wysokoprzepustowe pozwala na dogłębną charakteryzację genomów, transkryptomów oraz powiązanych z nimi procesów u tysięcy organizmów. W ciągu ostatnich lat opracowano metody, które pozwalają na sekwencjonowanie 3' końców bakteryjnych cząsteczek RNA w różnorodnych warunkach eksperymentalnych, umożliwiając tym samym precyzyjną identyfikację miejsc terminacji transkrypcji, a także stabilnych produktów degradacji czy przetwarzania RNA (Dar et al., 2016; Ju et al., 2019; Mondal et al.,

2016). Metody te mogą znaleźć bezpośrednie zastosowanie m.in. w adnotacji, mechanizmów warunkowej terminacji transkrypcji w odpowiedzi na stres antybiotykowy, a także identyfikacji nowych celów dla leków opartych na oddziaływaniach z mechanizmami fałdowania RNA.

Głównym celem niniejszej rozprawy jest identyfikacja i charakterystyka terminatorów transkrypcji u bakterii oraz udziału warunkowej terminacji transkrypcji w odpowiedzi na stres antybiotykowy u metycylinoopornych szczepów *S. aureus*. *S. aureus* jest obecnie jednym z najistotniejszych klinicznie gatunków bakterii, którego szczepy charakteryzują się bogatym portfolio mechanizmów oporności na bardzo wiele z obecnie stosowanych leków (Foster, 2017; Tong et al., 2015; Vestergaard et al., 2019). Szczególnie istotna jest zatem dogłębna analiza transkryptomu oraz odpowiedzi transkrypcyjnej na działanie różnych klas antybiotyków, a także identyfikacja mechanizmów opartych o cząsteczki RNA, które potencjalnie mogą stanowić cele dla nowo opracowywanych antybiotyków. W tym przypadku niezwykle obiecującymi kandydatami są mechanizmy pozwalające na warunkową regulację terminacji transkrypcji genów niezbędnych dla prawidłowego funkcjonowania komórki.

## 7.1. Centralny Dogmat Biologii Molekularnej

Przez wiele lat aktualna wiedza w dziedzinie biologii molekularnej oparta była o Centralny Dogmat, który objaśnia sposoby przepływu informacji genetycznej w systemach biologicznych. Materiał genetyczny w komórkach przechowywany jest w formie cząsteczek kwasu deoksyrybonukleinowego (DNA), z nielicznymi wyjątkami w postaci wirusów RNA, w których materiałem genetycznym jest kwas rybonukleinowy (RNA). Wspomniany dogmat definiuje w jaki sposób przetwarzana jest informacja genetyczna, prowadząc do syntezy funkcjonalnych produktów, takich jak białka czy cząsteczki RNA.

Fundamenty pod opisywaną teorię położył Francis Crick (1958), w publikacji pod tytułem *'On protein synthesis'*. Autor tego przełomowego dzieła umieścił w nim dwa kluczowe stwierdzenia:

*"In its simplest form it [the Sequence Hypothesis] assumes that the specificity of a piece of nucleic acid is expressed solely by the sequence of its bases, and that this sequence is a (simple) code for the amino-acid sequence of a particular protein",*

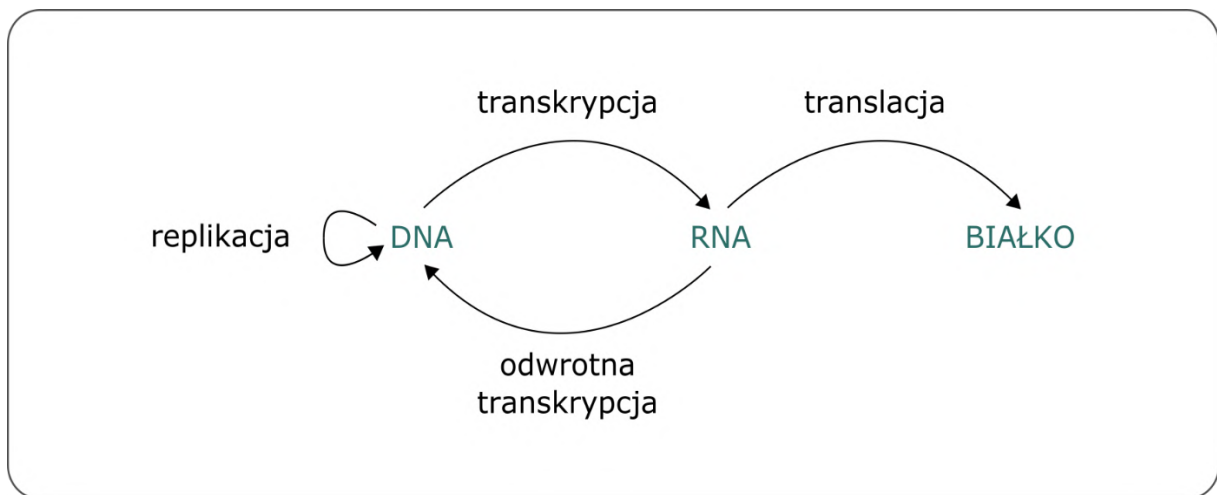
oraz

*"[The Central Dogma] states that once `information` has passed into protein it cannot get out again. In more detail, the transfer of information from nucleic acid to nucleic acid, or from nucleic acid to protein may be possible, but transfer from protein to protein, or from protein to nucleic acid is impossible" (Crick F, 1958).*

Powyższe założenia zmieniły sposób postrzegania mechanizmów przepływu informacji genetycznej. Crick (1958) zasugerował, że sekwencja nukleotydowa stanowi wzorzec do syntezy łańcuchów aminokwasowych, jednocześnie determinując ich sekwencję. Autor zauważył również, że przepływ informacji od białek do kwasów nukleinowych lub pomiędzy białkami nie jest możliwy. Dzisiejszy poziom zrozumienia omawianych procesów jest nieporównywalnie większy niż 65 lat temu, lecz mimo tego, założenia przedstawione przez Crick'a w dalszym ciągu uważane są w znacznym stopniu za zgodne z rzeczywistością (Koonin, 2012).

Obecny stan wiedzy zakłada, że głównym nośnikiem informacji genetycznej, zakodowanej w formie sekwencji nukleotydydowej, jest DNA. Jednoniciowe cząsteczki RNA są natomiast syntezowane na matrycy DNA w procesie nazywanym transkrypcją (Rys. 1). Pełne spektrum funkcji pełnionych przez cząsteczki RNA zapewne jeszcze przez długi czas nie zostanie poznane. Powszechnie jednak wiadomo, że są one odpowiedzialne za regulację i katalizowanie różnorodnych procesów biochemicznych zachodzących w komórkach (Waters & Storz, 2009) oraz służą jako wzorce do syntezy łańcuchów polipeptydowych w procesie nazywanym translacją (Moore & Steitz, 2011; Rodnina, 2018) (Rys. 1), mogą więc stanowić ostateczny produkt ekspresji informacji genetycznej.

Proces syntezy polipeptydów przebiega w oparciu o kody trójkowe, gdzie trzy kolejne nukleotydy (nazywane kodonem) determinują wstawiany do łańcucha białkowego aminokwas. Informacja określająca jaka kombinacja nukleotydydów odpowiada danemu aminokwasowi nazywana jest kodem genetycznym. Wiadomo, że różne kodony mogą kodować ten sam aminokwas, dlatego też kod genetyczny nazywa się często mianem zdegenerowanego, tzn. nie jest możliwe odtworzenie pierwotnej sekwencji nukleotydydowej na podstawie sekwencji aminokwasowej (Brown, 2002). W związku z powyższym i zgodnie z przewidywaniami, do dziś nie został odkryty mechanizm pozwalający na bezpośredni przepływ informacji genetycznej pomiędzy białkami a kwasami nukleinowymi (Koonin, 2012), co jest spójne z założeniami przedstawionymi przez Francis Crick'a już w 1958 roku. Poznano natomiast proces syntezy dwuniciowego DNA (tzw. cDNA, ang. complementary DNA) na matrycy jednoniciowego RNA, nazywany odwrotną transkrypcją i katalizowany przez specjalne enzymy – odwrotne transkryptazy (Temin, 1985) (Rys. 1).

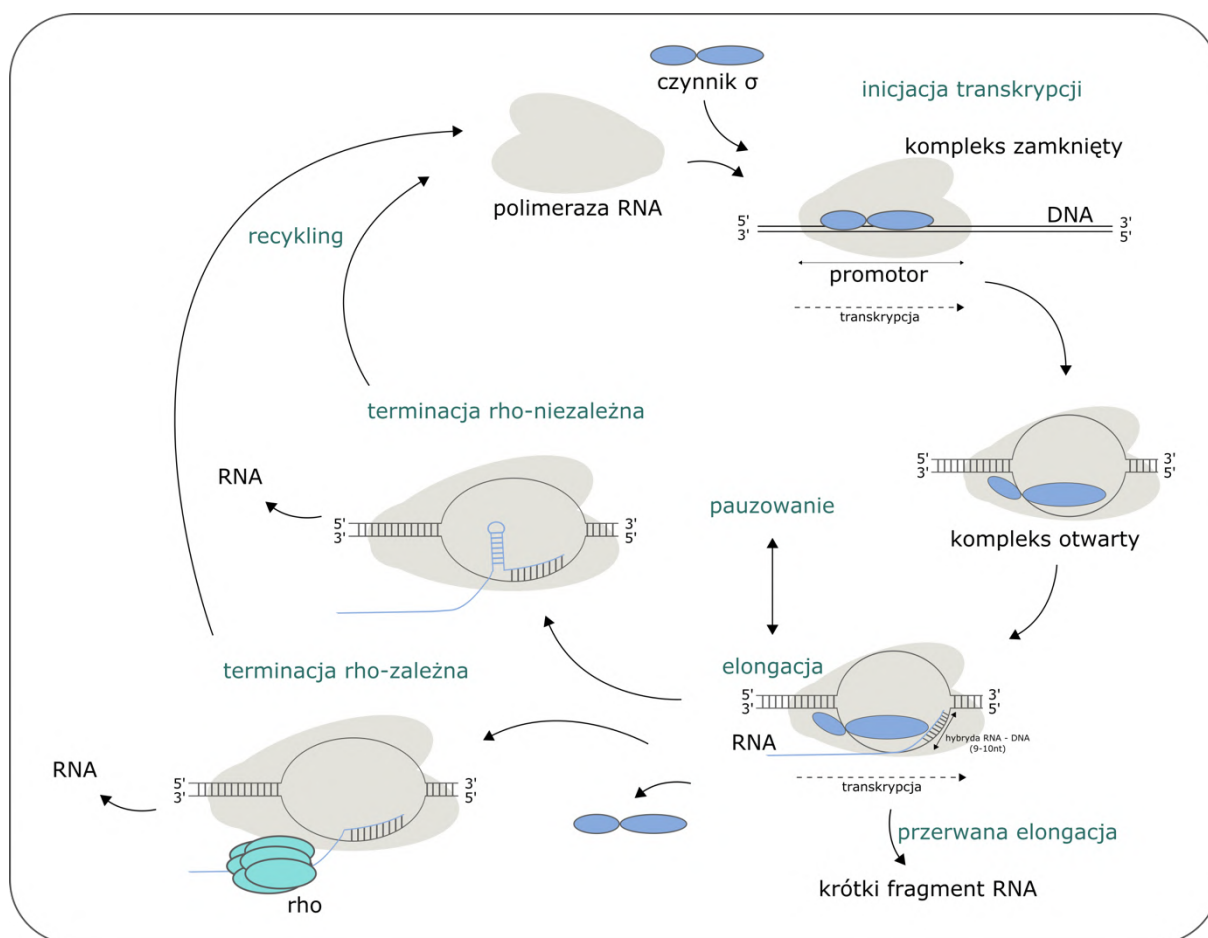


*Rysunek 1. Schematyczna reprezentacja współczesnej wersji Centralnego Dogmatu Biologii Molekularnej ukazującego przepływ informacji genetycznej u organizmów żywych.*

## 7.2. Synteza cząsteczek RNA u bakterii

Transkrypcja umożliwia ekspresję informacji genetycznej zawartej w formie DNA, prowadząc do syntezy jednoniciowej cząsteczki RNA. Proces ten przeprowadzany jest przez enzym nazywany polimerazą RNA w sposób selektywny i ściśle kontrolowany przez maszynę komórkową, której rolą jest określenie, które fragmenty DNA oraz w jakim momencie mają ulegać transkrypcji (Hermann, 2019). U bakterii transkrypcji podlegają głównie fragmenty DNA nazywane genami, które mogą wchodzić w skład, tzw. operonów. Operony stanowią zbiór genów zlokalizowanych w swoim bezpośrednim sąsiedztwie, transkrybowanych i regulowanych wspólnie jako jedna jednostka transkrypcyjna (Osborn & Field, 2009).

Umownie wyróżnia się trzy etapy transkrypcji: inicjację, elongację oraz terminację. Etap inicjacji zależny jest od powstania kompleksu polimerazy RNA oraz czynnika  $\sigma$  (sigma), który odpowiada za rozpoznanie specyficznej sekwencji DNA nazywanej promotorem, rozplecenie dwuniciowego DNA i inicjalizację transkrypcji w konkretnej pozycji (Hermann, 2019; Mejía-Almonte et al., 2020) (Rys. 2). Bakterie mogą posiadać od kilku do kilkudziesięciu różnych czynników  $\sigma$ , wiążących różne rodzaje promotorów, kontrolując tym samym ekspresję różnych genów (Karpen & deHaseth, 2015). Interakcja holoenzymu (kompleksu polimerazy RNA z czynnikiem  $\sigma$ ) z sekwencją promotorową powoduje rozplecenie nici DNA oraz umieszczenie nici matrycowej DNA (czyli nici, na podstawie, której będzie syntezowana nowopowstająca cząsteczka RNA) w centrum aktywnym polimerazy RNA. Powstający w ten sposób kompleks nazywany jest kompleksem otwartym (Browning & Busby, 2016), w którego skład wchodzi 12-14 rozplecionych par zasad DNA (Karpen & deHaseth, 2015) (Rys. 2). Centrum aktywne enzymu katalizuje powstawanie wiązań fosfodiesterowych łączących reszty cukrowe nukleotydów w nowopowstających cząsteczkach RNA (Mejía-Almonte et al., 2020). Niezbędnym substratem powyższej reakcji są trójfosforany nukleozydów, które włączane są do nowopowstającej nici RNA na zasadzie komplementarności do nici matrycowej (Karpen & deHaseth, 2015).



Rysunek 2. Schematyczna reprezentacja cyklu transkrypcyjnego u bakterii.

Najczęściej spotykane promotory bakteryjne reprezentowane są przez specyficzne regiony DNA, wśród których możemy wyróżnić charakterystyczne motywy sekwencyjne. Jednymi z takich regionów są 6-nukleotydowe sekwencje rozpoczynające się odpowiednio 35nt oraz 10nt powyżej miejsca startu transkrypcji. Są one nazywane odpowiednio regionami -35 oraz -10, a ich sekwencje konsensusowe to odpowiednio TTGACA oraz TATAAT (Karpen & deHaseth, 2015). Kolejnym charakterystycznym elementem wielu promotorów bakteryjnych jest ciąg nukleotydów G występujący bezpośrednio po regionie -10. (Karpen & deHaseth, 2015). Inne istotne i charakterystyczne motywy sekwencyjne zostały zidentyfikowane w kierunku 5' względem motywu -35 (tzw. element UP o sekwencji konsensusowej AAAWWTWTTTTNNAANN, ang. *upstream promoter*) oraz na pozycjach od -4 do +2 (tzw. element CRE; ang. *core-recognition element*) (Browning & Busby, 2016; Mejía-Almonte et al., 2020). Powyższe cechy nie są uniwersalnymi cechami wszystkich promotorów bakteryjnych, lecz uważa się, że promotory je posiadające pozwalają na szybsze i skuteczniejsze wiązanie holoenzymu (Karpen & deHaseth, 2015).

Polimeraza RNA jest białkiem składającym się z pięciu podjednostek, oznaczanych umownie greckimi symbolami  $\alpha$  (w skład holoenzymu wchodzi dwie identyczne podjednostki),  $\beta$ ,  $\beta'$  oraz  $\omega$ . Region UP promotora bakteryjnego jest wiązany przez domenę C-końca podjednostki  $\alpha$  polimerazy RNA. Pozostałe ze wspomnianych motywów są rozpoznawane i wiązane przez podjednostki  $\sigma$  (Browning & Busby, 2016).

Proces inicjacji transkrypcji może być regulowany na wielu różnych poziomach, m. in. poprzez modulowanie stężenia polimerazy, czynników  $\sigma$  czy innych substratów reakcji, jak również poprzez wiązanie specyficznych białek, nazywanych czynnikami transkrypcyjnymi, do DNA. W zależności od sposobu regulacji czynniki transkrypcyjne można podzielić na aktywatory i represory, które odpowiednio aktywują lub blokują proces transkrypcji poprzez wiązanie się do specyficznych sekwencji DNA (Mejía-Almonte et al., 2020). Proces inicjacji może być również regulowany przez inne rodzaje cząsteczek, takie jak małe regulatorowe RNA czy niewielkie ligandy. Wspomniane elementy regulatorowe mogą również wiązać się bezpośrednio do holoenzymu modyfikując jego zdolność do wiązania się z sekwencją promotora (Mejía-Almonte et al., 2020).

W momencie utworzenia kompleksu otwartego, kompleks polimerazy RNA może rozpocząć proces syntezy cząsteczki RNA. Etap ten nazywany jest elongacją (Rys. 2). Proces transkrypcji związany jest z występowaniem, tzw. bąbla transkrypcyjnego, czyli rozpleczonego fragmentu dwuniciowego DNA, którego długość u *E. coli* wynosi od 10 do 12nt (Mustaev et al., 2017). Umożliwia on syntezę cząsteczki RNA w oparciu o pojedynczą nić DNA na zasadzie komplementarności zasad, prowadząc podczas syntezy do powstawania hybrydy DNA i RNA.

Translokacja polimerazy wzdłuż nici matrycowej DNA związana jest z występowaniem dwóch stanów. Stan pre-translokacyjny, w którym długość hybrydy RNA:DNA wynosi 10nt, występuje po dołączeniu komplementarnego nukleozydu do nowopowstającej nici RNA, ale jeszcze przed przesunięciem holoenzymu na kolejną pozycję. W tym stanie 3' koniec RNA znajduje się w centrum aktywnym enzymu. O stanie po-translokacyjnym mówimy po przesunięciu polimerazy o jedną pozycję, gdy długość hybrydy RNA:DNA spada do 9nt (Mustaev et al., 2017). Translokacja kompleksu transkrypcyjnego związana jest z koniecznością rozplecenia kolejnej pary nukleotydów występujących w transkrybowanej nici DNA oraz odtworzeniem wiązania pomiędzy nukleotydami występującymi na pierwszej pozycji w obrębie bąbla transkrypcyjnego (Mustaev et al., 2017).

Proces transkrypcji nie jest jednostajny. Polimeraza RNA może się zatrzymywać lub spowalniać syntezę nowopowstającej cząsteczki RNA, wprowadzając tzw. pauzy (Mustaev et



al., 2017; Washburn & Gottesman, 2015) (Rys. 2). Pauzowanie transkrypcji jest zależne od sekwencji nukleotydowej i przypuszczalnie od wiązania białek regulatorowych do DNA (Mustaev et al., 2017). Możliwe jest także cofanie się kompleksu transkrypcyjnego w kierunku promotora (ang. *backtracking*) (Washburn & Gottesman, 2015). W przypadku cofania się holoenzymu długość hybrydy RNA:DNA pozostaje jednak stała i wynosi zawsze 9-10nt (Washburn & Gottesman, 2015).

Proces elongacji jest modulowany przez szereg białek regulatorowych. Do najważniejszych należą NusA oraz NusG, które stymulują powstawanie pauz oraz odgrywają istotną rolę w terminacji transkrypcji (Mandell et al., 2021). Białka UvrD oraz Mfd promują z kolei odpowiednio cofanie się kompleksu transkrypcyjnego oraz translokację w kierunku 3', biorąc udział w procesie naprawy DNA (Washburn & Gottesman, 2015). U *E. coli* występują również białka GreA/B, które przycinają wolny koniec 3' RNA powstały w wyniku cofnięcia się holoenzymu, umożliwiając tym samym wznowienie procesu transkrypcji (Washburn & Gottesman, 2015). Synteza cząsteczki RNA kończy terminację transkrypcji (Rys. 2).

Terminacja transkrypcji u bakterii może zachodzić z wykorzystaniem dwóch odrębnych mechanizmów. Tradycyjnie wyróżniamy, tzw. rho-zależną (ang. *rho-dependent, factor-dependent*) oraz rho-niezależną (ang. *rho-independent, factor-independent, intrinsic*) terminację transkrypcji, w zależności od udziału białka Rho w niniejszym procesie. Białko Rho jest homo-heksameryczną ATPazą posiadającą również aktywność helikazy RNA:DNA, która jak uważano do tej pory, rozpoczyna oddziaływanie z nowopowstającą cząsteczką RNA w obrębie, tzw. sekwencji *rut* (ang. *rho-utilization element*) (Chełkowska-Pauszek et al., 2021; Mandell et al., 2022), charakteryzującą się wysokim udziałem nukleotydów cytozynowych w stosunku do guaninowych oraz regularnym rozmieszczeniem dipyrimidynowych powtórzeń (Chełkowska-Pauszek et al., 2021; Mandell et al., 2022).

Tradycyjny model opisujący mechanizm rho-zależnej terminacji transkrypcji u bakterii zakładał, że aktywność translokazy ATP-zależnej pozwala białku Rho na przemieszczanie się w obrębie syntezywanej nici, umożliwiając „dogonienie” polimerazy RNA i w konsekwencji destabilizację kompleksu transkrypcyjnego oraz terminację transkrypcji (Hao, Epshtein, et al., 2021; Hao, Svetlov, et al., 2021). Obecnie pojawia się jednak coraz więcej dowodów, które wskazują na to, że Rho wiąże się do kompleksu elongacyjnego już na początkowych etapach transkrypcji, oddziałując ściśle z czynnikami transkrypcyjnymi NusA oraz NusG (Epshtein et al., 2010; Hao, Epshtein, et al., 2021; Hao, Svetlov, et al., 2021; Mandell, Zemba, et al., 2022; Said et al., 2021). Nowy model omawianego mechanizmu zakłada, że interakcja takiego kompleksu pre-terminacyjnego z sekwencją *rut* indukuje zmiany allosteryczne w obrębie

polimerazy RNA powodując jej przejście w stan nieaktywny, co w konsekwencji prowadzi do dysasocjacji nowopowstającej cząsteczki RNA z centrum aktywnego, skutkując tym samym terminacją transkrypcji (Epshtein et al., 2010; Hao, Epshtein, et al., 2021; Hao, Svetlov, et al., 2021; Mandell, Zemba, et al., 2022; Said et al., 2021).

Alternatywnym mechanizmem terminacji transkrypcji u bakterii jest, tzw. rho-niezależna terminacja transkrypcji, która historycznie uważana była za niezależną od jakichkolwiek czynników transkrypcyjnych, takich jak Rho, NusA czy NusG (Mandell, Zemba, et al., 2022). Rho-niezależna terminacja transkrypcji zakłada, że formująca się w obrębie kanału wyjściowego polimerazy RNA struktura typu spinka do włosów, prowadzi do rozplecenia hybrydy RNA:DNA oraz dysasocjacji kompleksu transkrypcyjnego (Chełkowska-Pauszek et al., 2021; Gusarov & Nudler, 1999; Mandell et al., 2022). Alternatywne modele zakładają hiper-translokację kompleksu transkrypcyjnego lub zmiany allosteryczne w obrębie polimerazy RNA indukowane formowaniem spinki terminacyjnej, w obu przypadkach prowadząc do efektywnej terminacji transkrypcji (Mandell, Zemba, et al., 2022). Niezwykle istotnym odkryciem okazuje się być fakt, iż w rzeczywistości proces ten nie jest całkowicie niezależny od czynników transkrypcyjnych. Udowodniono, że terminacja transkrypcji w obrębie około 97% terminatorów, które do tej pory były klasyfikowane jako rho-niezależne u *B. subtilis*, jest w istocie stymulowana poprzez co najmniej jedno z białek regulatorowych Rho, NusA lub NusG (Mandell et al., 2021; Mandell, Vishwakarma, et al., 2022).

Spinki terminacyjne, które leżą u podstaw, tzw. rho-niezależnej terminacji transkrypcji, występują zazwyczaj w obrębie bogatych w pary G-C fragmentów nowopowstających cząsteczek RNA (Ray-Soni et al., 2016; Roberts, 2019). Struktury te mają tendencję do formowania się w większości przypadków około 7-9 nukleotydów powyżej 3' końca RNA (Ray-Soni et al., 2016; Roberts, 2019). Niezwykle charakterystycznym elementem spinek terminacyjnych jest trakt o długości około 7-8 nukleotydów, bogaty w nukleotydy urydynowe, występujący bezpośrednio poniżej struktury drugorzędowej. Niektóre terminatory posiadają również analogiczny trakt bogaty w nukleotydy adeninowe, poprzedzający spinę, który pozwala na efektywną terminację w obu kierunkach (Chełkowska-Pauszek et al., 2021; Ray-Soni et al., 2016; Roberts, 2019).

### 7.3. Rola cząsteczek RNA w komórkach bakteryjnych

Pełny repertuar funkcji pełnionych przez cząsteczki RNA w komórkach nie został dotychczas całkowicie poznany, jednakże wiadomo, że różnorodność mechanizmów i procesów, w których odgrywają one kluczowe role, jest ogromna. Z pewnością jednym z najważniejszych procesów, do przeprowadzenia którego cząsteczki RNA są niezbędne, jest proces translacji (Moore & Steitz, 2011; Rodnina, 2018). We wspomnianym procesie pełnią one rolę matrycy, na bazie której syntezowany jest łańcuch polipeptydowy stanowiący podstawę do uformowania w pełni funkcjonalnego białka. Cząsteczki RNA stanowiące matrycę do syntezy białek nazywane są matrycowym RNA (mRNA). Istnieje również specjalna grupa cząsteczek, która jest odpowiedzialna za dostarczanie aminokwasów do rybosomów przeprowadzających proces syntezy peptydów, tzw. tRNA (ang. *transfer RNA*). Niezwykle ważną grupę stanowią także rybosomalne RNA (rRNA), będące integralną częścią wspomnianych rybosomów, składających się z szeregu białek i cząsteczek kwasów nukleinowych (Moore & Steitz, 2011). Dodatkowo RNA mogą pełnić w komórkach rozmaite funkcje regulatorowe, katalizować różnorodne reakcje biochemiczne oraz kontrolować ekspresję genów, jednocześnie nie ulegając translacji. Są to tzw. ncRNA (ang. *non-coding RNA*), czyli niekodujące cząsteczki RNA (Brown, 2002).

U bakterii cząsteczki RNA mogą modulować odpowiedzi na różnorodne sygnały komórkowe lub środowiskowe z wykorzystaniem szeregu dobrze scharakteryzowanych i poznanych mechanizmów. Jednym ze sposobów wspomnianej modulacji może być hybrydyzacja z innymi cząsteczkami RNA. Mechanizm ten najczęściej pozwala regulować procesy translacji czy degradacji RNA. Powszechnie jest również wiązanie RNA z białkami, potencjalnie wpływające na ich funkcjonowanie (Waters & Storz, 2009).

Funkcje regulatorowe RNA powiązane są często z możliwością tworzenia dynamicznie zmieniających się struktur drugorzędowych, które powstają na zasadzie spontanicznego parowania się bliskich lub odległych regionów wchodzących w skład tej samej jednoniciowej cząsteczki kwasu nukleinowego. Struktury drugorzędowe mogą modulować szereg procesów związanych z procesami regulacji ekspresji genów, zazwyczaj wpływając na działanie różnorodnych enzymów czy kompleksów białkowych, takich jak np. polimeraza RNA czy rybosom (Chełkowska-Pauszek et al., 2021). Często spotykanym mechanizmem działania jest zmiana fałdowania cząsteczki kwasu nukleinowego w odpowiedzi na różnorodne sygnały

środowiskowe, która w konsekwencji prowadzi do blokowania lub aktywacji ekspresji genu lub genów położonych poniżej (w kierunku 3') poprzez oddziaływanie struktury drugorzędowej RNA z maszyną komórkową zaangażowaną w procesy transkrypcji lub translacji (Chelkowska-Pauszek et al., 2021; Dersch et al., 2017).

#### 7.4. Wpływ struktur drugorzędowych RNA na regulację ekspresji genów

Szczegółowe opisy mechanizmów terminacji transkrypcji u bakterii oraz funkcji pełnionych przez struktury drugorzędowe RNA w regulacji ekspresji genów zostały opisane w poniższym artykule stanowiącym integralną część niniejszej pracy doktorskiej.

Chełkowska-Pauszek A, **Kosiński JG**, Marciniak K, Wysocka M, Bąkowska-Żywicka K, Żywicki M. *The Role of RNA Secondary Structure in Regulation of Gene Expression in Bacteria*. Int J Mol Sci. 2021 Jul 22;22(15):7845. doi: 10.3390/ijms22157845. PMID: 34360611; PMCID: PMC8346122

\* - Autorzy deklarują, że ich wkład w przygotowanie powyższej pracy był równy, a kolejność listowania ma porządek alfabetyczny

<https://www.mdpi.com/1422-0067/22/15/7845>

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8346122/>

## **7.5. Terminacja transkrypcji a problem lekooporności bakterii**

W wielu przypadkach procesy warunkujące lekooporność u bakterii wykorzystują różnorodne mechanizmy oparte o przedwczesną terminację transkrypcji. Warunkowa terminacja transkrypcji pozwala dynamicznie, w odpowiedzi na różnorodne sygnały komórkowe, regulować ekspresję genów, w tym również tych zaangażowanych w odpowiedź bakterii na działanie antybiotyków (Dersch et al., 2017).

Antybiotyki są substancjami wstrzymującymi lub hamującymi wzrost drobnoustrojów, w tym bakterii chorobotwórczych. Można je zdefiniować jako związki organiczne, które hamują rozwój drobnoustrojów lub prowadzą do ich śmierci poprzez określone interakcje z elementami komórki bakteryjnej, bez względu na pochodzenie związku (Davies & Davies, 2010). Antybiotyki mogą być otrzymywane metodami laboratoryjnymi lub wytwarzane przez niektóre organizmy, np. bakterie i grzyby w sposób naturalny (Davies & Davies, 2010; Fleming, 1980; Gaynes, 2017), a ich wykorzystanie okazało się największym przełomem w medycynie XX w., pozwalającym na uratowanie życia i zdrowia niezliczonej liczby osób (Davies & Davies, 2010). Kolejne lata i dekady rozwoju nauk medycznych prowadziły do

odkryć wielu nowych klas antybiotyków, stosowanych w leczeniu pacjentów do dnia dzisiejszego.

Zdolność niezwykle szybkiej adaptacji bakterii do zmieniających się warunków środowiskowych lub ekspozycji na określone substancje spowodowała jednak szybkie wykształcenie się mechanizmów oporności na działanie antybiotyków, skutkując jednocześnie spadkiem skuteczności wykorzystywanych dotychczas terapii. Szacuje się, że w samym roku 2019, infekcje wywołane przez bakterie odporne na obecnie stosowane metody leczenia spowodowały około pięć milionów przedwczesnych zgonów na Świecie i były bezpośrednią przyczyną ok. 1.3 miliona z nich (Murray et al., 2022). Nadmierne wykorzystanie antybiotyków w medycynie, rolnictwie i innych dziedzinach życia spowodowało, w wyniku nieustającej presji selekcyjnej, zwiększenie ilości opornych na nie organizmów (Dodds, 2017; Jian et al., 2021; Levy & Bonnie, 2004). Dodatkowo, trzeba zauważyć, że problem antybiotykoodporności bakterii narasta w alarmującym tempie. Prognozuje się, że do roku 2050 liczba zgonów, związanych z zakażeniami wywołanymi przez lekooporne szczepy bakterii, wzrośnie do ok. 10 milionów rocznie i spowoduje straty finansowe w wysokości sięgającej 10 miliardów USD (J. O'Neill, 2016). Istotnie rosnąca liczba szczepów bakterii opornych jednocześnie na wiele klas antybiotyków znacząco zwiększa szansę na powrót do ery pre-antybiotykowej, w której dla wielu chorób zakaźnych nie były znane efektywne metody leczenia (Davies & Davies, 2010).

Istnieją przekonujące dowody wskazujące na fakt, że mechanizmy oporności na antybiotyki istniały na długo przed rozpoczęciem ich stosowania w medycynie, o czym świadczyć może fakt, że wiele znanych genów warunkujących oporność na leki zostało zidentyfikowanych w próbkach DNA pochodzących m. in. z wiecznej zmarzliny lub silnie izolowanych ekosystemów (Jian et al., 2021; Waglechner & Wright, 2017). Przypuszcza się, że oporność bakterii jest ewolucyjnym wynikiem kontaktów z innymi drobnoustrojami zdolnymi do produkcji tego rodzaju związków (Jian et al., 2021; Munita & Cesar, 2016; Waglechner & Wright, 2017).

Do mechanizmów umożliwiających nabycie odporności można zaliczyć spontaniczne mutacje DNA, horyzontalny transfer genów oraz presję selekcyjną wywołaną obecnością leków (Davies & Davies, 2010; Levy & Bonnie, 2004; Martinez, 2014; Munita & Cesar, 2016). Horyzontalny transfer genów polega na transferze „mobilnego” materiału genetycznego (np. transpozonów, plazmidów lub ‘gołego’ DNA) pomiędzy bakteriami na drodze transformacji, transdukcji czy koniugacji (Jian et al., 2021; Munita & Cesar, 2016). Proces transformacji pozwala na pobieranie „wolnego” DNA ze środowiska. Tak pobrane DNA może być następnie dziedziczone przez komórki potomne. Mechanizm transdukcji polega z kolei na wprowadzeniu

nowego materiału genetycznego do komórki bakteryjnej przez bakteriofagi. Ostatnim poznany procesem horyzontalnego transferu genów jest koniugacja, czyli bezpośrednia wymiana materiału genetycznego między komórkami bakteryjnymi za pomocą specjalnych mostków cytoplazmatycznych nazwanych pilus. Wszystkie z wymienionych mechanizmów mogą prowadzić do nabycia oporności na drodze inkorporacji fragmentów DNA determinujących oporność na stosowane leki (Martinez, 2014; Munita & Cesar, 2016).

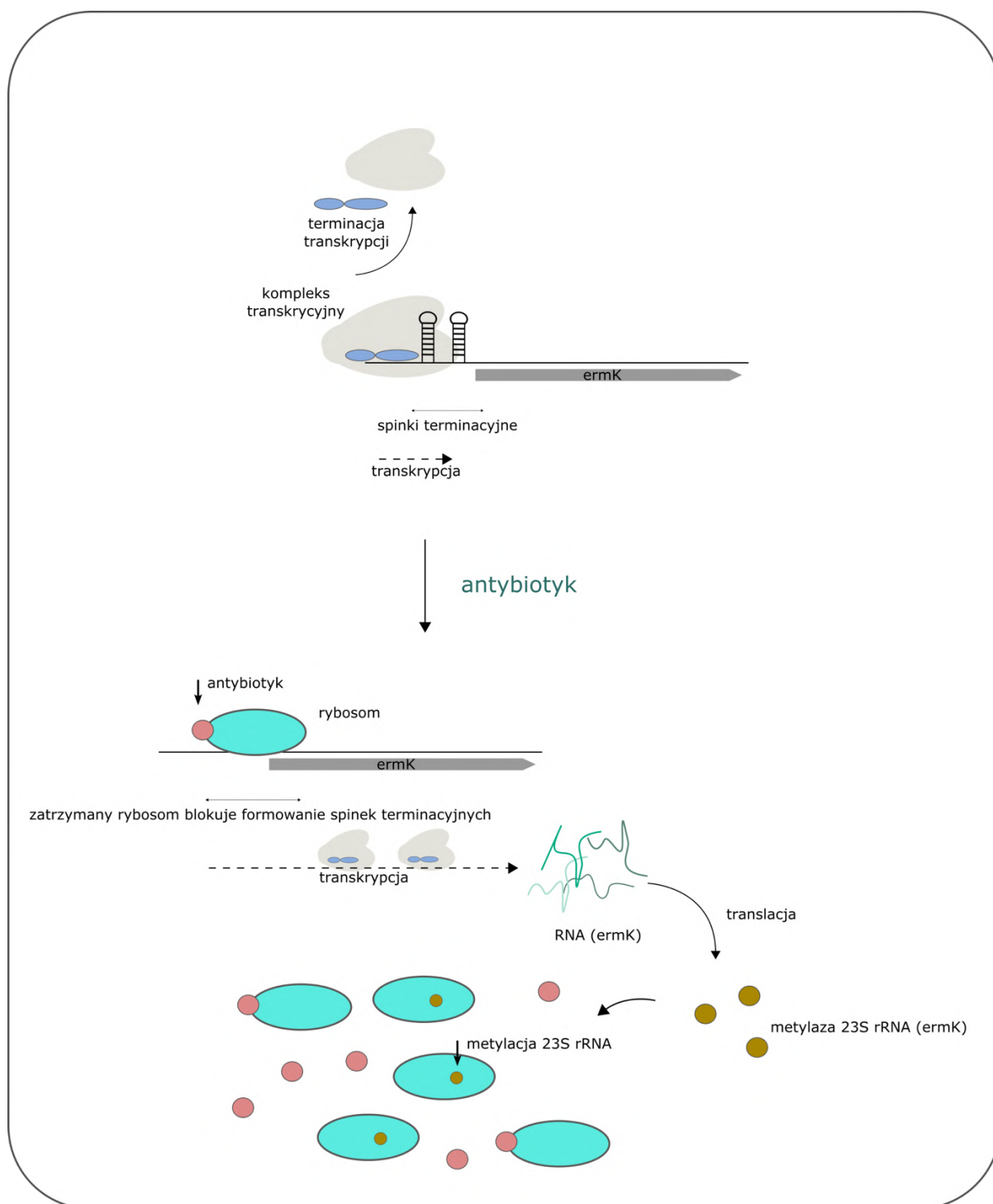
Munita oraz Arias (2016) zaproponowali czteroklasowy podział mechanistycznych sposobów uzyskiwania oporności na leki przez bakterie. Pierwszy z nich bazuje na degradacji antybiotyku (np. na drodze hydrolizy) lub jego enzymatycznej modyfikacji, prowadząc do zmiany sposobu jego oddziaływania na komórkę (Jian et al., 2021; Munita & Cesar, 2016; Waglechner & Wright, 2017). Drugi mechanizm opiera się na zapobieganiu by cząsteczki leku dotarły do swojego celu w komórce, np. poprzez jego aktywne usuwanie lub ograniczenie wnikania poprzez błony i ściany komórkowe (Jian et al., 2021; Munita & Cesar, 2016; Waglechner & Wright, 2017). Trzeci z nich polega na modyfikowaniu lub blokowaniu celu oddziaływania związku terapeutycznego. Z kolei czwarty typ mechanizmu uzyskiwania oporności na antybiotyki bazuje na wprowadzaniu zmian zwiększających całościowo możliwości adaptacyjnych komórki (Munita & Cesar, 2016).

Na przestrzeni ostatnich lat pojawia się coraz więcej doniesień naukowych wskazujących, że ekspresja wielu genów warunkujących zwiększoną oporność na antybiotyki u bakterii może być regulowana na poziomie transkrypcyjnym przy pomocy różnorodnych elementów strukturalnych RNA, takich jak ryboprzełączniki czy atenuatory (Dersch et al., 2017).

Jednym z przykładów tego rodzaju mechanizmu może być regulacja terminacji transkrypcji genu *ermK* u *Bacillus licheniformis* w odpowiedzi na traktowanie antybiotykami należącymi do grup makrolidów, linkozamidów czy streptogramin typu B (Kwon et al., 2006). Gen *ermK* odpowiada za syntezę metylazy 23S rRNA, która jest enzymem metylującym określone pozycje nukleotydowe w cząsteczce 23S rRNA (Kwak et al., 1991). Tego rodzaju modyfikacja potranskrypcyjna jednej z podjednostek rybosomu prokariotycznego pozwala zaburzyć mechanizmy oddziaływania wspomnianych wyżej klas leków z rybosomami. Mechanizmy te są odpowiedzialne za blokowanie procesu translacji i syntezy białek (Kwon et al., 2006). Tym samym ekspresja genu *ermK*, skutkująca syntezą omawianej metylazy, prowadzi do wykształcenia oporności na przedstawione powyżej klasy antybiotyków. W obrębie sekwencji liderowej genu *ermK* mogą powstawać dwie wzajemnie się wykluczające struktury drugorzędowe (Kwak et al., 1991). W skład jednej z nich wchodzi dwa terminatory



rho-niezależne, które pozwalają na przerwanie transkrypcji w warunkach nieuwzględniających oddziaływania antybiotyku (Kwak et al., 1991). Pełnej długości transkrypty genu *ermK* powstają natomiast w momencie traktowania komórek bakteryjnych przy pomocy dedykowanego leku, np. erytromycyny (Kwak et al., 1991). Jest to bezpośrednio związane z zatrzymaniem podążających za polimerazą RNA rybosomów przeprowadzających translację w regionie liderowym, będącego efektem interakcji z cząsteczkami antybiotyku. Zatrzymany rybosom blokuje możliwość uformowania się spinek terminacyjnych, prowadząc jednocześnie do syntezy pełnej długości produktów genu, pozwalających zwalczyć negatywne oddziaływanie leku na komórki (Kwak et al., 1991; Kwon et al., 2006) (Rys. 3).



Rysunek 3. Rola warunkowej terminacji transkrypcji w regulacji ekspresji genu *ermK* oraz wykształceniu oporności na antybiotyki.

Regulacja ekspresji genów związanych z antybiotykoopornością, oparta o zmiany struktury drugorzędowej RNA, może również wpływać na proces translacji. Można w tym przypadku posłużyć się przykładem innego genu z rodziny *erm*, a konkretnie genu *ermC* u *S.*

*aureus*. Podobnie jak w przypadku wyżej opisywanego genu *ermK*, gen *ermC* koduje metylazę 23S rRNA i posiada krótką otwartą ramkę odczytu w swoim regionie liderowym (Dersch et al., 2017). Aktywna translacja tego krótkiego peptydu prowadzi do powstania struktury drugorzędowej, która zawiera miejsce wiązania rybosomu (ang. *ribosome-binding site*; RBS) genu *ermC*, jednocześnie blokując jego dostępność dla rybosomu i uniemożliwiając translację. Oddziaływanie antybiotyku makrolidowego z rybosomem powoduje jego zatrzymanie w trakcie syntezy liderowego peptydu i fałdowanie alternatywnej struktury drugorzędowej, która eksponuje miejsce wiązania rybosomu, umożliwiając efektywną translację genu *ermC* (Dersch et al., 2017). Podobnie jak w przypadku genu *ermK*, synteza metylazy 23S rRNA prowadzi do zmniejszenia podatności bakterii na działanie antybiotyku (Dersch et al., 2017).

Jedną z metod ograniczania negatywnego wpływu antybiotyków na funkcjonowanie organizmów bakteryjnych jest aktywne usuwanie cząsteczek leku poza obszar komórki (Jian et al., 2021; Munita & Cesar, 2016; Waglechner & Wright, 2017). Proces ten wymaga wykorzystania specjalnych białek transportujących (tzw. pomp efflux), które wykorzystują odpowiednie źródła energii (np. ATP) do eksportu szkodliwych substancji przez kanały transbłonowe poza ustrój komórki (Li & Nikaido, 2009). Przykładem tego rodzaju transporterów jest rodzina białek ABC (ang. *ATP-binding cassette*), do których należą białka *BrmC* i *BrmD* u *B. subtilis* (Reilman et al., 2014; Torres et al., 2009).

Geny *BrmC* i *BrmD* wchodzi w skład pojedynczego operonu wraz z genem *BrmB*, a ich wzajemne ułożenie na genomie jest alfabetyczne (kolejno od 5' końca: *BrmB*, *BrmC*, *BrmD*) (Torres et al., 2009). Regulacja ekspresji operonu, na drodze przedwczesnej terminacji transkrypcji, warunkowana jest obecnością rybosomu przeprowadzającego syntezę białka, podążającego bezpośrednio za kompleksem polimerazy (Reilman et al., 2014). U bakterii, procesy transkrypcji i translacji nie są względem siebie rozłączne, a nowopowstały łańcuch RNA ulega translacji w momencie, kiedy nie została jeszcze zakończona transkrypcja danego genu lub operonu (Wang et al., 2020). W sytuacji, gdy cząsteczki antybiotyku są nieobecne gen *BrmB* ulega translacji bez przeszkód, a interakcja rybosomu z cząsteczką RNA prowadzi do uformowania terminatora, który jest w tym przypadku strukturą bardziej stabilną energetycznie (Reilman et al., 2014). Terminacja transkrypcji zachodząca poniżej genu *BrmB* powoduje, że produkty pozostałych genów wchodzących w skład operonu (w tym *BrmC* i *BrmD*) nie powstają. W obecności antybiotyków poziom translacji genu *BrmB* istotnie maleje, co pozwala na uformowanie struktury antyterminatora i efektywną transkrypcję genów *BrmC* i *BrmD*, a w konsekwencji syntezę białek transporterowych umożliwiających eksport cząsteczek antybiotyku poza ustrój komórki (Reilman et al., 2014). Przykładem leku oddziałującego na

powyżej opisany mechanizm jest linkomycyna oddziałująca z rybosomami obecnymi w komórce bakteryjnej (Reilman et al., 2014).

Podobny mechanizm regulacyjny opisano w przypadku genu *VmlR* u *B. subtilis* (Takada et al., 2022). Białko syntezowane na podstawie powyższego genu jest ATPazą, która jest w stanie zastąpić cząsteczki antybiotyków wiążące się do rybosomu, odpowiedzialne za blokowanie centrum peptydotransferazy rybosomu bakteryjnego, bądź też tunelu, przez który wydostaje się nowo syntezowany polipeptyd (Takada et al., 2022). Zatrzymanie rybosomu podczas translacji otwartej ramki odczytu zlokalizowanej w regionie liderowym genu *VmlR*, wywołane działaniem leku będącego inhibitorem translacji, zapobiega utworzeniu struktury anty-antyterminatora, prowadząc do uformowania alternatywnej struktury drugorzędowej RNA pełniącej rolę antyterminatora. Powyższy mechanizm pozwala na efektywną transkrypcję genu w momencie oddziaływania antybiotyku na komórkę i przedwczesną terminację transkrypcji w przypadku jego braku. Terminacja transkrypcji spowodowana jest utworzeniem struktury spinki terminacyjnej (Takada et al., 2022). Produkty genu *VmlR* zwiększają oporność bakterii na antybiotyki, takie jak pleuromutyliny, linkozamidy czy streptograminy typu A (Takada et al., 2022).

Istotna rola elementów struktury drugorzędowej w regulacji ekspresji genów bakteryjnych, zaangażowanych w zmniejszanie podatności komórki na antybiotyki, sprawia, że stanowią one potencjalnie bardzo obiecujące cele oddziaływania dla nowoprojektowanych klas antybiotyków (Dersch et al., 2017). Doskonałym przykładem mogą w tym przypadku być ryboprzełączniki.

Ryboprzełączniki stanowią funkcjonalne elementy struktury drugorzędowej RNA, które w odpowiedzi na wiązanie niewielkich metabolitów mogą modulować zmiany strukturalne w zależnych od siebie fragmentach cząsteczki kwasu rybonukleinowego. Ich zdolność do rozpoznawania i wiązania konkretnego typu małocząsteczkowego metabolitu oraz fakt, że kontrolują one ekspresję co najmniej 4% genów u bakterii Gram-dodatnich, stanowią o ich ogromnym potencjale jako celach dla nowo opracowywanych klas antybiotyków (Machtel et al., 2016). Dodatkowo nie obserwuje się występowania ryboprzełączników u człowieka i wyższych *Eukaryota*, co może być związane z mniejszą toksycznością potencjalnych leków bazujących na oddziaływaniu z ryboprzełącznikami (Machtel et al., 2016). Niewielkie rozmiary i prosta budowa chemiczna wiązanych ligandów również przemawiają na korzyść tego typu systemów, ułatwiając dostarczenie potencjalnych leków do komórek i zmniejszając nakłady finansowe potrzebne do ich efektywnej produkcji (Machtel et al., 2016). Powszechne jest również zjawisko regulacji wielu genów bakteryjnych przy pomocy tego samego rodzaju

ryboprzełączników (wiążącego ten sam rodzaj ligandu). Fakt ten jest bezpośrednio związany ze zmniejszonym prawdopodobieństwem nabycia oporności przez bakterie na drodze spontanicznych mutacji. Mutacja pojedynczej pozycji w genomie zapewne okazałaby się niewystarczająca by zniwelować oddziaływanie antybiotyku na wiele genów równocześnie (Machtel et al., 2016).

Jako przykład związku zabijającego bakterie poprzez oddziaływanie ze strukturą ryboprzełączników może posłużyć związek PC1 (ang. *pyrimidine compound 1*) (Mulhbacher et al., 2010). PC1 jest zdolny do oddziaływania z ryboprzełącznikami guaninowymi, które odpowiadają za regulację ekspresji genów wchodzących w skład szlaków metabolicznych biosyntezy puryn. Ryboprzełączniki guaninowe, po związaniu guaniny pełniącej tu rolę ligandu, pozwalają na stabilizację struktury drugorzędowej warunkującej atenuację transkrypcji genu *guaA* u *S. aureus*. W przypadku ograniczonej obecności nukleotydów guaninowych w komórce mechanizm ten pozwala na efektywne przeprowadzanie transkrypcji genów wchodzących w skład kontrolowanego operonu. Związek PC1 pozwala zahamować transkrypcję, nawet w przypadku, gdy podaż puryn w obrębie komórki jest niewystarczająca, prowadząc tym samym do śmierci komórki (Mulhbacher et al., 2010).

Do niedawna mechanizmy regulacji ekspresji genów oparte o atenuatory i ryboprzełączniki uważane były za stosunkowo rzadkie, jednak w ostatnich latach, wraz z pojawieniem się metod wysokoprzepustowego sekwencjonowania 3' końców cząsteczek RNA, okazało się, że są one bardzo powszechne w transkryptomach bakteryjnych (Dersch et al., 2017). Zaangażowanie tego rodzaju systemów regulatorowych w modulowanie rozlicznych procesów prowadzących do zwiększenia poziomów antybiotykoodporności u bakterii, spowodowało zwiększone zainteresowanie środowiska naukowego w kontekście wykorzystania ich jako potencjalnych celów dla nowych klas antybiotyków (Dersch et al., 2017; Machtel et al., 2016).

## 7.6. Molekularne mechanizmy lekooporności gronkowca złocistego

Niniejsza rozprawa dedykowana jest identyfikacji i charakterystyce rho-niezależnych terminatorów transkrypcji u bakterii oraz ich udziału w odpowiedzi na stres antybiotykowy u gronkowca złocistego (*S. aureus*). Gronkowiec złocisty należy obecnie do jednych z najistotniejszych klinicznie patogenów, charakteryzujących się wysokim poziomem oporności na większość stosowanych antybiotyków (Cheung et al., 2021). Dlatego też niezwykle ważne, w kontekście opracowania nowych terapii, jest dokładne zrozumienie sposobów działania mechanizmów regulacji ekspresji genów, m.in. warunkowej terminacji transkrypcji, zaangażowanych w odpowiedź komórkową *S. aureus* na traktowanie antybiotykami.

Gronkowiec złocisty jest Gram-dodatnią komensalną bakterią występującą w znacznej części populacji ludzkiej (~10-40%), głównie w obrębie błony śluzowej nosa, na skórze lub w obrębie dolnych odcinków układu pokarmowego (Cheung et al., 2021; Kwiecinski & Horswill, 2020; Lakhundi & Zhang, 2018; Tong et al., 2015). Do klinicznie istotnych zakażeń dochodzi zazwyczaj w przypadku naruszenia lub uszkodzenia błon śluzowych lub skóry, gdy bakterie mają możliwość przedostania się do innych tkanek lub krwioobiegu (Lee et al., 2018). Szczególnie narażone na infekcje są osoby z obniżoną odpornością, korzystające z inwazyjnych przyrządów medycznych (np. cewników), bądź poddawane zabiegom chirurgicznym (Lee et al., 2018).

Sumaryczna liczba poważnych zakażeń wywoływanych przez gronkowca złocistego jest trudna do oszacowania, ale przypuszcza się, że bakterie tego gatunku są odpowiedzialne za setki tysięcy lub nawet miliony takich sytuacji rocznie (Cheung et al., 2021). Omawiane bakterie mogą wywoływać m.in. niebezpieczne dla życia i zdrowia zapalenia dróg oddechowych (w tym płuc), infekcje skórne, tkanek miękkich, układu sercowo-naczyniowego, kostno-stawowego, a także części ciała będących celem rozmaitych operacji chirurgicznych (Cheung et al., 2021; Tong et al., 2015). Wspomniane zakażenia mogą prowadzić do bardzo groźnych stanów chorobowych, takich jak np. sepsa, zapalenie wsierdza czy bakteremia (Kwecinski & Horswill, 2020; Tong et al., 2015).

Gronkowiec złocisty jest bardzo pospolitym patogenem wywołującym zakażenia szpitalne oraz poza-szpitalne (Cheung et al., 2021; Lakhundi & Zhang, 2018). Wiele szczepów gronkowca złocistego wykazuje równoczesną oporność na wiele ze stosowanych klas leków,

czyniąc proces leczenia bardzo skomplikowanym oraz powodując wysoką śmiertelność u pacjentów (Lakhundi & Zhang, 2018). Liczba zakażeń kończących się śmiercią pacjenta jest znacząco wyższa w przypadku bakteriemii wywołanych przez gronkowca złocistego niż sumarycznie w przypadku takich chorób jak AIDS, gruźlica czy zapalenie wątroby (Cheung et al., 2021).

Najistotniejszym klinicznie szczepem gronkowca od lat pozostaje MRSA (ang. *methicilin-resistant S. aureus*), czyli szczep oporny na metycylinę (Cheung et al., 2021; Lakhundi & Zhang, 2018). Zakażenia wywołane przez MRSA charakteryzują się najwyższą śmiertelnością i zachorowalnością, a także najdłuższym czasem i kosztem leczenia (Cheung et al., 2021).

Pierwsze antybiotyko-oporne szczepy opisywanego gatunku zostały zidentyfikowane krótko po opracowaniu pierwszego antybiotyku - penicyliny, a dokładnie dwa lata po jej wprowadzeniu do użytku klinicznego w roku 1940 (Lakhundi & Zhang, 2018). Metycylina, będąca półsyntetycznym antybiotykiem b-laktamowym stosowanym powszechnie w leczeniu zakażeń wywołanych przez gronkowca złocistego, została opracowana w latach 50. XX w. (Lakhundi & Zhang, 2018). Pierwsze szczepy gronkowca oporne na metycylinę zostały natomiast wykryte już w roku 1961, co wskazuje na jego ogromne zdolności adaptacyjne (Jevons, 1961; Lakhundi & Zhang, 2018; Lee et al., 2018). Szczepy MRSA spotykane są obecnie na całym Świecie, jednak ich udział wśród pozostałych szczepów charakteryzuje się znacznym zróżnicowaniem geograficznym. Najwyższy obserwowany jest w krajach Ameryki Południowej oraz Azji, gdzie waha się często w granicach 50-70%, podczas gdy w krajach Skandynawskich wynosi on <5% (Lee et al., 2018).

Zdecydowana większość obecnie spotykanych szczepów gronkowca złocistego jest odporna na antybiotyki z rodziny penicylin. Ważnym elementem genomowym warunkującym oporność na penicylinę jest gen *blaZ*, który odpowiedzialny jest za kodowanie enzymu  $\beta$ -laktamazy.  $\beta$ -laktamazy katalizują hydrolizę pierścieni  $\beta$ -laktamowych, będących kluczowym elementem wspomnianej grupy antybiotyków, do której zaliczają się m.in. penicyliny (Pantosti et al., 2007; Vestergaard et al., 2019). Gen *blaZ* wchodzi w skład transpozonu, który może być zintegrowany z genomem chromosomowym, a także przenoszony w obrębie plazmidów, mogących również warunkować oporność na inne klasy leków (Vestergaard et al., 2019). Szerokie zastosowanie penicylin w leczeniu zakażeń wywoływanych przez gronkowca złocistego oraz wynikająca z tego nieustanna presja selekcyjna spowodowały, że w dniu dzisiejszym około 99% szczepów jest opornych na penicylinę (Vestergaard et al., 2019).

Nabycie oporności na antybiotyki  $\beta$ -laktamowe, do których należą między innymi metycylina, nafcylina czy oksacylina, jest związane u *S. aureus* z integracją mobilnego elementu genomowego o nazwie SCCmec (ang. *staphylococcal cassette chromosome mec*) na drodze horyzontalnego transferu genów (Lakhundi & Zhang, 2018). Do tej pory opisano 13 różnych typów kasety SCCmec posiadających gen *mecA* lub jego homologi *mecB*, *mecC* i *mecD* (Lakhundi & Zhang, 2018; Lee et al., 2018). Gen *mecA* koduje zmodyfikowaną wersję białka PBP2 (ang. *penicillin-binding protein*), która bierze udział w syntezie peptydoglikanu, stanowiącego istotny element budulcowy bakteryjnej ściany komórkowej (Foster, 2017). Białka PBP stanowią główny cel oddziaływania antybiotyków  $\beta$ -laktamowych, które wiążą się w obrębie centrum aktywnego domeny transpeptydazy PBP, blokując tym samym możliwość syntezy peptydoglikanu, a w konsekwencji prowadząc do śmierci komórki (Foster, 2017; Hodille et al., 2017; Vestergaard et al., 2019). Zmodyfikowana wersja białka PBP2 kodowana przez gen *mecA* nosi nazwę PBP2a. PBP2a, dzięki zmianom strukturalnym w obrębie centrum aktywnego, uniemożliwia wiązanie cząsteczek antybiotyku, pozwalając na efektywną syntezę peptydoglikanu podczas traktowania bakterii antybiotykiem  $\beta$ -laktamowym (np. metycyliną lub oksacyliną) (Foster, 2017). Pomimo że metycylina została już wycofana z użytku klinicznego, termin MRSA pozostał stosowany w odniesieniu do szczepów zawierających gen *mecA* (Vestergaard et al., 2019).

Kaseta SCCmec posiada zazwyczaj od 20 do ponad 60 tysięcy par zasad. W jej skład wchodzi kompleks genów *mec* składający się z genu *mecA* oraz genów regulatorowych *mecRI* oraz *mecI*, a także kompleks *ccr* zawierający geny kodujące rekombinazy odpowiedzialne za integrację mobilnego elementu z genomowym DNA (Pantosti et al., 2007; Vestergaard et al., 2019). Poszczególne typy SCCmec są również nośnikiem wielu innych genów lub elementów mobilnych warunkujących oporność na takie antybiotyki jak tobramycyna, kanamycyna, erytromycyna, spektynomycyna i inne (Vestergaard et al., 2019).

Obiecujące w leczeniu zakażeń wywołanych przez MRSA są podejścia oparte na połączeniu antybiotyku  $\beta$ -laktamowego z dodatkowymi substancjami zwiększającymi podatność bakterii na tego typu leczenie. Przykładem takiej substancji może być kwas klawulanowy działający jako inhibitor  $\beta$ -laktamaz (Vestergaard et al., 2019). W przeciągu ostatnich lat opracowano również antybiotyk  $\beta$ -laktamowy nowej generacji, który w chwili obecnej okazuje się skuteczny przeciwko MRSA w większości przypadków. Ceftarolina wiąże się do centrum allosterycznego białka PBP2a wywołując zmiany strukturalne prowadzące do ekspozycji centrum aktywnego i umożliwiające oddziaływanie drugiej cząsteczki antybiotyku



z centrum aktywnym. Krótko po wprowadzeniu leku do obiegu pojawiły się jednak mutanty o ograniczonej podatności na nowy antybiotyk ze względu na zmiany w obrębie białka PBP2a (Foster, 2017; Vestergaard et al., 2019).

Niezwykle istotna jest także, obserwowana u gronkowca złocistego, zwiększona oporność na wankomycynę oraz jej pochodne. Wankomycyna zaczęła być stosowana w leczeniu zakażeń wywołanych przez MRSA w latach 60. XX wieku. Przez długi czas wankomycyna była uważana za najskuteczniejszy antybiotyk przeciw MRSA (Hodille et al., 2017), jednak i w tym przypadku bakterie wykształciły mechanizmy oporności w odpowiedzi na stosowane leczenie (Lee et al., 2018). Wankomycyna jest antybiotykiem glikopeptydowym wiążącym się do białek PBP2a i PBP2. Zapobiega ona prawidłowemu fałdowaniu peptydoglikanu poprzez blokowanie reakcji transglikozytacji oraz transpeptydacji. Transglikozytacja oraz transpeptydacja to reakcje, które są w komórce katalizowane przez białka PBP (Vestergaard et al., 2019). Zwiększona oporność na wankomycynę (MIC ~4-8µg/ml) jest rezultatem spontanicznych mutacji pojawiających się w trakcie długotrwałej terapii antybiotykowej w obrębie genów biorących udział w syntezie ściany komórkowej i utrzymaniu homeostazy (Foster, 2017; Lee et al., 2018; McGuinness et al., 2017; Pantosti et al., 2007; Vestergaard et al., 2019). Szczepy gronkowca złocistego o zwiększonej oporności na wankomycynę nazywane są VISA (ang. *vancomycin-intermediate-resistant Staphylococcus aureus*).

Stosunkowo rzadko obserwowane są natomiast szczepy o wysokim poziomie oporności na wankomycynę (MIC  $\geq$  ~8-16µg/ml), nazywane VRSA (ang. *vancomycin-resistant Staphylococcus aureus*) (Guo et al., 2020; McGuinness et al., 2017). Obserwowany w tym przypadku determinant oporności, w postaci operonu *vanA*, został prawdopodobnie nabyty od bakterii z rodzaju *Escherichia* (McGuinness et al., 2017; Pantosti et al., 2007). Operon *vanA* zawiera szereg genów kodujących zmodyfikowane prekursorzy peptydoglikanu o znacznie niższym powinowactwie do wankomycyny, ale wciąż możliwych do wykorzystania przez PBP2 do syntezy ściany komórkowej (Foster, 2017; Pantosti et al., 2007). Operon *vanA* zawiera także geny regulatorowe, które umożliwiają syntezę alternatywnych produktów podczas ekspozycji na antybiotyk (McGuinness et al., 2017; Pantosti et al., 2007). Pomimo znacznych obaw związanych z rozprzestrzenianiem tych szczepów, okazały się one posiadać znacznie zredukowaną przeżywalność i nie stanowią obecnie bardzo istotnego zagrożenia klinicznego (Foster, 2017).

Wśród szczepów MRSA spotykana jest także oporność na lipoglikopeptydy, takie jak telawancyna, orytawancyna czy dalbawancyna, będące pochodnymi wankomycyny

(Vestergaard et al., 2019). Zawierają one dodatkowe reszty lipofilowe, które zwiększają czas półtrwania leku oraz ułatwiają interakcję z błonami komórkowymi, podczas gdy cząsteczki leku blokują procesy transglikozylacji oraz transpeptydacji (Vestergaard et al., 2019). Reszty lipofilowe mogą również prowadzić do zaburzenia integralności błon komórkowych i w konsekwencji do lizy komórki (Vestergaard et al., 2019).

Jednym z powszechnie stosowanych antybiotyków w terapii zakażeń wywołanych przez gronkowca złocistego jest daptomycyna, będąca lipopeptydem. Daptomycyna, oddziałując z jonami  $Ca^{2+}$ , penetruje błonę komórkową bakterii umożliwiając wstawienie lipopetydu w strukturę błony poprzez wiązanie leku z posiadającymi ładunek ujemny grupami fosfatydyloglicerolu (Foster, 2017; Guo et al., 2020). Powyższa kaskada reakcji prowadzi do depolaryzacji i permeabilizacji błony komórkowej, a także wycieku jonów poza ustrój komórkowy, co w konsekwencji prowadzi do śmierci komórki bakteryjnej (Foster, 2017; Vestergaard et al., 2019). Oporność na daptomycynę jest wywoływana na drodze spontanicznych mutacji zachodzących w obrębie genów odpowiedzialnych za obronę bakterii przed uszkodzeniami błony i ściany komórkowej, szczególnie podczas długotrwałego leczenia lub w przypadku wysokiego zagęszczenia patogenów (Foster, 2017).

Kolejnym popularnym lekiem stosowanym zarówno przeciwko szczepom podatnym i opornym na metycylinę jest linezolid (Guo et al., 2020). Cząsteczki antybiotyku wiążą się z podjednostką 23S rRNA, blokując tym samym możliwość prawidłowego oddziaływania rybosomu z aminoacylo-tRNA, co prowadzi do zahamowania procesów syntezy białek. Również w tym przypadku zidentyfikowano mutacje (w obrębie genów 23S rRNA), które znacząco zwiększają poziom odporności bakterii na opisywany lek (Hodille et al., 2017; Vestergaard et al., 2019). Oporność na linezolid może być również związana z nabyciem genu *cfz*, który bardzo często jest zasocjowany z mobilnymi elementami genomowymi lub plazmidami. Gen *cfz* koduje metylotransferazę, która jest odpowiedzialna za metylację pozycji A2503 w centrum peptydotransferazy 23S rRNA uniemożliwiając prawidłowe oddziaływanie antybiotyku (Vestergaard et al., 2019). Podobnie działającym lekiem jest tedizolid, który pozostaje skuteczny również w przypadku szczepów posiadających gen *cfz* (Vestergaard et al., 2019). Pomimo istnienia powyższych mechanizmów oporności, zdecydowana większość szczepów (>98%) gronkowca złocistego pozostaje podatna na działanie obu antybiotyków (Vestergaard et al., 2019).

Do grona istotnych klinicznie antybiotyków, ukierunkowanych przeciw szczepom gronkowca złocistego, należą również makrolidy, linkozamidy, streptograminy, czy pleuromutyliny (Foster, 2017; Vestergaard et al., 2019). Wszystkie ze wspomnianych grup

antybiotyków są odpowiedzialne za blokowanie syntezy białek poprzez oddziaływanie z centrum peptydotransferazy podjednostki 50S rybosomu prokariotycznego (Vestergaard et al., 2019). W tym przypadku mechanizmy oporności są zazwyczaj związane z mutacjami lub modyfikacjami w obrębie centrum peptydotransferazy, aktywnym usuwaniem antybiotyku lub jego inaktywacją (Vestergaard et al., 2019). Przykładami są tutaj opisane wcześniej mechanizmy oparte o geny z rodziny *erm* oraz *cfr*. Antybiotyki należące do powyższych grup mogą być aktywnie usuwane poza komórki przy pomocy białek transportujących Msr(A), LmrS czy vga(A) (Vestergaard et al., 2019). Innymi antybiotykami, blokującymi syntezę białek poprzez wiązanie w pobliżu centrum peptydotransferazy, są chloramfenikol i florfenikol, które jednak ze względu na wysoką toksyczność są stosowane jedynie miejscowo (Foster, 2017; Vestergaard et al., 2019).

Tetracykliny stanowią grupę antybiotyków, które blokują możliwość oddziaływania pomiędzy aminoacylo-tRNA a rybosomem poprzez wiązanie do podjednostki 30S (Hodille et al., 2017). Przykładami powszechnie stosowanych antybiotyków tetracyklinowych są tygecyklina i doksycyklina (Hodille et al., 2017). Oporność na leki w tym przypadku może być warunkowana działaniem pomp Tet(K) oraz Tet(L), bądź poprzez wiązanie białek TetO lub TetM do podjednostki 30S, które w konsekwencji prowadzi do uwolnienia związanej do rybosomu tetracykliny (Foster, 2017; Vestergaard et al., 2019).

Ciekawym mechanizmem działania charakteryzują się antybiotyki z rodziny aminoglikozydów, takie jak neomycyna, kanamycyna czy gentamycyna, które wiążąc się do 16S rRNA w obrębie podjednostki 30S rybosomu prokariotycznego, powodują częste błędy w interakcjach pomiędzy kodonem a antykodonem, co w konsekwencji skutkuje zwiększoną liczbą błędów na etapie translacji oraz niepoprawnie syntezowanymi białkami (Foster, 2017; Vestergaard et al., 2019). Najistotniejszy mechanizm oporności na aminoglikozydy jest oparty o metylotransferazę 16S rRNA (*rmtC*), której działanie uniemożliwia wiązanie cząsteczki antybiotyku do miejsca aktywnego (Vestergaard et al., 2019).

Kolejnym antybiotykem wykorzystywanym przeciwko gronkowcom jest mupirocyna. Mupirocyna jest antybiotykem wykorzystywanym miejscowo do leczenia zakażeń skórnych wywołanych przez gronkowca złocistego (Dadashi et al., 2020). Jej sposób działania bazuje na hamowaniu syntezy białek poprzez blokowanie syntetazy izoleucyno-tRNA (Dadashi et al., 2020; Foster, 2017; Vestergaard et al., 2019). Oporność na mupirocynę uzyskiwana jest na drodze mutacji w obrębie genu kodującego syntetazę aminoacylo-tRNA (*ileS*), które zmniejszają jej powinowactwo do cząsteczek antybiotyku (Dadashi et al., 2020; Foster, 2017;

Vestergaard et al., 2019). Możliwe jest również nabycie plazmidowego genu *mupA*, który koduje syntetazę odporną na działanie leku (Dadashi et al., 2020; Vestergaard et al., 2019).

Antybiotykiem o podobnym zastosowaniu do mupirocyny jest kwas fusydynowy, wiążący się do czynnika elongacyjnego G, który jest odpowiedzialny za translokację mRNA oraz tRNA w obrębie rybosomu. Opisywane wiązanie prowadzi do wstrzymania translacji, hamując tym samym rozwój bakterii (Foster, 2017; A. J. O'Neill & Chopra, 2006; Rayner & Munckhof, 2005; Vestergaard et al., 2019). W przypadku kwasu fusydynowego oporność może być wynikiem mutacji w obrębie genu *fusA*, który koduje czynnik elongacyjny G. Zmiana konformacji białka, będąca wypadkową mutacji, może zmniejszać powinowactwo leku do czynnika elongacyjnego G (Vestergaard et al., 2019). Inne znane mechanizmy uzyskiwania oporności przez bakterie uwzględniają mutacje w obrębie genu *fusE*, kodującego rybosomalne białko L6, a także genów *fusB* i *fusC*, których produkty wiążą się do czynnika elongacyjnego G, powodując wznowienie translacji (Foster, 2017; A. J. O'Neill & Chopra, 2006; Vestergaard et al., 2019).

Ryfampicyna należy do grupy antybiotyków, które działają poprzez hamowanie transkrypcji. Ryfampicyna wiąże się do podjednostki  $\beta$  prokariotycznej polimerazy RNA uniemożliwiając prawidłowe przeprowadzanie procesu transkrypcji (Foster, 2017; Rayner & Munckhof, 2005; Vestergaard et al., 2019). Odporność na opisywany antybiotyk jest spowodowana mutacją w obrębie genu *rpoB*, który koduje podjednostkę  $\beta$  polimerazy RNA (Rayner & Munckhof, 2005). Mutacja ta powoduje znaczne zmniejszenie powinowactwa ryfampicyny do polimerazy (Vestergaard et al., 2019).

Kolejną grupą antybiotyków mającą zastosowanie w leczeniu zakażeń wywoływanych przez gronkowca złocistego są fluorochinolony, których działanie oparte jest na oddziaływaniu z topoizomerazą IV oraz gyrazą DNA (topoizomeraza II), które to enzymy pełnią kluczowe role, m. in. w replikacji DNA (Foster, 2017; Pietrusinski & Stacek, 2006; Vestergaard et al., 2019). Gyraza DNA jest odpowiedzialna za wprowadzanie negatywnych superskrętów DNA, natomiast topoizomeraza IV bierze udział w rozplataniu chromosomów w procesach replikacji oraz relaksacji DNA (Pietrusinski & Stacek, 2006). Oba enzymy wprowadzają dwuniciowe nacięcia w obrębie DNA, które muszą być w kolejnych etapach reakcji ponownie połączone (Pietrusinski & Stacek, 2006). Powstanie kompleksu topoizomerazy i fluorochinolonu uniemożliwia ponowne łączenie nici DNA, prowadząc do powstania dwuniciowych przerw w obrębie DNA, co w konsekwencji powoduje śmierć komórki (Foster, 2017). Oporność na fluorochinolony jest powszechna i wynika z mutacji w obrębie genów kodujących opisywane

enzymy (np. *gyrA* oraz *grlA*), a także z nadekspresji pomp typu efflux *NorA*, *NorB* oraz *NorC* (Foster, 2017; Vestergaard et al., 2019).

Powyższa lista antybiotyków wykorzystywanych w leczeniu różnego rodzaju zakażeń wywoływanych przez gronkowca złocistego nie jest wyczerpująca, stanowiąc jedynie przegląd najbardziej istotnych klinicznie leków stosowanych w medycynie i weterynarii. Problem antybiotykoodporności w przypadku gronkowca złocistego jest znaczący, co znajduje wyraz w zidentyfikowaniu szczepów opornych na każdy z przedstawionych leków oraz takich, które charakteryzują się jednoczesną opornością na wiele różnych klas antybiotyków, znacząco utrudniając skuteczną i szybką terapię. Niezwykle istotna jest zatem konieczność opracowania nowych klas leków, dla których uzyskanie oporności przez bakterie będzie mniej prawdopodobne lub obarczone niskim poziomem przeżywalności.

## 7.7. Metody identyfikacji miejsc terminacji transkrypcji

Nieustanna konieczność opracowywania nowych klas antybiotyków, spowodowana szybko rozprzestrzeniającym się zjawiskiem lekooporności u bakterii, powoduje konieczność zastosowania nowoczesnych metod badawczych dedykowanych poszukiwaniu potencjalnych nowych celów terapeutycznych. Mechanizmy regulatorowe oparte o cząsteczki RNA, w tym o elementy struktury drugorzędowej, stanowią niezwykle obiecujące cele dla nowych leków. Szczególnie interesujące są te bazujące na warunkowej terminacji transkrypcji, które pozwalają regulować ekspresje genów w odpowiedzi na różnorodne sygnały komórkowe, w tym oddziaływanie konkretnych substancji na komórki (Chełkowska-Pauszek et al., 2021). Identyfikacja miejsc terminacji transkrypcji w skali całego transkryptomu wymaga zastosowania metod wysokoprzepustowego sekwencjonowania 3' końców RNA oraz dedykowanych metod obliczeniowych i analitycznych.

Ostatnie kilkanaście lat to nieprzerwany i bardzo dynamiczny rozwój metod sekwencjonowania nowej generacji, pozwalających na równoległe i zautomatyzowane sekwencjonowanie wielu milionów fragmentów DNA równocześnie. Obecnie jedna z najpopularniejszych metod sekwencjonowania nowej generacji jest oparta na platformie zaadoptowanej przez firmę Illumina. Wspomniana technologia bazuje na zastosowaniu specjalnych, szklanych płytek, na których kowalencyjnie związane są miliony oligonukleotydów komplementarnych do sekwencji adaptorowych przyłączanych do 5' i 3' końców krótkich cząsteczek DNA (zazwyczaj rzędu kilkuset nukleotydów) na etapie tworzenia tzw. 'biblioteki'. Istnieją dwa typy oligonukleotydów komplementarnych odpowiednio do sekwencji 5' i 3' adaptera.

Metoda wykorzystywana przez firmę Illumina opiera się na reakcji 'mostkowego PCR', prowadzącej do powstania klastrów identycznych fragmentów DNA zlokalizowanych w swoim bezpośrednim sąsiedztwie na płytce. Pierwszym etapem wspomnianej reakcji jest hybrydyzacja krótkich fragmentów DNA zawierających przyłączone sekwencje 5' i 3' adaptorów, do oligonukleotydów związanych z płytką. Następnie przeprowadzana jest synteza nici komplementarnej do danego fragmentu z wykorzystaniem mieszaniny zawierającej trójfosforany nukleotydów oraz polimerazę DNA.

Kolejnym etapem reakcji jest denaturacja DNA i usuwanie nici stanowiącej pierwotną matrycę. W ten sposób wszystkie sekwencjonowane fragmenty są związane kowalencyjnie

z płytką. Ze względu na gęste rozmieszczenie oligonukleotydów na płytce, nie może hybrydyzować przy pomocy wolnego regionu adaptorowego do drugiego typu oligonukleotydu, tworząc tzw. 'mostki'. Opisywane 'mostki' składają się z sekwencjonowanego fragmentu DNA połączonego w danym momencie z płytką na obu końcach. Polimeraza DNA syntezuje nić komplementarną, po czym następuje denaturacja nici, prowadząc tym samym do powielenia matrycy. Reakcja mostkowego PCR przeprowadzana jest wielokrotnie prowadząc do utworzenia grup (klastrów) identycznych fragmentów DNA związanych z płytką i zlokalizowanych w bezpośrednim sąsiedztwie względem siebie. Nici związane z jednym typem oligonukleotydu są następnie odmywane, a sekwencjonowanie przeprowadzane jest w cyklach na zasadzie syntezy.

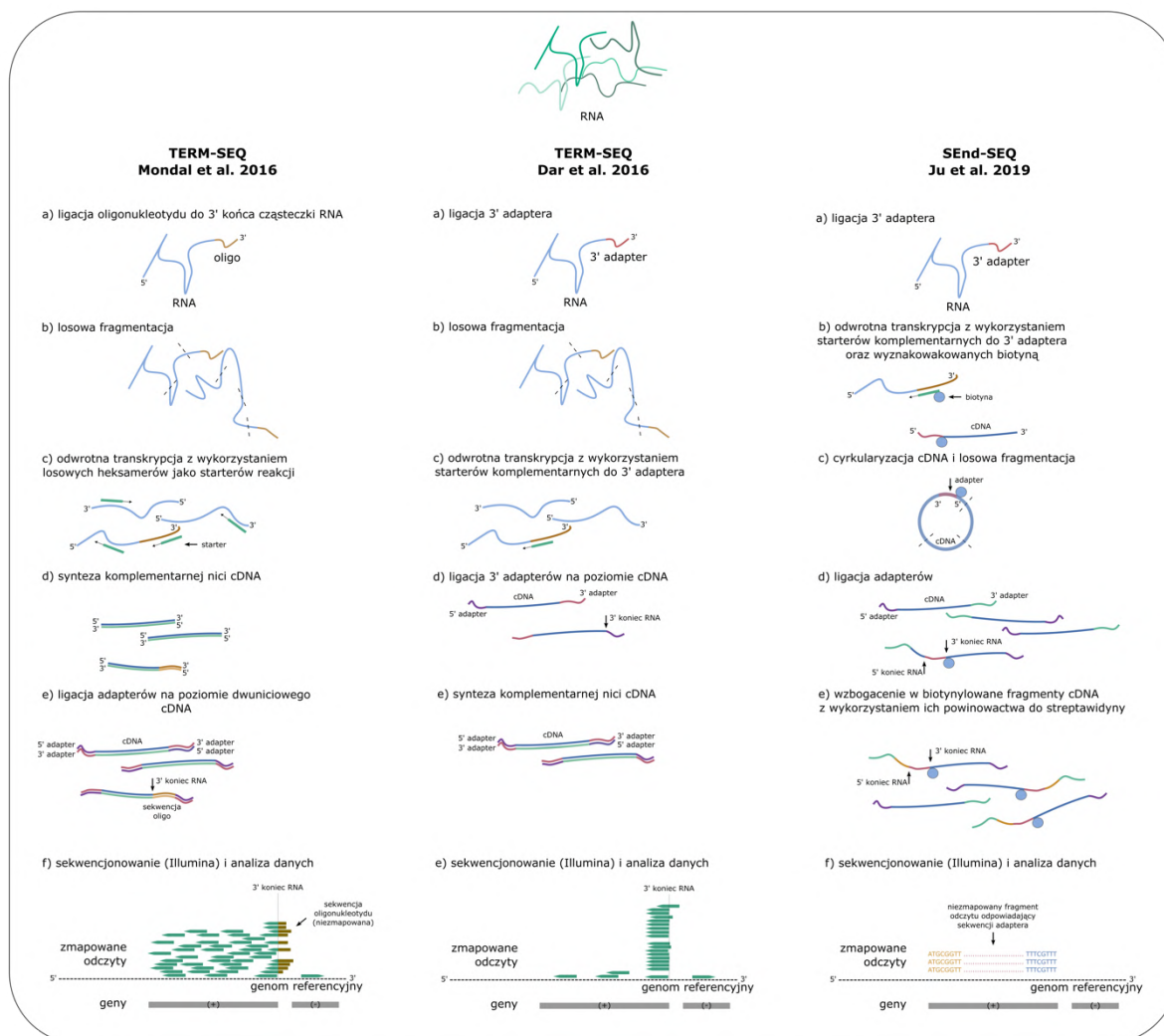
Inicjalizacja procesu jest możliwa dzięki hybrydyzacji starterów reakcji do regionu adaptorowego. W każdym kolejnym cyklu sekwencjonowania, dodawany jest dokładnie jeden wyznakowany fluorescencyjnie nukleotyd. Dodawane nukleotydy są chemicznie zablokowane, żeby uniemożliwić inkorporację w danym cyklu większej ich liczby. Wstawienie nukleotydu do nowo syntezowanej nici powoduje emisję sygnału świetlnego o kolorze przypisanym do danego typu nukleotydu. Sygnał jest wzmacniany dzięki obecności setek lub tysięcy identycznych fragmentów DNA znajdujących się obok siebie. Po odczytaniu sygnału świetlnego przy pomocy specjalnego sensora, nukleotydy są odblokowywane, umożliwiając tym samym przeprowadzenie kolejnego cyklu sekwencjonowania. Opisywana technologia umożliwia uzyskanie milionów, a nawet miliardów odczytów, zazwyczaj o długości do 300 nukleotydów (Hu et al., 2021; Slatko et al., 2018).

Jednym z powszechnych zastosowań platformy Illumina jest wysokoprzepustowe sekwencjonowanie transkryptomu przy pomocy protokołu RNA-seq (ang. *RNA sequencing*). Metoda RNA-seq pozwala uzyskać informacje o obecności i częstości występowania cząsteczek RNA w obrębie badanego transkryptomu. Sekwencjonowanie transkryptomu przy pomocy platformy Illumina wymaga przeprowadzenia fragmentacji RNA, odwrotnej transkrypcji uzyskanych fragmentów do cDNA oraz ligacji adapterów (Zhang et al., 2018). Możliwe jest również zastosowanie różnorodnych metod polegających na wzbogaceniu biblioteki we fragmenty pochodzące z wybranej frakcji transkryptomu, np. mRNA czy sRNA. RNA-seq jest stosowany powszechnie w celu szacowania poziomu ekspresji genów, badań alternatywnego splicingu, identyfikacji nowych izoform splicingowych czy fuzji genowych (Kukurba & Montgomery, 2015).

Na przestrzeni ostatnich lat opracowane zostały trzy popularne metody wysokoprzepustowego sekwencjonowania RNA pozwalające na precyzyjną identyfikację 3'

końców RNA u bakterii. Dwie ze wspomnianych metod powstały równolegle w roku 2016, funkcjonując do dziś pod nazwą Term-seq (Dar et al., 2016; Mondal et al., 2016). Protokół opracowany przez Mondal i in. (Mondal et al., 2016) (Rys. 4) opiera się na ligacji 2',3'-dideoxy oligonukleotydu do 3' końca cząsteczek RNA oraz ich losowej fragmentacji. Kolejnymi etapami procesu są odpowiednio odwrotna transkrypcja fragmentów RNA do cDNA przy użyciu losowych heksametrów pełniących rolę starterów, synteza nici komplementarnej, ligacja 5' i 3' adapterów oraz rybodeplecja. Biblioteka przygotowywana jest w oparciu o niciowo-specyficzny protokół Illumina TruSeq stranded i sekwencjonowana przy użyciu dedykowanej platformy (Mondal et al., 2016). Odczyty uzyskane w ten sposób można podzielić na dwa zestawy w oparciu o to, czy w ich obrębie można zidentyfikować sekwencję oligonukleotydu wykorzystanego w czasie przygotowania biblioteki (Mandell et al., 2021; Mondal et al., 2016). Grupa odczytów, w których sekwencja oligonukleotydu została zlokalizowana, może posłużyć do precyzyjnego wyznaczenia 3' końców RNA po uprzednim usunięciu jego sekwencji oraz zmapowaniu do genomu referencyjnego. Metoda została opracowana w ten sposób, że 5' koniec takiego odczytu wskazuje bezpośrednio na 3' koniec cząsteczki RNA. Pozostałe odczyty można wykorzystać jak te pochodzące z sekwencjonowania transkryptomu powszechnie stosowaną metodą RNA-seq, np. do profilowania poziomów ekspresji genów.





Rysunek 4. Schematyczna reprezentacja protokołów wysokoprzepustowego sekwencjonowania 3' końców RNA u bakterii.

Metoda zaproponowana przez Dar i in. (Dar et al., 2016) (Rys. 4) polega na ligacji adaptera do 3' końca cząsteczek RNA, losowej fragmentacji RNA i inicjacji odwrotnej transkrypcji z wykorzystaniem specjalnie zaprojektowanych oligonukleotydów komplementarnych do sekwencji adaptera i wykorzystywanych jako startery wspomnianej reakcji. Następnym wymaganym krokiem w ramach opisywanego protokołu jest ligacja 3' adaptera do jednoniciowego cDNA oraz synteza drugiej nici cDNA. W skład niciowo-specyficznej biblioteki cDNA wchodzi wyłącznie fragmenty RNA zawierające oba dołączone elementy. Sekwencjonowanie również oparte jest o platformę Illumina. Zastosowanie powyższej metody prowadzi do uzyskania odczytów, których 5' końce odpowiadają 3' końcom sekwencjonowanych cząsteczek RNA (Dar et al., 2016).

Kolejna metoda sekwencjonowania wysokoprzepustowego pozwalająca na jednoczesną identyfikację 5' i 3' końców cząsteczek RNA to protokół o nazwie SEnd-seq (ang. *simultaneous 5' and 3' end sequencing*, Rys. 4) bazujący na cyrkularyzacji cDNA przy pomocy dedykowanych ligaz, które preferencyjnie łączą końce 5' i 3' w obrębie tej samej cząsteczki RNA (Ju et al., 2019). Pierwszym etapem zaprezentowanej technologii jest ligacja 3' adaptera do cząsteczek RNA. cDNA jest syntezowane przy pomocy starterów komplementarnych do sekwencji 3' adaptera, dodatkowo wyznakowanych biotyną na 5' końcu. Kolejnym etapem jest, wspomniana już, cyrkularyzacja cząsteczek cDNA, po której następuje ich fragmentacja (Ju et al., 2019). Próbkę są wzbogacane we fragmenty cDNA wyznakowane biotyną, z wykorzystaniem ich powinowactwa do streptawidyny. Uzyskiwane w ten sposób odczyty zawierają sekwencje odpowiadające 5' i 3' końcom tej samej cząsteczki RNA, rozdzielone od siebie sekwencją 3' adaptera (Ju et al., 2019).

Zastosowanie omawianych protokołów znacząco ułatwiło przeszukiwanie transkryptomów bakteryjnych pod kątem występowania ryboprzełączników, atenuatorów, jak również innych potencjalnych regulatorów prowadzących do warunkowej terminacji transkrypcji. Wyniki otrzymane przy pomocy przedstawianych metod pozwoliły, m.in. na zidentyfikowanie atenuatora regulującego ekspresję transporterów ABC, warunkującego oporność na linkomycynę u *Listeria monocytogenes* (Dar et al., 2016).

## 8. Cel pracy

Głównym celem niniejszej pracy jest identyfikacja i charakterystyka rho-niezależnych terminatorów transkrypcji u bakterii oraz ich udziału w odpowiedzi na stres antybiotykowy u gronkowca złocistego. Szczególny nacisk zostanie położony na adnotację oraz porównanie charakterystyki wspomnianych terminatorów u 14 gatunków bakterii, tak by możliwe było dokładne określenie ich zróżnicowania w obrębie królestwa bakterii. Ponadto, zidentyfikowane zostaną terminatory zaangażowane w regulację ekspresji genów gronkowca złocistego w odpowiedzi na stres antybiotykowy, mogące stanowić potencjalne cele terapeutyczne dla nowo opracowywanych klas leków. W tym celu opracowane zostaną nowatorskie metody obliczeniowej analizy danych pochodzących z wysokoprzepustowych protokołów Term-seq oraz RNA-seq.

### Szczegółowe cele:

- opracowanie metod oraz oprogramowania bioinformatycznego pozwalającego na:
  - identyfikację oraz adnotację terminatorów na podstawie danych z wysokoprzepustowego sekwencjonowania 3' końców RNA (Term-seq)
  - przewidywanie terminatorów na podstawie danych z wysokoprzepustowego sekwencjonowania RNA (RNA-seq)
  - adnotację interwałów w obrębie transkryptów z wykorzystaniem referencyjnej adnotacji genomowej
- kompleksowa identyfikacja rho-niezależnych terminatorów transkrypcji u bakterii, ze szczególnym uwzględnieniem gronkowca złocistego, w oparciu o opracowane metody analizy danych z wysokoprzepustowego sekwencjonowania RNA
- charakterystyka rho-niezależnych terminatorów transkrypcji w królestwie bakterii oraz analiza różnic występujących pomiędzy taksonami
- identyfikacja rho-niezależnych terminatorów transkrypcji regulujących ekspresję genów w odpowiedzi na stres antybiotykowy u gronkowca złocistego

## 9. Metody

### 9.1. Wykorzystane zestawy danych

#### Term-seq

Publicznie dostępne dane pochodzące z sekwencjonowania 3' końców RNA metodą Term-seq, wykorzystane do identyfikacji rho-niezależnych terminatorów transkrypcji u *Escherichia coli*, *Bacillus subtilis*, *Enterococcus faecalis*, *Listeria monocytogenes*, *Zymomonas mobilis*, *Synechocystis sp*, *Streptomyces coelicolor*, *Streptomyces venezuelae*, *Streptomyces tsukubaensis*, *Streptomyces griseus*, *Streptomyces lividans*, *Streptomyces clavuligerus* oraz *Streptomyces avermitilis*, zostały szczegółowo przedstawione i opisane w ramach publikacji narzędzia TERMITE, która stanowi element niniejszej rozprawy doktorskiej (Rozdział 9.1).

Dane o numerze akcesyjnym GSE158830, na podstawie których przeprowadzone zostały analizy u szczepu dzikiego gronkowca złocistego (JKD6009), zostały pobrane z bazy danych *Gene Expression Omnibus* (<https://www.ncbi.nlm.nih.gov/geo/>).

#### RNA-seq

Przewidywanie rho-niezależnych terminatorów gronkowca złocistego, na podstawie danych RNA-seq, zostało przeprowadzone w oparciu o analizę 138 publicznie dostępnych zestawów danych (próbek) pochodzących z 7 niezależnych projektów (Choe et al., 2018; Gao et al., 2020; Mediati et al., 2022; Poudel et al., 2020; Rajput et al., 2021; Viering et al., 2022). W ramach opisywanych badań szereg metycylinoopornych szczepów gronkowca złocistego poddawany był działaniu subletalnych dawek antybiotyków, takich jak wankomycyna, linezolid, nafecylina czy oksacylina. Szczegółowy wykaz zestawów danych, wykorzystanych w ramach niniejszych badań, stanowi załącznik do niniejszej pracy (załącznik 1). Odczyty zostały pobrane z bazy danych Sequence Read Archive dostępnej pod adresem <https://www.ncbi.nlm.nih.gov/sra> (Leinonen et al., 2011).

## **termiRNAtor**

Na potrzeby wytrenowania i ewaluacji klasyfikatora programu termiRNAtor wykorzystano 128 publicznie dostępnych próbek pochodzących z sekwencjonowania 3' końców RNA metodą TERM-seq wraz z odpowiadającymi im próbami RNA-seq uzyskanymi w ramach tych samych warunków eksperymentalnych (załącznik 10).

## 9.2. Identyfikacja rho-niezależnych terminatorów transkrypcji na podstawie danych Term-seq

Proces identyfikacji rho-niezależnych terminatorów transkrypcji na podstawie danych Term-seq u *E. coli*, *B. subtilis*, *E. faecalis*, *L. monocytogenes*, *Z. mobilis*, *Synechocystis sp.*, *S. coelicolor*, *S. venezuelae*, *S. tsukubaensis*, *S. griseus*, *S. lividans*, *S. clavuligerus* oraz *S. avermitilis* został szczegółowo opisany w ramach publikacji narzędzia TERMITE (wersja 1.0.0), która stanowi element niniejszej rozprawy doktorskiej (Rozdział 9.1). Identyczna metodologia została wykorzystana do identyfikacji terminatorów u gronkowca złocistego.

Na potrzeby sporządzenia referencyjnego zestawu rho-niezależnych terminatorów transkrypcji gronkowca złocistego, wyselekcjonowane zostały wyłącznie terminatory posiadające zidentyfikowaną spinkę terminacyjną przewidzianą przez moduły *RNAfold* (spinka o minimalnej energii swobodnej struktury niższej niż -3 kcal/mol) lub *TransTermHP* programu TERMITE (szczegóły: rozdział 9.1). Na potrzeby porównań pomiędzy gatunkami wykorzystano natomiast kryteria filtrowania opisane w ramach publikacji narzędzia TERMITE (Rozdział 9.1).

### 9.3. Przewidywanie rho-niezależnych terminatorów transkrypcji na podstawie danych RNA-seq u gronkowca złocistego

Zastosowany w prezentowanej rozprawie proces adnotacji rho-niezależnych terminatorów transkrypcji u gronkowca złocistego wykorzystuje szereg metod obliczeniowych bazujących na analizie sekwencji nukleotydowej, metodach przewidywania struktury drugorzędowej RNA, przewidywaniu miejsc terminacji transkrypcji na podstawie danych RNA-seq oraz ich precyzyjnym określaniu na podstawie danych Term-seq. Wykorzystanie danych RNA-seq pochodzących z eksperymentów, w których metycylinooporne szczepy gronkowca złocistego podlegały traktowaniu subletalnymi dawkami wankomycyny, linezolidu, nafcyliny czy oksacyliny, pozwoliło na identyfikację terminatorów, których działanie jest indukowane odpowiedzią komórki na działanie leków.

Analiza jakości odczytów została przeprowadzona przy pomocy programów *FASTQC* (wersja 0.11.9) (Andrews, 2010) oraz *MultiQC* (wersja 1.14) (Ewels et al., 2016). Sekwencje adaptorowe zostały natomiast usunięte za pomocą programu *Trim Galore!* z zastosowaniem domyślnych parametrów (wersja 0.6.7) (Krueger, 2012).

W ramach kolejnego etapu odczyty zostały zmapowane do genomu referencyjnego gronkowca złocistego (GCF\_000013425.1) przy pomocy programu *bowtie* (wersja 1.2.0; użyte parametry: *--best*, *--strata*, *-a*) (Langmead et al., 2009), a format plików wynikowych zmieniony na format BAM z wykorzystaniem programu *samtools view* (Danecek et al., 2021). Zmapowane odczyty zostały posortowane wg. koordynatów genomowych za pomocą narzędzia *sort* z pakietu *samtools* (Danecek et al., 2021), natomiast analiza jakości mapowania została wykonana przy użyciu narzędzia *bamQC* z pakietu *QualiMap* (wersja 2.2.1) (Okonechnikov et al., 2016).

Kluczowy etap analizy danych stanowiło uruchomienie programu *termiRNator*, który został opracowany na potrzeby niniejszej rozprawy doktorskiej, a którego celem jest przewidywanie miejsc rho-niezależnej terminacji transkrypcji u bakterii na podstawie danych RNA-seq z wykorzystaniem metod uczenia maszynowego (został on opisany szczegółowo w rozdziale 9.2). Na potrzeby opisywanego procesu predykcji wykorzystano 138 zestawów danych, które opisane zostały we wcześniejszych fragmentach bieżącego rozdziału. Przewidywanie miejsc terminacji transkrypcji (użyte parametry *-u 150 -d 20*) zostało wykonane dla każdej grupy replikatów biologicznych niezależnie (załącznik 1). W celu dodatkowego

potwierdzenia, że zidentyfikowane przez *termiRNAtor* pozycje są wynikiem rho-niezależnej terminacji transkrypcji, wykorzystany został program *TransTermHP* (wersja 2.0.9) z domyślnymi parametrami (Kingsford et al., 2007). *TransTermHP* jest programem pozwalającym na identyfikację rho-niezależnych terminatorów transkrypcji w genomach bakteryjnych na podstawie sekwencji nukleotydowej podanego genomu referencyjnego.

Do końcowego zestawu rho-niezależnych terminatorów transkrypcji, przewidzianych na podstawie danych RNA-seq, zostały wybrane te, spośród znalezionych przez *TransTermHP* terminatorów, których spinki terminacyjne pokrywały pozycje przewidziane przez *termiRNAtor* lub znajdowały się od nich w odległości nie większej niż 10 nt powyżej (w obrębie tej samej nici). Ponadto, określone wyżej warunki musiały być spełnione w co najmniej 3 analizowanych próbkach. Powyższe filtrowanie zostało wykonane przy pomocy narzędzia *closest* należącego do pakietu *bedtools* (parametry -s -D) (Quinlan, 2014; Quinlan & Hall, 2010).



## 9.4. Utworzenie referencyjnego atlasu rho-niezależnych terminatorów transkrypcji u gronkowca złocistego

Utworzenie możliwie kompletnego zestawu rho-niezależnych terminatorów transkrypcji u gronkowca złocistego było możliwe dzięki zastosowaniu szeregu metod pozwalających na ich identyfikację, zarówno na podstawie wyników sekwencjonowania wysokoprzepustowego metodami RNA-seq oraz Term-seq, a także analizy sekwencji nukleotydowych oraz struktury drugorzędowej RNA. W ramach niniejszego atlasu połączone zostały zestawy terminatorów zidentyfikowanych na podstawie danych RNA-seq oraz Term-seq (rozdziały 8.3 oraz 8.4). W celu uniknięcia redundancji danych w obrębie tworzonego atlasu, terminatory które zostały przewidziane na podstawie danych RNA-seq, a mające swój odpowiednik w ramach terminatorów określonych z wykorzystaniem danych Term-seq, zostały odrzucone.

Referencyjny punkt terminacji (*POT - Point of Termination*) dla terminatorów znalezionych wyłącznie na podstawie danych RNA-seq został ustawiony na pozycję położoną 5 nukleotydów poniżej podstawy zidentyfikowanej spinki do włosów. Wartość ta została określona na podstawie mediany odległości punktów terminacji od podstawy spinek terminacyjnych, która z kolei została obliczona w ramach analizy terminatorów zidentyfikowanych z wykorzystaniem danych Term-seq (pierwszy kwartyl = 4 nt; trzeci kwartyl = 6nt; rozdział 9). Ponadto funkcje programu *TERMITE* oraz program *AGoutI* (będący częścią niniejszej rozprawy, rozdział 9.3, wersja 1.0.3) zostały wykorzystane do uzupełnienia brakujących informacji dla terminatorów znalezionych wyłącznie na podstawie danych RNA-seq przy pomocy programu *termiRNAtor*. Zastosowanie powyżej opisanego zestawu operacji umożliwiło uzyskanie identycznego zestawu adnotacji dla każdego terminatora w obrębie atlasu.

## 9.5. Identyfikacja miejsc warunkowej terminacji transkrypcji w odpowiedzi na stres antybiotykowy u gronkowca złocistego

Miejsca warunkowej terminacji transkrypcji, indukowane w odpowiedzi na stres antybiotykowy, zostały określone na podstawie danych RNA-seq (załącznik 1). W ramach przedstawionych badań prezentowany jest wpływ różnych stężeń antybiotyków, takich jak linezolid, wankomycyna, nafcylina, czy oksacylina, na wykorzystanie poszczególnych miejsc rho-niezależnej terminacji transkrypcji. Podstawę analizy stanowiło obliczenie efektywności terminacji dla każdego terminatora obecnego w opracowanym atlasie rho-niezależnych terminatorów transkrypcji u gronkowca złocistego.

Efektywność terminacji jest miarą obliczaną niezależnie dla każdego warunku eksperymentalnego (np. kontrola, traktowanie antybiotykiem), informującą użytkownika o zmianie poziomu transkrypcji występującej w regionie danego terminatora. Wartość ta określana jest na podstawie plików *bigWig* reprezentujących znormalizowane pokrycie każdej pozycji w genomie zmapowanymi odczytami pochodzącymi z wysokoprzepustowego sekwencjonowania metodą RNA-seq. Normalizacja pokrycia odbywa się poprzez określenie liczby odczytów przypadających na daną pozycję na każdy milion odczytów znajdujących się w obrębie danej biblioteki (CPM, ang. *counts per milion mapped reads*). Wspomniane pliki *bigWig* zostały utworzone przy pomocy narzędzia *bamCoverage* wchodzącego w skład pakietu *deeptools* (wersja 3.5.0; użyte parametry `--binSize 1, --normalizeUsing CPM`) (Ramírez et al., 2016) na podstawie plików w formacie BAM opisanych we wcześniejszych częściach niniejszego rozdziału.

Efektywność terminacji (%T) dana jest wzorem:

$$\%T = \max\left(\frac{u-d}{u} * 100, 0\right), \text{ gdzie}$$

$u$  reprezentuje średnie pokrycie odczytami 90 nukleotydów znajdujących się *powyżej*, natomiast  $d$  15 nukleotydów *poniżej* od referencyjnego punktu terminacji (POT). Powyższe wartości zostały dobrane w sposób empiryczny na podstawie wizualizacji danych w programie IGV (Thorvaldsdóttir et al., 2013). Istotnym faktem jest wyłączenie z opisanych obliczeń pokrycia 3 pierwszych pozycji zarówno *powyżej* jak i *poniżej* względem punktu terminacji.

Ponadto, wartość %T została wyznaczona jedynie dla terminatorów, u których nieznormalizowane średnie pokrycie odczytami *powyżej* wynosiło co najmniej 10 odczytów.

Efektywność terminacji (%T) została obliczona niezależnie dla każdego warunku eksperymentalnego (replikatu biologicznego). Ponadto, wartości %T zostały porównane pomiędzy każdym warunkiem eksperymentalnym (próby traktowane antybiotykiem), a odpowiadającą mu próbą kontrolną (załącznik 1), poprzez obliczenie wartości DT danej wzorem:

$$DT = t - c, \text{ gdzie}$$

*t* oznacza wartość średnią obliczoną dla replikatów reprezentujących próbę traktowaną antybiotykiem, natomiast *c* próbę kontrolną. Miary %T oraz DT, zostały już wcześniej zaproponowane do oceny efektywności terminacji w ramach publikacji naukowych (Mandell et al., 2021; Mondal et al., 2016).

Implementacja obustronnego testu T-Studenta *scipy.stats.ttest\_ind* (Virtanen et al., 2020) pozwoliła na weryfikację hipotezy zerowej mówiącej o równości średnich %T pomiędzy każdą parą porównywanych grup. Ze względu na znaczną liczbę przeprowadzonych testów statystycznych wykonano poprawkę na wielokrotne testowanie przy pomocy metody FDR (*ang. false discovery rate*) Benjamini’ego/Hochberg’a, zaimplementowaną w obrębie funkcji *statsmodels.stats.multitest.fdr\_correction* (Seabold & Perktold, 2010).

Na podstawie uzyskanych wartości, w ramach poszczególnych porównań, wyselekcjonowane zostały miejsca terminacji posiadające wartość  $|DT| \geq 30$  oraz  $FDR < 0.05$ . Uzyskana lista została poddana inspekcji w przeglądarce genomowej IGV w celu usunięcia wątpliwych przypadków, mogących stanowić obserwacje fałszywie pozytywne, m.in. warunkowane ekspresją sąsiadujących genów. Wynikiem opisanej analizy jest zestaw miejsc warunkowej terminacji transkrypcji aktywowanych lub deaktywowanych w odpowiedzi na stres antybiotykowy u gronkowca złocistego.

## 10. Wyniki i dyskusja

### 10.1. *TERMITE* – metoda identyfikacji i adnotacji stabilnych 3' końców RNA na podstawie wyników sekwencjonowania z zastosowaniem protokołu TERM-seq

Szczegółowy opis opracowanej metody został przedstawiony w ramach publikacji, stanowiącej integralną część niniejszej pracy doktorskiej

Kosiński JG, Ranaweera S, Chełkowska-Pauszek A, Kashlev M, Babitzke P, Żywicki M. *Characterization of bacterial intrinsic transcription terminators identified with TERMITE – a novel method for comprehensive analysis of Term-seq data*. bioRxiv. 2024 May. doi: 10.1101/2024.05.16.594410

<https://www.biorxiv.org/content/10.1101/2024.05.16.594410v1>

Opracowanie powyższej metody pozwoliło na identyfikację i adnotację bakteryjnych terminatorów rho-niezależnych, prowadząc do powstania najbardziej kompletnego atlasu miejsc terminacji transkrypcji utworzonego na podstawie danych Term-seq. Autorzy wykazali, że charakterystyka terminatorów rho-niezależnych pod wieloma względami różni się znacząco na przestrzeni drzewa filogenetycznego królestwa bakterii. W ramach niniejszej publikacji przedstawiono atlas terminatorów dla *E. coli*, *B. subtilis*, *E. faecalis*, *L. monocytogenes*, *Z. mobilis*, *Synechocystis sp.*, *S. coelicolor*, *S. venezuelae*, *S. coelicolor*, *S. tsukubensis*, *S. griseus*, *S. lividans*, *S. clavuligerus* oraz *S. avermitilis*.

*TERMITE* stanowi obecnie najbardziej kompletne narzędzie bioinformatyczne do analizy danych Term-seq dostępne dla środowiska naukowego. Oprogramowanie pozwala na identyfikację stabilnych 3' końców RNA (zachowawczych pomiędzy replikatami biologicznymi) oraz ich szczegółową adnotację, ze szczególnym uwzględnieniem adnotacji rho-niezależnych miejsc terminacji transkrypcji.

Materiały uzupełniające dostępne są pod adresem:

*<https://www.biorxiv.org/content/10.1101/2024.05.16.594410v1.supplementary-material>*

## 10.2. *termiRNAtor* – metoda przewidywania miejsc terminacji transkrypcji na podstawie danych RNA-seq wykorzystująca algorytmy uczenia maszynowego

W chwili pisania niniejszego rozdziału, w bazie danych *NCBI Sequence Read Archive* znajdowało się jedynie 291 zestawów danych uzyskanych przy pomocy protokołu Term-seq (zapytanie `(Term-seq) AND "bacteria"[orgn:\_\_txid2]`) oraz aż 93370 pochodzących z eksperymentów typu RNA-seq (zapytanie `(RNA-seq) AND "bacteria"[orgn:\_\_txid2]`). Pokazuje to jednoznacznie z jak niewielką liczbą publicznie dostępnych bibliotek, ukierunkowanych na sekwencjonowanie 3' końców RNA, dysponujemy obecnie. Brak tego typu danych znacząco ogranicza możliwość prowadzenia badań, nie tylko nad terminacją transkrypcji u bakterii, ale także nad procesami takimi jak przetwarzanie czy degradacja RNA (szczególnie u organizmów niemodelowych).

W związku z powyższym, na potrzeby prowadzonych badań, opracowane zostało oprogramowanie o nazwie *termiRNAtor*. Program pozwala na przewidywanie rho-niezależnych miejsc terminacji transkrypcji na podstawie danych RNA-seq z wykorzystaniem algorytmów uczenia maszynowego. Kod źródłowy wraz z odpowiednią dokumentacją zostały opublikowane na platformie github pod adresem <https://github.com/zywicki-lab/termiRNAtor> (wersja 0.0.1). Utworzenie tego rodzaju metody możliwe było dzięki opracowaniu modeli uczenia maszynowego w oparciu o biblioteki RNA sekwencjonowane równolegle przy użyciu protokołów RNA-seq oraz Term-seq.

Na podstawie danych z sekwencjonowania wysokoprzepustowego 3' końców RNA metodą Term-seq możliwa stała się identyfikacja rzeczywistych miejsc rho-niezależnej terminacji dzięki zastosowaniu programu *TERMITE* (rozdział 9.1). Wykorzystanie zestawu prawdziwych miejsc terminacji pozwoliło z kolei na wytrenowanie modelu lasu losowego, który na podstawie sekwencji nukleotydowej oraz pokrycia genomu odczytami pochodzącymi z sekwencjonowania tej samej biblioteki metodą RNA-seq, pozwala na wysoce czułe i precyzyjne przewidywanie rho-niezależnych terminatorów transkrypcji.

Wytrenowanie klasyfikatora na danych pochodzących od różnorodnych pod względem taksonomicznym bakterii umożliwiło utworzenie modelu, który w sposób uniwersalny może być stosowany do predykcji miejsc terminacji transkrypcji u niezwykle szerokiego spektrum gatunków. Model wykorzystany w niniejszej pracy został stworzony w oparciu o dane

pochodzące od *E. faecalis*, *E. coli*, *Synechocystis sp.*, *Z. mobilis* oraz *B. subtilis* (dane uczące), natomiast jego wydajność została oszacowana na podstawie danych uzyskanych dla *L. monocytogenes* (dane testowe), by zapobiec przeuczeniu modelu pod kątem danych obecnych w zestawie uczącym. Zastosowane dane zostały szczegółowo opisane w publikacji poświęconej narzędziu *TERMITE*. Sumarycznie, na potrzeby treningu oraz ewaluacji klasyfikatora, wykorzystano 128 próbek z sekwencjonowania typu Term-seq, wraz z odpowiadającymi im danymi RNA-seq (załącznik 10).

## Cechy programu *termiRNAtor*

*termiRNAtor* pozwala na wykorzystanie już wytrenowanego klasyfikatora do przewidywania miejsc rho-niezależnej terminacji na podstawie danych RNA-seq (moduł *predict\_TTS.py*), jak również na opracowanie własnego modelu (moduł *train\_classifier.py*) w oparciu o biblioteki RNA sekwencjonowane równoległe przy pomocy protokołów RNA-seq oraz Term-seq. Program korzysta z informacji zakodowanych w plikach BAM, opisujących odczyty zmapowane do genomu referencyjnego. Dane Term-seq wykorzystywane są jedynie na potrzeby treningu oraz ewaluacji klasyfikatora (umożliwiając określenie rzeczywistych miejsc terminacji transkrypcji), podczas gdy predykcja wykonywana jest wyłącznie w oparciu o dane RNA-seq.

Odczyty znajdujące się w odpowiednich plikach BAM muszą być posortowane po koordynatach genomowych, np. przy użyciu programu *samtools sort*. Program *termiRNAtor* do prawidłowego działania wymaga także danych wejściowych w postaci sekwencji genomu referencyjnego (format FASTA) oraz odpowiadającego im pliku zawierającego adnotację genów (format GTF lub GFF). Trening klasyfikatora wymaga wykorzystania co najmniej dwóch replikatów biologicznych (Term-seq), dla każdej próbki wchodzącej w skład zbioru uczącego bądź testowego. Jest to bezpośrednio związane z wymaganiami narzuconymi przez program *TERMITE*, który identyfikuje stabilne 3' końce RNA na podstawie poziomu zachowawczości sygnału pomiędzy replikatami biologicznymi.

Wszystkie wymagane dane muszą zostać odpowiednio opisane w ramach dedykowanego pliku konfiguracyjnego, co zostało szczegółowo opisane w ramach dokumentacji omawianego narzędzia (patrz pliki *config.csv* oraz *config\_prediction.csv* udostępnione odpowiednio dla modułów *train\_classifier.py* oraz *predict\_TTS.py* w repozytorium). Obecnie *termiRNAtor* wspiera wykorzystanie jedynie niciowo-specyficznych danych RNA-seq sekwencjonowanych z jednego końca.

## Identyfikacja rho-niezależnych terminatorów transkrypcji na podstawie danych Term-seq

*termiRNAtor* po wczytaniu wymaganych plików konfiguracyjnych oraz wejściowych rozpoczyna przetwarzanie danych. W ramach pierwszego kroku program analizuje pokrycie genomu 5' końcami odczytów, uzyskanymi z eksperymentów typu Term-seq (reprezentującymi 3' końce RNA). Pokrycie genomu jest następnie normalizowane przy użyciu metody CPM (ang. *counts per million mapped reads*) i zapisywane w formie plików *bigWig*. Oba etapy wykonywane są przy użyciu programu *bamCoverage* wchodzącego w skład pakietu *deeptools*. W ramach kolejnego kroku dla każdej analizowanej próbki program *TERMITe* (moduły *find\_stable\_rna\_ends* oraz *annotate*) wykorzystywany jest do identyfikacji stabilnych 3' końców RNA będących wynikiem rho-niezależnej terminacji transkrypcji. *termiRNAtor* do zestawów uczącego oraz testowego wybiera jedynie terminatory, które zostały potwierdzone równocześnie przy pomocy modułów *RNAfold* oraz *TransTermHP* zaimplementowanych w obrębie programu *TERMITe*.

### Tworzenie zbiorów uczącego i testowego

W ramach kolejnego etapu, dla każdego genu zaadnotowanego w obrębie pliku GTF/GFF, wyznaczany jest region rozpoczynający się domyślnie 20 nt powyżej, a kończący 150 nt poniżej 3' końca genu. Rozmiarem regionu można sterować odpowiednio za pomocą argumentów *-tts-upstream* oraz *-tts-downstream*. Region poddawany analizie może się nakładać z genem występującym poniżej w obrębie tej samej nici jedynie na przestrzeni maksymalnie 10 nt (*--tts-invasion*). Gdy ww. warunek nie jest spełniony, region jest odpowiednio skracany. Następnie, regiony o długości mniejszej niż 50 nukleotydów są wykluczane z analizy. W kolejnym etapie, dla każdego wyznaczonego regionu (dla każdego genu) przypisywane jest najliczniej wykorzystywane w danej próbie miejsce terminacji transkrypcji, określone na podstawie danych Term-seq w obrębie opisywanego regionu (jeżeli występuje). Regiony (geny), posiadające przypisane w ten sposób miejsce terminacji transkrypcji, wchodzi w skład zbiorów uczącego oraz testowego. Każda pozycja genomowa w obrębie wyznaczonego regionu traktowana jest jako przykład (uczący lub testowy), pozytywny w przypadku pozycji reprezentującej miejsce terminacji transkrypcji lub negatywny w przeciwnym przypadku.



Każdy przykład uczący oraz testowy opisywany jest przy pomocy zbioru cech, na podstawie których *termiRNator* przeprowadza procesy treningu klasyfikatora oraz predykcji. Wspomniane cechy reprezentują obecność poszczególnych nukleotydów na konkretnych pozycjach w regionie okalającym przykład uczący lub testowy oraz pokrycie każdej z tych pozycji odczytami z eksperymentu RNA-seq w obrębie danej próbki. Region okalający każdy z przykładów obejmuje domyślnie 150 nukleotydów powyżej (opcja *--upstream*) oraz 20 nt poniżej (opcja *--downstream*).

W ramach kolejnego etapu odfiltrowywane są geny charakteryzujące się niewielkim średnim pokryciem odczytami z sekwencjonowania typu RNA-seq, co jest bezpośrednio powiązane z niskim poziomem ekspresji danego genu. Ze zbioru uczącego oraz predykcyjnego wykluczane są geny, dla których średnie pokrycie pierwszych 20 nukleotydów (na podstawie, którego jesteśmy w stanie oszacować poziom ekspresji danego genu) w ramach analizowanego regionu nie przekracza mediany wartości obliczonych dla wszystkich genów. Następnie przeprowadzane jest balansowanie liczebności zbiorów przykładów pozytywnych i negatywnych przy użyciu próbkowania ze zwracaniem, tak by liczba przykładów negatywnych i pozytywnych była równa wartości sprecyzowanej w ramach parametru *--training\_set\_size* (domyślnie 2 000 000).

### **Trening klasyfikatora**

Klasyfikator trenowany jest w oparciu o dane pochodzące z gatunków określonych w ramach opcji *--train-on*, ewaluacja natomiast na tych zdefiniowanych flagą *--test-on*. *termiRNator* używa do predykcji modelu lasów losowych, zaimplementowanego w ramach metody *sklearn.ensemble.RandomForestClassifier* pochodzącej z pakietu *scikit-learn* (Pedregosa et al., 2011). Udostępniony i wykorzystany klasyfikator został opracowany w oparciu o las składający się z 1000 drzew decyzyjnych.

### **Przewidywanie miejsc terminacji transkrypcji na podstawie danych RNA-seq**

Predykcja przy użyciu opracowanego modelu dokonywana jest dla każdej pozycji w obrębie regionu określonego parametrami *-tts-upstream* oraz *-tts-downstream*, wyłącznie dla genów posiadających odpowiedni poziom ekspresji oszacowany na podstawie danych RNA-seq (jak wyżej). Dla każdej pozycji w ww. regionie klasyfikator przypisuje prawdopodobieństwo, że dana pozycja stanowi miejsce terminacji transkrypcji. Jako miejsce

terminacji wyznaczana jest pozycja o maksymalnym prawdopodobieństwie dla danego regionu, o ile jego wartość przekracza wartość *--min\_probability* (domyślnie 0.15, wyznaczona na podstawie Wykresu 1).

Tabela 1. Macierz pomyłek oraz podstawowe metryki pozwalające na ewaluację modelu

klasa rzeczywista	klasa przewidziana		
		pozytywna	negatywna
	pozytywna	PP: 2053	FN: 1295
negatywna	PN: 3319	FP: 757	

*PP* – obserwacje prawdziwie pozytywne, *PN* – obserwacje prawdziwie negatywne,  
*FN* – obserwacje fałszywie negatywne, *FP* – obserwacje fałszywie pozytywne

	Precyzja	Czułość	Liczba przykładów
Fałsz	0.72	0.81	4076
Prawda	0.73	0.61	3348

*Prawda* - przykłady posiadające terminator zidentyfikowany na podstawie danych Term-seq

*Fałsz* - przykłady nie posiadające terminatora zidentyfikowanego na podstawie danych Term-seq

*Precyzja* – odsetek obserwacji określonych jako pozytywne, będących w istocie pozytywnymi:

$$\text{precyzja} = TP / (TP + FP)$$

*Czułość* - odsetek obserwacji prawdziwie pozytywnych:

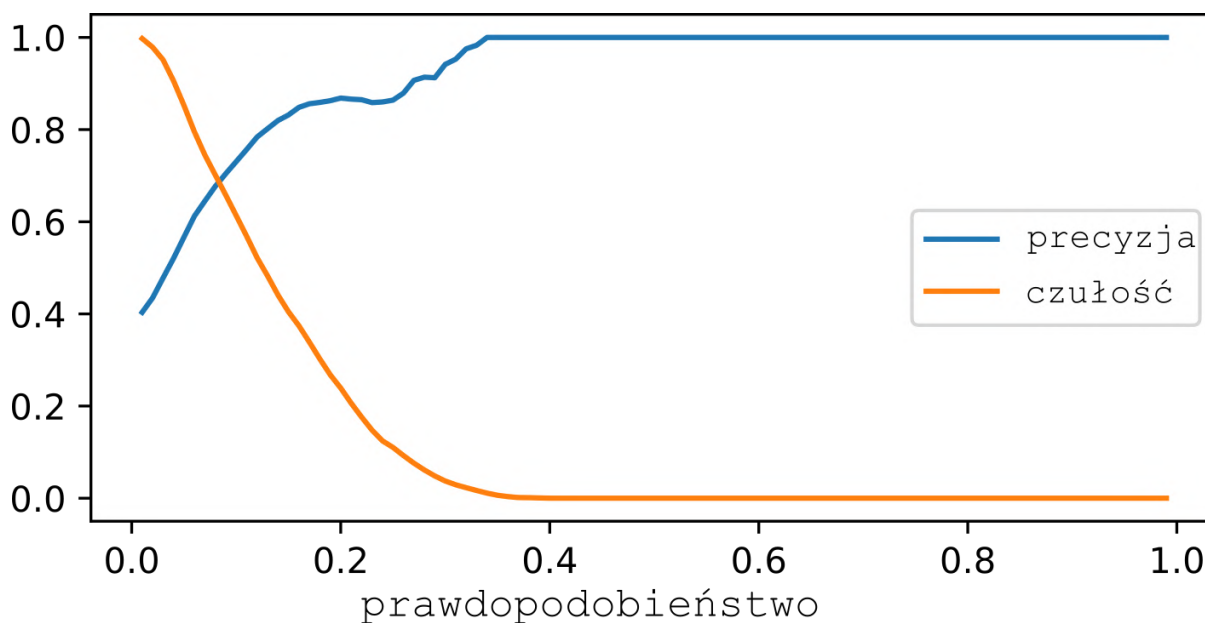
$$\text{czułość} = TP / (TP + FN)$$

## Ewaluacja klasyfikatora

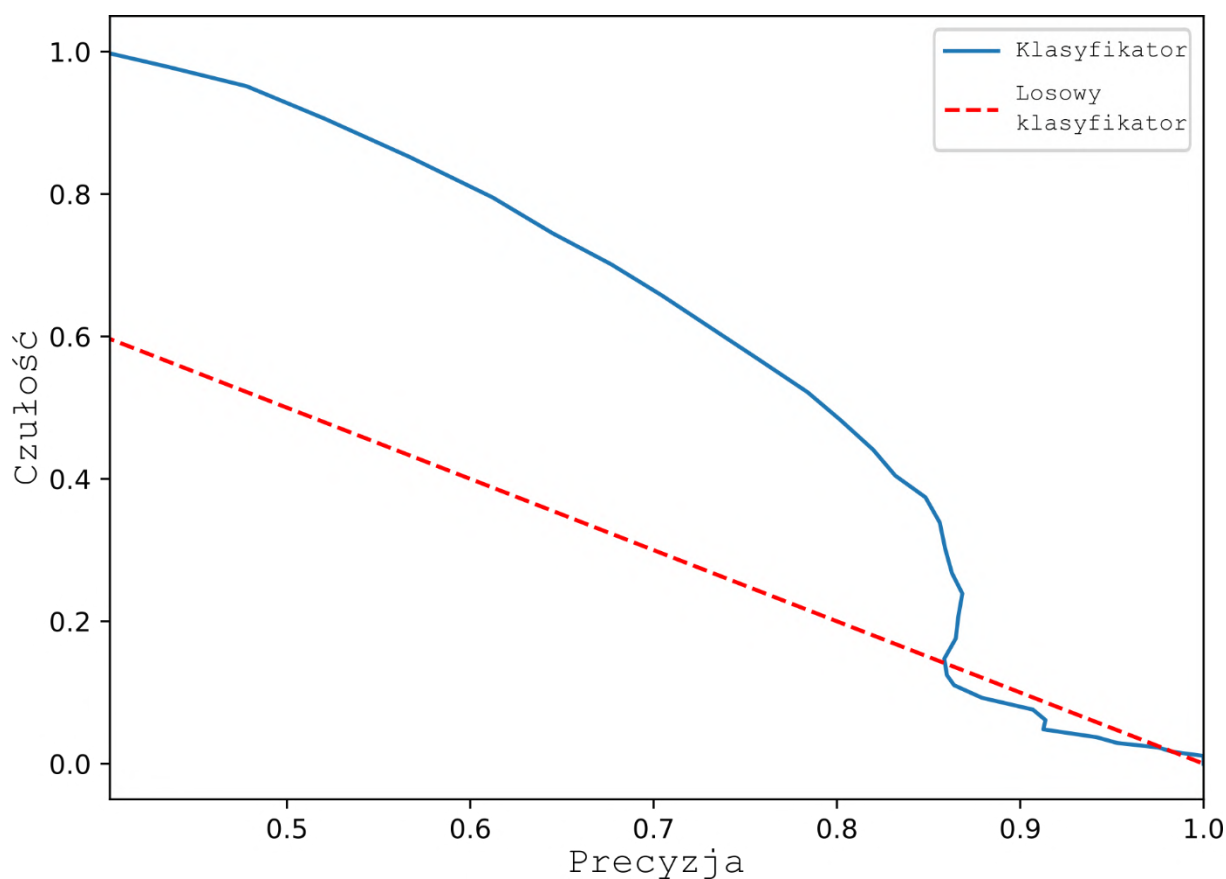
Ewaluacja utworzonego modelu została dokonana w oparciu o dane pochodzące od *L. monocytogenes*. Istotne jest, że dane dla tego gatunku nie zostały wykorzystane na etapie uczenia klasyfikatora. Podstawowe metryki pozwalające ocenić poprawność predykcji znajdują się w, tzw. macierzy pomyłek ukazującej liczbę obserwacji prawdziwie i fałszywie pozytywnych oraz prawdziwie i fałszywie negatywnych (Tabela 1). Liczba wartości prawdziwie pozytywnych jest tożsama z liczbą genów posiadających wyznaczone miejsce terminacji transkrypcji na podstawie danych Term-seq, dla których przewidziane miejsce terminacji transkrypcji znajduje się w odległości nie większej niż 5 nt od niego. Liczba wartości prawdziwie negatywnych jest rozumiana natomiast jako liczba genów, dla których nie zostało zidentyfikowane miejsce terminacji na podstawie danych Term-seq oraz nie zostało ono przewidziane na podstawie danych RNA-seq. Obserwacje fałszywie negatywne zostały określone jako liczba genów bez przewidzianego miejsca terminacji na podstawie danych

RNA-seq, podczas gdy zostało ono wyznaczone z wykorzystaniem danych Term-seq. Ostatnią wykorzystaną miarą jest liczba obserwacji fałszywie pozytywnych oznaczająca liczbę genów, dla których przewidywane miejsce terminacji jest oddalone od rzeczywistego o więcej niż 5 nt. Ponadto, w *Tabeli 1* można również znaleźć wartości opisujące precyzję (odsetek obserwacji określonych jako pozytywne, będących w istocie pozytywnymi) oraz czułość (odsetek obserwacji prawdziwie pozytywnych) dla opracowanego modelu. Analiza wyżej opisanych metryk jednoznacznie pokazuje, że opracowany model pozwala na wysoce czułe i precyzyjne przewidywanie miejsc rho-niezależnej terminacji transkrypcji u bakterii.

Jako miejsce terminacji oznaczana jest pozycja o maksymalnym prawdopodobieństwie obliczonym dla regionu reprezentującego określony gen, o ile wartość prawdopodobieństwa przekracza wartość progową opisaną parametrem `--min_probability` (domyślnie 0.15). Wartość progowa prawdopodobieństwa została dobrana tak by maksymalizować jednocześnie zarówno czułość jak i specyficzność modelu (Wykres 1). Zależność czułości od precyzji w porównaniu do klasyfikatora losowego również została przedstawiona na potrzeby ewaluacji modelu (Wykres 2).

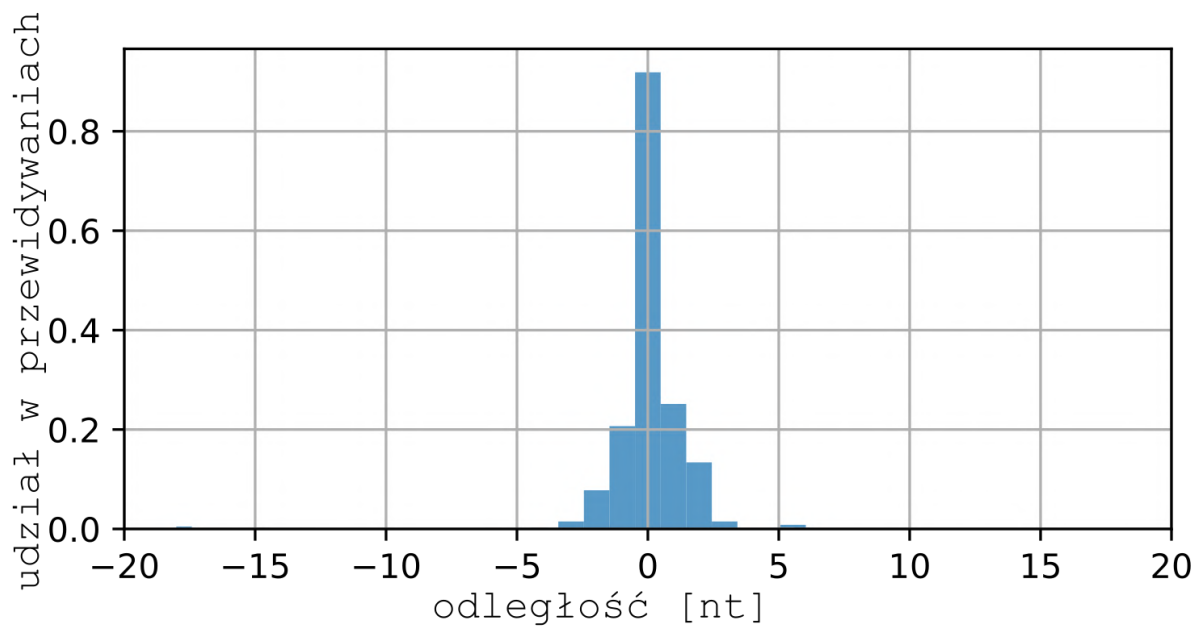


Wykres 1. Wpływ obranej progowej wartości prawdopodobieństwa (`--min-probability`) na czułość i specyficzność klasyfikacji.



Wykres 2. Relacja pomiędzy precyzją a czułością opracowanego klasyfikatora w porównaniu do modelu losowego.

Jakość modelu obrazuje także rozkład odległości pomiędzy rzeczywistymi a przewidzianymi miejscami rho-niezależnej terminacji transkrypcji u *L. monocytogenes* (Wykres 3). Analiza opisanego rozkładu pokazuje, że model w znacznej większości przypadków pozwala na wyznaczenie miejsc terminacji z dokładnością do maksymalnie 3 nukleotydów względem pozycji oznaczonych na podstawie danych Term-seq. Powyższe obserwacje podkreślają jak istotne mogą być dane pochodzące z eksperymentów RNA-seq w kontekście precyzyjnego oznaczania miejsc rho-niezależnej terminacji transkrypcji.



Wykres 3. Rozkład odległości pomiędzy rzeczywistymi a przewidzianymi miejscami rho-niezależnej terminacji transkrypcji u *L. monocytogenes*.

### **10.3. *AGouTI* – metoda adnotacji interwałów genomowych oraz transkryptomowych z wykorzystaniem informacji zawartych w plikach GTF lub GFF**

Szczegółowy opis opracowanej metody został przedstawiony w ramach publikacji, stanowiącej integralną część niniejszej pracy doktorskiej

**Kosiński JG, Żywicki M. *AGouTI* - Flexible Annotation of Genomic and Transcriptomic Intervals.** PLoS Comput Biol. 2023 Oct 18;19(10):e1011527. doi: 10.1371/journal.pcbi.1011527. PMID: 37851672; PMCID: PMC10621952.

<https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1011527>

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10621952/>

*AGouTI* pozwala na adnotację interwałów genomowych bądź transkryptomowych przy użyciu informacji zawartych w plikach GTF lub GFF. W ramach przedstawionej pracy metoda ta została wykorzystana do uzyskania szczegółowych informacji o zidentyfikowanych rho-niezależnych terminatorach transkrypcji u bakterii.

Materiały uzupełniające dostępne są pod adresem:

<https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1011527#sec01>

## 10.4. Atlas rho-niezależnych terminatorów transkrypcji u metycylinoopornych szczepów gronkowca złocistego

W niniejszej rozprawie przeprowadziłem kompleksową identyfikację rho-niezależnych terminatorów transkrypcji u metycylinoopornych szczepów gronkowca złocistego. Do tego celu wykorzystałem dane z wysokoprzepustowego sekwencjonowania transkryptomu metodą RNA-seq oraz sekwencjonowania 3' końców RNA metodą Term-seq. Przeanalizowałem łącznie 140 publicznie dostępnych próbek (138 pochodzących z eksperymentów RNA-seq oraz 2 z Term-seq), w tym próbki pochodzące z bakterii hodowanych w różnych warunkach eksperymentalnych oraz poddane traktowaniu subletalnymi dawkami antybiotyków, takimi jak wankomycyna, linezolid, nafcylina czy oksacylina (rozdział 8.1).

W wyniku badań zidentyfikowałem 1128 rho-niezależnych terminatorów transkrypcji (załącznik 11), które stanowią, obecnie najbardziej kompletny atlas terminatorów opracowany na podstawie danych pochodzących z wysokoprzepustowego sekwencjonowania dla metycylinoopornych szczepów gronkowca złocistego. Do identyfikacji terminatorów zastosowano specjalnie opracowane do tego celu narzędzia *TERMITE* oraz *termiRNAtor*, które mogą z powodzeniem być stosowane w badaniach terminacji transkrypcji u bakterii. W skład przygotowanego atlasu wchodzi 1083 terminatory zidentyfikowane na podstawie danych Term-seq z wykorzystaniem programu *TERMITE* oraz 412 terminatorów zidentyfikowanych na podstawie danych RNA-seq przy pomocy programu *termiRNAtor* (367 z nich zostało zaadnotowanych równocześnie przy pomocy obu metod).

Kluczowym aspektem opracowanej listy terminatorów, jest fakt, że zawiera ona terminatory zidentyfikowane w ramach szeregu warunków eksperymentalnych oraz stadiów wzrostu, w tym w warunkach stresowych wywołanych stosowaniem rozmaitych antybiotyków, podawanych w różnych dawkach oraz w różnym czasie (załącznik 1).

Dodatkowo, przeprowadzona została kompleksowa adnotacja zidentyfikowanych terminatorów, która pozwoliła na uzyskanie szeregu informacji dotyczących ich cech strukturalnych i funkcjonalnych. W tabeli wynikowej możemy znaleźć informacje o koordynatach genomowych regionu, a także wyznaczonym referencyjnym punkcie terminacji (POT, ang. *point of termination*), który odpowiada pozycji o najwyższym zagęszczeniu 3' końców RNA obserwowanych (lub przewidywanych) dla danego terminatora. Ponadto, w tabeli wynikowej (załącznik 11) znajduje się kolumna zawierająca sekwencję nukleotydową

okalającą referencyjny punkt terminacji, wraz z zaznaczonym tymże punktem. Dla terminatorów zidentyfikowanych na podstawie danych TERM-seq zawarto również informację o średniej wysokości oraz poziomie zachowawczości sygnału reprezentującego pokrycie danego terminatora 3' końcami RNA w ramach analizowanych replikatów biologicznych. W udostępnionych wynikach znajdują się także informacje świadczące o potencjalnym występowaniu zidentyfikowanego terminatora w obrębie sekwencji kodującej genu. Dodatkowo, w ramach przeprowadzonej adnotacji udostępnione zostały szczegółowe dane o najbliższych genach zlokalizowanych odpowiednio powyżej oraz poniżej, wraz z uwzględnieniem odpowiednich odległości.

Szczegółowa informacja o lokalizacji, strukturze oraz energii zidentyfikowanej struktury typu spinka do włosów, informacja o sekwencji nukleotydowej zidentyfikowanych traktów poly-A oraz poly-U, a także odległości pomiędzy zidentyfikowaną strukturą typu spinka do włosów a referencyjnym punktem terminacji, również zostały uwzględnione w ramach przedstawionej adnotacji dla każdego terminatora. Powyższe dane zostały ujęte w tabeli dla spinek terminacyjnych zidentyfikowanych przy pomocy wszystkich metod zaimplementowanych w programie *TERMITe*.

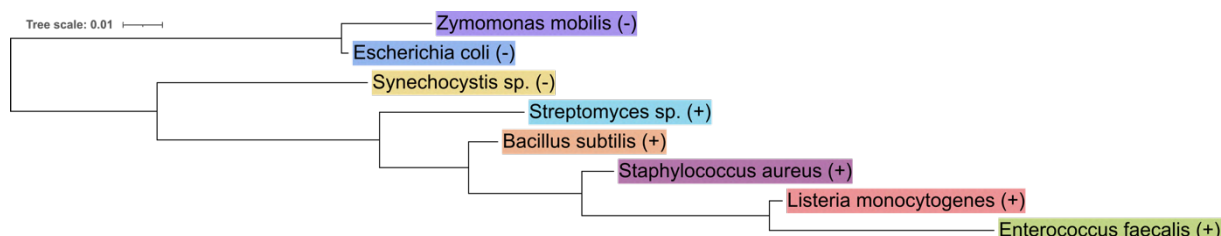
Wszystkie wspomniane informacje dostarczają cennych informacji na temat struktury i funkcji zidentyfikowanych terminatorów transkrypcji. Uzyskane w ten sposób dane pozwoliły na kompleksowe porównanie charakterystyki rho-niezależnych terminatorów transkrypcji występujących u metycylinoopornych szczepów gronkowca złocistego oraz innych gatunków bakterii.



## 10.5. Charakterystyka rho-niezależnych terminatorów transkrypcji u gronkowca złocistego

Charakterystyka terminatorów dla większości omawianych w rozprawie gatunków bakterii została przedstawiona w publikacji poświęconej narzędziu *TERMITE*, która stanowi integralną część niniejszej rozprawy (rozdział 9.1). W ramach wspomnianej publikacji uwzględniona została kompleksowa charakterystyka rho-niezależnych terminatorów transkrypcji zidentyfikowanych u *E. coli*, *B. subtilis*, *E. faecalis*, *L. monocytogenes*, *Z. mobilis*, *Synechocystis sp.*, *S. coelicolor*, *S. venezuelae*, *S. tsukubaensis*, *S. griseus*, *S. lividans*, *S. clavuligerus* oraz *S. avermitilis*. Ponadto, szczegółowo omówione i przeanalizowane zostały różnice występujące pomiędzy poszczególnymi taksonami. W poniższym rozdziale wyniki uzyskane dla wyżej wymienionych gatunków, zostaną zestawione z charakterystyką terminatorów u gronkowca złocistego.

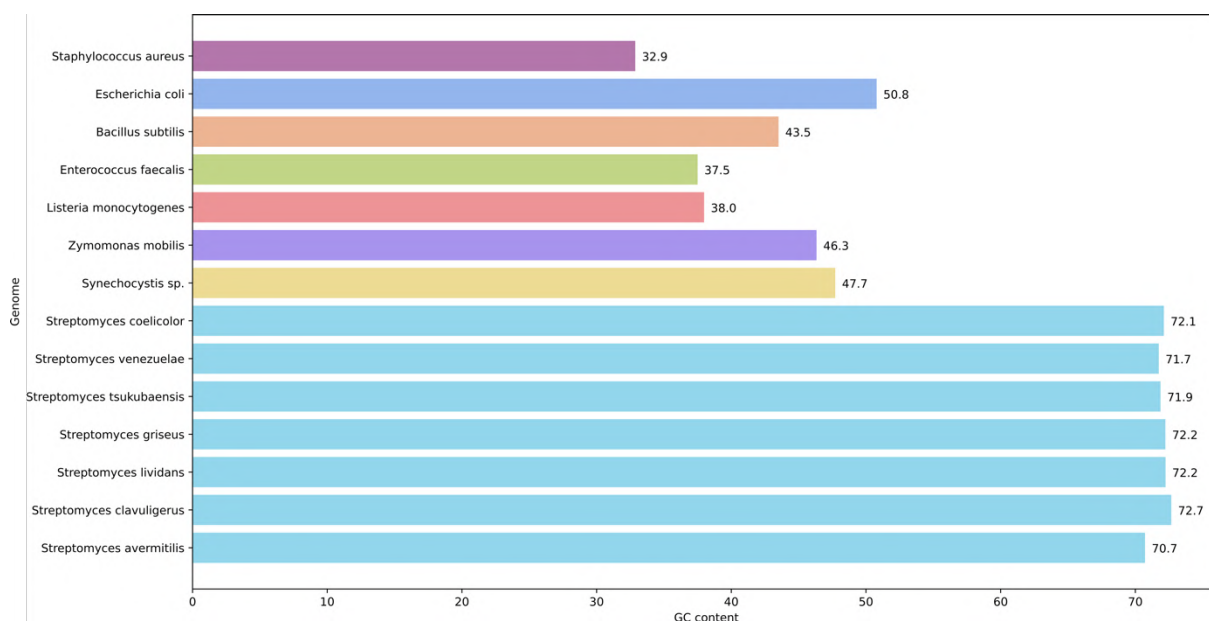
Zależności ewolucyjne pomiędzy gronkowcem a pozostałymi gatunkami zostały przedstawione w formie drzewa filogenetycznego (Wykres 4).



Wykres 4. Drzewo filogenetyczne reprezentujące analizowane gatunki bakterii. Wizualizacja została opracowana według metodologii opisanej w obrębie publikacji narzędzia *TERMITE* (rozdział 9.1). (+) oraz (-) oznaczają odpowiednio bakterie Gram-dodatnie oraz Gram-ujemne.

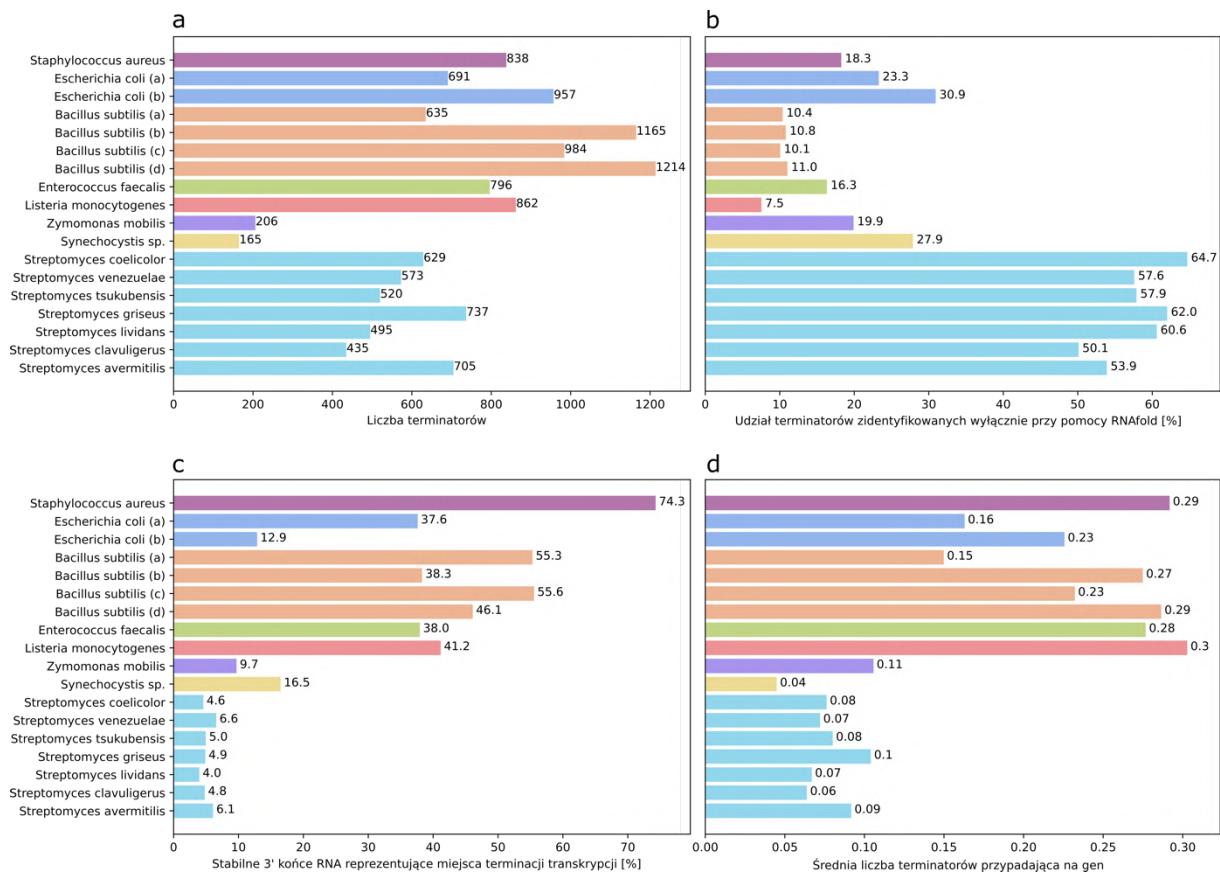
Do cech charakterystycznych rho-niezależnych terminatorów transkrypcji należy obecność sekwencji bogatej w U, poprzedzonej bogatą w pary GC strukturą typu spinka do włosów (Chełkowska-Pauszek et al., 2021). Opcjonalnie spinka terminacyjna może być również poprzedzona traktem bogatym w A, co jest szczególnie charakterystyczne dla dwukierunkowych terminatorów transkrypcji (Chełkowska-Pauszek et al., 2021). Ze względu na wysoką zależność terminatorów rho-niezależnych od sekwencji nukleotydowej, niezwykle ważne jest odniesienie obserwowanych wyników do średniej zawartości par GC w genomach analizowanych bakterii, którą przedstawiono w rozdziale 9.1, a z uwzględnieniem *S. aureus*

w obrębie Wykresu 5. *S. aureus* charakteryzuje się najniższą zawartością par GC w genomie spośród analizowanych gatunków bakterii (32.9%).



Wykres 5. Średnia zawartość par GC w analizowanych genomach. Przedstawione wartości zostały obliczone na podstawie plików FASTA zawierających referencyjne sekwencje genomowe wykorzystane w prezentowanych badaniach.

By uniknąć różnic wynikających z zastosowanych protokołów analitycznych, atlas terminatorów uzyskanych dla gronkowca złocistego został przefiltrowany przy użyciu kryteriów zastosowanych w ramach publikacji *TERMITE* (rozdział 9.1). W związku z powyższym, w porównaniu uwzględnione zostały wyłącznie terminatory, dla których referencyjny punkt terminacji znajduje się co najwyżej 200nt od najbliższej sekwencji kodującej zlokalizowanej *powyżej*. Dodatkowo, w przypadku terminatorów potwierdzonych wyłącznie przy pomocy protokołu opartego o program *RNAfold* (patrz algorytm narzędzia *TERMITE*), usunięte zostały te posiadające minimalną energię swobodną struktury wyższą niż -3 kcal/mol. Zastosowanie powyższych kryteriów pozwoliło na wyodrębnienie zestawu 838 rho-niezależnych terminatorów u gronkowca złocistego (Wykres 6a), co odpowiada aż 74.3% spośród zidentyfikowanych na podstawie danych Term-seq stabilnych 3' końców RNA (Wykres 6c).



Wykres 6. **a)** Liczba terminatorów rho-niezależnych zidentyfikowanych na podstawie danych z wysokoprzepustowego sekwencjonowania u *S. aureus* oraz gatunków bakterii analizowanych w ramach rozdziału 9.1. **b)** Liczba terminatorów rho-niezależnych zidentyfikowanych wyłącznie przy pomocy protokołu analitycznego opartego o RNAfold (zaimplementowanego w ramach narzędzia TERMITE). Metoda oparta o RNAfold zakłada istnienie jedynie spinki terminacyjnej o odpowiedniej energii swobodnej struktury. Metoda ta nie wymaga występowania traktów poli-U ani poli-A, pozwalając na identyfikacje terminatorów niekanonicznych. **c)** Procentowy udział stabilnych 3' końców RNA będących wynikiem rho-niezależnej terminacji transkrypcji. **d)** Średnia liczba zidentyfikowanych terminatorów rho-niezależnych przypadająca na gen.

Wyniki zaprezentowane w rozdziale 9.1, po zestawieniu ich z rezultatami uzyskanymi dla *S. aureus*, świadczą o wysokim udziale sygnałów odpowiadających rho-niezależnej terminacji transkrypcji, w stosunku do ogólnej liczby zidentyfikowanych 3' końców RNA, u tego gatunku bakterii (Wykres 6c). Powyższa obserwacja może sugerować, że jest to dominujący sposób terminacji transkrypcji u *S. aureus* (w przeciwieństwie np. do rodzaju *Streptomyces*). Liczba terminatorów, które zostały zidentyfikowane u gronkowca złocistego, jest wysoce zbliżona do liczby obserwowanych u innych, stosunkowo blisko spokrewnionych,

gatunków bakterii Gram-dodatnich, takich jak *B. subtilis*, *L. monocytogenes* czy *E. faecalis* (Wykres 6a). Analizowane gatunki różnią się jednak istotnie pod względem rozmiaru genomu oraz liczby genów. W związku z powyższym, by dokładniej zrozumieć przedstawiane liczebności terminatorów rho-niezależnych, należy liczbę terminatorów odnieść do liczby genów występujących u danego gatunku (Wykres 6d). U *S. aureus* na jeden gen przypadło średnio 0.29 terminatora, co jest wartością podobną do uzyskanych dla *B. subtilis* (0.15-0.29), *E. faecalis* (0.28), *L. monocytogenes* (0.30) czy w mniejszym stopniu do *E. coli* (0.16-0.23). Dla pozostałych gatunków obserwowane wartości są zauważalnie niższe, tj. 0.07-0.10 dla rodzaju *Streptomyces*, 0.04 dla *Synechocystis* czy 0.11 dla *Z. mobilis* (rozdział 9.1).

Warto zauważyć, że znaczący wpływ na liczbę obserwowanych terminatorów może mieć szereg czynników natury technicznej, takich jak sposób przygotowania bibliotek, czy głębokość sekwencjonowania. Obecność powyższych artefaktów utrudnia wyciąganie bezpośrednich wniosków dotyczących bezwzględnych liczb terminatorów rho-niezależnych występujących w genomach bakteryjnych. Mimo wszystko, wyżej przedstawione wartości (Wykres 6) pozwalają określić rzetelność przeprowadzonej w niniejszym rozdziale kompleksowej charakterystyki terminatorów bakteryjnych.

Identyfikacja miejsc terminacji transkrypcji na podstawie danych Term-seq została przeprowadzona przy użyciu programu *TERMITE*. Dla każdego zidentyfikowanego stabilnego 3' końca RNA, *TERMITE* poszukuje w regionie zlokalizowanym bezpośrednio powyżej cech rho-niezależnego terminatora, takich jak obecność struktury spinki terminacyjnej czy traktów poli-A oraz poli-U. W tym celu wykorzystywane są programy *RNAfold* oraz *TransTermHP*. Metoda oparta o działanie narzędzia *TranTermHP* zakłada występowanie traktu poli-U oraz spinki terminacyjnej w obrębie każdego terminatora rho-niezależnego, natomiast metoda oparta o *RNAfold* zakłada istnienie jedynie spinki terminacyjnej o odpowiedniej energii swobodnej struktury (patrz rozdział 9.1). Liczba terminatorów zidentyfikowanych wyłącznie na podstawie protokołu analitycznego opartego o *RNAfold* informuje zatem o udziale niekanonicznych terminatorów wśród wszystkich zidentyfikowanych dla danego gatunku (Wykres 6b). Udział terminatorów niekanonicznych u *S. aureus* wynosi 18.3%, co świadczy o stosunkowo wysokim udziale terminatorów uznawanych powszechnie za kanoniczne.

Wyniki przeprowadzonych badań pokazują jednoznacznie duże zróżnicowanie w budowie i charakterystyce spinek terminacyjnych będących kluczowym elementem analizowanych rho-niezależnych terminatorów transkrypcji (Wykres 7, załączniki 2-5, rozdział 9.1). Istotnie statystycznie różnice wykazane zostały zarówno w ramach porównań minimalnej energii swobodnej struktury wyznaczonej za pomocą programu *RNAfold* (Kruskal-Wallis:

2591.3,  $p = 0.00$ ; Wykres 7a) oraz wartości *hairpin score* stanowiącej jej heurystyczny odpowiednik obliczany przez program *TransTermHP* (Kruskal-Wallis: 3612.8,  $p = 0.00$ ; Wykres 7b). Ponadto, istotne statystycznie różnice znalezione zostały w długości spinek terminacyjnych występujących u poszczególnych taksonów (Kruskal-Wallis: 3517.7,  $p = 0.00$ ; Wykres 7c) oraz długości pętli wchodzących w ich skład (Kruskal-Wallis: 60.9,  $p = 0.00$ ; Wykres 7d). Wyniki dla wszystkich analizowanych gatunków z wyłączeniem gronkowca złocistego, zostały szczegółowo omówione w publikacji poświęconej narzędziu *TERMITE* (rozdział 9.1).

W celu weryfikacji hipotezy zerowej o braku przesunięcia porównywanych rozkładów pomiędzy każdą możliwą parą analizowanych gatunków zastosowano test U Manna-Whitneya wg. protokołu opisanego w ramach publikacji narzędzia *TERMITE*. Szczegółowe wyniki dla wszystkich przedstawionych w niniejszym rozdziale parametrów są dostępne w tabeli znajdującej się w załącznikach (załącznik 9).

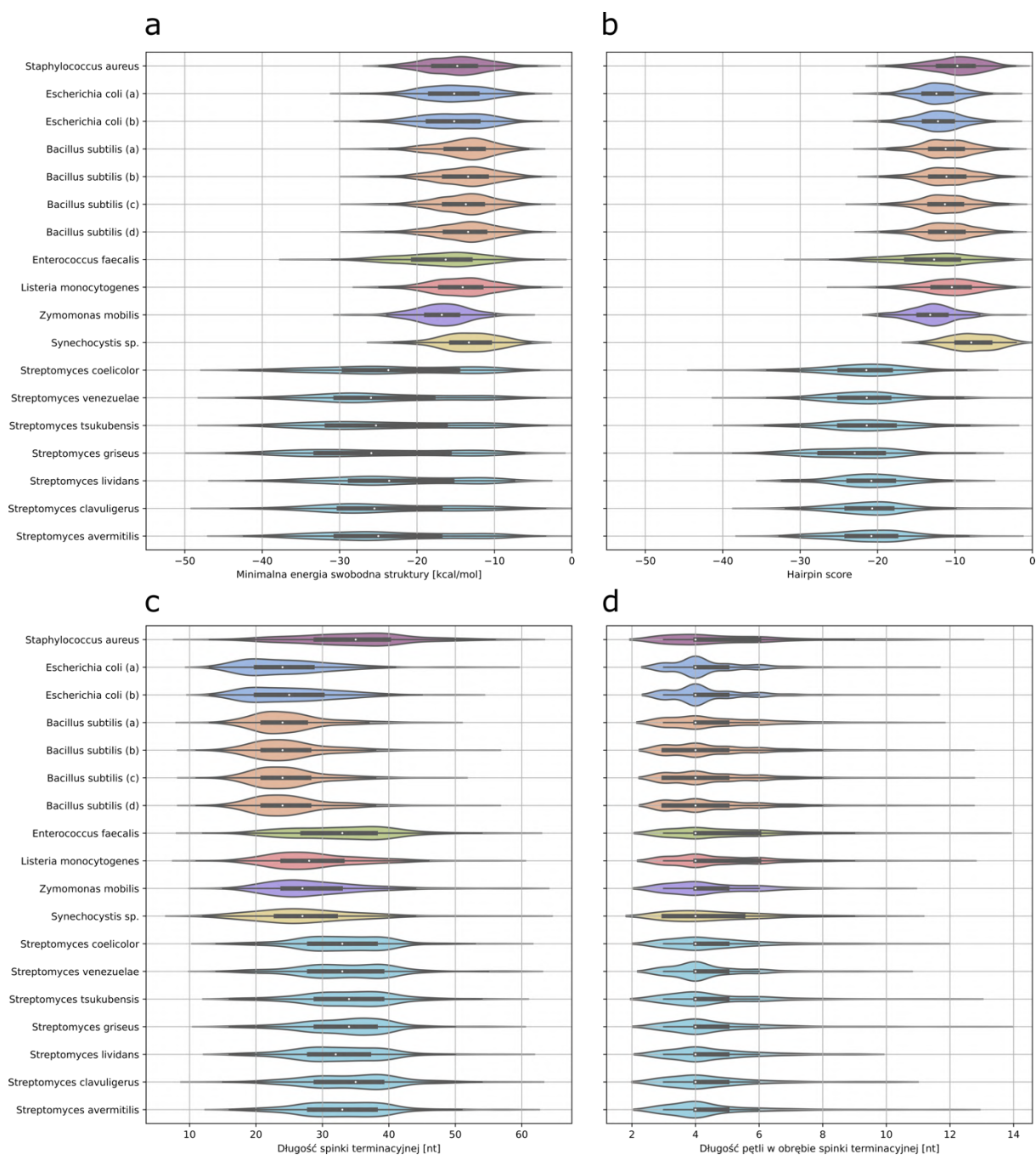
Charakterystyka terminatorów zidentyfikowanych u gronkowca złocistego jest niezwykle zbliżona do obserwowanej u stosunkowo blisko spokrewnionych ewolucyjnie gatunków, takich jak *L. monocytogenes* oraz *E. faecalis*. Na podstawie prezentowanych wyników zaobserwowano tendencję terminatorów u ww. gatunków bakterii do posiadania znacząco dłuższych spinek terminacyjnych oraz występujących w nich pętli, niż u gatunków powszechnie uznawanych za modelowe, tj. *E. coli* i *B. subtilis* (Wykres 7c, 7d, załączniki 4-5).

Pod względem energetycznym spinki terminacyjne, zidentyfikowane dla wszystkich analizowanych gatunków bakterii w ramach rozdziału 9.1 oraz *S. aureus*, posiadają zbliżone rozkłady z wyjątkiem rodzaju *Streptomyces*, który charakteryzuje się tendencją do posiadania znacznie niższych wartości minimalnej energii swobodnej struktury oraz jej odpowiednika w postaci heurystycznie obliczanej przez program *TransTermHP* wartości o nazwie *hairpin score* (Wykres 7a, 7b, załączniki 2-3). U gronkowca spinki terminacyjne charakteryzują się zazwyczaj minimalną energią swobodną struktury pomiędzy -17.9 kcal/mol oraz -12.4 kcal/mol (rozstęp pomiędzy pierwszym a trzecim kwartylem). Minimalna obserwowana wartość to -24.9, natomiast maksymalna wynosi -3.6 kcal/mol (załącznik 2). Wartości parametru *hairpin score* oscylują zazwyczaj w przedziale od -12.2 do -7.6, z wartościami minimalną i maksymalną wynoszącymi odpowiednio -19.7 i -2.2 (załącznik 3).

Długość spinki terminacyjnej u gronkowca złocistego różni się od wartości, przedstawionych w rozdziale 9.1 dla innych gatunków bakterii, wynosząc zazwyczaj od 29 do 40 nukleotydów (rozstęp pomiędzy pierwszym a trzecim kwartylem). U *E. coli* i *B. subtilis* długość ta to odpowiednio od 20 do 30 i od 21 do 28 nukleotydów. U bliżej spokrewnionych

gatunków, tj. *L. monocytogenes* oraz *E. faecalis* wartości te wynoszą kolejno od 24 do 33 i od 27 do 38 nukleotydów (załącznik 4). Pod względem długości spinki terminacyjnej, podobne do gronkowca złocistego wartości obserwujemy także dla rodzaju *Streptomyces*, gdzie odpowiednio wynoszą one od 28 do 39 nukleotydów, co stanowi niezwykle interesującą obserwację przez wzgląd na skrajnie różną zawartość par GC w omawianych genomach (Wykres 5).

Najmniejsze różnice pomiędzy gatunkami zaobserwowane zostały dla długości pętli wchodzącej w skład spinki terminacyjnej (Wykres 7d, załącznik 5). Podobnie jak w przypadku pozostałych porównań, tak również i tu rozkład długości obserwowany u gronkowca jest najbardziej zbliżony do tych charakteryzujących *L. monocytogenes* oraz *E. faecalis*.

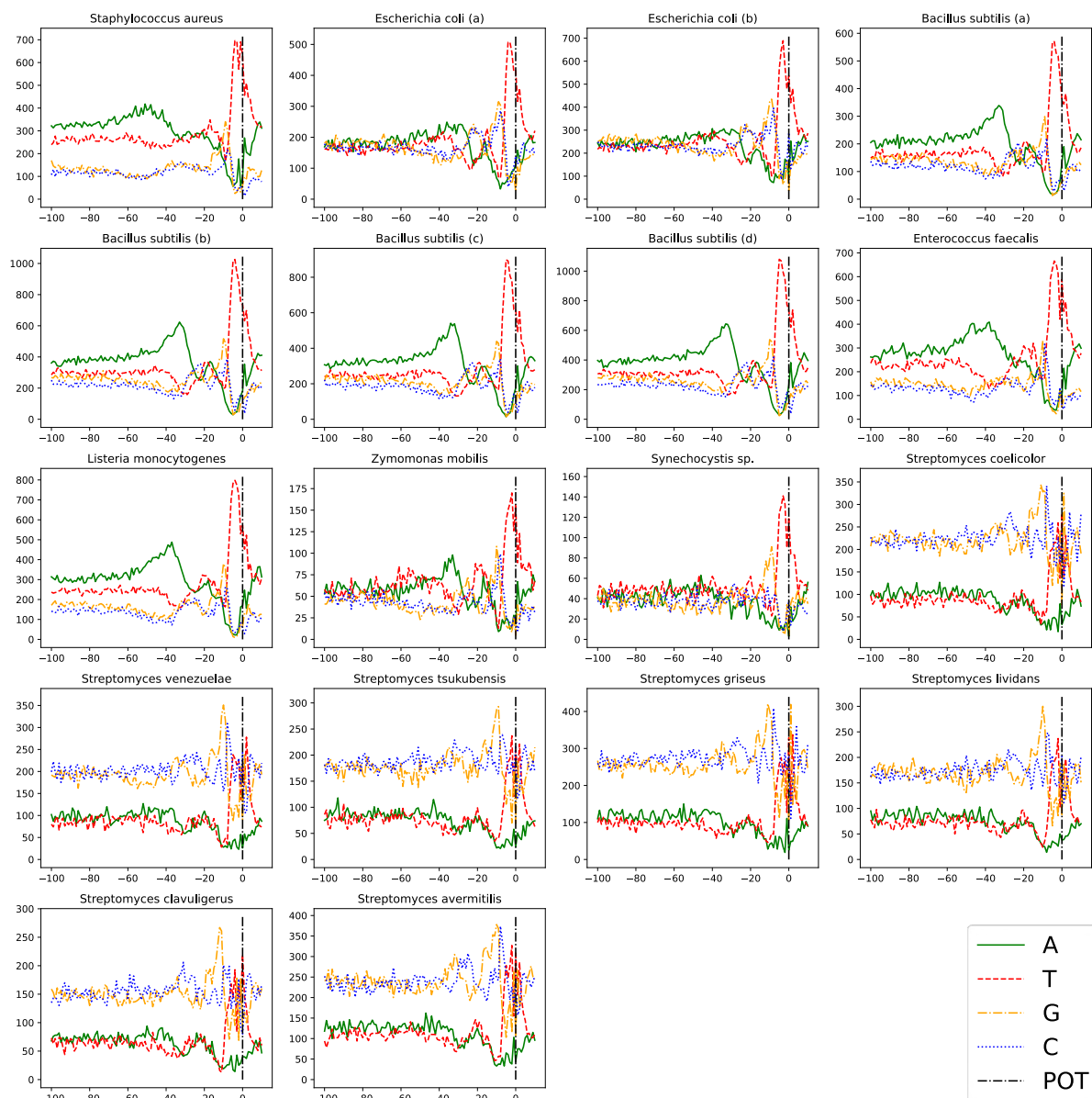


Wykres 7. Porównanie spinek terminacyjnych zidentyfikowanych u *S. aureus* oraz gatunków bakterii analizowanych w ramach rozdziału 9.1. **a)** rozkład wartości minimalnej energii swobodnej struktury typu spinka do włosów, zaadnotowanej przy pomocy programu RNAfold (patrz algorytm programu TERMITE) **b)** rozkład wartości `hairpin score` określonej przy pomocy programu TransTermHP (patrz algorytm programu TERMITE) dla zidentyfikowanych spinek terminacyjnych. Omawiany współczynnik stanowi heurystyczny odpowiednik minimalnej energii swobodnej struktury **c)** rozkład długości zidentyfikowanych spinek terminacyjnych **d)** rozkład długości pętli zidentyfikowanych w obrębie zidentyfikowanych spinek terminacyjnych

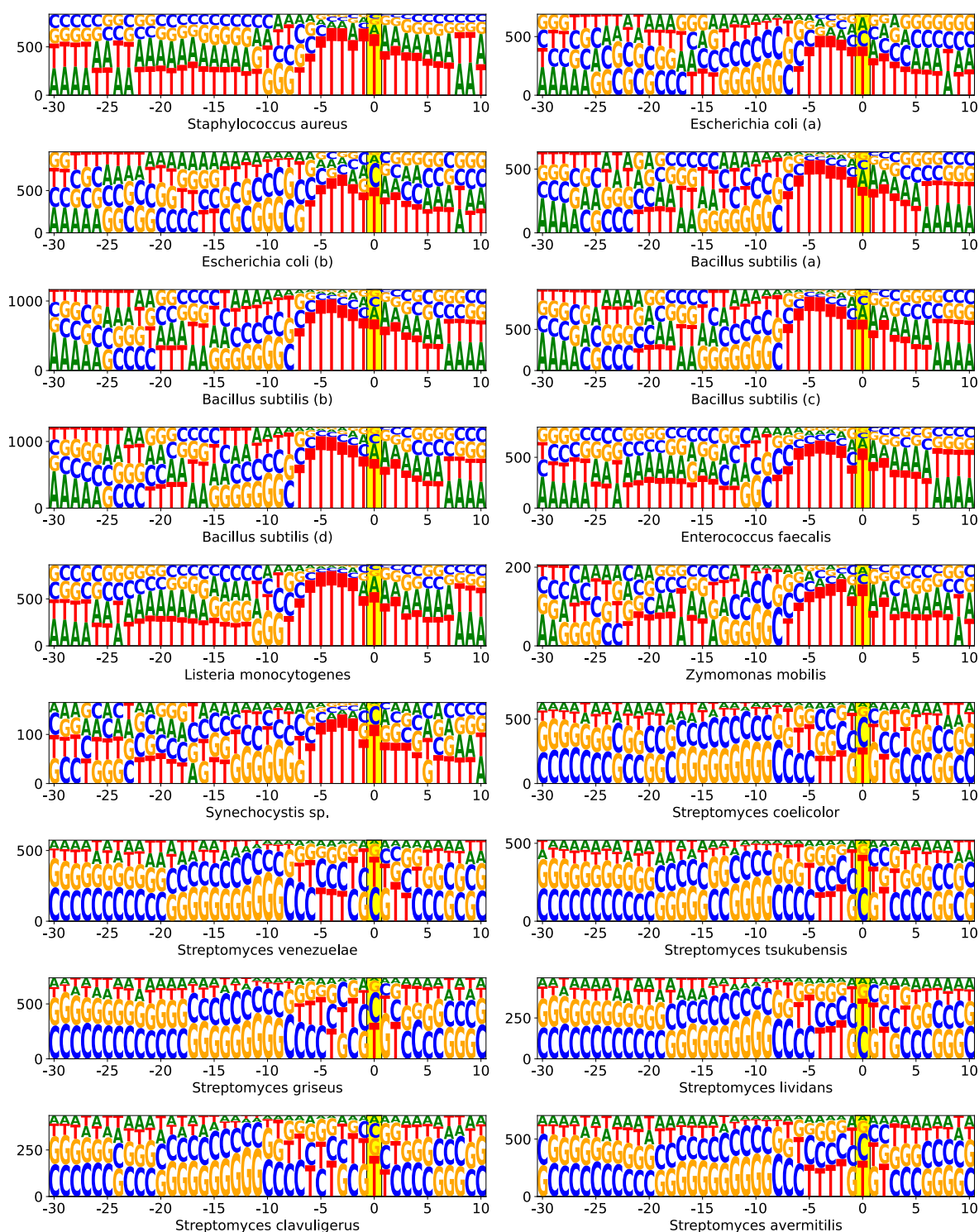
W celu dokładniejszego zrozumienia różnic obserwowanych w obrębie zidentyfikowanych terminatorów rho-niezależnych, dokonano analizy kompozycji nukleotydowej sekwencji występujących *powyżej* oraz *poniżej* względem wyznaczonego referencyjnego punktu terminacji (Wykres 8). Podobnie jak w powyższych akapitach, różnice pomiędzy poszczególnymi gatunkami, z wyłączeniem *S. aureus*, zostały szczegółowo omówione w ramach publikacji poświęconej narzędziu TERMITE (rozdział 9.1).

*S. aureus* charakteryzuje się podobną sygnaturą nukleotydową miejsc terminacji do *L. monocytogenes* oraz *E. faecalis*, z zauważalnie niższym stosunkiem A do T w regionie od 60 do 30 nukleotydów *powyżej* punktu terminacji (Wykres 8). Powyższa obserwacja może świadczyć o niższym udziale terminatorów dwukierunkowych niż u pozostałych *Firmicutes*. W zgodzie z powyższym terminatory u *S. aureus* wciąż posiadają trakt znacznie wzbogacony w A znajdujący się od 60 do 40 nukleotydów *powyżej*, a także region niezwykle bogaty w U pokrywający referencyjny punkt terminacji (występujący pomiędzy 7nt *powyżej* a 7nt *poniżej*, Wykres 9). Podobnie jak u pozostałych gatunków bakterii obserwowane jest charakterystyczne 3 nukleotydowe wzbogacenie w G tuż przed rozpoczęciem traktu bogatego w U (Wykresy 8 i 9). Istotnym jest także fakt, że struktura spinki do włosów u *S. aureus* charakteryzuje się najniższym udziałem G oraz C spośród wszystkich analizowanych bakterii (Wykres 5), co może być bezpośrednio związane z najwyższą obserwowaną zawartością A oraz T w obrębie genomu.





Wykres 8. Porównanie kompozycji nukleotydowej sekwencji występujących powyżej oraz poniżej względem referencyjnego punktu terminacji (POT) u *S. aureus* oraz gatunków bakterii analizowanych w ramach rozdziału 9.1. Oś x reprezentuje odległość od referencyjnego punktu terminacji odpowiednio powyżej (wartości ujemne) oraz poniżej (wartości dodatnie). Oś y odpowiada natomiast liczbie terminatorów, u których na danej pozycji obserwowany jest określony nukleotyd.



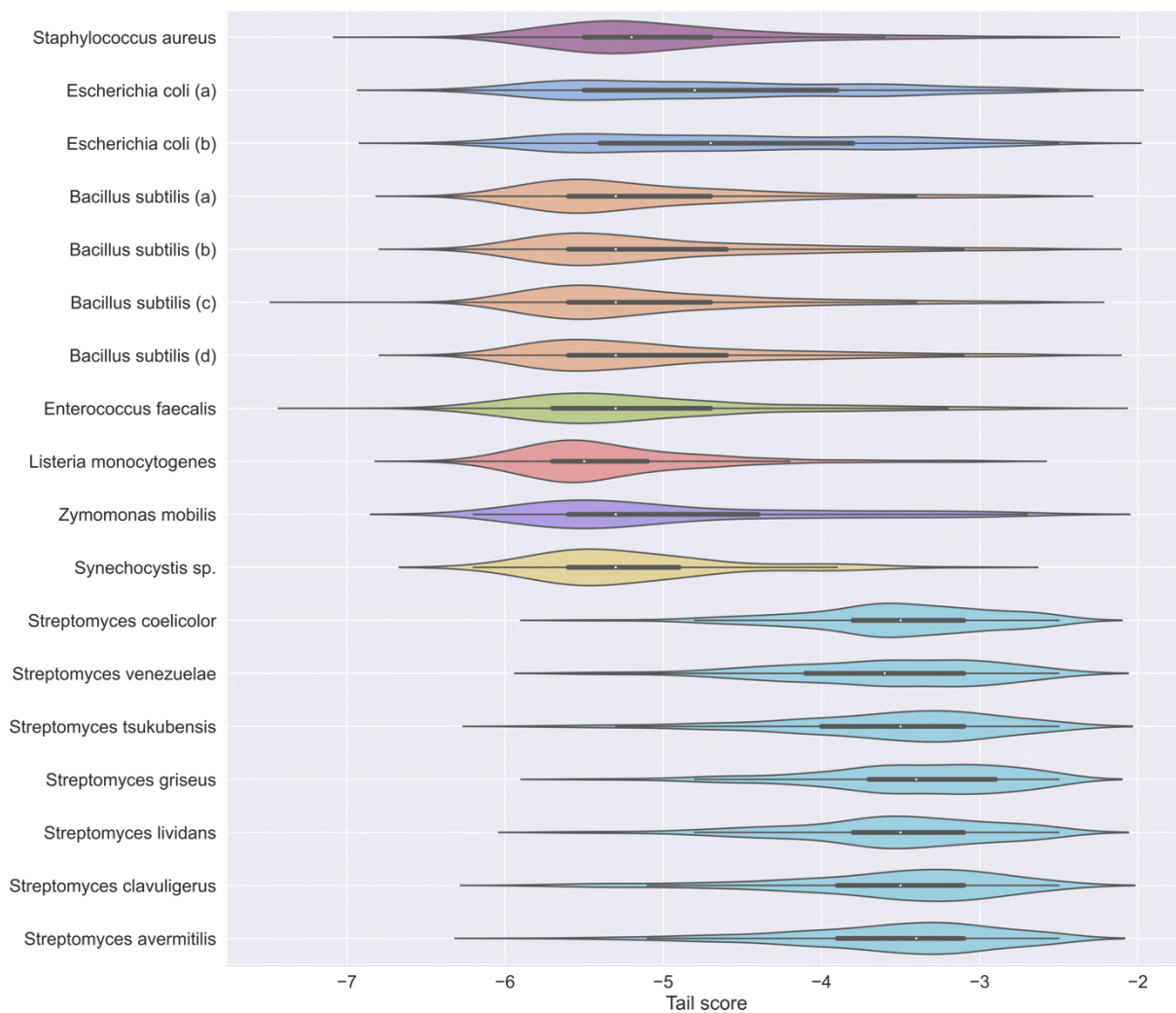
Wykres 9. Porównanie kompozycji nukleotydujowej sekwencji występujących powyżej oraz poniżej względem referencyjnego punktu terminacji (pozycja oznaczona kolorem żółtym) u *S. aureus* oraz gatunków bakterii analizowanych w ramach rozdziału 9.1. Oś x reprezentuje odległość od referencyjnego punktu terminacji odpowiednio powyżej (wartości ujemne) oraz poniżej (wartości dodatnie). Oś y odpowiada natomiast liczbie terminatorów, u których na danej pozycji obserwowany jest określony nukleotyd.

Pod względem kompozycji nukleotydowej traktu bogatego w U *S. aureus* również istotnie przypomina pozostałe bakterie przynależące do typu *Firmicutes*. Powyższego dowodzi analiza wykonana w oparciu o obliczenie udziału U w ramach sekwencji o długości 10 nukleotydów występującej bezpośrednio poniżej względem wyznaczonej spinki terminacyjnej (Wykres 10). Podobnie jak u pozostałych *Firmicutes* oraz *Synechocystis sp.* wartość ta wynosi, w znakomitej większości przypadków, od 7 do 9 nukleotydów. Mniej niż 10% terminatorów u *S. aureus* posiada w obrębie analizowanej sekwencji mniej niż 6 nukleotydów uracylowych.

U *S. aureus* rozkład wartości tail score (TS) (wykres 11, załącznik 6, rozdział 9.1) zawiera się w przedziale od -6.7 do -2.5 z medianą wynoszącą -5.2, średnią -5.02 oraz wariancją na poziomie 0.5. Zgodnie z oczekiwaniami wartości te są niezwykle zbliżone do innych bakterii należących do *Firmicutes*. Podobną charakterystyką wykazują się również terminatory obserwowane u *Synechocystis sp.* oraz *Z. mobilis*. Wykazane różnice w ramach wyżej opisanych rozkładów są istotne statystycznie (Kruskal-Wallis: 3015.6,  $p = 0.00$ ; Mann-Whitney U – załącznik 9).



Wykres 10. Udział nukleotydów uracylowych (U) w sekwencji obejmującej 10nt poniżej względem wyznaczonej spinki terminacyjnej u *S. aureus* oraz gatunków bakterii analizowanych w ramach rozdziału 9.1.



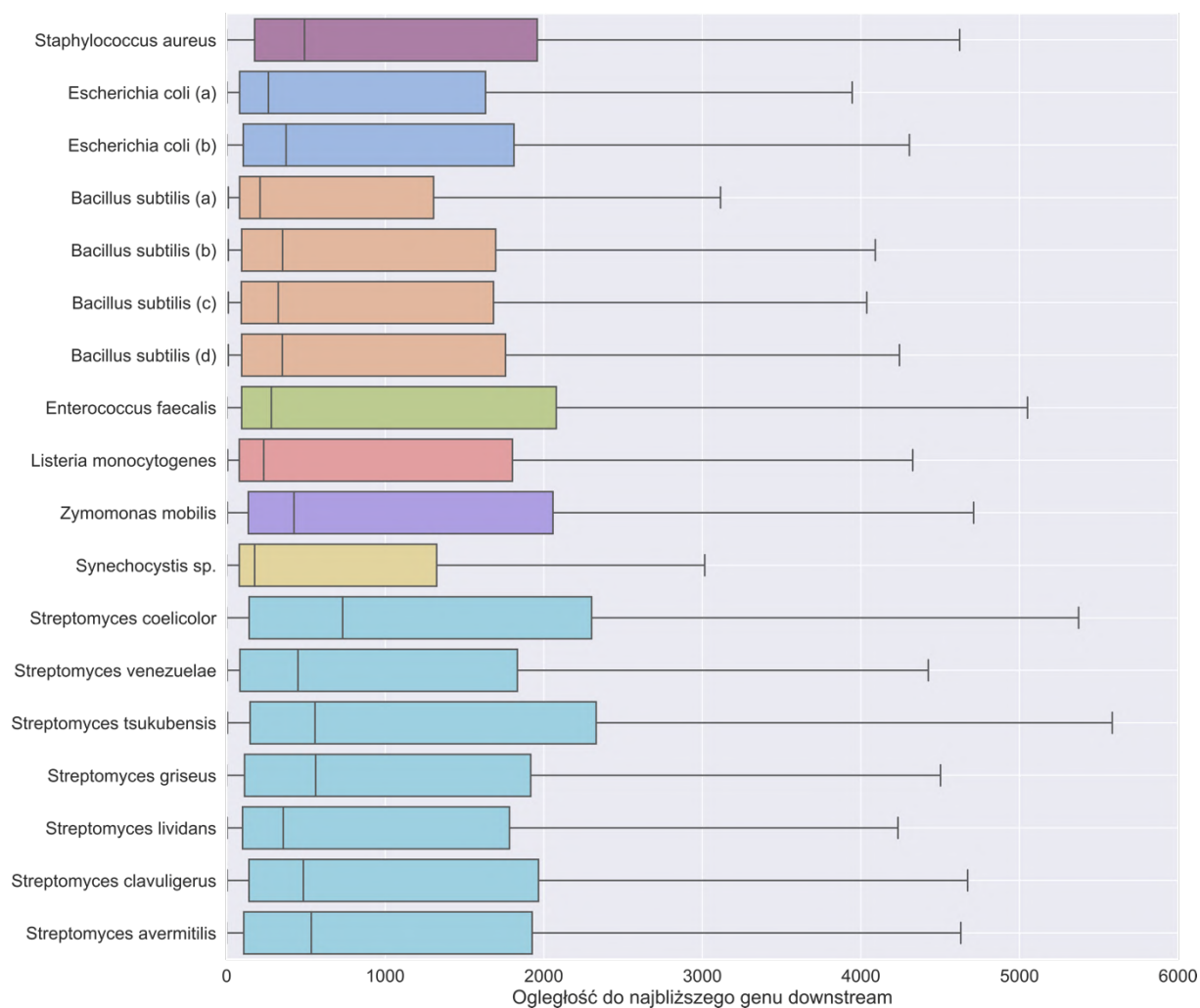
Wykres 11. Rozkład wartości `tail score` (TS) obliczanych przez program TransTermHP dla rho-niezależnych terminatorów zidentyfikowanych u *S. aureus* oraz gatunków bakterii analizowanych w ramach rozdziału 9.1.

Niezwykle interesującym aspektem funkcjonowania terminatorów rho-niezależnych u bakterii jest także ich lokalizacja genomowa. W ramach niniejszej analizy zbadana została odległość pomiędzy referencyjnymi punktami terminacji, wyznaczonymi dla poszczególnych terminatorów, a najbliższymi genami zaadnotowanymi odpowiednio *poniżej* (Wykres 12, załącznik 7) oraz *powyżej* (Wykres 13, załącznik 8) w obrębie tej samej nici. Podobnie jak w powyższych przypadkach różnice, zidentyfikowane pomiędzy poszczególnymi gatunkami bakterii, z wyjątkiem *S. aureus*, zostały szczegółowo omówione w ramach publikacji poświęconej narzędziu *TERMITe* (rozdział 9.1).

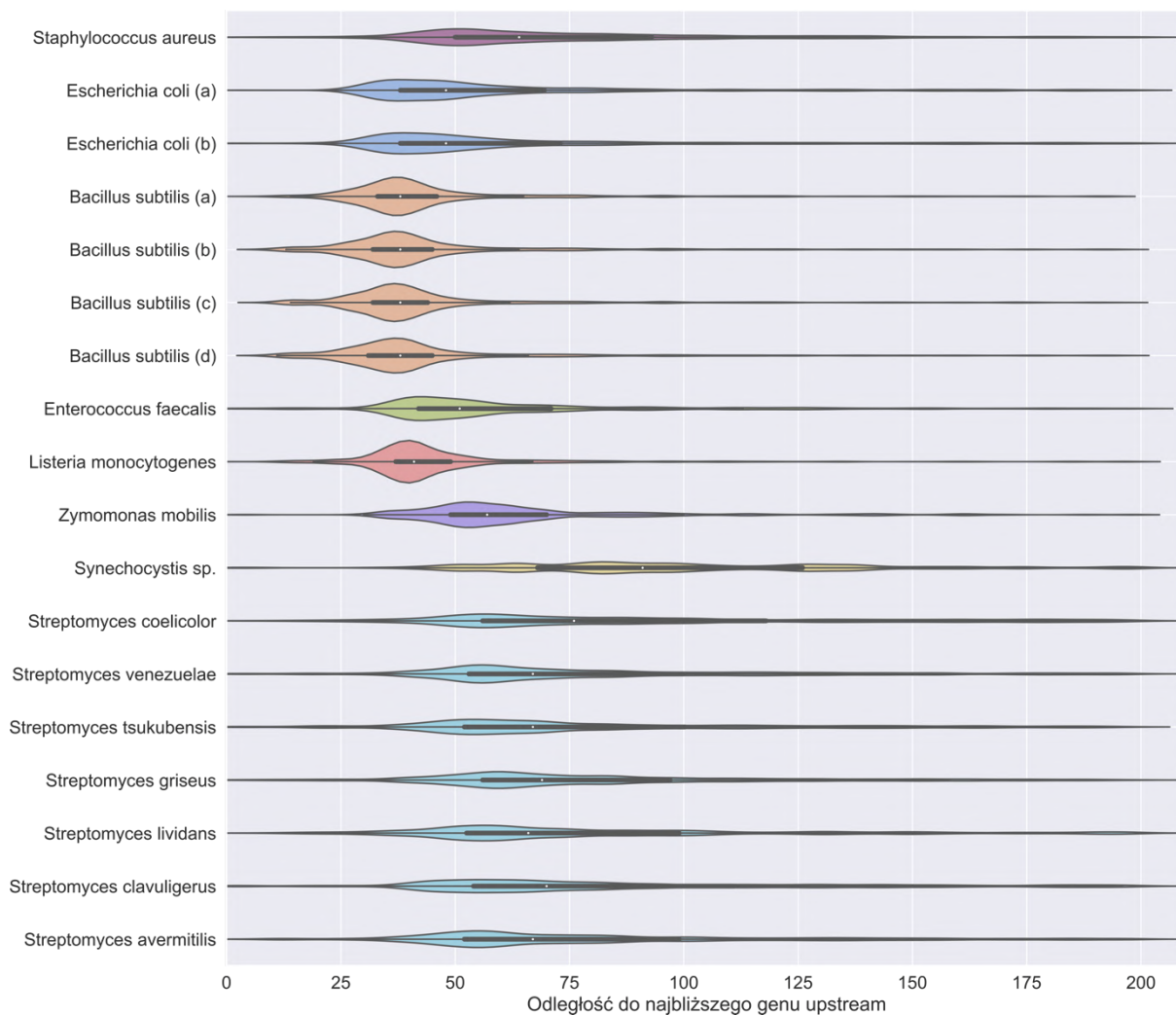
Różnice w rozkładach odległości pomiędzy punktami terminacji a genami zlokalizowanymi *powyżej* oraz *poniżej* okazały się istotne statystycznie (powyżej - Kruskal-Wallis: 3801.4,  $p = 0.0$ ; poniżej - Kruskal-Wallis: 124.5,  $p = 0.0$ ; Mann-Whitney U – załącznik

9). Obserwowane punkty terminacji u *S. aureus* znajdują się średnio 2161 nukleotydów od najbliższego genu zlokalizowanego poniżej w obrębie tej samej nici (Wykres 12, załączniki Tabela 7). Mediana odległości wynosi 491 nukleotydów. Przedstawione wartości są najwyższe spośród analizowanych bakterii z wyjątkiem rodzaju *Streptomyces*.

Okazuje się, że terminatory rho-niezależne występujące u *S. aureus* posiadają tendencję do występowania w znacznej odległości również od genów znajdujących się powyżej (Wykres 13, załącznik 8). Średnia odległość wynosi aż 76 nukleotydów (mediana 64), podczas gdy dla pozostałych analizowanych gatunków *Firmicutes* wartość ta nie przekracza 62 nukleotydów obserwowanych u *E. faecalis* (mediana 51), a często jest znacznie niższa wynosząc np. ok. 45 nukleotydów u *B. subtilis* (mediana 38) czy 47 nukleotydów u *L. monocytogenes* (mediana 41). Pod względem omawianych odległości *S. aureus* ustępuje jedynie rodzajowi *Streptomyces* oraz *Synechocystis*, dla których wartości średnie wynoszą odpowiednio od 79.4 do 90.3 oraz 98.7 nukleotydów.



Wykres 12. Rozkład odległości pomiędzy referencyjnym punktem terminacji a najbliższym genem zlokalizowanym poniżej w obrębie tej samej nici u *S. aureus* oraz gatunków bakterii analizowanych w ramach rozdziału 9.1.



Wykres 13. Rozkład odległości pomiędzy referencyjnym punktem terminacji a najbliższym genem zlokalizowanym powyżej w obrębie tej samej nici u *S. aureus* oraz gatunków bakterii analizowanych w ramach rozdziału 9.1.

Na wykresie 14 (Wykres 14) została przedstawiona zachowawczość struktury rho-niezależnego terminatora transkrypcji utworzona z wykorzystaniem programu R2R (Weinberg & Breaker, 2011) (patrz rozdział 9.1). Podsumowując, struktura i sekwencja terminatorów rho-niezależnych u *S. aureus* jest niezwykle zbliżona do kanonicznej charakterystyki terminatorów poznanej na podstawie badań *E. coli* oraz *B. subtilis* (Chełkowska-Pauszek et al., 2021). Szczególne podobieństwo obserwowane jest w stosunku do pozostałych bakterii z typu *Firmicutes*.





## 10.6. Warunkowa terminacja transkrypcji w odpowiedzi na stres antybiotykowy

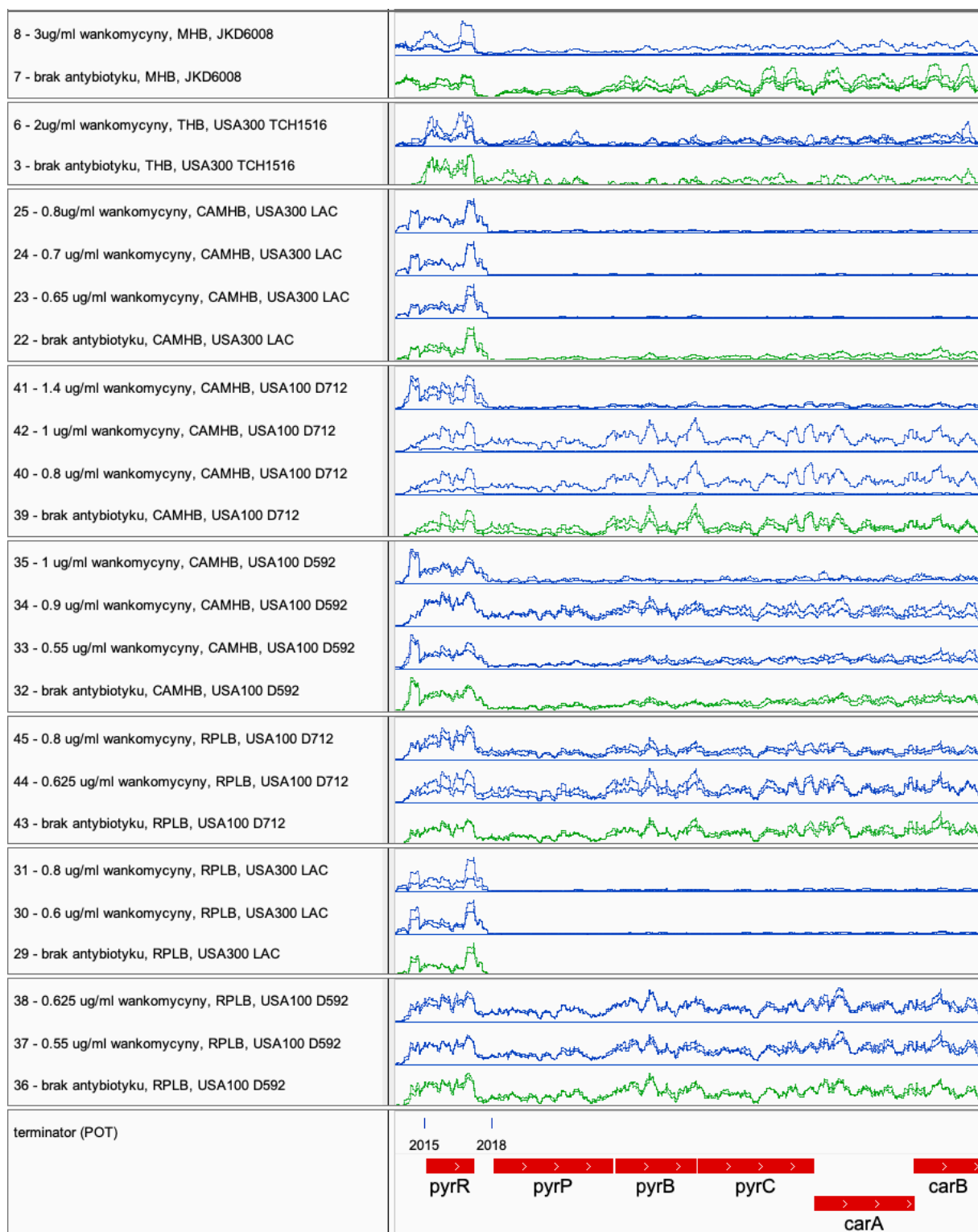
Analiza warunkowej terminacji transkrypcji w odpowiedzi na stres antybiotykowy u metycylinoopornych szczepów *S. aureus* została wykonana w oparciu o 142 publicznie dostępne próbki z sekwencjonowania metodą RNA-seq (rozdział 8.1, załącznik 1). Wykorzystane zbiory danych (pochodziły z 7 niezależnych eksperymentów polegających na traktowaniu bakterii subletalnymi dawkami antybiotyków, do których należą odpowiednio wankomycyna, nafcylina, linezolid i oksacylina. Zmiany efektywność terminacji (DT) pomiędzy replikatami traktowanymi antybiotykami, a odpowiadającymi im próbkami kontrolnymi, zostały obliczone dla wszystkich terminatorów adnotowanych w ramach atlasu opisanego w rozdziale 9.4. Szczegółowy opis protokołu analitycznego, opracowanego na potrzeby identyfikacji miejsc warunkowej terminacji transkrypcji indukowanej działaniem leków, znajduje się w rozdziale 8.5. Zastosowanie obranej metodologii umożliwiło zidentyfikowanie 11 terminatorów wykazujących istotne różnice w poziomach terminacji w odpowiedzi na stres antybiotykowy u metycylinoopornych szczepów *S. aureus* (Tabela 3).

Spośród zidentyfikowanych miejsc warunkowej terminacji transkrypcji, na potrzeby dalszej analizy, wyselekcjonowane zostały wyłącznie te zidentyfikowane w więcej niż w ramach jednego porównania prób dla tego samego antybiotyku. Powyższy warunek został spełniony wyłącznie dla terminatora oznaczonego jako `WT\_3prime\_signal\_forward\_2018` (oznaczany również jako `2018`). Przedstawione wyniki ukazują, że terminator `2018`, zlokalizowany w obrębie operonu *pyr*, jest zaangażowany w odpowiedź komórek bakteryjnych na stres antybiotykowy wywołany przez linezolid (Wykres 16) oraz nafcylinę (Wykres 18), a być może także w ograniczonym stopniu przez inne antybiotyki (Wykresy 15 i 17). Znacznie zwiększony poziom terminacji transkrypcji operonu *pyr* wywołany jest przez traktowanie komórek bakteryjnych linezolidem (1µg/ml, THB, szczep USA300 TCH1516, FDR = 0.014; Wykres 16) oraz nafcyliną (1µg/ml (FDR>0.05), 0.5µg/ml (FDR=0.000), 0.15µg/ml (FDR=0.013) w pożywce CAMHB; 0.031µg/ml (FDR>0.05), 0.063µg/ml (FDR>0.05) w pożywce RPLB, szczep USA300 LAC, Wykres 18, Tabela 3). Interesująca jest także niezwykle wysoka, w porównaniu do pozostałych analizowanych szczepów, efektywność terminacji transkrypcji terminatorów `2015` i `2018` u szczepu USA300 LAC, obserwowana zarówno w próbach traktowanych antybiotykiem oraz kontrolnych (Wykresy 15 i 18).

Powyższe obserwacje pokazują, że regulacja biosyntezy puryn jest jednym z mechanizmów indukowanych działaniem antybiotyków, być może stanowiącym istotny element odpowiedzi komórkowej na tego rodzaju warunki stresowe, o potencjalnym znaczeniu klinicznym.

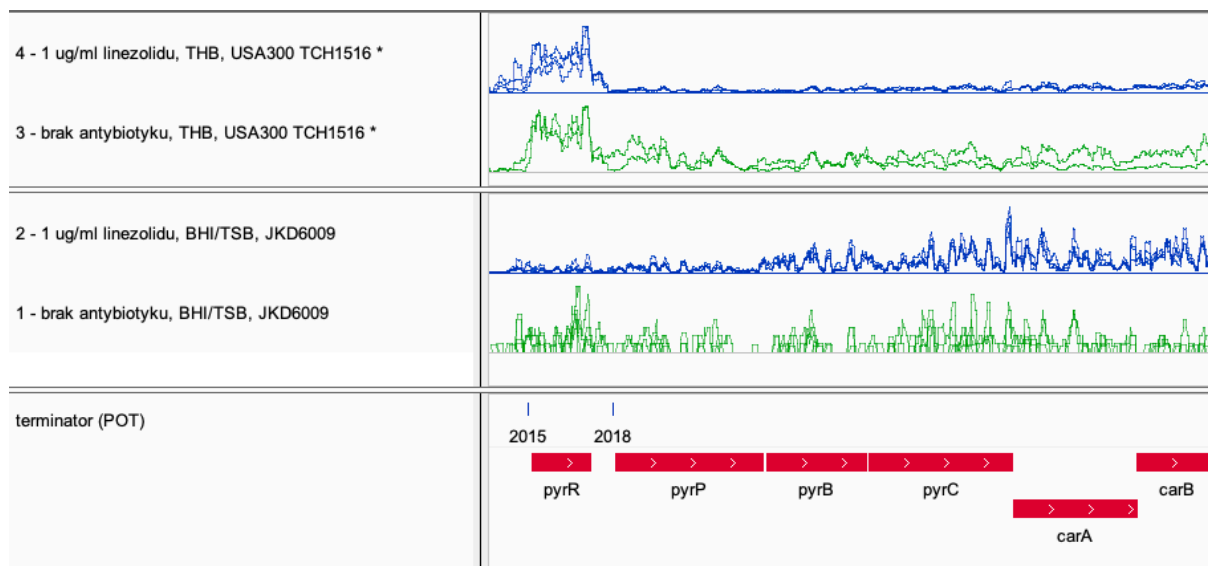
*Tabela 3. Terminatory wykazujące istotne różnice w poziomach terminacji transkrypcji w odpowiedzi na stres antybiotykowy u metycylinoopornych szczepów S. aureus. Pełna tabela, wraz z efektywnością terminacji wyznaczoną dla każdego z replikatów biologicznych, znajduje się w załączniku 12. Objaśnienia kolumn: %T – efektywność terminacji, DT – różnica pomiędzy średnią efektywnością terminacji wyznaczoną dla replikatów biologicznych reprezentujących próbę badaną oraz próbę kontrolną, statystyka T – wartość statystyki T (test T-Studenta), wartość p (test T-Studenta), FDR – False Discovery Rate. Szczegółowy opis prób i porównań znajduje się w załączniku 1.*

terminator	średnia %T [próba traktowana]	średnia %T [próba kontrolna]	DT	statystyka T	wartość p	FDR
identyfikator projektu: PRJEB23980 1 µg/ml linezolidu (warunek 4) vs brak antybiotyku (warunek 3)						
WT_3prime_signal_forward_2018	62,5	2,5	60,0	13,6	0,001	0,014
identyfikator projektu: PRJNA576951 1 µg/ml linezolidu (warunek 2) vs brak antybiotyku (warunek 1)						
WT_3prime_signal_reverse_1086	53,5	89,7	-36,2	-5,2	0,002	0,022
identyfikator projektu: GSE158830 3 µg/ml wankomycyny (warunek 8) vs brak antybiotyku (warunek 7)						
WT_3prime_signal_forward_632	60,5	16,3	44,2	4,9	0,008	0,032
WT_3prime_signal_forward_999	78,3	43,6	34,7	6,3	0,003	0,017
WT_3prime_signal_forward_1059	45,2	0,0	45,2	13,1	0,000	0,003
WT_3prime_signal_forward_1325	53,7	3,0	50,7	10,5	0,000	0,005
WT_3prime_signal_forward_2172	37,4	0,9	36,4	17,3	0,000	0,002
WT_3prime_signal_forward_2417	82,7	26,2	56,5	11,4	0,000	0,004
WT_3prime_signal_reverse_1020	71,1	24,0	47,1	13,8	0,000	0,003
WT_3prime_signal_reverse_2497	47,7	0,0	47,7	12,5	0,000	0,003
WT_3prime_signal_reverse_2940	37,9	7,2	30,7	19,0	0,000	0,001
identyfikator projektu: PRJNA470935 0.15 µg/ml nafcyliny (warunek 11) vs brak antybiotyku (warunek 9)						
WT_3prime_signal_forward_2018	55,9	3,6	52,3	9,4	0,000	0,000
identyfikator projektu: PRJNA470935 0.5 µg/ml nafcyliny (warunek 10) vs brak antybiotyku (warunek 9)						
WT_3prime_signal_forward_2018	49,0	3,6	45,4	6,6	0,000	0,013

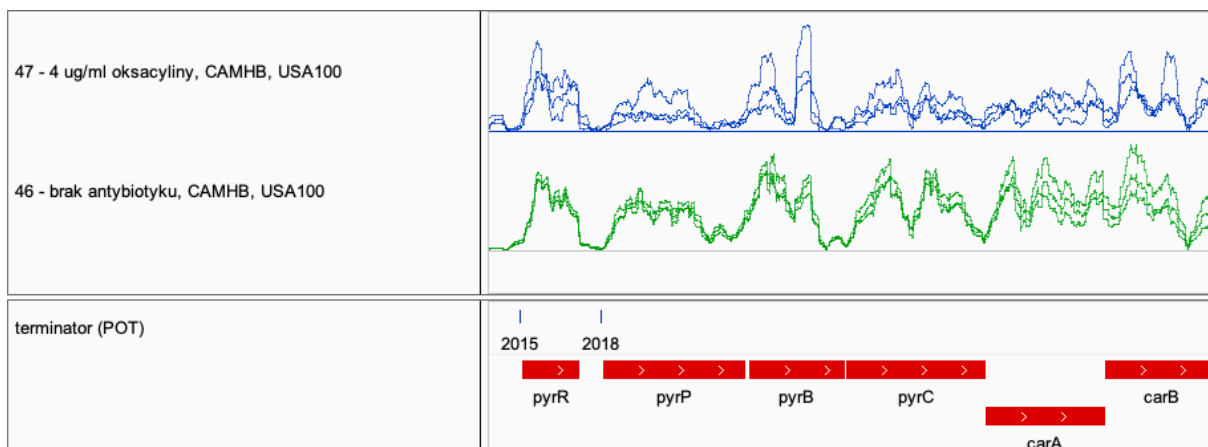


Wykres 15. Wpływ wankomycyny na regulację ekspresji operonu *pyrBI* u metycylinoopornych szczepów *S. aureus*. Wizualizacja została utworzona przy pomocy programu IGV. Niebieskie i zielone ścieżki reprezentują odpowiednio znormalizowane (CPM) pokrycie genomu odczytami z sekwencjonowania RNA-seq próbek traktowanych i nietraktowanych antybiotykiem (wyłącznie nie +). Po lewej stronie każdej ze ścieżek znajduje się kolejno adnotacja informująca o identyfikatorze warunku

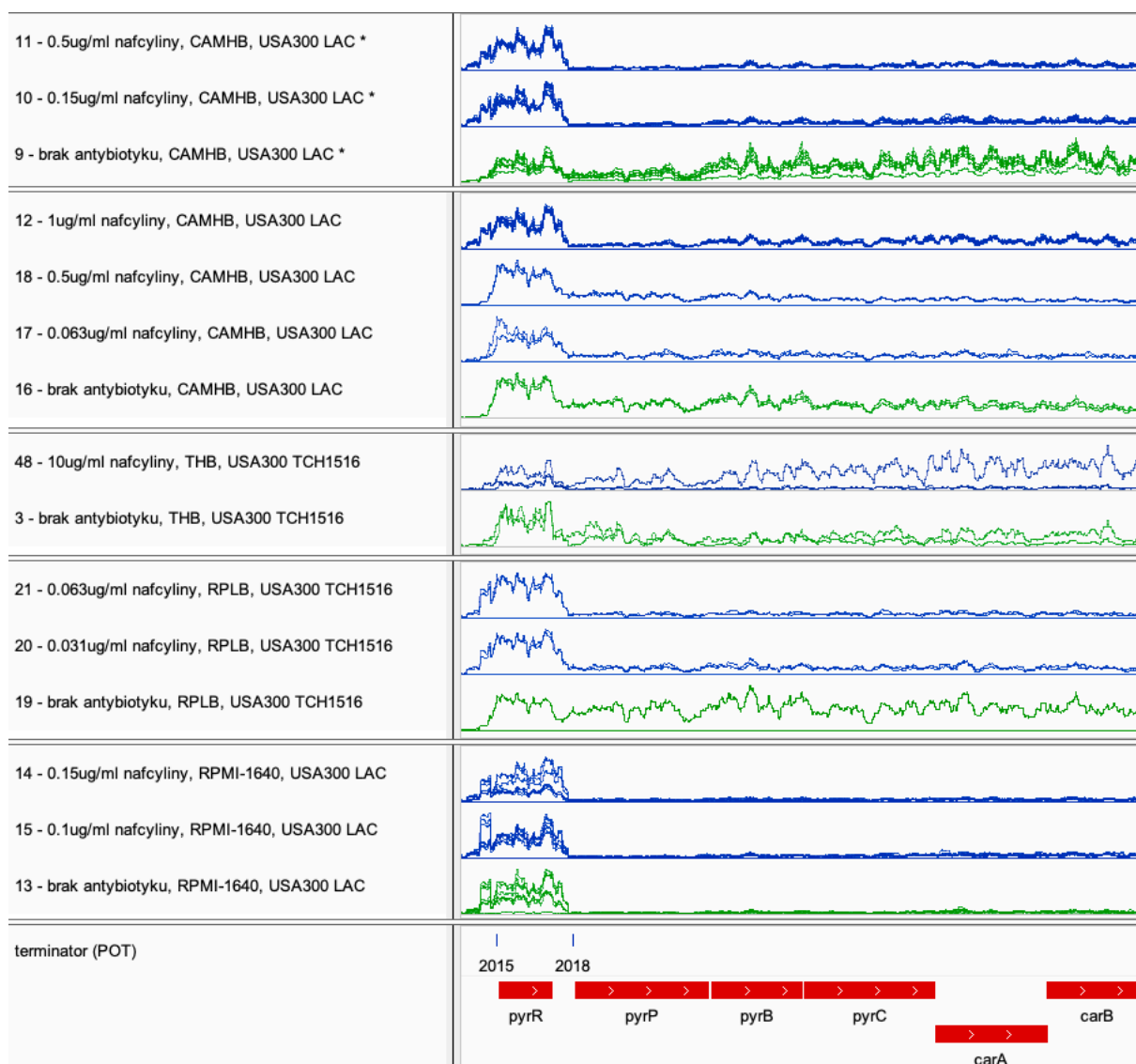
eksperymentalnego (załącznik 1), dawce antybiotyku, zastosowanym podłożu do hodowli bakteryjnej (**BHI/TS** - Brain Heart Infusion/Tryptone Soy, **THB** - Todd Hewitt Broth, **CAMHB** - Cation Adjust Mueller Hinton, **MHB** - Mueller Hinton, **RPLB** - RPMI-1640 suplementowana 10% LB) oraz szczepie. Poszczególne eksperymenty (porównania) zostały przedstawione przy pomocy osobnych paneli. Lokalizacja genów w obrębie analizowanego operonu ukazana została kolorem czerwonym, terminatorów zaś kolorem niebieskim (ścieżka terminator (POT)). Identyfikatory przedstawionych terminatorów nie zawierają przedrostka `WT\_3prime\_signal\_forward\_`. Nie zidentyfikowano istotnych statystycznie różnic w efektywności terminacji transkrypcji terminatorów, zlokalizowanych w obrębie operonu pyr, w odpowiedzi na traktowanie bakterii wankomycyną.



Wykres 16. Wpływ linezolidu na regulację ekspresji operonu *pyrBI* u metycylinoopornych szczepów *S. aureus* (patrz objaśnienia Wykres 15). Porównania, w ramach których zaobserwowane różnice w poziomach terminacji transkrypcji zostały uznane za istotne statystycznie w obrębie operonu *pyr*, oznaczono przy pomocy \*.



Wykres 17. Wpływ oksacyliny na regulację ekspresji operonu *pyrBI* u metycylinoopornych szczepów *S. aureus* (patrz objaśnienia Wykres 15). Nie zidentyfikowano istotnych statystycznie różnic w efektywności terminacji transkrypcji terminatorów, zlokalizowanych w obrębie operonu *pyr*, w odpowiedzi na traktowanie bakterii oksacyliną.



Wykres 18. Wpływ nafcyliny na regulację ekspresji operonu *pyrBI* u metycylinyopornych szczepów *S. aureus* (patrz objaśnienia Wykres 15). Porównania, w ramach których zaobserwowane różnice w poziomach terminacji transkrypcji zostały uznane za istotne statystycznie w obrębie operonu *pyr*, oznaczono przy pomocy \*.

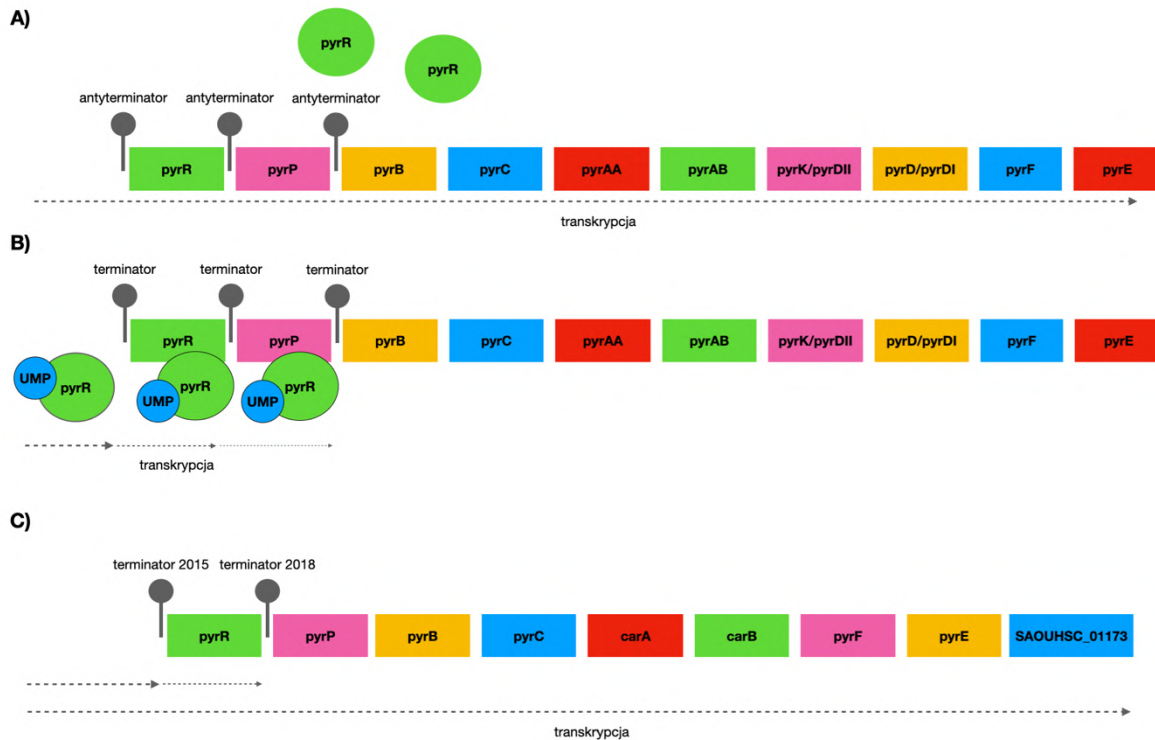


Terminator 2018 oraz sąsiedni terminator 2015 (*WT\_3prime\_signal\_forward\_2015*) występują w pozycjach niezwykle zbliżonych do opisanych w literaturze atenuatorów regulujących ekspresję operonu *pyr* u *B. subtilis*. Operon *pyr* u *B. subtilis* koduje szereg białek niezbędnych do biosyntezy pyrimidyn, w tym urydyno-5'-monofosforanu (Charles L. Turnbough & Switzer, 2008). Transkrypcja operonu rozpoczyna się z pojedynczego promotora i jest ściśle regulowana za pomocą trzech atenuatorów zlokalizowanych odpowiednio w regionie 5' liderowym operonu i regionach międzygenowych pomiędzy genami *pyrR* a *pyrP* oraz *pyrP* a *pyrB* (Charles L. Turnbough & Switzer, 2008; Patel, 2001) (Wykres 19). RNA, w obrębie regionów atenuatorowych, posiada tendencję do formowania dwóch alternatywnych struktur drugorzędowych. Pierwszą z możliwych konformacji stanowi struktura terminatora rho-niezależnego, w którego skład wchodzi spinka terminacyjna oraz trakt poli-U zlokalizowany bezpośrednio poniżej względem jej podstawy (Wykres 19B) (Charles L. Turnbough & Switzer, 2008). Alternatywnie RNA może przyjąć w tym regionie strukturę antyterminatora mającą formę długiej spinki do włosów, która zawiera szereg pętli wewnętrznych i wyklucza możliwość powstania terminatora poprzez wykorzystanie tych samych elementów sekwencji (Wykres 19A) (Charles L. Turnbough & Switzer, 2008). Kluczowym elementem regulacji operonu *pyr* jest białko *PyrR*, będące produktem genu o tej samej nazwie. *PyrR* posiada zdolność wiązania się do syntezowanego mRNA operonu *pyr*, faworyzując tym samym obieranie przez RNA konformacji odpowiadającej strukturze rho-niezależnych terminatorów transkrypcji (Charles L. Turnbough & Switzer, 2008).

Kolejnym istotnym elementem regulacji ekspresji omawianego operonu jest aktywacja białka *PyrR*, która umożliwia wiązanie do mRNA. Aktywacja przebiega na drodze wiązania rybonukleotydów urydynowych (urydyno-5'-monofosforan, urydyno-5'-trójfosforan) przez białko, podczas gdy przyłączanie rybonukleotydów guanozynowych (guanyzino-5'-monofosforan, guanyzino-5'-difosforan, guanyzino-5'-trójfosforan) skutecznie zapobiega wiązaniu *PyrR* do mRNA (Charles L. Turnbough & Switzer, 2008). Miejsca wiązania białka *PyrR* do mRNA znajdują się bezpośrednio powyżej względem sekwencji antyterminatora każdego z atenuatorów (Charles L. Turnbough & Switzer, 2008).

Podobne cechy operonu *pyr* obserwujemy u *S. aureus*, gdzie zidentyfikowane zostały struktury dwóch rho-niezależnych terminatorów transkrypcji, występujące odpowiednio w regionie 5' liderowym operonu (terminator 2015) i regionie międzygenowym pomiędzy genami *pyrR* a *pyrP* (terminator 2018). Nie obserwujemy natomiast występowania aktywnej struktury terminatorowej pomiędzy genami *pyrP* oraz *pyrB*. Istotny jest także fakt, że terminacja transkrypcji w obu rejonach genomowych jest wzajemnie skorelowana, tj., gdy

terminacja zachodzi w obrębie jednego terminatora, to również drugi z nich jest funkcjonalny (Wykresy 15-18). Powyższa obserwacja jest zgodna z modelem regulacji operonu znanym u *B. subtilis* (Wykres 19) poprzez wiązania aktywowanych białek *PyrR* do mRNA, co sugeruje, że u *S. aureus* mamy do czynienia z podobnym mechanizmem regulacji biosyntezy pirymidyn z wyłączeniem atenuatora występującego pomiędzy genami *pyrP* oraz *pyrB* (Wykres 19C).



Wykres 19. Schemat mechanizmu regulacji ekspresji operonu *pyr* u *B. subtilis* oraz organizacja operonu *pyr* u *S. aureus*. **A)** u *B. subtilis* w regionie 5' liderowym operonu *pyr* i regionach międzygenowych, pomiędzy genami *pyrR* a *pyrP*, oraz *pyrP* a *pyrB*, formują się struktury antyterminatorów zapobiegające formowaniu spinek terminacyjnych, **B)** u *B. subtilis* w przypadku wysokiego stężenia pirymidyn (UTP, UMP) w komórce, białko *pyrR*, będące produktem kodowanym przez gen o tej samej nazwie, wiąże się z cząsteczką urydyno-5'-monofosforanu lub urydyno-5'-trójfosforanu, stymulując tym samym wiązanie kompleksu w regionie znajdującym się bezpośrednio poniżej względem struktury antyterminatora. Wspomniane wiązanie promuje formowanie alternatywnej struktury rho-niezależnego terminatora transkrypcji i w konsekwencji hamowanie transkrypcji operonu. Ryciny A i B opracowane na podstawie (Patel, 2001). **C)** organizacja operonu *pyr* u *S. aureus* wraz z obserwowanymi rho-niezależnymi terminatorami transkrypcji, pozwalającymi na warunkową terminację transkrypcji operonu *pyr*.

## 11. Podsumowanie

W ramach niniejszej rozprawy doktorskiej, zostały opracowane trzy niezwykle istotne narzędzia bioinformatyczne. Pierwsze z nich, o nazwie *TERMiTe*, pozwala na identyfikację i adnotację stabilnych 3' końców RNA u bakterii na podstawie danych TERM-seq, stanowiąc najbardziej kompletne oprogramowanie do analizy tego rodzaju danych, dostępne obecnie dla społeczności naukowej. Sygnały identyfikowane przez program mogą reprezentować 3' końce stabilnych produktów przetwarzania lub degradacji RNA, a także rzeczywiste miejsca terminacji transkrypcji. Stabilność końców określa jest na podstawie oceny zachowawczości sygnału pomiędzy replikatami biologicznymi, pozwalając tym samym na odfiltrowanie pośrednich produktów przetwarzania i degradacji RNA, a także artefaktów procesu sekwencjonowania oraz mapowania odczytów. Precyzyjna adnotacja 3' końców RNA pozwala na uzyskanie szeregu informacji dotyczących, między innymi ich lokalizacji genomowej, kompozycji nukleotydowej okalającego regionu, czy też jego potencjału do tworzenia struktur drugorzędowych RNA, umożliwiając tym samym identyfikację natywnych miejsc terminacji transkrypcji, ze szczególnym uwzględnieniem tych określanych mianem rho-niezależnych.

Drugie ze wspomnianych narzędzi, o nazwie *termiRNAtor*, pozwala na czułe i precyzyjne przewidywanie miejsc rho-niezależnej terminacji transkrypcji u bakterii na podstawie danych RNA-seq, umożliwiając wykorzystanie w nowym charakterze ogromnej liczby publicznie dostępnych danych wysokoprzepustowych, wygenerowanych przy pomocy tegoż protokołu. Wspomniane oprogramowanie bazuje na algorytmie lasów losowych, który na podstawie profilu pokrycia genomu odczytami RNA-seq oraz na podstawie sekwencji genomowej, umożliwia predykcję miejsc terminacji transkrypcji. Opracowanie opisywanego modelu było możliwe poprzez analizę prób sekwencjonowanych zarówno z zastosowaniem protokołu RNA-seq oraz Term-seq. Na podstawie danych Term-seq, możliwe było dokładne określenie rzeczywistych miejsc rho-niezależnej terminacji transkrypcji (programem *TERMiTe*). Miejsca te posłużyły jako przykłady uczące do wytrenowania odpowiedniego modelu predykcyjnego, który swoje przewidywania opiera wyłącznie na podstawie zestawu cech pochodzących z danych RNA-seq oraz sekwencji genomowej. Zastosowanie wspomnianego oprogramowania pozwala zniwelować problem, wynikający z niewielkiej liczby publicznie dostępnych zestawów danych, generowanych przy pomocy protokołów

dedykowanej identyfikacji 3' końców RNA u bakterii, pozwalając w tym celu wykorzystać ogromną liczbę dostępnych bibliotek RNA-seq.

Trzecie z zaprezentowanych narzędzi, o nazwie AGouTI, służy do adnotacji interwałów genomowych oraz transkryptomowych przy pomocy informacji genomowych zawartych w plikach GTF lub GFF. Wykorzystanie opracowanych narzędzi pozwoliło na utworzenie niezwykle kompletnych i szczegółowych atlasów miejsc rho-niezależnej terminacji transkrypcji u bakterii. Na podstawie publicznie dostępnych zestawów danych, możliwe okazało się wykonanie adnotacji miejsc terminacji transkrypcji dla *E. coli*, *B. subtilis*, *E. faecalis*, *L. monocytogenes*, *Z. mobilis*, *Synechocystis sp.*, *S. coelicolor*, *S. venezuelae*, *S. tsukubaensis*, *S. griseus*, *S. lividans*, *S. clavuligerus*, *S. avermitilis* oraz *S. aureus*.

Niniejsza rozprawa doktorska przedstawia także i dyskutuje istotne różnice w charakterystyce terminatorów rho-niezależnych występujących w królestwie bakterii. Istotne różnice zaobserwowane zostały w kompozycji nukleotydowej, strukturach drugorzędowych, budowie, a także lokalizacji genomowej terminatorów występujących u omawianych taksonów. Przedstawione wyniki zdecydowanie wzbogacają dotychczasowy stan wiedzy naukowej o procesie terminacji transkrypcji, który został oparty w zdecydowanej większości o badania wykonywane na organizmach modelowych, takich jak *E. coli* oraz *B. subtilis*.

W ramach rozprawy, dokonano również identyfikacji warunkowej terminacji transkrypcji w odpowiedzi na stres antybiotykowy u metycylinoopornych szczepów *S. aureus*. Zaobserwowana została w szczególności regulacja ekspresji operonu *pyr*, w obrębie którego zidentyfikowane zostały dwa rho-niezależne terminatory transkrypcji, występujące odpowiednio w regionie 5' liderowym operonu oraz w regionie międzygenowym, pomiędzy genami *pyrR* a *pyrP*. Operon *pyr* jest odpowiedzialny za kodowanie białek, biorących udział w procesach biosyntezy pirymidyn. Regulacja ekspresji wyżej wymienionego operonu została zaobserwowana w odpowiedzi na traktowanie takimi antybiotykami jak linezolid czy nafcylina. W tym przypadku pokazano, że stres antybiotykowy może prowadzić do przedwczesnej terminacji transkrypcji operonu, redukując tym samym ekspresję produktów biosyntezy pirymidyn u metycylinoopornych szczepów *S. aureus*. Obserwacja ta może mieć znaczenie kliniczne, w szczególności w kontekście opracowywania nowych klas antybiotyków.

Podsumowując, niniejsza praca przedstawia szereg metod analitycznych, nowych narzędzi bioinformatycznych oraz opis nowatorskiej metodologii, które pozwalają na precyzyjną identyfikację i adnotację miejsc rho-niezależnej terminacji transkrypcji, a także rozpoznanie sygnałów warunkowej terminacji w odpowiedzi na stres indykowany zastosowaniem antybiotyków w leczeniu zakażeń bakteryjnych. Praca ta stanowi tym samym znakomite źródło

wiedzy i inspiracji dla osób zainteresowanym badaniem procesu terminacji transkrypcji oraz jej roli w regulacji ekspresji genów u bakterii.

## 12. Spis literatury

1. Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data. Available online at <http://www.bioinformatics.babraham.ac.uk/projects/fastqc> (0.11.9).
2. Brown, T. A. (2002). Genomes. UK: Wiley-Liss Manchester, 572. <https://www.ncbi.nlm.nih.gov/books/NBK21128/>
3. Browning, D. F., & Busby, S. J. W. (2016). Local and global regulation of transcription initiation in bacteria. *Nature Reviews. Microbiology*, 14(10), 638–650. <https://doi.org/10.1038/NRMICRO.2016.103>
4. Charles L. Turnbough, Jr., & Switzer, R. L. (2008). Regulation of Pyrimidine Biosynthetic Gene Expression in Bacteria: Repression without Repressors. *Microbiology and Molecular Biology Reviews : MMBR*, 72(2), 266. <https://doi.org/10.1128/MMBR.00001-08>
5. Chełkowska-Pauszek, A., Kosiński, J. G., Marciniak, K., Wysocka, M., Bakowska-żywicka, K., & Żywicki, M. (2021). The Role of RNA Secondary Structure in Regulation of Gene Expression in Bacteria. *International Journal of Molecular Sciences*, 22(15). <https://doi.org/10.3390/IJMS22157845>
6. Cheung, G. Y. C., Bae, J. S., & Otto, M. (2021). Pathogenicity and virulence of *Staphylococcus aureus*. *Virulence*, 12(1), 547–569. <https://doi.org/10.1080/21505594.2021.1878688>
7. Choe, D., Szubin, R., Dahesh, S., Cho, S., Nizet, V., Palsson, B., & Cho, B. K. (2018). Genome-scale analysis of Methicillin-resistant *Staphylococcus aureus* USA300 reveals a tradeoff between pathogenesis and drug resistance. *Scientific Reports*, 8(1). <https://doi.org/10.1038/S41598-018-20661-1>
8. Crick F. (1958). *Symposia of the Society for Experimental Biology*. *Symposia of the Society for Experimental Biology*, 12, 138–163.
9. Dadashi, M., Hajikhani, B., Darban-Sarokhalil, D., van Belkum, A., & Goudarzi, M. (2020). Mupirocin resistance in *Staphylococcus aureus*: A systematic review and meta-analysis. *Journal of Global Antimicrobial Resistance*, 20, 238–247. <https://doi.org/10.1016/J.JGAR.2019.07.032>
10. Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, 10(2). <https://doi.org/10.1093/gigascience/giab008>
11. Dar, D., Shamir, M., Mellin, J. R., Koutero, M., Stern-Ginossar, N., Cossart, P., & Sorek, R. (2016). Term-seq reveals abundant ribo-regulation of antibiotics resistance in bacteria. *Science (New York, N.Y.)*, 352(6282). <https://doi.org/10.1126/SCIENCE.AAD9822>
12. Davies, J., & Davies, D. (2010). Origins and evolution of antibiotic resistance. *Microbiology and Molecular Biology Reviews : MMBR*, 74(3), 9–16. <https://doi.org/10.1128/MMBR.00016-10>
13. Dersch, P., Khan, M. A., Mühlen, S., & Görke, B. (2017). Roles of Regulatory RNAs for Antibiotic Resistance in Bacteria and Their Potential Value as Novel Drug Targets. *Frontiers in Microbiology*, 8(MAY). <https://doi.org/10.3389/FMICB.2017.00803>
14. Dodds, D. R. (2017). Antibiotic resistance: A current epilogue. *Biochemical Pharmacology*, 134, 139–146. <https://doi.org/10.1016/J.BCP.2016.12.005>
15. Epshtein, V., Dutta, D., Wade, J., & Nudler, E. (2010). An Allosteric Mechanism of Rho-dependent Transcription Termination. *Nature*, 463(7278), 245. <https://doi.org/10.1038/NATURE08669>

16. Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 32(19), 3047. <https://doi.org/10.1093/BIOINFORMATICS/BTW354>
17. Fleming, A. (1980). Classics in infectious diseases: on the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of *B. influenzae* by Alexander Fleming, Reprinted from the *British Journal of Experimental Pathology* 10:226-236, 1929. *Reviews of Infectious Diseases*, 2(1), 129–139.
18. Foster, T. J. (2017). Antibiotic resistance in *Staphylococcus aureus*. Current status and future prospects. *FEMS Microbiology Reviews*, 41(3), 430–449. <https://doi.org/10.1093/FEMSRE/FUX007>
19. Gao, W., Guérillot, R., Lin, Y. H., Tree, J., Beaume, M., François, P., Monk, I. R., Seemann, T., Schrenzel, J., Howden, B. P., & Stinear, T. P. (2020). Comparative Transcriptomic and Functional Assessments of Linezolid-Responsive Small RNA Genes in *Staphylococcus aureus*. *MSystems*, 5(1). <https://doi.org/10.1128/MSYSTEMS.00665-19>
20. Gaynes, R. (2017). The Discovery of Penicillin—New Insights After More Than 75 Years of Clinical Use. *Emerging Infectious Diseases*, 23(5), 849. <https://doi.org/10.3201/EID2305.161556>
21. Guo, Y., Song, G., Sun, M., Wang, J., & Wang, Y. (2020). Prevalence and Therapies of Antibiotic-Resistance in *Staphylococcus aureus*. *Frontiers in Cellular and Infection Microbiology*, 10, 107. <https://doi.org/10.3389/FCIMB.2020.00107>
22. Gusarov, I., & Nudler, E. (1999). The mechanism of intrinsic transcription termination. *Molecular Cell*, 3(4), 495–504. [https://doi.org/10.1016/S1097-2765\(00\)80477-3](https://doi.org/10.1016/S1097-2765(00)80477-3)
23. Hao, Z., Epshtein, V., Kim, K. H., Proshkin, S., Svetlov, V., Kamarthapu, V., Bharati, B., Mironov, A., Walz, T., & Nudler, E. (2021). Pre-Termination Transcription Complex: Structure and Function. *Molecular Cell*, 81(2), 281. <https://doi.org/10.1016/J.MOLCEL.2020.11.013>
24. Hao, Z., Svetlov, V., & Nudler, E. (2021). Rho-dependent transcription termination: a revisionist view. *Transcription*, 12(4), 171. <https://doi.org/10.1080/21541264.2021.1991773>
25. Helmann, J. D. (2019). Where to begin? Sigma factors and the selectivity of transcription initiation in bacteria. *Molecular Microbiology*, 112(2), 335–347. <https://doi.org/10.1111/MMI.14309>
26. Hodille, E., Rose, W., Diep, B. A., Goutelle, S., Lina, G., & Dumitrescu, O. (2017). The Role of Antibiotics in Modulating Virulence in *Staphylococcus aureus*. *Clinical Microbiology Reviews*, 30(4), 887. <https://doi.org/10.1128/CMR.00120-16>
27. Hu, T., Chitnis, N., Monos, D., & Dinh, A. (2021). Next-generation sequencing technologies: An overview. *Human Immunology*, 82(11), 801–811. <https://doi.org/10.1016/J.HUMIMM.2021.02.012>
28. Jevons, M. P. (1961). “Celbenin” - resistant *Staphylococci*. *British Medical Journal*, 1(5219), 124. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1952888/>
29. Jian, Z., Zeng, L., Xu, T., Sun, S., Yan, S., Yang, L., Huang, Y., Jia, J., & Dou, T. (2021). Antibiotic resistance genes in bacteria: Occurrence, spread, and control. *Journal of Basic Microbiology*, 61(12), 1049–1070. <https://doi.org/10.1002/JOBM.202100201>
30. Ju, X., Li, D., & Liu, S. (2019). Full-length RNA profiling reveals pervasive bidirectional transcription terminators in bacteria. *Nature Microbiology*, 4(11), 1907–1918. <https://doi.org/10.1038/S41564-019-0500-Z>

31. Karpen, M. E., & deHaseth, P. L. (2015). Base flipping in open complex formation at bacterial promoters. *Biomolecules*, 5(2), 668–678. <https://doi.org/10.3390/BIOM5020668>
32. Kingsford, C. L., Ayanbule, K., & Salzberg, S. L. (2007). Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. *Genome Biology*, 8(2). <https://doi.org/10.1186/GB-2007-8-2-R22>
33. Koonin, E. v. (2012). Does the central dogma still stand? *Biology Direct*, 7(1), 1–7. <https://doi.org/10.1186/1745-6150-7-27/COMMENTS>
34. Kosinski, J. G. (2024). Identification and characteristics of the intrinsic transcription terminators in bacteria and their contribution to the antibiotic stress response - supplementary materials. Zenodo. <https://doi.org/10.5281/zenodo.11392921>
35. Krueger, F. (2012). Trimgalore. Available online at [https://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore/](https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) (0.6.7).
36. Kukurba, K. R., & Montgomery, S. B. (2015). RNA Sequencing and Analysis. *Cold Spring Harbor Protocols*, 2015(11), 951–969. <https://doi.org/10.1101/PDB.TOP084970>
37. Kwak, J. H., Choi, E. C., & Weisblum, B. (1991). Transcriptional attenuation control of ermK, a macrolide-lincosamide-streptogramin B resistance determinant from *Bacillus licheniformis*. *Journal of Bacteriology*, 173(15), 4725–4735. <https://doi.org/10.1128/jb.173.15.4725-4735.1991>
38. Kwiecinski, J. M., & Horswill, A. R. (2020). Staphylococcus aureus bloodstream infections: pathogenesis and regulatory mechanisms. *Current Opinion in Microbiology*, 53, 51–60. <https://doi.org/10.1016/J.MIB.2020.02.005>
39. Kwon, A.-R., Min, Y.-H., Yoon, E.-J., Kim, J.-A., Shim, M.-J., & Choi, E.-C. (2006). ErmK leader peptide : amino acid sequence critical for induction by erythromycin. *Archives of Pharmacal Research*, 29(12), 1154–1157. <https://doi.org/10.1007/BF02969307>
40. Lakhundi, S., & Zhang, K. (2018). Methicillin-Resistant Staphylococcus aureus: Molecular Characterization, Evolution, and Epidemiology. *Clinical Microbiology Reviews*, 31(4). <https://doi.org/10.1128/CMR.00020-18>
41. Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, 10(3), R25. <https://doi.org/10.1186/gb-2009-10-3-r25>
42. Lee, A. S., de Lencastre, H., Garau, J., Kluytmans, J., Malhotra-Kumar, S., Peschel, A., & Harbarth, S. (2018). Methicillin-resistant Staphylococcus aureus. *Nature Reviews. Disease Primers*, 4. <https://doi.org/10.1038/NRDP.2018.33>
43. Leinonen, R., Sugawara, H., & Shumway, M. (2011). The Sequence Read Archive. *Nucleic Acids Research*, 39(Database issue), D19. <https://doi.org/10.1093/NAR/GKQ1019>
44. Levy, S. B., & Bonnie, M. (2004). Antibacterial resistance worldwide: causes, challenges and responses. *Nature Medicine* 2004 10:12, 10(12), S122–S129. <https://doi.org/10.1038/nm1145>
45. Li, X.-Z., & Nikaido, H. (2009). Efflux-mediated drug resistance in bacteria: an update. *Drugs*, 69(12), 1555–1623. <https://doi.org/10.2165/11317030-000000000-00000>



46. Machtel, P., Bąkowska-Żywicka, K., & Żywicki, M. (2016). Emerging applications of riboswitches - from antibacterial targets to molecular tools. *Journal of Applied Genetics*, 57(4), 531–541. <https://doi.org/10.1007/S13353-016-0341-X>
47. Mandell, Z. F., Oshiro, R. T., Yakhnin, A. V., Vishwakarma, R., Kashlev, M., Kearns, D. B., & Babitzke, P. (2021). NusG is an intrinsic transcription termination factor that stimulates motility and coordinates gene expression with NusA. *ELife*, 10. <https://doi.org/10.7554/ELIFE.61880>
48. Mandell, Z. F., Vishwakarma, R. K., Yakhnin, H., Murakami, K. S., Kashlev, M., & Babitzke, P. (2022). Comprehensive transcription terminator atlas for *Bacillus subtilis*. *Nature Microbiology*, 7(11), 1918–1931. <https://doi.org/10.1038/S41564-022-01240-7>
49. Mandell, Z. F., Zemba, D., & Babitzke, P. (2022). Factor-stimulated intrinsic termination: getting by with a little help from some friends. *Transcription*, 13(4–5), 96. <https://doi.org/10.1080/21541264.2022.2127602>
50. Martinez, J. L. (2014). General principles of antibiotic resistance in bacteria. *Drug Discovery Today. Technologies*, 11(1), 33–39. <https://doi.org/10.1016/J.DDTEC.2014.02.001>
51. McGuinness, W. A., Malachowa, N., & DeLeo, F. R. (2017). Vancomycin Resistance in *Staphylococcus aureus*. *The Yale Journal of Biology and Medicine*, 90(2), 269–281. <https://doi.org/10.1201/9780849340574-15>
52. Mediati, D. G., Wong, J. L., Gao, W., McKellar, S., Pang, C. N. I., Wu, S., Wu, W., Sy, B., Monk, I. R., Biazik, J. M., Wilkins, M. R., Howden, B. P., Stinear, T. P., Granneman, S., & Tree, J. J. (2022). RNase III-CLASH of multi-drug resistant *Staphylococcus aureus* reveals a regulatory mRNA 3'UTR required for intermediate vancomycin resistance. *Nature Communications*, 13(1). <https://doi.org/10.1038/S41467-022-31177-8>
53. Mejía-Almonte, C., Busby, S. J. W., Wade, J. T., van Helden, J., Arkin, A. P., Stormo, G. D., Eilbeck, K., Pálsson, B. O., Galagan, J. E., & Collado-Vides, J. (2020). Redefining fundamental concepts of transcription initiation in bacteria. *Nature Reviews. Genetics*, 21(11), 699–714. <https://doi.org/10.1038/S41576-020-0254-8>
54. Mondal, S., Yakhnin, A. v., Sebastian, A., Albert, I., & Babitzke, P. (2016). NusA-dependent transcription termination prevents misregulation of global gene expression. *Nature Microbiology*, 1(1). <https://doi.org/10.1038/NMICROBIOL.2015.7>
55. Moore, P. B., & Steitz, T. A. (2011). The Roles of RNA in the Synthesis of Protein. *Cold Spring Harbor Perspectives in Biology*, 3(11). <https://doi.org/10.1101/CSHPERSPECT.A003780>
56. Mulhbach, J., Brouillette, E., Allard, M., Fortier, L. C., Malouin, F., & Lafontaine, D. A. (2010). Novel Riboswitch Ligand Analogs as Selective Inhibitors of Guanine-Related Metabolic Pathways. *PLoS Pathogens*, 6(4), 1–11. <https://doi.org/10.1371/JOURNAL.PPAT.1000865>
57. Munita, J. M., & Cesar, A. A. (2016). Mechanisms of Antibiotic Resistance. *Microbiology Spectrum*, 4(2), 464–472. <https://doi.org/10.1128/MICROBIOLSPEC.VMBF-0016-2015>
58. Murray, C. J., Ikuta, K. S., Sharara, F., Swetschinski, L., Robles Aguilar, G., Gray, A., Han, C., Bisignano, C., Rao, P., Wool, E., Johnson, S. C., Browne, A. J., Chipeta, M. G., Fell, F., Hackett, S., Haines-Woodhouse, G., Kashef Hamadani, B. H., Kumaran, E. A. P., McManigal, B., ... Naghavi, M. (2022). Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *The Lancet*, 399(10325), 629–655. [https://doi.org/10.1016/S0140-6736\(21\)02724-0](https://doi.org/10.1016/S0140-6736(21)02724-0)

59. Mustaev, A., Roberts, J., & Gottesman, M. (2017). Transcription elongation. *Transcription*, 8(3), 150. <https://doi.org/10.1080/21541264.2017.1289294>
60. Okonechnikov, K., Conesa, A., & García-Alcalde, F. (2016). Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics*, 32(2), 292. <https://doi.org/10.1093/BIOINFORMATICS/BTV566>
61. O'Neill, A. J., & Chopra, I. (2006). Molecular basis of fusB-mediated resistance to fusidic acid in *Staphylococcus aureus*. *Molecular Microbiology*, 59(2), 664–676. <https://doi.org/10.1111/J.1365-2958.2005.04971.X>
62. O'Neill, J. (2016). Tackling drug-resistant infections globally: final report and recommendations. Downloaded from <https://Amr-Review.Org/>, 1–16.
63. Osbourn, A. E., & Field, B. (2009). Operons. *Cellular and Molecular Life Sciences*, 66(23), 3755. <https://doi.org/10.1007/S00018-009-0114-3>
64. Pantosti, A., Sanchini, A., & Monaco, M. (2007). Mechanisms of antibiotic resistance in *Staphylococcus aureus*. *Future Microbiology*, 2(3), 323–334. <https://doi.org/10.2217/17460913.2.3.323>
65. Patel, M. V. (2001). The regulatory roles of PyrR and Crc in pyrimidine metabolism in *Pseudomonas aeruginosa*. <https://digital.library.unt.edu/ark:/67531/metadc2875/>
66. Pedregosa, F., Michel, V., Grisel OLIVIERGRISEL, O., Blondel, M., Prettenhofer, P., Weiss, R., Vanderplas, J., Cournapeau, D., Pedregosa, F., Varoquaux, G., Gramfort, A., Thirion, B., Grisel, O., Dubourg, V., Passos, A., Brucher, M., Perrot and Édouardand, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12(85), 2825–2830. <http://jmlr.org/papers/v12/pedregosa11a.html>
67. Pietrusinski, M., & Staczek, P. (2006). Badania struktury i funkcji prokariotycznych topoizomeraz klasy II szansa dla poszukiwania i syntezy nowych leków przeciwbakteryjnych. *Postępy Biochemii*, 52(3), 271–282.
68. Poudel, S., Tsunemoto, H., Seif, Y., Sastry, A. V., Szubin, R., Xu, S., Machado, H., Olson, C. A., Anand, A., Pogliano, J., Nizet, V., & Palsson, B. O. (2020). Revealing 29 sets of independently modulated genes in *Staphylococcus aureus*, their regulators, and role in key physiological response. *Proceedings of the National Academy of Sciences of the United States of America*, 117(29), 17228–17239. <https://doi.org/10.1073/PNAS.2008413117>
69. Quinlan, A. R. (2014). BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Current Protocols in Bioinformatics*, 47, 11.12.1-11.12.34. <https://doi.org/10.1002/0471250953.BI1112S47>
70. Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)*, 26(6), 841–842. <https://doi.org/10.1093/BIOINFORMATICS/BTQ033>
71. Rajput, A., Poudel, S., Tsunemoto, H., Meehan, M., Szubin, R., Olson, C. A., Seif, Y., Lamsa, A., Dillon, N., Vrbanac, A., Sugie, J., Dahesh, S., Monk, J. M., Dorrestein, P. C., Knight, R., Pogliano, J., Nizet, V., Feist, A. M., & Palsson, B. O. (2021). Identifying the effect of vancomycin on health care-associated methicillin-resistant *Staphylococcus aureus* strains using bacteriological and physiological media. *GigaScience*, 10(1). <https://doi.org/10.1093/GIGASCIENCE/GIAA156>
72. Ramírez, F., Ryan, D. P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., Heyne, S., Dündar, F., & Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research*, 44(W1), W160–W165. <https://doi.org/10.1093/NAR/GKW257>

73. Rayner, C., & Munckhof, W. J. (2005). Antibiotics currently used in the treatment of infections caused by *Staphylococcus aureus*. *Internal Medicine Journal*, 35 Suppl 2(SUPPL. 2). <https://doi.org/10.1111/J.1444-0903.2005.00976.X>
74. Ray-Soni, A., Bellecourt, M. J., & Landick, R. (2016). Mechanisms of Bacterial Transcription Termination: All Good Things Must End. *Annual Review of Biochemistry*, 85, 319–347. <https://doi.org/10.1146/ANNUREV-BIOCHEM-060815-014844>
75. Reilman, E., Mars, R. A. T., van Dijl, J. M., & Denham, E. L. (2014). The multidrug ABC transporter BmrC/BmrD of *Bacillus subtilis* is regulated via a ribosome-mediated transcriptional attenuation mechanism. *Nucleic Acids Research*, 42(18), 11393–11407. <https://doi.org/10.1093/nar/gku832>
76. Roberts, J. W. (2019). Mechanisms of Bacterial Transcription Termination. *Journal of Molecular Biology*, 431(20), 4030–4039. <https://doi.org/10.1016/J.JMB.2019.04.003>
77. Rodnina, M. v. (2018). Translation in Prokaryotes. *Cold Spring Harbor Perspectives in Biology*, 10(9). <https://doi.org/10.1101/CSHPERSPECT.A032664>
78. Said, N., Hilal, T., Sunday, N. D., Khatri, A., Bürger, J., Mielke, T., Belogurov, G. A., Loll, B., Sen, R., Artsimovitch, I., & Wahl, M. C. (2021). Steps toward translocation-independent RNA polymerase inactivation by terminator ATPase  $\rho$ . *Science (New York, N.Y.)*, 371(6524). <https://doi.org/10.1126/SCIENCE.ABD1673>
79. Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and Statistical Modeling with Python. PROC. OF THE 9th PYTHON IN SCIENCE CONF. <http://statsmodels.sourceforge.net/>
80. Slatko, B. E., Gardner, A. F., & Ausubel, F. M. (2018). Overview of Next-Generation Sequencing Technologies. *Current Protocols in Molecular Biology*, 122(1). <https://doi.org/10.1002/CPMB.59>
81. Takada, H., Mandell, Z. F., Yakhnin, H., Glazyrina, A., Chiba, S., Kurata, T., Wu, K. J. Y., Tresco, B. I. C., Myers, A. G., Aktinson, G. C., Babitzke, P., & Haurlyliuk, V. (2022). Expression of *Bacillus subtilis* ABCF antibiotic resistance factor VmlR is regulated by RNA polymerase pausing, transcription attenuation, translation attenuation and (p)ppGpp. *Nucleic Acids Research*, 50(11), 6174–6189. <https://doi.org/10.1093/NAR/GKAC497>
82. Temin, H. M. (1985). Reverse transcription in the eukaryotic genome: retroviruses, pararetroviruses, retrotransposons, and retrotranscripts. *Molecular Biology and Evolution*, 2(6), 455–468. <https://doi.org/10.1093/OXFORDJOURNALS.MOLBEV.A040365>
83. Thorvaldsdóttir, H., Robinson, J. T., & Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, 14(2), 178–192. <https://doi.org/10.1093/BIB/BBS017>
84. Tong, S. Y. C., Davis, J. S., Eichenberger, E., Holland, T. L., & Fowler, V. G. (2015). *Staphylococcus aureus* infections: epidemiology, pathophysiology, clinical manifestations, and management. *Clinical Microbiology Reviews*, 28(3), 603–661. <https://doi.org/10.1128/CMR.00134-14>
85. Torres, C., Galián, C., Freiberg, C., Fantino, J.-R., & Jault, J.-M. (2009). The YheI/YheH heterodimer from *Bacillus subtilis* is a multidrug ABC transporter. *Biochimica et Biophysica Acta*, 1788(3), 615–622. <https://doi.org/10.1016/j.bbamem.2008.12.012>
86. Vestergaard, M., Frees, D., & Ingmer, H. (2019). Antibiotic Resistance and the MRSA Problem. *Microbiology Spectrum*, 7(2). <https://doi.org/10.1128/MICROBIOLSPEC.GPP3-0057-2018>

87. Viering, B., Cunningham, T., King, A., Blackledge, M. S., & Miller, H. B. (2022). Brominated Carbazole with Antibiotic Adjuvant Activity Displays Pleiotropic Effects in MRSA's Transcriptome. *ACS Chemical Biology*, 17(5). <https://doi.org/10.1021/ACSCHEMBIO.2C00168>
88. Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., ... SciPy 1.0 Contributors. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3), 261–272. <https://doi.org/10.1038/s41592-019-0686-2>
89. Waglechner, N., & Wright, G. D. (2017). Antibiotic resistance: it's bad, but why isn't it worse? *BMC Biology*, 15(1). <https://doi.org/10.1186/S12915-017-0423-1>
90. Wang, C., Molodtsov, V., Firlar, E., Kaelber, J. T., Blaha, G., Su, M., & Ebright, R. H. (2020). Structural basis of transcription-translation coupling. *Science (New York, N.Y.)*, 369(6509), 1359–1365. <https://doi.org/10.1126/science.abb5317>
91. Washburn, R. S., & Gottesman, M. E. (2015). Regulation of Transcription Elongation and Termination. *Biomolecules*, 5(2), 1063. <https://doi.org/10.3390/BIOM5021063>
92. Waters, L. S., & Storz, G. (2009). Regulatory RNAs in Bacteria. *Cell*, 136(4), 615. <https://doi.org/10.1016/J.CELL.2009.01.043>
93. Weinberg, Z., & Breaker, R. R. (2011). R2R - software to speed the depiction of aesthetic consensus RNA secondary structures. *BMC Bioinformatics*, 12(1), 1–9. <https://doi.org/10.1186/1471-2105-12-3/FIGURES/7>
94. Zhang, H., He, L., & Cai, L. (2018). Transcriptome Sequencing: RNA-Seq. *Methods in Molecular Biology (Clifton, N.J.)*, 1754, 15–27. [https://doi.org/10.1007/978-1-4939-7717-8\\_2](https://doi.org/10.1007/978-1-4939-7717-8_2)

## 13. Załączniki

### Załącznik 1

**Tabela 1.** Wykaz publicznie dostępnych zbiorów danych pochodzących z sekwencjonowania transkryptomu metodą RNA-seq wykorzystanych do przewidywania miejsc rho-niezależnej terminacji transkrypcji u gronkowca złocistego oraz miejsc warunkowej terminacji transkrypcji w odpowiedzi na działanie stresu antybiotykowego. Identyfikatory projektów i próbek pochodzą z bazy danych Sequence Read Archive (<https://www.ncbi.nlm.nih.gov/sra/>) oraz Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo/>). Próbkę oznaczoną \* zostały wykorzystane wyłącznie w celu identyfikacji miejsc warunkowej terminacji transkrypcji w odpowiedzi na działanie stresu antybiotykowego przez wzgląd na brak niciowej specyficzności odczytów (4 próbki). Próbki posiadające identyczną wartość w kolumnie Warunek reprezentują replikaty biologiczne reprezentujące ten sam warunek eksperymentalny.

Identyfikator projektu	Identyfikator próbki	Krótki opis	Warunek
PRJNA576951	SRR10272845	MRSA szczep JKD6009; brak antybiotyku; pożywka Brain Heart Infusion (Becton Dickinson) lub Tryptone Soy (Oxoid)	1
PRJNA576951	SRR10272846	MRSA szczep JKD6009; brak antybiotyku; pożywka Brain Heart Infusion (Becton Dickinson) lub Tryptone Soy (Oxoid)	1
PRJNA576951	SRR10272847	MRSA szczep JKD6009; brak antybiotyku; pożywka Brain Heart Infusion (Becton Dickinson) lub Tryptone Soy (Oxoid)	1
PRJNA576951	SRR10272848	MRSA szczep JKD6009; brak antybiotyku; pożywka Brain Heart Infusion (Becton Dickinson) lub Tryptone Soy (Oxoid)	1
PRJNA576951	SRR10272849	MRSA szczep JKD6009; 1 µg/ml linezolid; pożywka Brain Heart Infusion (Becton Dickinson) lub Tryptone Soy (Oxoid)	2
PRJNA576951	SRR10272850	MRSA szczep JKD6009; 1 µg/ml linezolid; pożywka Brain Heart Infusion (Becton Dickinson) lub Tryptone Soy (Oxoid)	2
PRJNA576951	SRR10272851	MRSA szczep JKD6009; 1 µg/ml linezolid; pożywka Brain Heart Infusion (Becton Dickinson) lub Tryptone Soy (Oxoid)	2
PRJNA576951	SRR10272852	MRSA szczep JKD6009; 1 µg/ml linezolid; pożywka Brain Heart Infusion (Becton Dickinson) lub Tryptone Soy (Oxoid)	2
PRJEB23980	SRR5432014*	MRSA szczep USA300 TCH1516; brak antybiotyku; pożywka Todd Hewitt Broth (Hardy Diagnostics)	3
PRJEB23980	SRR5432015*	MRSA szczep USA300 TCH1516; brak antybiotyku; pożywka Todd Hewitt Broth (Hardy Diagnostics)	3

<b>PRJEB23980</b>	SRR5432019	MRSA szczep USA300 TCH1516; 1µg/ml linezolid; pożywka Todd Hewitt Broth (Hardy Diagnostics)	4
<b>PRJEB23980</b>	SRR5432020	MRSA szczep USA300 TCH1516; 1µg/ml linezolid; pożywka Todd Hewitt Broth (Hardy Diagnostics)	4
<b>PRJEB23980</b>	SRR5432021	MRSA szczep USA300 TCH1516; 1µg/ml linezolid; pożywka Todd Hewitt Broth (Hardy Diagnostics)	4
<b>PRJNA470935</b>	SRR6391831	MRSA szczep USA300 TCH1516; 1µg/ml nafcyliny; pożywka Cation Adjust Mueller Hinton	5
<b>PRJNA470935</b>	SRR6391832	MRSA szczep USA300 TCH1516; 1µg/ml nafcyliny; pożywka Cation Adjust Mueller Hinton	5
<b>PRJNA470935</b>	SRR6391846	MRSA szczep USA300 TCH1516; 1µg/ml nafcyliny; pożywka Cation Adjust Mueller Hinton	5
<b>PRJEB23980</b>	SRR5432016*	MRSA szczep USA300 TCH1516; 2µg/ml wankomycyny; pożywka Todd Hewitt Broth (Hardy Diagnostics)	6
<b>PRJEB23980</b>	SRR5432017	MRSA szczep USA300 TCH1516; 2µg/ml wankomycyny; pożywka Todd Hewitt Broth (Hardy Diagnostics)	6
<b>PRJEB23980</b>	SRR5432018*	MRSA szczep USA300 TCH1516; 2µg/ml wankomycyny; pożywka Todd Hewitt Broth (Hardy Diagnostics)	6
<b>GSE158830</b>	SRR17697171	MRSA szczep JKD6008; brak antybiotyku; pożywka Mueller Hinton (Merck)	7
<b>GSE158830</b>	SRR17697172	MRSA szczep JKD6008; brak antybiotyku; pożywka Mueller Hinton (Merck)	7
<b>GSE158830</b>	SRR17697173	MRSA szczep JKD6008; brak antybiotyku; pożywka Mueller Hinton (Merck)	7
<b>GSE158830</b>	SRR12748682	MRSA szczep JKD6008; 3µg/ml wankomycyny; pożywka Mueller Hinton (Merck)	8
<b>GSE158830</b>	SRR12748683	MRSA szczep JKD6008; 3µg/ml wankomycyny; pożywka Mueller Hinton (Merck)	8
<b>GSE158830</b>	SRR12748684	MRSA szczep JKD6008; 3µg/ml wankomycyny; pożywka Mueller Hinton (Merck)	8
<b>PRJNA470935</b>	SRR8195171	MRSA szczep USA300 LAC; brak antybiotyku; pożywka Cation Adjust Mueller Hinton	9
<b>PRJNA470935</b>	SRR8195174	MRSA szczep USA300 LAC; brak antybiotyku; pożywka Cation Adjust Mueller Hinton	9
<b>PRJNA470935</b>	SRR8195173	MRSA szczep USA300 LAC; brak antybiotyku; pożywka Cation Adjust Mueller Hinton	9
<b>PRJNA470935</b>	SRR8195198	MRSA szczep USA300 LAC; brak antybiotyku; pożywka Cation Adjust Mueller Hinton	9
<b>PRJNA470935</b>	SRR8195197	MRSA szczep USA300 LAC; brak antybiotyku; pożywka Cation Adjust Mueller Hinton	9
<b>PRJNA470935</b>	SRR8195196	MRSA szczep USA300 LAC; brak antybiotyku; pożywka Cation Adjust Mueller Hinton	9
<b>PRJNA470935</b>	SRR7146315	MRSA szczep USA300 LAC; 0.15µg/ml nafcylina; pożywka Cation Adjust Mueller Hinton	10
<b>PRJNA470935</b>	SRR7146317	MRSA szczep USA300 LAC; 0.15µg/ml nafcylina; pożywka Cation Adjust Mueller Hinton	10
<b>PRJNA470935</b>	SRR7146318	MRSA szczep USA300 LAC; 0.15µg/ml nafcylina; pożywka Cation Adjust Mueller Hinton	10
<b>PRJNA470935</b>	SRR8195176	MRSA szczep USA300 LAC; 0.15µg/ml nafcylina; pożywka Cation Adjust Mueller Hinton	10
<b>PRJNA470935</b>	SRR8195182	MRSA szczep USA300 LAC; 0.15µg/ml nafcylina; pożywka Cation Adjust Mueller Hinton	10

<b>PRJNA470935</b>	SRR8195178	MRSA szczep USA300 LAC; 0.15µg/ml nafcylina; pożywka Cation Adjust Mueller Hinton	10
<b>PRJNA470935</b>	SRR8195195	MRSA szczep USA300 LAC; 0.15µg/ml nafcylina; pożywka Cation Adjust Mueller Hinton	10
<b>PRJNA470935</b>	SRR8195186	MRSA szczep USA300 LAC; 0.15µg/ml nafcylina; pożywka Cation Adjust Mueller Hinton	10
<b>PRJNA470935</b>	SRR8195185	MRSA szczep USA300 LAC; 0.15µg/ml nafcylina; pożywka Cation Adjust Mueller Hinton	10
<b>PRJNA470935</b>	SRR7146312	MRSA szczep USA300 LAC; 0.5µg/ml nafcylina; pożywka Cation Adjust Mueller Hinton	11
<b>PRJNA470935</b>	SRR7146319	MRSA szczep USA300 LAC; 0.5µg/ml nafcylina; pożywka Cation Adjust Mueller Hinton	11
<b>PRJNA470935</b>	SRR7146320	MRSA szczep USA300 LAC; 0.5µg/ml nafcylina; pożywka Cation Adjust Mueller Hinton	11
<b>PRJNA470935</b>	SRR8195177	MRSA szczep USA300 LAC; 0.5µg/ml nafcylina; pożywka Cation Adjust Mueller Hinton	11
<b>PRJNA470935</b>	SRR8195180	MRSA szczep USA300 LAC; 0.5µg/ml nafcylina; pożywka Cation Adjust Mueller Hinton	11
<b>PRJNA470935</b>	SRR8195179	MRSA szczep USA300 LAC; 0.5µg/ml nafcylina; pożywka Cation Adjust Mueller Hinton	11
<b>PRJNA470935</b>	SRR8195161	MRSA szczep USA300 LAC; 0.5µg/ml nafcylina; pożywka Cation Adjust Mueller Hinton	11
<b>PRJNA470935</b>	SRR8195162	MRSA szczep USA300 LAC; 0.5µg/ml nafcylina; pożywka Cation Adjust Mueller Hinton	11
<b>PRJNA470935</b>	SRR8195159	MRSA szczep USA300 LAC; 0.5µg/ml nafcylina; pożywka Cation Adjust Mueller Hinton	11
<b>PRJNA470935</b>	SRR7146300	MRSA szczep USA300 LAC; 1µg/ml nafcylina; pożywka Cation Adjust Mueller Hinton	12
<b>PRJNA470935</b>	SRR7146301	MRSA szczep USA300 LAC; 1µg/ml nafcylina; pożywka Cation Adjust Mueller Hinton	12
<b>PRJNA470935</b>	SRR7146311	MRSA szczep USA300 LAC; 1µg/ml nafcylina; pożywka Cation Adjust Mueller Hinton	12
<b>PRJNA470935</b>	SRR8195194	MRSA szczep USA300 LAC; 1µg/ml nafcylina; pożywka Cation Adjust Mueller Hinton	12
<b>PRJNA470935</b>	SRR8195193	MRSA szczep USA300 LAC; 1µg/ml nafcylina; pożywka Cation Adjust Mueller Hinton	12
<b>PRJNA470935</b>	SRR8195192	MRSA szczep USA300 LAC; 1µg/ml nafcylina; pożywka Cation Adjust Mueller Hinton	12
<b>PRJNA470935</b>	SRR8195160	MRSA szczep USA300 LAC; 1µg/ml nafcylina; pożywka Cation Adjust Mueller Hinton	12
<b>PRJNA470935</b>	SRR8195165	MRSA szczep USA300 LAC; 1µg/ml nafcylina; pożywka Cation Adjust Mueller Hinton	12
<b>PRJNA470935</b>	SRR8195166	MRSA szczep USA300 LAC; 1µg/ml nafcylina; pożywka Cation Adjust Mueller Hinton	12
<b>PRJNA470935</b>	SRR7146298	MRSA szczep USA300 LAC; brak antybiotyku; pożywka RPMI-1640, suplementowana 10% LB	13
<b>PRJNA470935</b>	SRR7146304	MRSA szczep USA300 LAC; brak antybiotyku; pożywka RPMI-1640, suplementowana 10% LB	13
<b>PRJNA470935</b>	SRR7146305	MRSA szczep USA300 LAC; brak antybiotyku; pożywka RPMI-1640, suplementowana 10% LB	13
<b>PRJNA470935</b>	SRR8195167	MRSA szczep USA300 LAC; brak antybiotyku; pożywka RPMI-1640, suplementowana 10% LB	13
<b>PRJNA470935</b>	SRR8195168	MRSA szczep USA300 LAC; brak antybiotyku; pożywka RPMI-1640, suplementowana 10% LB	13
<b>PRJNA470935</b>	SRR8195188	MRSA szczep USA300 LAC; brak antybiotyku; pożywka RPMI-1640, suplementowana 10% LB	13
<b>PRJNA470935</b>	SRR8195187	MRSA szczep USA300 LAC; brak antybiotyku; pożywka RPMI-1640, suplementowana 10% LB	13

<b>PRJNA470935</b>	SRR7146306	MRSA szczep USA300 LAC; 0.15µg/ml nafcylina; pożywka RPMI-1640, suplementowana 10% LB	14
<b>PRJNA470935</b>	SRR7146309	MRSA szczep USA300 LAC; 0.15µg/ml nafcyliny; pożywka RPMI-1640, suplementowana 10% LB	14
<b>PRJNA470935</b>	SRR7146310	MRSA szczep USA300 LAC; 0.15µg/ml nafcyliny; pożywka RPMI-1640, suplementowana 10% LB	14
<b>PRJNA470935</b>	SRR8195170	MRSA szczep USA300 LAC; 0.15µg/ml nafcyliny; pożywka RPMI-1640, suplementowana 10% LB	14
<b>PRJNA470935</b>	SRR8195169	MRSA szczep USA300 LAC; 0.15µg/ml nafcyliny; pożywka RPMI-1640, suplementowana 10% LB	14
<b>PRJNA470935</b>	SRR8195181	MRSA szczep USA300 LAC; 0.15µg/ml nafcyliny; pożywka RPMI-1640, suplementowana 10% LB	14
<b>PRJNA470935</b>	SRR8195175	MRSA szczep USA300 LAC; 0.15µg/ml nafcyliny; pożywka RPMI-1640, suplementowana 10% LB	14
<b>PRJNA470935</b>	SRR7146302	MRSA szczep USA300 LAC; 0.1µg/ml nafcyliny; pożywka RPMI-1640, suplementowana 10% LB	15
<b>PRJNA470935</b>	SRR7146303	MRSA szczep USA300 LAC; 0.1µg/ml nafcyliny; pożywka RPMI-1640, suplementowana 10% LB	15
<b>PRJNA470935</b>	SRR7146307	MRSA szczep USA300 LAC; 0.1µg/ml nafcyliny; pożywka RPMI-1640, suplementowana 10% LB	15
<b>PRJNA470935</b>	SRR8195184	MRSA szczep USA300 LAC; 0.1µg/ml nafcyliny; pożywka RPMI-1640, suplementowana 10% LB	15
<b>PRJNA470935</b>	SRR8195183	MRSA szczep USA300 LAC; 0.1µg/ml nafcyliny; pożywka RPMI-1640, suplementowana 10% LB	15
<b>PRJNA470935</b>	SRR8195190	MRSA szczep USA300 LAC; 0.1µg/ml nafcyliny; pożywka RPMI-1640, suplementowana 10% LB	15
<b>PRJNA470935</b>	SRR8195189	MRSA szczep USA300 LAC; 0.1µg/ml nafcyliny; pożywka RPMI-1640, suplementowana 10% LB	15
<b>PRJNA470935</b>	SRR6391835	MRSA szczep USA300 TCH1516; brak antybiotyku; pożywka Cation Adjust Mueller Hinton	16
<b>PRJNA470935</b>	SRR6391850	MRSA szczep USA300 TCH1516; brak antybiotyku; pożywka Cation Adjust Mueller Hinton	16
<b>PRJNA470935</b>	SRR6391840	MRSA szczep USA300 TCH1516; 0.063µg/ml nafcyliny; pożywka Cation Adjust Mueller Hinton	17
<b>PRJNA470935</b>	SRR6391842	MRSA szczep USA300 TCH1516; 0.063µg/ml nafcyliny; pożywka Cation Adjust Mueller Hinton	17
<b>PRJNA470935</b>	SRR6391833	MRSA szczep USA300 TCH1516; 0.5µg/ml nafcyliny; pożywka Cation Adjust Mueller Hinton	18
<b>PRJNA470935</b>	SRR6391839	MRSA szczep USA300 TCH1516; 0.5µg/ml nafcyliny; pożywka Cation Adjust Mueller Hinton	18
<b>PRJNA470935</b>	SRR6391838	MRSA szczep USA300 TCH1516; brak antybiotyku; pożywka RPMI-1640, suplementowana 10% LB	19
<b>PRJNA470935</b>	SRR6391837	MRSA szczep USA300 TCH1516; 0.031µg/ml nafcyliny; pożywka RPMI-1640, suplementowana 10% LB	20
<b>PRJNA470935</b>	SRR6391847	MRSA szczep USA300 TCH1516; 0.031µg/ml nafcyliny; pożywka RPMI-1640, suplementowana 10% LB	20
<b>PRJNA470935</b>	SRR6391841	MRSA szczep USA300 TCH1516; 0.031µg/ml nafcyliny; pożywka RPMI-1640, suplementowana 10% LB	20
<b>PRJNA470935</b>	SRR6391836	MRSA szczep USA300 TCH1516; 0.063µg/ml nafcyliny; pożywka RPMI-1640, suplementowana 10% LB	21



<b>PRJNA470935</b>	SRR6391848	MRSA szczep USA300 TCH1516; 0.063µg/ml nafcyliny; pożywka RPMI-1640, suplementowana 10% LB	21
<b>PRJNA526539</b>	SRR8710732	MRSA szczep USA300 LAC; brak antybiotyku; pożywka Cation Adjust Mueller Hinton	22
<b>PRJNA526539</b>	SRR8710737	MRSA szczep USA300 LAC; brak antybiotyku; pożywka Cation Adjust Mueller Hinton	22
<b>PRJNA526539</b>	SRR8710735	MRSA szczep USA300 LAC; 0.65µg/ml wankomycyny; pożywka Cation Adjust Mueller Hinton	23
<b>PRJNA526539</b>	SRR8710736	MRSA szczep USA300 LAC; 0.65µg/ml wankomycyny; pożywka Cation Adjust Mueller Hinton	23
<b>PRJNA526539</b>	SRR8710731	MRSA szczep USA300 LAC; 0.7µg/ml wankomycyny; pożywka Cation Adjust Mueller Hinton	24
<b>PRJNA526539</b>	SRR8710734	MRSA szczep USA300 LAC; 0.7µg/ml wankomycyny; pożywka Cation Adjust Mueller Hinton	24
<b>PRJNA526539</b>	SRR8710730	MRSA szczep USA300 LAC; 0.8µg/ml wankomycyny; pożywka Cation Adjust Mueller Hinton	25
<b>PRJNA526539</b>	SRR8710733	MRSA szczep USA300 LAC; 0.8µg/ml wankomycyny; pożywka Cation Adjust Mueller Hinton	25
<b>PRJNA526539</b>	SRR8710726	MRSA szczep USA300 LAC; brak antybiotyku; pożywka RPMI-1640, suplementowana 10% LB	29
<b>PRJNA526539</b>	SRR8710739	MRSA szczep USA300 LAC; brak antybiotyku; pożywka RPMI-1640, suplementowana 10% LB	29
<b>PRJNA526539</b>	SRR8710729	MRSA szczep USA300 LAC; 0.6µg/ml wankomycyny; pożywka RPMI-1640, suplementowana 10% LB	30
<b>PRJNA526539</b>	SRR8710738	MRSA szczep USA300 LAC; 0.6µg/ml wankomycyny; pożywka RPMI-1640, suplementowana 10% LB	30
<b>PRJNA526539</b>	SRR8710727	MRSA szczep USA300 LAC; 0.8µg/ml wankomycyny; pożywka RPMI-1640, suplementowana 10% LB	31
<b>PRJNA526539</b>	SRR8710728	MRSA szczep USA300 LAC; 0.8µg/ml wankomycyny; pożywka RPMI-1640, suplementowana 10% LB	31
<b>PRJNA638628</b>	SRR11976310	MRSA szczep USA100 D592; brak antybiotyku; pożywka Cation Adjust Mueller Hinton	32
<b>PRJNA638628</b>	SRR11976321	MRSA szczep USA100 D592; brak antybiotyku; pożywka Cation Adjust Mueller Hinton	32
<b>PRJNA638628</b>	SRR11976306	MRSA szczep USA100 D592; 0.55µg/ml wankomycyny; pożywka Cation Adjust Mueller Hinton	33
<b>PRJNA638628</b>	SRR11976309	MRSA szczep USA100 D592; 0.55µg/ml wankomycyny; pożywka Cation Adjust Mueller Hinton	33
<b>PRJNA638628</b>	SRR11976305	MRSA szczep USA100 D592; 0.9µg/ml wankomycyny; pożywka Cation Adjust Mueller Hinton	34
<b>PRJNA638628</b>	SRR11976308	MRSA szczep USA100 D592; 0.9µg/ml wankomycyny; pożywka Cation Adjust Mueller Hinton	34

<b>PRJNA638628</b>	SRR11976304	MRSA szczep USA100 D592; 1µg/ml wankomycyny; pożywka Cation Adjust Mueller Hinton	35
<b>PRJNA638628</b>	SRR11976307	MRSA szczep USA100 D592; 1µg/ml wankomycyny; pożywka Cation Adjust Mueller Hinton	35
<b>PRJNA638628</b>	SRR11976328	MRSA szczep USA100 D592; brak antybiotyku; pożywka RPMI-1640, suplementowana 10% LB	36
<b>PRJNA638628</b>	SRR11976331	MRSA szczep USA100 D592; brak antybiotyku; pożywka RPMI-1640, suplementowana 10% LB	36
<b>PRJNA638628</b>	SRR11976325	MRSA szczep USA100 D592; 0.55µg/ml wankomycyny; pożywka RPMI-1640, suplementowana 10% LB	37
<b>PRJNA638628</b>	SRR11976330	MRSA szczep USA100 D592; 0.55µg/ml wankomycyny; pożywka RPMI-1640, suplementowana 10% LB	37
<b>PRJNA638628</b>	SRR11976327	MRSA szczep USA100 D592; 0.625µg/ml wankomycyny; pożywka RPMI-1640, suplementowana 10% LB	38
<b>PRJNA638628</b>	SRR11976329	MRSA szczep USA100 D592; 0.625µg/ml wankomycyny; pożywka RPMI-1640, suplementowana 10% LB	38
<b>PRJNA638628</b>	SRR11976320	MRSA szczep USA100 D712; brak antybiotyku; pożywka Cation Adjust Mueller Hinton	39
<b>PRJNA638628</b>	SRR11976324	MRSA szczep USA100 D712; brak antybiotyku; pożywka Cation Adjust Mueller Hinton	39
<b>PRJNA638628</b>	SRR11976319	MRSA szczep USA100 D712; 0.8µg/ml wankomycyny; pożywka Cation Adjust Mueller Hinton	40
<b>PRJNA638628</b>	SRR11976323	MRSA szczep USA100 D712; 0.8µg/ml wankomycyny; pożywka Cation Adjust Mueller Hinton	40
<b>PRJNA638628</b>	SRR11976316	MRSA szczep USA100 D712; 1.4µg/ml wankomycyny; pożywka Cation Adjust Mueller Hinton	41
<b>PRJNA638628</b>	SRR11976322	MRSA szczep USA100 D712; 1.4µg/ml wankomycyny; pożywka Cation Adjust Mueller Hinton	41
<b>PRJNA638628</b>	SRR11976317	MRSA szczep USA100 D712; 1µg/ml wankomycyny; pożywka Cation Adjust Mueller Hinton	42
<b>PRJNA638628</b>	SRR11976326	MRSA szczep USA100 D712; 1µg/ml wankomycyny; pożywka Cation Adjust Mueller Hinton	42
<b>PRJNA638628</b>	SRR11976313	MRSA szczep USA100 D712; brak antybiotyku; pożywka RPMI-1640, suplementowana 10% LB	43
<b>PRJNA638628</b>	SRR11976315	MRSA szczep USA100 D712; brak antybiotyku; pożywka RPMI-1640, suplementowana 10% LB	43
<b>PRJNA638628</b>	SRR11976312	MRSA szczep USA100 D712; 0.625µg/ml wankomycyny; pożywka RPMI-1640, suplementowana 10% LB	44
<b>PRJNA638628</b>	SRR11976318	MRSA szczep USA100 D712; 0.625µg/ml wankomycyny; pożywka RPMI-1640, suplementowana 10% LB	44
<b>PRJNA638628</b>	SRR11976311	MRSA szczep USA100 D712; 0.8µg/ml wankomycyny; pożywka RPMI-1640, suplementowana 10% LB	45

<b>PRJNA638628</b>	SRR11976314	MRSA szczep USA100 D712; 0.8µg/ml wankomycyny; pożywka RPMI-1640, suplementowana 10% LB	45
<b>GSM5799537</b>	SRR17535775	MRSA szczep USA100; brak antybiotyku; pożywka Cation Adjust Mueller Hinton	46
<b>GSM5799537</b>	SRR17535779	MRSA szczep USA100; brak antybiotyku; pożywka Cation Adjust Mueller Hinton	46
<b>GSM5799537</b>	SRR17535783	MRSA szczep USA100; brak antybiotyku; pożywka Cation Adjust Mueller Hinton	46
<b>GSM5799537</b>	SRR17535774	MRSA szczep USA100; 4µg/ml oksacyliny; pożywka Cation Adjust Mueller Hinton	47
<b>GSM5799537</b>	SRR17535778	MRSA szczep USA100; 4µg/ml oksacyliny; pożywka Cation Adjust Mueller Hinton	47
<b>GSM5799537</b>	SRR17535782	MRSA szczep USA100; 4µg/ml oksacyliny; pożywka Cation Adjust Mueller Hinton	47
<b>PRJEB23980</b>	SRR5432011	MRSA szczep USA300 TCH1516; 10µg/ml nafcyлина; pożywka Todd Hewitt Broth (Hardy Diagnostics)	48
<b>PRJEB23980</b>	SRR5432012	MRSA szczep USA300 TCH1516; 10µg/ml nafcyлина; pożywka Todd Hewitt Broth (Hardy Diagnostics)	48
<b>PRJEB23980</b>	SRR5432013	MRSA szczep USA300 TCH1516; 10µg/ml nafcyлина; pożywka Todd Hewitt Broth (Hardy Diagnostics)	48

## Załącznik 2

*Tabela 2. Charakterystyka rozkładów minimalnej energii swobodnej struktur spinek terminacyjnych, zidentyfikowanych dla poszczególnych gatunków bakterii przy pomocy protokołu opartego o program RNAfold.*

	obs	Q1	Q2	Q3	min	maks	średnia	wariancja	skośność	kurtoza
<b>S. aureus</b>	725	-17.9	-14.8	-12.4	-24.9	-3.6	-15.0	15.09	0.01	-0.33
<b>E. coli (a)</b>	650	-18.3	-15.2	-12.2	-28.9	-4.9	-15.29	17.65	-0.22	-0.05
<b>E. coli (b)</b>	904	-18.6	-15.2	-12.08	-28.5	-3.9	-15.36	18.74	-0.22	-0.44
<b>B. subtilis (a)</b>	564	-16.3	-13.5	-11.4	-27.9	-5.6	-13.94	14.19	-0.55	0.32
<b>B. subtilis (b)</b>	1005	-16.5	-13.4	-11.0	-27.9	-4.0	-13.77	15.84	-0.36	0.15
<b>B. subtilis (c)</b>	867	-16.5	-13.7	-11.5	-27.9	-4.1	-13.99	14.36	-0.43	0.25
<b>B. subtilis (d)</b>	1053	-16.4	-13.4	-11.2	-27.9	-4.0	-13.8	15.17	-0.39	0.21
<b>E. faecalis</b>	673	-20.5	-16.3	-13.1	-34.9	-3.6	-16.94	28.0	-0.36	-0.13
<b>L. monocytogenes</b>	754	-17.0	-14.1	-11.7	-26.2	-3.3	-14.23	15.25	-0.12	-0.09
<b>Z. mobilis</b>	189	-18.8	-16.8	-14.7	-28.4	-7.2	-16.89	11.54	-0.29	0.55
<b>Synechocystis sp.</b>	154	-15.58	-13.3	-10.6	-23.8	-5.3	-13.24	13.01	-0.27	-0.08
<b>S. coelicolor</b>	607	-29.45	-23.7	-14.7	-43.0	-4.2	-22.91	80.12	-0.09	-0.99
<b>S. venezuelae</b>	548	-30.53	-25.95	-17.9	-43.4	-3.4	-24.45	74.69	0.29	-0.73
<b>S. tsukubensis</b>	499	-31.65	-25.3	-16.3	-43.0	-3.2	-24.15	84.63	0.2	-1.03
<b>S. griseus</b>	701	-33.1	-25.9	-15.8	-44.7	-6.1	-24.91	94.04	0.1	-1.18
<b>S. lividans</b>	471	-28.65	-23.6	-15.45	-42.1	-7.4	-22.68	68.33	0.03	-0.92
<b>S. clavuligerus</b>	414	-30.1	-25.5	-16.97	-44.1	-3.4	-23.78	72.95	0.26	-0.75
<b>S. avermitilis</b>	667	-30.5	-25.0	-17.0	-42.4	-3.4	-23.82	73.72	0.19	-0.91

### **Objaśnienia kolumn:**

*obs* – liczba obserwacji/terminatorów

*Q1* – pierwszy kwartyl

*Q2* – drugi kwartyl/mediana

*Q3* – trzeci kwartyl

*min* – wartość minimalna

*maks* – wartość maksymalna

*średnia* – wartość średnia

*wariancja* – wariancja

*skośność* – skośność rozkładu

*kurtoza* – kurtoza rozkładu

### Załącznik 3

**Tabela 3.** Charakterystyka rozkładów wartości `hairpin score` obliczanej dla spinek terminacyjnych wyznaczonych przez TransTermHP w ramach algorytmu zaimplementowanego w programie TERMIte. Wartość `hairpin score` stanowi heurystyczny odpowiednik minimalnej energii swobodnej struktur spinek terminacyjnych. Charakterystyki przedstawiono dla wszystkich analizowanych gatunków bakterii.

	obs	Q1	Q2	Q3	min	maks	średnia	wariancja	skośność	kurtoza
<b>S. aureus</b>	685	-12.2	-9.7	-7.6	-19.7	-2.2	-9.98	11.13	-0.42	-0.17
<b>E. coli (a)</b>	530	-14.08	-12.4	-10.4	-21.5	-3.0	-12.28	8.35	-0.01	0.36
<b>E. coli (b)</b>	661	-14.0	-12.2	-10.3	-21.5	-3.0	-12.13	8.85	-0.02	0.16
<b>B. subtilis (a)</b>	569	-13.2	-11.2	-9.0	-21.3	-2.6	-11.1	10.26	-0.1	0.07
<b>B. subtilis (b)</b>	1039	-13.2	-11.1	-8.8	-20.9	-2.3	-11.03	11.03	-0.08	-0.08
<b>B. subtilis (c)</b>	885	-13.3	-11.3	-9.1	-22.4	-2.4	-11.28	10.77	-0.09	0.14
<b>B. subtilis (d)</b>	1080	-13.2	-11.2	-8.9	-21.3	-2.4	-11.11	10.8	-0.08	-0.04
<b>E. faecalis</b>	666	-16.3	-12.7	-9.5	-29.3	-2.4	-12.96	24.53	-0.26	-0.37
<b>L. monocytogenes</b>	797	-12.9	-10.4	-8.1	-24.6	-2.2	-10.59	12.96	-0.27	0.01
<b>Z. mobilis</b>	165	-14.7	-13.2	-11.1	-19.8	-2.9	-12.95	8.75	0.04	0.28
<b>Synechocystis sp.</b>	119	-9.8	-7.9	-5.45	-14.6	-2.1	-7.72	8.42	-0.09	-0.72
<b>S. coelicolor</b>	222	-24.9	- 21.45	-18.3	-40.9	-8.1	-21.8	29.07	-0.4	0.75
<b>S. venezuelae</b>	243	-24.95	-21.4	-18.5	-37.8	-3.1	-21.38	28.85	0.31	0.56
<b>S. tsukubensis</b>	219	-24.95	-21.4	-17.8	-37.4	-5.6	-21.41	31.81	0.04	0.14
<b>S. griseus</b>	280	-27.5	- 22.95	-19.2	-42.7	-7.4	-23.47	31.9	-0.11	-0.01
<b>S. lividans</b>	195	-23.75	-20.8	-17.9	-32.4	-8.1	-20.91	21.74	-0.01	0.27
<b>S. clavuligerus</b>	217	-24.0	-20.7	-18.1	-35.2	-2.1	-20.78	26.87	0.5	1.49
<b>S. avermitilis</b>	325	-24.0	-20.8	-17.6	-35.1	-4.4	-20.91	26.11	0.02	-0.03

#### Objaśnienia kolumn:

*obs* – liczba obserwacji/terminatorów

*Q1* – pierwszy kwartyl

*Q2* – drugi kwartyl/mediana

*Q3* – trzeci kwartyl

*min* – wartość minimalna

*maks* – wartość maksymalna

*średnia* – wartość średnia

*wariancja* – wariancja

*skośność* – skośność rozkładu

*kurtoza* – kurtoza rozkładu

## Załącznik 4

Tabela 4. Charakterystyka rozkładów długości spinek terminacyjnych zidentyfikowanych dla poszczególnych gatunków bakterii.

	obs	Q1	Q2	Q3	min	maks	średnia	wariancja	skośność	kurtoza
<b>S. aureus</b>	838	29.0	35.0	40.0	12	59	34.74	73.07	0.14	-0.09
<b>E. coli (a)</b>	691	20.0	24.0	28.5	13	56	24.6	44.97	0.97	1.45
<b>E. coli (b)</b>	957	20.0	25.0	30.0	13	51	25.47	45.97	0.6	0.08
<b>B. subtilis (a)</b>	635	21.0	24.0	27.5	11	48	24.97	31.08	0.97	1.21
<b>B. subtilis (b)</b>	1165	21.0	24.0	28.0	11	54	25.12	33.96	0.91	1.16
<b>B. subtilis (c)</b>	984	21.0	24.0	28.0	11	49	24.97	30.97	1.0	1.43
<b>B. subtilis (d)</b>	1214	21.0	24.0	28.0	11	54	25.1	34.09	1.02	1.66
<b>E. faecalis</b>	796	27.0	33.0	38.0	12	59	32.86	58.21	0.17	-0.19
<b>L. monocytogenes</b>	862	24.0	28.0	33.0	11	57	28.66	48.13	0.76	0.74
<b>Z. mobilis</b>	206	24.0	27.0	32.75	15	59	29.0	54.92	1.31	2.41
<b>Synechocystis sp.</b>	165	23.0	27.0	32.0	12	59	27.52	61.01	0.87	1.52
<b>S. coelicolor</b>	629	28.0	33.0	38.0	14	58	32.83	44.99	0.05	0.38
<b>S. venezuelae</b>	573	28.0	33.0	39.0	14	59	33.36	54.89	0.26	0.79
<b>S. tsukubensis</b>	520	29.0	34.0	39.0	16	57	33.84	49.51	0.0	0.34
<b>S. griseus</b>	737	29.0	34.0	38.0	14	57	33.4	44.89	-0.1	0.62
<b>S. lividans</b>	495	28.0	32.0	37.0	16	58	32.38	46.49	0.12	0.28
<b>S. clavuligerus</b>	435	29.0	35.0	39.0	13	59	34.37	53.36	0.14	0.26
<b>S. avermitilis</b>	705	28.0	33.0	38.0	16	59	33.01	47.06	0.29	0.63

### Objaśnienia kolumn:

*obs* – liczba obserwacji/terminatorów

*Q1* – pierwszy kwartyl

*Q2* – drugi kwartyl/mediana

*Q3* – trzeci kwartyl

*min* – wartość minimalna

*maks* – wartość maksymalna

*średnia* – wartość średnia

*wariancja* – wariancja

*skośność* – skośność rozkładu

*kurtoza* – kurtoza rozkładu

## Załącznik 5

*Tabela 5. Charakterystyka rozkładów długości pętli w ramach spinek terminacyjnych zidentyfikowanych dla poszczególnych gatunków bakterii.*

	obs	Q1	Q2	Q3	min	maks	średnia	wariancja	skośność	kurtoza
<i>S. aureus</i>	685	4.0	4.0	6.0	3	12	4.96	3.88	1.3	1.35
<i>E. coli (a)</i>	530	4.0	4.0	5.0	3	11	4.4	1.47	1.46	3.12
<i>E. coli (b)</i>	661	4.0	4.0	5.0	3	11	4.38	1.53	1.51	3.24
<i>B. subtilis (a)</i>	569	4.0	4.0	5.0	3	11	4.61	2.27	1.17	1.35
<i>B. subtilis (b)</i>	1039	3.0	4.0	5.0	3	12	4.58	2.37	1.42	2.61
<i>B. subtilis (c)</i>	885	3.0	4.0	5.0	3	12	4.59	2.29	1.29	2.12
<i>B. subtilis (d)</i>	1080	3.0	4.0	5.0	3	12	4.59	2.4	1.4	2.56
<i>E. faecalis</i>	666	4.0	4.0	6.0	3	13	4.87	2.93	1.4	2.81
<i>L. monocytogenes</i>	797	4.0	4.0	6.0	3	12	4.74	2.48	1.29	2.26
<i>Z. mobilis</i>	165	4.0	4.0	5.0	3	10	4.45	1.75	1.06	1.18
<i>Synechocystis sp.</i>	119	3.0	4.0	5.5	3	10	4.61	2.37	1.08	0.92
<i>S. coelicolor</i>	222	4.0	4.0	5.0	3	11	4.56	2.13	1.45	2.62
<i>S. venezuelae</i>	243	4.0	4.0	5.0	3	10	4.35	1.53	1.62	3.45
<i>S. tsukubensis</i>	219	4.0	4.0	5.0	3	12	4.55	2.35	1.67	3.78
<i>S. griseus</i>	280	4.0	4.0	5.0	3	13	4.55	2.31	2.03	6.1
<i>S. lividans</i>	195	4.0	4.0	5.0	3	9	4.5	1.78	1.25	1.59
<i>S. clavuligerus</i>	217	4.0	4.0	5.0	3	10	4.48	2.19	1.66	3.22
<i>S. avermitilis</i>	325	4.0	4.0	5.0	3	12	4.43	2.27	2.09	5.75

### **Objaśnienia kolumn:**

*obs* – liczba obserwacji/terminatorów

*Q1* – pierwszy kwartyl

*Q2* – drugi kwartyl/mediana

*Q3* – trzeci kwartyl

*min* – wartość minimalna

*maks* – wartość maksymalna

*średnia* – wartość średnia

*wariancja* – wariancja

*skośność* – skośność rozkładu

*kurtoza* – kurtoza rozkładu

## Załącznik 6

*Tabela 6. Charakterystyka rozkładów wartości tail score obliczonych przez program TransTermHP dla rho-niezależnych terminatorów zidentyfikowanych u poszczególnych gatunków bakterii.*

	obs	Q1	Q2	Q3	min	maks	średnia	wariancja	skośność	kurtoza
<i>S. aureus</i>	685	-5.5	-5.2	-4.7	-6.7	-2.5	-5.02	0.5	1.05	1.13
<i>E. coli (a)</i>	530	-5.5	-4.8	-3.9	-6.4	-2.5	-4.65	0.87	0.39	-0.9
<i>E. coli (b)</i>	661	-5.4	-4.7	-3.8	-6.4	-2.5	-4.58	0.91	0.28	-1.04
<i>B. subtilis (a)</i>	569	-5.6	-5.3	-4.7	-6.4	-2.7	-5.12	0.54	1.16	0.99
<i>B. subtilis (b)</i>	1039	-5.6	-5.3	-4.6	-6.4	-2.5	-5.05	0.63	1.06	0.54
<i>B. subtilis (c)</i>	885	-5.6	-5.3	-4.7	-7.1	-2.6	-5.11	0.56	1.13	0.97
<i>B. subtilis (d)</i>	1080	-5.6	-5.3	-4.6	-6.4	-2.5	-5.04	0.63	1.05	0.51
<i>E. faecalis</i>	666	-5.7	-5.3	-4.7	-7.0	-2.5	-5.13	0.63	1.0	0.64
<i>L. monocytogenes</i>	797	-5.7	-5.5	-5.1	-6.5	-2.9	-5.31	0.37	1.37	2.27
<i>Z. mobilis</i>	165	-5.6	-5.3	-4.4	-6.2	-2.7	-4.95	0.81	0.97	-0.18
<i>Synechocystis sp.</i>	119	-5.6	-5.3	-4.9	-6.2	-3.1	-5.18	0.37	1.06	0.86
<i>S. coelicolor</i>	222	-3.8	-3.5	-3.1	-5.5	-2.5	-3.52	0.35	-0.52	0.22
<i>S. venezuelae</i>	243	-4.1	-3.6	-3.1	-5.5	-2.5	-3.59	0.43	-0.57	-0.09
<i>S. tsukubensis</i>	219	-4.0	-3.5	-3.1	-5.8	-2.5	-3.56	0.47	-0.86	0.58
<i>S. griseus</i>	280	-3.7	-3.4	-2.9	-5.5	-2.5	-3.43	0.38	-0.77	0.37
<i>S. lividans</i>	195	-3.8	-3.5	-3.1	-5.6	-2.5	-3.55	0.4	-0.65	0.35
<i>S. clavuligerus</i>	217	-3.9	-3.5	-3.1	-5.8	-2.5	-3.57	0.5	-1.01	0.7
<i>S. avermitilis</i>	325	-3.9	-3.4	-3.1	-5.9	-2.5	-3.54	0.44	-0.76	0.38

### **Objaśnienia kolumn:**

*obs* – liczba obserwacji/terminatorów

*Q1* – pierwszy kwartyl

*Q2* – drugi kwartyl/mediana

*Q3* – trzeci kwartyl

*min* – wartość minimalna

*maks* – wartość maksymalna

*średnia* – wartość średnia

*wariancja* – wariancja

*skośność* – skośność rozkładu

*kurtoza* – kurtoza rozkładu



## Załącznik 7

Tabela 7. Charakterystyka rozkładów odległości pomiędzy referencyjnym punktem terminacji a najbliższym genem zlokalizowanym poniżej w obrębie tej samej nici.

	obs	Q1	Q2	Q3	min	maks	średnia	wariancja	skośność	kurtoza
<i>S. aureus</i>	836	176.75	491.0	1959.25	1.0	54666.0	2161.02	23115190.68	5.1	34.39
<i>E. coli (a)</i>	691	82.0	264.0	1633.0	3.0	40597.0	1469.56	8905260.06	5.59	52.15
<i>E. coli (b)</i>	957	105.0	375.0	1812.0	2.0	40598.0	1704.31	10829139.36	4.69	33.97
<i>B. subtilis (a)</i>	635	82.0	211.0	1305.0	11.0	58833.0	1671.35	20704051.81	7.27	72.94
<i>B. subtilis (b)</i>	1165	95.0	353.0	1697.0	11.0	76739.0	1833.31	21231131.38	8.1	98.32
<i>B. subtilis (c)</i>	984	92.75	325.5	1683.0	11.0	76739.0	1882.39	23045489.27	8.14	97.11
<i>B. subtilis (d)</i>	1214	95.0	351.5	1759.0	11.0	76739.0	1864.88	21777302.75	7.76	90.79
<i>E. faecalis</i>	794	95.0	282.0	2079.0	2.0	30260.0	2155.74	20104345.95	3.56	14.12
<i>L. monocytogenes</i>	862	80.0	234.0	1802.5	7.0	77755.0	2303.26	34549442.77	6.29	56.09
<i>Z. mobilis</i>	206	137.25	425.0	2058.25	5.0	14228.0	1707.09	6762451.13	2.18	4.72
<i>Synechocystis sp.</i>	165	80.0	177.0	1325.0	1.0	8612.0	1125.82	3665028.21	2.35	5.08
<i>S. coelicolor</i>	629	142.0	732.0	2302.0	1.0	31398.0	1985.13	10383444.1	3.4	17.33
<i>S. venezuelae</i>	573	84.0	450.0	1834.0	4.0	31480.0	1664.19	9388157.6	3.96	22.85
<i>S. tsukubensis</i>	519	148.0	558.0	2330.5	6.0	71482.0	2139.07	21523262.55	7.8	98.54
<i>S. griseus</i>	737	113.0	562.0	1918.0	1.0	68511.0	1907.85	18322104.81	8.62	111.62
<i>S. lividans</i>	495	100.5	357.0	1784.0	4.0	42700.0	1671.99	11546888.55	5.53	48.7
<i>S. clavuligerus</i>	435	141.0	484.0	1967.0	4.0	23628.0	1724.57	7674920.0	2.91	12.09
<i>S. avermitilis</i>	705	108.0	534.0	1927.0	1.0	78901.0	1751.21	16617207.04	11.07	186.94

### Objaśnienia kolumn:

*obs* – liczba obserwacji/terminatorów

*Q1* – pierwszy kwartyl

*Q2* – drugi kwartyl/mediana

*Q3* – trzeci kwartyl

*min* – wartość minimalna

*maks* – wartość maksymalna

*średnia* – wartość średnia

*wariancja* – wariancja

*skośność* – skośność rozkładu

*kurtoza* – kurtoza rozkładu

## Załącznik 8

*Tabela 8. Charakterystyka rozkładów odległości pomiędzy referencyjnym punktem terminacji a najbliższym genem zlokalizowanym powyżej w obrębie tej samej nici.*

	obs	Q1	Q2	Q3	min	maks	średnia	wariancja	skośność	kurtoza
<i>S. aureus</i>	838	50.0	64.0	93.0	1.0	200.0	76.46	1455.07	1.19	1.02
<i>E. coli (a)</i>	691	38.0	48.0	69.5	2.0	199.0	62.76	1492.51	1.83	2.65
<i>E. coli (b)</i>	957	38.0	48.0	73.0	1.0	200.0	63.9	1640.91	1.7	2.12
<i>B. subtilis (a)</i>	635	33.0	38.0	46.0	6.0	193.0	46.28	814.43	2.93	9.56
<i>B. subtilis (b)</i>	1165	32.0	38.0	45.0	8.0	196.0	45.03	797.73	2.82	9.19
<i>B. subtilis (c)</i>	984	32.0	38.0	44.0	8.0	196.0	44.13	750.1	3.03	10.56
<i>B. subtilis (d)</i>	1214	31.0	38.0	45.0	8.0	196.0	44.75	824.32	2.81	9.02
<i>E. faecalis</i>	796	42.0	51.0	71.0	1.0	200.0	62.76	1216.6	1.77	3.12
<i>L. monocytogenes</i>	862	37.0	41.0	49.0	1.0	199.0	47.73	676.87	3.29	13.16
<i>Z. mobilis</i>	206	49.0	57.0	70.0	4.0	198.0	66.34	918.05	1.96	4.22
<i>Synechocystis sp.</i>	165	68.0	91.0	126.0	1.0	200.0	98.7	1568.61	0.52	-0.02
<i>S. coelicolor</i>	629	56.0	76.0	118.0	4.0	200.0	90.29	2121.29	0.78	-0.42
<i>S. venezuelae</i>	573	53.0	67.0	100.0	1.0	200.0	80.66	1658.47	1.08	0.61
<i>S. tsukubensis</i>	520	52.0	67.0	100.25	1.0	198.0	79.47	1719.81	0.99	0.33
<i>S. griseus</i>	737	56.0	69.0	97.0	1.0	200.0	80.42	1494.59	1.04	0.69
<i>S. lividans</i>	495	52.5	66.0	99.0	7.0	200.0	81.11	1794.79	1.12	0.6
<i>S. clavuligerus</i>	435	54.0	70.0	111.0	2.0	200.0	84.89	1943.18	0.93	0.1
<i>S. avermitilis</i>	705	52.0	67.0	99.0	1.0	200.0	79.9	1637.05	1.1	0.63

### **Objaśnienia kolumn:**

*obs* – liczba obserwacji/terminatorów

*Q1* – pierwszy kwartyl

*Q2* – drugi kwartyl/mediana

*Q3* – trzeci kwartyl

*min* – wartość minimalna

*maks* – wartość maksymalna

*średnia* – wartość średnia

*wariancja* – wariancja

*skośność* – skośność rozkładu

*kurtoza* – kurtoza rozkładu

## Załącznik 9

**Tabela 9.** Porównanie rozkładów analizowanych cech terminatorów rho-niezależnych dla każdej możliwej pary gatunków/eksperymentów. Porównania wykonane zostały przy pomocy testu U Manna-Whitneya. Testowana hipoteza zerowa zakłada brak przesunięcia porównywanych rozkładów analizowanych cech pomiędzy każdą możliwą parą analizowanych gatunków/eksperymentów.

1. [https://zenodo.org/records/11392921/files/hairpin\\_energies.csv?download=1](https://zenodo.org/records/11392921/files/hairpin_energies.csv?download=1) (porównania rozkładów minimalnej energii swobodnej struktury spinki terminacyjnej)
2. [https://zenodo.org/records/11392921/files/hairpin\\_scores.csv?download=1](https://zenodo.org/records/11392921/files/hairpin_scores.csv?download=1) (porównania rozkładów wartości hairpin score)
3. [https://zenodo.org/records/11392921/files/hairpin\\_lengths.csv?download=1](https://zenodo.org/records/11392921/files/hairpin_lengths.csv?download=1) (porównania rozkładów długości spinki terminacyjnej)
4. [https://zenodo.org/records/11392921/files/loop\\_lengths.csv?download=1](https://zenodo.org/records/11392921/files/loop_lengths.csv?download=1) (porównania rozkładów długości pętli wchodzącej w skład spinki terminacyjnej)
5. [https://zenodo.org/records/11392921/files/tail\\_scores.csv?download=1](https://zenodo.org/records/11392921/files/tail_scores.csv?download=1) (porównania rozkładów wartości tail score)
6. [https://zenodo.org/records/11392921/files/distances\\_to\\_downstream\\_genes.csv?download=1](https://zenodo.org/records/11392921/files/distances_to_downstream_genes.csv?download=1) (porównania rozkładów odległości pomiędzy referencyjnym punktem terminacji a najbliższym genem zlokalizowanym poniżej w obrębie tej samej nici)
7. [https://zenodo.org/records/11392921/files/distances\\_to\\_upstream\\_genes.csv?download=1](https://zenodo.org/records/11392921/files/distances_to_upstream_genes.csv?download=1) (porównania rozkładów odległości pomiędzy referencyjnym punktem terminacji a najbliższym genem zlokalizowanym powyżej w obrębie tej samej nici)
8. [https://zenodo.org/records/11392921/files/hairpin\\_energies-column-description.txt?download=1](https://zenodo.org/records/11392921/files/hairpin_energies-column-description.txt?download=1) (objaśnienia kolumn)

DOI: [10.5281/zenodo.11392921](https://doi.org/10.5281/zenodo.11392921) (Kosinski, 2024)

## Załącznik 10

**Tabela 10.** Wykaz publicznie dostępnych zbiorów danych wykorzystanych do wytrenowania i ewaluacji klasyfikatora programu termiRNator. W skład zbior testowego wchodzi próbki pochodzące z *L. monocytogenes*, uczącego natomiast pozostałe. Identyfikatory próbek pochodzą z bazy danych Sequence Read Archive (<https://www.ncbi.nlm.nih.gov/sra/>). W tabeli wyszczególnione są próby pochodzące z sekwencjonowania 3' końców RNA metodą Term-seq, oraz uzyskane w ramach tego samego warunku eksperymentalnego próby RNA-seq

Numer próbki	Term-seq	RNA-seq	Gatunek
1	SRR17335818	SRR17335818	<i>B. subtilis</i>
2	SRR17335819	SRR17335819	<i>B. subtilis</i>
3	SRR17335820	SRR17335820	<i>B. subtilis</i>
4	SRR17335821	SRR17335821	<i>B. subtilis</i>
5	SRR17335822	SRR17335822	<i>B. subtilis</i>
6	SRR17335823	SRR17335823	<i>B. subtilis</i>
7	SRR17335824	SRR17335824	<i>B. subtilis</i>
8	SRR17335825	SRR17335825	<i>B. subtilis</i>
9	SRR17335826	SRR17335826	<i>B. subtilis</i>
10	SRR17335827	SRR17335827	<i>B. subtilis</i>
11	SRR17335828	SRR17335828	<i>B. subtilis</i>
12	SRR17335829	SRR17335829	<i>B. subtilis</i>
13	ERR1248362	ERR1248364	<i>B. subtilis</i>
14	ERR1248363	ERR1248364	<i>B. subtilis</i>
15	ERR1248370	ERR1248393	<i>B. subtilis</i>
16	ERR1248371	ERR1248393	<i>B. subtilis</i>
17	ERR1248372	ERR1248394	<i>B. subtilis</i>
18	ERR1248373	ERR1248394	<i>B. subtilis</i>
19	ERR1248374	ERR1248394	<i>B. subtilis</i>
20	ERR1248375	ERR1248395	<i>B. subtilis</i>
21	ERR1248376	ERR1248395	<i>B. subtilis</i>
22	ERR1248377	ERR1248395	<i>B. subtilis</i>
23	ERR1248378	ERR1248396	<i>B. subtilis</i>
24	ERR1248379	ERR1248396	<i>B. subtilis</i>
25	ERR1248380	ERR1248396	<i>B. subtilis</i>
26	ERR1248381	ERR1248397	<i>B. subtilis</i>
27	ERR1248382	ERR1248397	<i>B. subtilis</i>
28	ERR1248383	ERR1248397	<i>B. subtilis</i>
29	ERR1248384	ERR1248398	<i>B. subtilis</i>
30	ERR1248385	ERR1248398	<i>B. subtilis</i>
31	ERR1248386	ERR1248398	<i>B. subtilis</i>
32	ERR1248387	ERR1248399	<i>B. subtilis</i>

33	ERR1248388	ERR1248399	<i>B. subtilis</i>
34	ERR1248389	ERR1248399	<i>B. subtilis</i>
35	ERR1248390	ERR1248400	<i>B. subtilis</i>
36	ERR1248391	ERR1248400	<i>B. subtilis</i>
37	ERR1248392	ERR1248400	<i>B. subtilis</i>
38	ERR1248436	ERR1248439	<i>L. monocytogenes</i>
39	ERR1248437	ERR1248439	<i>L. monocytogenes</i>
40	ERR1248438	ERR1248439	<i>L. monocytogenes</i>
41	ERR1248440	ERR1248461	<i>L. monocytogenes</i>
42	ERR1248441	ERR1248461	<i>L. monocytogenes</i>
43	ERR1248442	ERR1248461	<i>L. monocytogenes</i>
44	ERR1248443	ERR1248462	<i>L. monocytogenes</i>
45	ERR1248444	ERR1248462	<i>L. monocytogenes</i>
46	ERR1248445	ERR1248462	<i>L. monocytogenes</i>
47	ERR1248446	ERR1248463	<i>L. monocytogenes</i>
48	ERR1248447	ERR1248463	<i>L. monocytogenes</i>
49	ERR1248448	ERR1248463	<i>L. monocytogenes</i>
50	ERR1248449	ERR1248464	<i>L. monocytogenes</i>
51	ERR1248450	ERR1248464	<i>L. monocytogenes</i>
52	ERR1248451	ERR1248464	<i>L. monocytogenes</i>
53	ERR1248452	ERR1248465	<i>L. monocytogenes</i>
54	ERR1248453	ERR1248465	<i>L. monocytogenes</i>
55	ERR1248454	ERR1248465	<i>L. monocytogenes</i>
56	ERR1248455	ERR1248466	<i>L. monocytogenes</i>
57	ERR1248456	ERR1248466	<i>L. monocytogenes</i>
58	ERR1248457	ERR1248466	<i>L. monocytogenes</i>
59	ERR1248458	ERR1248467	<i>L. monocytogenes</i>
60	ERR1248459	ERR1248467	<i>L. monocytogenes</i>
61	ERR1248460	ERR1248467	<i>L. monocytogenes</i>
62	ERR1248401	ERR1248404	<i>E. faecalis</i>
63	ERR1248402	ERR1248404	<i>E. faecalis</i>
64	ERR1248403	ERR1248404	<i>E. faecalis</i>
65	ERR1248409	ERR1248429	<i>E. faecalis</i>
66	ERR1248410	ERR1248429	<i>E. faecalis</i>
67	ERR1248411	ERR1248429	<i>E. faecalis</i>
68	ERR1248412	ERR1248430	<i>E. faecalis</i>
69	ERR1248413	ERR1248430	<i>E. faecalis</i>
70	ERR1248414	ERR1248430	<i>E. faecalis</i>
71	ERR1248415	ERR1248431	<i>E. faecalis</i>
72	ERR1248416	ERR1248431	<i>E. faecalis</i>
73	ERR1248417	ERR1248431	<i>E. faecalis</i>
74	ERR1248418	ERR1248432	<i>E. faecalis</i>

75	ERR1248419	ERR1248432	<i>E. faecalis</i>
76	ERR1248420	ERR1248432	<i>E. faecalis</i>
77	ERR1248421	ERR1248433	<i>E. faecalis</i>
78	ERR1248422	ERR1248433	<i>E. faecalis</i>
79	ERR1248423	ERR1248433	<i>E. faecalis</i>
80	ERR1248424	ERR1248434	<i>E. faecalis</i>
81	ERR1248425	ERR1248434	<i>E. faecalis</i>
82	ERR1248426	ERR1248435	<i>E. faecalis</i>
83	ERR1248427	ERR1248435	<i>E. faecalis</i>
84	ERR1248428	ERR1248435	<i>E. faecalis</i>
85	SRR1918928	SRR1918928	<i>B. subtilis</i>
86	SRR1918929	SRR1918929	<i>B. subtilis</i>
87	SRR1918930	SRR1918930	<i>B. subtilis</i>
88	SRR1918931	SRR1918931	<i>B. subtilis</i>
89	SRR1918932	SRR1918932	<i>B. subtilis</i>
90	SRR1918933	SRR1918933	<i>B. subtilis</i>
91	SRR1918934	SRR1918934	<i>B. subtilis</i>
92	SRR1918935	SRR1918935	<i>B. subtilis</i>
93	SRR1918936	SRR1918936	<i>B. subtilis</i>
94	SRR1918937	SRR1918937	<i>B. subtilis</i>
95	SRR1918938	SRR1918938	<i>B. subtilis</i>
96	SRR1918939	SRR1918939	<i>B. subtilis</i>
97	SRR12763770	SRR12763774	<i>Synechocystis</i>
98	SRR12763771	SRR12763775	<i>Synechocystis</i>
99	SRR12232093	SRR12232093	<i>B. subtilis</i>
100	SRR12232094	SRR12232094	<i>B. subtilis</i>
101	SRR12232095	SRR12232095	<i>B. subtilis</i>
102	SRR12232096	SRR12232096	<i>B. subtilis</i>
103	SRR12232097	SRR12232097	<i>B. subtilis</i>
104	SRR12232098	SRR12232098	<i>B. subtilis</i>
105	SRR12232099	SRR12232099	<i>B. subtilis</i>
106	SRR12232100	SRR12232100	<i>B. subtilis</i>
107	SRR10400939	SRR10400958	<i>Z. mobilis</i>
108	SRR10400940	SRR10400959	<i>Z. mobilis</i>
109	SRR10400941	SRR10400960	<i>Z. mobilis</i>
110	SRR10400942	SRR10400961	<i>Z. mobilis</i>
111	SRR10400943	SRR10400962	<i>Z. mobilis</i>
112	SRR10400944	SRR10400963	<i>Z. mobilis</i>
113	SRR10400945	SRR10400964	<i>Z. mobilis</i>
114	SRR10400946	SRR10400965	<i>Z. mobilis</i>
115	SRR10400947	SRR10400966	<i>Z. mobilis</i>
116	SRR10400948	SRR10400967	<i>Z. mobilis</i>

<b>117</b>	SRR10400949	SRR10400968	<i>Z. mobilis</i>
<b>118</b>	SRR10400950	SRR10400969	<i>Z. mobilis</i>
<b>119</b>	SRR10400951	SRR10400970	<i>Z. mobilis</i>
<b>120</b>	SRR10400952	SRR10400971	<i>Z. mobilis</i>
<b>121</b>	SRR10400953	SRR10400974	<i>Z. mobilis</i>
<b>122</b>	SRR10400954	SRR10400977	<i>Z. mobilis</i>
<b>123</b>	SRR10400955	SRR10400980	<i>Z. mobilis</i>
<b>124</b>	SRR10400956	SRR10400983	<i>Z. mobilis</i>
<b>125</b>	SRR10400957	SRR10400986	<i>Z. mobilis</i>
<b>126</b>	SRR22439140	SRR22439140	<i>E. coli</i>
<b>127</b>	SRR22439141	SRR22439141	<i>E. coli</i>
<b>128</b>	SRR22439142	SRR22439142	<i>E. coli</i>

## **Załącznik 11**

*Tabela 11. Atlas rho-niezależnych terminatorów transkrypcji u metycylinoopornych szczepów gronkowca złocistego*

<https://zenodo.org/records/11392921/files/MRSA-intrinsic-terminators-atlas.tsv?download=1> (główna tabela)

<https://zenodo.org/records/11392921/files/MRSA-intrinsic-terminators-atlas-column-description.txt?download=1> (objaśnienia kolumn)

DOI: **10.5281/zenodo.11392921** (Kosinski, 2024)



## **Załącznik 12**

*Tabela 12. Wyniki analizy różnicowej terminacji transkrypcji w odpowiedzi na stres antybiotykowy u *S. aureus*. Objaśnienia kolumn: %T – efektywność terminacji. Szczegółowy opis prób i porównań znajduje się w załączniku 1.*

*<https://zenodo.org/records/11392921/files/differential-termination-analysis.xlsx?download=1>*

DOI: **10.5281/zenodo.11392921** (Kosinski, 2024)

## 14. Oświadczenia współautorów

**Kosiński JG**, Ranaweera S, Chelkowska-Pauszek A, Kashlev M, Babitzke P, Żywicki M. *Characterization of bacterial intrinsic transcription terminators identified with TERMITE – a novel method for comprehensive analysis of Term-seq data*. bioRxiv. 2024 May. doi: 10.1101/2024.05.16.594410

Poznań, 14.06.2024.....

Jan Kosiński, Msc  
Department of Computational Biology  
Institute of Molecular Biology and Biotechnology  
Faculty of Biology, Adam Mickiewicz University in Poznań  
Uniwersytetu Poznańskiego 6  
61-614 Poznań

### Co-authorship statement


As a co-author of the publication

Kosiński JG, Ranaweera S, Chełkowska-Pauszek A, Kashlev M, Babitzke P, Żywicki M.  
*Characterization of bacterial intrinsic transcription terminators identified with TERMITE – a novel method for comprehensive analysis of Term-seq data.* bioRxiv. 2024 May. doi: 10.1101/2024.05.16.594410

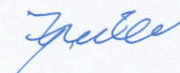
I hereby confirm my contributions to this publication as follows:

1. **Algorithm Design and Implementation:** I led the design and implementation of the algorithm
2. **Computational Analyses of Bacterial Intrinsic Terminators:** I conducted computational analyses, excluding those related to secondary structure conservation
3. **Data Visualization:** I played a pivotal role in creating effective data visualizations
4. **Results Interpretation and Algorithm Evaluation:** I actively participated in interpreting results and evaluating algorithm performance
5. **Manuscript Composition:** I led the preparation of the initial version of the manuscript
6. **Final Manuscript Preparation:** I contributed to preparing the final version of the manuscript

Signature:



Confirmed by Marek Żywicki, PhD:



University Park, Jun 6<sup>th</sup>, 2024

Sandeevani Ranaweera  
Department of Biochemistry and Molecular Biology  
Penn State University  
University Park, PA 16802

**Co-authorship statement**

As a co-author of the publication

Kosiński JG, Ranaweera S, Chełkowska-Pauszek A, Kashlev M, Babitzke P, Żywicki M.  
*Characterization of bacterial intrinsic transcription terminators identified with TERMITE –  
a novel method for comprehensive analysis of Term-seq data.* bioRxiv. 2024 May. doi:  
10.1101/2024.05.16.594410

I hereby confirm my contributions to this publication as follows:

1. ***E. coli* sequencing libraries preparation:** I was responsible for conducting the process of *E. coli* sequencing libraries preparation
2. **Manuscript Preparation:** I contributed to preparing the initial version of the manuscript

  
Signature:

Poznań, 10.06.2024.....

Agnieszka Chełkowska-Pauszek, MSc  
Department of Computational Biology  
Institute of Molecular Biology and Biotechnology  
Faculty of Biology, Adam Mickiewicz University in Poznań  
Uniwersytetu Poznańskiego 6  
61-614 Poznań

#### Co-authorship statement

As a co-author of the publication

Kosiński JG, Ranaweera S, Chełkowska-Pauszek A, Kashlev M, Babitzke P, Żywicki M.  
*Characterization of bacterial intrinsic transcription terminators identified with TERMITE –  
a novel method for comprehensive analysis of Term-seq data.* bioRxiv. 2024 May. doi:  
10.1101/2024.05.16.594410

I hereby confirm my contributions to this publication as follows:

1. **Data analysis:** I analyzed the secondary structure conservation of the terminators
2. **Data Visualization:** I played a pivotal role in creating data visualizations

Signature:

Agnieszka  
Chełkowska-Pauszek

06-05-2024

Frederick MD, .....

Mikhail Kashlev, Ph.D.  
Senior Principal Investigator and Deputy Lab Chief RNA Biology Laboratory  
National Cancer Institute  
Frederick MD, 21702  
USA

### Co-authorship statement

As a co-author of the publication

Kosiński JG, Ranaweera S, Chelkowska-Pauszek A, Kashlev M, Babitzke P, Żywicki M.  
*Characterization of bacterial intrinsic transcription terminators identified with TERMITE –  
a novel method for comprehensive analysis of Term-seq data.* bioRxiv. 2024 May. doi:  
10.1101/2024.05.16.594410

I hereby confirm my contributions to this publication as follows:

1. **Review:** I was responsible for the critical review of both the initial and the final version of the manuscript
2. **Final Manuscript Preparation:** I contributed to preparing the final version of the manuscript
3. **Results Interpretation:** I actively participated in interpreting results

Signature:



University Park, Jun 6<sup>th</sup>, 2024

Paul Babitzke, PhD  
Stanley R. Person Professor of Molecular Biology  
Director, Center for RNA Molecular Biology  
Department of Biochemistry and Molecular Biology  
Penn State University  
University Park, PA 16802

### Co-authorship statement

As a co-author of the publication

Kosiński JG, Ranaweera S, Chelkowska-Pauszek A, Kashlev M, Babitzke P, Żywicki M. *Characterization of bacterial intrinsic transcription terminators identified with TERMITE – a novel method for comprehensive analysis of Term-seq data*. bioRxiv. 2024 May. doi: 10.1101/2024.05.16.594410

I hereby confirm my contributions to this publication as follows:

1. ***E. coli* sequencing libraries preparation:** I led the process of *E. coli* sequencing libraries preparation
2. **Final Manuscript Preparation:** I contributed to preparing the final version of the manuscript
3. **Review:** I reviewed both the initial and the final versions of the manuscript
4. **Results Interpretation and Algorithm Evaluation:** I actively participated in interpreting results and evaluating algorithm performance

Signature:



Poznań, 19.06.2024

Marek Żywicki, PhD  
Department of Computational Biology  
Institute of Molecular Biology and Biotechnology  
Faculty of Biology, Adam Mickiewicz University in Poznań  
Uniwersytetu Poznańskiego 6  
61-614 Poznań

#### Co-authorship statement


As a co-author of the publication

Kosiński JG, Ranaweera S, Chełkowska-Pauszek A, Kashlev M, Babitzke P, Żywicki M.  
*Characterization of bacterial intrinsic transcription terminators identified with TERMITE – a novel method for comprehensive analysis of Term-seq data.* bioRxiv. 2024 May. doi:  
10.1101/2024.05.16.594410

I hereby confirm my contributions to this publication as follows:

1. **Algorithm Design:** I participated in the design of the algorithm
2. **Data Visualization:** I participated in creating effective data visualizations
3. **Results Interpretation and Algorithm Evaluation:** I actively participated in interpreting results and evaluating algorithm performance
4. **Manuscript Composition:** I led the preparation of the final version of the manuscript
5. **Review:** I reviewed both the initial and the final versions of the manuscript

Signature:





**Kosiński JG, Żywicki M. *AGouTI - Flexible Annotation of Genomic and Transcriptomic Intervals*. PLoS Comput Biol. 2023 Oct 18;19(10):e1011527. doi: 10.1371/journal.pcbi.1011527. PMID: 37851672; PMCID: PMC10621952.**

Poznań, 14.06.2024...

Jan Kosiński, Msc  
Department of Computational Biology  
Institute of Molecular Biology and Biotechnology  
Faculty of Biology, Adam Mickiewicz University in Poznań  
Uniwersytetu Poznańskiego 6  
61-614 Poznań

#### Co-authorship statement

As a co-author of the publication

Kosiński JG, Żywicki M. AGouTI-Flexible Annotation of Genomic and Transcriptomic Intervals. PLoS Comput Biol. 2023 Oct 18;19(10):e1011527. doi: 10.1371/journal.pcbi.1011527. PMID: 37851672; PMCID: PMC10621952.


I hereby confirm my contributions to this publication as follows:

1. participated in conceptualizing the algorithm, acquiring funding, designing the methodology, preparing visualizations, validating and testing the software
2. developed the software
3. prepared the initial version of the manuscript

Signature:



Confirmed by Marek Żywicki, PhD:



Poznań, 14.06.2024

Marek Żywicki, PhD  
Department of Computational Biology  
Institute of Molecular Biology and Biotechnology  
Faculty of Biology, Adam Mickiewicz University in Poznań  
Uniwersytetu Poznańskiego 6  
61-614 Poznań

#### Co-authorship statement

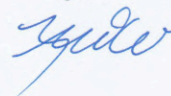
As a co-author of the publication

Kosiński JG, Żywicki M. AGouTI-Flexible Annotation of Genomic and Transcriptomic Intervals. PLoS Comput Biol. 2023 Oct 18;19(10):e1011527. doi: 10.1371/journal.pcbi.1011527. PMID: 37851672; PMCID: PMC10621952.

I hereby confirm my contributions to this publication as follows:

1. participated in conceptualizing the algorithm, acquiring funding, designing the methodology, preparing visualizations, validating and testing the software
2. wrote the final version of the manuscript
3. supervised the project

Signature:



Chełkowska-Pauszek A, **Kosiński JG**, Marciniak K, Wysocka M, Bąkowska-Żywicka K, Żywicki M. *The Role of RNA Secondary Structure in Regulation of Gene Expression in Bacteria*. Int J Mol Sci. 2021 Jul 22;22(15):7845. doi: 10.3390/ijms22157845. PMID: 34360611; PMCID: PMC8346122

Poznań, 14.06.2024

Jan Kosiński, Msc  
Department of Computational Biology  
Institute of Molecular Biology and Biotechnology  
Faculty of Biology, Adam Mickiewicz University in Poznań  
Uniwersytetu Poznańskiego 6  
61-614 Poznań

### Co-authorship statement

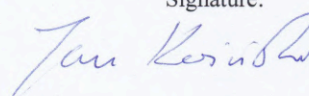
As a co-author of the publication

Chelkowska-Pauszek A, Kosiński JG, Marciniak K, Wysocka M, Bąkowska-Żywicka K, Żywicki M. *The Role of RNA Secondary Structure in Regulation of Gene Expression in Bacteria*. Int J Mol Sci. 2021 Jul 22;22(15):7845. doi: 10.3390/ijms22157845. PMID: 34360611; PMCID: PMC8346122,

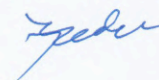
I hereby confirm my contributions to this publication as follows:

1. **Data Visualization:** I played a pivotal role in creating data visualizations
2. **Manuscript Preparation:** I was jointly responsible for writing the manuscript

Signature:



Confirmed by Marek Żywicki, PhD:



Poznań, 10.06.2024.....

Agnieszka Chelkowska-Pauszek, Msc  
Department of Computational Biology  
Institute of Molecular Biology and Biotechnology  
Faculty of Biology, Adam Mickiewicz University in Poznań  
Uniwersytetu Poznańskiego 6  
61-614 Poznań

### Co-authorship statement

As a co-author of the publication

Chelkowska-Pauszek A, Kosiński JG, Marciniak K, Wysocka M, Bąkowska-Żywicka K, Żywicki M. *The Role of RNA Secondary Structure in Regulation of Gene Expression in Bacteria*. Int J Mol Sci. 2021 Jul 22;22(15):7845. doi: 10.3390/ijms22157845. PMID: 34360611; PMCID: PMC8346122,

I hereby confirm my contributions to this publication as follows:

1. **Data Visualization:** I played a pivotal role in creating data visualizations
2. **Manuscript Preparation:** I was jointly responsible for writing the manuscript

Signature:

Agnieszka  
Chelkowska-Pauszek

Poznań, .....

17. 06. 2024

Klementyna Marciniak  
Institute of Bioorganic Chemistry  
Polish Academy of Sciences  
Poznań, Poland

### Co-authorship statement

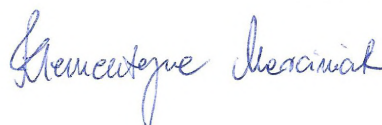
As a co-author of the publication

Chelkowska-Pauszek A, Kosiński JG, Marciniak K, Wysocka M, Bąkowska-Żywicka K, Żywicki M. *The Role of RNA Secondary Structure in Regulation of Gene Expression in Bacteria*. Int J Mol Sci. 2021 Jul 22;22(15):7845. doi: 10.3390/ijms22157845. PMID: 34360611; PMCID: PMC8346122,

I hereby confirm my contributions to this publication as follows:

1. **Manuscript Preparation:** I was jointly responsible for writing the manuscript

Signature:



Poznań, ..... 6.06.2024

Marta Szustakowska (Wysocka), Msc  
Department of Computational Biology  
Institute of Molecular Biology and Biotechnology  
Faculty of Biology, Adam Mickiewicz University in Poznań  
Uniwersytetu Poznańskiego 6  
61-614 Poznań

#### Co-authorship statement


As a co-author of the publication

Chełkowska-Pauszek A, Kosiński JG, Marciniak K, Wysocka M, Bąkowska-Żywicka K, Żywicki M. *The Role of RNA Secondary Structure in Regulation of Gene Expression in Bacteria*. Int J Mol Sci. 2021 Jul 22;22(15):7845. doi: 10.3390/ijms22157845. PMID: 34360611; PMCID: PMC8346122,

I hereby confirm my contributions to this publication as follows:

1. **Data Visualization:** I played a pivotal role in creating data visualizations
2. **Manuscript Preparation:** I was jointly responsible for writing the manuscript

Signature:





Poznań, 11.06.2024

Kamilla Grzywacz (Bąkowska-Żywicka)  
Institute of Bioorganic Chemistry  
Polish Academy of Sciences  
Poznań, Poland

#### Co-authorship statement

As a co-author of the publication

Chelkowska-Pauszek A, Kosiński JG, Marciniak K, Wysocka M, Bąkowska-Żywicka K, Żywicki M. *The Role of RNA Secondary Structure in Regulation of Gene Expression in Bacteria*. Int J Mol Sci. 2021 Jul 22;22(15):7845. doi: 10.3390/ijms22157845. PMID: 34360611; PMCID: PMC8346122,

I hereby confirm my contributions to this publication as follows:

1. **Conceptualization:** I was jointly responsible for conceptualizing the manuscript and providing advisory support during its preparation
2. **Review:** I played a pivotal role in the review of the manuscript

Signature:

Kamilla Grzywacz

Poznań, 17.06.2025

Marek Żywicki, PhD  
Department of Computational Biology  
Institute of Molecular Biology and Biotechnology  
Faculty of Biology, Adam Mickiewicz University in Poznań  
Uniwersytetu Poznańskiego 6  
61-614 Poznań

#### Co-authorship statement

As a co-author of the publication

Chełkowska-Pauszek A, Kosiński JG, Marciniak K, Wysocka M, Bąkowska-Żywicka K, Żywicki M. *The Role of RNA Secondary Structure in Regulation of Gene Expression in Bacteria*. Int J Mol Sci. 2021 Jul 22;22(15):7845. doi: 10.3390/ijms22157845. PMID: 34360611; PMCID: PMC8346122,

I hereby confirm my contributions to this publication as follows:

1. **Conceptualization:** I was jointly responsible for conceptualizing the manuscript and providing advisory support during its preparation
2. **Review:** I played a pivotal role in the review of the manuscript

Signature:

