

การพัฒนากระแสนางเชิงคอมพิวเตอร์เพื่อการระบุอาร์เอ็นเอที่เติมพอลิอะดีนีนในโพรคาริโอตจาก  
ข้อมูลลำดับเบสอาร์เอ็นเอที่เติมพอลิยูรีดีนด้วยนาโนพอร์โดยตรง



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต  
สาขาวิชาชีวสารสนเทศศาสตร์และชีววิทยาเชิงคอมพิวเตอร์ (สหสาขาวิชา) สหสาขาวิชาชีวสารสนเทศ  
ศาสตร์และชีววิทยาทางคอมพิวเตอร์  
บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย  
ปีการศึกษา 2564  
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

COMPUTATIONAL WORKFLOW DEVELOPMENT FOR IDENTIFICATION OF  
POLYADENYLATED RNA IN PROKARYOTE FROM DIRECT RNA NANOPORE SEQUENCING  
DATA WITH POLYURIDYLATION



A Thesis Submitted in Partial Fulfillment of the Requirements  
for the Degree of Master of Science in Bioinformatics and Computational Biology  
Inter-Department of Bioinformatics and Computational Biology

GRADUATE SCHOOL

Chulalongkorn University

Academic Year 2021

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การพัฒนากระแสนงานเชิงคอมพิวเตอร์เพื่อการระบุอาร์เอ็นเอที่เติมพอลิอะดีนีนไนโพรคาร์ีโอตจากข้อมูลลำดับเบสอาร์เอ็นเอที่เติมพอลิยูรีดีนด้วยนาโนพอร์โดยตรง
โดย	นายจิรัชติ กรัณยากรณ์
สาขาวิชา	ชีวสารสนเทศศาสตร์และชีววิทยาเชิงคอมพิวเตอร์ (สหสาขาวิชา)
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	อาจารย์ ดร.พรชัย แก้วทรัพย์ศักดิ์
อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม	ผู้ช่วยศาสตราจารย์ ดร.มนนันทน์ พงษ์พานิช อาจารย์ ดร.นฤมล ประทานวณิช

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้บัณฑิตวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

..... คณบดีบัณฑิตวิทยาลัย  
(รองศาสตราจารย์ ดร.ยุพธนา ฉัยพรรณรัตน์)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ  
(ศาสตราจารย์ ดร.ศุภจิตรา ชัชวาลย์)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก  
(อาจารย์ ดร.พรชัย แก้วทรัพย์ศักดิ์)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม  
(ผู้ช่วยศาสตราจารย์ ดร.มนนันทน์ พงษ์พานิช)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม  
(อาจารย์ ดร.นฤมล ประทานวณิช)

..... กรรมการ  
(รองศาสตราจารย์ ดร.สัญญาชัย พยุงกร)

..... กรรมการภายนอกมหาวิทยาลัย  
(อาจารย์ ดร.พิรุณ เจนเจริญพันธ์)

จิรัชติ กรณ์ยากรณ์ : การพัฒนากระแสงานเชิงคอมพิวเตอร์เพื่อการระบุอาร์เอ็นเอที่เติมพอลิอะดีนีน  
ในโพรแคริโอตจากข้อมูลลำดับเบสอาร์เอ็นเอที่เติมพอลิยูรีดีนด้วยนาโนพอร์โดยตรง. (

COMPUTATIONAL WORKFLOW DEVELOPMENT FOR IDENTIFICATION OF  
POLYADENYLATED RNA IN PROKARYOTE FROM DIRECT RNA NANOPORE SEQUENCING  
DATA WITH POLYURIDYLATION) อ.ที่ปรึกษาหลัก : อ. ดร.พรชัย แก้วทรัพย์ศักดิ์, อ.ที่ปรึกษาร่วม  
: ผศ. ดร.มนนัทธ์ พงษ์พานิช, อ. ดร.นฤมล ประทานวณิช

กระบวนการเติมพอลิอะดีนีนเป็นการตัดแปรรหัสอาร์เอ็นเอที่ปลาย 3' หลังการถอดรหัสหรือเรียกว่าหางพอลิอะดีนีน ซึ่งหางพอลิอะดีนีนมีหน้าที่ในการควบคุมเสถียรภาพของอาร์เอ็นเอและนำไปสู่การควบคุมการแสดงออกของยีน ปัจจุบันการศึกษาความยาวและหน้าที่ของหางพอลิอะดีนีนในอาร์เอ็นเอของโพรแคริโอตมีน้อยมาก เนื่องจากความยาวของหางพอลิอะดีนีนในอาร์เอ็นเอของโพรแคริโอตสั้นมาก และทำหน้าที่เป็นสัญญาณการย่อยสลายอาร์เอ็นเอ ส่งผลให้ตรวจพบได้ยาก ในการศึกษาครั้งนี้ได้ทำการพัฒนาวิธีทางคอมพิวเตอร์เพื่อศึกษาหางพอลิอะดีนีนความยาวสั้นในอาร์เอ็นเอของโพรแคริโอต โดยใช้ข้อมูลอาร์เอ็นเอจากการหาลำดับเบสยุคที่สามด้วยอ็อกซ์ฟอร์ดนาโนพอร์ที่สามารถหาลำดับเบสอาร์เอ็นเอสายยาวได้โดยตรง โดยทำการปรับปรุงขั้นตอนการเตรียมไลบรารีสำหรับการหาลำดับเบสโดยใช้เอนไซม์เพื่อสังเคราะห์พอลิยูรีดีนที่ปลาย 3' ของสายลำดับเบส เพื่ออนุรักษ์ความยาวหางพอลิอะดีนีนดั้งเดิมไว้ และใช้สำหรับการเชื่อมต่อกับพอลิอะดีนีนอะแดปเตอร์สำหรับการหาลำดับเบส วิธีนี้ไม่จำเป็นต้องเพิ่มปริมาณสารพันธุกรรม หรือถอดรหัสย้อนกลับ ที่อาจส่งผลต่อความยาวดั้งเดิมของหางพอลิอะดีนีน จากการศึกษาด้วยอาร์เอ็นเอสังเคราะห์ที่มีการกำหนดความยาวหางพอลิอะดีนีนไว้คงที่พบว่าโปรแกรมสำหรับการประมาณความยาวหางพอลิอะดีนีน (nanopolish และ tailfindr) มีความคลาดเคลื่อนในการประมาณความยาวของหางพอลิอะดีนีนสายสั้น โดยมีค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ยอยู่ในช่วง 0.33 ถึง 7.49 นิวคลีโอไทด์ สำหรับ nanopolish และ 6.38 ถึง 16.89 นิวคลีโอไทด์ สำหรับ tailfindr หลังจากการปรับปรุงข้อมูลด้วยการกำจัดสัญญาณไม่ปกติและใช้ข้อมูลจากการหาลำดับเบสเข้าช่วยพบว่าโปรแกรมสามารถหาความยาวของหางพอลิอะดีนีนสายสั้นได้ใกล้เคียงมากขึ้น โดยมีค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ยลดลงอยู่ในช่วง 0.43 ถึง 7.28 นิวคลีโอไทด์ สำหรับ nanopolish และ 3.28 ถึง 5.96 นิวคลีโอไทด์ สำหรับ tailfindr และเมื่อนำวิธีการศึกษาดังกล่าวไปประยุกต์ในอาร์เอ็นเอของ *Escherichia coli* K-12 ได้ตรวจพบหางพอลิอะดีนีนในยีนซึ่งเคยมีรายงานจากงานวิจัยก่อนหน้านี้ เช่น *rpsO*, *rplQ* และ *ompA* เป็นต้น นอกจากนี้ยังพบยีนที่มีความเป็นไปได้ที่จะมีหางพอลิอะดีนีน แต่ยังไม่มียารายงานมาก่อน เช่น *apt* และ *ppa* เป็นต้น การศึกษาความยาวหางพอลิอะดีนีนในโพรแคริโอตสามารถเชื่อมโยงกับการควบคุมการแสดงออกของยีนซึ่งจะช่วยพัฒนาองค์ความรู้ด้านการควบคุมการแสดงออกของยีนในแบคทีเรีย และสามารถนำไปประยุกต์ได้ในอนาคต

สาขาวิชา	ชีวสารสนเทศศาสตร์และชีววิทยา เชิงคอมพิวเตอร์ (สหสาขาวิชา)	ลายมือชื่อนิสิต .....
ปีการศึกษา	2564	ลายมือชื่อ อ.ที่ปรึกษาหลัก .....
		ลายมือชื่อ อ.ที่ปรึกษาร่วม .....
		ลายมือชื่อ อ.ที่ปรึกษาร่วม .....

# # 6380139020 : MAJOR BIOINFORMATICS AND COMPUTATIONAL BIOLOGY

KEYWORD: Poly(A) tail / Direct RNA Nanopore sequencing / RNA / Escherichia coli

Jirachote Karunyakorn : COMPUTATIONAL WORKFLOW DEVELOPMENT FOR IDENTIFICATION OF POLYADENYLATED RNA IN PROKARYOTE FROM DIRECT RNA NANOPORE SEQUENCING DATA WITH POLYURIDYLATION. Advisor: PORNCHAI KAEWSAPSAK, Ph.D. Co-advisor: Asst. Prof. MONNAT PONGPANICH, Ph.D., NARUEMON PRATANWANICH, Ph.D.

Polyadenylation or poly(A) tail is a post-transcriptional modification at the 3' end of RNA. Its main function is to regulate RNA stability and control gene expression. Currently a little is known about the length and function of poly(A) tail in prokaryotic RNAs as they are challenging to detect due to very short poly(A) tail length and its function as RNA degradation signal. In this study, a computational method was developed to study short length poly(A) tail in prokaryotic RNA using Oxford Nanopores Technologies (ONT) direct RNA sequencing, a third-generation sequencing technique, which can provide long-read sequencing data. In library preparation, poly(U) polymerase was used to synthesize polyuridine or poly(U) tail at the 3' end of the sequencing strand to bind to the custom poly(A) sequencing adapter, while still preserve the native poly(A) tail length. This method does not need reverse transcription or amplification that might affect the native length poly(A) tail. Using synthetic RNAs with pre-defined poly(A) length, the programs for estimating the length of poly(A) tail, nanopolish and tailfindr, have noticeable error in estimating the length of short poly(A) tail with mean absolute error were ranging from 0.33 to 7.49 nucleotides for the nanopolish and 6.38 to 16.89 nucleotides for the tailfindr. After removing reads with unusual poly(A) electrical signal and reassigning underestimated poly(A) tail using basecalled sequencing data, our developed poly(A) tail reassignment workflow improved the accuracy of estimating short poly(A) tail length with the mean absolute errors ranging from 0.43 to 7.28 nucleotides for the nanopolish and 3.28 to 5.96 nucleotides for the tailfindr. Applying our method to *Escherichia coli* RNA, previously reported genes with poly(A) tails such as *rpsO*, *rplQ* and *ompA* were found. Furthermore, the novel genes with potential polyadenylation such as *apt* and *ppa* were also discovered. Study of poly(A) tail length in prokaryotes can be linked to gene expression and regulation, which can lead various applications such as drug development for pathogens or bacterial engineering in the future.

Field of Study: Bioinformatics and  
Computational Biology

Academic Year: 2021

Student's Signature .....

Advisor's Signature .....

Co-advisor's Signature .....

Co-advisor's Signature .....

## กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ ด้วยความกรุณาเมตตาจากบุคคลผู้มีพระคุณหลายท่านที่ได้ช่วยเหลือ ให้คำแนะนำ ข้อชี้แนะ คำปรึกษา ความคิดเห็นทางวิชาการอันประโยชน์ต่องานวิจัยอย่างยิ่ง และความร่วมมือของบุคคลจากหลายฝ่ายที่ผู้วิจัยมีอาจหาค่าเปรียบได้

ขอกราบขอบพระคุณอาจารย์ ดร. พรชัย แก้วทรัพย์ศักดิ์ ที่ให้ความอนุเคราะห์ และสละเวลาอันมีค่ารับเป็นอาจารย์ที่ปรึกษา และ ผศ. ดร. มนันท์ พงษ์พานิช และ อาจารย์ ดร. นฤมล ประทานวณิช อาจารย์ที่ปรึกษาร่วม ที่ได้กรุณาให้คำปรึกษา คำชี้แนะ และตรวจพิจารณาแก้ไขวิทยานิพนธ์อันเป็นประโยชน์ต่อความถูกต้องและความสมบูรณ์ของเนื้อหาทางวิชาการ

ขอขอบพระคุณ ศ.ดร.ศุภจิตรา ชัชวาลย์ ที่ได้กรุณาเสียสละเวลาอันมีค่าเป็นประธานกรรมการสอบวิทยานิพนธ์ และกรุณาให้คำปรึกษาทางวิชาการ และคำแนะนำด้านทุนการศึกษาเพื่อประโยชน์สูงสุดแก่นิสิตผู้รับทุน

ขอขอบคุณทุนอุดหนุนการศึกษาระดับบัณฑิตศึกษาจาก บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย เพื่อเฉลิมฉลองวโรกาสที่พระบาทสมเด็จพระเจ้าอยู่หัว ทรงเจริญพระชนมายุครบ ๗๒ พรรษา และทุน ๙๐ ปี จุฬาลงกรณ์มหาวิทยาลัย กองทุนรัชดาภิเษกสมโภช รวมถึง ทุนพัฒนาอาจารย์ใหม่/ นักวิจัยใหม่ กองทุนรัชดาภิเษกสมโภช ที่ได้มอบทุนการศึกษา และ ทุนสำหรับงานวิจัยส่งผลให้งานวิจัยสามารถสำเร็จลุล่วง

ขอขอบพระคุณขอขอบคุณอาจารย์ ดร. สิริระ ศรีสวัสดิ์ และ รศ.ดร.สัญชัย พงษ์ภร ที่จัดหาเครื่องมืออำนวยความสะดวกสำหรับงานวิจัย ทั้งห้องปฏิบัติการ และ เครื่องเซฟเวอร์สำหรับให้ในงานวิจัยส่งผลให้งานวิจัยสามารถสำเร็จลุล่วง

ขอขอบพระคุณคุณอาจารย์ อ. ดร. ธิดาทิพย์ วงศ์สุรวัดน์ และ อ. ดร. พิรุณ เจนเจริญพันธ์ จากคณะแพทยศิริราช มหาวิทยาลัยมหิดล ที่ให้คำปรึกษา และข้อชี้แนะเพิ่มเติมทางวิชาการสำหรับการปรับปรุงพัฒนางานวิจัยให้มีประสิทธิภาพมากขึ้น

ขอกราบขอบพระคุณท่านอาจารย์ทุกท่านในหลักสูตรชีวสารสนเทศ และชีววิทยาเชิงคอมพิวเตอร์ จุฬาลงกรณ์มหาวิทยาลัย รวมถึงท่านอาจารย์จาก ภาคชีววิทยา มหาวิทยาลัยมหิดลทุกท่าน ที่ได้ถ่ายทอดความรู้มาตลอดการศึกษา และให้คำชี้แนะในระดับอุดมศึกษาของผู้วิจัย

ขอขอบพระคุณอาจารย์ ดร. วรุฒ ศิริวุฒิ อาจารย์ที่ปรึกษาในระดับปริญญาตรีของผู้วิจัย ผู้ให้คำแนะนำ ข้อชี้แนะ และคำปรึกษาทางด้านการวิจัย และทุนการศึกษา รวมถึงกรุณาเขียนจดหมายคำแนะนำสำหรับการสมัครทุนการศึกษา

ผู้วิจัยขอขอบคุณเจ้าหน้าที่บัณฑิตวิทยาลัย ที่ช่วยประสานงาน ให้คำปรึกษา และสนับสนุนข้อมูลต่างๆอย่างรวดเร็ว เป็นผลให้การดำเนินการทางเอกสารลุล่วงได้ด้วยดี

สุดท้ายนี้ผู้วิจัยขอกราบแทบเท้าขอบพระคุณบิดา มารดา และขอบคุณครอบครัวของผู้วิจัย ผู้ให้การสนับสนุนในทุกๆด้าน และคอยอยู่เคียงข้างเสมอในช่วงเวลาที่ยากลำบาก และขอบคุณมิตรสหาย และรุ่นพี่ของผู้วิจัยทั้งในหลักสูตรชีวสารสนเทศ และชีววิทยาเชิงคอมพิวเตอร์ จุฬาลงกรณ์มหาวิทยาลัย และภาคชีววิทยา มหาวิทยาลัยมหิดล รวมถึงบุคคลที่ผู้วิจัยรู้จัก ที่คอยช่วยเหลือให้คำปรึกษา และข้อคิดเห็นทั้งด้านวิชาการ และคอยเป็นกำลังใจให้เสมอมา และขอขอบคุณจุฬาลงกรณ์มหาวิทยาลัยที่มีโอกาส และทุนการศึกษาที่มีอาจประเมินค่าได้แก่ผู้วิจัย

## สารบัญ

	หน้า
.....	ค
บทคัดย่อภาษาไทย.....	ค
.....	ง
บทคัดย่อภาษาอังกฤษ.....	ง
กิตติกรรมประกาศ.....	จ
สารบัญ.....	ฉ
สารบัญตาราง.....	ช
สารบัญรูป.....	ฌ
บทที่ 1 บทนำ.....	1
1.1 ที่มาและความสำคัญ.....	1
1.2 วัตถุประสงค์.....	3
1.3 ขอบเขตการวิจัย.....	3
1.4 ประโยชน์ที่คาดว่าจะได้รับ.....	4
บทที่ 2 ทบทวนวรรณกรรม.....	5
2.1 ทางพอลิอะดีนีน.....	5
2.2 ความแตกต่างของการสังเคราะห์และหน้าที่ในทางพอลิอะดีนีนในโพรแครีโอตและยูแคริโอต.....	5
2.3 การศึกษาทางพอลิอะดีนีนในโพรแครีโอต.....	9
2.4 เทคโนโลยีการหาลำดับเบสด้วยอีอกซ์ฟอร์ดนาโนพอร์.....	11
บทที่ 3 วิธีการดำเนินงานวิจัย.....	18
3.1 กระบวนการสำหรับการหาความยาวทางพอลิอะดีนีน.....	18
3.2 กระบวนการสำหรับการปรับปรุงข้อมูลความยาวทางพอลิอะดีนีน.....	20

3.3 โปรแกรมสำหรับการแบ่งช่วงลำดับเบสอาร์เอ็นเอเพื่อการตรวจสอบอาร์เอ็นเอที่มีการแตกหัก ที่ปลาย 3' .....	23
3.4 การตรวจสอบช่วงสัญญาณไฟฟ้าในสายลำดับเบส.....	24
บทที่ 4 ผลการวิจัย .....	25
4.1 การทดสอบวิธีการเตรียมไลบรารีและการพัฒนากระแสน้ำสำหรับการหาลำดับเบสด้วย อี อ็อกซ์ฟอร์ดนาโนพอร์ด้วยการสังเคราะห์พอลิยูรีดินโดยใช้อาร์เอ็นเอสังเคราะห์ที่มีความยาว ทางพอลิอะดีนีนคั้งที่ .....	25
4.1.1 อาร์เอ็นเอสังเคราะห์ที่มีทางพอลิอะดีนีนความยาวคั้งที่ใช้ในงานวิจัย และค่าสถิติของ ลำดับเบสแต่ละชุดการทดลอง.....	26
4.1.2 การทดสอบประสิทธิภาพโปรแกรมประมาณความยาวทางพอลิอะดีนีนสายสั้น .....	27
4.1.3 กระแสน้ำสำหรับการปรับปรุงความยาวทางพอลิอะดีน .....	28
4.2 การตรวจหายีนที่มีทางพอลิอะดีนีนในแบคทีเรีย <i>Escherichia coli</i> K-12 ด้วยไลบรารีสำหรับ การหาลำดับเบสด้วยอ็อกซ์ฟอร์ดนาโนพอร์ด้วยการสังเคราะห์พอลิยูรีดิน .....	30
4.2.1 ค่าสถิติของลำดับเบสในแต่ละชุดการทดลอง .....	30
4.2.2 คุณลักษณะอาร์เอ็นเอของยีนใน <i>E. coli</i> K-12 ที่มีความเป็นไปได้ของการมีทางพอลิอะ ดีนีนที่ถูกตรวจพบ.....	31
บทที่ 5 อภิปรายผลการทดลอง.....	37
5.1 การทดสอบการเตรียมไลบรารีด้วยการสังเคราะห์พอลิยูรีดินและกระแสน้ำสำหรับการ ตรวจหาทางพอลิอะดีนีนโดยการหาลำดับเบสด้วยอ็อกซ์ฟอร์ดนาโนพอร์ .....	37
5.2 คุณลักษณะที่มักพบในอาร์เอ็นเอที่มีทางพอลิอะดีนีนใน <i>E. coli</i> .....	38
5.3 ประสิทธิภาพของการหาลำดับเบสอาร์เอ็นเอด้วยอ็อกซ์ฟอร์ดนาโนพอร์ .....	40
5.4 วิสัยทัศน์ในการศึกษาในอนาคต .....	42
บทที่ 6 สรุปผลการวิจัยและข้อเสนอแนะ .....	45
บรรณานุกรม.....	46
ประวัติผู้เขียน.....	55



## สารบัญตาราง

	หน้า
ตารางที่ 1 ลำดับเบสของอาร์เอ็นเอสังเคราะห์และตัวแปรควบคุม .....	25
ตารางที่ 2 ค่าสถิติของลำดับเบสของอาร์เอ็นเอสังเคราะห์ในแต่ละการทดลองจาก nanoplot .....	26
ตารางที่ 3 ค่าสถิติการประมาณความยาวทางพอลิอะดีนีนก่อนขั้นตอนการปรับปรุงข้อมูลความยาวทางพอลิอะดีนีน .....	27
ตารางที่ 4 ค่าสถิติการประมาณความยาวทางพอลิอะดีนีนหลังขั้นตอนการปรับปรุงข้อมูลความยาวทางพอลิอะดีนีน .....	29
ตารางที่ 5 ค่าสถิติของลำดับเบสของอาร์เอ็นเอ E. coli K-12 ในแต่ละการทดลองจาก nanoplot.	30
ตารางที่ 6 ตัวอย่างยื่นที่ความเป็นไปได้ที่จะมีทางพอลิอะดีนีนในงานวิจัยนี้ .....	32

## สารบัญรูป

	หน้า
รูปที่ 1 การสังเคราะห์ทางพอลิเมอไดเมอร์เพื่อหาลำดับเบสอาร์เอ็นเอในสิ่งมีชีวิตโพรแคริโอต.....	8
รูปที่ 2 รูปแบบของอาร์เอ็นเอตามตำแหน่งของทางพอลิเมอไดเมอร์ที่ถูกเคยถูกรายงาน .....	8
รูปที่ 3 การสังเคราะห์ทางพอลิเมอไดเมอร์สายยาวในโพรแคริโอตโดยการจับกับโปรตีน Hfq ที่มี ความสามารถช่วยยับยั้งการทำงานของเอนไซม์เอ็กโซไรโบนิวคลีเอส .....	9
รูปที่ 4 ความแตกต่างระหว่างการหาลำดับเบสแบบแซงเจอร์และการหาลำดับเบสแบบ NGS.....	12
รูปที่ 5 กระบวนการของหาลำดับเบสโดยใช้เอกซ์ฟอร์ดนาโนพอร์โดยสรุป.....	14
รูปที่ 6 ช่วงสัญญาณไฟฟ้าของอาร์เอ็นเอที่ผ่านรูนาโนพอร์แสดงถึงแต่ละช่วงลำดับเบส .....	15
รูปที่ 7 การเตรียมไลบรารีสำหรับการหาลำดับเบสอาร์เอ็นเอด้วยเอกซ์ฟอร์ดนาโนพอร์.....	16
รูปที่ 8 การหาลำดับเบสสายยาวด้วยเทคโนโลยี SMRT.....	17
รูปที่ 9 แผนผังการเตรียมไลบรารีเพื่อหาลำดับเบสด้วยเอกซ์ฟอร์ดนาโนพอร์โดยการสังเคราะห์พอลิยู ริดีน.....	19
รูปที่ 10 แผนผังการทำงานของวิธีการทางคอมพิวเตอร์ในการประมาณความยาวทางพอลิเมอไดเมอร์สาย สั้น.....	20
รูปที่ 11 ช่วงสัญญาณไฟฟ้าที่ถูกทำนายและประมาณความยาวทางพอลิเมอไดเมอร์ A. มีพอลิยูริดีน และ ทางพอลิเมอไดเมอร์ B. มีเพียงทางพอลิเมอไดเมอร์เท่านั้น .....	21
รูปที่ 12 การกระจายตัวของค่าเฉลี่ยของสัญญาณไฟฟ้าในช่วงทางพอลิเมอไดเมอร์ที่ถูกประมาณ .....	21
รูปที่ 13 การกระจายตัวของส่วนเบี่ยงเบนมาตรฐานของสัญญาณไฟฟ้าในช่วงทางพอลิเมอไดเมอร์ที่ถูก ประมาณในช่วงของส่วนเบี่ยงเบนมาตรฐานที่ต่างกันออกไป.....	21
รูปที่ 14 ขั้นตอนการทำงานของโปรแกรมสำหรับการแบ่งช่วงลำดับเบสที่ถูกเตรียมด้วยไลบรารีที่มี การสังเคราะห์พอลิยูริดีนและตรวจสอบการแตกหักลำดับเบสที่ปลาย 3'.....	23
รูปที่ 15 สัญญาณไฟฟ้าของอาร์เอ็นเอในแต่ละช่วง A. สัญญาณไฟฟ้าจากลำดับเบสในชุดข้อมูล GFP 10 minutes B. สัญญาณไฟฟ้าจากลำดับเบสในชุดข้อมูล GFP 30 minutes.....	25

รูปที่ 16 กราฟแสดงความยาวหางพอลิอะดีนีนที่ถูกประมาณก่อนขั้นตอนการปรับปรุงข้อมูลความยาวหางพอลิอะดีนีน A. nanopolish และ B. tailfindr..... 28

รูปที่ 17 กราฟแสดงความยาวหางพอลิอะดีนีนที่ถูกประมาณหลังขั้นตอนการปรับปรุงข้อมูลความยาวหางพอลิอะดีนีน A. nanopolish และ B. tailfindr..... 29

รูปที่ 18 ตัวอย่างสัญญาณไฟฟ้าในแต่ละช่วงของลำดับเบส rpsO ..... 33

รูปที่ 19 ตัวอย่างตำแหน่งของนิวคลีโอไทด์ในตำแหน่งของสัญญาณไฟฟ้าในลำดับเบสของยีน rpsO ..... 33

รูปที่ 20 A. รูปแบบการแสดงออกของ rpsO B. โครงสร้างทุติยภูมิสำหรับการยุติการถอดรหัสของ rpsO ที่ถูกทำนาย ..... 34

รูปที่ 21 ตัวอย่างสัญญาณไฟฟ้าในแต่ละช่วงของยีนอื่นๆที่มีรายงานหางพอลิอะดีนีนในงานวิจัยก่อนหน้า..... 35

รูปที่ 22 ตัวอย่างตำแหน่งของนิวคลีโอไทด์ในตำแหน่งของสัญญาณไฟฟ้าในลำดับเบสของยีน rlpQ35

รูปที่ 23 A. รูปแบบการแสดงออกของ rlpQ B. โครงสร้างทุติยภูมิสำหรับการยุติการถอดรหัสของ rlpQ ที่ถูกทำนาย..... 35

รูปที่ 24 ตัวอย่างตำแหน่งของนิวคลีโอไทด์ในตำแหน่งของสัญญาณไฟฟ้าในลำดับเบส apt ..... 36

รูปที่ 25 A. รูปแบบการแสดงออกของ apt B. โครงสร้างทุติยภูมิสำหรับการยุติการถอดรหัสของ apt ที่ถูกทำนาย ..... 36

# บทที่ 1

## บทนำ

### 1.1 ที่มาและความสำคัญ

หางพอลิโอตินีนคือการดัดแปรหรือแต่งเติมของอาร์เอ็นเอหลังการถอดรหัสดีเอ็นเอจากการสังเคราะห์อะดีนีนนิวคลีโอไทด์ต่อกันเป็นสายยาวติดต่อกันที่ปลาย 3' ของอาร์เอ็นเอ (August, Ortiz, & Hurwitz, 1962) โดยมีบทบาทและหน้าที่ที่เกี่ยวข้องกับการแสดงออกของยีนผ่านการควบคุมเสถียรภาพของอาร์เอ็นเอ และมีความแตกต่างกันออกไปในแต่ละชนิดของสิ่งมีชีวิตและประเภทของอาร์เอ็นเอ หางพอลิโอตินีนในยูแคริโอตทำหน้าที่ในการเพิ่มความเสถียรของอาร์เอ็นเอ และมีความยาวสูงได้ถึง 200 นิวคลีโอไทด์ (Zhao, Hyman, & Moore, 1999) ในขณะที่ในหางพอลิโอตินีนในโพรแคริโอต ทำหน้าที่ลดความเสถียรของอาร์เอ็นเอซึ่งนำไปสู่การย่อยสลายของอาร์เอ็นเอ และมีความยาวอยู่ที่ประมาณ 1-20 นิวคลีโอไทด์เท่านั้น (Marujo et al., 2000; O'Hara et al., 1995; Sarkar, 1997)

การศึกษากระบวนการสร้างและความยาวหางพอลิโอตินีนจะนำไปสู่การเข้าใจในการควบคุมการแสดงออกของยีนภายในเซลล์ผ่านการควบคุมความเสถียรของอาร์เอ็นเอ และสามารถนำไปประยุกต์ใช้เพื่อควบคุมการแสดงออกของยีนได้ แต่เนื่องจากหางพอลิโอตินีนในโพรแคริโอตเป็นสายสั้น นอกจากนี้หางพอลิโอตินีนยังเป็นสัญญาณของการย่อยสลายของอาร์เอ็นเอ ทำให้ตรวจพบอาร์เอ็นเอที่มีหางพอลิโอตินีนได้ปริมาณน้อย (Marujo et al., 2000; Sarkar, 1997) อีกทั้งยังมีการปรับแต่งอาร์เอ็นเอที่ปลาย 3' ชนิดอื่นเช่น เฮทเทอโรพอลิเมอร์ (Mohanty & Kushner, 2006) ทำให้การศึกษาหางพอลิโอตินีนในโพรแคริโอตมีความท้าทายสูง วิธีการวัดความยาวหางพอลิโอตินีนของยีนในโพรแคริโอตแบบดั้งเดิมทำได้โดยการคัดเลือกยีนโดยใช้ไพร์เมอร์จำเพาะต่อชนิดของยีนที่นำมาศึกษา และทำการถอดรหัสอาร์เอ็นเอย้อนกลับ และวัดความยาวของสายลำดับเบสด้วยเจลอิเล็กโตรโฟเรซิส (Karnik, Taljanidisz, Sasvari-Szekely, & Sarkar, 1987; Taljanidisz, Karnik, & Sarkar, 1987) ซึ่งสามารถวัดดังกล่าวทำได้เพียงทีละยีน, ใช้แรงงานสูง และระยะเวลานาน ในปัจจุบันได้มีการศึกษาที่มีหางพอลิโอตินีนในโพรแคริโอตระดับทรานสคริปโตม แต่เป็นเพียงการทำนายการมีอยู่ของหางพอลิโอตินีนเท่านั้น ไม่ได้รายงานความยาวหางพอลิโอตินีนในแต่ละยีนโดยตรง อีกทั้งการใช้เทคโนโลยีลำดับเบสยุคใหม่ (Next-Generation sequencing, NGS) ในการศึกษาหางพอลิโอตินีน อาจส่งผลให้เสี่ยงเกิดความคลาดเคลื่อน โดยเฉพาะขั้นตอนที่ต้องพึ่งพาการเพิ่มปริมาณสารพันธุกรรมด้วยวิธีพีซีอาร์ ซึ่งอาจส่งผลให้ความยาวของหางพอลิโอตินีนอาจผิดเพี้ยนไปจากเดิม (Jain,

Olsen, Paten, & Akesson, 2016) อีกทั้งผลลำดับเบสที่ได้เป็นสายสั้น ทำให้อาจไม่ครอบคลุมทั้งโอเปอร์รอนของยีนในโพรแคริโอต

ในงานวิจัยนี้จึงได้ประยุกต์การหาลำดับเบสด้วยอ็อกซ์ฟอร์ดนาโนพอร์มาเพื่อศึกษาหาความยาวของหางพอลิอะดีนีนในอาร์เอ็นเอของโพรแคริโอต โดยใช้ *Escherichia coli* K-12 พันธุ์ป่า (wild type) เป็นสิ่งมีชีวิตต้นแบบสำหรับการศึกษาเพื่อจุดประสงค์ในการทราบความยาวของหางพอลิอะดีนีนในสภาวะปกติของ *E. coli* K-12 พันธุ์ป่า และมุ่งเน้นไปที่การศึกษาหางพอลิอะดีนีนที่ปรากฏที่ปลาย 3' ของยีน เนื่องจากอ็อกซ์ฟอร์ดนาโนพอร์สามารถหาลำดับเบสอาร์เอ็นเอสายยาวได้โดยตรงจากการถอดรหัสลำดับเบสจากสัญญาณไฟฟ้า โดยไม่ต้องทำการเพิ่มจำนวนสารพันธุกรรม ทำให้สามารถหาลำดับเบสความยาวของอาร์เอ็นเอได้โดยตรง และไม่ส่งผลกระทบต่อความยาวดั้งเดิมของหางพอลิอะดีนีน (Galalde et al., 2018; Jain et al., 2016)

อย่างไรก็ตามเนื่องจากในขั้นตอนการเตรียมไลบรารีสำหรับการหาลำดับเบสอาร์เอ็นเอสายยาวโดยตรงด้วยอ็อกซ์ฟอร์ดนาโนพอร์ต้องการอาร์เอ็นเอที่มีความยาวหางพอลิอะดีนีนมากพอที่จะเชื่อมต่อกับอะแด็ปเตอร์พอลิไทมีนสำหรับการหาลำดับเบสได้ แต่หางพอลิอะดีนีนในอาร์เอ็นเอของ *E. coli* มีความยาวไม่เพียงพอที่จะเชื่อมต่อกับอะแด็ปเตอร์พอลิไทมีนได้ เว้นแต่จะทำการสังเคราะห์พอลิอะดีนีนที่ปลาย 3' ของอาร์เอ็นเอ แต่จะส่งผลให้ความยาวหางพอลิอะดีนีนดั้งเดิมผิดเพี้ยน ในการศึกษาวิจัยนี้จึงได้ทำการประยุกต์เตรียมไลบรารีโดยสังเคราะห์พอลิยูริดีนลงไปที่ปลาย 3' ของอาร์เอ็นเอแทนด้วยเอนไซม์ Poly(U) polymerase เพื่อสังเคราะห์พอลิยูริดีนในอาร์เอ็นเอที่ปลาย 3' และใช้อะแด็ปเตอร์พอลิอะดีนีนเพื่อเชื่อมกับพอลิยูริดีนที่ทำการสังเคราะห์ก่อนหน้านี้ ด้วยวิธีนี้ทำให้สามารถอนุรักษ์ความยาวของหางพอลิอะดีนีนดั้งเดิมของอาร์เอ็นเอไว้ได้เมื่อทำการหาลำดับเบส ซึ่งสามารถใช้แทนการเตรียมไลบรารีแบบมาตรฐานโดยสังเคราะห์หางพอลิอะดีนีนตามที่กล่าวข้างต้น และทำการพัฒนากระแสนงานทางคอมพิวเตอร์ในการตรวจพบหางพอลิอะดีนีนด้วยวิธีการเตรียมไลบรารีดังกล่าว โดยรวมถึงการใช้โปรแกรมสำหรับการประมาณความยาวของหางพอลิอะดีนีนของอาร์เอ็นเอที่จากการหาลำดับเบสด้วยอ็อกซ์ฟอร์ดนาโนพอร์โดยเฉพาะ (Krause et al., 2019; Loman, Quick, & Simpson, 2015)

จากการทดสอบกับอาร์เอ็นเอสังเคราะห์ที่มีหางพอลิอะดีนีนความยาวคงที่ เพื่อทำการทดสอบความสามารถเอนไซม์ Poly(U) polymerase ในการเตรียมไลบรารีสำหรับการหาลำดับเบสด้วยอ็อกซ์ฟอร์ดนาโนพอร์แทนวิธีแบบมาตรฐานที่ใช้เอนไซม์ Poly(A) polymerase เพื่อสังเคราะห์พอลิอะดีนีน พบว่าการเตรียมไลบรารีด้วยการสังเคราะห์พอลิยูริดีนสามารถใช้ในการหาลำดับเบสได้ แต่จากการทดสอบประสิทธิภาพของโปรแกรมประมาณความยาวหางพอลิอะดีนีนที่มีอยู่ในปัจจุบัน

(Krause et al., 2019; Loman et al., 2015) ในการประมาณความยาวทางพอลิอะดีนีนสั้นที่มีความยาวคงที่ พบว่าเกิดความคลาดเคลื่อนจากความยาวทางพอลิอะดีนีนที่แท้จริง จึงได้พัฒนากระบวนการเพื่อปรับปรุงข้อมูลจากการกักความยาวทางพอลิอะดีนีนจากความยาวพอลิอะดีนีนในสายลำดับเบส และการกรองสายลำดับเบสที่มีความยาวทางพอลิอะดีนีนยาวผิดปกติออกโดยอ้างอิงจากค่าสถิติ ทำให้ได้ผลลัพธ์ความยาวทางพอลิอะดีนีนที่ใกล้เคียงกับความยาวที่แท้จริงมากขึ้น และมีค่าความคลาดเคลื่อนที่ต่ำลงเมื่อเทียบกับก่อนการใช้กระบวนการปรับปรุงข้อมูล

จากการประยุกต์การเตรียมไลบรารีและกระบวนการเพื่อใช้กับอาร์เอ็นเอจาก *Escherichia coli* ส่งผลให้สามารถประมาณความยาวทางพอลิอะดีนีนได้ครอบคลุมทุกอาร์เอ็นเอของ *Escherichia coli* ที่สามารถสกัดออกมาได้ในการศึกษาครั้งนี้ โดยสามารถตรวจพบอาร์เอ็นเอที่มีความเป็นไปได้ที่จะพบทางพอลิอะดีนีนทั้งในยีนที่มีรายงานก่อนหน้านี้ และยีนที่ยังไม่เคยมีรายงานมาก่อนของทางพอลิอะดีนีน ส่งผลให้ได้ข้อมูลความยาวทางพอลิอะดีนีนดั้งเดิมของ *Escherichia coli* K-12 แบบลำดับเบสสายยาวเป็นครั้งแรก ซึ่งเป็นองค์ความรู้เพื่อนำไปพัฒนาต่อยอด เพื่อประยุกต์ใช้ต่อในงานวิจัยที่เกี่ยวข้องกับการควบคุมการแสดงออกของยีนโพรแคริโอตในอนาคต

## 1.2 วัตถุประสงค์

เพื่อพัฒนากระบวนการในการตรวจหาทางพอลิอะดีนีน และตรวจหายีนที่มีทางพอลิอะดีนีนในโพรแคริโอตโดยใช้ *Escherichia coli* K-12 พันธุ์ป่า เป็นสิ่งมีชีวิตต้นแบบโดยใช้อ็อกซ์ฟอร์ดนาโนพอร์ในการหาลำดับเบสด้วยวิธีการเตรียมไลบรารีโดยการสังเคราะห์พอลิยูรีดีน

## 1.3 ขอบเขตการวิจัย

ทำการพัฒนากระบวนการในการตรวจหาพอลิอะดีนีนโดยใช้สายอาร์เอ็นเอสังเคราะห์จาก green fluorescent protein (GFP) และอาร์เอ็นเอจาก *Escherichia coli* K-12 พันธุ์ป่า ในสภาวะปกติที่ถูกเตรียมไลบรารีด้วยอ็อกซ์ฟอร์ดนาโนพอร์ด้วยการสังเคราะห์พอลิยูรีดีนด้วยเอนไซม์ Poly(U) polymerase

#### 1.4 ประโยชน์ที่คาดว่าจะได้รับ

การศึกษาในครั้งนี้สามารถสร้างกระบวนการศึกษาทางพอลิอะดีนีน และช่วยส่งเสริมองค์ความรู้พื้นฐานของความยาวพอลิอะดีนีนในโพรแคริโอตซึ่งไม่เคยได้ศึกษาความยาวทางพอลิอะดีนีนดั้งเดิมโดยการหาลำดับเบสสายยาวด้วยอ็อกซ์ฟอร์ดนาโนพอร์มาก่อน และเนื่องจากความยาวของทางพอลิอะดีนีนมีหน้าที่ควบคุมความเสถียรของอาร์เอ็นเอ จึงสามารถนำไปใช้ประยุกต์กับสิ่งมีชีวิตโพรแคริโอตสายพันธุ์อื่นได้ เพื่อนำไปพัฒนาวิธีทางเลือกสำหรับการควบคุมการแสดงออกของยีนในโพรแคริโอตและประยุกต์ใช้กับหลากหลายสาขาเช่น การแพทย์ เทคโนโลยีชีวภาพ และการเกษตร เป็นต้น



## บทที่ 2

### ทบทวนวรรณกรรม

#### 2.1 ทางพอลิอะดีนีน

ทางพอลิอะดีนีนเป็นฮอโมพอลิเมอร์ชนิดหนึ่งจากการตัดแปรรหัสอาร์เอ็นเอหลังจากการกระบวนการถอดรหัส ซึ่งถูกค้นพบในปี ค.ศ. 1960 (Edmonds & Abrams, 1960) การสังเคราะห์ทางพอลิอะดีนีนมักพบในแมสเซ็นเจอร์อาร์เอ็นเอก่อนถูกแปลรหัสเป็นโปรตีนต่อไป โดยอาร์เอ็นเอจะถูกเติมอะดีนีนนิวคลีโอไทด์เข้ากันเป็นสายยาวที่ปลาย 3' ด้วยเอนไซม์ Poly(A) polymerase และใช้อะดีโนซีนไตรฟอสเฟต (adenosine triphosphate, ATP) เป็นสารตั้งต้น (August et al., 1962) กระบวนการนี้สามารถพบได้ทั้งในสิ่งมีชีวิตยูแคริโอต (Hadjivassiliou & Brawerman, 1966), โพรแคริโอต (Nakazato, Venkatesan, & Edmonds, 1975) และอาร์เคีย (Brown & Reeve, 1985) และพบได้ทั้งในไซโทพลาสซึมและออร์แกเนลล์อย่างไมโทคอนเดรีย (Nagaike, Suzuki, & Ueda, 2008) และพลาสติด (Komine, Kwong, Anguera, Schuster, & Stern, 2000) อย่างไรก็ตามไม่ใช่โมเลกุลของทุกอาร์เอ็นเอจะมีทางพอลิอะดีนีน เช่น ฮิสโตนอาร์เอ็นเอของสัตว์เลี้ยงลูกด้วยนม (Dávila López & Samuelsson, 2008) โดยหน้าที่หลักของทางพอลิอะดีนีนจะเกี่ยวข้องกับการพัฒนาของอาร์เอ็นเอ การส่งออกของอาร์เอ็นเอจากนิวเคลียสสู่ไซโทพลาสซึม (Hector et al., 2002) และการควบคุมการแสดงออกของยีน (Guhaniyogi & Brewer, 2001)

#### 2.2 ความแตกต่างของการสังเคราะห์และหน้าที่ในทางพอลิอะดีนีนในโพรแคริโอตและยูแคริโอต

ทางพอลิอะดีนีนในโพรแคริโอตและยูแคริโอตนั้นมีความแตกต่างกันทั้งในกระบวนการสังเคราะห์ตลอดไปจนถึงความยาวและหน้าที่ กระบวนการสร้างทางพอลิอะดีนีนในยูแคริโอตนั้นต้องการลำดับเบสอนุรักษ์เพื่อเป็นสัญญาณให้เอนไซม์ทำการสังเคราะห์สายทางพอลิอะดีนีนที่ปลาย 3' (Proudfoot, 1991) ในขณะที่โพรแคริโอตลำดับเบสอนุรักษ์ที่เป็นสัญญาณสำหรับกระบวนการสังเคราะห์ทางพอลิอะดีนีนยังไม่สามารถระบุได้แน่ชัด (Sarkar, 1997) ความยาวของทางพอลิอะดีนีนในยูแคริโอตมักมีความยาวตั้งแต่ 80 ถึง 200 นิวคลีโอไทด์ (Zhao et al., 1999) และมักทำให้อาร์เอ็นเอมีความเสถียรมากขึ้น โดยเพิ่มครึ่งชีวิตของอาร์เอ็นเอ (Guhaniyogi & Brewer, 2001) และช่วยป้องกันการย่อยสลายระหว่างการลำเลียงสายอาร์เอ็นเอจากนิวเคลียสสู่ไซโทพลาสซึมเพื่อแปลรหัสอาร์เอ็นเอเป็นโปรตีน (Hector et al., 2002) ในขณะที่โพรแคริโอตมักมีความยาวสั้นกว่ามาก



เพียงแค่ประมาณ 1 ถึง 20 นิวคลีโอไทด์โดยส่วนใหญ่ และสามารถพบได้สูงสุดเพียง 50 นิวคลีโอไทด์ (Sarkar, 1997) และมักเป็นสัญญาณที่ก่อให้เกิดความไม่เสถียรแก่อาร์เอ็นเอ นำไปสู่กระบวนการสลายของอาร์เอ็นเอ (O'Hara et al., 1995) โดยหน้าที่อื่นของหางพอลิอะดีนีนในโพรแคริโอต นอกจากเป็นสัญญาณสำหรับการย่อยสลายอาร์เอ็นเอแล้วยังเกี่ยวข้องกับการจับโปรตีน ribosomal subunit protein S1 ซึ่งคล้ายกับโปรตีน poly(A) binding proteins ที่จับกับหางพอลิอะดีนีนในอาร์เอ็นเอของยูแคริโอตที่ช่วยเพิ่มความเสถียรอาร์เอ็นเอและช่วยในกระบวนการเริ่มต้นการแปลรหัส (Munroe & Jacobson, 1990) แม้ยังไม่สามารถยืนยันถึงบทบาทและหน้าที่ในการจับของโปรตีน ribosomal subunit protein S1 กับหางพอลิอะดีนีนในโพรแคริโอตได้แน่ชัด แต่คาดว่ามีส่วนในการช่วยกระตุ้นการเริ่มต้นกระบวนการแปลรหัสเช่นเดียวกับในยูแคริโอต (Feng, Huang, Liao, & Cohen, 2001) อีกทั้งคาดว่าหางพอลิอะดีนีนอาจมีส่วนเกี่ยวข้องกับการควบคุมจำนวนพลาสมิด เนื่องจากในการศึกษาก่อนหน้าพบว่าการกลายพันธุ์ของยีน *pcnB* ที่ถูกแปลรหัสเป็นเอนไซม์ PAP I สำหรับการสังเคราะห์หางพอลิอะดีนีนในโพรแคริโอตทำให้พลาสมิด ColE1 มีจำนวนลดลง (Lopilato, Bortner, & Beckwith, 1986)

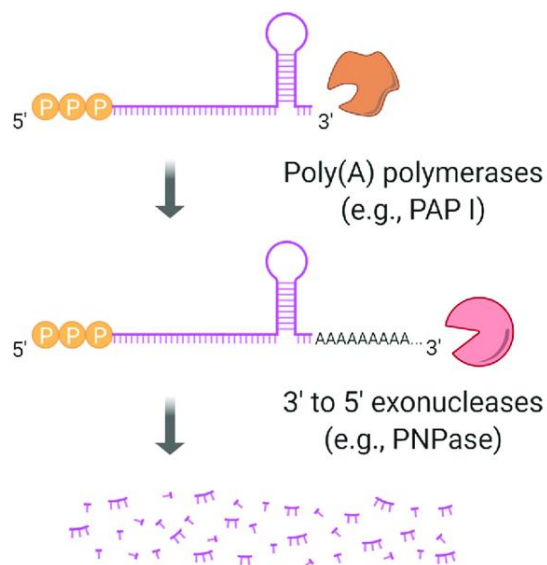
การสังเคราะห์หางพอลิอะดีนีนอาศัยเอนไซม์ Polynucleotide adenylyltransferase หรือ Poly(A) polymerase ที่มีคุณสมบัติสังเคราะห์อะดีนีนนิวคลีโอไทด์ที่ปลาย 3' ของอาร์เอ็นเอ โดยใน *E. coli* มีเอนไซม์ที่เกี่ยวข้อง 2 ชนิด คือ Poly(A) polymerase I (PAP I) (Edmonds & Abrams, 1960) ซึ่งหางพอลิอะดีนีนมักถูกสังเคราะห์ด้วยเอนไซม์นี้ โดยถูกแปลรหัสมาจากยีน *pcnB* และอีกเอนไซม์หนึ่งคือ Poly(A) polymerase II (PAP II) ที่ถูกแปลรหัสมาจากยีน *f310* (Kalapos, Cao, Kushner, & Sarkar, 1994) โดยยังอยู่ขั้นตอนของการศึกษาเพื่อเข้าใจบทบาทและการทำงานของเอนไซม์ชนิดนี้ การสังเคราะห์พอลิอะดีนีนในโพรแคริโอตมักเกิดขึ้นหลังการถอดรหัสทันที เนื่องจากโพรแคริโอตไม่มีเยื่อหุ้มนิวเคลียสและมีการแสดงออกแบบโอเปอเรรอนคือมีการแสดงออกหลายยีนออกมาพร้อมกันจากหนึ่งโปรโมเตอร์ภายในการแสดงออกหนึ่งครั้ง จึงต้องมีการควบคุมการแสดงออกอย่างทันทั่วทั้งที่ เพื่อไม่ให้มีแสดงออกของยีนมากเกินไปโดยการย่อยสลายอาร์เอ็นเอด้วยเอนไซม์ไรโบนิวคลีเอสซึ่งมักเกิดขึ้นหลังการสังเคราะห์หางพอลิอะดีนีนเสร็จสิ้น (Sarkar, 1997)

โดยเอนไซม์ที่ทำหน้าที่ในการย่อยสลายอาร์เอ็นเอในโพรแคริโอตมีสองประเภท ประเภทแรกคือเอ็กโซไรโบนิวคลีเอสที่ย่อยอาร์เอ็นเอจากปลาย 3' เช่น PNPase RNase R RNase II (Marujo et al., 2000) และประเภทที่สองคือเอนไซม์เอ็นโดไรโบนิวคลีเอสที่ย่อยสลายอาร์เอ็นเอภายในสาย เช่น RNase E (Goodrich & Steege, 1999) อีกทั้งอาร์เอ็นเอสามารถถูกย่อยสลายจากการจับกับ

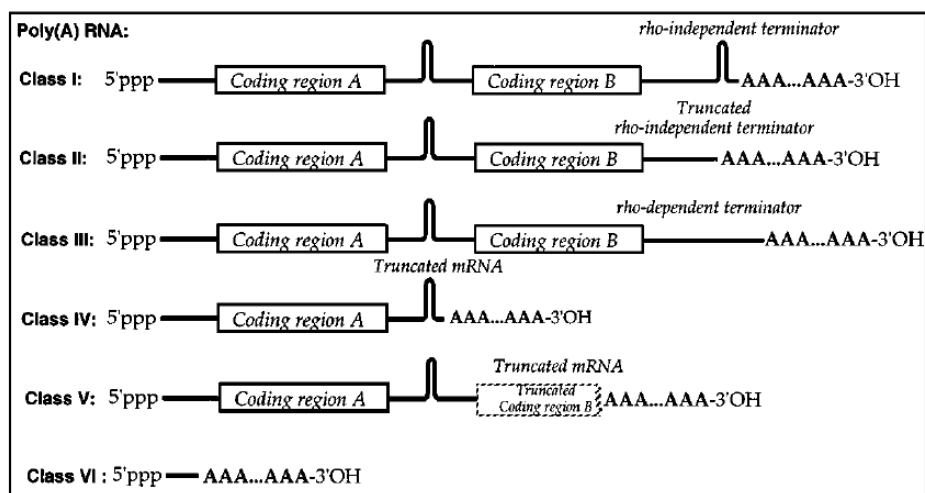
degradosome ซึ่งประกอบด้วย RNase E Rh1B PNPase และโปรตีนอื่นๆ สำหรับย่อยสลายอาร์เอ็นเอ (Miczak, Kaberdin, Wei, & Lin-Chao, 1996)

ตัวอย่างการควบคุมการแสดงออกของยีนผ่านทางพอลิอะดีนีนใน *E. coli* คือยีน *CspA* ถูกแปลรหัสเป็น cold shock protein ที่จะมีการแสดงออกมากขึ้นเพื่อตอบสนองต่อสภาพแวดล้อมเมื่ออุณหภูมิลดลง ทำให้สามารถพบอาร์เอ็นเอที่ถูกแปลรหัสเป็นโปรตีนกลุ่มนี้จำนวนมากในสภาวะดังกล่าว แต่เมื่อสิ้นสุดการกระบวนการปรับตัว หรืออุณหภูมิกลับสู่สภาวะปกติ จะพบการแสดงออกของอาร์เอ็นเอในกลุ่ม cold shock protein อย่าง *CspA* น้อยลง ในขณะที่ *pnp* ที่ถูกแปลรหัสเป็นเอนไซม์ PNPase สำหรับการย่อยสลายอาร์เอ็นเอผ่านทางพอลิอะดีนีนมีการแสดงออกมากขึ้น (Yamanaka & Inouye, 2001) เพื่อควบคุมการแสดงออกของ *CspA*

ทางพอลิอะดีนีนมักพบในอาร์เอ็นเอที่มีโครงสร้างทุติยภูมิสำหรับการสิ้นสุดการถอดรหัส (Rho-independent terminator) ที่ปลาย 3' เนื่องจากสามารถยับยั้งการทำงานของเอนไซม์เอ็กโซไรโบนิวคลีเอสได้ จึงต้องทำการสังเคราะห์ทางพอลิอะดีนีนเพื่อให้เอนไซม์เอ็กโซไรโบนิวคลีเอสสามารถจับกับสายอาร์เอ็นเอเพื่อย่อยสลายโครงสร้างทุติยภูมิสำหรับการหยุดการถอดรหัสทำให้อาร์เอ็นเอสามารถถูกย่อยสลายได้ดังรูปที่ 1 (Vargas-Blanco & Shell, 2020) ซึ่งพบว่าการสังเคราะห์ทางพอลิอะดีนีนสามารถเกิดได้มากกว่าหนึ่งครั้ง จนกว่าโครงสร้างทุติยภูมิของอาร์เอ็นเอจะคลายตัว (Hajnsdorf & Kaberdin, 2018; Sarkar, 1997) อย่างไรก็ตามมีรายงานการพบทางพอลิอะดีนีนตำแหน่งอื่นของอาร์เอ็นเอได้ดังรูปที่ 2 (Sarkar, 1997) เช่น ยีน *lacZ* ใน *E. coli* ที่ถูกแปลรหัสเป็นโปรตีน beta-galactosidase โดยมีการแสดงออกแบบโอเพอร์รอนและอยู่ลำดับแรกของการแสดงออก (รูปที่ 2 Class IV) โดยพบทางพอลิอะดีนีนอยู่หลังโครงสร้างทุติยภูมิภายในสายลำดับเบส (Sarkar, 1997) และบางยีนที่ไม่พบโครงสร้างทุติยภูมิสำหรับการหยุดการถอดรหัส (รูปที่ 2 Class III) เช่น *tolA*, *mhc* และ *galF* เป็นต้น (Mohanty & Kushner, 2006)



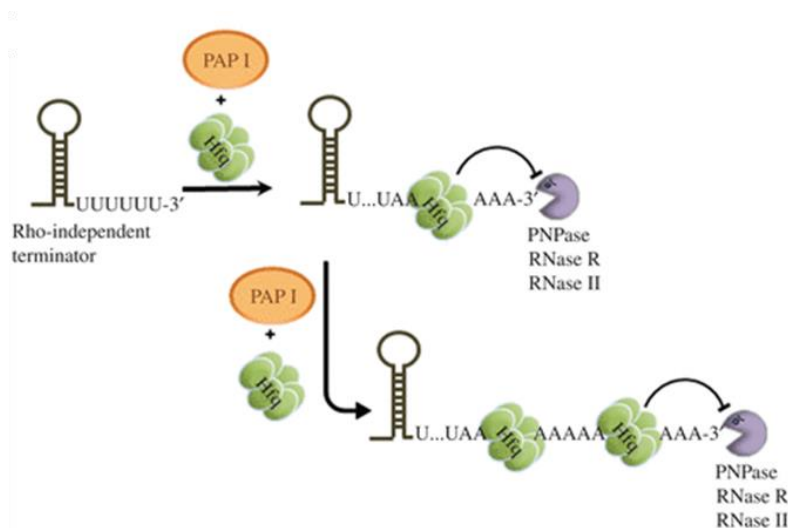
รูปที่ 1 การสังเคราะห์หางพอลิอะดีนีนเพื่อย่อยสลายอาร์เอ็นเอในสิ่งมีชีวิตโพรแคริโอต  
(Vargas-Blanco & Shell, 2020)



รูปที่ 2 รูปแบบของอาร์เอ็นเอตามตำแหน่งของหางพอลิอะดีนีนที่ถูกเคยถูกรายงาน  
(Sarkar, 1997)

ในปกติแล้วหางพอลิอะดีนีนในอาร์เอ็นเอของโพรแคริโอตมีความยาวเพียง 1 ถึง 20 นิวคลีโอไทด์ตามที่กล่าวข้างต้น อย่างไรก็ตามมีรายงานการพบหางพอลิอะดีนีนสายยาวถึง 50 นิวคลีโอไทด์ในอาร์เอ็นเอของบางยีน เช่น *rpsO* ใน *E. coli* (Hajnsdorf & Régnier, 2000) โดยพบว่าหางพอลิอะดีนีนได้ไปจับกับโปรตีน Hfq ที่บริเวณปลาย 3' ทำให้สามารถยับยั้งการจับของเอนไซม์เอ็กโซไรโบนิวคลีเอสไม่ให้ย่อยสลายอาร์เอ็นเอเช่นเดียวกับโครงสร้างทุติยภูมิสำหรับการยุติการถอดรหัสทำให้หาง

พอลิอะดีนีนมีความยาวมากขึ้นดังรูปที่ 3 (Walsh et al., 2001) โดยรูปแบบการจับของ Hfq มักจับกับอาร์เอ็นเอในบริเวณที่มีอะดีนีนหรือยูรีดีนอยู่มาก เช่นหางพอลิอะดีนีน หรือ ออลิโกยูรีดีน (Hajnsdorf & Kaberdin, 2018; Otaka, Ishikawa, Morita, & Aiba, 2011; Senear & Steitz, 1976)



รูปที่ 3 การสังเคราะห์หางพอลิอะดีนีนสายยาวในโพรแคริโอตโดยการจับกับโปรตีน Hfq ที่มีความสามารถช่วยยับยั้งการทำงานของเอนไซม์เอ็กโซไรโบนิวคลีเอส (Hajnsdorf & Kaberdin, 2018)

### 2.3 การศึกษาหางพอลิอะดีนีนในโพรแคริโอต

แม้ว่าหางพอลิอะดีนีนจะถูกค้นพบตั้งแต่ปี ค.ศ. 1960 (Edmonds & Abrams, 1960) แต่การศึกษาหางพอลิอะดีนีนในโพรแคริโอตอย่าง *E. coli* (Nakazato et al., 1975) และ *Caulobacter crescentus* (Ohta, Sanders, & Newton, 1975) ถูกศึกษาในอีก 15 ปีต่อมา ในการศึกษาเกี่ยวกับหางพอลิอะดีนีนแรกเริ่มมีการใช้ oligo(dT)-cellulose chromatography เป็นเม็دتรงกลมขนาดเล็กที่มีสายนิวคลีโอไทด์ไทมีนติดอยู่ ซึ่งสามารถจับกับอาร์เอ็นเอที่มีหางพอลิอะดีนีนได้ และทำการคัดแยกอาร์เอ็นเอที่มีหางพอลิอะดีนีนมาศึกษาเพื่อตรวจสอบยีน และวัดขนาดความยาวหางพอลิอะดีนีนของอาร์เอ็นเอในยีนโดยเปรียบเทียบความยาวลำดับเบสสายพันธุที่มีและไม่มีหางพอลิอะดีนีน โดยมีการทำการศึกษาคู่ขนานใช้ยีนที่มีคุณลักษณะต่างกัน 2 ชนิดที่พบหางพอลิอะดีนีนในอาร์เอ็นเอ คือ เช่น *lpp* ที่ถูกแปลรหัสเป็น outer membrane lipoprotein โดยมีการแสดงออกแบบเดี่ยวไม่เป็นโอเปอร์รอน และมีความเสถียรทำให้สามารถสกัดออกมาได้ปริมาณค่อนข้างมาก และ ยีน *trpA* ที่ถูกแปลรหัสเป็น alpha-subunit of tryptophan synthase โดยมีการแสดงออกแบบโอเปอร์รอนและอยู่ลำดับสุดท้ายของการแสดงออก และเป็นหนึ่งในยีนที่มีครึ่งชีวิตที่สั้นมาก

ที่สุด ใน *E. coli* (Sarkar, 1997) โดยพบว่าผลการศึกษาทางพอลิอะดีนีนทั้ง *lpp* และ *trpA* ก่อนข้าง คล้ายคลึงกันโดย 40–50% ของอาร์เอ็นเอของยีน *lpp* และ *trpA* พบทางพอลิอะดีนีน และมีความ ยาวอยู่ในช่วง 10-20 นิวคลีโอไทด์ (Karnik et al., 1987; Taljanidisz et al., 1987)

ในการศึกษาทางพอลิอะดีนีนในโพรแคริโอตมักไม่ใช้พันธุ์ป่าของแบคทีเรียชนิดนั้น แต่เป็น สายพันธุ์ที่ถูกทำยีนนิ่งออกเอาต์ที่เกี่ยวข้องกับผลิตภัณฑ์เอนไซม์ในการย่อยสลายอาร์เอ็นเอที่มีทางพอลิอะ ดีนีน เช่น *pnp* ที่ผลิต PNPase เพื่อไม่ให้อาร์เอ็นเอที่มีทางพอลิอะดีนีนเข้าสู่กระบวนการย่อยสลาย ทำให้สามารถรักษาความยาวทางพอลิอะดีนีนไว้ได้ โดยในการศึกษาพอลิอะดีนีนยุคแรกเริ่มสามารถ ทำได้เพียงที่ละยีน และใช้ทรัพยากรและเวลามากทำให้มีเพียงไม่กี่ยีนของ *E. coli* ที่ถูกรายงานความ ยาวของทางพอลิอะดีนีน เช่น *lpp*, *trpA* และ *rpsO* เป็นต้น (Hajnsdorf, Braun, Haugel-Nielsen, & Régnier, 1995; Karnik et al., 1987; Senear & Steitz, 1976; Taljanidisz et al., 1987) ต่อมาจึงได้มีการพัฒนาในการตรวจหายีนที่มีทางพอลิอะดีนีนระดับจีโนมด้วยวิธีซีดีเอ็นเอมาโคร อาร์เรย์ (cDNA macroarray) และทำการเปรียบเทียบพันธุ์ป่า กับพันธุ์ *pcnB+* ที่ถูกเหนี่ยวนำให้ *pcnB* มีการแสดงออกเพิ่มมากขึ้นส่งผลให้เอนไซม์ PAP I ที่สังเคราะห์ทางพอลิอะดีนีนถูกแปลรหัสมากขึ้น ทำให้เพิ่มโอกาสในการเจออาร์เอ็นเอในขณะที่มีทางพอลิอะดีนีนอยู่ และได้รายงานยีนที่มีทางพอลิอะ ดีนีนมากกว่า 100 ยีน ซึ่งพบว่าโดยส่วนใหญ่กว่า 70% เป็นยีนหรือโอเปอเรรอนที่มีโครงสร้างทุติยภูมิ สำหรับการสิ้นสุดการถอดรหัสจะถูกสังเคราะห์ทางพอลิอะดีนีน ในการศึกษาแล้วยังพบว่าไม่ใช่แค่ทาง พอลิอะดีนีนที่เป็นการแต่งเติมหลังการถอดรหัสที่ปลาย 3' เท่านั้น แต่ยังพบการแต่งเติมอื่นที่เรียกว่า เฮทเทอโรพอลิเมอร์ซึ่งถูกสังเคราะห์ด้วยเอนไซม์ PNPase ที่แม้ว่าจะจะเป็นเอนไซม์ที่ทำหน้าที่ในการ ย่อยสลายอาร์เอ็นเอก็ตาม โดยสามารถสังเคราะห์นิวคลีโอไทด์ได้ทั้ง 4 ชนิดสู่ปลาย 3' ของอาร์เอ็นเอ ได้ถึง 100 นิวคลีโอไทด์ (Mohanty & Kushner, 2006) ซึ่งสูงกว่าการสังเคราะห์ทางพอลิอะดีนีน ด้วย PAP I ที่มีความยาวได้สูงสุดเพียง 50 นิวคลีโอไทด์และต้องการโปรตีน Hfq มาช่วยจับกับสาย ทางพอลิอะดีนีนที่กำลังถูกสังเคราะห์เพื่อยับยั้งไม่ให้เอนไซม์เฮ็กโซไรโบนิวคลีโอเอสสามารถย่อยสลาย อาร์เอ็นเอได้ (Walsh et al., 2001) อัตราส่วนนิวคลีโอไทด์ของเฮทเทอโรพอลิเมอร์พบปริมาณอะ ดีนีนนิวคลีโอไทด์สูงสุดที่สุด โดยยีนที่อาร์เอ็นเอถูกสังเคราะห์เฮทเทอโรพอลิเมอร์ที่ปลาย 3' ที่ได้รับการ ยืนยันแล้วคือ *pcnB* ส่วนอาร์เอ็นเอของยีนอื่นสามารถพบเฮทเทอโรพอลิเมอร์ในสัดส่วนน้อยกว่า ทางพอลิอะดีนีน เช่น *ompA* อย่างไรก็ตามบทบาทและหน้าที่ของการแต่งเติมในรูปแบบนี้ยังไม่ทราบ แน่ชัด คาดว่าอาจเกี่ยวข้องกับการย่อยสลายของอาร์เอ็นเอเช่นเดียวกับทางพอลิอะดีนีนที่เป็นฮอโม พอลิเมอร์ (Mohanty & Kushner, 2006)

ต่อมาได้มีการประยุกต์ใช้เทคโนโลยีการหาลำดับเบสยุคใหม่ (Next Generation Sequencing, NGS) และนำมาใช้กับการตรวจหายีนที่มีทางพอลิอะดีนีน โดยทำการหาลำดับเบสอาร์

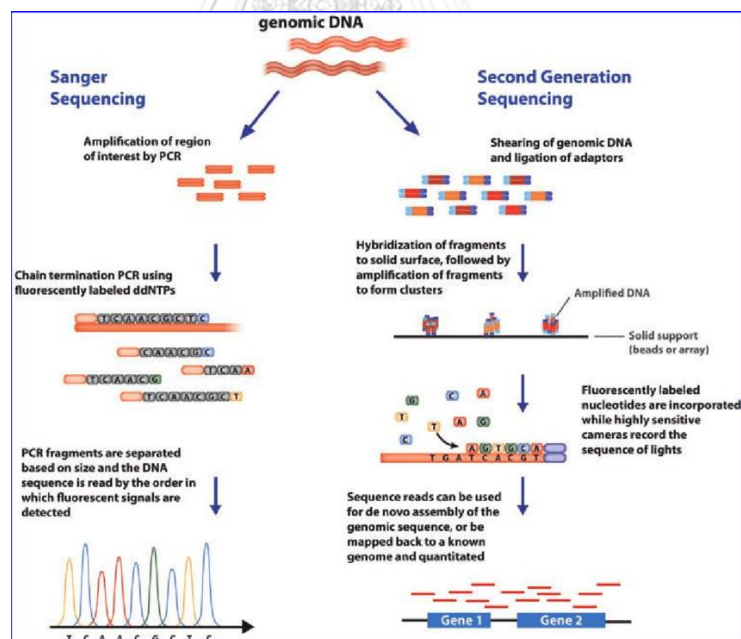
เอ็นเอสสายสั้นซึ่งทำการศึกษาทั้งทรานสคริปโตม และทำการเปรียบเทียบระหว่างพันธุ์ป่า และสายพันธุ์ *pcnB*- ที่ถูกกลายพันธุ์ทำให้มีการแสดงออกที่น้อยลง ส่งผลให้พบอาร์เอ็นเอที่มีทางพอลิอะดีนีนที่น้อยลง นอกจากนี้ยังได้มีการพัฒนาวิธีในการตรวจหาทางพอลิอะดีนีนโดยการประมาณความเป็นได้ของสายทางพอลิอะดีนีนจากนิวคลีโอไทด์ที่ไม่ปรากฏบนจีโนมอ้างอิงที่ปลาย 3' ซึ่งหมายถึงอาจถูกแต่งเติมเข้ามาหลังจากขั้นตอนการถอดรหัส และได้รายงานถึงรายชื่อของยีนที่มีโอกาสพบทางพอลิอะดีนีนทั้งที่พบใหม่และพบในงานวิจัยก่อนหน้า รวมถึงรายงานกลุ่มหน้าที่ของยีน (Functional annotation) ที่พบทางพอลิอะดีนีนที่มีการแสดงออกเปลี่ยนแปลงไปทั้งแสดงออกมาขึ้น และน้อยลง เมื่อ *pcnB* มีการแสดงออกน้อยลง (Maes et al., 2017)

## 2.4 เทคโนโลยีการหาลำดับเบสด้วยอ็อกซ์ฟอร์ดนาโนพอร์

การหาลำดับเบสของนิวคลีโอไทด์คือกระบวนการที่ประยุกต์นำเทคโนโลยีเพื่อมาใช้ในการหาลำดับเบสของสายนิวคลีโอไทด์ดีเอ็นเอหรืออาร์เอ็นเอเพื่อศึกษาข้อมูลพันธุกรรมของสิ่งมีชีวิตเพื่อประยุกต์ใช้ในทางการแพทย์, การเกษตร และเทคโนโลยีชีวภาพ โดยเทคโนโลยีการหาลำดับเบสในยุคแรกเริ่มถูกพัฒนาโดย Frederick Sanger ในปี ค.ศ. 1977 โดยใช้หลักการต่อสายลำดับเบสหลายขนาดบนเจลพอลิอะครีลาไมด์จากลำดับเบสที่ถูกหยุดการสังเคราะห์ลำดับเบสด้วยไดดีออกซีนิวคลีโอไทด์ไตรฟอสเฟต (dideoxynucleotide triphosphates, ddNTPs) แต่ละชนิด ซึ่งวิธีการนี้เรียกว่าเทคโนโลยีการหาลำดับเบสแบบแซงเจอร์ (sanger sequencing) หรือการหาลำดับเบสในยุคที่หนึ่ง (First-generation sequencing) (Sanger, Nicklen, & Coulson, 1977) โดยสามารถทำได้ทั้งในดีเอ็นเอและอาร์เอ็นเอ อย่างไรก็ตามข้อจำกัดของการหาลำดับเบสด้วยแซงเจอร์คือทำได้ทีละน้อย, ใช้เวลานาน และมีค่าใช้จ่ายสูง (Crossley et al., 2020)

ต่อมาเทคโนโลยีการหาลำดับเบสได้มีการพัฒนาเพื่อหาที่สามารถวิเคราะห์สายลำดับเบสหลายๆสายพร้อมกันจำนวนมากในเวลาที่รวดเร็ว เรียกว่าการหาลำดับเบสในยุคใหม่ (Next-generation sequencing, NGS) โดยทำการตัดสายดีเอ็นเอออกมาเป็นชิ้นเล็กๆ และเพิ่มปริมาณลำดับเบสด้วยวิธีการพีซีอาร์ จากนั้นทำการหาลำดับเบสซึ่งมีหลากหลายวิธี เช่นการสังเคราะห์สายลำดับเบสคู่ขนานโดยใช้นิวคลีโอไทด์ที่ติดฉลากเรืองแสงฟลูออเรสเซนต์ที่ต่างกันและบันทึกค่าแสงระหว่างสังเคราะห์ออกมาเป็นสายลำดับเบส เรียกว่า bridge amplification (Mardis, 2008) หรือการวัดการเปลี่ยนแปลงค่า pH จากการปล่อยโปรตรอนระหว่างการสังเคราะห์นิวคลีโอไทด์สายคู่ขนาน ซึ่งจะแตกต่างกันไปในนิวคลีโอไทด์แต่ละชนิด เรียกว่า Ion torrent sequencing (Merriman & Rothberg, 2012) ซึ่งสายลำดับเบสที่ได้จะเป็นสายสั้น และต้องการประกอบสาย

ลำดับเบส (assembly) ให้เป็นลำดับเบสสายยาวจากช่วงปลายลำดับเบสที่ซ้ำกัน โดยความแตกต่างระหว่างการหาลำดับเบสแบบแซงเจอร์ และ NGS สรุปในรูปแบบที่ 4 (Bunnik & Le Roch, 2013) โดยในปัจจุบันมี 4 บริษัทที่พัฒนาแพลตฟอร์มสำหรับการหาลำดับเบส คือ ThermoFisher (Ion Torrent), illumina (Solexa/HiSeq/MiSeq), AB SOLiD (Life Technology) และ 454 (Roche) เนื่องจากแต่ละรูปแบบมีการเตรียมไลบรารีและการหาลำดับเบสที่ต่างกัน ทำให้ผลลัพธ์ที่ได้มีความต่างกัน เช่น จำนวนสายดีเอ็นเอ, ความยาวสูงสุด, ระดับความถูกต้องของข้อมูล และเวลาที่ใช้ในการหาลำดับเบส ทำให้ต้องเลือกใช้เทคโนโลยีที่เหมาะสมต่อการประยุกต์ใช้งาน (Kuczynski et al., 2012) การหาลำดับเบสอาร์เอ็นเอด้วยเทคโนโลยี NGS จะแตกต่างจากดีเอ็นเอ เพราะอาร์เอ็นเอส่วนใหญ่เป็นโมเลกุลสายเดี่ยว และมีโครงสร้างพื้นฐานเป็นโรโบนิวคลีโอไทด์ ไม่ใช่ดีออกซีโรโบนิวคลีโอไทด์แบบดีเอ็นเอ ทำให้การหาลำดับเบสด้วยอาร์เอ็นเอมีการเตรียมไลบรารีที่ยุงยากกว่า โดยต้องการการถอดรหัสย้อนกลับ, การสังเคราะห์สายลำดับเบสตรงข้ามเพื่อให้เป็นสายคู่โดยใช้ไพรเมอร์แบบสุ่ม และการเพิ่มปริมาณลำดับเบสซึ่งอาจเกิดความคลาดเคลื่อนได้ (Z. Wang, Gerstein, & Snyder, 2009)



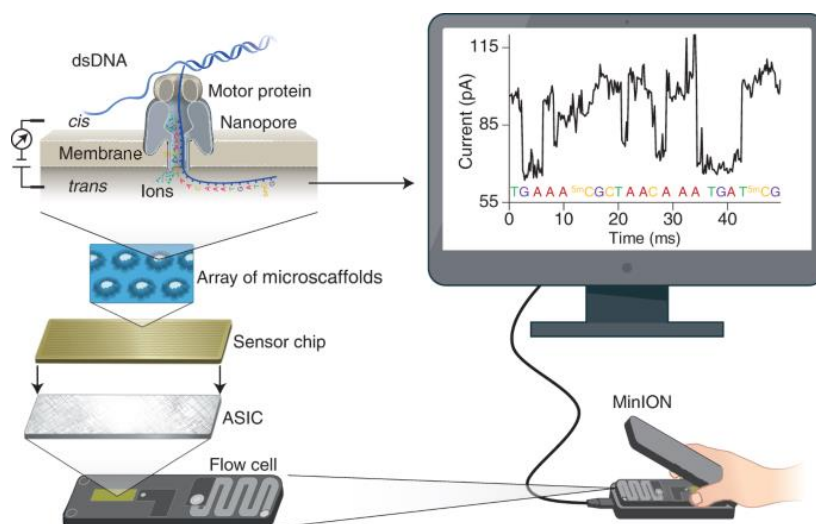
รูปที่ 4 ความแตกต่างระหว่างการหาลำดับเบสแบบแซงเจอร์และการหาลำดับเบสแบบ NGS (Bunnik & Le Roch, 2013)

ในปัจจุบันเทคโนโลยีการหาลำดับเบสได้เข้าสู่ยุคที่สาม (Third-generation sequencing, TGS) โดยไม่จำเป็นต้องทำการพีซีอาร์เพื่อเพิ่มปริมาณลำดับเบส แต่ใช้โมเลกุลดั้งเดิมที่ไม่ได้สังเคราะห์ขึ้นเพื่อหาลำดับเบสจากโมเลกุลเดี่ยวแต่ละโมเลกุล ดังนั้นข้อมูลลำดับเบสที่ได้จึงไม่มีความคลาดเคลื่อนอันเกิดจากการเพิ่มปริมาณลำดับเบส นอกจากนี้ความยาวลำดับเบสที่ได้รับจะมีความยาวที่สูงขึ้นมากซึ่งสามารถสูงได้ถึงสิบกิโลเบส จึงถูกเรียกอีกชื่อว่าการหาลำดับเบสสายยาว (Long read sequencing) นอกจากนี้ยังใช้ในเวลาในการหาลำดับเบสสั้นลงมากโดยมักไม่เกิน 1 วัน ทำให้สามารถทำการวิเคราะห์ได้ตามเวลาจริง (real-time) อย่างไรก็ตามข้อเสียหลักของการหาลำดับเบสสายยาวคือความคลาดเคลื่อนสูงกว่าแบบการหาลำดับเบสในยุคก่อนหน้าอย่างมาก (van Dijk, Jaszczyszyn, Naquin, & Thermes, 2018)

การหาลำดับเบสด้วยอ็อกซ์ฟอร์ดนาโนพอร์เป็นหนึ่งในการหาลำดับเบสในยุคที่สามพัฒนาโดยบริษัท Oxford Nanopore Technologies (ONT) โดยการประยุกต์ใช้ช่องในระดับนาโน เช่น แอลฟาฮีโมไลซิน (Alpha hemolysin) ซึ่งมีคุณสมบัติเป็นตัวรับรู้ชีวภาพ (biosensor) (van Dijk et al., 2018) ที่ติดอยู่บนแผ่นโพลีเมอร์ที่มีความต้านต่อกระแสไฟฟ้า โดยมีการปล่อยกระแสไฟฟ้าที่มีแรงดันไฟฟ้าคงที่ให้ไหลผ่านรูนาโนพอร์ และเมื่อลำดับเบสที่ถูกเตรียมไลบรารีสำหรับการหาลำดับเบสด้วยอ็อกซ์ฟอร์ดนาโนพอร์เข้าสู่รูนาโนพอร์จากปลาย 3' โดยส่วนมอเตอร์โปรตีนที่มีคุณสมบัติคล้ายคลึงกับเอนไซม์ helicase จะทำการแยกสายลำดับเบสคู่ให้กลายเป็นสายลำดับเบสเดี่ยว ทำให้ลำดับเบสผ่านเข้าไปในรูนาโนพอร์ได้ โดยลำดับเบสสายเดี่ยวที่มีประจุลบจะไหลจากฝั่งที่มีประจุลบ (*cis*) ไปยังประจุบวก (*trans*) (Y. Wang, Zhao, Bollas, Wang, & Au, 2021) เมื่อแต่ละลำดับนิวคลีโอไทด์ไหลผ่านรูนาโนพอร์จะเกิดการเปลี่ยนแปลงของกระแสไฟฟ้าและตัวรับรู้ชีวภาพจะจับเก็บสัญญาณไฟฟ้าเหล่านั้นบันทึกไว้ในไฟล์ชื่อ "FAST5" จากนั้นสัญญาณไฟฟ้าที่ถูกจับเก็บและจะถูกแปลงเป็นลำดับเบส (basecalling) ด้วยซอฟต์แวร์ "basecaller" เช่น Albacore และ Guppy ซึ่งจับเก็บในไฟล์ "FASTQ" โดยในปัจจุบันมีเพียง Guppy เท่านั้นที่ถูกยังถูกพัฒนาต่อเนื่อง เนื่องจากสามารถทำงานได้บนหน่วยประมวลผลภาพ ซึ่งมีความเร็วในการแปลงสัญญาณเหนือกว่าการใช้หน่วยประมวลผลกลาง (Wick, Judd, & Holt, 2019) โดยกระบวนการหาลำดับเบสด้วยอ็อกซ์ฟอร์ดนาโนพอร์ถูกสรุปในรูปแบบที่ 5 (Y. Wang et al., 2021) ในการแปลงสัญญาณไฟฟ้าเป็นสายลำดับเบส ซอฟต์แวร์ basecaller จะพิจารณารูปแบบของช่วงสัญญาณที่เปลี่ยนแปลงไป โดยมีรูปแบบช่วงสัญญาณแตกต่างกันตามรูปแบบของลำดับเบสที่ต่อเนื่องกัน เรียกว่า "k-mer" โดย k แทนจำนวนนิ

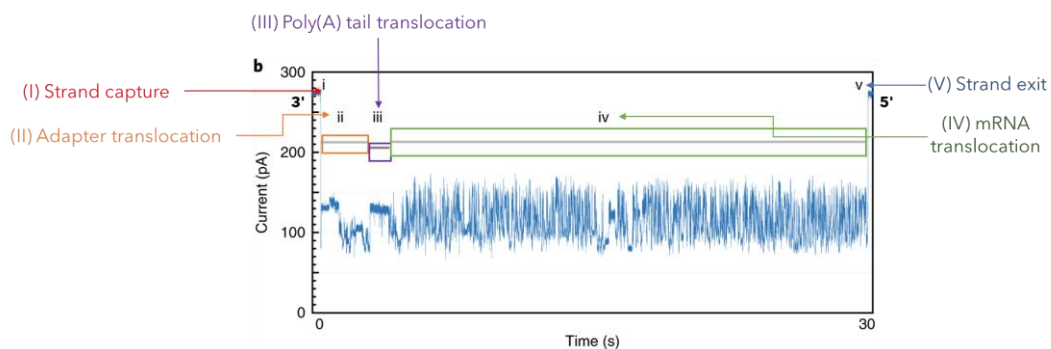


วคลีโอไทดต์ต่อ 1 ช่วงสัญญาณ (Deamer, Akesson, & Branton, 2016) โดยแสดงแต่ละช่วงสัญญาณของสายลำดับเบสอาร์เอ็นเอสรุปในรูปที่ 6 (Workman et al., 2019)



รูปที่ 5 กระบวนการของหาลำดับเบสโดยใช้อ็อกซ์ฟอร์ดนาโนพอร์โดยสรุป (Y. Wang et al., 2021)

อ็อกซ์ฟอร์ดนาโนพอร์ได้วางขายแพลตฟอร์มสำหรับการหาลำดับเบสออกมาหลายรุ่น เช่น MinION, GridION, PromethION เป็นต้น โดยจะแตกต่างกันในเรื่องของความสะดวกในการพกพา, จำนวนรูนาโนพอร์สำหรับการหาลำดับเบส, ความเร็วในของสายลำดับเบสที่ผ่านรูนาโนพอร์ และปริมาณข้อมูลที่ได้รับ ทำให้สามารถเลือกใช้ตามลักษณะการใช้งานที่ต้องการ โดยแพลตฟอร์ม MinION ได้รับความนิยมในการใช้งาน เนื่องจากพกพาสะดวก, ให้ข้อมูลในปริมาณที่เพียงพอต่อการศึกษา และมีราคาที่เหมาะสม ผล ซึ่งมีโฟลว์เซลล์ (flow cell) แบ่งพื้นที่เป็น 512 ช่อง (channel) แต่ละช่องมี 4 รู (pore) ดังนั้น Minion จะมีรูนาโนพอร์ทั้งสิ้นจำนวน 2,048 รู ที่สามารถใช้หาลำดับเบส ในขณะที่เครื่องรุ่นอื่น เช่น PromethION ที่ใช้สำหรับการหาลำดับเบสปริมาณมากที่มีถึง 48 โพล์เซลล์ หรือ 144,000 ช่อง ส่งผลให้สามารถให้ข้อมูลสายลำดับเบสได้ถึง 8.6 เทระไบต์ แต่ไม่สะดวกสำหรับสำหรับการพกพานอกสถานที่ (Y. Wang et al., 2021)



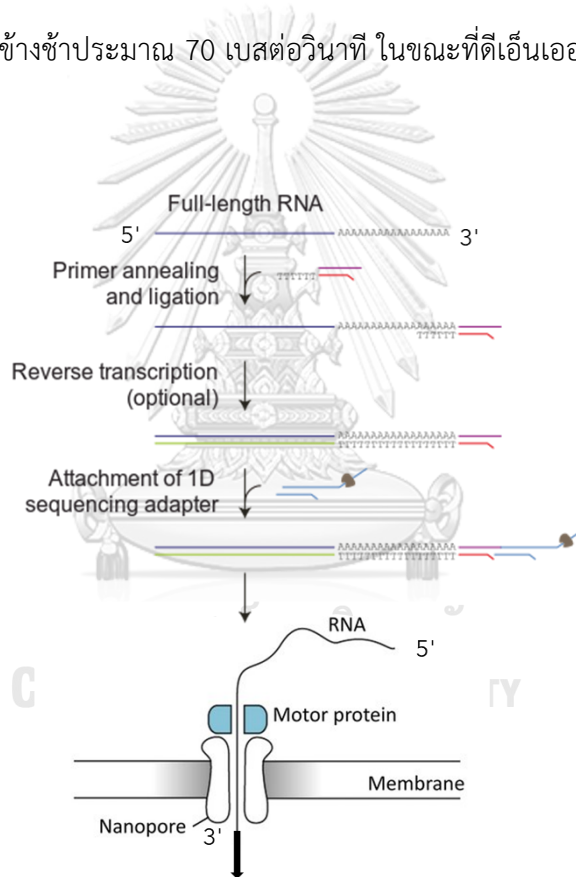
รูปที่ 6 ช่วงสัญญาณไฟฟ้าของอาร์เอ็นเอที่ผ่านรูนาโนพอร์แสดงถึงแต่ละช่วงลำดับเบส

(I) เมื่อสายลำดับเบสเข้าสู่รูนาโนพอร์ (II) ช่วงอะแดปเตอร์ (III) ช่วงหางพอลิอะดีน (IV) ช่วงเอ็มอาร์เอ็นเอ (V) เมื่อสายลำดับเบสออกจากรูนาโนพอร์ (Workman et al., 2019)

การหาลำดับเบสด้วยนาโนพอร์สามารถนำไปประยุกต์ใช้ศึกษาหลายอย่าง เช่น การทำ *de novo* assembly, metagenomic, metatranscriptomic และอื่นๆ อีกมาก (Y. Wang et al., 2021) และเนื่องจากแพลตฟอร์มของอ็อกซ์ฟอร์ดนาโนพอร์ไม่จำเป็นต้องใช้ระบบประมวลผลที่สูงมาก และสามารถเก็บข้อมูลได้ง่ายโดยไม่ต้องใช้ทักษะขั้นสูง โดยมีซอฟต์แวร์ MinKNOW ซึ่งใช้ควบคุมฮาร์ดแวร์ของอ็อกซ์ฟอร์ดนาโนพอร์ โดยสามารถตั้งค่าการหาลำดับเบส และการติดตามข้อมูลของตัวอย่างทำให้สามารถวิเคราะห์ผลลำดับเบสได้แบบทันที อย่างไรก็ตามในปัจจุบันอ็อกซ์ฟอร์ดนาโนพอร์ยังมีข้อจำกัดในการศึกษา เช่น อัตราความผิดพลาดของการหาลำดับอาร์เอ็นเอยังคงสูงถึงเกือบ 10% เมื่อเทียบกับการหาลำดับเบสด้วยวิธี NGS ที่มีความผิดพลาดไม่ถึง 1% (Delahaye & Nicolas, 2021) อีกทั้งอ็อกซ์ฟอร์ดนาโนพอร์ไม่สามารถทำการหาลำดับเบสที่มีนิวคลีโอไทด์ซ้ำกันเป็นสายยาวได้ (Rang, Kloosterman, & de Ridder, 2018)

อ็อกซ์ฟอร์ดนาโนพอร์ได้พัฒนาวิธีการสำหรับการหาลำดับเบสอาร์เอ็นเอสายยาวโดยตรง (direct RNA sequencing) ซึ่งเป็นการใช้สายอาร์เอ็นเอที่ได้จากการสกัดมาทำการหาลำดับเบส โดยไม่จำเป็นต้องทำการถอดรหัสย้อนกลับให้เป็นซีดีเอ็นเอ หรือเพิ่มปริมาณสารพันธุกรรมเหมือนวิธี NGS ซึ่งอาจนำไปสู่ความผิดพลาดของข้อมูล โดยการเตรียมไลบรารีสำหรับการหาลำดับเบสเริ่มจากการเชื่อมต่อพอลิโทไมโนอะแดปเตอร์กับหางพอลิอะดีนีนของอาร์เอ็นเอ และถอดรหัสย้อนกลับสายอาร์เอ็นเอเพื่อคลายโครงสร้างทุติยภูมิของอาร์เอ็นเอทำให้สามารถเข้าไปในรูนาโนพอร์ได้สะดวก และสุดท้ายต่อกับมอเตอร์โปรตีนที่ปลาย 3' (Garalde et al., 2018) โดยกระบวนการเตรียมไลบรารีถูกรูปในรูปที่ 7 (Jonkhout et al., 2017) ซึ่งการเตรียมตัวอย่างเพื่อให้ได้ข้อมูลสายลำดับเบสที่มีประสิทธิภาพจะต้องมีการวางแผนในการเตรียมไลบรารีที่ดี เพื่อให้ได้ข้อมูลปริมาณมาก และสายลำดับเบสยาว ซึ่ง

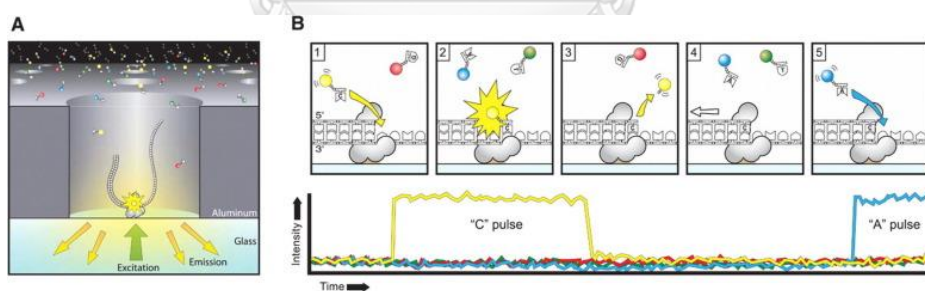
ควรใช้ระยะเวลาในการเตรียมไลบรารีที่สั้น และตัวอย่างไม่ควรมีการปนเปื้อน (Y. Wang et al., 2021) เนื่องจากอาร์เอ็นเอเป็นโมเลกุลสายเดี่ยวจึงมีความเสถียรต่ำ (Rich, 2009) ทำให้มีความอ่อนไหวสูงส่งผลให้สามารถเกิดการแตกหักได้ง่ายกว่าสายดีเอ็นเอที่เป็นโมเลกุลสายคู่กว่ามาก การหาลำดับเบสอาร์เอ็นเอสายยาวด้วยอ็อกซ์ฟอร์ดนาโนพอร์ได้รับความนิยมในการประยุกต์ไปใช้ทำให้ตรวจหาไอโซฟอร์มที่หายาก หรือตรวจหาการปรับแต่งอาร์เอ็นเอบนสายยาว เช่น เมทิลเลชัน (methylation) หรือ ซูโดยูริดีลเลชัน (pseudouridylation) เป็นต้น (Galalde et al., 2018; Jain et al., 2016) อย่างไรก็ตามการหาลำดับเบสอาร์เอ็นเอสายยาวโดยตรงความแม่นยำเฉลี่ยเพียง 83% ถึง 86% และได้ข้อมูลสายลำดับเบสประมาณ 1 ล้านสาย ต่อ 1 โพล์เซลล์เท่านั้น เนื่องจากมีความเร็วการหาลำดับเบสค่อนข้างช้าประมาณ 70 เบสต่อวินาที ในขณะที่ดีเอ็นเออยู่ที่ 430 เบสต่อวินาที (Y. Wang et al., 2021)



รูปที่ 7 การเตรียมไลบรารีสำหรับการหาลำดับเบสอาร์เอ็นเอด้วยอ็อกซ์ฟอร์ดนาโนพอร์ (Jonkhout et al., 2017)

โดยเมื่อเทียบกับการหาลำดับเบสด้วยเทคโนโลยี SMRT (Single Molecule, Real-Time) ซึ่งเป็นอีกหนึ่งทำวิธีการหาลำดับเบสสายยาวพัฒนาโดยบริษัท Pacific Biosciences ซึ่งเป็นการวิเคราะห์สายลำดับเบสโมเลกุลเดี่ยวโดยไม่จำเป็นต้องเพิ่มปริมาณลำดับเบส และสามารถทำการวิเคราะห์ได้ตามเวลาจริงเช่นเดียวกับ ONT แต่ Pacbio จะใช้เทคนิคการเรืองแสงฟลูออเรสเซนซ์เมื่อ

เกิดการสังเคราะห์นิวคลีโอไทด์ที่สายลำดับเบสต้นแบบ ซึ่งคล้ายกับเทคโนโลยีการหาลำดับเบสในยุคก่อนหน้าแบบ bridge amplification โดยกระบวนการหาลำดับเบส SMRT จะเริ่มจากการเชื่อมต่อลำดับเบสกับอะแท็ปเตอร์ที่เชื่อมทั้ง 2 สายเข้าด้วยกัน ทำให้มีลักษณะคล้ายลำดับเบสรูปวงแหวน (circular nucleotide sequence) และทำการใส่ไพรเมอร์และเอนไซม์ DNA polymerase จากนั้นแต่ละสายลำดับเบสจะเข้าไปในบ่อที่เรียกว่า zero-move wave guide (ZMW) ที่มีไบโอดีคยอตรึงลำดับเบสให้อยู่กันบ่อ โดยในบ่อจะมีนิวคลีโอไทด์ติดฉลากเรืองแสงฟลูออเรสเซนต์สำหรับการสังเคราะห์นิวคลีโอไทด์อยู่ และเนื่องจากบ่อ ZMW มีขนาดเล็กทำให้การเรืองแสงฟลูออเรสเซนต์ขณะที่เอนไซม์ DNA polymerase ทำการสังเคราะห์นิวคลีโอไทด์ที่ติดฉลากเรืองแสงฟลูออเรสเซนต์จะอยู่ที่บริเวณก้นบ่อเท่านั้น (Levene et al., 2003) โดยกระบวนการหาลำดับเบสด้วย SMRT ถูกสรุปในรูปที่ 8 (Rhoads & Au, 2015) สำหรับการหาลำดับเบสด้วยอาร์เอ็นเอด้วยเทคโนโลยี SMRT จำเป็นจะต้องทำการถอดรหัสย้อนกลับเพื่อสร้างสายลำดับเบสสายคู่ที่เป็นลูกผสมอาร์เอ็นเอ/ดีเอ็นเอเพื่อใช้ในการกระบวนการหาลำดับเบสดังกล่าว เนื่องจากการวิเคราะห์การเรืองแสงฟลูออเรสเซนต์ของนิวคลีโอไทด์ที่ถูกสังเคราะห์มีความแม่นยำกว่าแปลงสัญญาณไฟฟ้าเป็นสายลำดับเบสด้วย ONT ทำให้การหาลำดับเบสด้วยเทคโนโลยี SMRT ความคลาดเคลื่อนค่อนข้างต่ำกว่า แต่มีค่าใช้จ่ายที่สูงกว่า, ไม่สามารถพกพาได้ และใช้ระยะเวลาในการหาลำดับเบสนานกว่าในบางฮาร์ดแวร์ (Weirather et al., 2017)



รูปที่ 8 การหาลำดับเบสสายยาวด้วยเทคโนโลยี SMRT

(Rhoads & Au, 2015)

ในปัจจุบันทั้ง ONT และ Pacbio ยังคงแข่งขันเพื่อพัฒนาฮาร์ดแวร์และซอฟต์แวร์อย่างต่อเนื่องเพื่อเพิ่มประสิทธิภาพและความแม่นยำสำหรับการหาลำดับเบสแบบสายยาวโดยประยุกต์ใช้เทคโนโลยีใหม่เพื่อลดข้อเสียที่ทำให้เกิดความคลาดเคลื่อนที่จะเกิดขึ้น รวมถึงการพัฒนาเครื่องมือทางชีวสารสนเทศสำหรับการวิเคราะห์ลำดับเบสเพื่อนำไปประยุกต์ใช้ในการวิจัยศึกษา และอุตสาหกรรม

## บทที่ 3

### วิธีการดำเนินงานวิจัย

#### 3.1 กระบวนการสำหรับการหาความยาวทางพอลิอะดีนีน

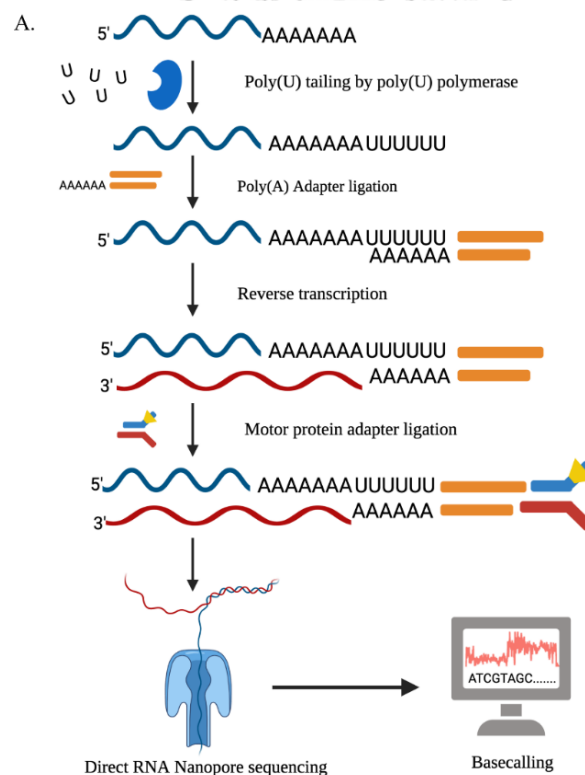
หลังจากการหาลำดับเบสด้วยอ็อกซ์ฟอร์ดนาโนพอร์โดยการเตรียมไลบรารีด้วยการสังเคราะห์พอลิยูรีดีนซึ่งสรุปในรูปแบบที่ 9 ทำการแปลงสัญญาณไฟฟ้าในไฟล์ FAST5 เป็นไฟล์ FASTQ ที่บรรจุข้อมูลลำดับเบส (basecalling) ด้วยโปรแกรม Guppy เวอร์ชัน 5.0.11 โดยใช้ configuration file rna r9.4.1 70bps hac.cfg ที่ตรงกลับรุ่นโฟลว์เซลล์ (FLO-MIN106) และชุดสำหรับการหาลำดับเบสอาร์เอ็นเอด้วยอ็อกซ์ฟอร์ดนาโนพอร์สายตรง (SQK-RNA002) และกำหนดพารามิเตอร์ --reverse\_sequence yes เพื่อทำการย้อนลำดับเบสกลับสำหรับการหาลำดับเบสด้วยอาร์เอ็นเอ --trim\_strategy rna เพื่อทำการตัดแต่งลำดับเบสอาร์เอ็นเอ --fast5\_out เพื่อสร้างไฟล์ FAST5 ที่มี event move table สำหรับการประมาณความยาวพอลิอะดีนีนด้วย tailfindr และใช้ NanoPlot เวอร์ชัน 1.36.2 (De Coster, D’Hert, Schultz, Cruts, & Van Broeckhoven, 2018) เพื่อตรวจสอบค่าสถิติของลำดับเบส และทำการคัดเลือกชุดข้อมูลที่มีคุณภาพที่อยู่ในโฟลเดอร์ pass ที่ถูกสร้างในกระบวนการแปลงสัญญาณไฟฟ้ามาใช้ในการศึกษาเท่านั้น

จากนั้นทำการจัดตำแหน่งสายลำดับเบสกับลำดับเบสอ้างอิง (sequence alignment) โดยใช้ Minimap2 เวอร์ชัน 2.18-r1015 (Li, 2018) โดยกำหนดพารามิเตอร์ -a -x map-ont สำหรับลำดับเบสที่ถูกหาลำดับเบสด้วยอ็อกซ์ฟอร์ดนาโนพอร์ ในส่วนของอาร์เอ็นเอสังเคราะห์นั้นลำดับเบสอ้างอิงที่ใช้คือลำดับนิวคลีโอไทด์ 100 ตัวสุดท้ายที่ปลาย 3' ของอาร์เอ็นเอสังเคราะห์ซึ่งรวมถึงบาร์โค้ดแต่ไม่รวมทางพอลิอะดีนีนเพื่อใช้สำหรับการคัดแยกบาร์โค้ดที่กำหนดแต่ละความยาวทางพอลิอะดีนีน ในส่วนลำดับเบสที่ไม่มีบาร์โค้ดนั้นจะไม่สามารถตำแหน่งกับลำดับเบสอ้างอิงได้และถูกกรองออกจากการวิเคราะห์เนื่องจากการแตกหักที่ปลาย 3' และอาจก่อให้เกิดความผิดพลาดในการประมาณความยาวทางพอลิอะดีนีน และในส่วนของอาร์เอ็นเอของ *E. coli* K-12 ใช้จีโนมอ้างอิงและทรานสคริปโตมอ้างอิง NC\_000913.3 (Freddolino, Amini, & Tavazoie, 2012) เพื่อหาคำแหน่งของแต่ละสายลำดับเบสบนจีโนมและระบุชื่อยีนของแต่ละสายลำดับเบสตามลำดับ จากนั้นใช้ samtools เวอร์ชัน 1.12 (Li et al., 2009) ในการจัดการไฟล์ SAM และ BAM และทำการคัดแยกเฉพาะข้อมูลคุณภาพที่มีผล Mapping quality มากกว่าเท่ากับ 10 และ Alignment score มากกว่า

เท่ากับ 150 สำหรับการวิเคราะห์ต่อไป และใช้ IGV เวอร์ชัน 2.9 (Robinson et al., 2011) เพื่อตรวจสอบการครอบคลุมและความลึกของลำดับเบสกับลำดับเบสอ้างอิง

ในชุดข้อมูลอาร์เอ็นเอของ *E. coli* K-12 จะถูกทำการคัดกรองลำดับเบสที่แตกหักที่ปลาย 3' ออก โดยใช้โปรแกรมที่ถูกเขียนขึ้นทำการแบ่งช่วงลำดับเบสโดยใช้ข้อมูลจากอักขระ CIGAR ในไฟล์ SAM โดยถ้าหากลำดับเบสนั้นไม่มีรหัสหยุดสำหรับการแปลรหัสในช่วงลำดับเบสที่ถูกแปลรหัส และไม่มีช่วงลำดับเบสที่ไม่ถูกแปลรหัสที่ปลาย 3' หมายความว่าลำดับเบสนั้นมีการแตกหักที่ปลาย 3' และจะถูกคัดทิ้ง จากนั้นทำการประมาณความยาวทางพอลิอะดีนีนด้วย nanopolish เวอร์ชัน 0.11.3 (Loman et al., 2015) และ tailfindr เวอร์ชัน 1.2 (Krause et al., 2019) และทำการรวบรวมข้อมูลความยาวทางพอลิอะดีนีนไว้ในไฟล์ SAM

จากนั้นทำการโหลดไฟล์ SAM ที่มีคอลัมน์ผลลัพธ์การประมาณความยาวทางพอลิอะดีนีน เข้าสู่กระแสน้ำที่ถูกพัฒนาขึ้นสำหรับปรับปรุงความยาวทางพอลิอะดีนีน สุดท้ายทำการสร้างแผนภูมิแผนภูมิเพื่อแสดงผลด้วย R เวอร์ชัน 4.1.0 และ python เวอร์ชัน 3.8.5

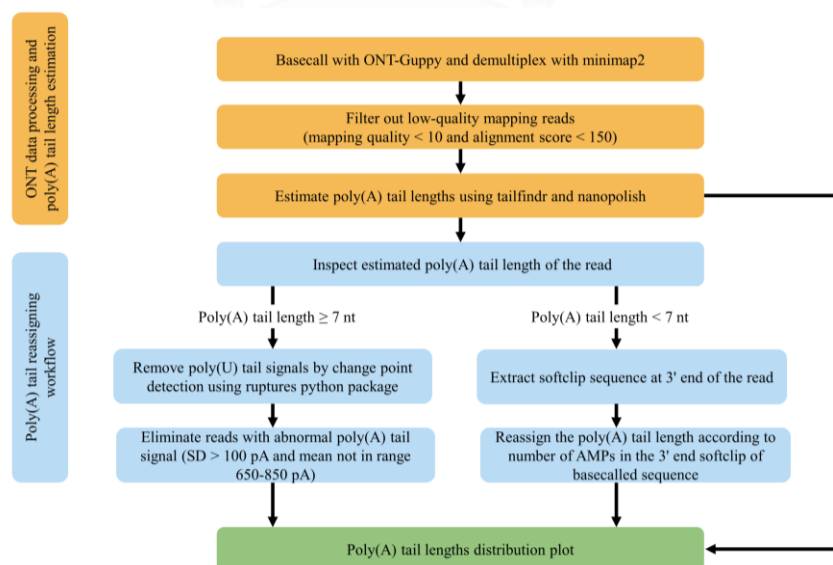


รูปที่ 9 แผนผังการเตรียมไลบรารีเพื่อหาลำดับเบสด้วยอ็อกซ์ฟอร์ดนาโนพอร์โดยการสังเคราะห์พอลิยูริดีน

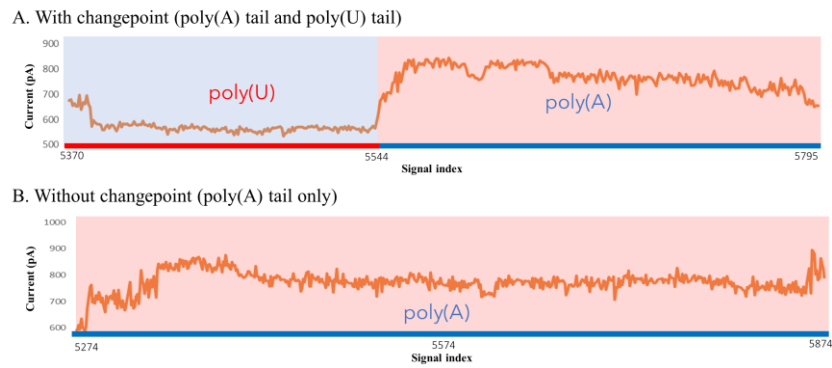
### 3.2 กระบวนการสำหรับการปรับปรุงข้อมูลความยาวหางพอลิอะดีนีน

เนื่องจากอัลกอริทึมสำหรับการประมาณความยาวหางพอลิอะดีนีนที่มีอยู่ในปัจจุบันมักใช้ในการประมาณความยาวหางพอลิอะดีนีนในอาร์เอ็นเอของยูแคริโอตซึ่งมีความยาวสูง ทำให้การประมาณความยาวหางพอลิอะดีนีนนั้นมีความท้าทาย และอาจส่งผลให้เกิดความคาดเคลื่อนสูงกว่าการประมาณความยาวหางพอลิอะดีนีนที่ยาว จึงได้ทำการพัฒนากระบวนการขึ้นเพื่อปรับปรุงข้อมูลความยาวหางพอลิอะดีนีนที่ถูกประมาณเพื่อให้ใกล้เคียงกับความยาวหางพอลิอะดีนีนที่แท้จริง ทั้งลำดับเบสที่หางพอลิอะดีนีนที่ถูกประมาณความยาวสั้นและยาวกว่าความยาวที่แท้จริง โดยกระบวนการถูกสรุปในรูปที่ 10

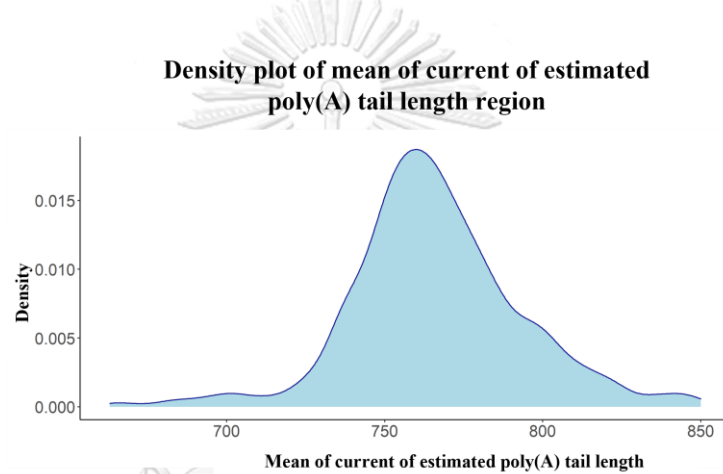
โดยกระบวนการเริ่มจากการตรวจสอบความยาวหางพอลิอะดีนีนที่ถูกประมาณของแต่ละลำดับเบส ในกรณีที่หางพอลิอะดีนีนถูกประมาณความยาวสั้นกว่าปกติ ซึ่งมักพบอาร์เอ็นเอในกลุ่มที่มีความยาวหางพอลิอะดีนีนเท่ากับ 5-10 นิวคลีโอไทด์ จะถูกแก้ไขข้อมูลความยาวหางพอลิอะดีนีนโดยการใช้ความยาวหางพอลิอะดีนีนที่พบในลำดับเบสแทน และถึงแม้ว่าอ็อกซ์ฟอร์ดนาโนพอร์จะไม่สามารถหาลำดับเบสฮอโมโพลิเมอร์สายยาวทั้งสายได้ (Rang et al., 2018) แต่มีประสิทธิภาพเพียงพอในการหาลำดับเบสฮอโมโพลิเมอร์สายสั้นทั้งสายได้ในความยาวอยู่ที่ 7 ถึง 12 นิวคลีโอไทด์ ซึ่งในการศึกษาครั้งนี้ตั้งเกณฑ์ไว้ว่า ถ้าหากลำดับเบสที่มีความยาวหางพอลิอะดีนีนที่ถูกประมาณมีความยาวต่ำกว่า 7 นิวคลีโอไทด์ ลำดับเบสนั้นจะถูกตรวจสอบความยาวของหางพอลิอะดีนีน โดยทำการตรวจสอบช่วงลำดับเบสที่ปลาย 3' ที่ไม่อยู่ในลำดับเบสอ้างอิง หรือ softclip และแทนที่ความยาวหางพอลิอะดีนีนที่ถูกประมาณด้วยโปรแกรมด้วยจำนวนอะดีนีนนิวคลีโอไทด์ที่นับได้สูงสุดใน softclip



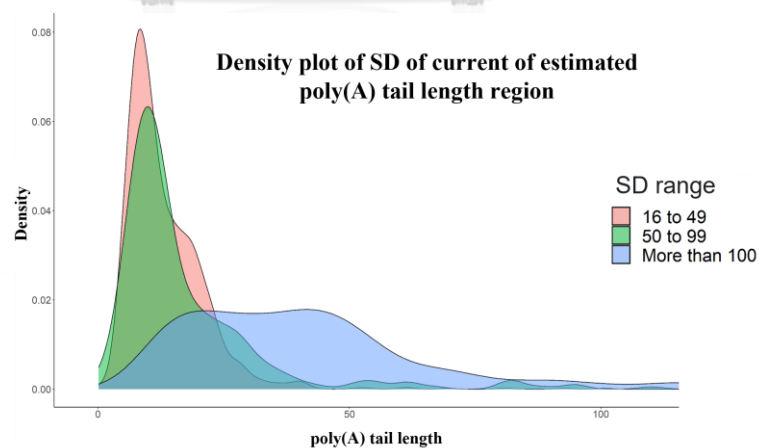
รูปที่ 10 แผนผังการทำงานของวิธีการทางคอมพิวเตอร์ในการประมาณความยาวหางพอลิอะดีนีนสายสั้น



รูปที่ 11 ช่วงสัญญาณไฟฟ้าที่ถูกทำนายและประมาณความยาวหางพอลิอะดีนีน A. มีพอลิยูรีดีน และหางพอลิอะดีนีน B. มีเพียงหางพอลิอะดีนีนเท่านั้น



รูปที่ 12 การกระจายตัวของค่าเฉลี่ยของสัญญาณไฟฟ้าในช่วงหางพอลิอะดีนีนที่ถูกประมาณ



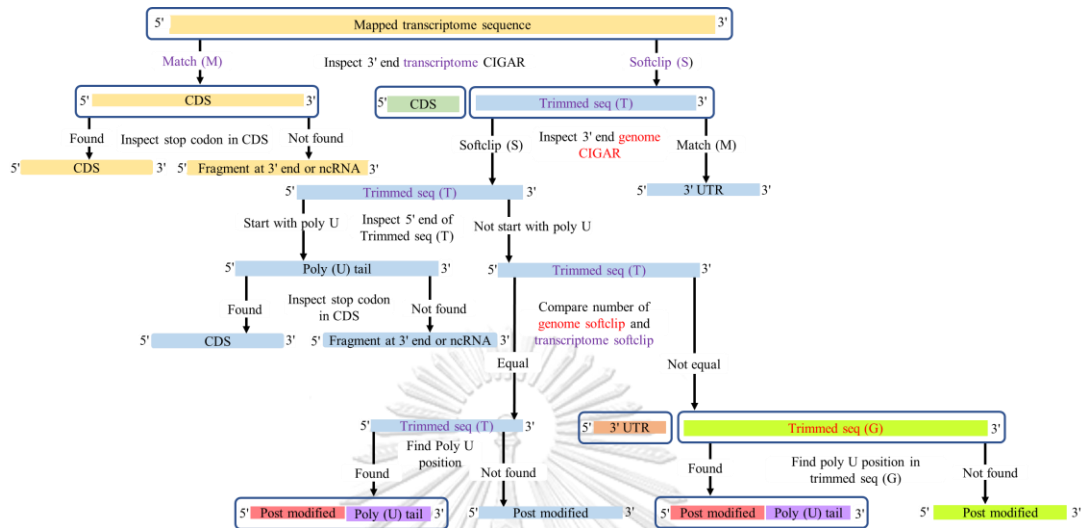
รูปที่ 13 การกระจายตัวของส่วนเบี่ยงเบนมาตรฐานของสัญญาณไฟฟ้าในช่วงหางพอลิอะดีนีนที่ถูกประมาณในช่วงของส่วนเบี่ยงเบนมาตรฐานที่ต่างกันออกไป



ในส่วนของลำดับเบสที่หางพอลิอะดีนีนถูกประมาณความยาวสูงกว่าหรือเท่ากับ 7 นิวคลีโอไทด์จะถูกตรวจสอบช่วงสัญญาณไฟฟ้าของหางพอลิอะดีนีนที่ถูกประมาณ และหาจุดเปลี่ยนแปลงของสัญญาณระหว่างพอลิยูรีดีนที่ถูกสังเคราะห์และหางพอลิอะดีนีน เพื่อตัดสัญญาณไฟฟ้าของในส่วนของพอลิยูรีดีนที่อาจจะติดมาจากการประมาณความยาวหางพอลิอะดีนีนทำให้ความยาวสูงผิดปกติตั้งรูปที่ 11 โดยใช้แพ็คเกจ rupture ใน python (Truong, Oudre, & Vayatis, 2020) โดยหากพบจุดเปลี่ยนระหว่างสัญญาณ ซึ่งหมายความว่าพบช่วงสัญญาณไฟฟ้าพอลิยูรีดีนในช่วงสัญญาณไฟฟ้าของหางพอลิอะดีนีน จะทำการคำนวณความยาวหางพอลิอะดีนีนของสายลำดับเบสนั้นใหม่ โดยใช้ข้อมูลจากโปรแกรมในการหาความยาวหางพอลิอะดีนีนคำนวณ โดยใช้ดัชนีของสัญญาณไฟฟ้าตำแหน่งสุดท้ายในช่วงสัญญาณไฟฟ้าของหางพอลิอะดีนีนที่ถูกประมาณ ลบด้วยดัชนีของสัญญาณไฟฟ้าที่ถูกทำนายว่าเป็นจุดเปลี่ยนแปลงระหว่างสัญญาณหางพอลิอะดีนีนกับพอลิยูรีดีนที่ถูกสังเคราะห์ จากนั้นหารด้วยอัตราจำนวนสัญญาณต่อหนึ่งนิวคลีโอไทด์

จากการตรวจสอบค่าสถิติของช่วงสัญญาณไฟฟ้าของความยาวหางพอลิอะดีนีนที่ถูกประมาณ พบว่าค่าเฉลี่ยอยู่ในช่วง 650-850 พิโคแอมแปร์ตั้งรูปที่ 12 และค่าเบี่ยงเบนมาตรฐานที่มากกว่า 100 พิโคแอมแปร์จะมีการกระจายตัวของความยาวหางพอลิอะดีนีนกว้างตั้งรูปที่ 13 จึงได้กำหนดเกณฑ์สำหรับการคัดสายลำดับเบสออกอ้างอิงจากค่าสถิติของช่วงสัญญาณไฟฟ้าของความยาวหางพอลิอะดีนีนที่ถูกประมาณ คือค่าเฉลี่ยของไม่อยู่ในช่วง 650-850 พิโคแอมแปร์ หรือค่าเบี่ยงเบนมาตรฐานเกินกว่า 100 พิโคแอมแปร์

### 3.3 โปรแกรมสำหรับการแบ่งช่วงลำดับเบสอาร์เอ็นเอเพื่อการตรวจสอบอาร์เอ็นเอที่มีการแตกหักที่ปลาย 3'



รูปที่ 14 ขั้นตอนการทำงานของโปรแกรมสำหรับการแบ่งช่วงลำดับเบสที่ถูกเตรียมด้วยไลบรารีที่มีการสังเคราะห์พอลิยูรีดินและตรวจสอบการแตกหักลำดับเบสที่ปลาย 3'

เนื่องจากอาร์เอ็นเอเป็นโมเลกุลที่มักพบเป็นสายเดี่ยวส่งผลให้เป็นโมเลกุลที่มีความเสถียรค่อนข้างต่ำ (Rich, 2009) ส่งผลให้สามารถเกิดการแตกหักของสายลำดับเบสได้ระหว่างขั้นตอนการเตรียมไลบรารีสำหรับการหาลำดับเบส และถ้าหากเกิดการแตกหักที่ปลาย 3' ของอาร์เอ็นเอ อาจส่งผลให้การวิเคราะห์ความยาวทางพอลิอะดีนีนเกิดความคลาดเคลื่อนขึ้น จึงได้ทำการเขียนโปรแกรมในการแบ่งช่วงลำดับเบสโดยอัตโนมัติ โดยจะแบ่งได้ดังนี้ (1) ช่วงที่มีการแปลรหัส (2) ช่วงที่ไม่มีการแปลรหัสที่ปลาย 3' (3) ช่วงที่ถูกเติมแต่งหลังการถอดรหัสที่ปลาย 3' (4) ช่วงลำดับเบสอื่นที่ปลาย 3' เช่น พอลิยูรีดินที่ถูกสังเคราะห์ หรืออะดีปเตออร์เป็นต้น โดยใช้อักขระ CIGAR จากขั้นตอนจัดตำแหน่งลำดับเบสกับจีโนมและทรานสคริปโตมอ้างอิงในไฟล์ SAM

โดยโปรแกรมจะเริ่มต้นตรวจสอบอักขระ CIGAR ที่ปลาย 3' จากการจัดตำแหน่งลำดับเบสด้วยทรานสคริปโตมอ้างอิง หากพบเป็น match แสดงว่าลำดับเบสนี้มีแต่ช่วงที่ถูกแปลรหัส ซึ่งจะถูกนำไปตรวจสอบลำดับเบสสำหรับการหยุดการแปลรหัส เพื่อตรวจสอบว่ามีการแตกหักที่ปลาย 3' หรือไม่ โดยถ้าหากพบว่าไม่มีรหัสหยุด แสดงว่ามีการแตกหักที่ปลาย 3' และจะไม่ถูกนำมาวิเคราะห์ต่อ เพื่อลดความคลาดเคลื่อนที่อาจเกิดในขั้นตอนการวิเคราะห์ความยาวทางพอลิอะดีนีน แต่ถ้าหากพบ softclip แทน ลำดับเบสจะถูกแบ่งช่วงครั้งแรกโดยใช้หมายเลขของ softclip จากผลการจัดตำแหน่งลำดับเบสด้วยทรานสคริปโตมอ้างอิงที่เป็นตัวระบุจำนวนนิวคลีโอไทด์ที่ใช้ในการแบ่งลำดับ

เบส ซึ่งจะได้เป็นสองช่วงลำดับเบสคือ ช่วงที่ถูกถอดรหัส และช่วงที่แบ่งออกมาตามหมายเลขของ softclip จากผลการจัดตำแหน่งลำดับเบสด้วยทรานสคริปโตมอ้างอิง

สำหรับช่วงที่ถูกแบ่งออกมาตามหมายเลขของ softclip จากผลการจัดตำแหน่งลำดับเบสด้วยทรานสคริปโตมอ้างอิง จะถูกนำไปตรวจสอบอักขระ CIGAR ที่ปลาย 3' จากการจัดตำแหน่งลำดับเบสด้วยจีโนมอ้างอิง หากเป็น match หมายความว่าช่วงลำดับเบสที่ถูกแบ่งออกมดังกล่าวนั้น เป็นช่วงที่ไม่ถูกแปลรหัสเท่านั้น แต่ถ้าหากอักขระ CIGAR ที่ปลาย 3' จากการจัดตำแหน่งลำดับเบสด้วยจีโนมอ้างอิงเป็น softclip ช่วงที่ถูกแบ่งออกมดังกล่าวจะถูกนำไปทำการตรวจสอบว่าพบพอลิยูริดีนที่ถูกสังเคราะห์อยู่ที่จุดเริ่มต้นที่ปลาย 5' ของลำดับเบสถูกแบ่งออกมาหรือไม่ หากพบพอลิยูริดีนที่จุดเริ่มต้นของลำดับเบสที่ปลาย 5' หมายความว่าช่วงลำดับเบสที่แบ่งออกมดังกล่าวไม่มีช่วงที่ไม่ถูกแปลรหัสที่ปลาย 3' และการแต่งเติมหลังถอดรหัส ถ้าหากไม่พบพอลิยูริดีนที่จุดเริ่มต้น ปลาย 5' ของช่วงลำดับเบสดังกล่าวจะต้องถูกนำไปตรวจสอบต่อว่าหมายเลขของ softclip ในทรานสคริปโตม และจีโนมอ้างอิงที่ปลาย 3' มีจำนวนเท่ากันหรือไม่ หากเท่ากันหมายความว่าไม่มีช่วงที่ไม่ถูกแปลรหัส ซึ่งจะพบช่วงที่ถูกการแต่งเติมหลังถอดรหัส แต่อาจมีพอลิยูริดีนหรือไม่มีก็ได้ สามารถตรวจสอบได้จากการหาตำแหน่งเริ่มต้นของพอลิยูริดีนในสายลำดับเบสที่ถูกแบ่งออกมา

ในกรณีที่หมายเลข softclip จากการจัดตำแหน่งลำดับเบสในทรานสคริปโตมและจีโนมอ้างอิงที่ปลาย 3' ไม่เท่ากัน หมายความว่ามีความยาวช่วงที่ไม่ถูกแปลรหัสที่ปลาย 3' อยู่ในช่วงลำดับเบสที่แบ่งออกมาตามหมายเลขของ softclip จากผลการจัดตำแหน่งลำดับเบสด้วยทรานสคริปโตมอ้างอิง โดยจะทำการแบ่งช่วงลำดับเบสครั้งที่สอง โดยใช้จำนวนนิวคลีโอไทด์ของ softclip ที่ปลาย 3' จากการจัดตำแหน่งลำดับเบสกับจีโนมอ้างอิง ซึ่งจะได้ช่วงที่ไม่ถูกแปลรหัสที่ปรากฏในจีโนมอ้างอิง และช่วงที่มีการแต่งเติมหลังการถอดรหัส ซึ่งจะถูกตรวจการมีอยู่ของพอลิยูริดีนที่ถูกสังเคราะห์เช่นกันโดยใช้วิธีตามที่กล่าวข้างต้นโดยกระบวนการทำงานของโปรแกรมถูกสรุปในรูปที่ 14

### 3.4 การตรวจสอบช่วงสัญญาณไฟฟ้าในสายลำดับเบส

ทำการตรวจสอบสัญญาณไฟฟ้าของแต่ละสายลำดับเบสที่มีลักษณะแบบฮอโมโพลิเมอร์และต่อกับสายพอลิยูริดีนที่ถูกสังเคราะห์ที่ปลาย 3' ของอาร์เอ็นเอ ซึ่งฮอโมโพลิเมอร์ดังกล่าวมีความเป็นไปได้ที่จะเป็นหางพอลิอะดีนีน และทำการจัดตำแหน่งสัญญาณไฟฟ้ากับตำแหน่งลำดับเบสด้วย nanopolish event align (Loman et al., 2015) เพื่อระบุตำแหน่งของนิวคลีโอไทด์กับตำแหน่งสัญญาณไฟฟ้าของสายลำดับเบส

# บทที่ 4

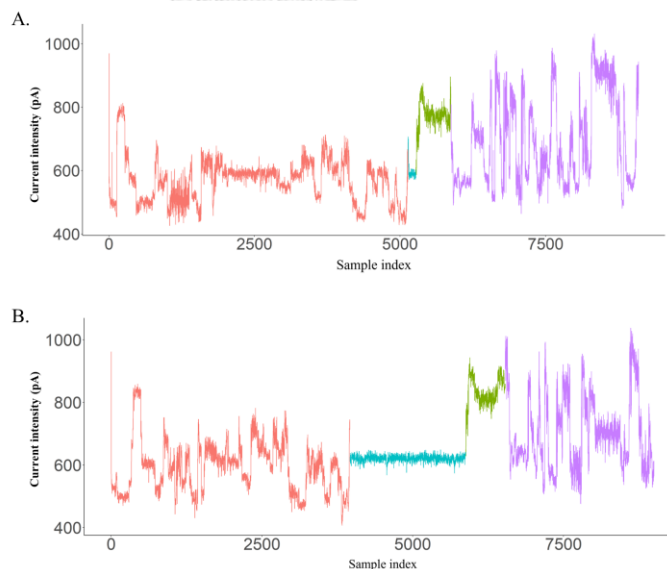
## ผลการวิจัย

### 4.1 การทดสอบวิธีการเตรียมไลบรารีและการพัฒนากระแสนงานสำหรับการหาลำดับเบสด้วย อ็อกซ์ฟอร์ดนาโนพอร์ด้วยการสังเคราะห์พอลิยูรีตินโดยใช้อาร์เอ็นเอสังเคราะห์ที่มีความยาวทาง พอลิอะดีนีนคิงที่

ตารางที่ 1 ลำดับเบสของอาร์เอ็นเอสังเคราะห์และตัวแปรควบคุม

บาร์โค้ด	ลำดับเบสบางส่วนจากปลาย 3' ของอาร์เอ็นเอ
GFP-NB01	TACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCTGCTGGAGTTCGTGACCGCCCGGGATCACTCTCGGCATGGACGAGCTGTACA AGCACAAAGACCCGACAACTTTCTT
GFP-NB23-5A	CTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCTGCTGGAGTTCGTGACCGCCCGGGATCACTCTCGGCATGGACGAGCTGTAC AAGCTTACTACCCAGTGAACCTCCTCGAAAAA
GFP-NB10-10A	CTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCTGCTGGAGTTCGTGACCGCCCGGGATCACTCTCGGCATGGACGAGCTGTAC AAGGAGAGGACAAAGTTTCAACGCTTAAAAAAAAA
GFP-NB12-20A	CTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCTGCTGGAGTTCGTGACCGCCCGGGATCACTCTCGGCATGGACGAGCTGTAC AAGTCCGATTCGTCTTCTTCTACCTGAAAAAAAAAAAAAAAAA
YHR174W_mRNA ENO2 (Control)	CTTCAAGACCGCTGGTATCCAAATGTTGCTGATGACTTGACTGTCCACCAACCCAGTAGAATTGCTACCGCCATCGAAAAAGGCTGCTGACGCTTTGTTGTTGAAGGTTAACC AAATCGGTACCTTGTCTGAATCCATCAAGGCTGCTCAAGACTCTTTCGCTGCCAACTGGGGTGTATGGTTTCCCACAGATCTGGTAACTGAAGACACTTTCATTGCTGACTTG GTTGTCCGTTTGAGAACTGGTCAATCAAGACTGGTCTCCAGCTAGATCCGAAAGATTGGCTAAGTTGAACCAATTGTTGAGAATCGAAGAAGAAATGGGTGACAAGGCTGTCTA CGCCGGTGAAAACTTCCACCACGGTGACAAGTTGTAA

ลำดับเบสที่ขีดเส้นใต้คือบาร์โค้ดที่ใช้ในการแยกแต่ละความยาวของหางพอลิอะดีนีนแต่ละรูปแบบ



รูปที่ 15 สัญญาณไฟฟ้าของอาร์เอ็นเอในแต่ละช่วง A. สัญญาณไฟฟ้าจากลำดับเบสในชุดข้อมูล GFP  
10 minutes B. สัญญาณไฟฟ้าจากลำดับเบสในชุดข้อมูล GFP 30 minutes

โดย (I) อะแดปเตอร์ (II) พอลิยูรีตินที่ถูกสังเคราะห์ (III) หางพอลิอะดีนีน (IV) สายอาร์เอ็นเอ  
สังเคราะห์และบาร์โค้ด

ตารางที่ 2 ค่าสถิติของลำดับเบสของอาร์เอ็นเอสังเคราะห์ในแต่ละการทดลองจาก nanoplot

ชุดข้อมูล	จำนวนสายลำดับเบส	ค่าเฉลี่ยความยาวสาย	ค่าเฉลี่ยคุณภาพของสาย	ค่ามัธยฐานของความยาวสายลำดับเบส	ค่ามัธยฐานของคุณภาพของสายลำดับเบส
		ลำดับเบส	ลำดับเบส		
GFP 10 mins	20090	437.6	9.8	394	10.2
GFP 30 mins (REP1)	11935	400.4	8.1	342	8.8
GFP 30 mins (REP2)	9632	223.6	5.7	78	4.4

#### 4.1.1 อาร์เอ็นเอสังเคราะห์ที่มีหางพอลิอะดีนีนความยาวคงที่ที่ใช้ในงานวิจัย และค่าสถิติของลำดับเบสแต่ละชุดการทดลอง

อาร์เอ็นเอสังเคราะห์ green fluorescent protein (GFP) ที่มีหางพอลิอะดีนีนความยาวคงที่ที่ใช้ในการศึกษามี 4 รูปแบบตามความยาวของหางพอลิอะดีนีนโดยใช้บาร์โค้ดในการระบุแต่ละความยาวของหางพอลิอะดีนีนในอาร์เอ็นเอ ประกอบด้วย 1. ไม่มีหางพอลิอะดีนีน (GFP-NB01) 2. มีความยาว 5 นิวคลีโอไทด์ (GFP-NB23) 3. มีความยาว 10 นิวคลีโอไทด์ (GFP-NB10) 4. มีความยาว 20 นิวคลีโอไทด์ (GFP-NB12) และใช้อาร์เอ็นเอของยีสต์เป็นตัวแปรควบคุมในการทดลองซึ่งแสดงในตารางที่ 1

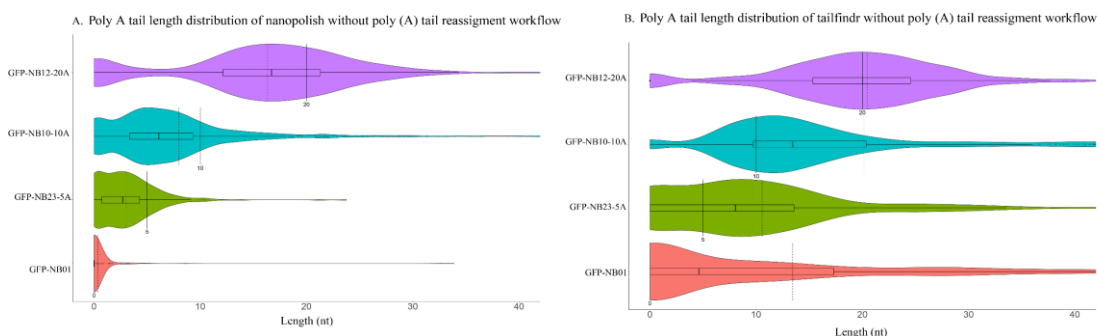
การทดลองแบ่งเป็น 2 ชุด ตามเวลาที่ใช้ในการสังเคราะห์พอลิยูริดีนคือ 10 นาที และ 30 นาที ซึ่งทำ 2 ซ้ำ ซึ่งพบว่าวิธีเตรียมไลบรารีที่เตรียมสำหรับการหาลำดับเบสอาร์เอ็นเอด้วยอ็อกซ์ฟอร์ดนาโนพอร์โดยใช้เอนไซม์ Poly(U) polymerase แทน Poly(A) polymerase สามารถทำให้ในการลำดับเบสได้สำเร็จโดยยืนยันได้จากรูปที่ 15 ที่แสดงถึงช่วงสัญญาณไฟฟ้าที่พบพอลิยูริดีนที่ถูกสังเคราะห์โดยต่อจากหางพอลิอะดีนีน ในการเพิ่มระยะเวลาในการสังเคราะห์พอลิยูริดีนที่ปลาย 3' ของอาร์เอ็นเอ ส่งผลให้ความยาวของช่วงสัญญาณพอลิยูริดีนสูงขึ้นอย่างเห็นได้ชัด จากการเปรียบเทียบสัญญาณไฟฟ้าของลำดับเบสในชุดข้อมูลที่ใช้เวลาในการสังเคราะห์พอลิยูริดีน 10 นาที (รูปที่ 15A) และสัญญาณของสายลำดับเบสในชุดข้อมูลที่ใช้เวลาในการสังเคราะห์พอลิยูริดีน 30 นาที (รูปที่ 15B) อย่างไรก็ตามจากตารางที่ 2 แสดงค่าสถิติของลำดับเบสโดย nanoplot พบว่าสายลำดับเบสในชุดข้อมูลที่ใช้เวลาในการสังเคราะห์พอลิยูริดีน 10 นาที มีความยาวและคุณภาพสูงกว่าชุดข้อมูลที่ใช้เวลาในการสังเคราะห์พอลิยูริดีน 30 นาทีทั้ง 2 ซ้ำ

#### 4.1.2 การทดสอบประสิทธิภาพโปรแกรมประมาณความยาวทางพอลิอะดีนีนสายสั้น

หลังจากการจำแนกบาร์โค้ดตามความยาวทางพอลิอะดีนีนที่คงที่ และคัดเลือกสายลำดับเบสที่มีคุณภาพ Mapping quality มากกว่าเท่ากับ 10 และ Alignscore มากกว่าเท่ากับ 150 และการประมาณความยาวของทางพอลิอะดีนีนด้วยโปรแกรม tailfindr และ nanopolish ผลลัพธ์หลังการประมาณความยาวทางพอลิอะดีนีนแสดงในตารางที่ 3 และรูปที่ 16 โดยพบว่าการประมาณความยาวทางพอลิอะดีนีนด้วย nanopolish ประมาณความยาวทางพอลิอะดีนีนของอาร์เอ็นเอที่มีทางพอลิอะดีนีนคงที่ 0, 5 และ 10 นิวคลีโอไทด์ ได้ใกล้เคียงกับความยาวจริงมากกว่า tailfindr ในขณะที่ tailfindr สามารถประมาณความยาวทางพอลิอะดีนีนของอาร์เอ็นเอที่มีความยาวทางพอลิอะดีนีน 20 นิวคลีโอไทด์ได้อย่างแม่นยำอ้างอิงจากทั้งค่าเฉลี่ยและค่ามัธยฐานในตารางที่ 3 อย่างไรก็ตามหากพิจารณารูป 16A ที่แสดงถึงการกระจายตัวของความยาวทางพอลิอะดีนีนที่ถูกประมาณด้วย nanopolish จะพบว่าการกระจายตัวของความยาวทางพอลิอะดีนีนต่ำผิดปกติในอาร์เอ็นเอที่มีความยาวทางพอลิอะดีนีน 5 และ 10 นิวคลีโอไทด์ นอกจากนี้เมื่อพิจารณาที่ค่าเฉลี่ยของความยาวทางพอลิอะดีนีนที่ประมาณได้พบว่าโดยส่วนใหญ่แล้ว nanopolish ประมาณความยาวทางพอลิอะดีนีนต่ำกว่าความยาวจริง ในขณะที่ tailfindr ประมาณความยาวทางพอลิอะดีนีนสูงกว่าปกติซึ่งมีค่าเฉลี่ยของการประมาณความยาวที่สูงกว่าในอาร์เอ็นเอทุกรูปแบบความยาว ดังตารางที่ 3 อีกทั้งมีการกระจายตัวของความยาวทางพอลิอะดีนีนที่ถูกประมาณสูงกว่าปกติดังรูปที่ 16B

ตารางที่ 3 ค่าสถิติการประมาณความยาวทางพอลิอะดีนีนก่อนขั้นตอนการปรับปรุงข้อมูลความยาวทางพอลิอะดีนีน

บาร์โค้ด/อัลกอริทึม	จำนวนสายลำดับเบส	ค่าเฉลี่ย		ค่ามัธยฐาน		ส่วนเบี่ยงเบนมาตรฐาน		ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย	
		nanopolish	tailfindr	nanopolish	tailfindr	nanopolish	tailfindr	nanopolish	tailfindr
GFP-NB01	374	0.33	13.45	0.0	4.60	2.52	21.17	0.33	16.89
GFP-NB23-5A	293	3.04	10.56	2.70	8.05	2.97	11.98	2.90	5.52
GFP-NB10-10A	1169	7.99	20.14	6.11	13.45	9.57	20.23	6.21	6.69
GFP-NB12-20A	523	16.35	20.45	16.71	20.01	10.29	12.66	7.49	6.38



รูปที่ 16 กราฟแสดงความยาวหางพอลิอะดีนีนที่ถูกประมาณก่อนขั้นตอนการปรับปรุงข้อมูลความยาวหางพอลิอะดีนีน A. nanopolish และ B. tailfindr

เส้นประคือค่าเฉลี่ยความยาวหางพอลิอะดีนีนที่ถูกประมาณ เส้นหนาคือความยาวที่แท้จริง

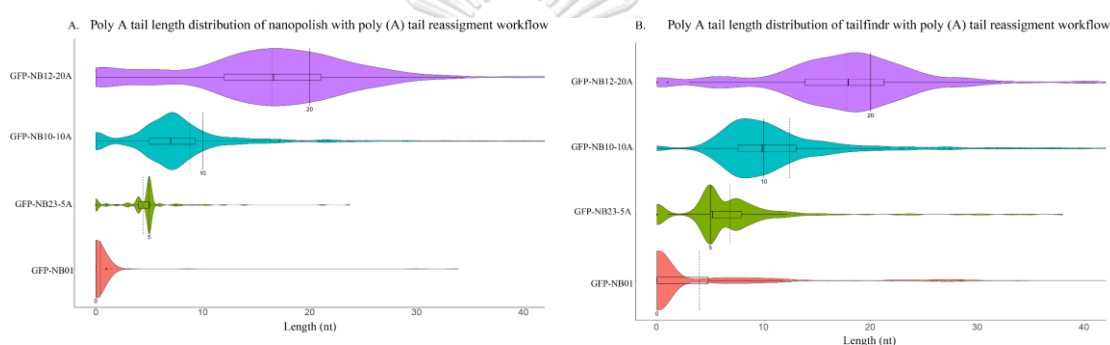
#### 4.1.3 กระแสงานสำหรับการปรับปรุงความยาวหางพอลิอะดีนีน

หลังจากทำการปรับปรุงความยาวพอลิอะดีนีนโดยทำการกู้ความยาวหางพอลิอะดีนีนของลำดับเบสที่ถูกประมาณความยาวหางพอลิอะดีนีนต่ำกว่าปกติ โดยใช้ความยาวหางพอลิอะดีนีนในที่ปรากฏในสายลำดับเบสแทน และทำการคำนวณความยาวหางพอลิอะดีนีนใหม่จากลำดับเบสที่ถูกประมาณความยาวหางพอลิอะดีนีนร่วมกับพอลิยูริดีนที่ถูกสังเคราะห์โดยใช้แพ็คเกจ rupture ใน python (Truong et al., 2020) และคัดลำดับเบสที่มีความยาวสูงผิดปกติออกโดยอ้างอิงจากค่าสถิติของช่วงสัญญาณหางพอลิอะดีนีนถูกประมาณ โดยค่าสถิติและการกระจายตัวของความยาวหางพอลิอะดีนีนหลังจากการใช้กระแสงานในการปรับปรุงความยาวหางพอลิอะดีนีนได้ถูกสรุปลงในตารางที่ 4 และรูปที่ 17 ตามลำดับ โดยพบว่าค่าเฉลี่ยของหางพอลิอะดีนีนที่ในกลุ่มอาร์เอ็นเอที่ไม่มีหางพอลิอะดีนีนถูกประมาณด้วย tailfindr ลดลงอย่างมากจาก 13.45 สู่ 3.96 และสำหรับค่ามัธยฐานจาก 4.60 สู่ 0 ในขณะที่ค่าเฉลี่ยใน nanopolish เพิ่มขึ้นเล็กน้อย สำหรับอาร์เอ็นเอที่มีหางพอลิอะดีนีนยาว 5 ถึง 10 นิวคลีโอไทด์ ทั้ง nanopolish และ tailfindr ให้ผลลัพธ์ในการประมาณหางพอลิอะดีนีนได้ใกล้เคียงกับความยาวจริงมากขึ้น โดยทั้งค่าเฉลี่ยและค่ามัธยฐานเพิ่มขึ้นทั้งใน nanopolish และ tailfindr ในส่วนของอาร์เอ็นเอที่มีความยาวหางพอลิอะดีนีน 20 นิวคลีโอไทด์ ค่าเฉลี่ยของ nanopolish เพิ่มขึ้นเล็กน้อย ในขณะที่ค่าเฉลี่ยของและค่ามัธยฐานของ tailfindr ลดลง อาจเนื่องมาจากการได้กรองสายลำดับเบสที่มีความยาวหางพอลิอะดีนีนผิดปกติทิ้ง และการตัดสัญญาณของพอลิยูริดีนที่อาจติดตามทำให้ความยาวหางพอลิอะดีนีนสั้นลง นอกจากนี้ผลลัพธ์จากการคำนวณค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ยหลังการปรับปรุงข้อมูลมีค่าลดลงกว่าก่อนทำการปรับปรุงทุกชุดข้อมูล และทุกโปรแกรม ยกเว้นในชุดข้อมูลที่ไม่มีความยาวหางพอลิอะดีนีนที่ถูกประมาณความยาวด้วย nanopolish เท่านั้น นอกจากนี้ยังพบว่าหางพอลิอะดีนีนที่ถูกประมาณด้วย tailfindr มักมีค่า

ส่วนเบี่ยงเบนมาตรฐานสูงกว่า nanopolish แม้หลังการใช้อัลกอริทึมช่วยเหลือ อีกทั้งจำนวนลำดับเบสที่ถูกกรองทิ้งไปยังมีจำนวนมากกว่า nanopolish ในทุกชุดข้อมูล (ตารางที่ 4)

ตารางที่ 4 ค่าสถิติการประมาณความยาวหางพอลิอะดีนีนหลังขั้นตอนการปรับปรุงข้อมูลความยาวหางพอลิอะดีนีน

บาร์โค้ด/ อัลกอริทึม	จำนวนสายลำดับเบส		ค่าเฉลี่ย		ค่ามัธยฐาน		ส่วนเบี่ยงเบนมาตรฐาน		ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย	
	<i>nanopolish</i>	<i>tailfindr</i>	<i>nanopolish</i>	<i>tailfindr</i>	<i>nanopolish</i>	<i>tailfindr</i>	<i>nanopolish</i>	<i>tailfindr</i>	<i>nanopolish</i>	<i>tailfindr</i>
GFP-NB01	374	272	0.43	3.96	0	0	2.40	9.01	0.43	3.96
GFP-NB23-5A	293	251	4.39	6.86	5	5.21	2.63	5.59	1.41	3.28
GFP-NB10-10A	1165	976	8.81	12.41	7	9.85	8.86	10.68	5.21	5.15
GFP-NB12-20A	521	505	16.47	17.73	16.62	17.9	9.95	9.02	7.28	5.96



รูปที่ 17 กราฟแสดงความยาวหางพอลิอะดีนีนที่ถูกประมาณหลังขั้นตอนการปรับปรุงข้อมูลความยาวหางพอลิอะดีนีน A. nanopolish และ B. tailfindr

เส้นประคือค่าเฉลี่ยความยาวหางพอลิอะดีนีน เส้นหนาคือความยาวที่แท้จริง



## 4.2 การตรวจหา ยีนที่มีหางพอลิอะดีนีนในแบคทีเรีย *Escherichia coli* K-12 ด้วยไลบรารี สำหรับการหาลำดับเบสด้วยออกซ์ฟอร์ดนาโนพอร์ด้วยการสังเคราะห์พอลิยูริดีน

### 4.2.1 ค่าสถิติของลำดับเบสในแต่ละชุดการทดลอง

ในการศึกษาทำการแบ่งชุดการทดลองเป็น 2 กลุ่ม ได้แก่ 1. กลุ่มที่ใช้เอนไซม์ Poly(A) polymerase ในการเตรียมไลบรารีซึ่งเป็นการเตรียมแบบมาตรฐานและเป็นตัวแปรควบคุมในการศึกษา จำนวน 1 ซ้ำ และ 2. กลุ่มที่ใช้เอนไซม์ Poly(U) polymerase ในการเตรียมไลบรารี โดยแบ่งย่อยเป็นใช้เวลาในขั้นตอนการสังเคราะห์พอลิยูริดีน 30 นาที จำนวน 5 ซ้ำ และใช้เวลาในขั้นตอนการสังเคราะห์พอลิยูริดีน 90 นาที จำนวน 1 ซ้ำ โดยค่าสถิติหลังการหาลำดับเบสสรุปในตารางที่ 5

ตารางที่ 5 ค่าสถิติของลำดับเบสของอาร์เอ็นเอ *E. coli* K-12 ในแต่ละการทดลองจาก nanoplot

กลุ่มข้อมูล	จำนวนสายลำดับเบส	ค่าเฉลี่ยความยาวสายลำดับเบส	ค่าเฉลี่ยคุณภาพของสายลำดับเบส	ค่ามัธยฐานของความยาวสายลำดับเบส	ค่ามัธยฐานของคุณภาพของสายลำดับเบส	จำนวนยีนที่พบ
<i>E. coli</i> poly(A)	1117023	226.7	7.4	114	7.6	3200
<i>E. coli</i> poly U 30 mins (REP1)	2022	185.3	7.6	124	7.6	32
<i>E. coli</i> poly U 30 mins (REP2)	7170	225.6	5.7	82	4.6	161
<i>E. coli</i> poly U 30 mins (REP3)	8781	145.9	5	79	4.5	132
<i>E. coli</i> poly U 30 mins (REP4)	1476	159.9	3.9	67.5	3.7	7
<i>E. coli</i> poly U 30 mins (REP5)	9643	239.6	5.2	85	4.4	123
<i>E. coli</i> poly U 90 mins	18795	95.2	4.3	76	4.3	6

จากตารางที่ 5 พบว่ากลุ่มที่ใช้เอนไซม์ Poly(A) polymerase ในการเตรียมไลบรารี ให้จำนวนสายลำดับเบสมากกว่า 1 ล้านสาย และพบ 3200 ยีน จาก 4401 ยีนทั้งหมดใน *E. coli* ในขณะที่กลุ่มที่ใช้เอนไซม์ Poly(U) polymerase ในการเตรียมไลบรารีพบลำดับเบสจำนวนสูงสุดในชุดการทดลองซ้ำไม่ถึง 2 หมื่นสาย ซึ่งน้อยกว่าเกือบ 100 เท่า และพบจำนวนยีนไม่ถึง 10% ของยีนทั้งหมดใน *E. coli* โดยความยาวเฉลี่ยของสายลำดับเบสที่ได้ทั้ง 2 รูปแบบการเตรียมไลบรารีอยู่ที่ประมาณ 150-240 นิวคลีโอไทด์ ยกเว้นในกลุ่มที่ใช้เวลาสังเคราะห์พอลิยูริดีนนานถึง 90 นาที มีความยาวเฉลี่ยไม่ถึง 100 นิวคลีโอไทด์ และค่าเฉลี่ยคุณภาพของสายลำดับเบสอยู่ในช่วง 4-7.5 ซึ่งมีค่าสถิติของลำดับเบสต่ำกว่าเมื่อเปรียบเทียบกับค่าสถิติของลำดับเบสจากงานวิจัยก่อนหน้านี้ทำการ

หาลำดับเบสอาร์เอ็นเอ *E. coli* K-12 ด้วยอ็อกซ์ฟอร์ดนาโนพอร์เช่นกัน โดยใช้การเตรียมไลบรารีด้วย เอนไซม์ Poly(A) polymerase แบบมาตรฐาน ซึ่งให้ปริมาณสายลำดับเบส ประมาณ 9 แสนสาย โดยมีความยาวสายลำดับเบสเฉลี่ยอยู่ที่ 700-900 นิวคลีโอไทด์ และค่าเฉลี่ยคุณภาพอยู่ที่ประมาณ 9-10 (Grünberger, Ferreira-Cerca, & Grohmann, 2022)

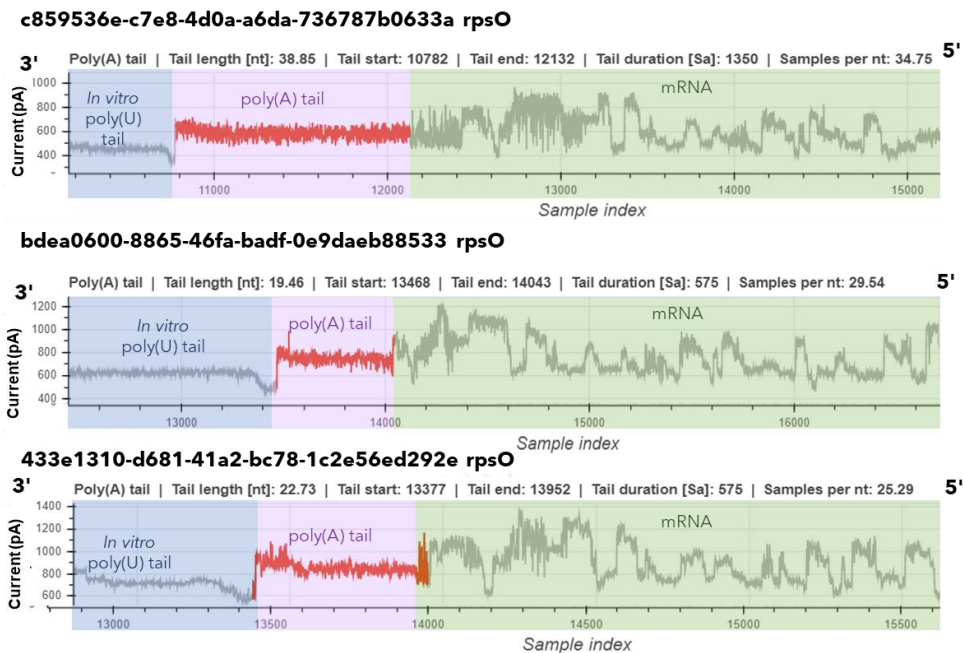
#### 4.2.2 คุณลักษณะอาร์เอ็นเอของยีนใน *E. coli* K-12 ที่มีความเป็นไปได้ของการมีหางพอลิอะดีนีนที่ถูกตรวจพบ

หลังจากการวิเคราะห์ข้อมูลผ่านกระแสน้ำทางคอมพิวเตอร์ที่ถูกพัฒนา ในการศึกษาครั้งนี้ ได้ตรวจพบยีนที่มีความเป็นไปได้ที่จะพบหางพอลิอะดีนีนจำนวน 18 ยีน โดยแบ่งเป็น 15 ยีนที่มีเคยรายงาน และ 3 ยีนที่ไม่เคยมีรายงานวิจัยก่อนหน้านี้ของการมีหางพอลิอะดีนีนดังนี้ *acpP*, *CspA*, *fdhE*, *lpp*, *ompA*, *rplQ*, *rplT*, *rpmE*, *rpsG*, *rpsI*, *rpsO*, *rpsT*, *secY*, *tufA*, *tufB*, *apt*, *ompC*, *ppa* โดยบางยีนถูกแสดงคุณลักษณะไว้ในตารางที่ 6

ตารางที่ 6 ตัวอย่างยีนที่ความเป็นไปได้ที่จะมีหางพอลิอะดีนีนในงานวิจัยนี้

รายชื่อยีน	โครงสร้างทุติยภูมิสำหรับการสิ้นสุดการถอดรหัส	การจับกับโปรตีน Hfq	โอเปอร์รอน	จำนวนหางพอลิอะดีนีนที่พบในงานวิจัยนี้	จำนวนหางพอลิอะดีนีนที่พบในงานก่อนหน้า	สายพันธุ์ที่เคยศึกษา	เอกสารอ้างอิงการรายงานหางพอลิอะดีนีน
<i>ompA</i>	มี	(Vytvytska et al., 1998)	ไม่มี	6	ไม่ได้ระบุ	<i>pcnB+</i> <i>PH</i> <i>pcnB-</i> <i>pnp</i> <i>rne</i>	(O'Hara et al., 1995)
<i>rpsO</i>	มี	(Le Derout et al., 2003)	มี	19-38	1-50	<i>pnp</i> <i>Rll</i> <i>PH</i> <i>pcnB-</i> <i>pcnB+</i>	(Hajnsdorf et al., 1995)
<i>rplQ</i>	มี	(Mildenhall et al., 2016)	มี อยู่ลำดับสุดท้าย	2-4	ไม่ได้ระบุ	<i>pcnB-</i> <i>pcnB+</i> <i>rne</i>	(Mohanty & Kushner, 2006)
<i>CspA</i>	มี	(Hankins, Denroche, & Mackie, 2010)	ไม่มี	6	ไม่ได้ระบุ	<i>pnp</i>	(Yamanaka & Inouye, 2001)
<i>rpmE</i>	มี	ไม่มี	ไม่มี	9	ไม่ได้ระบุ	<i>pcnB-</i>	(Mohanty & Kushner, 2006)
<i>fdhE</i>	มี	ไม่มี	มี อยู่ลำดับสุดท้าย	5	ไม่ได้ระบุ	<i>pcnB-</i>	(Maes et al., 2017)
<i>rpsI</i>	มี	ไม่มี	ไม่มี	3	ไม่ได้ระบุ	<i>pcnB+</i>	(Mohanty & Kushner, 2006)
<i>tufB</i>	มี	(Sittka, Pfeiffer, Tedin, & Vogel, 2007)	ไม่มี	3	ไม่ได้ระบุ	<i>pcnB-</i>	(Maes et al., 2017)
<i>ompC</i>	มี	(Douchin, Bohn, & Bouloc, 2006)	ไม่มี	6	-	WT	งานวิจัยนี้
<i>ppa</i>	ไม่มี	ไม่มี	ไม่มี	3	-	WT	งานวิจัยนี้
<i>apt</i>	มี	ไม่มี	ไม่มี	2-5	-	WT	งานวิจัยนี้

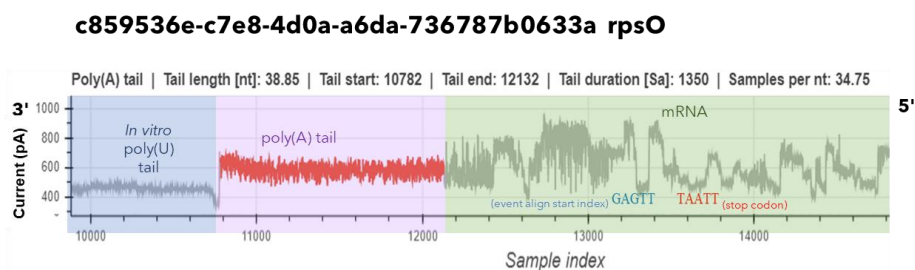
โดย WT แทนพันธุ์ป่าที่มีการแสดงออกของยีนที่ถูกแปลรหัสเป็นเอนไซม์ไรโบนิวคลีโอเอสปกติ ส่วน *pcnB-*, *pnp*, *Rll*, *rne* และ *PH* แทนสายพันธุ์ที่มีการแสดงออกของยีนดังกล่าวน้อยลงซึ่งส่งผลต่อเอนไซม์ที่ถูกแปลรหัสขาดแคลนคือ เอนไซม์ PAP I, PNPase, RNase II, RNase E, RNase PH ตามลำดับ ส่วน *pcnB+* แทนสายพันธุ์ที่มีการแสดงออกของยีนดังกล่าวมากขึ้นทำให้เอนไซม์ PAP I ถูกแปลรหัสมากขึ้น โดยเอนไซม์ PAP I สังเคราะห์หางพอลิอะดีนีน ส่วนเอนไซม์ PNPase, RNase II, RNase E, RNase PH เป็นเอนไซม์ไรโบนิวคลีโอเอสสำหรับการย่อยสลายอาร์เอ็นเอผ่านหางพอลิอะดีนีน



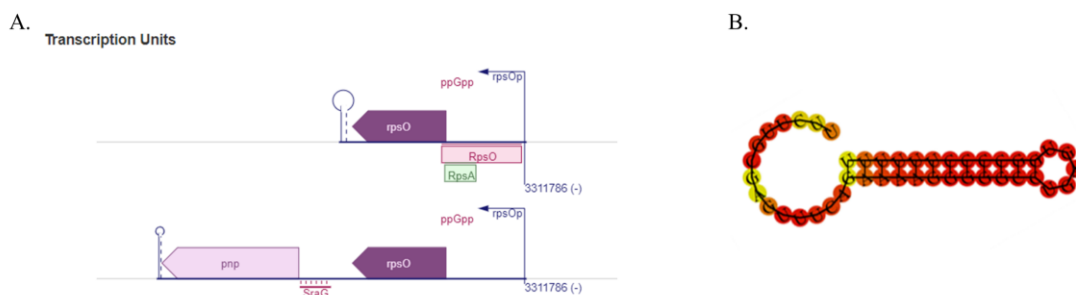
รูปที่ 18 ตัวอย่างสัญญาณไฟฟ้าในแต่ละช่วงของลำดับเบส *rpsO*

หนึ่งในอาร์เอ็นเอของยีนที่มีสัญญาณหางพอลิอะดีนีนที่ตรวจพบในการศึกษารั้งนี้คือ *rpsO* ซึ่งยึดต้นแบบในการศึกษาหางพอลิอะดีนีนในโพรแคริโอต เนื่องจากในอาร์เอ็นเอของ *E. coli* สามารถพบหางพอลิอะดีนีนที่มีความยาวยาวมากกว่ายีนอื่น โดยพบได้สูงสุดถึง 50 นิวคลีโอไทด์ (O'Hara et al., 1995) แต่โดยปกติในอาร์เอ็นเอของ *rpsO* ใน *E. coli* พันธุ์ป่ามีโอกาสพบหางพอลิอะดีนีนได้เพียง 10% (Marujo et al., 2000) และมีความยาวพอลิอะดีนีนอยู่ประมาณ 1-3 นิวคลีโอไทด์เท่านั้น

*rpsO* สามารถมีการแสดงออกเป็นโอเปอร์รอน โดยสามารถใช้โปรโมเตอร์ร่วมกับ *pnp* หรือสามารถแสดงออกแบบเดี่ยวได้ ดังรูปที่ 20A โดยมีโครงสร้างทุติยภูมิในการยุติการถอดรหัสของตัวเอง ดังรูปที่ 20 จากงานวิจัยก่อนหน้าพบว่าโปรตีน Hfq สามารถจับกับ *rpsO* ได้และมีส่วนร่วมในการสังเคราะห์หางพอลิอะดีนีนให้มีความยาวมากขึ้น (Hajnsdorf & Régnier, 2000; Le Derout et al., 2003)



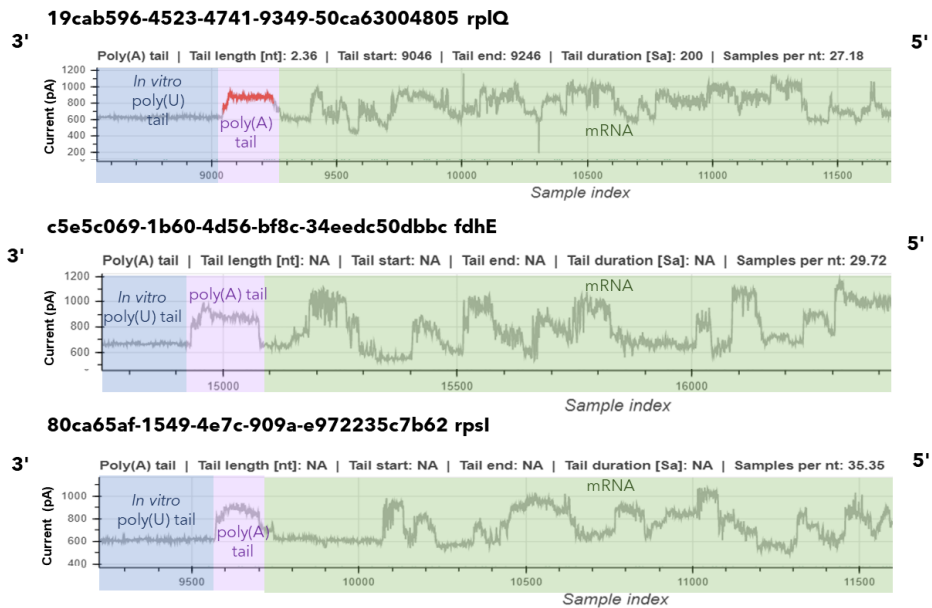
รูปที่ 19 ตัวอย่างตำแหน่งของนิวคลีโอไทด์ในตำแหน่งของสัญญาณไฟฟ้าในลำดับเบสของยีน *rpsO*



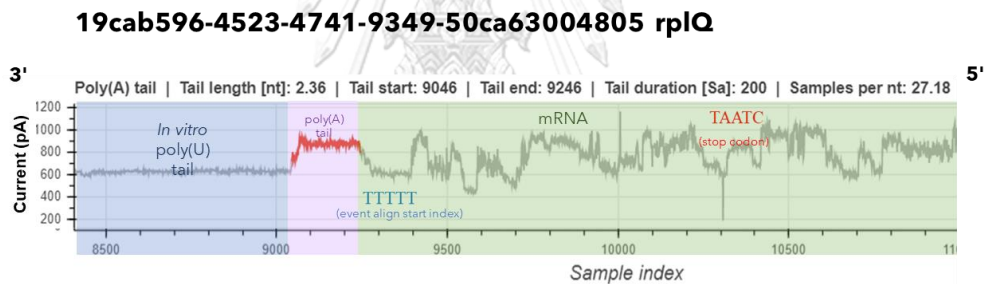
รูปที่ 20 A. รูปแบบการแสดงออกของ *rpsO* B. โครงสร้างทุติยภูมิสำหรับการยุติการถอดรหัสของ *rpsO* ที่ถูกทำนาย

ในการศึกษาครั้งนี้พบสัญญาณไฟฟ้าฮอโมพอลิเมอร์ที่ต่อจากพอลิยูริดีนที่ถูกสังเคราะห์อย่างชัดเจน ดังรูปที่ 18 อย่างไรก็ตามไม่พบทางพอลิอะดีนีนในลำดับเบส ในขณะที่พอลิยูริดีนที่ถูกสังเคราะห์ในอาร์เอ็นเอพบเป็นสายยาวในลำดับเบส และในการทำจัดตำแหน่งสัญญาณไฟฟ้ากับตำแหน่งนิวคลีโอไทด์บนลำดับเบส ไม่สามารถระบุตำแหน่งสุดท้ายของส่วนที่ปรากฏในจีโนมที่อยู่ก่อนหน้าฮอโมพอลิเมอร์สายยาวดังกล่าว ส่งผลให้ไม่สามารถระบุตำแหน่งที่แน่ชัดได้ของนิวคลีโอไทด์ในสายลำดับเบสบนตำแหน่งสัญญาณไฟฟ้าได้ ดังรูปที่ 19

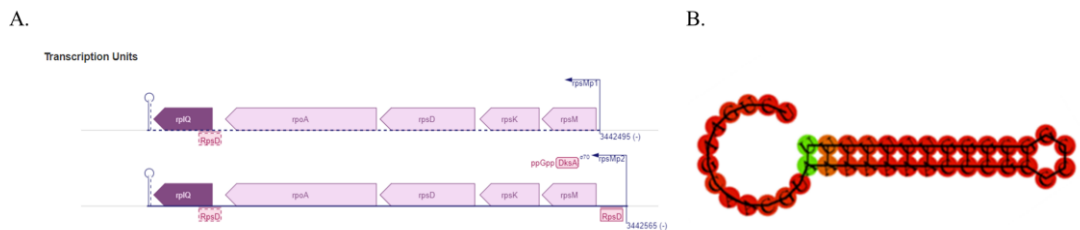
นอกจากอาร์เอ็นเอของยีน *rpsO* แล้วยังพบอาร์เอ็นเอของยีนอื่นที่มีหางพอลิอะดีนีนและมีรายงานก่อนหน้าของหางพอลิอะดีนีน เช่นในอาร์เอ็นเอของยีน *rplQ*, *fdhE* และ *rpsI* โดยแสดงรูปแบบสัญญาณฟ้าในรูปที่ 21 โดยยีนเหล่านี้สามารถพบหางพอลิอะดีนีนได้ที่ปลาย 3' ของลำดับเบส โดยถูกจัดให้เป็นช่วงของการแต่งเติมหลังการถอดรหัสจากโปรแกรมการแบ่งช่วงลำดับเบสที่ถูกเขียนขึ้น แม้ว่าอาร์เอ็นเอของยีน *rpsI* และ *fdhE* นั้นไม่สามารถถูก tailfindr ประมาณช่วงหางพอลิอะดีนีนได้ก็ตาม ตัวอย่างที่เห็นชัดที่สุดคือในอาร์เอ็นเอของยีน *rplQ* โดยเป็นยีนที่มีการแสดงออกแบบโอเปอร์รอน แต่อยู่ลำดับสุดท้ายดังรูปที่ 23A และมีโครงสร้างทุติยภูมิสำหรับการยุติการถอดรหัสในช่วงลำดับเบสที่ไม่ได้ถูกแปลรหัสที่ปลาย 3' ดังรูปที่ 23 อีกทั้งมีการรายงานถึงการจับกับโปรตีน Hfq (Mildenhall et al., 2016) ในการศึกษาครั้งนี้พบสัญญาณไฟฟ้าหางพอลิอะดีนีนที่ต่อพอลิยูริดีนที่ถูกสังเคราะห์ โดยทั้ง tailfindr และ nanopolish สามารถประมาณความยาวพอลิอะดีนีนที่สอดคล้องกันอยู่ที่ 2-3 นิวคลีโอไทด์ ซึ่งสอดคล้องกับความยาวหางพอลิอะดีนีนในช่วงลำดับเบสที่เป็นการแต่งเติมหลังถอดรหัส นอกจากนี้ผลการจัดตำแหน่งสัญญาณไฟฟ้าและลำดับเบสแสดงให้เห็นถึงจุดเริ่มต้นของลำดับเบสที่ปรากฏบนจีโนมอ้างอิงอยู่หลังสัญญาณของฮอโมพอลิเมอร์ที่ถูกทำนายเป็นหางพอลิอะดีนีนในสัญญาณไฟฟ้างดังรูปที่ 22



รูปที่ 21 ตัวอย่างสัญญาณไฟฟ้าในแต่ละช่วงของยีนอื่นๆที่มีรายงานทางพอลิอะดีนีนในงานวิจัยก่อนหน้า

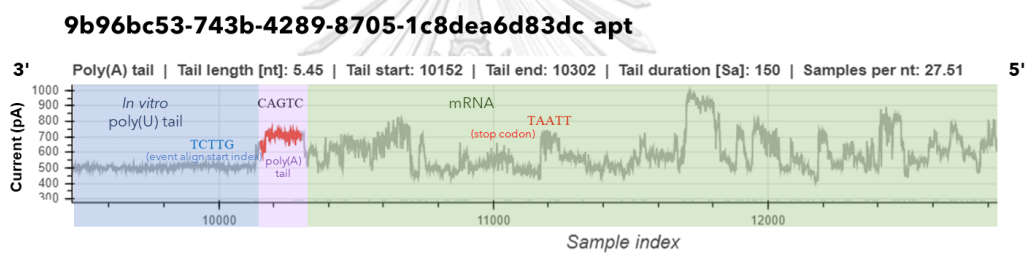


รูปที่ 22 ตัวอย่างตำแหน่งของนิวคลีโอไทด์ในตำแหน่งของสัญญาณไฟฟ้าในลำดับเบสของยีน *rplQ*



รูปที่ 23 A. รูปแบบการแสดงออกของ *rplQ* B. โครงสร้างทุติยภูมิสำหรับการยุติการถอดรหัสของ *rplQ* ที่ถูกทำนาย

ในการศึกษาครั้งนี้ยังพบทางพอลิอะดีนีนในอาร์เอ็นเอของยีนที่ไม่เคยมีรายงานทางพอลิอะดีนีนในอาร์เอ็นเอมาก่อน จากสัญญาณไฟฟ้าที่มีลักษณะเป็นฮอโมพอลิเมอร์ที่ปลาย 3' ต่อจากสัญญาณพอลิยูรีดีนที่ถูกสังเคราะห์ดังรูปที่ 24 และการมีอยู่ของโครงสร้างทุติยภูมิสำหรับการยุติการถอดรหัสเช่นในอาร์เอ็นเอของยีน *apt* ดังรูปที่ 25B เป็นสัญญาณบ่งบอกความเป็นไปได้ของทางพอลิอะดีนีน นอกจากนี้ยังพบว่าในอาร์เอ็นเอของยีน *apt* พบทางพอลิอะดีนีนในช่วงของลำดับเบสที่ถูกแต่งเติมหลังถอดรหัสที่ปลาย 3' และทั้ง nanopolish และ tailfindr สามารถประมาณความยาวทางพอลิอะดีนีนได้ ในช่วง 2-5 นิวคลีโอไทด์ อย่างไรก็ตามการพบว่ามีอาร์เอ็นเอของยีน *apt* มีการจัดลำดับสัญญาณไฟฟ้าและลำดับเบสไม่ตรงตามที่คาดหวังดังรูปที่ 24 โดยจุดเริ่มต้นของลำดับเบสบนตำแหน่งสัญญาณไฟฟ้าไปอยู่ในช่วงพอลิยูรีดีนที่ถูกสังเคราะห์ ทำให้สัญญาณฮอโมพอลิเมอร์ที่ทำนายว่าเป็นทางพอลิอะดีนีนถูกจัดตำแหน่งเป็นลำดับเบสอื่นที่ปรากฏบนจีโนมอ้างอิง



รูปที่ 24 ตัวอย่างตำแหน่งของนิวคลีโอไทด์ในตำแหน่งของสัญญาณไฟฟ้าในลำดับเบส *apt*

A.

B.



รูปที่ 25 A. รูปแบบการแสดงออกของ *apt* B. โครงสร้างทุติยภูมิสำหรับการยุติการถอดรหัสของ *apt* ที่ถูกทำนาย

## บทที่ 5

### อภิปรายผลการทดลอง

#### 5.1 การทดสอบการเตรียมไลบรารีด้วยการสังเคราะห์พอลิยูริดีนและกระแสน้ำสำหรับการตรวจหาทางพอลิอะดีนีนโดยการหาลำดับเบสด้วยอ็อกซ์ฟอร์ดนาโนพอร์

ในการศึกษานี้ ได้แสดงให้เห็นว่าการเตรียมไลบรารีด้วยเอนไซม์ Poly(U) polymerase สามารถใช้สำหรับการหาลำดับเบสอาร์เอ็นเอด้วยอ็อกซ์ฟอร์ดนาโนพอร์แทนการเตรียมไลบรารีแบบมาตรฐานด้วย Poly(A) polymerase ได้ โดยในการทดลองได้แบ่งเป็น 2 กลุ่มหลัก ตามเวลาที่ใช้ในขั้นตอนการสังเคราะห์พอลิยูริดีนในอาร์เอ็นเอ โดยการเพิ่มเวลาในขั้นตอนการสังเคราะห์พอลิยูริดีนส่งผลให้ความยาวของช่วงสัญญาณพอลิยูริดีนสูงขึ้นดังรูปที่ 15 แต่เนื่องจากอาร์เอ็นเอเป็นโมเลกุลที่มีความไม่เสถียรสูงเพราะมีโครงสร้างแบบสายเดี่ยว (Rich, 2009) ทำให้การเพิ่มระยะเวลาในการเตรียมไลบรารีเวลาอาจส่งผลในการเพิ่มโอกาสให้สายลำดับเบสมีการแตกหักเพิ่มมากขึ้น ทำให้ความยาวสั้นลง และยังอาจนำไปสู่คุณภาพของลำดับเบสที่ต่ำลงด้วย ดังตารางที่ 2

แม้ในปัจจุบันได้มีการพัฒนาโปรแกรมสำหรับการศึกษาความยาวของทางพอลิอะดีนีนสำหรับอ็อกซ์ฟอร์ดนาโนพอร์ แต่โปรแกรมที่ดังกล่าวถูกพัฒนาเพื่อประมาณความยาวทางพอลิอะดีนีนของอาร์เอ็นเอที่มีทางพอลิอะดีนีนสายยาวอย่างเช่นอาร์เอ็นเอในยูแคริโอต ซึ่งสามารถตรวจพบพอลิอะดีนีนได้ง่ายกว่าและให้ผลลัพธ์ที่มีความถูกต้องสูงกว่า จากการศึกษาครั้งนี้พบว่าถึงแม้ในการตรวจหาทางพอลิอะดีนีนความยาวสั้นในอาร์เอ็นเอมีความท้าทายสูง แต่โปรแกรมที่มีอยู่ในปัจจุบันสามารถประมาณความยาวทางพอลิอะดีนีนได้ค่อนข้างแม่นยำ และสาเหตุที่ผลลัพธ์การประมาณความยาวทางพอลิอะดีนีนต่างกัน มาจากแต่ละโปรแกรมมีอัลกอริทึมสำหรับการประมาณความยาวที่แตกต่างกัน โดย nanopolish ทำการประมาณความยาวทางพอลิอะดีนีนจากข้อมูลการจัดตำแหน่งลำดับเบสกับลำดับเบสอ้างอิงและสัญญาณไฟฟ้า ในขณะที่ tailfindr ใช้สัญญาณไฟฟ้าเพียงอย่างเดียวในการประมาณสายความยาวทางพอลิอะดีนีน (Krause et al., 2019; Workman et al., 2019) ซึ่งอาจส่งผลให้การประมาณทางพอลิอะดีนีนสายสั้นที่มีความยาวช่วงสัญญาณไฟฟ้าสั้นอาจเกิดความคลาดเคลื่อนได้ ส่งผลให้ประมาณความยาวทางพอลิอะดีนีนผิดช่วงสัญญาณ และช่วงสัญญาณที่ถูกประมาณได้มักยาวจนผิดปกติ อีกทั้งพบว่าความยาวของทางพอลิอะดีนีนที่ถูกประมาณด้วย tailfindr มีค่าเบี่ยงเบนมาตรฐานที่สูงเกือบทุกกลุ่มข้อมูล และจำนวนสายลำดับเบสที่ถูกกรองที่มากกว่าเมื่อเทียบกับ nanopolish ซึ่งอาจหลักฐานของการประมาณความยาวทางพอลิอะดีนีนที่สูงผิดปกติของโปรแกรม tailfindr จากตารางที่ 3 และตารางที่ 4



จากการทดสอบประสิทธิภาพพบว่า nanopolish มีประสิทธิภาพที่สุดหากใช้ในการแยกอาร์เอ็นเอที่ไม่มีกับมีหางพอลิอะดีนีน แต่หากเป็นอาร์เอ็นเอสายสั้นที่มีความยาว 1 ถึง 10 นิวคลีโอไทด์นั้น มีโอกาสที่จะประมาณความยาวอาร์เอ็นเอต่ำกว่าปกติ จึงแนะนำว่าควรใช้ nanopolish ร่วมกับกระแสน้ำที่ถูกพัฒนาส่งผลให้ได้ข้อมูลที่ใกล้เคียงกับความยาวจริงมากกว่า ส่วนอาร์เอ็นเอที่หางพอลิอะดีนีนมีความยาว 20 นิวคลีโอไทด์นั้น ควรใช้ผลลัพธ์การประมาณความยาวหางพอลิอะดีนีนจาก tailfindr แบบเดี่ยวจะได้ผลลัพธ์ที่ใกล้เคียงความยาวจริงที่สุด ซึ่งกระแสน้ำในการช่วยเหลือการประมาณหางพอลิอะดีนีนช่วยสามารถประมาณหางพอลิอะดีนีนได้ใกล้เคียงกับความยาวที่แท้จริงมากยิ่งขึ้น และช่วยลดความคลาดเคลื่อนที่เกิดขึ้น โดยใช้ข้อมูลหางพอลิอะดีนีนจากลำดับเบสที่ไม่อยู่ลำดับเบสอ้างอิงแทน ,การกำจัดสัญญาณไฟฟ้าพอลิยูรีตินออก และการกรองสายลำดับเบสที่มีช่วงสัญญาณไฟฟ้าของหางพอลิอะดีนีนที่ถูกประมาณผิดพลาดออก อย่างไรก็ตามในการศึกษานี้ยังมีข้อจำกัด เนื่องจากกระบวนการนี้พึ่งพาข้อมูลลำดับเบสและสัญญาณไฟฟ้าของอ็อกซ์ฟอร์ดนาโนพอร์ ซึ่งยังมีความผิดพลาดสูงมากเมื่อเทียบกับการหาลำดับเบสด้วยวิธี NGS (Y. Wang et al., 2021) ซึ่งอาจส่งผลให้ความถูกต้องของนิวคลีโอไทด์ในสายลำดับเบสเกิดความคลาดเคลื่อนรวมถึงในส่วนของหางพอลิอะดีนีนด้วย

## 5.2 คุณลักษณะที่มักพบในอาร์เอ็นเอที่มีหางพอลิอะดีนีนใน *E. coli*

ในงานศึกษาอื่นที่มีหางพอลิอะดีนีนระดับจีโนมก่อนหน้านี้ ได้รายงานอื่นที่มีความโอกาสพบหางพอลิอะดีนีนมากกว่า 100 ยีน (Maes et al., 2017; Mohanty & Kushner, 2006) อย่างไรก็ตามยีนที่ไม่ได้รับการรายงาน ไม่ได้หมายความว่าไม่มีหางพอลิอะดีนีนเสมอไป แต่อาจไม่สามารถตรวจพบได้ในสภาวะนั้น เนื่องจากความยาวหางพอลิอะดีนีนที่สั้นในอาร์เอ็นเอโพคริโอต และถูกย่อยสลายแทบจะทันทีหลังมีการสังเคราะห์หางพอลิอะดีนีนในอาร์เอ็นเอ (Sarkar, 1997) จากการศึกษาก่อนหน้านี้ที่ตรวจพบเกี่ยวกับกลุ่มหน้าที่ของยีนที่มีการแสดงออกเปลี่ยนแปลงในสายพันธุ์ *pcnB*- และจากการนำยีนที่มีเคยมีรายงานการมีอยู่หางพอลิอะดีนีนในอาร์เอ็นเอทุกสายพันธุ์จัดกลุ่มตามหน้าที่ด้วย Gene ontology แบบ biological process บน DAVID (Sherman et al., 2022) พบว่าส่วนใหญ่เป็นยีนที่เกี่ยวข้องกับการแปลรหัส นอกจากนี้ยังพบว่าเกี่ยวข้องกับการตอบสนองต่อสภาพแวดล้อมที่เปลี่ยนไป และการควบคุมการแสดงออกของยีน รวมถึงการตอบสนองต่อการต้านยาปฏิชีวนะ ซึ่งหากยีนเหล่านี้มีหากการแสดงออกมากเกินไปอาจส่งผลให้เกิดความผิดปกติได้จึงมีการควบคุมการแสดงออกยีนเหล่านี้ผ่านหางพอลิอะดีนีน

ทางพอลิอะดีนีนมักพบในอาร์เอ็นเอที่มีโครงสร้างทุติยภูมิสำหรับการยุติการถอดรหัส เพื่อให้เอนไซม์เอ็กโซไรโบนิวคลีเอสสามารถเริ่มต้นจับสายอาร์เอ็นเอและย่อยสลายโครงสร้างทุติยภูมิที่สามารถป้องกันการดำเนินงานของเอ็กโซไรโบนิวคลีเอสได้ และมักพบในยีนที่สามารถแสดงออกแบบเดี่ยวได้ หรือในลำดับสุดท้ายของโอเปอร์รอนซึ่งติดอยู่กับโครงสร้างทุติยภูมิสำหรับการยุติการถอดรหัส อย่างไรก็ตามจากการศึกษาก่อนหน้ารายงานว่ายีนที่ไม่ได้อยู่ลำดับสุดท้ายในโอเปอร์รอนและไม่พบโครงสร้างทุติยภูมิสำหรับการยุติการถอดรหัสระหว่างช่วงโอเปอร์รอนนั้นสามารถพบทางพอลิอะดีนีนได้เช่นกัน (Mohanty & Kushner, 2006) ซึ่งไม่น่าแปลกใจเพราะอาร์เอ็นเอที่ไม่ได้โครงสร้างทุติยภูมิสำหรับการยุติการถอดรหัสสามารถพบทางพอลิอะดีนีนได้เช่นกันแต่พบในสัดส่วนน้อยกว่า (Mohanty & Kushner, 2006) นอกจากนี้โครงสร้างทุติยภูมิสำหรับการยุติการถอดรหัสแล้วอาร์เอ็นเอที่มีการจับของโปรตีน Hfq มีความมีโอกาสที่จะพบทางพอลิอะดีนีน เนื่องจากโปรตีน Hfq ช่วยส่งเสริมในการสังเคราะห์ทางพอลิอะดีนีนให้มีความยาวมากขึ้นโดยสามารถยับยั้งการทำงานของเอ็กโซไรโบนิวคลีเอสเช่นเดียวกับโครงสร้างทุติยภูมิ โดย Hfq จะมีการจับที่ลำดับเบสมีอะดีนีนหรือยูรีดีนซ้ำกันมาก (Walsh et al., 2001) แต่ในปัจจุบันรายงานเกี่ยวกับอาร์เอ็นเอทั้งหมดที่มีการจับกับโปรตีน Hfq ยังไม่เป็นที่กระจ่าง และยังมีความขัดแย้งกันในบางรายงาน (Bilusic, Popitsch, Rescheneder, Schroeder, & Lybecker, 2014)

ในการศึกษาทางพอลิอะดีนีนก่อนหน้ามักทำการใช้แบคทีเรียกลายพันธุ์ยีนที่ถูกแปลรหัสเอนไซม์ไรโบนิวคลีเอสในการย่อยสลายอาร์เอ็นเอมีการแสดงออกน้อยลงเพื่อทำให้ตรวจพบทางพอลิอะดีนีนได้ง่ายขึ้น เช่น ยีน *pnp* ที่ถูกแปลรหัสเป็น PNPase ที่สามารถย่อยอาร์เอ็นเอที่ปลาย 3' หรือ ยีน *rne* ที่ถูกแปลรหัสเป็นเอนไซม์ Ribonuclease E ซึ่งเกี่ยวข้องกับการตัดสายอาร์เอ็นเอภายใน หรือใช้แบคทีเรียกลายพันธุ์ยีน *pcnB* ที่ถูกแปลรหัส PAP I สำหรับการสังเคราะห์ทางพอลิอะดีนีนส่งผลให้ทางพอลิอะดีนีนบนอาร์เอ็นเอมีการแสดงออกน้อยลง (Maes et al., 2017) ส่งผลสามารถพบอาร์เอ็นเอของยีนที่ตกเป็นเป้าหมายของไรโบนิวคลีเอสสำหรับการย่อยสลายผ่านการทางพอลิอะดีนีนมากขึ้น แต่อาจกระทบต่อความยาวทางพอลิอะดีนีน และยังมีการทดลองที่ใช้สายพันธุ์ยีน *pcnB* มีการแสดงออกมากขึ้น (Mohanty & Kushner, 2006) ทำให้เอนไซม์ PAP I ถูกแปลรหัสมากขึ้น แม้จะส่งผลให้มีความยาวทางพอลิอะดีนีนมากขึ้น แต่การที่อาร์เอ็นเอมีทางพอลิอะดีนีนยาวมากขึ้นอาจทำให้ตกเป็นเป้าหมายในการถูกย่อยสลายจากเอนไซม์ไรโบนิวคลีเอสมากขึ้นด้วย ทำให้ปริมาณการแสดงออกอาจต่ำลง

การศึกษาครั้งนี้ใช้ *E. coli* K-12 พันธุ์ป่าในสภาวะปกติเท่านั้น ซึ่งอาจส่งผลให้การตรวจหาอาร์เอ็นเอที่มีหางพอลิอะดีนีนได้จำนวนน้อยกว่า และอาจได้ความยาวที่สั้นกว่าด้วย เนื่องจากเอนไซม์ไรโบนิวคลีเอสสำหรับการย่อยสลายสามารถทำงานได้ปกติ ทำให้การย่อยอาร์เอ็นเอผ่านทางพอลิอะดีนีนสามารถเกิดขึ้นได้ ฉะนั้นในการตรวจหาหางพอลิอะดีนีนในแบคทีเรียพันธุ์ป่าจึงมีความท้าทายอย่างมาก ซึ่งจากงานวิจัยก่อนหน้าพบว่าในอาร์เอ็นเอของ *E. coli* พันธุ์ป่ามีเพียงอาร์เอ็นเอ 10% ของยีน *rpsO* เท่านั้นที่จะสามารถพบหางพอลิอะดีนีน และในอาร์เอ็นเอของ *rpsO* ที่พบหางพอลิอะดีนีน มักพบยาวเพียง 1-3 นิวคลีโอไทด์เท่านั้น (Marujo et al., 2000) นอกจากนี้หางพอลิอะดีนีนไม่ได้มีอยู่แค่ที่ปลาย 3' แต่สามารถเกิดขึ้นที่ตรงกลางของอาร์เอ็นเอหลังจากถูกตัดด้วยเอนไซม์เอนโดไรโบนิวคลีเอส เพื่อกำจัดโครงสร้างทุติยภูมิที่อยู่ภายในสายอาร์เอ็นเอ ทำให้การตรวจสอบหางพอลิอะดีนีนสามารถทำได้ยากมากขึ้นและความซับซ้อนในการศึกษาขึ้น (Sarkar, 1997)

### 5.3 ประสิทธิภาพของการหาลำดับเบสอาร์เอ็นเอด้วยอ็อกซ์ฟอร์ดนาโนพอร์

การหาลำดับเบสด้วยนาโนพอร์ยังคงเป็นที่ถูกพูดถึงในปัจจุบันในแง่ของการที่แลกมากับสายลำดับเบสที่ได้ยาวมากขึ้นกับความคลาดเคลื่อนที่สูงขึ้น เนื่องจากความผิดพลาดในการหาลำดับเบสที่สูงเกือบ 10% เมื่อเทียบกับ NGS ที่ไม่ถึง 1% (Delahaye & Nicolas, 2021) ส่งผลให้เกือบทุกงานวิจัยที่ศึกษาเกี่ยวกับการหาลำดับเบสอาร์เอ็นเอด้วยนาโนพอร์ต้องใช้ข้อมูลจากการหาลำดับเบสสายสั้นด้วย NGS ควบคู่ไปด้วยเพื่อเป็นหลักฐานรองรับ โดยเฉพาะในงานที่เกี่ยวข้องกับการศึกษาเชิงเกลินิวคลีโอไทด์โพลีมอร์ฟิซึม (single-nucleotide polymorphism, SNP) ซึ่งความผิดพลาดของลำดับเบสเพียงเล็กน้อยสามารถส่งผลให้ข้อมูลความคลาดเคลื่อนได้ (Ahsan, Liu, Fang, & Wang, 2021) อีกทั้งอาร์เอ็นเอซึ่งมีความเสถียรต่ำ เนื่องจากโครงสร้างที่เป็นสายเดี่ยวและหมู่ไฮดรอกซิลที่คาร์บอนตำแหน่งที่สองที่มีความใกล้ชิดกับหมู่ฟอสเฟต ทำให้สามารถเกิดพันธะต่ออ็อกซิเจนที่ฟอสเฟตได้ ส่งผลให้สายอาร์เอ็นเอเกิดการแตกหักได้ง่ายกว่าในดีเอ็นเอ (Rich, 2009) ซึ่งส่งผลให้ความยาวของสายอาร์เอ็นเอสั้นลงกว่าปกติ ส่งผลให้จำนวนสายลำดับเบสได้หลังการหาลำดับเบสมักน้อยกว่าในการศึกษาดีเอ็นเอ นอกจากนี้ความไม่เสถียรของอาร์เอ็นเออาจส่งผลให้สัญญาณไฟฟ้าที่เปลี่ยนแปลงมีรูปแบบที่แตกต่างจากดีเอ็นเอ เนื่องจากโครงสร้างของโมเลกุลอาร์เอ็นเอต่างจากดีเอ็นเอ และอาจส่งผลให้เกิดปัญหาในบางโปรแกรม เช่น nanopolish event align (Loman et al., 2015) ที่สามารถจัดตำแหน่งสายลำดับเบสกับตำแหน่งสัญญาณไฟฟ้า ซึ่งพบว่าเมื่อใช้ในการศึกษาอาร์เอ็นเอมีการจัดผิดตำแหน่ง โดยอาจเนื่องมาจากการหาช่วงสัญญาณไฟฟ้าที่ตรงกับลำดับเบสที่เป็นไปได้มาก

ที่สุดที่ปลาย 3' ทำให้คลาดเคลื่อนไปตรงกับช่วงสัญญาณก่อนที่จะถึงช่วงของที่ปรากฏในจีโนมเช่นในอาร์เอ็นเอของยีน *apt* ที่ไปตรงในช่วงพอลิยูริดีนที่ถูกสังเคราะห์ดังรูปที่ 24 ในส่วนของประสิทธิภาพของโปรแกรมที่ใช้ในการประมาณความยาวของหางพอลิอะดีนีนในการศึกษาครั้งนี้พบว่าโปรแกรม *tailfindr* ที่พึ่งพาสัญญาณไฟฟ้าในการประมาณความยาวหางพอลิอะดีนีนเพียงอย่างเดียว มักประมาณความยาวหางพอลิอะดีนีนในอาร์เอ็นเอของ *E. coli* สูงผิดปกติ อันเนื่องมาจากการประมาณสัญญาณที่ผิดพลาด เพราะความยาวสัญญาณไฟฟ้าของช่วงหางพอลิอะดีนีนที่อาจสั้นมากหรือไม่มีเลย ในขณะที่โปรแกรม *nanopolish* ที่ใช้ข้อมูลการจัดตำแหน่งลำดับเบสควบคู่ไปกับสัญญาณไฟฟ้า ซึ่งให้ผลลัพธ์แม่นยำสูงกว่าเมื่อตรวจสอบหางพอลิอะดีนีนสายสั้น โดยประมาณว่าอาร์เอ็นเอส่วนใหญ่ใน *E. coli* ไม่มีหางพอลิอะดีนีน อย่างไรก็ตามโปรแกรมอาจมีความผิดพลาด โดยการประมาณความยาวหางพอลิอะดีนีนสั้นกว่าปกติในบางครั้ง

จากการตรวจสอบฮอโมพอลิเมอร์ที่เชื่อมต่อระหว่างช่วงสัญญาณเอ็มอาร์เอ็นเอและพอลิยูริดีนที่ถูกสังเคราะห์ ซึ่งสัญญาณไฟฟ้างี้สามารถนำไปสู่การสันนิษฐานถึงการมีอยู่ของหางพอลิอะดีนีนได้ และการตรวจสอบมีอยู่ของโครงสร้างทุติยภูมิสำหรับการยุติการถอดรหัส ด้วยจากการสืบค้นในฐานข้อมูล *Ecocyc* (Keseler et al., 2011) หรือทำนายจากลำดับเบสที่ปลาย 3' ที่ไม่ถูกแปลรหัสด้วย *RNAfold* (Lorenz et al., 2011) รวมถึงสืบค้นรายงานการจับของ *Hfq* ในแต่ละยีน โดยคุณลักษณะเหล่านี้มักพบในอาร์เอ็นเอของยีนที่มีหางพอลิอะดีนีน โดยในการศึกษาครั้งนี้พบสายลำดับเบสของยีน *rplQ* ที่ตรงกับสมมติฐานทุกประการ โดย *tailfindr* และ *nanopolish* สามารถประมาณความยาวหางพอลิอะดีนีนได้สอดคล้องกัน และพบหางพอลิอะดีนีนในช่วงลำดับเบสที่ไม่ปรากฏในจีโนม รวมถึงพบสัญญาณฮอโมพอลิเมอร์ที่ต่อจากพอลิยูริดีนที่ถูกสังเคราะห์ และมีผลลัพธ์จากการจัดตำแหน่งลำดับเบสและสัญญาณไฟฟ้าตรงตามตำแหน่งที่คาดหวังคือจัดตำแหน่งลำดับเบสเริ่มต้นหลังจากสัญญาณหางพอลิอะดีนีนที่ถูกทำนาย หรือในช่วงเริ่มต้นของช่วงสัญญาณไฟฟ้าที่มีลำดับเบสปรากฏในจีโนมอ้างอิง

จากการศึกษาค่าสถิติของลำดับเบสหลังพบว่าปริมาณจำนวนสายลำดับเบสต่ำในชุดข้อมูลที่เตรียมไลบรารีด้วยเอนไซม์ Poly(U) polymerase ซึ่งให้ผลลัพธ์จำนวนสายลำดับเบสทั้งหมดที่น้อยกว่าถึง 100-1000 เท่า เมื่อเทียบกับชุดข้อมูลที่เตรียมไลบรารีด้วยเอนไซม์ Poly(A) polymerase ซึ่งเป็นการเตรียมไลบรารีแบบมาตรฐาน และนำไปสู่จำนวนยีนที่ถูกพบที่ไม่ถึง 10% ของยีนทั้งหมดใน *E. coli* ในชุดข้อมูลของ Poly(U) polymerase ซึ่งคาดว่ามาจากประสิทธิภาพในการของการเติมยูริดีนด้วยเอนไซม์ Poly(U) polymerase โดยแม้เพิ่มเวลาในขั้นตอนการสังเคราะห์พอลิยูริดีนเป็น 90

นาที่ ถึงจะสามารถเพิ่มจำนวนสายลำดับเบสที่ได้รับ แต่พบว่าคุณภาพและความยาวสายลำดับเบสลดลงไปด้วยดังตารางที่ 5 ซึ่งเวลาในการเตรียมไลบรารีที่มากขึ้นอาจส่งผลให้เพิ่มโอกาสในการแตกหักของอาร์เอ็นเอที่มีความเสถียรต่ำ โดยการเกิดแตกหักของอาร์เอ็นเอพบทุกในชุดข้อมูล ส่งผลให้ไม่ได้ครอบคลุมทั้งยีนโดยเฉพาะที่ปลาย 3' ซึ่งเป็นจุดที่มีการสังเคราะห์ทางพอลิอะดีนีน ทำให้อาจส่งผลให้การประมาณความยาวทางพอลิอะดีนีนเกิดความผิดพลาดขึ้น

นอกจากนี้คุณภาพและความยาวของลำดับเบสของอาร์เอ็นเอยังถือว่าไม่เป็นไปตามที่คาดหวังแม้ในชุดข้อมูลที่เตรียมไลบรารีด้วย Poly(A) polymerase ที่เป็นมาตรฐาน เมื่อเทียบการวิจัยที่คล้ายคลึงกันก่อนหน้านี้ ซึ่งได้ค่าเฉลี่ยคุณภาพอยู่ที่ประมาณ 9-10 และ ความยาวเฉลี่ยอยู่ 700-900 นิวคลีโอไทด์ (Grünberger et al., 2022) ซึ่งอาจหมายถึงประสิทธิภาพในการทดลองอาจต้องมีการปรับให้เหมาะสมมากขึ้น

#### 5.4 วิสัยทัศน์ในการศึกษาในอนาคต

ในปัจจุบันได้มีการควบคุมการแสดงออกของยีนในโพรแคริโอตมากมายด้วยหลากหลายกรรมวิธี เช่น CRISPR-Cas, ยีนน็อกเอาต์, ยีนน็อกดาวน์ หรือพันธุวิศวกรรม และอื่นๆ แต่กับทางพอลิอะดีนีนที่มีความเกี่ยวข้องกับการแสดงออกของยีนโดยตรง โดยเป็นสัญญาณของการย่อยสลายอาร์เอ็นเอในโพรแคริโอต แต่กลับไม่เคยถูกนำมาใช้จริงในงานทางการประยุกต์แม้จะมีศักยภาพ เช่น การนำไปประยุกต์ใช้ยับยั้งการแสดงออกของยีนก่อโรค หรือยีนต้านยา ด้วยการเพิ่มความยาวของทางพอลิอะดีนีนในอาร์เอ็นเอแบบที่เรียกก่อโรค เพื่อเร่งการย่อยสลายผลิตภัณฑ์จากยีนดังกล่าว เนื่องจากการศึกษายังมีข้อจำกัดสูงมาก ส่วนหนึ่งมาจากในแบคทีเรียสังเคราะห์ทางความยาวพอลิอะดีนีนสั้น และเป็นสัญญาณของการย่อยสลายของอาร์เอ็นเอ (O'Hara et al., 1995; Sarkar, 1997) ทำให้พบอาร์เอ็นเอที่มีทางพอลิอะดีนีนได้น้อย (Marujo et al., 2000; Sarkar, 1997) อีกทั้งอาร์เอ็นเอในโพรแคริโอตยังมีการเติมแต่งอาร์เอ็นเอที่ปลาย 3' หลังการถอดรหัสนอกเหนือจากทางพอลิอะดีนีน เช่น การแต่งเติมแบบเฮเทอโรพอลิเมอร์เป็นต้น (Mohanty & Kushner, 2006) ทำให้มีการซับซ้อนในการศึกษามากขึ้น ส่งผลให้ในปัจจุบันยังไม่มีรายงานความยาวของทางพอลิอะดีนีนที่ครอบคลุมทุกยีนในโพรแคริโอตในฐานะข้อมูล

การมาถึงของอ็อกซ์ฟอร์ดนาโนพอร์ทำให้สามารถช่วยในการหาลำดับเบสของอาร์เอ็นเอแบบสายยาวได้ และให้ข้อมูลในระดับทั้งจีโนม ซึ่งมีศักยภาพมาใช้ประยุกต์ในการหาความยาวทางพอลิอะ

ดีนีนสายดั้งเดิม ซึ่งเป็นปัญหาที่พบในงานก่อนหน้าที่มีการเพิ่มสารพันธุกรรมซึ่งกระทบต่อความยาวดั้งเดิมของหางพอลิอะดีนีน (Jain et al., 2016; Krause et al., 2019) อย่างไรก็ตามข้อจำกัดที่เพิ่มขึ้นมาในการหาลำดับเบสอาร์เอ็นเอด้วยนาโนพอร์คือต้องการหางพอลิอะดีนีนที่มีความยาวมากพอเพื่อสามารถต่อกับอะแดปเตอร์ที่มีไทมีนนิวคลีโอไทด์เป็นองค์ประกอบเพื่อใช้ในการหาลำดับเบส ซึ่งเชื่อว่าอาร์เอ็นเอทุกชนิดจะมีหางพอลิอะดีนีน (Yang, Duff, Graveley, Carmichael, & Chen, 2011) หรือมีหางพอลิอะดีนีนยาวเพียงพอที่จะสามารถต่อกับไทมีนอะแดปเตอร์ได้ ในงานวิจัยก่อนหน้าได้มีการปรับปรุงขั้นตอนการเตรียมไลบรารีสำหรับหาลำดับเบสอาร์เอ็นเอด้วยนาโนพอร์ โดยทำการสังเคราะห์พอลิไอโนซีนแทนหางพอลิอะดีนีนที่ปลาย 3' และใช้ไซโตซีนอะแดปเตอร์เพื่อเชื่อมต่อกับพอลิไอโนซีนที่ถูกสังเคราะห์ ซึ่งขั้นตอนการเตรียมไลบรารีประสบความสำเร็จและสามารถใช้สำหรับการหาลำดับเบสได้ และให้ผลการทดลองใกล้เคียงกับการเตรียมไลบรารีแบบมาตรฐาน (Vo et al., 2021) อย่างไรก็ตามไอโนซีนเป็นนิวคลีโอไทด์ที่มีเบสเป็นรูปแบบพิวรีนเช่นเดียวกับอะดีนีนส่งผลให้สัญญาณไฟฟ้าจะคล้ายคลึงกัน ทำให้การประมาณความยาวหางพอลิอะดีนีนด้วยสัญญาณไฟฟ้าอาจเกิดความผิดพลาด และความยาวหางพอลิอะดีนีนอาจสูงกว่าปกติ เนื่องจากไม่สามารถแยกออกระหว่างพอลิไอโนซีนและหางพอลิอะดีนีน

การศึกษานี้จึงได้ปรับปรุงการเตรียมไลบรารีโดยใช้การสังเคราะห์พอลิยูรีดีนด้วย Poly(U) polymerase ซึ่งสามารถนำมาใช้แทนการเตรียมไลบรารีมาตรฐาน เนื่องจากยูรีดีนเป็นนิวคลีโอไทด์ที่มีเบสไพริมิดีนซึ่งแตกต่างจากอะดีนีนที่เป็นนิวคลีโอไทด์ที่มีเบสรูปแบบพิวรีน ทำให้สามารถตรวจหาช่วงสัญญาณได้สะดวกกว่าโดยพิจารณาจากจุดเปลี่ยนสัญญาณ ฮอโมพอลิเมอร์ และอาจช่วยให้โปรแกรมประมาณความยาวหางพอลิอะดีนีนสามารถทำงานทรงประสิทธิภาพมากขึ้น พบว่าการเตรียมไลบรารีโดยการสังเคราะห์พอลิยูรีดีนสามารถใช้ในการหาลำดับเบสได้ทั้งในอาร์เอ็นเอสังเคราะห์ และในอาร์เอ็นเอของสิ่งมีชีวิต

หางพอลิอะดีนีนได้รับความสนใจในการศึกษามากขึ้น โดยเริ่มมีการนำออกซ์ฟอร์ดนาโนพอร์มาประยุกต์ใช้ในการหาความยาวหางพอลิอะดีนีน (Workman et al., 2019) และได้มีการพัฒนาโปรแกรมสำหรับการประมาณความยาวหางพอลิอะดีนีนโดยเฉพาะ (Krause et al., 2019; Loman et al., 2015) ซึ่งมักนำไปใช้ศึกษากับหางพอลิอะดีนีนในยูแคริโอต เนื่องจากหางพอลิอะดีนีนมีความยาวที่สูง อย่างไรก็ตามโปรแกรมหังกล่าวสามารถประมาณหางพอลิอะดีนีนความยาวสั้นได้เช่นกัน แม้ประสิทธิภาพจะยังไม่แม่นยำเท่าการประมาณความยาวหางพอลิอะดีนีนที่มีความยาวสูง ในการศึกษาครั้งนี้จึงทำการพัฒนากระแสน้ำเพื่อปรับปรุงข้อมูลความยาวพอลิอะดีนีนที่ถูกประมาณให้มีความ

ใกล้เคียงกับความยาวที่แท้จริงมากขึ้นโดยเมื่อทดสอบกับอาร์เอ็นเอสังเคราะห์ที่มีความยาวหางพอลิอะดีนีนคิงที่ พบว่าผลลัพธ์ใกล้เคียงกับความยาวของหางพอลิอะดีนีนที่แท้จริง และมีค่าความคลาดเคลื่อนที่น้อยลงกว่าแบบที่ไม่ใช้กระแสนงานปรับปรุง อย่างไรก็ตามในอาร์เอ็นเอในสิ่งมีชีวิตมีความซับซ้อนกว่ามาก รวมถึงการมีอยู่ของหางพอลิอะดีนีนในโพรแคริโอตสามารถพบได้ยาก ทำให้กระแสนงานในการประมาณหางพอลิอะดีนีนอาจต้องมีการปรับปรุงเพื่อลดความคลาดเคลื่อนให้มากที่สุด

อ็อกซ์ฟอร์ดนาโนพอร์มีศักยภาพในการพัฒนาและสามารถเป็นเครื่องมือมาตรฐานที่ใช้ในการหาลำดับเบสสายยาวในอนาคตได้ ซึ่งปัจจุบันยังมีการวิจัยและพัฒนาทั้งฮาร์ดแวร์และซอฟต์แวร์นี้อย่างต่อเนื่อง อีกทั้งการศึกษาหางพอลิอะดีนีนด้วยวิธีการหาลำดับเบสสายยาวโดยตรงด้วยอ็อกซ์ฟอร์ดนาโนพอร์ในโพรแคริโอตครั้งนี้สามารถนำไปสู่การศึกษาเชิงลึกเกี่ยวกับความสัมพันธ์ระหว่างความยาวหางพอลิอะดีนีนและการแสดงออกของยีน โดยช่วยส่งเสริมองค์ความรู้พื้นฐานของความยาวหางพอลิอะดีนีนในโพรแคริโอต และสามารถนำไปประยุกต์ความเข้าใจถึงระบบวิวัฒนาการของอิมูโนภายในเซลล์เพื่อตอบสนองต่อสภาพแวดล้อม อีกทั้งยังมีศักยภาพที่สามารถนำไปใช้ประยุกต์กับสิ่งมีชีวิตโพรแคริโอตสายพันธุ์อื่นซึ่งมีความสามารถในก่อโรคได้ นำไปสู่การควบคุมการแสดงออกของยีนต่างๆ เช่น ยีนที่ก่อให้เกิดการติดเชื้อ หรือยีนที่ก่อให้เกิดโรคเพื่อใช้ประโยชน์ทางการแพทย์ได้ในอนาคต เป็นต้น

## บทที่ 6

### สรุปผลการวิจัยและข้อเสนอแนะ

ในงานวิจัยนี้แสดงให้เห็นว่าการเตรียมไลบรารีเพื่อหาลำดับเบสอาร์เอ็นเอด้วยอ็อกซ์ฟอร์ดนาโนพอร์แบบสายยาวด้วยการเตรียมไลบรารีโดยการสังเคราะห์พอลิยูรีดินด้วยเอนไซม์ Poly(U) polymerase สามารถทำได้ในอาร์เอ็นเอของสายสังเคราะห์และในอาร์เอ็นเอของสิ่งมีชีวิตโพรแคริโอต และได้พัฒนากระแสนงานทางคอมพิวเตอร์สำหรับการตรวจหาทางพอลิอะดีนีนสำหรับอาร์เอ็นเอที่ถูกหาลำดับเบสด้วยอ็อกซ์ฟอร์ดนาโนพอร์ นอกจากนี้ยังได้พัฒนากระแสนงานเพื่อปรับปรุงข้อมูลความยาวพอลิอะดีนีนที่ถูกประมาณให้มีความใกล้เคียงกับความยาวที่แท้จริงมากขึ้น และมีค่าความคลาดเคลื่อนที่น้อยลง

ในการศึกษาครั้งนี้ได้ตรวจพบอาร์เอ็นเอที่มีโอกาสพบทางพอลิอะดีนีนในอาร์เอ็นเอของ *E. coli* K-12 พันธุ์ป่าทั้งในยีนทรานสคริปต์และไม่เคยถูกทรานสคริปต์มาก่อนในการพบทางพอลิอะดีนีน โดยมักพบอาร์เอ็นเอของยีนที่มีโครงสร้างทุติยภูมิสำหรับการยุติการถอดรหัส หรือ ยีนที่อาร์เอ็นเอที่มีรายงานการจับโปรตีน Hfq หรือ ยีนที่มีการแสดงออกแบบเดี่ยว อย่างไรก็ตามเนื่องจากทางพอลิอะดีนีนในอาร์เอ็นเอของ *E. coli* พบได้ยากเพราะมักเข้ากระบวนการย่อยสลายไปแล้วด้วยไรโบนิวคลีเอส นอกจากนี้จำนวนลำดับเบสทั้งหมดที่ได้รับในการศึกษาครั้งนี้มีจำนวนน้อยกว่าปกติ ทำให้จำนวนของยีนที่ตรวจพบและจำนวนลำดับเบสในแต่ละยีนต่ำกว่าการเตรียมไลบรารีแบบมาตรฐาน ส่งผลให้จำนวนอาร์เอ็นเอที่มีทางพอลิอะดีนีนที่ตรวจพบมีจำนวนน้อยมาก

ถึงแม้การศึกษาครั้งนี้จะประสบความสำเร็จในการหาลำดับเบสอาร์เอ็นเอด้วยอ็อกซ์ฟอร์ดนาโนพอร์ โดยการเตรียมไลบรารีด้วย Poly(U) polymerase เพื่อสังเคราะห์พอลิยูรีดินและสามารถรักษาและรายงานความยาวดั้งเดิมของทางพอลิอะดีนีนในอาร์เอ็นเอของยีนของ *E. coli* K-12 พันธุ์ป่าในสภาวะปกติ อย่างไรก็ตามวิธีการทดลองยังต้องมีการปรับปรุงคุณภาพเพื่อเพิ่มจำนวนข้อมูลสายลำดับเบสสำหรับการวิเคราะห์มากขึ้น ถึงกระนั้นการศึกษานี้ได้แสดงให้เห็นถึงความเป็นไปได้ถึงการศึกษาคความยาวทางพอลิอะดีนีนในโพรแคริโอต และพัฒนาเพื่อใช้ในการศึกษาต่อเป็นทางเลือกในการควบคุมการแสดงออกของยีนผ่านทางพอลิอะดีนีนไม่ใช่แค่ใน *E. coli* แต่รวมถึงสิ่งมีชีวิตโพรแคริโอตชนิดอื่น และสามารถนำไปต่อยอดเพื่อประยุกต์ใช้ในหลากหลายสาขาเช่น การแพทย์ เทคโนโลยีชีวภาพ และการเกษตร เป็นต้น



## บรรณานุกรม

- Ahsan, M. U., Liu, Q., Fang, L., & Wang, K. (2021). NanoCaller for accurate detection of SNPs and indels in difficult-to-map regions from long-read sequencing by haplotype-aware deep neural networks. *Genome Biology*, 22(1), 261. doi:10.1186/s13059-021-02472-2
- August, J. T., Ortiz, P. J., & Hurwitz, J. (1962). Ribonucleic acid-dependent ribonucleotide incorporation. I. Purification and properties of the enzyme. *Journal of Biological Chemistry*, 237, 3786-3793.
- Bilusic, I., Popitsch, N., Rescheneder, P., Schroeder, R., & Lybecker, M. (2014). Revisiting the coding potential of the E. coli genome through Hfq co-immunoprecipitation. *RNA biology*, 11(5), 641-654. doi:10.4161/rna.29299
- Brown, J. W., & Reeve, J. N. (1985). Polyadenylated, noncapped RNA from the archaeobacterium *Methanococcus vannielii*. *Journal of Bacteriology*, 162(3), 909-917. doi:10.1128/jb.162.3.909-917.1985
- Bunnik, E. M., & Le Roch, K. G. (2013). An Introduction to Functional Genomics and Systems Biology. *Adv Wound Care (New Rochelle)*, 2(9), 490-498. doi:10.1089/wound.2012.0379
- Crossley, B. M., Bai, J., Glaser, A., Maes, R., Porter, E., Killian, M. L., . . . Toohey-Kurth, K. (2020). Guidelines for Sanger sequencing and molecular assay monitoring. *Journal of Veterinary Diagnostic Investigation*, 32(6), 767-775. doi:10.1177/1040638720905833
- Dávila López, M., & Samuelsson, T. (2008). Early evolution of histone mRNA 3' end processing. *RNA (New York, N.Y.)*, 14(1), 1-10. doi:10.1261/rna.782308
- De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M., & Van Broeckhoven, C. (2018). NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics*, 34(15), 2666-2669. doi:10.1093/bioinformatics/bty149
- Deamer, D., Akeson, M., & Branton, D. (2016). Three decades of nanopore sequencing. *Nature Biotechnology*, 34(5), 518-524. doi:10.1038/nbt.3423
- Delahaye, C., & Nicolas, J. (2021). Sequencing DNA with nanopores: Troubles and biases.

- PLoS One*, 16(10), e0257521. doi:10.1371/journal.pone.0257521
- Douchin, V., Bohn, C., & Bouloc, P. (2006). Down-regulation of Porins by a Small RNA Bypasses the Essentiality of the Regulated Intramembrane Proteolysis Protease RseP in *Escherichia coli*\*. *Journal of Biological Chemistry*, 281(18), 12253-12259. doi:https://doi.org/10.1074/jbc.M600819200
- Edmonds, M., & Abrams, R. (1960). Polynucleotide biosynthesis: formation of a sequence of adenylate units from adenosine triphosphate by an enzyme from thymus nuclei. *Journal of Biological Chemistry*, 235, 1142-1149.
- Feng, Y., Huang, H., Liao, J., & Cohen, S. N. (2001). *Escherichia coli* poly(A)-binding proteins that interact with components of degradosomes or impede RNA decay mediated by polynucleotide phosphorylase and RNase E. *Journal of Biological Chemistry*, 276(34), 31651-31656. doi:10.1074/jbc.M102855200
- Freddolino, P. L., Amini, S., & Tavazoie, S. (2012). Newly identified genetic variations in common *Escherichia coli* MG1655 stock cultures. *Journal of Bacteriology*, 194(2), 303-306. doi:10.1128/jb.06087-11
- Garalde, D. R., Snell, E. A., Jachimowicz, D., Sipos, B., Lloyd, J. H., Bruce, M., . . . Turner, D. J. (2018). Highly parallel direct RNA sequencing on an array of nanopores. *Nature Methods*, 15(3), 201-206. doi:10.1038/nmeth.4577
- Goodrich, A. F., & Steege, D. A. (1999). Roles of polyadenylation and nucleolytic cleavage in the filamentous phage mRNA processing and decay pathways in *Escherichia coli*. *RNA (New York, N.Y.)*, 5(7), 972-985. doi:10.1017/s1355838299990398
- Grünberger, F., Ferreira-Cerca, S., & Grohmann, D. (2022). Nanopore sequencing of RNA and cDNA molecules in *Escherichia coli*. *RNA (New York, N.Y.)*, 28(3), 400-417. doi:10.1261/rna.078937.121
- Guhaniyogi, J., & Brewer, G. (2001). Regulation of mRNA stability in mammalian cells. *Gene*, 265(1-2), 11-23. doi:10.1016/s0378-1119(01)00350-x
- Hadjivassiliou, A., & Brawerman, G. (1966). Polyadenylic acid in the cytoplasm of rat liver. *Journal of Molecular Biology*, 20(1), 1-7. doi:10.1016/0022-2836(66)90112-4
- Hajnsdorf, E., Braun, F., Haugel-Nielsen, J., & Régnier, P. (1995). Polyadenylation destabilizes the rpsO mRNA of *Escherichia coli*. *Proceedings of the National*

*Academy of Sciences of the United States of America*, 92(9), 3973-3977.

doi:10.1073/pnas.92.9.3973

Hajnsdorf, E., & Kaberdin, V. R. (2018). RNA polyadenylation and its consequences in prokaryotes. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 373(1762). doi:10.1098/rstb.2018.0166

Hajnsdorf, E., & Régnier, P. (2000). Host factor Hfq of *Escherichia coli* stimulates elongation of poly(A) tails by poly(A) polymerase I. *Proceedings of the National Academy of Sciences of the United States of America*, 97(4), 1501-1505.

doi:10.1073/pnas.040549897

Hankins, J. S., Denroche, H., & Mackie, G. A. (2010). Interactions of the RNA-Binding Protein Hfq with *cspA* mRNA, Encoding the Major Cold Shock Protein. *Journal of Bacteriology*, 192(10), 2482-2490. doi:10.1128/JB.01619-09

Hector, R. E., Nykamp, K. R., Dheur, S., Anderson, J. T., Non, P. J., Urbinati, C. R., . . .

Swanson, M. S. (2002). Dual requirement for yeast hnRNP Nab2p in mRNA poly(A) tail length control and nuclear export. *The EMBO Journal*, 21(7), 1800-1810. doi:10.1093/emboj/21.7.1800

Jain, M., Olsen, H. E., Paten, B., & Akeson, M. (2016). The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biology*, 17(1), 239. doi:10.1186/s13059-016-1103-0

Jonkhout, N., Tran, J., Smith, M. A., Schonrock, N., Mattick, J. S., & Novoa, E. M. (2017). The RNA modification landscape in human disease. *RNA (New York, N.Y.)*, 23(12), 1754-1769. doi:10.1261/rna.063503.117

Kalapos, M. P., Cao, G. J., Kushner, S. R., & Sarkar, N. (1994). Identification of a Second Poly(A) Polymerase in *Escherichia coli*. *Biochemical and Biophysical Research Communications*, 198(2), 459-465. doi:https://doi.org/10.1006/bbrc.1994.1067

Karnik, P., Taljanidisz, J., Sasvari-Szekely, M., & Sarkar, N. (1987). 3'-terminal polyadenylate sequences of *Escherichia coli* tryptophan synthetase  $\alpha$ -subunit messenger RNA. *Journal of Molecular Biology*, 196(2), 347-354.

doi:https://doi.org/10.1016/0022-2836(87)90695-4

Keseler, I. M., Collado-Vides, J., Santos-Zavaleta, A., Peralta-Gil, M., Gama-Castro, S.,

- Muñiz-Rascado, L., . . . Karp, P. D. (2011). EcoCyc: a comprehensive database of *Escherichia coli* biology. *Nucleic acids research*, *39*(Database issue), D583-590. doi:10.1093/nar/gkq1143
- Komine, Y., Kwong, L., Anguera, M. C., Schuster, G., & Stern, D. B. (2000). Polyadenylation of three classes of chloroplast RNA in *Chlamydomonas reinhardtii*. *RNA (New York, N.Y.)*, *6*(4), 598-607. doi:10.1017/s1355838200992252
- Krause, M., Niazi, A. M., Labun, K., Torres Cleuren, Y. N., Müller, F. S., & Valen, E. (2019). tailfindr: alignment-free poly(A) length measurement for Oxford Nanopore RNA and DNA sequencing. *RNA (New York, N.Y.)*, *25*(10), 1229-1241. doi:10.1261/rna.071332.119
- Kuczynski, J., Lauber, C. L., Walters, W. A., Parfrey, L. W., Clemente, J. C., Gevers, D., & Knight, R. (2012). Experimental and analytical tools for studying the human microbiome. *Nature Reviews Genetics*, *13*(1), 47-58. doi:10.1038/nrg3129
- Le Derout, J., Folichon, M., Briani, F., Dehò, G., Régnier, P., & Hajnsdorf, E. (2003). Hfq affects the length and the frequency of short oligo(A) tails at the 3' end of *Escherichia coli* rpsO mRNAs. *Nucleic acids research*, *31*(14), 4017-4023. doi:10.1093/nar/gkg456
- Levene, M. J., Korlach, J., Turner, S. W., Foquet, M., Craighead, H. G., & Webb, W. W. (2003). Zero-mode waveguides for single-molecule analysis at high concentrations. *Science*, *299*(5607), 682-686. doi:10.1126/science.1079700
- Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094-3100. doi:10.1093/bioinformatics/bty191
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., . . . Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078-2079. doi:10.1093/bioinformatics/btp352
- Loman, N. J., Quick, J., & Simpson, J. T. (2015). A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nature Methods*, *12*(8), 733-735. doi:10.1038/nmeth.3444
- Lopilato, J., Bortner, S., & Beckwith, J. (1986). Mutations in a new chromosomal gene of *Escherichia coli* K-12, pcnB, reduce plasmid copy number of pBR322 and its derivatives. *Molecular & general genetics*, *205*(2), 285-290.

doi:10.1007/bf00430440

Lorenz, R., Bernhart, S. H., Höner Zu Siederdisen, C., Tafer, H., Flamm, C., Stadler, P. F., & Hofacker, I. L. (2011). ViennaRNA Package 2.0. *Algorithms for Molecular Biology*, 6, 26-26. doi:10.1186/1748-7188-6-26

Maes, A., Gracia, C., Innocenti, N., Zhang, K., Aurell, E., & Hajnsdorf, E. (2017). Landscape of RNA polyadenylation in *E. coli*. *Nucleic acids research*, 45(5), 2746-2756. doi:10.1093/nar/gkw894

Mardis, E. R. (2008). Next-generation DNA sequencing methods. *Annual Review of Genomics and Human Genetics*, 9, 387-402. doi:10.1146/annurev.genom.9.081307.164359

Marujo, P. E., Hajnsdorf, E., Le Derout, J., Andrade, R., Arraiano, C. M., & Régnier, P. (2000). RNase II removes the oligo(A) tails that destabilize the rpsO mRNA of *Escherichia coli*. *RNA (New York, N.Y.)*, 6(8), 1185-1193. doi:10.1017/s135583820000073x

Merriman, B., & Rothberg, J. M. (2012). Progress in ion torrent semiconductor chip based sequencing. *Electrophoresis*, 33(23), 3397-3417. doi:10.1002/elps.201200424

Miczak, A., Kaberdin, V. R., Wei, C. L., & Lin-Chao, S. (1996). Proteins associated with RNase E in a multicomponent ribonucleolytic complex. *Proceedings of the National Academy of Sciences of the United States of America*, 93(9), 3865-3869. doi:10.1073/pnas.93.9.3865

Mildenhall, K. B., Wiese, N., Chung, D., Maples, V. F., Mohanty, B. K., & Kushner, S. R. (2016). RNase E-based degradosome modulates polyadenylation of mRNAs after Rho-independent transcription terminators in *Escherichia coli*. *Molecular microbiology*, 101(4), 645-655. doi:10.1111/mmi.13413

Mohanty, B. K., & Kushner, S. R. (2006). The majority of *Escherichia coli* mRNAs undergo post-transcriptional modification in exponentially growing cells. *Nucleic acids research*, 34(19), 5695-5704. doi:10.1093/nar/gkl684

Munroe, D., & Jacobson, A. (1990). mRNA poly(A) tail, a 3' enhancer of translational initiation. *Molecular and Cellular Biology*, 10(7), 3441-3455. doi:10.1128/mcb.10.7.3441-3455.1990

Nagaike, T., Suzuki, T., & Ueda, T. (2008). Polyadenylation in mammalian mitochondria:

- insights from recent studies. *Biochimica et Biophysica Acta*, 1779(4), 266-269.  
doi:10.1016/j.bbagr.2008.02.001
- Nakazato, H., Venkatesan, S., & Edmonds, M. (1975). Polyadenylic acid sequences in *E. coli* messenger RNA. *Nature*, 256(5513), 144-146. doi:10.1038/256144a0
- O'Hara, E. B., Chekanova, J. A., Ingle, C. A., Kushner, Z. R., Peters, E., & Kushner, S. R. (1995). Polyadenylation helps regulate mRNA decay in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, 92(6), 1807-1811. doi:10.1073/pnas.92.6.1807
- Ohta, N., Sanders, M., & Newton, A. (1975). Poly(adenylic acid) sequences in the RNA of *Caulobacter crescentus*. *Proceedings of the National Academy of Sciences of the United States of America*, 72(6), 2343-2346. doi:10.1073/pnas.72.6.2343
- Otaka, H., Ishikawa, H., Morita, T., & Aiba, H. (2011). PolyU tail of rho-independent terminator of bacterial small RNAs is essential for Hfq action. *Proceedings of the National Academy of Sciences of the United States of America*, 108(32), 13059-13064. doi:10.1073/pnas.1107050108
- Proudfoot, N. (1991). Poly(A) signals. *Cell*, 64(4), 671-674. doi:10.1016/0092-8674(91)90495-k
- Rang, F. J., Kloosterman, W. P., & de Ridder, J. (2018). From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. *Genome Biology*, 19(1), 90. doi:10.1186/s13059-018-1462-9
- Rhoads, A., & Au, K. F. (2015). PacBio Sequencing and Its Applications. *Genomics Proteomics Bioinformatics*, 13(5), 278-289. doi:10.1016/j.gpb.2015.08.002
- Rich, A. (2009). The era of RNA awakening: structural biology of RNA in the early years. *Quarterly Reviews of Biophysics*, 42(2), 117-137. doi:10.1017/s0033583509004776
- Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, 29(1), 24-26. doi:10.1038/nbt.1754
- Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, 74(12), 5463-5467. doi:10.1073/pnas.74.12.5463
- Sarkar, N. (1997). Polyadenylation of mRNA in prokaryotes. *Annual Review of*

- Biochemistry*, 66, 173-197. doi:10.1146/annurev.biochem.66.1.173
- Senear, A. W., & Steitz, J. A. (1976). Site-specific interaction of Qbeta host factor and ribosomal protein S1 with Qbeta and R17 bacteriophage RNAs. *Journal of Biological Chemistry*, 251(7), 1902-1912.
- Sherman, B. T., Hao, M., Qiu, J., Jiao, X., Baseler, M. W., Lane, H. C., . . . Chang, W. (2022). DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic acids research*. doi:10.1093/nar/gkac194
- Sittka, A., Pfeiffer, V., Tedin, K., & Vogel, J. (2007). The RNA chaperone Hfq is essential for the virulence of *Salmonella typhimurium*. *Molecular microbiology*, 63(1), 193-217. doi:10.1111/j.1365-2958.2006.05489.x
- Taljanidisz, J., Karnik, P., & Sarkar, N. (1987). Messenger ribonucleic acid for the lipoprotein of the *Escherichia coli* outer membrane is polyadenylated. *Journal of Molecular Biology*, 193(3), 507-515. doi:10.1016/0022-2836(87)90263-4
- Truong, C., Oudre, L., & Vayatis, N. (2020). Selective review of offline change point detection methods. *Signal Processing*, 167, 107299. doi:https://doi.org/10.1016/j.sigpro.2019.107299
- van Dijk, E. L., Jaszczyszyn, Y., Naquin, D., & Thermes, C. (2018). The Third Revolution in Sequencing Technology. *Trends in Genetics*, 34(9), 666-681. doi:10.1016/j.tig.2018.05.008
- Vargas-Blanco, D. A., & Shell, S. S. (2020). Regulation of mRNA Stability During Bacterial Stress Responses. *Front Microbiol*, 11, 2111. doi:10.3389/fmicb.2020.02111
- Vo, J. M., Mulrone, L., Quick-Cleveland, J., Jain, M., Akesson, M., & Ares, M., Jr. (2021). Synthesis of modified nucleotide polymers by the poly(U) polymerase Cid1: application to direct RNA sequencing on nanopores. *RNA (New York, N.Y.)*, 27(12), 1497-1511. doi:10.1261/ma.078898.121
- Vytvytska, O., Jakobsen Jimmy, S., Balcunaite, G., Andersen Jens, S., Baccarini, M., & von Gabain, A. (1998). Host factor I, Hfq, binds to *Escherichia coli* ompA mRNA in a growth rate-dependent fashion and regulates its stability. *Proceedings of the National Academy of Sciences*, 95(24), 14118-14123. doi:10.1073/pnas.95.24.14118

- Walsh, A. P., Tock, M. R., Mallen, M. H., Kaberdin, V. R., von Gabain, A., & McDowall, K. J. (2001). Cleavage of poly(A) tails on the 3'-end of RNA by ribonuclease E of *Escherichia coli*. *Nucleic acids research*, *29*(9), 1864-1871. doi:10.1093/nar/29.9.1864
- Wang, Y., Zhao, Y., Bollas, A., Wang, Y., & Au, K. F. (2021). Nanopore sequencing technology, bioinformatics and applications. *Nature Biotechnology*, *39*(11), 1348-1365. doi:10.1038/s41587-021-01108-x
- Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, *10*(1), 57-63. doi:10.1038/nrg2484
- Weirather, J. L., de Cesare, M., Wang, Y., Piazza, P., Sebastiano, V., Wang, X. J., . . . Au, K. F. (2017). Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. *F1000Research*, *6*, 100. doi:10.12688/f1000research.10571.2
- Wick, R. R., Judd, L. M., & Holt, K. E. (2019). Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biology*, *20*(1), 129. doi:10.1186/s13059-019-1727-y
- Workman, R. E., Tang, A. D., Tang, P. S., Jain, M., Tyson, J. R., Razaghi, R., . . . Timp, W. (2019). Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nature Methods*, *16*(12), 1297-1305. doi:10.1038/s41592-019-0617-2
- Yamanaka, K., & Inouye, M. (2001). Selective mRNA degradation by polynucleotide phosphorylase in cold shock adaptation in *Escherichia coli*. *Journal of Bacteriology*, *183*(9), 2808-2816. doi:10.1128/JB.183.9.2808-2816.2001
- Yang, L., Duff, M. O., Graveley, B. R., Carmichael, G. G., & Chen, L.-L. (2011). Genomewide characterization of non-polyadenylated RNAs. *Genome Biology*, *12*(2), R16. doi:10.1186/gb-2011-12-2-r16
- Zhao, J., Hyman, L., & Moore, C. (1999). Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiology and Molecular Biology Reviews*, *63*(2), 405-445.





จุฬาลงกรณ์มหาวิทยาลัย  
**CHULALONGKORN UNIVERSITY**

## ประวัติผู้เขียน

ชื่อ-สกุล	Jirachote Karunyakorn
วัน เดือน ปี เกิด	29 June 1997
สถานที่เกิด	Thailand
วุฒิการศึกษา	Mahidol University, B.sc.
ที่อยู่ปัจจุบัน	62/152 Parkert Nontaburi 11120



จุฬาลงกรณ์มหาวิทยาลัย  
CHULALONGKORN UNIVERSITY