
ชีวสารสนเทศ : การประยุกต์ใช้ในงานวิจัยวิทยาศาสตร์ชีวภาพ
Bioinformatics : Application in Biological Science Research

ชมภูนุช วิรุณานนท์ และ วรวุฒิ จุฬาลักษณ์านุกูล*
ภาควิชาพฤกษศาสตร์ คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
Chompunuch Virunanon and Warawut Chulalaksananukul*
Department of Botany, Faculty of Science, Chulalongkorn University.

บทคัดย่อ

ชีวสารสนเทศเป็นเครื่องมือสำคัญในการศึกษาอนุชีววิทยาและเป็นศาสตร์ที่ผนวกความรู้หลายด้านทั้งทางชีววิทยา เคมี คณิตศาสตร์และคอมพิวเตอร์ชีววิทยา ซึ่งมีความเกี่ยวเนื่องกันอย่างไม่สามารถแยกจากกันได้ ปัจจุบันโปรแกรมทางชีวสารสนเทศจำนวนมากเกิดขึ้นเพื่อรองรับความต้องการการใช้งานประยุกต์ต่างๆเนื่องจากข้อมูลที่ได้จากการศึกษาทางชีววิทยาทั่วโลกที่มีเพิ่มขึ้นเป็นจำนวนมากขึ้นในแต่ละวัน การเรียนรู้พื้นฐานและความเข้าใจในการใช้งานจึงมีความสำคัญและเป็นเครื่องมือสู่ความสำเร็จในการทำงานของนักอนุชีววิทยาและผู้ทำงานวิจัยที่เกี่ยวข้อง โดยในบทความนี้กล่าวถึงวัตถุประสงค์หลักสามอย่าง โดยมีการวิเคราะห์ข้อมูลทางพันธุกรรม การออกแบบยา รวมถึงประยุกต์ใช้ในทางอนุชีววิทยาและวิศวกรรมเมแทบอลิก เป็นประเด็นหลัก

คำสำคัญ : ชีวสารสนเทศ อนุชีววิทยา

Abstract

Bioinformatics is an important tool in molecular biology studies and is a multidisciplinary subject that involving several biology, chemistry, mathematics and computational biology which could not be disassociated. Currently, a lot of bioinformatics programs are available for many purposes. Moreover, raw data from biological research are constantly increasing. Learning of basic and principal of bioinformatics are very important things to use these tools for molecular biologist's and relevant works. This topic described the main three objectives. The application in genetic analysis and annotation, medicine design, molecular biology and metabolic engineering are focused.

Keywords : bioinformatics, molecular biology

Corresponding author. E-mail: warawut.c@chula.ac.th

เมื่อก้าวถึงกระบวนการทางสารสนเทศ อาจกล่าวได้ว่ามีส่วนเกี่ยวข้องกับชีวิตประจำวันของคนในยุคสมัยนี้เป็นอย่างมาก ต่างจากในอดีตที่อาจอยู่ในรูปแบบสื่อสิ่งพิมพ์หรืออื่นๆ ในปัจจุบันการพัฒนาทางสารสนเทศส่งผลทำให้การทำงานและดำเนินชีวิตเป็นไปอย่างสะดวกรวดเร็วมากขึ้น สารสนเทศที่รู้จักกันคืออยู่ในรูปแบบของการให้บริการแบบเครือข่ายที่ใช้ในการสื่อสารเป็นสากลทั้งในรูปแบบของอินเทอร์เน็ต รวมถึงสื่ออิเล็กทรอนิกส์ต่างๆ เป็นต้น เครือข่ายเหล่านี้ได้อำนวยความสะดวกให้แก่ผู้ใช้ ตัวอย่างที่ชัดเจนพบเห็นได้ในการส่งจดหมายอิเล็กทรอนิกส์ ซึ่งปัจจุบันเพียงไม่กี่วินาทีก็สามารถติดต่อสื่อสารกับคนอีกซีกโลกได้ตามใจปรารถนา อีกทั้งค่าบริการเหล่านี้มีแนวโน้มมีเสถียรภาพใช้จ่ายน้อยลงเนื่องมาจากเหตุผลการแข่งขันทางธุรกิจ

กระบวนการทางสารสนเทศเป็นอย่างไร

สารสนเทศ (information) คือ ผลลัพธ์ของกระบวนการจัดการข้อมูล ซึ่งมักจะอยู่ในรูปฐานข้อมูล (database) ข้อมูลที่ยังไม่ได้ผ่านกระบวนการทางสารสนเทศจะอยู่ในรูปของข้อมูลดิบ เครือข่ายเป็นส่วนหนึ่งของสารสนเทศ สารสนเทศ (information) ที่ให้ความหมายโดย Debons et al., 1981 มีสองนัยคือ

1. Data Drive หมายถึง ข้อมูลดิบ
2. Cognitive Driven คือ การรับรู้และเรียนรู้ ซึ่งมีความหมายของข้อมูลมาเกี่ยวข้อง นำไปสู่การวิเคราะห์และสังเคราะห์ พัฒนาไปสู่ความรู้ได้ (wisdom)

รูปแบบของกระบวนการให้บริการทางสารสนเทศโดยผ่านทางอินเทอร์เน็ตและโปรแกรมคอมพิวเตอร์มีหลักการโดยย่อคือ จะต้องมีการบริการ (server) ซึ่งมีผู้ที่ทำหน้าที่บริหารจัดการ (administrator) และแจกจ่ายข้อมูล ซึ่งจะถูกจัดเก็บอยู่ในฐานข้อมูล (database) ให้แก่ผู้รับบริการ (client) ซึ่งสามารถติดต่อสื่อสารกับผู้ใช้บริการได้จากเครือข่าย (networks) ทั้งแบบมีสายและไร้สาย ส่วนประกอบหลักมีอยู่สี่ส่วนด้วยกัน คือ ฮาร์ดแวร์ ซอฟต์แวร์ ข้อมูล และบุคลากร ผู้รับบริการสามารถติดต่อและได้รับการบริการข้อมูล รวมถึงวิเคราะห์ประมวลผลข้อมูลจากผู้ให้บริการได้ และผู้จัดการระบบ (admin) จะทำหน้าที่ในการตรวจสอบและจัดเก็บข้อมูลให้ถูกต้องชัดเจนอยู่ในรูปแบบฐานข้อมูล การจัดการที่ดีจะทำให้ข้อมูลที่จัดเก็บมีความคงที่ สามารถเรียกกลับมาใช้ได้ง่าย ข้อมูลมีความทันสมัย และไม่ซ้ำซ้อน และทั้งหมดนี้จะต้องอยู่ในเงื่อนไขความปลอดภัย (security) เพื่อไม่ให้มีผู้ไม่ประสงค์ดีสามารถเข้ามาทำการดัดแปลงข้อมูลได้โดยง่าย ตัวอย่างเช่น การทำธุรกรรม

การเงินผ่านระบบอินเทอร์เน็ต ผู้ร้องขออาจต้องทำการแจ้งรหัสผ่าน (password) ต่อผู้ให้บริการก่อนที่จะเข้าไปตรวจสอบยอดเงินในบัญชีของตนเอง ซึ่งระบบความปลอดภัยนี้เป็นผลลัพธ์จากการตั้งเงื่อนไขและความคุ้มครองคำสั่งปฏิบัติการโดยผู้จัดการระบบ ข้อมูลดิบที่ผ่านเข้าสู่กระบวนการสารสนเทศแล้ว จะต้องมีความถูกต้องและมีรูปแบบที่ดูง่ายยิ่งขึ้นกว่าเดิม มีประโยชน์ตรงตามความต้องการของผู้ใช้ ซึ่งกระบวนการจัดการดังกล่าวก็เกิดขึ้นเช่นเดียวกันในการจัดการข้อมูลทางชีววิทยาที่เพิ่มปริมาณขึ้นมากมายในแต่ละวัน จึงเป็นที่มาของศาสตร์ทางด้าน “ชีวสารสนเทศ” ซึ่งจะกล่าวในหัวข้อการนำเทคโนโลยีสารสนเทศมาใช้ในงานวิจัยทางชีววิทยา

การนำเทคโนโลยีสารสนเทศมาใช้ในงานวิจัยทางชีววิทยา

เทคโนโลยีสารสนเทศ (Information technology) หมายถึงหมายถึงกระบวนการต่างๆ และระบบงานที่จัดการให้ได้สารสนเทศที่ต้องการ มักอยู่ในลักษณะของฐานข้อมูล เช่น ฐานข้อมูลผู้ป่วยในโรงพยาบาล ฐานข้อมูลการลงทะเบียนเรียนของนักศึกษาในมหาวิทยาลัย เป็นต้น ลักษณะของข้อมูลดิบดังกล่าว เมื่อปราศจากการจัดการด้วยกระบวนการทางสารสนเทศ มักจะมีปริมาณมาก ไม่เป็นระเบียบและยากต่อการเข้าถึง ปัจจุบันการศึกษาทางชีววิทยาปัจจุบันในระดับโมเลกุล นักวิจัยทั่วโลกที่ทำงานทางอนุชีววิทยาก็มีเป็นจำนวนมาก ทำให้ข้อมูลการวิจัยปริมาณมหาศาลถูกผลิตขึ้นในแต่ละวันไม่รู้จบสิ้นเช่นเดียวกัน ดังนั้นข้อมูลเหล่านี้ก็ต้องการวิธีการและระบบจัดการที่ดีเช่นกัน

ชีวสารสนเทศ เป็นศาสตร์การจัดการข้อมูลทางชีววิทยาซึ่งมีอยู่เป็นปริมาณมากเพื่อการประยุกต์ มีจุดกำเนิดมาจากโครงการ The Human Genome Project ในปี 2533 โดยการสนับสนุนของสำนักสุขภาพแห่งชาติและกรมพลังงานแห่งสหรัฐอเมริกา โครงการนี้เป็นโครงการทางเทคโนโลยีทางชีวภาพที่สร้างความเปลี่ยนแปลงครั้งสำคัญอันหนึ่งของมนุษยชาติ มีการร่วมมือในระดับนานาชาติเพื่อศึกษาวิเคราะห์ลำดับนิวคลีโอไทด์ในจีโนมมนุษย์ และต่อมาได้เกิดโครงการศึกษาจีโนมในสิ่งมีชีวิตอื่นอีกหลายชนิด เช่น จีโนมข้าว จีโนมหนู จีโนมยีสต์ เป็นต้น ทั้งหมดนี้ก่อให้เกิดการพัฒนาเทคโนโลยีรวมถึงวิทยาการคอมพิวเตอร์ ควบคู่กันจนเป็นที่มาของศาสตร์ชีวสารสนเทศ โดยอาศัยการนำผลการทดลองจากการวิจัยในห้องปฏิบัติการ (*in vitro* informations) บันทึกลงในฐานข้อมูลที่แยกประเภทชัดเจน และเมื่อต้องการออกแบบการทดลองใหม่หรือตั้งสมมติฐานใหม่ ก็นำข้อมูล

ดังกล่าวกลับมาช่วยวิเคราะห์ ทำนายสมมติฐานหรือสร้างต้นแบบ การทดลองใหม่ด้วยโปรแกรมคอมพิวเตอร์ (*in silico* modeling) ล่วงหน้าก่อนการลงมือทำปฏิบัติการจริงได้ นับว่าเป็นการพัฒนาการที่ล้ำคดียิ่ง ช่วยในการทำงานบางประเภทของ นักวิทยาศาสตร์ที่ในอดีตอาจเป็นไปได้ไม่ได้เลย ตลอดจนลด ค่าใช้จ่ายและเวลาในการทำงาน ปัจจุบันมีฐานข้อมูลทางชีววิทยา ให้เลือกใช้หลากหลายตามการใช้งาน ตัวอย่างเช่น ข้อมูลของ ลำดับนิวคลีโอไทด์ในจีโนมของจุลชีพ ข้อมูลลำดับกรดอะมิโนของ เอนไซม์เซลล์ในเรา ข้อมูลวิถีเมแทบอลิซึมของน้ำตาลเพนโทส ในแบคทีเรีย เป็นต้น ซึ่งถ้าจะแบ่งออกเป็นหมวดหมู่ก็อาจจำแนก เพื่อให้เห็นขอบเขตกว้างๆ ดังนี้ คือ

1. ฐานข้อมูลที่เกี่ยวข้องกับการศึกษาจีโนม ข้อมูล โครงสร้างและหน้าที่ของยีน
2. ฐานข้อมูลที่เกี่ยวข้องกับการศึกษาทรานสคริปโตม ข้อมูลการแปลรหัสของยีน
3. ฐานข้อมูลที่เกี่ยวข้องกับการศึกษาโปรตีโอม หรือข้อมูล จากการศึกษโปรตีนที่อยู่ในสิ่งมีชีวิต
4. ฐานข้อมูลที่เกี่ยวข้องกับการศึกษาเมตาโบโลม วิถีเมแทบอลิซึม

สำหรับนักชีววิทยาแล้ว การเลือกค้นข้อมูลจากฐาน ข้อมูลที่มีความเฉพาะเจาะจงกับงาน ถือเป็นเรื่องที่สำคัญยิ่ง และไม่ควรละเลย เพราะจะเพิ่มโอกาสได้ข้อมูลที่ สามารถตอบสนองความต้องการของผู้ใช้ได้ตรงตามวัตถุประสงค์การ ใช้งานได้ดีขึ้น หรืออาจเรียกว่ามีความเฉพาะเจาะจงมากขึ้น เนื่องจากฐานข้อมูลเหล่านี้ได้มีการคัดกรองข้อมูลแล้วในระดับหนึ่ง เช่น เมื่อมีความต้องการค้นข้อมูลลำดับของกรดอะมิโน (amino acid) ก็ควรใช้ข้อมูลในฐานข้อมูล SWISSPROT หรือ ExPASy จะเหมาะสมกว่า หรือหากจะสืบค้นหาลำดับเบสที่เกี่ยวข้องกับโรค Haemophilia B ก็ควรจะใช้ฐานข้อมูล HAEMB จะเข้าถึงข้อมูล ได้รวดเร็วกว่าใช้ Google ตามปกติ ตารางที่ 1 แสดงตัวอย่าง ฐานข้อมูลที่เฉพาะเจาะจงกับงานเฉพาะทางทางอนุชีววิทยา และพันธุศาสตร์

โดยทั่วไปฐานข้อมูลจะมีการติดต่อกับผู้ให้บริการในการ รับข้อมูล ให้บริการข้อมูล และวิเคราะห์ข้อมูลที่ได้รับเทียบกับ ฐานข้อมูลโดยมีบริการหลายอย่างเช่น ในฐานข้อมูล The National Center for Biotechnology Information (NCBI) มีบริการ PubMed ซึ่งบริการ Basic Local Alignment Search Tool (BLAST) แก่ผู้ใช้โดยให้ผู้บริการกรอกข้อมูลผลการอ่านลำดับ เบสของสารพันธุกรรมของงานวิจัยตนเองลงไปเพื่อทำการเทียบ

ลำดับความเหมือน ระหว่างข้อมูลของผู้รับบริการและข้อมูลลำดับ สารพันธุกรรมที่จัดเก็บในฐานข้อมูล เพื่อช่วยในการตัดสินใจหรือ วิเคราะห์ผลของผู้รับบริการ ปัจจุบันมีเว็บไซต์เป็นจำนวนมากที่ เกี่ยวข้องกับทางชีวสารสนเทศ ซึ่งมีจำนวนมากที่ให้บริการข้อมูล โดยไม่คิดค่าใช้จ่าย (ตารางที่ 2)

บทความนี้จะกล่าวถึงตัวอย่างการประยุกต์ใช้ใน 3 ลักษณะคือ วิเคราะห์ข้อมูลทางพันธุกรรมเพื่อการศึกษาทาง วิทยาศาสตร์บริสุทธิ์ (pure science) การออกแบบยา และ วิศวกรรมเมแทบอลิก

การวิเคราะห์ข้อมูลทางพันธุกรรม

การจัดเรียงลำดับนิวคลีโอไทด์หรือกรดอะมิโนจากสาย โพลีเปปไทด์หลายๆ สายของทั้งหมดหรือส่วนหนึ่งของสาย โพลีเปปไทด์ (alignment) เป็นข้อมูลพื้นฐานในการทำนาย โครงสร้างโปรตีน บริเวณที่จัดเรียงกันกันได้ดีที่สุดในโครงสร้าง จะมีการแสดงผลออกมาโดยการทำงานของโปรแกรมซึ่งกำหนดค่า โดยผู้ใช้ การจัดเรียงที่ดีที่สุดนั้นมักเป็นการพยายามที่จะทำให้ ทั้งจำนวนของช่องว่างและลำดับกรดอะมิโนหรือดีเอ็นเอที่เข้าคู่ผิด (mismatch) มีค่าน้อยที่สุด (เจษฎา เด่นดวงบริพันธ์, 2545)

การวิเคราะห์ข้อมูลทางพันธุกรรมจากการจัดเรียงลำดับ นิวคลีโอไทด์ เป็นประโยชน์ต่อการทำงานทางชีวโมเลกุลมาก ทางด้านการวิเคราะห์หายีน โครงสร้างของยีน เช่น ส่วนที่ทำหน้าที่ ควบคุมการทำงานของยีน (regulatory elements) ส่วนที่จะ ทำการแปลรหัสเพื่อให้ได้โปรตีน (coding sequence) ซึ่งสามารถ ประยุกต์ใช้ได้ในระบบการโคลนเพื่อการสร้างโปรตีนลูกผสม ใหม่ๆ ขึ้น หรือแม้กระทั่งการจัดจำแนกและชี้เฉพาะสปีชีส์ของ สิ่งมีชีวิตโดยเทียบความคล้ายคลึงจากลำดับเบส และสร้าง แผนภาพต้นไม้แสดงความสัมพันธ์ของสิ่งมีชีวิตขึ้นมา เป็นต้น

สำหรับในงานทางด้านโปรตีน มักเป็นการทำงานที่ เกี่ยวกับการเปรียบเทียบโครงสร้างสามมิติของโปรตีน วิธีการที่ง่ายที่สุดประกอบด้วย การเทียบบริเวณหนึ่งบริเวณ หรือมากกว่าของโปรตีนเดียวกันแต่มาจากแหล่งที่มาที่ต่างกัน สามารถให้ข้อมูลที่เป็ประโยชน์ได้เช่น

1. ความเหมือนหรือความต่างระหว่างการทำนาย โครงสร้างผลึก (crystal structure) ของสองลำดับกรดอะมิโน ที่อิสระต่อกัน
2. ความเหมือนหรือความต่างระหว่างภายในกลุ่มของ โครงสร้างหนึ่งๆ ซึ่งมักจะได้มาด้วยการวิเคราะห์ NMR spectroscopy

ตารางที่ 1 สารบัญตัวอย่างสำหรับฐานข้อมูลลำดับเบสแบบเจาะจง (ดัดแปลงจาก Barnes และ Gray, 2003)

Name	Description and web address
General databases	
EMBLCHANGE	Sequence change features from EMBL http://www.ebi.ac.uk/embl/
SWISSCHANGE	Sequence change features from SWISS-PROT http://www.expasy.ch/sprot/
OMIMALLELE	Alleles from OMIM http://www.ncbi.nlm.nih.gov/omim
Mitochondrial genome	
HUMAN__MITBASE	Human mitochondrial DNA variants http://www3.ebi.ac.uk/Research/Mitbase
Locus-specific mutation databases	
P53LINK	P53 mutations database http://www-p53.iarc.fr/
BTKBASE	Bruton's tyrosine kinase mutations http://bioinf.uta.fi/BTKbase/
VWF	Von Willebrand factor gene variations http://www.vwf.group.shef.ac.uk/
CFTR	Cystic fibrosis mutation database http://www.genet.sickkids.on.ca/app
PAH	Phenylalanine hydroxylase mutations http://www.pahdb.mcgill.ca/
Central databases	
HGVbase	Human Genome Variation database (SNPs and mutations) http://hgibase.cgb.ki.se/

3. ความเหมือนหรือความต่างระหว่างโครงสร้างที่ได้มาจากการใช้ข้อมูลทาง X-ray diffraction และ NMR spectroscopy

4. ความเหมือนหรือความต่างที่เกิดขึ้นระหว่าง อะโป- (apo-) หรือ ฮาโล- (halo-) โปรตีน

5. การเปลี่ยนแปลงในโครงสร้างโปรตีนที่เกิดขึ้นเมื่อมีการจับกับลิแกนด์ (ligand) โคแฟคเตอร์ (cofactor) หรือโลหะ

โปรแกรมเปรียบเทียบโครงสร้างโปรตีนคู่เหมือนส่วนใหญ่มักถูกออกแบบเพื่อให้มาเทียบโครงสร้างโปรตีนคู่เหมือน (homologous protein) ที่มาจากสิ่งมีชีวิตต่างชนิดกัน มีโปรแกรมจำนวนมากและมีหลายยุทธวิธีที่ใช้กันในปัจจุบัน ในบทความนี้จะขอยกตัวอย่างโดยสังเขป เพื่อประกอบความเข้าใจเท่านั้น ตัวอย่างที่ยกมานี้เป็นเพียงหลักการเบื้องต้นของการทำงานที่นักอณูชีววิทยาใช้ในงานวิจัย โดยส่วนใหญ่กระบวนการนี้มีสองแบบ (Higgins and Taylor, 2000)

1. สำหรับการประยุกต์เพื่อวัตถุประสงค์เชิงวิวัฒนาการวิวัฒนาการหรือที่เรียกกันว่าวิวัฒนาการชาติพันธุ์ (phylogeny) จำเป็นที่จะต้องสนใจจำนวนข้อมูลที่จะนำมาเทียบ วัตถุประสงค์ก็เพื่อที่จะเพิ่มความเป็นไปได้ระหว่างการจัดเรียงลำดับนิวคลีโอไทด์

หรือกรดอะมิโน (alignment of nucleotide or amino acid sequence) ที่ใกล้เคียงกันที่ใกล้เคียงความจริง และเพื่อลดความคลาดเคลื่อนทางสถิติให้มากที่สุดเช่นกัน

2. สำหรับงานที่ต้องการหาโปรตีนที่ใกล้เคียงกับข้อมูลลำดับกรดอะมิโนที่มีอยู่ นักวิทยาศาสตร์มักจะใช้โปรแกรมที่สามารถทำการเทียบลำดับโดยหาบริเวณที่เหมาะสมได้โดยอัตโนมัติ โดยโปรแกรมจะทำการเลือกบริเวณการเริ่มเทียบโดยอัตโนมัติได้ทันที เพื่อเปรียบเทียบคุณสมบัติหรือโครงสร้างของโปรตีนที่ต้องการศึกษากับโปรตีนที่มีอยู่ในฐานข้อมูล

การออกแบบยา

ในอดีต การคิดค้นยาเพื่อรักษาอาการเจ็บป่วยของมนุษย์เป็นไปตามการเรียนรู้และสังเกตของแพทย์ อาจเป็นบันทึกตำรายาต่างๆ เช่น สมุนไพรที่ได้จากธรรมชาติ เมื่อวิทยาการทางเคมีและเภสัชมีความก้าวหน้ามากขึ้น การออกแบบยาจึงอยู่บนพื้นฐานของโครงสร้างทางเคมีซึ่งออกฤทธิ์ด้านต่อสารหรือจุลชีพที่ก่อโรค ในปัจจุบัน เมื่อวิทยาการทางอณูชีววิทยาและชีวสารสนเทศมีความก้าวหน้ามากขึ้น การออกแบบยาจึงสามารถดำเนินการโดยอาศัยพื้นฐานความรู้ต่างๆ โดยในที่นี้จะขอกกล่าวถึง

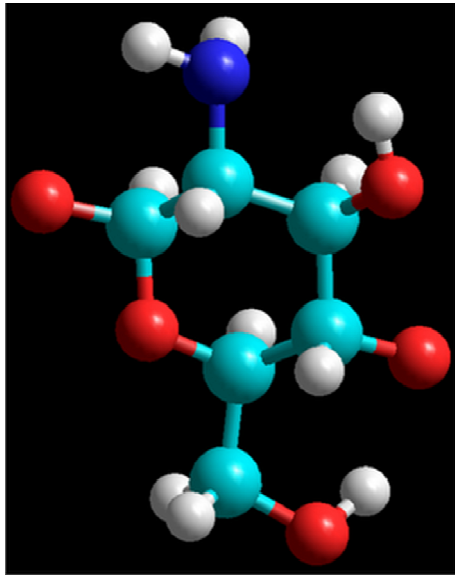
ตารางที่ 2 ตัวอย่างเว็บไซต์ที่ให้บริการข้อมูลชีวสารสนเทศ (Campbell and Heyer, 2003) เพื่อการสืบค้นข้อมูลการกลายของจีโนมและโปรตีน

Description of databases	Web address
Sequence variation databases	
OMIM	http://www.ncbi.nlm.nih.gov/Omim/
HGMD	http://www.hgmd.org
GDB Mutation Waystation	http://www.centralmutation.org
Central databases (SNPs and mutation) HGVbase	http://hgvdbase.cgb.ki.se
Sequence variation database (SRS)	http://srs.ebi.ac.uk
dbSNP	http://www.ncbi.nlm.nih.gov/SNP/
The SNP consortium (TSC)	http://snp.cshl.org/
Genetic marker maps (microsatellites, STSs other markers)	
Marchfield maps	http://research.marchfieldclinic.org/genetics/
Genome database (GDB)	http://www.gdb.org
dbSTS	http://www.ncbi.nlm.nih.gov/STS/
UniSTS	http://www.ncbi.nlm.nih.gov/genome/STS
Gene-oriented SNP and mutation visualization	
LocusLink	http://www.ncbi.nlm.nih.gov/LocusLink/
Protein Mutation Database	http://www.genome.ad.jp/htbin/
GoIPoly	http://61.139.84.5/gopoly/
SNPper	http://bio.chip.org:8080/bio/snpenter-enter
GeneSNPs	http://www.genome.utah.edu/genesnps/
Genome-oriented for SNP and mutation visualization	
Ensembl	http://www.ensembl.org
Human Genome Browser (UCSC-HGB)	http://www.genome.ucsc.edu/index.html

ขั้นตอนโดยคร่าวๆ 4 ขั้นตอน (จิรภรณ์ อังวิทยาธร, 2544) คือ การค้นพบและตรวจสอบจุดเป้าหมาย เป็นการตั้งสมมุติฐานถึงโมเลกุลเมื่อได้รับการยับยั้งหรือกระตุ้นจะส่งผลต่อการรักษาโรค ขั้นที่สองคือ การค้นพบตัวนำ (carrier) ที่เข้าจับกับโมเลกุลเป้าหมายได้และทำให้เหมาะสม ซึ่งเป็นกระบวนการคัดกรองหาสารเคมีที่มีฤทธิ์ต่อโมเลกุลเป้าหมาย กระบวนการพัฒนาก่อนคลินิกและกระบวนการพัฒนาในคลินิก ให้มีโครงสร้างของยาที่เหมาะสมกับการเข้าจับกับโมเลกุลต่างๆ ไม่ว่าจะเป็นโปรตีนที่ก่อโรค เช่น ตัวรับ (receptor) ที่เฉพาะเจาะจงบนผิวเซลล์ของไวรัส ในปัจจุบัน

การออกแบบยาอาศัยโปรแกรมทางชีวสารสนเทศในส่วนของ การทำแบบจำลองระหว่างลิแกนด์และโมเลกุลของยาหรือโปรตีนที่ก่อโรคในรูปแบบสามมิติ (3-Dimensional structure) ซึ่งมีบทบาทมากในส่วนของกระบวนการออกแบบโมเลกุลยาให้เหมาะสมกับเป้าหมายหรือโปรตีนที่ต้องการยับยั้ง ขั้นตอนการศึกษาเหล่านี้ สามารถใช้โปรแกรมที่มีอยู่ในเครื่องโดยใช้ฐานข้อมูลโครงสร้างสามมิติที่มีอยู่โดยไม่จำเป็นต้องมีการเชื่อมต่อกับอินเทอร์เน็ตตลอดเวลาก็ได้ เช่น กระบวนการออกแบบยารักษาโรคเอดส์ของ Merck Research Laboratories

และยาบรรเทาหวัดที่ผลิตโดย Sterling Winthrop (ธีร์เกียรติ เกิดเจริญ, 2539) โดยใช้ซอฟต์แวร์ที่มีความสามารถในการจำลอง ด้วยคอมพิวเตอร์ (Computer Simulations) หรือใช้ซอฟต์แวร์ สำหรับโมเดลกลศาสตร์เชิงโมเลกุลที่ชื่อว่า HYPERCHEM เป็นต้น (ภาพที่ 1)



ภาพที่ 1 โครงสร้างสามมิติของโคโคซานโดยใช้ซอฟต์แวร์ สำหรับโมเดลกลศาสตร์เชิงโมเลกุลที่ชื่อว่า HYPER-CHEM เพื่อช่วยในการจำลองรูปโมเลกุลสามมิติ (ดัดแปลงจาก Lai และคณะ 2008)

วิศวกรรมเมแทบอลิก

เป็นการประยุกต์กรรมวิธีในการวิเคราะห์และสังเคราะห์ ข้อมูลทางโปรตีนและวิถีเมแทบอลิซึมเพื่อปรับปรุงทิศทางในการ ผลิตภายในวิถีเมแทบอลิซึม ศาสตร์นี้กล่าวถึงกระบวนการทาง ชีวภาพ (bioprocess) ที่มีความซับซ้อนเนื่องจากหลายกรณี เช่น ปฏิกิริยาหนึ่งอาจมาจากการเร่งด้วยเอนไซม์หลายขั้นตอน หรือแม้แต่เอนไซม์เดียวอาจเร่งหลายปฏิกิริยาที่มีสารตั้งต้น เช่นเดียวกันได้เช่นกัน การศึกษาจึงอาจทำได้โดยการสร้างแบบ จำลองขึ้น อาศัยข้อมูลทางคณิตศาสตร์ควบคู่ไปกับการวิเคราะห์ ผลผลิตทั้งภายนอกและภายในเซลล์ รวมทั้งสมดุลรีดอกซ์ซึ่งจะ กระทำในระหว่างก่อนและหลังการดัดแปลงเซลล์ด้วยกระบวนการ ทางพันธุวิศวกรรมเพื่อควบคุมฟลักซ์ (flux) หรือวิถีเมแทบอลิซึม ในสิ่งมีชีวิต โดยหลักการ การผนวกพื้นฐานความรู้หรือข้อมูลทาง สารสนเทศและวิศวกรรมเมแทบอลิกควรจะพัฒนาควบคู่กันไป เพื่อปรับปรุงกระบวนการผลิตให้สอดคล้องกันมากที่สุด

เป้าหมายโดยทั่วไปของการทำวิศวกรรมเมแทบอลิกคือ ทำอย่างไรเพื่อกระตุ้นให้เซลล์มีการเจริญและสร้างผลผลิต เป้าหมาย ให้เป็นไปตามต้องการและเพื่อให้บรรลุวัตถุประสงค์ ในที่นี้จะกล่าวถึงการนำจุลชีพมาประยุกต์ใช้ในงานวิศวกรรมทาง เมแทบอลิก ซึ่งเป็นที่นิยมมากเนื่องจากหน่วยพันธุกรรมมีขนาดเล็ก สามารถดัดแปลงได้ง่าย สามารถเติบโตได้เร็ว สามารถเพิ่มจำนวน ได้ทวีคูณในระยะเวลาอันสั้น ทำให้สร้างโปรตีนหรือผลผลิตของ เซลล์ได้มากเมื่อควบคุมการเลี้ยงในภาวะปัจจัยที่มีความเหมาะสม แบบจำลองทางคณิตศาสตร์ที่อธิบายหรือเพื่อใช้เป็นตัวแทน พฤติกรรมของจุลชีพในกระบวนการจึงมีความจำเป็นอย่างยิ่ง เนื่องจากความซับซ้อนของกระบวนการภายในเซลล์และ ความสัมพันธ์แบบไม่เป็นเส้นตรง รวมถึงกิจกรรมภายในเซลล์ ดังนั้นการออกแบบจำลองที่ดีจึงเป็นพื้นฐานที่สำคัญเป็นอย่างยิ่ง สำหรับการตัดสินใจดัดแปลงพันธุกรรมของสิ่งมีชีวิตหรือจุลชีพ ในขั้นตอนถัดไป เหล่านี้ล้วนเป็นเหตุผลในการสร้างแบบจำลอง วิเคราะห์วิถีเมแทบอลิซึมในสิ่งมีชีวิต ซึ่งการออกแบบแบบจำลอง ดังกล่าวก็อาศัยโปรแกรมทางชีวสารสนเทศเป็นตัวช่วยในการ ประมวลผลนั่นเอง

เนื่องจากข้อมูลทางเมแทบอลอโลยีมีความเกี่ยวข้องโดยตรง กับกิจกรรมในเซลล์และมีความสัมพันธ์เชื่อมโยงต่อพันธุกรรม กิจกรรมของเซลล์และพีโนไทป์ของจุลชีพในกระบวนการชีวภาพ การวิเคราะห์จุลชีพโดยอาศัยพื้นฐานทางเครือข่ายเมแทบอลิซึม (metabolism network) ซึ่งทั้งหมดนี้ช่วยสนับสนุนให้กระบวนการ ดัดแปลงพันธุกรรมและกระบวนการผลิตมีประสิทธิภาพ วิศวกรรม เมแทบอลิซึมจึงเป็นงานใหม่สำหรับวิศวกรทางด้านกระบวนการ ชีวภาพ ปัจจุบัน มีโปรแกรมและเว็บไซต์ที่ให้บริการข้อมูลเป็น จำนวนมากเพื่อเก็บรักษาข้อมูลทาง mass spectroscopy ของโปรตีนชนิดต่างๆ รวมทั้งโปรแกรมทางการค้า เช่น PEAKS analysis สำหรับทำการเปรียบเทียบข้อมูลในรูปแบบของ mass spectroscopy ที่ได้จากโปรตีนที่ต้องการตรวจสอบเทียบกับ รูปแบบที่มีอยู่ในฐานข้อมูล ประกอบกับข้อมูลที่ได้จากการวิเคราะห์ อื่นเช่น แอคทิวิตีของเอนไซม์จากฐานข้อมูลเอนไซม์ BRENDA (TU Braunschweig Department of Bioinformatics, 2010) เหล่านี้ล้วนมีการจำลองแบบเครือข่ายวิถีเมแทบอลิซึมขึ้นมาเพื่อ ทำการวิเคราะห์ฟลักซ์ของวิถีเมแทบอลิซึมคู่ไปกับผลการทดลอง ในทางปฏิบัติเพื่อให้ได้แบบจำลองที่ถูกต้องมากที่สุด กระบวนการ ทางวิศวกรรมเมแทบอลิกมีข้อประยุกต์ใช้ที่น่าสนใจโดยเฉพาะ การผลิตเคมีภัณฑ์ที่มีมูลค่าสูง โดยเฉพาะทางอุตสาหกรรมอาหาร และพลังงาน เหล่านี้เป็นการผนวกศาสตร์ทางด้านพันธุวิศวกรรม

ตารางที่ 3 ประโยชน์ของกระบวนการทางวิศวกรรมเมแทบอลิกในโปรแกรมประเภท MMT, Gepasi และ E-Cell (ดัดแปลงจาก Shimizu, 2002)

กระบวนการ	ประโยชน์
(1) Metabolic flux analysis (MFA)	เพื่อการเข้าใจสรีรวิทยาของเซลล์
(2) Cell capability analysis	เพื่อวิเคราะห์ประสิทธิภาพของเซลล์
(3) On-line metabolic flux analysis	ใช้เพื่อการพัฒนากระบวนการปฏิบัติการกระบวนการชีวภาพ
(4) Metabolic control analysis (MCA)	การออกแบบระบบของการทำโมเลกุลผสม
(5) ¹³ C-NMR	เพื่อการชี้จุดที่เกิดผลผลิตของกระบวนการ
(6) Kinetic analysis	เพื่อวัดการเปลี่ยนแปลงของวิถีเมแทบอลิซึม
(7) Integration of bioinformatic data	เข้าใจเซลล์อย่างเป็นระบบและการตัดแปลงวงจรหรือวิถีเมแทบอลิซึม

(genetic engineering) สรีรวิทยา (physiology) และวิศวกรรมระบบ (system engineering) เข้าด้วยกัน วัตถุประสงค์เพื่อการออกแบบและสร้างสิ่งมีชีวิตที่สามารถสร้างผลิตภัณฑ์ที่ต้องการได้โดยกระบวนการทางพันธุศาสตร์เพื่อความคุ้มค่าและเป้าหมายของเอนไซม์หรือโปรตีนในเซลล์ กิจกรรมเหล่านี้ ทำให้เกิดคำว่าชีววิทยาสังเคราะห์ (synthetic biology) ขึ้นมา (Liao and Higashide, 2008) อย่างไรก็ตามวิศวกรรมเมแทบอลิซึมให้ความสำคัญกับตัวผลิตภัณฑ์ที่เกิดขึ้นจากการผลิตโดยจุลินทรีย์มากกว่าพฤติกรรมของเซลล์ซึ่งเน้นการศึกษากิจกรรมต่างๆ ภายในเซลล์ซึ่งเป็นศาสตร์ทางชีววิทยาสังเคราะห์มากกว่า การทำวิศวกรรมเมแทบอลิซึมก่อให้เกิดงานวิจัยเพื่อการประยุกต์ใช้หลากหลายเนื่องจากแม้จะมีการศึกษาเครือข่ายเมแทบอลิซึมในสิ่งมีชีวิตเป็นจำนวนมาก แต่ในรายละเอียดของกระบวนการเหล่านี้ยังไม่เป็นที่ทราบแน่ชัดทั้งหมด การศึกษาทางด้านเมแทบอลิซึมของเซลล์จึงยังเป็นการท้าทายความสามารถของนักวิทยาศาสตร์และวิศวกรต่อไป บทบาทของกระบวนการทางชีวสารสนเทศที่นำมาประยุกต์ใช้มีหลายรูปแบบดังแสดงในตารางที่ 3

บทสรุป

ความก้าวหน้าของการวิจัยทางอณูชีววิทยาและอณูพันธุศาสตร์ก่อให้เกิดข้อมูลทางชีววิทยาจำนวนมากชีวสารสนเทศเป็นศาสตร์ประยุกต์ที่นำระบบการจัดการข้อมูลและเครือข่ายของคอมพิวเตอร์มาจัดการเพื่อการจัดเก็บวิเคราะห์และประเมินผล ทำให้การวิจัยในเรื่องที่ซับซ้อนและหาคำตอบได้ยาก เช่น กลไกการควบคุมเมแทบอลิซึมในเซลล์สามารถทำได้อย่างมีประสิทธิภาพมากขึ้น โดยอาศัยข้อมูลที่

จัดเก็บในฐานข้อมูลสากล ร่วมกับการใช้โปรแกรมประยุกต์ในการวิเคราะห์แบบจำลองที่สร้างขึ้นในทางทฤษฎี ว่ามีความสอดคล้องกับผลการทดลองที่เกิดจากการปฏิบัติจริงมากน้อยเพียงใด รวมถึงทำให้สามารถค้นหาคำตอบที่ทำได้ยาก เช่นกลไกในการตอบสนองต่อสิ่งแวดล้อมของเซลล์ หรือแม้แต่การออกแบบยารักษาโรคซึ่งในอดีตอาจต้องใช้เวลาอันนับเป็นสิบปีในการคิดค้น ในปัจจุบันอาจใช้เวลาเพียงไม่กี่ปีในการพัฒนาเนื่องจากสามารถค้นหาข้อมูลของความผิดปกติของยีนหรือแม้แต่โครงสร้างโปรตีนในจุลินทรีย์ที่เป็นสาเหตุให้เกิดโรคและทำการทดสอบโดยอาศัยสมมุติฐานที่สร้างขึ้นจากแบบจำลอง (*in silico model*) ที่สร้างขึ้นโดยโปรแกรมคอมพิวเตอร์ประยุกต์ ซึ่งทำให้ประหยัดเวลา ค่าใช้จ่าย และทำให้นักวิจัยทำงานได้โดยมีข้อมูลสนับสนุนการทำงานและสมมุติฐานได้อย่างตรงประเด็นมากขึ้น ชีวสารสนเทศจึงเป็นศาสตร์ใหม่แขนงหนึ่งที่มีบทบาทและความสำคัญต่อวงการวิทยาศาสตร์ การแพทย์และวิศวกรรมศาสตร์ รวมถึงการวิจัยพัฒนาในเชิงอุตสาหกรรมและพาณิชย์

เอกสารอ้างอิง

- ธีรเกียรติ เกิดเจริญ. (2539). คอมพิวเตอร์ช่วยการออกแบบโมเลกุลกับการพัฒนาเทคโนโลยี. *วารสารเพื่อการพัฒนาเทคโนโลยี วัสดุ*, 4, 17-22.
- จิรภรณ์ อังวิทย์วร. (2544). *Molecular modeling in drug design*. ใน *ชีวสารสนเทศศาสตร์*. กรุงเทพฯ : สถาบันบัณฑิตวิทยาศาสตร์และเทคโนโลยีไทย สำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ.

เจษฎา เต็นตวงบริพันธ์. (2545). การสร้างวงษ์วานวิวัฒนาการ
ขึ้นมาใหม่ขั้นพื้นฐาน. ในหนังสือการประชุมวิชาการเรื่อง
Molecular techniques for biodiversity research.
กรุงเทพฯ : ภาควิชาชีววิทยา คณะวิทยาศาสตร์
จุฬาลงกรณ์มหาวิทยาลัย.

Barnes, M.R., & Gray, I.C. (2003). *Bioinformatics for
geneticists.* Chichester : John Wiley & Sons.

Campbell, A.M., & Heyer, L.J. (2003). *Discovering
genomics, proteomics, and bioinformatics.* San
Francisco : Cold Spring Harbor Laboratory Press
and Benjamin Cummings.

Debons, A., King, D.W., Mansfield, U., & Shirey, D.L.
(1981). *The information professional: Survey of
an emerging field.* New York : Marcel Dekker, Inc.

Higgins, D., & Taylor, W. (2000). *Bioinformatics:
Sequence, structure, and databanks.* Newyork :
Oxford University Press.

Lai, S., Shirlin, C., Kean, W., Gopal, V., & Goh, K.L.
(2010). Polyhedral oligomeric silsesquioxanes
(POSS)-Chitosan composite fibre for biomedical
applications. Available from: [http://sites.google.com/
site/fibrecomposites/research-fibre-composites-
stress-transfer-mechanisms-extracellular-matrices-
structure-property-carbon-nanotubes-hydroxyapa-
tite-chitosan/polyhedral-oligomeric-silsesquiox-
anes-poss--chitosan-composite-fibre-for-biomed-
ical-applications-nanoparticulate-hydroxyapatite](http://sites.google.com/site/fibrecomposites/research-fibre-composites-stress-transfer-mechanisms-extracellular-matrices-structure-property-carbon-nanotubes-hydroxyapatite-chitosan/polyhedral-oligomeric-silsesquioxanes-poss--chitosan-composite-fibre-for-biomedical-applications-nanoparticulate-hydroxyapatite)
[10 June 2010].

Liao, J.C., & Higashide, W. (2008). Metabolic
engineering of next-generation biofuels. *Journal
of Chemical Engineering Progress*, 104(8), S19–S23.

Shimizu, H. (2002). Metabolic engineering - Integrating
methodologies of molecular breeding and biopro-
cess systems engineering. *Journal of Biosciences
and Bioengineering*. 94(6), 563-573.

TU Braunschweig Department of Bioinformatics,
Germany. (2010). Available from: [http://www.
brenda-enzymes.org](http://www.brenda-enzymes.org) [16 February 2010].