บทที่ 2

ทษฤษฎีที่เกี่ยวข้อง

2.1 ความเป็นมาเกี่ยวกับสายพันธุกรรม

ในปี 1865 บาทหลวง Gregor Mendel ทำการทดลองผสมพันธุ์ต้นถั่ว ซึ่งพบว่าต้นถั่วพันธุ์แท้จะมียีนชนิดเดียวกันอยู่เป็นคู่ ในขณะพันธุ์ทางจะมียีนเป็นคู่แต่มีลักษณะต่างกันเล็กน้อย และยีนเด่นจะแสดงออกให้เห็นได้ด้วยตา ส่วนยีนด้อยจะแสดงออกได้ต่อเมื่อเป็นพันธุ์แท้เท่านั้น ดังนั้น Mendel ได้ตั้งสมมติฐานว่าในเซลล์สืบพันธุ์จะใช้ยีนเพียงชุดเดียว และเมื่อปฏิสนธิจะรวมกันเป็นคู่อีกครั้ง

หลังจาก Mendel เสียชีวิต วงการวิทยาศาสตร์ได้มีการใช้กล้องจุลทรรศน์ส่องเกตดูเซลล์ พบว่าเซลล์ของสิ่งมีชีวิตแต่ละชนิดมีจำนวนโครโมโซม (Chromosome) คงที่ ทำให้เกิดข้อสันนิษฐานว่าโครโมโซมเป็นที่อยู่ของยีน ต่อมาไม่นานได้มีการค้นพบว่าทุกครั้งในการแบ่งตัว (mitosis) โครโมโซมจะถูกคัดลอกและถ่ายทอดไปอยู่ในเซลล์ใหม่เสมอ และค้นพบว่าโครโมโซมทั้งหมดมีคู่หนึ่งที่ใช้กำหนดเพศ คือโครโมโซม X และโครโมโซม Y ซึ่งคู่ XX จะเป็นเพศหญิง และคู่ XY จะเป็นเพศชาย

การทดลองของ Thomas Morgan ในช่วงต้นคริสต์ศตวรรษ ใช้แมลงหวี่ในการทดลองโดยพยายามหาความสัมพันธ์ของยีนและโครโมโซม พบว่าลักษณะของแมลงหวี่ที่เกิดจากการผสมพันธุ์เป็นไปตามกฎของ Mendel และจากการสังเกตพบว่าสีตาของแมลงหวี่กับลักษณะทางเพศมักอยู่ด้วยกัน ดังนั้นจึงคิดได้ว่ายีนที่ระบุสีตาของแมลงหวี่จะอยู่บนโครโมโซมเพศด้วย และยีนลักษณะอื่นๆน่าจะอยู่บนโครโมโซมอื่นด้วย

ในช่วงต้นทศวรรษที่ 1920 คณะวิจัยของ Oswald Avery ทดลองพบว่าจุลินทรีย์ไม่มีพิษเมื่อนำไปรวมกับจุลินทรีย์มีพิษที่ตายแล้วจะกลายเป็นจุลินทรีย์มีพิษ จึงคิดว่ามีการถ่ายทอดยีนหรือสายพันธุกรรมจากจุลินทรีย์มีพิษไปยังจุลินทรีย์ไม่มีพิษ และในช่วงทศวรรษที่ 1940 ยังพบว่าสายพันธุกรรมไม่ถูกย่อยสลายด้วยเอนไซม์ที่ใช้ย่อยโปรตีน แต่ถูกย่อยด้วยเอมไซม์ที่ใช้ย่อยดีเอ็นเอ (DNA) จากการทดลองจึงยืนยันได้ว่ายีน หรือสายพันธุกรรม คือดีเอ็นเอ

ช่วงทศววรษที่ 1950 พบว่าสัดส่วนของเบสที่เป็นส่วนประกอบของดีเอ็นเอของสิ่งมีชีวิตมีปริมาณ A ต่อ T และ G ต่อ C คงที่ และเริ่มมีการศึกษาโครงสร้างของดีเอ็นเอด้วยรังสีเอ็ก (x-ray)

ในปี 1953 James Watson และ Francis Crick ค้นพบโครงสร้างดีเอ็นเอที่มีลักษณะเหมือนบันไดบิดเป็นเกลียว โดยส่วนประกอบเป็นราวบันไดคือน้ำตาลดีออกซีไรโบสซ์ (Deoxylibose) กับฟอสเฟต และขั้นบันได้คือเบสที่จับคู่กันระหว่าง A กับ T และ G กับ C เสมอ หลังจากนั้น Watson และ Crick สร้างสมมติฐานว่าซีกหนึ่งของราวบันไดดีเอ็นเอทำหน้าที่เป็นต้นแบบให้อีกซีกหนึ่ง นอกจากดีเอ็นเอที่พบในนิวเคลียสของเซลล์ ยังพบอาร์เอ็นเอ (Ribonucieic acid) ที่พบในไซโตพลาสซึม (Cytoplasm) ซึ่ง Watson และ Crick เสนอแนวคิดว่า รหัสพันธุกรรมจากดีเอ็นเอจะถูกคัดลองไปยังอาร์เอ็นเอ เพื่อพาออกจากนิวเคลียสสู่ไซโตพลาสซึมที่มีการสังเคราะห์โปรตีน และน่าจะต้องมีโมเลกุลที่เป็นตัวนำในการพากรดอะมิโนมาประกอบเป็นโปรตีนตามรหัสในอาร์เอ็นเอ

จากการวิจัยของ Watson และ Crick ทำให้การวิจัยทางด้านชีววิทยาระดับโมเลกุลมีความก้าวหน้าอย่างมาก จึงทำให้เกิดการศึกษารหัสพันธุกรรมทั้งหมด หรือ จีโนม (Genome) ในสิ่งมีชีวิตทั้งหมด โดยในช่วงทศวรษที่ 1990 เกิดโครงการวิจัยจีโนมขึ้น และโครงการที่ใหญ่ที่สุดคือโครงการจีโนมมนุษย์ [27]

2.2 **ชีวสารสนเทศศาสตร์** (Bioinformatics)

ชีวสารสนเทศน์ศาสตร์เป็นศาสตร์ที่เกี่ยวกับการจัดเก็บข้อมูลทางชีววิทยาให้เป็นระบบ โดยอาศัยเทคโนโลยีสารสนเทศในการวิเคราะห์ข้อมูล และหาคำตอบ จากการใช้เทคโนโลยีเข้ามาช่วยคำนวณทำให้นักวิทยาศาสร์สามารถสืบค้น และวิเคราะห์ข้อมูลได้อย่างรวดเร็ว

ยุคของชีวสารสนเทศศาสตร์เกิดขึ้นพร้อมโครงการจีโนมมนุษย์เพื่อศึกษาโมเลกุลและกระบวนการต่างๆภายในเซลล์ นักวิจัยมีจุดประสงค์ในการใช้เทคโนโลยีในการจัดเก็บและวิเคราะห์ผลข้อมูลจีโนไทป์ (Genotype) เพื่อทำนายลักษณะฟีโนไทป์ (Phenotype) เพื่อลดการทำงานในห้องปฏิบัติการ ซึ่งฟีโนไทป์มีการแสดงออกหลายระดับตั้งแต่ระดับที่สามารถมองเห็นได้ไปจนถึงระดับที่ไม่สามารถมองเห็นได้ เช่น การแสดงออกของยีนในระดับโครงสร้าง (Structural genomics) เป็นต้น

เมื่อข้อมูลทางชีวภาพสามารถรวบรวมได้ทั้งหมด นักวิจัยจะสามารถดึงข้อมูลเพื่อนำไปวินิจฉัย และป้องกันการเสี่ยงในการเกิดโรคที่เกิดจากยีนได้

ความเป็นมาของชีวสารสนเทศศาสตร์แบ่งได้เป็น 3 ช่วงใหญ่ๆคือ

2.2.1 ช่วงเริ่มต้นของชีวสารสนเทศศาตร์

จากการทิ้งระเบิดปรมาณูในญี่ปุ่นทำให้มนุษย์เกิดการเปลี่ยนแปลงในระดับสายพันธุกรรม ประเทศอเมริกาจึงเสนอโครงการวิจัยจีโนมมนุษย์ (Human genome project) ซึ่งโครงการนี้ทำการถอดรหัสสารพันธุกรรมดีเอ็นเอของมนุษย์เป็นหลัก โดยมีเหตุผลว่ามนุษย์ต้องมีความรู้พื้นฐานเกี่ยวกับรหัสทางพันธุกรรมของยีนทั้งหมดก่อนจึงสามารถเข้าใจปฏิสัมพันธ์ระหว่างคนกับสิ่งต่างๆได้

โครงการจีโนมมนุษย์ทำให้เกิดชีวสารสนเทศศาสตร์ขึ้นมาเพื่อจัดการข้อมูลจีโนมที่มีเพิ่มมากขึ้น เช่น ลำดับเบสของสารพันธุกรรม หรือลำดับกรดอะมิโน โดยเปลี่ยนข้อมูลให้อยู่ในรูปข้อมูลดิจิตอล และเก็บในฐานข้อมูลอย่างเป็นระเบียบ

2.2.2 ช่วงก่อนโครงการจีโนมมนุษย์เสร็จ

ช่วงก่อนโครงการจีโนมมนุษย์เสร็จ จะให้ความสำคัญในการถอดรหัสพันธุกรรมของมนุษย์และสิ่งมีชีวิตต้นแบบที่สามารถใช้ทดลองแทนมนุษย์ได้ เช่น หนู เป็นต้น โดยเน้นเรื่องการจัดทำฐานข้อมูลลำดับเบส และข้อมูลจีโนมอย่างเป็นระเบียบ

2.2.3 ช่วงหลังโครงการจีโนมมนุษย์

ช่วงหลังโครงการจีโนมมนุษย์ จะเน้นการแปลความหมายของรหัสพันธุกรรม ว่าเกี่ยวข้องกับมนุษย์อย่างไรบ้าง โดยมีการศึกษา 5 ส่วนหลักๆดังนี้

1 การแสดงออกของเซลล์ (Function genomics)

ศึกษาการแสดงออกของยีนโดยการตรวจชนิดและปริมาณเอ็มอาร์เอ็นเอ (mRNA) ที่ถูกสร้างขึ้นในเซลล์ในแต่ละช่วงเวลา หรือถูกสร้างขึ้นเพื่อตอบสนองต่อสิ่งเร้าภายนอก วิธีที่นิยมใช้คือการใช้ cDNA Microarray การแสดงออกของเซลล์จะช่วยให้เข้าใจการทำงานของยีนเนื่องจากเซลล์แต่ละชนิดจะมีการสร้างเอ็มอาร์เอ็นเอที่ต่างกัน เช่นเซลล์กระดูกจะมีจำนวนยีน จำนวนเบส และการเรียงตัวที่ไม่เหมือนเซลล์บริเวณอื่นเป็นต้น

2 โครงสร้างและหน้าที่ของโปรตีน (Structural genomics)

เป็นการใช้ข้อมูลในการทำนายโครงสร้างโปรตีน 3 มิติ โดยใช้ข้อมูลจากลำดับเบสของดีเอ็นเอ หรือลำดับของกรดอะมิโน เนื่องจากลำดับเบสของดีเอ็นเอเป็นตัวกำหนดโครงสร้างโปรตีน

ปัจจุบันการทำนายโครงสร้างยังมีความถูกต้องไม่มากนัก ยังจำเป็นต้องอาศัยการเปรียบเทียบกับโครงสร้างโปรตีนที่ถูกต้อง เนื่องจากยังไม่เข้าใจการพับงอของโปรตีน (Protein folding) ดีนัก

หากสามารถทำนายโครงสร้างโปรตีนได้ จะช่วยให้สร้างยา หรือสารกระตุ้นระบบภูมิคุ้มกันในสิ่งมีชีวิตได้ โดยการจำลองปฏิกิริยาการจับกันระหว่างยากับโปรตีน (Molecular docking) เป็นต้น

3 วิจัยประเภทและปริมาณของโปรตีน (Proteomics)

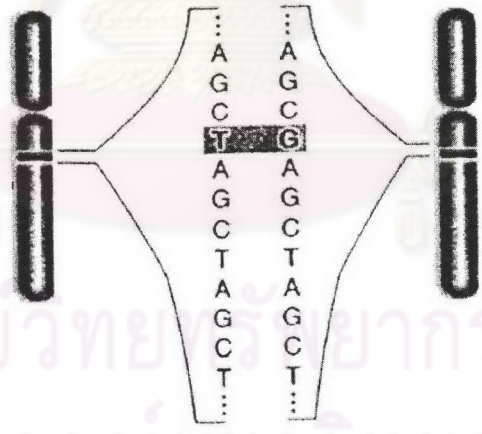
ทำการศึกษาชนิด คุณสมบัติ และหน้าที่ของโปรตีนที่ถูกสร้างขึ้น การศึกษามักเริ่มต้นด้วยการทำ 2DPAGE จากนั้นจะนำไปแยกด้วยเอ็นไซม์เพื่อนำไปวิเคราะห์

4 กระบวนการสร้างและสลายของเซลล์ (Metabolomics)

เป็นการศึกษากระบวนการสร้างและสลายของเซลล์ทั้งเซลล์ ปัจจุบันพยายามนำข้อมูลที่เกี่ยวกับกระบวนการมาสร้างการจำลอง (Simulation) โดยอาศัยเทคโนโลยีชีวสารสนเทศบนเครื่องคอมพิวเตอร์ประสิทธิภาพสูง

5 การบำบัดด้วยยา (Pharmacogenomics)

ทำการศึกษายาที่ตอบสนองจำเพาะต่อบุคคล เพื่อเพิ่มประสิทธิภาพ และลดผลข้างเคียงต่อผู้ป่วย การบำบัดด้วยยานำความรู้ทางเภสัชวิทยา ชีวเคมี ยีน จีโนม และ SNP (Single nucleotide polymorphisms) มาใช้ร่วมกัน



รูปที่2.1 ตัวอย่างแสดง SNP [1]

ในโครงการจีโนมค้นพบว่าลำดับเบสร้อยละ 99.9 ในมนุษย์มีความเหมือนกัน และที่ต่างกันร้อยละ 0.1 เป็นส่วนที่ต่างกันในแต่ละบุคคล หรือที่เรียกว่า SNP นักวิจัยมีความเชื่อว่าปัจจุบันผู้ป่วยอาจกินยาที่ไม่เกิดประโยชน์ให้กับร่างกาย ดังนั้นจึงนำความรู้เรื่อง SNP มาใช้เพื่อสร้างยาให้เกิดผลประโยชน์สูงสุดในแต่ละบุคคล

---

[1] http://www.1ban.co.jp/tachibana/mini/bio/images/23.gif

2.3 ดีเอ็นเอ อาร์เอ็นเอ และโปรตีน

เนื่องจากองค์ประกอบ และโครงสร้างของดีเอ็นเอ อาร์เอ็นเอ และโปรตีนมีความแตกต่างกัน ในหัวข้อนี้จึงแจกแจงลักษณะต่างๆ โดยแบ่งเป็นส่วนของดีเอ็นเอ ส่วนของอาร์เอ็นเอ และส่วนของโปรตีนดังนี้

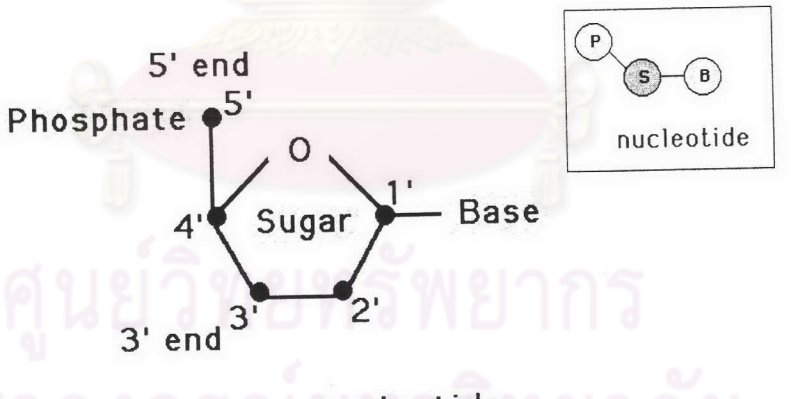
2.3.1 องค์ประกอบและโครงสร้างของดีเอ็นเอ

2.3.1.1 องค์ประกอบทางเคมีของดีเอ็นเอ

กรดดีออกซิโรโบนิวคลีอิก (Deoxyribonucleic acid) หรือดีเอ็นเอเป็นโพลีเมอร์ของนิวคลีโอไทด์ โดยแต่ละหน่วยย่อยนิวคลีโอไทด์ประกอบด้วย น้ำตาลดีออกซีไรโบส เบส และ หมู่ฟอสเฟต ความแตกต่างของดีเอ็นเอเกิดจากการเรียงตัวของเบสบนสายดีเอ็นเอที่มีความแตกต่างกัน เบสที่พบในดีเอ็นเอมี 4 ชนิดโดยแบ่งเป็น 2 กลุ่มตามความแตกต่างทางองค์ประกอบ และโครงสร้างคือ

1 เบสพิริมิดีน โครงสร้างหลักประกอบด้วยวงแหวน 1 วง เบสในกลุ่มนี้ได้แก่ ไซโทซีน (Cytosine: C) และ ไธมีน (Thymine: T)

2 เบสพิวรีน โครงสร้างหลักประกอบด้วยวงแหวน 2 วง เบสกลุ่มนี้ได้แก่ อะดินีน (Adenine: A) และ กัวนีน (Guanine: G)

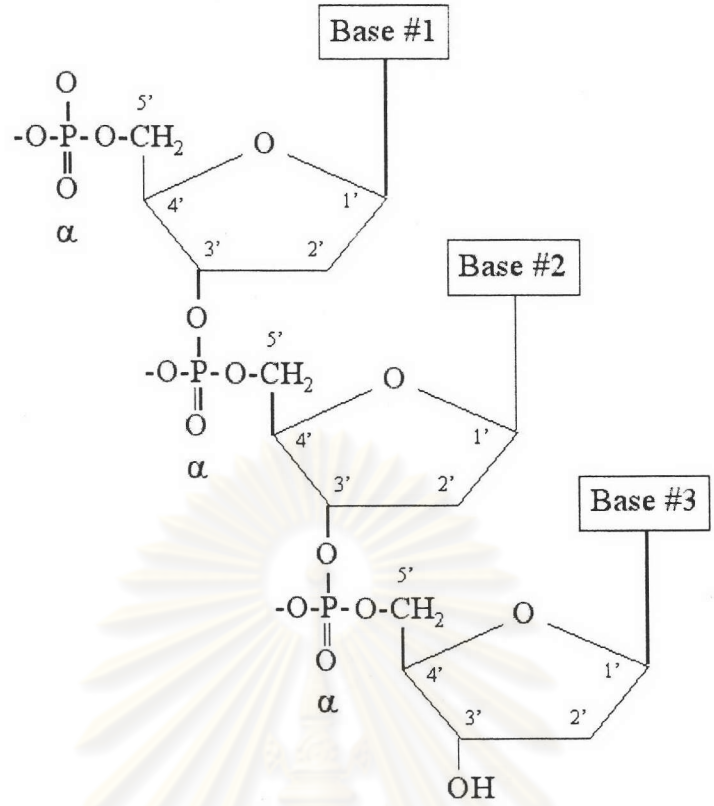


รูปที่2.2 องค์ประกอบของดีเอ็นเอ [2]

เบสในดีเอ็นเอจะต่ออยู่กับน้ำตาลดีออกซีไรโบสซึ่งมีคาร์บอน 5 อะตอม โดยมีหมู่ฟอสเฟตเป็นตัวเชื่อมระหว่างคาร์บอนตำแหน่งที่ 3 ของน้ำตาลโมเลกุลแรกกับคาร์บอนตำแน่งที่ 5 ของน้ำตาลโมเลกุลที่สอง ทำให้สายพอลีนิวคลีโอไทด์มีทิศทางที่แน่นอน คือจากปลาย 5 ไปปลาย 3

---

[2] http://www.irn.pdx.edu/~newmanl/nucleotide.GIF



รูปที่2.3 สายนิวคลีโอไทด์ที่เชื่อมด้วยพันธะฟอสโฟไดเอสเตอร์จากปลาย 5 ไปปลาย 3 [3]

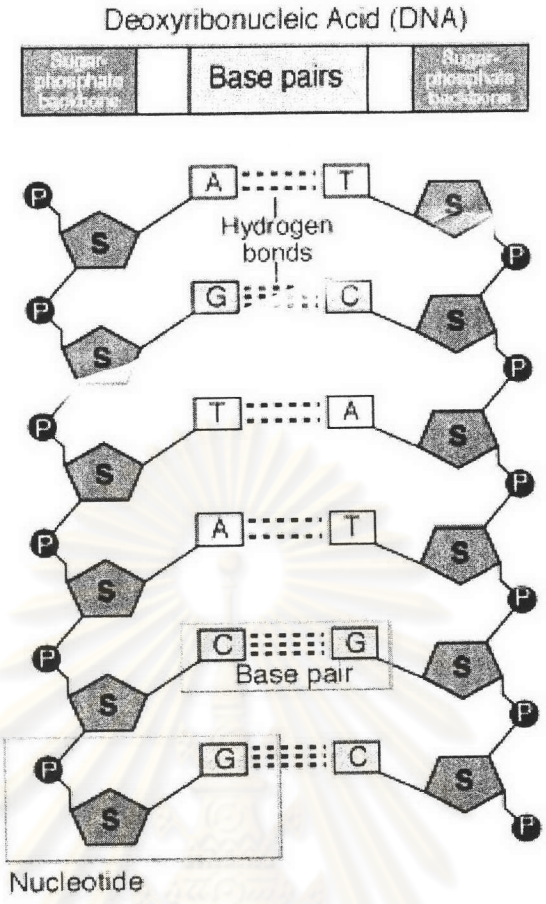
หมู่ฟอสเฟตทำให้ดีเอ็นเอมีประจุลบและมีคุณสมบัติเป็นกรด ความยาวของสายดีเอ็นเอเกิดจากการเรียงตัวของเบส 4 ชนิด โดยการเรียงตัวที่ต่างกันในโมเลกุลดีเอ็นเอจะกำหนดคุณสมบัติของสิ่งมีชีวิต

2.3.1.2 **โครงสร้างของดีเอ็นเอ**

Watson และ Crick เสนอโครงสร้างสามมิติของดีเอ็นเอ โดยอาศัยข้อมูลจากภาพเอ็กซเรย์ดิฟแฟรกชั่น (X-ray diffraction pattern) ของผลึกดีเอ็นเอ และจากข้อมูลองค์ประกอบทางเคมีของดีเอ็นเอ โดยเสนอว่าดีเอ็นเอมีโครงสร้างเป็นเกลียวคล้ายขดลวดสปริง เรียกว่าฮีลิกซ์ (Helix) และเกลียวของดีเอ็นเอเป็นเกลียวคู่เวียนขวา เกลียวคู่คือสายพอลีนิวคลีโอไทด์สองสายโดยมีน้ำตาล และหมู่ฟอสเฟตเป็นแกนของเกลียว (DNA backbone) แต่ละรอบเกลียวประกอบด้วยเบส 10 คู่ ซึ่งเกลียวคู่ของดีเอ็นเอถูกยึดด้วยพันธะไฮโดรเจนระหว่างเบสที่อยู่บนสายตรงข้ามกัน โดย A จับคู่กับ T ด้วยพันธะไฮโดรเจน 2 พันธะ และ C จับคู่กับ G ด้วยพันธะไฮโดรเจน 3 พันธะ นอกจากนั้นยังมีแรงไฮโดรโฟบิค (Hydrophobic interaction) ที่เกิดระหว่างเบสในสายเดียวกันเพื่อช่วยยึดโครงสร้างเกลียวคู่

---

[3] http://wine1.sb.fsu.edu/bch5425/lect02/IMG00006.GIF



รูปที่2.4 โครงสร้างของดีเอ็นเอ [4]

2.3.1.3 **หน้าที่ของดีเอ็นเอ**

ดีเอ็นเอมีหน้าที่ภายในเซลล์ที่สำคัญ 2 อย่างคือ

1 จำลองตัวเอง (DNA replication) เมื่อเกิดการแบ่งเซลล์เพื่อสร้างดีเอ็นเอให้เซลล์ใหม่

2 ถ่ายทอดข้อมูลผ่านอาร์เอ็นเอ (RNA) เพื่อกำหนดการเรียงตัวของกรดอะมิโนในโปรตีน กระบวนการถอดรหัสดีเอ็นเอเพื่อสร้างอาร์เอ็นเอ (Transcription) และการแปลรหัส (Translation)

แม้ร่างกายจะมีดีเอ็นเอเหมือนกันทุกเซลล์ แต่เซลล์แต่ละตัวจะมีการแสดงออกของยีนในลักษณะที่ต่างกันออกไป ทำให้อวัยวะส่วนต่างๆมีคุณสมบัติแตกต่างกัน

2.3.1.4 **ความแตกต่างของดีเอ็นเอ** (DNA polymophisms)

ความแตกต่างของลำดับเบสบนสายดีเอ็นเอที่ตำแหน่งต่างๆในจีโนมมนุษย์พบได้ระหว่างบุคคล หรือแม้แต่ในบุคคลเดียวกัน มีการประมาณว่าในแต่ละบุคคลจะพบความแตกต่างในดีเอ็นเอทุก 500-1000 เบส โดยสาเหตุที่ทำให้เกิดความแตกต่างคือ

---

[4] http://coris.noaa.gov/glossary/nucleotide_186.jpg

1 การเปลี่ยนของเบสตัวเดียว

การเปลี่ยนพบได้ทั้งการแทนที่เบสด้วยเบสอื่น (Nucleotide substitution) และการเพิ่มหรือหายของเบสเพียง 1 ตัว โดยสาเหตุที่ทำให้เกิดการเปลี่ยนของเบสตัวเดียวมาจากความผิดพลาดระหว่างการจำลองตัวของดีเอ็นเอ หรือเกิดจากการถูกกระตุ้นให้เปลี่ยนแปลงด้วยสิ่งก่อการกลาย (Mutagen)

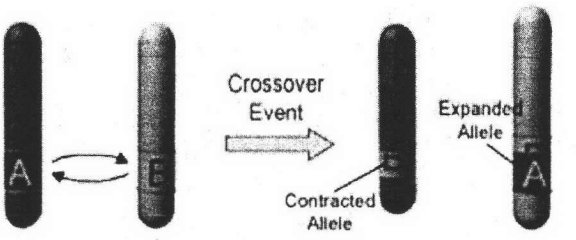
2 การขาดหาย หรือการสอดแทรกของดีเอ็นเอ

การขาดหายของดีเอ็นเอ (Deletion) พบในโรค เช่น โรคธาลัสซีเมีย และโรคดูเชนนี มัสคูลาร์ ดิสโทรฟีย์ (Duchenny muscular dystrophy) ซึ่งการขาดหายของดีเอ็นเอมีสาเหตุเกิดจากความผิดปกติเมื่อมีการแลกเปลี่ยนส่วนของดีเอ็นเอในกระบวนการไขว้เปลี่ยน ส่วนความแตกต่างที่เกิดจากการแทรก (Insertion) จะมีโอกาสพบได้น้อยกว่า

รูปที่2.5 การกลายพันธุ์บนสายดีเอ็นเอ

3 การขาดหาย หรือการขยายจำนวนของดีเอ็นเอในกระบวนการไขว้เปลี่ยน (Unequal crossing over)

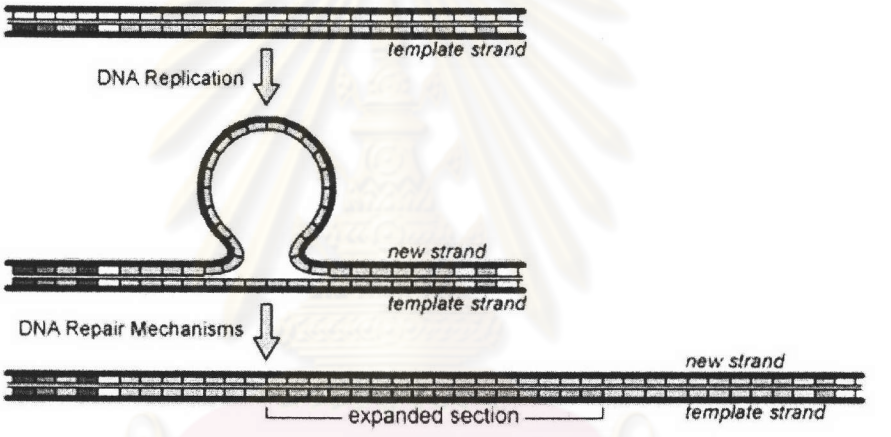
ความผิดปกติชนิดนี้มักพบในส่วนของจีโนมที่มีลำดับเบสดีเอ็นเอคล้ายกัน เช่น ในยีนกลุ่มของยีนเดียวกัน (Multigene families) หรือในส่วนของดีเอ็นเอที่ไม่ใช่รหัส โดยเมื่อเกิดการแลกเปลี่ยนส่วนของดีเอ็นเอระหว่างการไขว้เปลี่ยนของโครโมโซมคู่เหมือน ซึ่งอาจมีการเรียงตัวผิด (Mis-align) ของลำดับเบสที่คล้ายกัน ทำให้เกิดการไขว้เปลี่ยนของโครมาทิดที่ไม่เท่ากัน เป็นผลให้เกิดการขาดหาย หรือการขยายจำนวนของดีเอ็นเอ



รูปที่2.6 การขาดหาย หรือการขยายจำนวนของดีเอ็นเอในกระบวนการไขว้เปลี่ยน [5]

4 การขยายจำนวนของดีเอ็นเอที่เกิดจากการเลื่อนของเบส (Slippage bases)

ในบริเวณเบสซ้ำที่มีเบสแกนขนาดเล็กเรียงตัวกันหลายหน่วยอาจมีการเลื่อนของเบสทำให้เกิดการกลายพันธุ์ (Mutation) ของดีเอ็นเอ การเลื่อนทำให้เกิดการซ้อนกันของเบส และเกิดปลายยื่นออกมา จึงต้องมีการเติมเบสโดยเอมไซม์พอลิเมอเรส ทำให้เกิดการขยายจำนวนเบสซ้ำในบริเวณดังกล่าว และเกิดความแตกต่างของดีเอ็นเอขึ้น



รูปที่2.7 การขยายจำนวนของดีเอ็นเอที่เกิดจากการเลื่อนของเบส [6]

2.3.2 องค์ประกอบและโครงสร้างของอาร์เอ็นเอ

อาร์เอ็นเอเป็นกรดนิวคลีอิกชนิดหนึ่งมีโครงสร้าง และมีองค์ประกอบคล้ายดีเอ็นเอ โดยประกอบจาก 3 โมเลกุลของน้ำตาลไรโบส (CHO) เบส และกรดฟอสโฟริก (H_3PO_4) เบสที่พบแบ่งได้เป็น 2 ประเภทคือ

1 เบสพิริมิดีน ได้แก่ ไซโทซีน (Cytosine: C) และ ยูราซิล (Uracil: U)

2 เบสพิวรีน ได้แก่ อะดินีน (Adenine: A) และ กัวนีน (Guanine: G)

อาร์เอ็นเอแบ่งได้ 3 ชนิดตามหน้าที่

---

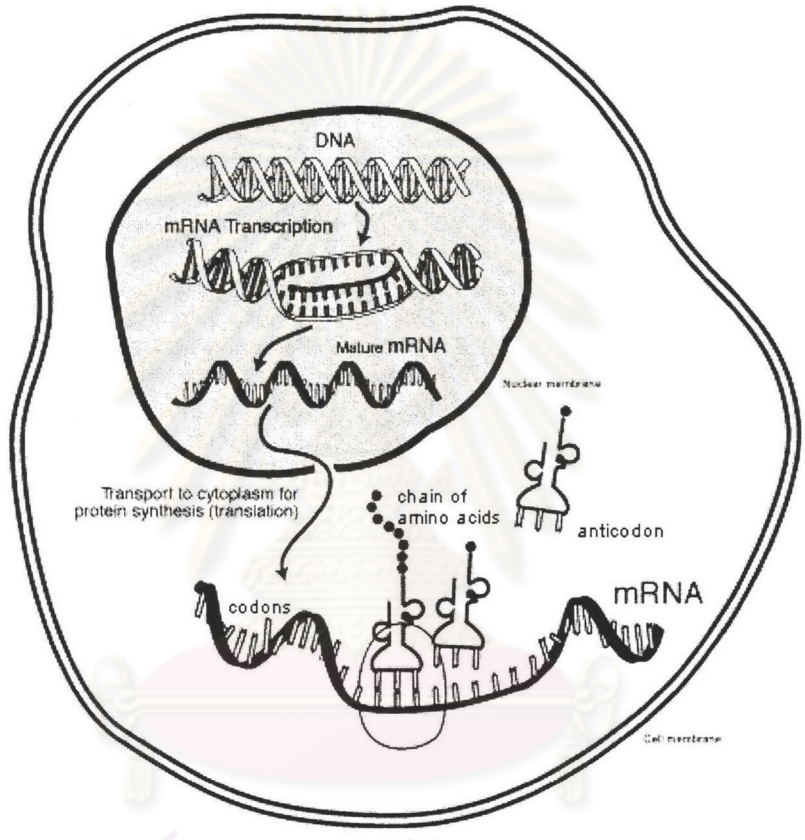
[5] http://www.stanford.edu/group/hopes/causes/mutation/f_q03unequal.jpg

[6] http://www.stanford.edu/group/hopes/causes/mutation/f_q05slippage.gif

1 เอ็มอาร์เอ็นเอ (messenger RNA, mRNA) ทำหน้าที่จำลองสายดีเอ็นเอ (Transcription) และนำรหัสพันธุกรรมออกจากนิวเคลียสไปไซโตพลาสซึม

2 ทีอาร์เอ็นเอ (transfer RNA, tRNA) เป็นอาร์เอ็นเอที่มีขนาดเล็ก และสำคัญในการสร้างโปรตีน โดยทำการคัดลอกรหัสจากเอ็มอาร์เอ็นเอ (Codon) และทำการรวมรหัสเข้าเป็นโพลีเปปไทด์

3 อาร์อาร์เอ็นเอ (ribosomal RNA, rRNA) ทำหน้าที่รวมตัวกับโปรตีนเป็นไรโบโซม



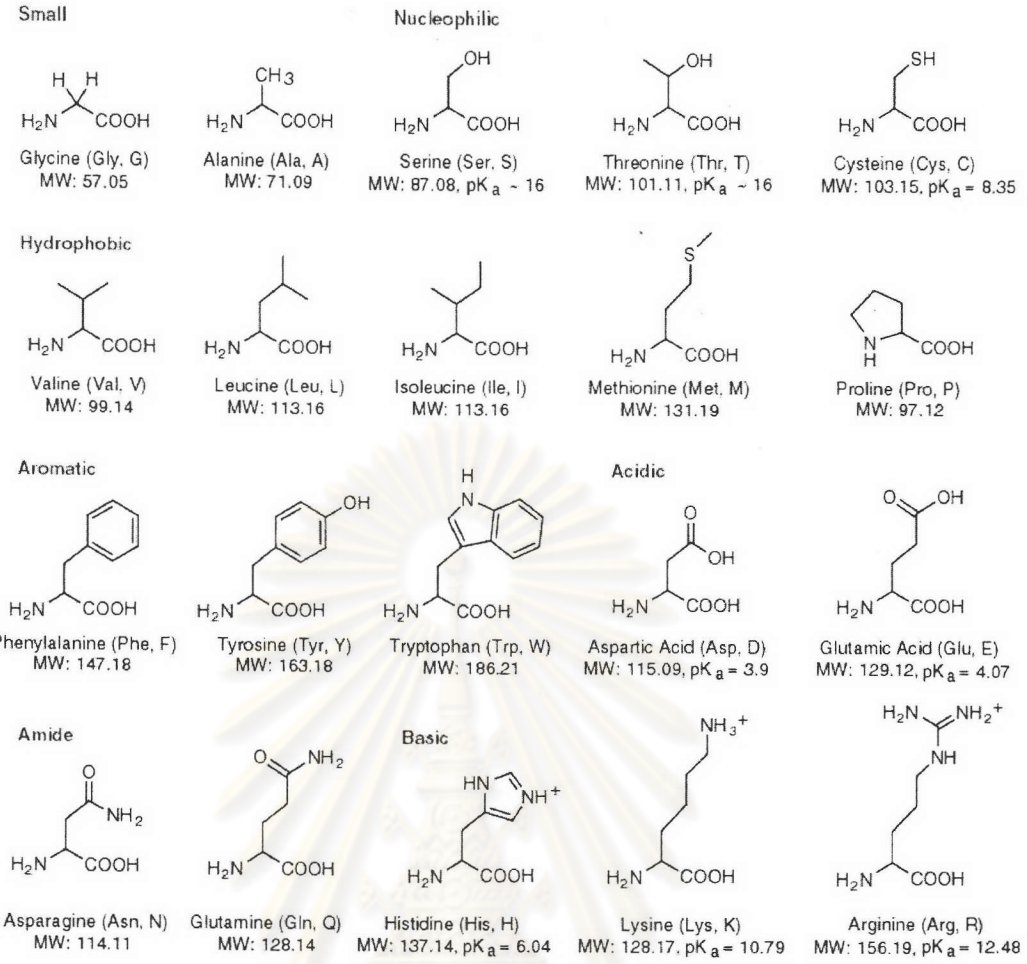
รูปที่2.8 หน้าที่ของอาร์เอ็นเอ [7]

2.3.3 องค์ประกอบและโครงสร้างของโปรตีน

โปรตีนประกอบด้วยกรดอะมิโน 20 ชนิดที่เกิดจากสายพันธุกรรมที่มีความยาว 3 ตัวอักษร (Codon) มาเชื่อมกันเป็นสายยาวด้วยพันธะเปปไทด์ (Peptide bond) ซึ่งเป็นพันธะเอไมด์ (Amide bond, CO-NH) ที่เป็นการรวมตัวกันระหว่างหมู่คาร์บอกซิลิค (Carboxylic group, COOH) กับหมู่อะมิโน (Amino group, NH_2) ของกรดอะมิโนแต่ละตัว

---

[7] http://jura.wi.mit.edu/bio/education/BOA-2000/BOA/mrna.gif



รูปที่2.9 โครงสร้างกรดอะมิโนทั้ง 20 ชนิด [8]

โครงสร้างโปรตีนแบ่งได้เป็น 4 ระดับคือ

1 โครงสร้างปฐมภูมิ (Primary structure) เป็นการเรียงลำดับของกรดอะมิโนที่เป็นสายโพลีเปปไทด์เพียงเส้นเดียว

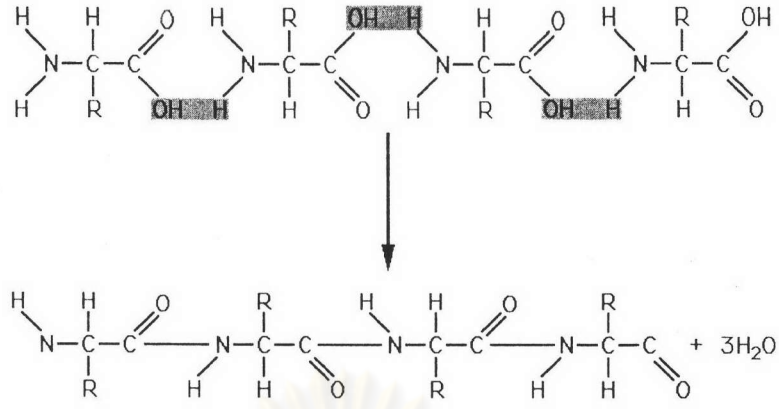
2 โครงสร้างทุติยภูมิ (Secondary structure) เป็นโครงสร้างการพันกันของสายโพลีเปปไทด์ โดยมีลักษณะบิดเป็นเกลียว โดยมีโครงสร้างที่สำคัญได้แก่เกลียวอัลฟา (α -helix) และแผ่นจีบบีตา (β -pleat)

3 โครงสร้างตติยภูมิ (Tertiary Structure) เป็นการพันกันของโปรตีนในระดับโครงสร้างทุติยภูมิให้อยู่ในสถาวะสามมิติ

4 โครงสร้างจตุรภูมิ (Quaternary Structure) เป็นโครงสร้างโปรตีนซึ่งประกอบด้วยสายโพลีเปปไทด์หลายสายรวมตัวเป็น 1 โมเลกุลโปรตีน

---

[8] http://www.neb.com/nebecomm/tech_reference/images/amino.gif



รูปที่2.10 โครงสร้างโปรตีน [9]

2.4 ปัญหาการจัดเรียงลำดับเบสหลายลำดับ

ปัญหาการจัดเรียงลำดับเบสหลายลำดับเป็นการเปรียบเทียบ และจัดเรียงรหัสดีเอ็นเอ หรือรหัสกรดอะมิโนเพื่อหาความสัมพันธ์กันของวิวัฒนาการในสิ่งมีชีวิต หรือเพื่อหาโครงสร้างสามมิติของโปรตีน ในการแก้ปัญหาที่แต่ละลำดับมีสายพันธุ์ใกล้เคียงกัน (High similarity) จะทำการจัดเรียงให้ในแต่ละแถวมีตัวอักษรเหมือนกันมากที่สุด โดยเมื่อแต่ละลำดับมีความยาวไม่เท่ากันจะมีการเพิ่มแก็ป (Gaps) เข้าไปในการจัดเรียง

2.4.1 รูปแบบข้อมูลที่ใช้กับปัญหา

เนื่องจากรหัสดีเอ็นเอ หรือรหัสกรดอะมิโนยังไม่อยู่ในรูปข้อมูลดิจิตอล จึงต้องทำการเข้ารหัสก่อน โดยรหัสดีเอ็นเอกำหนดให้มี 4 ตัวอักษร (A, T, C, G) และ รหัสกรดอะมิโนมี 20 ตัวอักษร (C, S, T, P, A, G, N, D, E, Q, H, R, K, M, I, L, V, F, Y, W) การเข้ารหัสมีรูปแบบแตกต่างกันตามการใช้งาน ในที่นี้จะทำการอธิบายข้อมูลแบบเอ็มเอสเอฟ (MSF format)

```
hmgl_trybr      .kkdsnapkr amtsfmffss ....dfrskh sdlsi.vems kaagaawkel
hmgt_mouse      ......kpkr prsayniyvs esfqeakdds aqgkl..... klvneawknl
hmgb_chite      ....adkpkr plsaymlwln saresikren pdfkv.teva kkggelwrgl
hmgl_wheat      ...dpnkpkr apsaffvfmg efreefkqkn pknksvaavg kaagerwksl


hmgl_trybr      gpeerkvyee maekdkeryk rem..... ..
hmgt_mouse      speekqayiq lakddriryd nemksweeqm ae
hmgb_chite      ..kdksewea kaatakqnyi ralqeyerng g.
hmgl_wheat      sesekapyva kanklkgeyn kaiaaynkge sa
```

รูปที่2.11 ตัวอย่างข้อมูลรูปแบบเอ็มเอสเอฟของรหัสกรดอะมิโน

---

[9] http://www.agen.ufl.edu/~chyn/age2062/lect/lect_02/3_20.gif

จากรูปที่2.11 รูปแบบเอ็มเอสเอฟจะขึ้นต้นด้วยชื่อลำดับกรดอะมิโนของแต่ละลำดับ ตามด้วยรหัสกรดอะมิโนที่มีตัวอักษรแตกต่างกันทั้งหมด 20 ตัวอักษร โดยลำดับกรดอะมิโนแต่ละลำดับมีความยาว 50 ตัวอักษร โดยทุกๆ 10 ตัวอักษรจะมีช่องว่างกั้น เมื่อมีตัวอักษรเกิน 50 ตัวในลำดับ จะทำการตัดส่วนที่เหลือขึ้นบรรทัดใหม่

2.4.2 คะแนนที่ใช้คำนวณความถูกต้อง

ในการแก้ปัญหาการจัดลำดับเบสหลายลำดับต้องการจัดเรียงให้แถวมีตัวอักษรเหมือนกันมากที่สุดจึงต้องมีการหาวิธีการคำนวณที่ใช้ในการแก้ปัญหา ซึ่งวิธีการคำนวณที่นิยมใช้จะได้มาจากตารางการแทนที่ได้จากทฤษฎีความน่าจะเป็นมาคำนวณหาสมการผลรวมคู่เบส หรืออีกวิธีหนึ่งคือการคำนวณหาเอ็นโทรปีต่ำสุด

2.4.2.1 สมการผลรวมคู่เบส (Sum-of-pairs) และตารางการแทน (Substitution matrix)

วิธีในการคำนวณการจัดเรียงลำดับเบสสองลำดับจะใช้สัดส่วนของความน่าจะเป็นที่จะเกิดตัวอักษรคู่กันในตำแหน่งเดียวกัน กับความน่าจะเป็นที่จะเกิดตัวอักษรคู่กันในตำแหน่งสุ่มใดก็ได้ในลำดับ

กำหนดให้มีสองลำดับเป็น x และ y ที่มีความยาว n และ m ตามลำดับ ตัว i กำหนดตำแหน่ง x_i และ j กำหนดตำแหน่ง y_j กำหนดตัวอักษร a และ b เป็นตัวอักษรในดีเอ็นเอหรือกรดอะมิโน

ให้ R เป็นตัวอักษรคู่กันในตำแหน่งสุ่มทั้งหมดของลำดับ และกำหนดให้ความถี่ตัวอักษร a ที่มีความอิสระต่อกัน (Independent) อยู่บนลำดับมีค่า q_a ความน่าจะเป็นที่จะเกิดตัวอักษรคู่กันในตำแหน่งใดก็ได้ในลำดับจะเป็นดังสมการ 2.1

$$P(x,y|R)=\prod_i q_{x_i}\prod_j q_{y_j} \tag{2.1}$$

ให้ M เป็นตัวอักษรคู่กันในตำแหน่งเดียวกันทั้งหมดของลำดับ และความน่าจะเป็นที่จะเกิดตัวอักษรคู่กันมีค่า p_{ab} ความน่าจะเป็นที่จะเกิดตัวอักษรคู่กันในตำแหน่งเดียวกันจะเป็นดังสมการ 2.2

$$P(x,y|M)=\prod_i p_{x_i y_i} \tag{2.2}$$

สัดส่วนของความน่าจะเป็นคือ

$$\frac{P(x,y|M)}{P(x,y|R)}=\frac{\prod_i p_{x_i y_i}}{\prod_i q_{x_i}\prod_i q_{y_i}}=\prod_i \frac{p_{x_i y_i}}{q_{x_i} q_{y_i}} \tag{2.3}$$

ทำการใช้ลอการิทึม (Logarithm) เพื่อง่ายแก่การคำนวณ โดยให้ S เป็นค่าคะแนนทั้งหมดที่ได้จากการจัดเรียงลำดับเบสสองลำดับ หรือเป็นค่าผลรวมคู่เบส (Sum-of-pairs) สมการที่ได้จะอยู่ในรูปสมการที่ 2.4

$$S = \sum_i s\left(x_i, y_i\right) \tag{2.4}$$

เมื่อ

$$s\left(a,b\right) = \log\left(\frac{p_{ab}}{q_a q_b}\right) \tag{2.5}$$

ตารางการแทนที่ได้จะมีค่ามาจาก $s\left(a,b\right)$ ถ้าเป็นตารางการแทนของดีเอ็นเอ ตารางจะมีลักษณะเป็นตาราง 2 มิติขนาด 4×4 และถ้าเป็นตารางการแทนของกรดอะมิโนตารางจะมีลักษณะเป็นตาราง 2 มิติขนาด 20×20 โดยที่ $s\left(a_i, a_j\right)$ เป็นค่าที่ตำแหน่ง i และ j ในตาราง ตารางการแทนที่นิยมใช้คือตาราง PAM [17] พัฒนาโดย Dayhoff, Schwartz และ Orcutt ในปี 1978 หรือตาราง BLOSUM [18] พัฒนาโดย Gonnet, Cohen และ Benner ในปี1992

2.4.2.2 **วิธีการคำนวณหาเอนโทรปีต่ำสุด** (Minimum entropy)

วิธีการหาเอนโทรปีต่ำสุดเป็นอีกวิธีในการคำนวณหาค่าความถูกต้องจากการจัดเรียงลำดับเบสโดยจะคำนวณจากเอ็นโทรปีในแต่ละแถว

กำหนดให้ m_i^j เป็นตัวอักษรในแถวที่ i ลำดับที่ j c_{ia} เป็นจำนวนตัวอักษร a ในแถวที่ i p_{ia} ความน่าจะเป็นของตัวอักษร a ในแถวที่ i และ a เป็นตัวอักษรหนึ่งของกรดอะมิโนที่มีทั้งหมด 20 ตัว ({C,S,T,P,A,G,N,D,E,Q,H,R,K,M,I,L,V,F,Y,W })

รูปที่2.12 ตัวอย่างลำดับเบสหลายลำดับสำหรับคำนวณเอ็นโทรปี

ค่า p_{ia} พอจะประมาณค่าได้ดังสมการ 2.6

$$P_{ia} = \frac{c_{ia}}{\sum_b c_{ib}} \tag{2.6}$$

ยกตัวอย่างการคำนวณจากรูปที่2.12 p_{iY} ในแถว m_i จะมีค่าเท่ากับ

$$P_{iY} = \frac{c_{iY}}{c_{iA} + c_{iC} + c_{iD} + ... + c_{iY}} = \frac{1}{4+1} = \frac{1}{5} \tag{2.7}$$

เพราะฉะนั้นค่าความน่าจะเป็นของแต่ละแถวจะมีลักษณะดังสมการที่ 2.8

$$\mathrm{P}\left(m_i\right) = \prod_a p_{ia}^{c_{ia}} \tag{2.8}$$

ใช้ลอการิทึมเปลี่ยนเป็นสมการผลรวมการบวก ดังสมการที่ 2.9

$$S\left(m_i\right) = \sum_a c_{ia} \log p_{ia} \tag{2.9}$$

สมการที่ใช้ในการคำนวณเอ็นโทรปีทุกแถวคือ

$$S = \sum_i S\left(m_i\right) \tag{2.10}$$

การจัดเรียงลำดับเบสที่ดีควรคำนวณได้ค่า S ไม่มาก

2.4.2.3 การหักคะแนนเมื่อมีแก็ป

การจัดเรียงลำดับอาจมีการเพิ่มแก็ปได้หลายวิธี ดังนั้นจึงต้องมีวิธีบ่งบอกว่าการเพิ่มแก็ปวิธีไหนดีที่สุด สมการที่นิยมใช้คือ

$$GapPenalties = g + hl \tag{2.11}$$

เมื่อ $GapPenalties$ คือการหักคะแนนเมื่อมีแก็ปในลำดับ g คือค่าคงที่ในการหักคะแนนเมื่อพบแก็ปแรก l คือความยาวกลุ่มของแก็ปที่ตามหลังแก็ปตัวแรก และ h คือค่าคงที่ในการหักคะแนนแก็ปที่ตามหลังแก็ปตัวแรก

Sequence1	AGTCACAT	Sequence1	AGTCACAT
Sequence2	AGT - - - AT	Sequence2	AGT - A - - T
แบบที่1		แบบที่2	

รูปที่2.13 ตัวอย่างการคำนวณการหักคะแนนเมื่อมีแก็ป

จากรูปที่2.13 สมมุติให้ค่าคงที่แก็ปเริ่มต้นเท่ากับ 10 ค่าคงที่แก็ปตามหลังแก็ปตัวแรกเท่ากับ 1 การจัดเรียงแบบที่1 มีค่าการหักคะแนนเท่ากับ (10+1*2) = 12 ส่วนการจัดเรียงแบบที่2

มีค่าการหักคะแนนเท่ากับ (10)+(10+1*1) = 21 การจัดเรียงแบบที่ 1 มีค่าการหักคะแนนน้อยกว่า จึงเป็นการจัดเรียงที่ดีกว่าการจัดเรียงแบบที่ 2

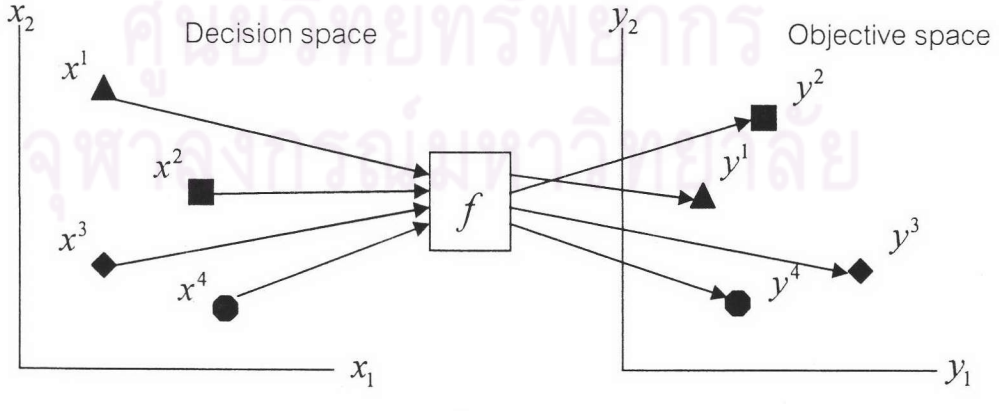
การหักคะแนนเมื่อมีแก็ปเป็นที่นิยมใช้ในการคำนวณร่วมกับสมการผลรวมคู่เบส (Sum-of-pairs) ในทางปฏิบัติแล้วการกำหนดค่าคงที่แก็ปเริ่มต้น และค่าคงที่แก็ปตามหลังแก็ปตัวแรก สามารถกำหนดได้ตามต้องการ และจากสมการผลรวมคู่เบสร่วมกับการหักคะแนนเมื่อมีแก็ป (affine gap penalties) เมื่อมีคะแนนมากจะมีแนวโน้มในการจัดเรียงถูกต้องสูง แต่การมีคะแนนมากบางครั้งอาจไม่ได้บ่งบอกถึงการจัดเรียงที่ถูกต้อง

2.5 **การแก้ปัญหาแบบหลายวัตถุประสงค์** (Multiple objective optimization)

การแก้ปัญหาในโลกจริง (Real-world problems) นั้นส่วนใหญ่แล้วมีปัจจัยในการเลือกเพื่อแก้ปัญหาได้หลายรูปแบบ ซึ่งแต่ละรูปแบบมีลักษณะที่แตกต่างกันออกไป การแก้ปัญหาต้องเลือกปัจจัยต่างๆอย่างเหมาะสม โดยไม่ขึ้นอยู่กับปัจจัยเพียงตัวเดียว (Trade-offs) เพื่อจะได้คำตอบที่มีประสิทธิภาพสูงสุด

2.5.1 **ทฤษฎีเบื้องต้นของการแก้ปัญหาแบบหลายวัตถุประสงค์**

การแก้ปัญหาแบบหลายวัตถุประสงค์สามารถมองในรูปแบบของเวกเตอร์ (Vector) สมมุติให้เวกเตอร์ในการตัดสินใจ (Decision vector) ประกอบไปด้วยสมาชิก $(x_1, x_2, ..., x_n)$ อยู่ในช่วง (Decision space) X จากนั้นกำหนดให้มีฟังก์ชัน $f : X \to Y$ ทำการเปลี่ยนค่าเวกเตอร์ในการตัดสินใจเป็นเวกเตอร์วัตถุประสงค์ (Objective vector) $(y_1, y_2, ..., y_n)$ ซึ่งอยู่ในช่วง (Objective space) Y



รูปที่2.14 กราฟแสดงปัญหาแบบสองวัตถุประสงค์

จากรูปที่2.14 จะพิจารณาให้มีเวกเตอร์ในการตัดสินใจ และเวกเตอร์วัตถุประสงค์อยู่ในรูปแกนสองมิติ โดยเวกเตอร์ x^i ประกอบไปด้วยสมาชิก (x_1^i, x_2^i) และ y^i ประกอบไปด้วยสมาชิก (y_1^i, y_2^i)

ในกรณีการเปรียบเทียบคำตอบ x^1 และ x^2 ตามแนวคิดพาเรโต (Pareto dominance) เมื่อเวกเตอร์วัตถุประสงค์ y^1 จะเด่นกว่าเวกเตอร์วัตถุประสงค์ y^2 ($y^1 \succ y^2$) ก็ต่อเมื่อไม่มีค่าของแต่ละสมาชิกใน y^1 น้อยกว่าค่าของแต่ละสมาชิกใน y^2 และสมาชิกไม่น้อยกว่าหนึ่งตัวใน y^1 มีค่ามากว่าในสมาชิก y^2 สามารถคิดในอีกทางหนึ่งได้ว่าคำตอบ x^1 ดีกว่า x^2 ($x^1 \succ x^2$) ถ้า $f(x^1)$ เด่นกว่า $f(x^2)$

เซตของคำตอบที่ดีในช่วง X หมายถึง พาเรโตเซต X^* (Pareto set) โดยที่ $X^* \subseteq X$ และเซตของคำตอบที่ดีในช่วง Y หมายถึง พาเรโตฟรอนต์ Y^* (Pareto front) โดยที่ $Y^* = f(X^*) \subseteq Y$

2.6 **ขั้นตอนวิธีเชิงวิวัฒน์** (Evolutionary Algorithm)

ขั้นตอนวิธีเชิงวิวัฒน์เหมาะสำหรับการแก้ปัญหาแบบหลายวัตถุประสงค์ เพราะในการปรับปรุงคำตอบแต่ละรอบจะผลิตคำตอบออกมาหลายค่า ซึ่งทำให้สามารถหาค่าเซตของคำตอบที่ดีได้ในการคำนวณเพียงครั้งเดียว

ขั้นตอนวิธีเชิงวิวัฒน์สามารถแยกย่อยได้เป็น ขั้นตอนวิธีเชิงพันธุกรรม (Genetic algorithm), การโปรแกรมเชิงพันธุกรรม (Genetic programming) และกลยุทธ์เชิงวิวัฒน์ (Evolutionary strategy) เป็นต้น ซึ่งในงานวิจัยนี้จะขอกล่าวถึงขั้นตอนวิธีเชิงพันธุกรรมเป็นตัวอย่าง

2.6.1 **ขั้นตอนวิธีพันธุกรรม** (Genetic Algorithms)

ขั้นตอนวิธีเชิงพันธุกรรม [19] ซึ่งถูกคิดค้นขึ้นโดย Holland (1975) จัดอยู่ในกลุ่มของการคำนวณเชิงวิวัฒน์เป็นวิธีการที่ใช้ในการค้นหา และค้นหาผลเฉลยที่ดีที่สุด (Optimization) วิธีการหนึ่ง ซึ่งขั้นตอนวิธีเชิงพันธุกรรมนั้นมีการค้นหาคำตอบโดยเลียนแบบพฤติกรรมการวิวัฒนาการตามธรรมชาติตามหลักทฤษฎีของดาวิน (Darwin) ซึ่งพ่อแม่สามารถถ่ายทอดลักษณะต่างไปยังรุ่นลูกได้ โดยรุ่นลูกนั้นจะมีลักษณะของรุ่นพ่อ หรือ/และรุ่นแม่อยู่

ขั้นตอนวิธีเชิงพันธุกรรมจะทำการแทนผลเฉลยของปัญหาต่างๆในรูปของโครโมโซม (Chromosome) โดยแต่ละโครโมโซมจะมีลักษณะเป็นสายอักขระเลขฐานสอง (Bit String) ซึ่งสายอักขระดังกล่าวแต่ละสายรวมกันจะถูกเรียกว่าประชากร (Population) การประเมินค่าความ

ถูกต้องของผลเฉลยต่าง ๆ จะกระทำโดยแปลงสายอักขระหรือประชากรดังกล่าวกลับไปเป็นผลเฉลยในรูปแบบที่เหมาะสม แล้วจึงทำการประเมินค่า

ชุดของผลเฉลยทั้งหมดที่ทำการประเมินค่าและทำการเลือกประชากรที่ดีออกมานั้น แต่ละชุดจะถูกเรียกว่ารุ่น (Generation) ในแต่ละรุ่นของประชากรจะต้องผ่านกระบวนการต่อไปนี้ คือ การประเมินค่า การเลือกประชากรที่มีลักษณะดี และการสร้างชุดผลเฉลยชุดใหม่ขึ้น

ขั้นตอนวิธีเชิงพันธุกรรมจะมีขั้นตอนวิธี (Algorithm) ดังรูปที่2.15 เมื่อ t คือจำนวนรุ่น และ T คือจำนวนรุ่นที่มากที่สุด

ขั้นที่1: กำหนดค่าเริ่มต้นโดยการสุ่มเป็นประชากรเริ่มต้น

$t = 0$

ขั้นที่2: คำนวนหาค่าความแข็งแรงให้กลุ่มประชากร

ขั้นที่3: ตรวจเงื่อนไขการหยุดโปรแกรมเมื่อ $t \geq T$

ขั้นที่4: ทำการกลายพันธุ์ หรือการไขว้เปลี่ยน แล้วนำผลที่ได้เป็นประชากรรุ่นถัดไป

ขั้นที่5: คำนวณฟังก์ชันวัตถุประสงค์ให้กับกลุ่มประชากรรุ่นถัดไป

$t = t + 1$

ทำซ้ำขั้นที่3

รูปที่2.15 รหัสเทียมของขั้นตอนวิธีเชิงพันธุกรรม

วิธีการสร้างชุดผลเฉลยใหม่นั้น จะอาศัยตัวดำเนินการทางพันธุกรรมมากระทำกับสายอักขระ โดยตัวดำเนินการที่ใช้ส่วนมาก มีอยู่ 3 ตัว คือ การกลายพันธุ์ (Mutation) การไขว้เปลี่ยน (Crossover) และการสืบพันธุ์ ซึ่งจะกล่าวในรายละเอียดต่อไป

2.6.2 **รายละเอียดของขั้นตอนวิธีเชิงพันธุกรรม**

ขั้นตอนในการค้นหาผลเฉลยของขั้นตอนวิธีเชิงพันธุกรรมนั้น ประกอบด้วย การสร้างกลุ่มประชากรของผลเฉลยเริ่มต้น การตรวจสอบค่าความแข็งแรงของผลเฉลย การสร้างกลุ่มประชากรของผลเฉลยรุ่นใหม่ และการค้นหาคำตอบ มีคำอธิบายรายละเอียดดังนี้

2.6.2.1 **การสร้างกลุ่มประชากรของผลเฉลยตั้งต้น**

ขั้นตอนวิธีเชิงพันธุกรรมจะทำการสุ่มอักขระเลขฐานสองให้กับผลเฉลยเริ่มต้นทุกตัวในกลุ่มประชากร โดยสายอักขระเลขฐานสองของผลเฉลยเริ่มต้นนั้นจะมีขนาดที่แน่นอน

2.6.2.2 การตรวจสอบค่าความแข็งแรงของผลเฉลย

ในขั้นตอนนี้ เป็นการนำเอาผลเฉลยต่างๆมาทำการทดสอบเพื่อวัดความถูกต้อง หรือประสิทธิภาพของผลเฉลย ค่าความถูกต้องนี้จะเรียกว่าค่าความแข็งแรงของผลเฉลย (Fitness) จะถูกนำไปใช้ในการคัดเลือกผลเฉลยที่จะนำไปเป็นพื้นฐานในการสร้างผลเฉลยรุ่นถัดไป

การประเมินค่าความแข็งแรงจะทำโดยการแปลงผลเฉลยซึ่งอยู่ในรูปของสายอักขระให้เป็นผลเฉลยในรูปแบบของแต่ละปัญหาเพื่อทำการทดสอบ และนำผลเฉลยดังกล่าวไปทำการประเมินค่าตามข้อกำหนดของปัญหาต่อไป

ค่าความแข็งแรงที่ดีต้องบอกถึงความแตกต่างของผลเฉลยได้เป็นอย่างดี โดยความแตกต่างของผลเฉลยนั้นจะเป็นตัวกำหนดว่าผลเฉลยตัวนั้นนั้นดีหรือไม่ดีเป็นค่าเท่าไหร่

2.6.2.3 การสร้างกลุ่มประชากรของผลเฉลยรุ่นใหม่

การสร้างกลุ่มประชากรใหม่นั้นจะเกิดจากการคัดเลือกผลเฉลยที่ได้จากการคัดเลือก แล้วนำมาสร้างผลเฉลยใหม่จากกระบวนการการสืบพันธุ์ (Reproduction) การกลายพันธุ์ (Mutation) และ การไขว้เปลี่ยน (Crossover) โดยผลเฉลยใหม่ที่ได้นั้นจะมีลักษณะบางส่วนของผลเฉลยที่ถูกคัดเลือกมาอยู่ด้วย

1 การสืบพันธุ์

การสืบพันธุ์เป็นการสร้างผลเฉลยใหม่ที่มีลักษณะเหมือนกับผลเฉลยต้นแบบทุกประการ โดยไม่มีการเปลี่ยนแปลงใด ๆ ตัวดำเนินการนี้มักนิยมใช้กับผลเฉลยที่มีค่าความแข็งแรง (Fitness) สูงที่สุดในรุ่น เพื่อคงลักษณะที่ดีที่สุดไว้

2 การกลายพันธุ์

การกลายพันธุ์เป็นการสร้างผลเฉลยใหม่จากผลเฉลยต้นแบบจำนวน 1 ผลเฉลย โดยทำการเปลี่ยนแปลงผลเฉลยไปบางส่วน การเปลี่ยนแปลงจะทำโดยสุ่มเลือกตำแหน่งในสายอักขระของผลเฉลย แล้วทำการเปลี่ยนแปลงค่าของอักขระในตำแหน่งของสายอักขระนั้น การเปลี่ยนแปลงค่าที่จะทำการกลายพันธุ์นั้นจะขึ้นอยู่กับอัตราการกลายพันธุ์ โดยอัตราการกลายพันธุ์นั้นสามารถที่จะใช้อัตราส่วนที่เท่ากันได้ หรือจะแปรเปลี่ยนค่าเพื่อให้สอดคล้องกับแต่ละปัญหา

1 0 1 1 0 0

(ก)

1 0 0 1 0 0

(ข)

รูปที่2.16 แสดงตัวอย่างการกลายพันธุ์จากผลเฉลยต้นแบบ (ก) ไปเป็นผลเฉลยใหม่ (ข)

3 การไขว้เปลี่ยน

การไขว้เปลี่ยนเป็นการสร้างผลเฉลยใหม่จากผลเฉลยต้นแบบจำนวน 2 ผลเฉลย โดยทำการแลกเปลี่ยนส่วนประกอบบางส่วนของผลเฉลยทั้งสองซึ่งกันและกัน การแลกเปลี่ยนจะทำการสุ่มเลือกตำแหน่งในสายอักขระของผลเฉลยทั้งสอง จากนั้นทำการตัดแบ่งสายอักขระทั้งสองที่ตำแหน่งนั้น และทำการสลับส่วนที่ตัดของสายอักขระแรก กับสายอักขระที่สอง แล้วทำการต่อส่วนที่ทำการสลับนั้นเข้าไป ซึ่งจะทำให้ได้ผลเฉลยใหม่สองผลเฉลย ที่มีลักษณะร่วมของผลเฉลยต้นแบบทั้งสอง

100001	10|0001
001100	00|1100
(ก)	(ข)

101100

000001

(ค)

รูปที่2.17 ตัวอย่างการไขว้เปลี่ยน จากผลเฉลยต้นแบบ 2 ผลเฉลย

(ก) ผลเฉลยต้นแบบ

(ข) การตัดแบ่งผลเฉลย

(ค) ผลเฉลยใหม่ที่ได้จากการไขว้เปลี่ยน

2.6.2.4 การค้นหาคำตอบ

การค้นหาคำตอบโดยขั้นตอนวิธีเชิงพันธุกรรมนั้นจะทำการปรับปรุงผลเฉลย ซ้ำไปเรื่อย ๆ โดยวงรอบกระบวนการนี้จะหยุดเมื่อผลเฉลยตรงกับคำตอบที่ต้องการ หรือทำจนครบรอบเท่ากับจำนวนรอบที่ได้กำหนดไว้

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย