



บทที่ 1 บทนำ

โปรตีนเป็นสารที่เป็นลวณประกอบต่างๆ ภายในเซลล์ของสิ่งมีชีวิต โปรตีนทำหน้าที่ เป็นโครงสร้างของเซลล์ เป็นตัวเร่งปฏิกิริยา เป็นตัวขนส่ง เป็นตัวควบคุมอัตราการเกิด เมตาโบลิซึมและทำหน้าที่อื่นๆอีกมากมาย ดังนั้นโปรตีนจึงเป็นสารประกอบอินทรีย์ที่มีความสำคัญอย่างยิ่ง

โปรตีนประกอบด้วยหน่วยย่อยที่เรียกว่า กรดอะมิโน (amino acid) ซึ่งมีอยู่ 20 ชนิด เชื่อมต่อกันด้วยพันธะเปปไทด์ (peptide bonds) โดยลำดับของกรดอะมิโนที่ใช้ในการสร้างโปรตีนจะถูกเก็บเป็นรหัสบนดีเอ็นเอ

ในการศึกษาโปรตีน นักวิทยาศาสตร์จะแยกสกัดโปรตีนจากสิ่งมีชีวิตและนำมาศึกษาจนทราบว่าโปรตีนตัวนั้นทำงานและมีสมบัติอย่างไร จากนั้นจึงนำโปรตีนไปหา ลำดับกรดอะมิโน แต่ในปัจจุบันด้วยความรู้ทางพันธุวิศวกรรม และชีวเคมี ทำให้ทราบ รหัสบนดีเอ็นเอ และสามารถแปลรหัสเหล่านั้นจนได้ลำดับของกรดอะมิโนของโปรตีน มากมาย และด้วยโครงการอ่านรหัสบนสารพันธุกรรมของมนุษย์ (Human Genome Project) และสิ่งมีชีวิตต่างๆ อีกหลายชนิดส่งผลให้ฐานข้อมูลของลำดับของกรดอะมิโน มีขนาดใหญ่ขึ้นอย่างรวดเร็ว ดังนั้นนักวิทยาศาสตร์จึงได้หันมาสนใจศึกษาโปรตีนจาก ลำดับของกรดอะมิโนแทน โดยหลักการที่ว่าโปรตีนที่ฮ่อมอโลยีกันจะมีหน้าที่และโครงสร้างคล้ายกัน ดังนั้นในการศึกษาโปรตีนจากลำดับของกรดอะมิโน จะทำโดยการนำ ลำดับกรดอะมิโนของโปรตีนที่ไม่ทราบโครงสร้างและหน้าที่ไปทดสอบฮ่อมอโลยีกับ โปรตีนที่ทราบโครงสร้างและหน้าที่แล้ว ถ้าโปรตีน 2 ชนิดนั้นมีฮ่อมอโลยี ก็น่าจะมีโครงสร้างและทำหน้าที่คล้ายกันหรือเหมือนกัน

การทดสอบฮ่อมอโลยีของโปรตีนจากลำดับของกรดอะมิโนที่ได้จากการอ่านรหัส บนสารพันธุกรรมยังมีข้อจำกัดอยู่มาก วิธีการที่นิยมใช้และเป็นมาตรฐานในปัจจุบันจะ ใช้วิธีการทดสอบความคล้ายกันของลำดับของกรดอะมิโน (sequence alignment) ดังนั้นจึงมีการศึกษาหาวิธีทดสอบความคล้ายกันของลำดับของกรดอะมิโนในโปรตีนมาก

มายหลายวิธีไม่ว่าจะเป็นวิธี maximum matching (Needleman., 1969), hidden markov models (Kevin., et al ,1999), Basic Local Alignment Search Tool (BLAST) (Karlin and Altschul,1990-1993) เป็นต้น นอกจากนี้ยังมีวิธีการ Hydrophobic cluster analysis (Karlin and Altschul., 1987) :ซึ่งสามารถใช้ได้ทั้งการทำนายโครงสร้างทุติยภูมิ และทดสอบความคล้ายกันของลำดับของกรดอะมิโน

แต่การศึกษาโปรตีนจากความคล้ายลำดับของกรดอะมิโนยังประสบปัญหาต่างๆ อันเกิดจากธรรมชาติของโปรตีนอันได้แก่เป็นการเว้นช่วง (gap) ของลำดับของกรดอะมิโนเพราะกรดอะมิโนในโปรตีนจะขดตัวเป็นโครงสร้างมีลักษณะเป็น domain หรือ cassette และเชื่อมต่อกัน ดังนั้นกรดอะมิโนที่ทำหน้าที่เป็นส่วนเชื่อมต่อกันจึงมีจำนวนต่างกันได้โดยไม่รบกวนการทำงานของโปรตีน การเปลี่ยนแปลงของกรดอะมิโนบางตัวในลำดับโดยไม่เปลี่ยนแปลงการทำงานของโปรตีนเนื่องจากโปรตีนที่มีอยู่ในสิ่งมีชีวิตที่ต่างกันจึงทำให้ลำดับของกรดอะมิโนต่างกันบ้างตามลำดับวิวัฒนาการที่ต่างกันของสิ่งมีชีวิต หรือความแตกต่างกันของขนาดโปรตีนและลำดับของกรดอะมิโนทำให้ความคล้ายกันของลำดับของกรดอะมิโนมีน้อยมาก ซึ่งจะทำให้ไม่สามารถทดสอบหาลำดับจากความคล้ายกันของลำดับของกรดอะมิโนได้ งานวิจัยนี้จึงได้เปลี่ยนลำดับของกรดอะมิโนให้อยู่ในรูปแผนภูมิ 2 มิติและเชื่อว่าแผนภูมิ 2 มิติจะเป็นเครื่องมือในการทดสอบหาลำดับของโปรตีนและจะลดปัญหาการทดสอบหาลำดับของโปรตีนในเรื่องของช่องว่างและการที่ลำดับของกรดอะมิโนที่นำมาทดสอบมี sequence identity ต่ำได้

การวิเคราะห์ลำดับของกรดอะมิโนโดยปรับข้อมูลให้อยู่ในรูปแผนภูมิ 2 มิติเป็นเทคนิคที่มีผู้นำมาใช้ได้อย่างได้ผลในการทำนายโครงสร้างทุติยภูมิ (secondary structure) ของโปรตีน โครงการนี้ได้รับเอาแนวคิดนี้และจะมาผนวกกับข้อมูลทางเคมี และชีวเคมี เช่น hydrophobicity, charge density, pK_a of Ionizing side chain หรือ surface area ของกรดอะมิโนแต่ละตัวแล้ว สร้างแผนภูมิ 2 มิติใหม่ที่มีศักยภาพสูงในการเปรียบเทียบหาลำดับของโปรตีน ด้วยแนวคิดที่ว่าโปรตีนหรือชิ้นส่วนของโปรตีน 2 ชนิดที่

ขอมอโลกัสกันจะให้แผนภูมิ 2 มิติที่เหมือนกันหรือคล้ายกัน และสร้างโครงข่ายประสาทเทียม (neural network) ช่วยในการพิจารณาความเหมือนกันหรือคล้ายกันของแผนภูมิ 2 มิติใหม่ที่สร้างขึ้น

โครงข่ายประสาทเทียม (neural networks) เป็นสาขาหนึ่งของ Artificial Intelligence ที่พยายามเลียนแบบการทำงานของสมองมนุษย์ ซึ่งในปัจจุบันมีการนำมาใช้อย่างกว้างขวางในงานหลายๆ ด้านเช่น งานรู้จำรูปแบบ (pattern recognition) การควบคุมเครื่องจักร การจำเสียงพูด การทำนายทางธุรกิจ การพยากรณ์อากาศ เป็นต้น

สมบัติของโครงข่ายประสาทเทียมที่น่าสนใจคือ มีความยืดหยุ่นสูงจนสามารถจำลองขบวนการของปัญหาได้ สามารถจำชุดของ input-output ที่มีความซับซ้อนมากๆ สามารถตอบสนองกับข้อมูลที่ไม่เคยเห็นได้ และสามารถประเมินผล output ได้อย่างรวดเร็ว และที่สำคัญไม่จำเป็นต้องทราบวิธีการในการที่จะเปลี่ยน input ให้กลายเป็น output เพียงแต่ให้ตัวอย่างของชุด input และ output ที่ถูกต้องเพื่อการ train โครงข่ายก็เพียงพอแล้ว ดังนั้นโครงข่ายประสาทเทียมจึงเป็นวิธีการหนึ่งที่เหมาะสมที่จะช่วยในการเปรียบเทียบความเหมือนหรือต่างกันของแผนภูมิ 2 มิติของกรดอะมิโนที่สร้างขึ้น โดยในโครงการนี้จะใช้โครงข่ายประสาทเทียมที่มีการเรียนรู้ (learning function) แบบ backpropagation (Rumelhart., et al., 1986) ซึ่งเป็นที่นิยมใช้ในการแก้ปัญหาต่างๆ ในปัจจุบัน และใช้โปรแกรมสำเร็จ SNNS (Stuttgart Neural Networks Simulation) ช่วยในการสร้างโครงข่ายประสาทเทียม

สำหรับงานวิจัยนี้จะสร้างแผนภูมิ 2 มิติของกรดอะมิโนเพื่อช่วยในการพิจารณาขอมอโลยีระหว่างโปรตีนโดยสมมติฐานที่ว่าโปรตีนที่ขอมอโลกัสกันจะให้แผนภูมิ 2 มิติของกรดอะมิโนที่คล้ายกันหรือเหมือนกัน และแก้ปัญหาการทดสอบขอมอโลยีของโปรตีนในเรื่องของช่องว่างและ sequence identity ต่ำได้และทำการสร้างโครงข่ายประสาทเทียม เพื่อช่วยในการเปรียบเทียบชั้นของแผนภูมิที่ได้ เพื่อนำไปเป็นเครื่องมือในการพิจารณาขอมอโลยีระหว่างโปรตีน 2 ชนิด และวางแนวทางพัฒนาโปรแกรมเพื่อให้พิจารณาความเหมือนหรือคล้ายกันระหว่างโปรตีนมากกว่า 2 ชนิด (multiple alignment)ต่อไป เพื่อใช้ได้กับฐานข้อมูลขนาดใหญ่ในอนาคต