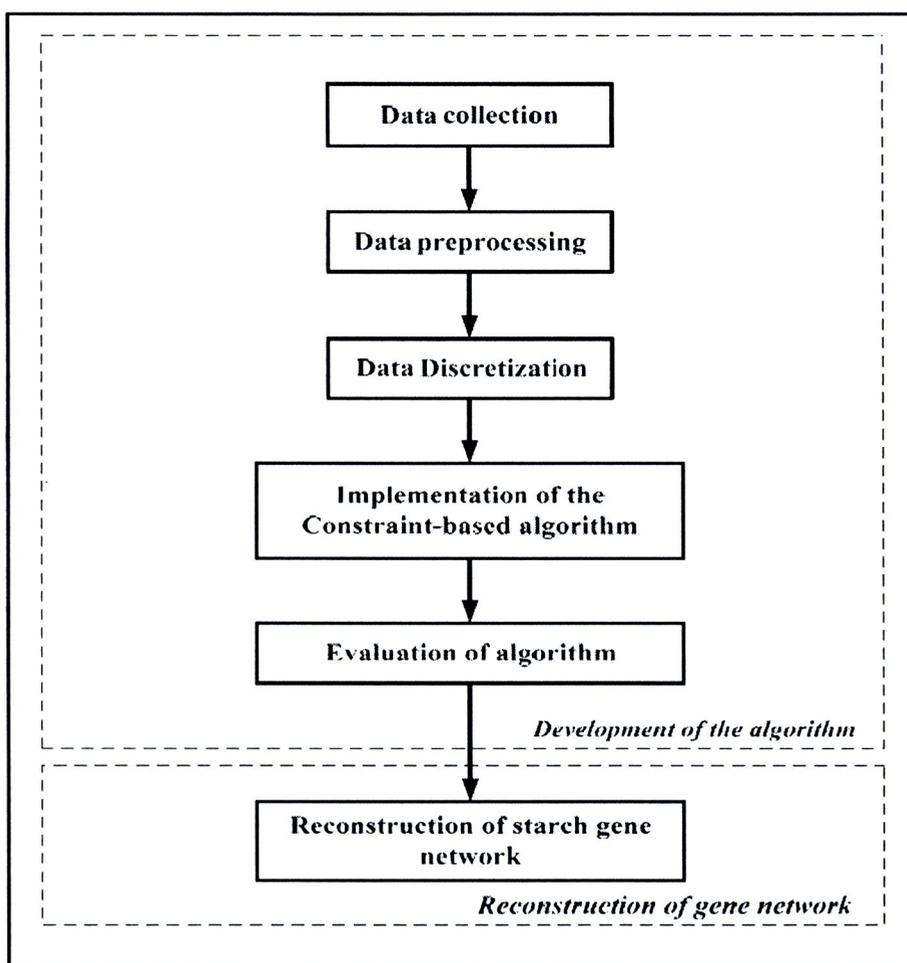


บทที่ 2 ระเบียบวิธีวิจัย

การศึกษาวิจัยแบ่งออกเป็น 6 ขั้นตอนหลัก ดังแสดงในรูปที่ 2.1 ซึ่งมีรายละเอียดดังต่อไปนี้

2.1 การเก็บและรวบรวมข้อมูล (Data collection)

การทดลองนี้จะเก็บรวบรวมข้อมูลไมโครอะเรย์ 2 แบบ คือ cDNA microarrays ของยีสต์ และ oligonucleotide microarray (Affymetrix GeneChips) จากฐานข้อมูล Stanford Microarray และ NASC's International Affymetrix Service (NASCArrays) ของพืช *Arabidopsis thaliana*



รูปที่ 2.1 วิธีการดำเนินวิจัยโดยสังเขป

2.2 การเตรียมข้อมูล (Data preprocessing)

ก่อนที่จะนำข้อมูลการแสดงออกของยีนไปใช้นั้น ต้องมีการเตรียมข้อมูลให้เหมาะสม รวมทั้งมีการปรับค่าการแสดงออกของยีนให้อยู่ในมาตรฐานเดียวกันตามเทคโนโลยีไมโครอะเรย์ที่ใช้ เพื่อสามารถนำมาเปรียบเทียบหรือวิเคราะห์ในขั้นตอนต่อไป ข้อมูลที่ใช้ในงานวิจัยนี้มี 2 แบบ จาก cDNA

microarrays และ oligonucleotide microarray (Affymetrix GeneChips) การเตรียมข้อมูลก็จะแตกต่างกัน ข้อมูลการแสดงออกของยีนที่ผ่านกระบวนการเตรียมข้อมูลแล้วนั้นจะแปลงเป็น 2 ระดับ (binary values) คือ เป็น 0 หรือ 1 สำหรับวิเคราะห์ด้วยเทคนิคบูลีนต่อไป

2.3 ประเมินวิธีการที่ใช้ในการแปลงข้อมูลการแสดงออกของยีน (Data Discretization)

ขั้นตอนที่สำคัญในการวิเคราะห์เครือข่ายควบคุมการแสดงออกของยีนด้วยเทคนิคบูลีน คือ การแปลงค่าการแสดงออกของยีนเป็นข้อมูล 0 และ 1 เพื่อใช้ในการสร้างฟังก์ชันบูลีน (Boolean function) ในการหาความสัมพันธ์ระหว่างคู่ยีนว่าเป็นความสัมพันธ์แบบกระตุ้น (activation) หรือยับยั้ง (inhibition) ซึ่งวิธีการทางสถิติที่ใช้ในการแปลงข้อมูลเป็น 0 หรือ 1 นั้นมีหลายแบบ จึงควรมีการประเมินหาวิธีการที่เหมาะสมใช้ในการแปลงข้อมูลการแสดงออกของยีน เพื่อให้ได้เครือข่ายควบคุมการแสดงออกของยีนที่มีความสอดคล้องกับระบบของสิ่งมีชีวิต โดยจะประเมินวิธีการด้วยการเปรียบเทียบค่าความเหมือน (Similarity score) ระหว่างเดนโดแกรม (dendrogram) ของ raw data และ binary values ว่ามีความเหมือนหรือต่างกันเพียงใด โดยเดนโดแกรมนั้นสร้างจากวิธีการ hierarchical clustering วิธีการที่ให้ค่า similarity score ระหว่าง dendrogram ของ raw data และ binary values สูงสุด ควรจะเป็นวิธีการที่ดีที่สุดที่จะใช้ในการแปลงข้อมูล และสามารถช่วยสร้างเครือข่ายควบคุมการแสดงออกของยีนด้วยเทคนิคบูลีนที่สอดคล้องกับข้อมูลทางชีววิทยามากที่สุด [13]

2.4 พัฒนาโปรแกรมที่ใช้ในการสร้างเครือข่ายการแสดงออกของยีนแบบมีเงื่อนไข

(Constraint-based Boolean network)

ด้วยข้อจำกัดของเทคนิคบูลีนที่อาจสร้างเครือข่ายควบคุมการแสดงออกของยีนที่มีความผิดพลาด เพราะสร้างมาจากข้อมูลเพียง 2 ระดับ คือ 0 และ 1 ดังนั้น นอกจากการพิจารณาถึงความสำคัญของ discretization method ก่อนการวิเคราะห์ด้วยเทคนิคบูลีนแล้ว ควรปรับปรุงเทคนิคบูลีนให้สามารถสร้างเครือข่ายควบคุมการแสดงออกของยีนที่มีความซับซ้อนน้อยลง ลดความผิดพลาดในการทำนาย การใส่เงื่อนไขทางชีววิทยาให้กับการวิเคราะห์ด้วยเทคนิคบูลีน น่าจะลดความซับซ้อนและลดการทำนายที่ผิดพลาดที่มักเกิดขึ้นจากการวิเคราะห์ด้วยวิธีนี้ ทำให้ง่ายต่อการศึกษาและวิเคราะห์เครือข่ายที่สร้างขึ้นได้ต่อไป ดังนั้น งานวิจัยนี้จึงพัฒนาวิธีการที่ใช้สร้างเครือข่ายการแสดงออกของยีนแบบมีเงื่อนไข เรียกว่า Constraint-based Boolean network ด้วยภาษา C++ ซึ่งจะใช้อุณหภูมิในการพัฒนาโปรแกรม

2.5 การทดสอบและประเมินโปรแกรม (Model validation)

หลังจากการพัฒนาโปรแกรมคอมพิวเตอร์เพื่อสร้างเครือข่ายควบคุมการแสดงออกของยีนแบบมีเงื่อนไข โปรแกรมที่พัฒนาขึ้นนั้นจำเป็นต้องมีการทดสอบเพื่อวัดประสิทธิภาพของโปรแกรมที่พัฒนาขึ้น โดยการเปรียบเทียบเครือข่ายควบคุมการแสดงออกของยีนอ้างอิง [14] ใน galactose pathway และเครือข่ายควบคุมการแสดงออกของยีนที่สร้างขึ้นจาก Constraint-based Boolean network ด้วยค่าสถิติ เช่น ค่าความถูกต้อง (accuracy) ค่าความแม่นยำ (precision) เป็นต้น

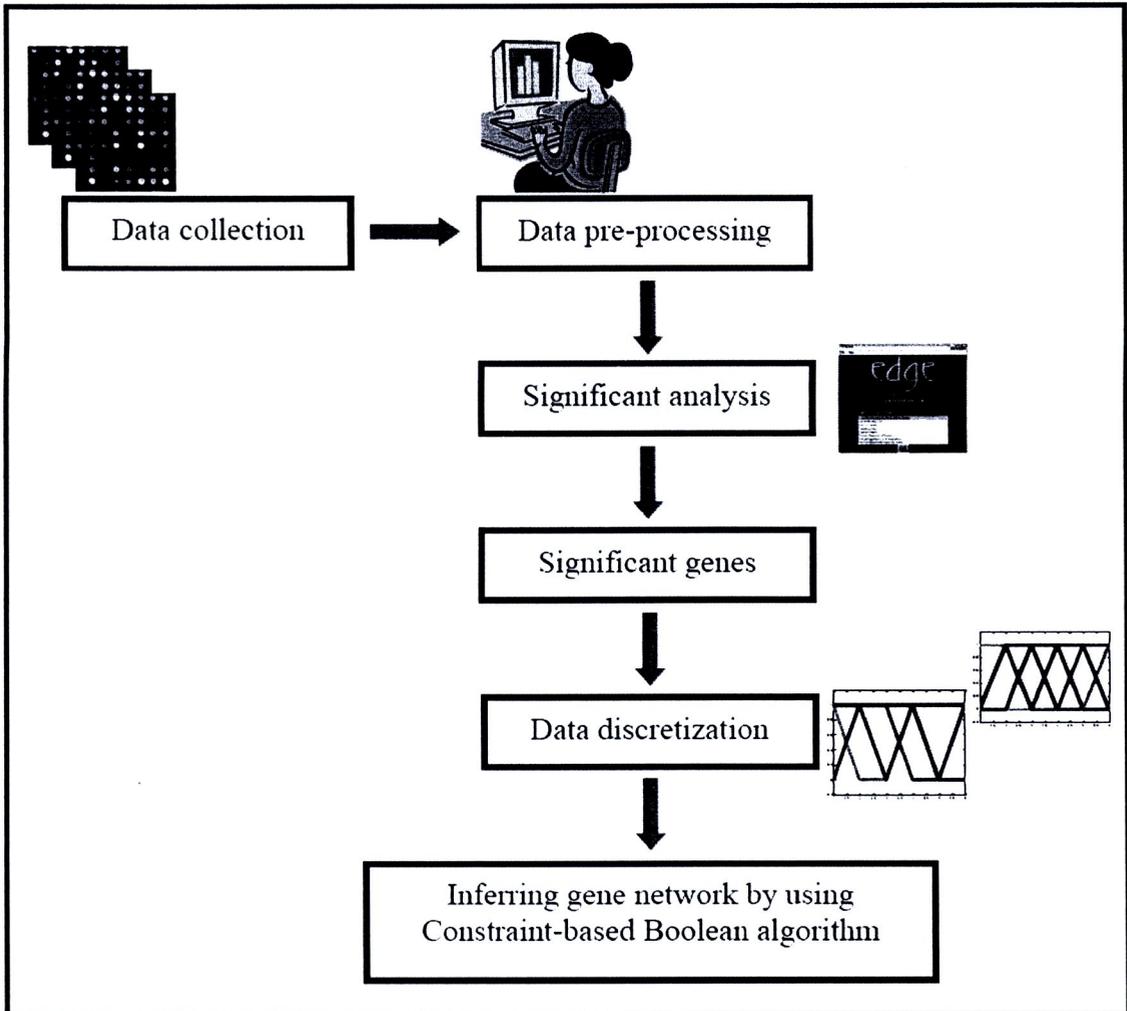
2.6 การประยุกต์ใช้โปรแกรมเพื่อสร้างและศึกษาเครือข่ายควบคุมการแสดงออกของยีนใน

กระบวนการสังเคราะห์แป้ง (Case study for starch application)

ในขั้นการทดลองนี้ จะนำเอาโปรแกรมที่พัฒนาขึ้นมาใช้สร้างเครือข่ายควบคุมการทำงานของยีนในกระบวนการสังเคราะห์แป้ง เพื่อทำความเข้าใจเกี่ยวกับกลไกการทำงานของกระบวนการสังเคราะห์แป้งมากขึ้น โดยเลือกเอาข้อมูลการแสดงออกของยีนในพืชต้นแบบ (plant model) นั่นคือ *Arabidopsis thaliana* ซึ่งข้อมูลการแสดงออกของยีนต้องผ่านกระบวนการ pre-processing และหา



ยีนที่มีการแสดงออกอย่างมีนัยสำคัญ (Significant genes) และหายีนที่เกี่ยวข้องกับกระบวนการสังเคราะห์แป้งโดยใช้ข้อมูลยีนในงานวิจัยของ [14] ข้อมูลที่ผ่านการ discretization จะนำไปสร้างเครือข่ายควบคุมการทำงานของยีนในกระบวนการสังเคราะห์แป้งด้วย Constraint-based Boolean network



รูปที่ 2.2 ขั้นตอนการวิเคราะห์เครือข่ายควบคุมการแสดงออกของยีนในกระบวนการสังเคราะห์แป้ง