บทที่ 4
การอนุมานแบบจำลองการควบคุมกันระหว่างยีนด้วยเครือข่ายเบย์เซียน โดยอาศัยวิธีการมาร์คอฟเชนมอนติคาร์โล

ยีนของเซลล์สิ่งมีชีวิตจะมีหน้าที่ที่เฉพาะเจาะจงของมันภายในกระบวนการทำงานของเซลล์ การแสดงออกของยีนจะถูกควบคุมจากปัจจัยต่างๆ เพื่อทำการสำเนารหัสพันธุกรรม (Transcribe) เป็นอาร์เอ็นเอ (RNA) จากนั้นจึงเข้าสู่กระบวนการแปลรหัสพันธุกรรม (Translate) เป็นโปรตีนที่ทำหน้าที่ต่าง ๆ โดยเฉพาะ (Alberts, 2002) ระดับการแสดงออกของยีนสามารถวัดออกมาได้ด้วยเทคโนโลยีดีเอ็นเอไมโครอาร์เรย์ ที่เป็นตัวบ่งบอกว่ายีนมีการตอบสนองต่อสิ่งแวดล้อมมากน้อยเพียงใด โดยการแสดงออกของยีนนี้จะถูกวัดออกมาเพื่อทำการศึกษาภายใต้สภาวะหนึ่ง ๆ ซึ่งในการแสดงออกของยีนนี้จะต้องมีกลุ่มของโปรตีนเข้ามาเกี่ยวข้องด้วยในลักษณะของการควบคุมกันระหว่างยีนในระหว่างกระบวนการแสดงออกของยีน ได้แก่ ทรานสคริปท์ชัน-นอลแฟกเตอร์ แอกทิเวเตอร์ (Activators) และโคแอกทิเวเตอร์ (Co– Activators) เป็นต้น ในทางชีวภาพกลไกการควบคุมกันระหว่างยีนนั้นมีความสำคัญ และเป็นวิธีการศึกษาพื้นฐานในการที่จะนำไปสู่การเข้าใจพฤติกรรมของยีนที่มีความซับซ้อน

การศึกษาโดยทั่วไปที่ไม่ได้อาศัยความรู้ทางด้านชีววิทยาเข้ามาเกี่ยวข้อง เช่น การวิเคราะห์ข้อมูลดีเอ็นเอไมโครอาร์เรย์ด้วยเทคนิควิธีการแบ่งกลุ่มข้อมูล (Cluster analysis) ตามลักษณะการแสดงออกของยีนมักจะถูกนำมาใช้ในการอนุมานหากลุ่มยีนที่ถูกควบคุมด้วยปัจจัยต่างๆ ร่วมกัน (Dirisi *et al.*, 1997) แต่อย่างไรก็ตามเป็นการยากที่จะทำการพิสูจน์ความถูกต้องของผลลัพธ์ที่ได้จากการแบ่งกลุ่มข้อมูลนี้ ดังนั้นวิธีการอนุมานทางสถิติหลายๆ วิธีการจึงถูกนำเสนอขึ้นมาเพื่อทำการอนุมานเครือข่ายการควบคุมกันระหว่างยีน ได้แก่ โมเดลเกาส์เซียนเชิงกราฟ (Graphical Gaussian Model) (Wu and Subramanian, 2003) เครือข่ายความน่าจะเป็นแบบ บูลีน (Probabilistic Boolean Networks) (Shmulevith *et al.*, 2002) และเครือข่ายเบย์เซียน (Bayesian Networks) (Hartemink, 2001 and Husmeier, 2003) ซึ่งด้วยวิธีการเหล่านี้อาจไม่สามารถทำการพิสูจน์ได้ว่าผลลัพธ์ที่ได้มีความถูกต้องหรือสมบูรณ์ครบถ้วนในทางชีววิทยา

ในการอนุมานเครือข่ายการควบคุมกันระหว่างยีนด้วยเครือข่ายเบย์เซียน ซึ่งอาจต้องมีการสร้างเครือข่ายที่ต้องการเปรียบเทียบขึ้นเอง หรือทำการหาเครือข่ายที่เป็นไปได้ทั้งหมดไว้ก่อน แล้วจึงทำการพิจารณาเครือข่ายที่ดีที่สุดโดยใช้วิธีการคิดคะแนนให้กับแต่ละเครือข่าย (Hertemink *et al.*, 2001) แต่ในทางปฏิบัติข้อมูลไมโครอาร์เรย์สามารถถูกสร้างออกมาได้เป็นจำนวนน้อยเมื่อ

เทียบกับจำนวนยีนที่กำลังศึกษา ซึ่งการวิเคราะห์ด้วยข้อมูลการแสดงออกของยีนที่มีอยู่เป็นจำนวนน้อยนี้อาจทำให้ได้ผลการวิเคราะห์ที่มีความน่าเชื่อถือต่ำ (Husmeier, 2003)

ดังนั้นในงานวิจัยนี้จึงได้นำวิธีการหนึ่งของมาร์คอฟเชนมอนติคาร์โล (Markov Chain Monte Carlo: MCMC) นั่นก็คือ อัลกอริทึมเมโทรโพลิส – แฮสติงส์ (Metropolis – Hastings algorithm) มาทำงานร่วมกับเครือข่ายเบย์เซียนในการสร้างแบบจำลองที่เป็นไปได้ของเครือข่ายการควบคุมกันระหว่างยีน โดยเครือข่ายการควบคุมระหว่างยีนที่มีการรายงานไว้แล้วในทางชีววิทยาจะถูกกำหนดเป็นโมเดลเริ่มต้นให้กับเครือข่ายเบย์เซียน จากนั้นความสัมพันธ์ระหว่างสองยีนในเครือข่ายใหม่จะถูกสร้างขึ้นมาแบบสุ่มด้วยการแจกแจงทวินาม (Binomial distribution) โดยที่ค่าความน่าจะเป็นของความสัมพันธ์ระหว่างยีนแต่ละคู่จะสร้างมาจากการแจกแจง Dirichlet (Dirichlet distriubtion) อีกทีหนึ่ง โดยพารามิเตอร์ของการแจกแจง Dirichlet นี้จะเป็นจำนวนความถี่ของความสัมพันธ์ที่เกิดขึ้นในเครือข่ายทั้งหมดที่ยอมรับไว้ก่อนหน้า จากนั้นจึงทำการคำนวณความน่าจะเป็นที่ยอมรับได้ด้วยอัลกอริทึมเมโทรโพลิส – แฮสติงส์ สำหรับใช้ในการพิจารณายอมรับเครือข่ายใหม่ที่ถูกสร้างขึ้นมา ความน่าจะเป็นที่ยอมรับได้นี้ประกอบด้วยเทอมการแจกแจงที่นำเสนอ และการแจกแจงที่คงที่ที่อยู่ในรูปแบบของการแจกแจงโพสทีเรีย (Posterior distribution) ของเครือข่ายก่อนหน้าและเครือข่ายที่ถูกสร้างขึ้นมาใหม่ โดยการแจกแจงที่นำเสนอที่ใช้ในงานวิจัยนี้จะเหมือนกับที่ Husmeier (2003) ได้เสนอไว้ ส่วนการแจกแจงโพสทีเรียจะคำนวณมาจากค่าความน่าจะเป็นที่มีเงื่อนไข (Conditional distribution) ของข้อมูลการแสดงออกของยีนที่ถูกแปลงเป็นข้อมูลชนิดไม่ต่อเนื่องแล้ว และงานวิจัยนี้ได้ใช้ชุดข้อมูลตัวอย่างในการทดลอง เป็นการแสดงออกของยีน CYC1 ที่ถูกควบคุมร่วมด้วย HAP2 HAP3 HAP4 ซึ่งเป็นข้อมูลยีนของยีสต์แซคคาโรไมซิส เซริวิซิเย

4.1 ชุดข้อมูลการทดลอง

การแสดงออกของยีน CYC1 (Cytochrome c, isoform 1) ซึ่งถูกควบคุมร่วมด้วยโค-แอกทิเวเตอร์ HAP2 HAP3 HAP4 จะถูกนำมาศึกษาในงานวิจัยนี้ ยีน CYC1 จะทำหน้าที่เป็นตัวนำอิเล็กตรอน (Electron carrier) ในผนังเซลล์ชั้นในของไมโตคอนเดรีย (Mitochondrial intermembrance space) ที่ส่งอิเล็กตรอนจาก ubiquinone-cytochrome c oxidoreductase ไปยัง cytochrome c oxidase ในระหว่างกระบวนการหายใจของเซลล์ของยีสต์แซคคาโรไมซิส เซริวิซิเย (Hortner *et al.*, 1982) ซึ่งได้มีรายงานไว้แล้วในงานวิจัยว่ายีนทั้งสี่ตัวนี้มีความเกี่ยวข้องกันโดยกลุ่มโปรตีน HAP2/3/4 ต่างไปร่วมกันควบคุมการแสดงออกของยีน CYC1 โดยที่ HAP2

และ HAP3 เป็นโปรตีนควบคุมที่เข้าไปจับกับส่วนโปรโมเตอร์ที่เฉพาะเจาะจงของยีน CYC1 และโปรตีน HAP4 จะเป็นตัวกระตุ้นเพื่อให้ยีน CYC1 เริ่มทำงานได้ (Olesen and Guarente, 1990)

ข้อมูลการแสดงออกของยีนในทั้งสี่นี้เป็นข้อมูลชนิดต่อเนื่องแสดงดังตาราง 4.1 ซึ่งในการศึกษาครั้งนี้ข้อมูลการแสดงออกยีนชนิดต่อเนื่องนี้จะถูกแปลงเป็นข้อมูลชนิดไม่ต่อเนื่องก่อนที่จะนำไปทำการทดลองดังตาราง 4.2 สำหรับเงื่อนไขที่ใช้ในการพิจารณาแปลงข้อมูลการแสดงออกของยีนจากข้อมูลชนิดต่อเนื่องให้เป็นข้อมูลชนิดไม่ต่อเนื่องจะพิจารณาจากการแสดงออกที่เพิ่มขึ้น ลดลง หรือคงที่ เมื่อทำการเปรียบเทียบกับค่าคงที่ที่กำหนดไว้ ดังนี้

เมื่อกำหนดให้ x คือข้อมูล เรากำหนดเงื่อนไขดังต่อไปนี้ คือ

- ถ้า x > 0.2 หมายถึง ยีนมีการแสดงออกเพิ่มขึ้น แทนด้วยเครื่องหมาย ↑
- ถ้า -0.2 <= x <= 0.2 หมายถึง ยีนมีการแสดงออกคงที่ แทนด้วยเครื่องหมาย – และ
- ถ้า x < -0.2 หมายถึง ยีนมีการแสดงออกลดลง แทนด้วยเครื่องหมาย ↓

ในการศึกษาครั้งนี้ได้เลือกใช้ค่า 0.2 และ -0.2 เป็นขอบเขตของข้อมูลที่ใช้ในการแปลงข้อมูลการแสดงของยีนจากข้อมูลชนิดต่อเนื่องให้เป็นข้อมูลชนิดไม่ต่อเนื่อง ซึ่งเป็นผลจากการสังเกตชุดข้อมูลการแสดงออกของทั้งสี่นี้ แล้วพบว่าค่าที่เลือกใช้นี้สามารถแบ่งลักษณะการแสดงออกที่แตกต่างกันออกเป็นการแสดงออกที่เพิ่มขึ้น ลดลง และคงที่ได้

ตาราง 4.1 ข้อมูลการแสดงออกของยีน CYC1, HAP2, HAP3 และ HAP4

CYC1	HAP2	HAP3	HAP4
-0.14	-0.15	-0.29	0.24
-0.22	0.03	0.24	0.58
-0.01	-0.12	0.12	1.04
-0.09	-0.32	-0.14	0.66
0.31	-0.23	-0.36	0.62
1.18	-0.04	-0.15	2.54
1.75	0.51	-0.09	3.13

ตาราง 4.2 ระดับการแสดงออกของยีน CYC1, HAP2, HAP3 และ HAP4 ที่ผ่านการแปลงให้เป็นข้อมูลชนิดไม่ต่อเนื่องแล้ว

CYC1	HAP2	HAP3	HAP4
–	–	↓	↑
↓	–	↑	↑
–	–	–	↑
–	↓	–	↑
↑	↓	↓	↑
↑	–	–	↑
↑	↑	–	↑

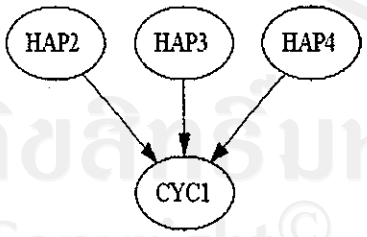
ยกตัวอย่างเช่นข้อมูลในคอลัมน์ที่ 1 แถวที่ 1 ของตาราง 4.1 จะได้ว่า x = -0.14 ตรงตามเงื่อนไขที่สองคือ -0.2 <= x <= 0.2 จึงได้ว่า x มีการแสดงออกคงที่ (–) เป็นต้น

4.2 วิธีการและการทดลอง

ในการศึกษาครั้งนี้ได้นำวิธีการของอัลกอริทึมเมโทรโพลิส — แฮสติงส์ ซึ่งเป็นวิธีการหนึ่งของวิธีการมาร์คอฟเชนมอนติคาร์โล มาผนวกกับเครือข่ายเบย์เซียน เพื่อทำการสร้างเครือข่ายการควบคุมกันระหว่างยีนของกลุ่มยีนตัวอย่าง ตามหลักการของเครือข่ายเบย์เซียนโครงสร้างของเครือข่าย (Network structure) จะเป็นแบบจำลองเชิงกราฟที่มีความสัมพันธ์ระหว่างโหนด (node) แบบระบุทิศทางและไม่มีลักษณะเป็นวงจร (Directed Acyclic Graph) หรือที่เรียกว่า DAG โดยที่แต่ละโหนดของเครือข่ายจะเป็นตัวแทนของแต่ละยีนในเครือข่าย และเส้นเชื่อม (edge) แทนความสัมพันธ์ที่บ่งบอกถึงการควบคุมกันระหว่างยีน ในงานวิจัยนี้ได้ใช้เมทริกซ์เครือข่าย (Network matrix) ขนาด n×n (n = จำนวนยีนในเครือข่าย) เป็นตัวแทนแสดงความสัมพันธ์ของการควบคุมกันระหว่างยีนต่าง ๆ ในเครือข่าย โดยกำหนดค่าความสัมพันธ์ระหว่างคู่ยีนในเครือข่ายดังต่อไปนี้

$$M[i,j] = \begin{cases} 0 & \text{หมายความว่า ไม่มีเส้นเชื่อม หรือไม่มีความสัมพันธ์ระหว่างยีนแม่ตัวที่ } i \text{ และยีนลูกตัวที่ } j \\ 1 & \text{หมายความว่า มีเส้นเชื่อม หรือมีความสัมพันธ์ระหว่างยีนแม่ตัวที่ } i \text{ และยีนลูกตัวที่ } j \end{cases}$$

(ก)



(ข)

Parent node

	CYC1	HAP2	HAP3	HAP4
CYC1	0	0	0	0
HAP2	1	0	0	0
HAP3	1	0	0	0
HAP4	1	0	0	0

Child node

รูป 4.1 เครือข่ายเบย์เซียนแสดงการควบคุมกันระหว่างยีน CYC1 HAP2 HAP3 และ HAP4 โดย (ก) แบบจำลองเชิงกราฟ และ (ข) เมทริกซ์เครือข่ายของแบบจำลอง

สามารถแสดงตัวอย่างของเครือข่ายการควบคุมกันที่มีการรายงานเอาไว้แล้วของกลุ่มยีนตัวอย่าง CYC1 HAP2 HAP3 และ HAP4 ที่นำมาศึกษาในครั้งนี้ได้ดังรูป 4.1(ก) และใช้เมทริกซ์เครือข่ายเป็นตัวแทนแบบจำลองเชิงกราฟแสดงการควบคุมกันระหว่างยีนได้ดังรูป 4.1(ข)

โดยยีนในแนวแถวของเมทริกซ์เครือข่ายจะแทนยีนที่เป็นโหนดพ่อแม่ (parent nodes) หรือตัวควบคุม (Regulators) และยีนในแนวคอลัมน์จะแทนยีนที่เป็นโหนดลูก (child node) หรือยีนที่ถูกควบคุม (Regulated genes)

4.2.1 อัลกอริทึมเมโทรโพลิส – แฮสติงส์

งานวิจัยนี้จะอาศัยหลักการของอัลกอริทึมเมโทรโพลิส – แฮสติงส์ (Gilks, 1996) ในการเรียนรู้เครือข่ายการควบคุมกันระหว่างยีนที่ถูกสร้างขึ้นมาแบบสุ่มเพื่อทำการสร้างเครือข่ายการควบคุมกันระหว่างกลุ่มยีนที่เราสนใจ โดยในตอนเริ่มต้นของการทดลองจะทำการกำหนดเครือข่ายความสัมพันธ์ระหว่างยีนเริ่มต้น M_{old} ด้วยเครือข่ายจริงทางชีววิทยาของยีนทั้งสี่ดังรูป 4.1 หลังจากนั้นอัลกอริทึมเมโทรโพลิส – แฮสติงส์จะทำการคำนวณแบบวนซ้ำเพื่อสร้างเครือข่ายการควบคุมกันระหว่างยีนแบบสุ่มขึ้นมาเป็นเครือข่ายใหม่ M_{new}

ค่าความสัมพันธ์ของเส้นเชื่อมที่แต่ละตำแหน่งของเมทริกซ์เครือข่ายใหม่นี้จะถูกสร้างขึ้นแบบสุ่มด้วยการแจกแจงทวินาม (Binomial distribution) (Milton and Tsokos, 1983) แสดงดังสมการที่ (17)

$$M_{new}[i,j] \sim binomial(edge\_prob[i,j]) \qquad (17)$$

โดย M_{new}[i,j] คือค่าความสัมพันธ์ระหว่างตัวควบคุมตัวที่ i และยีนที่ถูกควบคุมตัวที่ j ในเมทริกซ์เครือข่ายใหม่ โดยค่าที่ได้จากสุ่มจะเป็น 0 หรือ 1 เพียงค่าเดียวเท่านั้น พารามิเตอร์ *edge_prob*[i,j] ในสมการที่ (17) คือค่าความน่าจะเป็นของการสุ่ม (Binormial success probability) สำหรับสร้างเส้นเชื่อมที่ตำแหน่ง (i,j) ในเมทริกซ์เครือข่ายใหม่ M_{new}

สำหรับค่าของพารามิเตอร์ *edge_prob* นี้จะถูกสุ่มมาจากการแจกแจง Dirichlet (Regazzini, 2002) อีกทีหนึ่ง แสดงการแจกแจง Dirichlet ได้ดังสมการที่ (18)

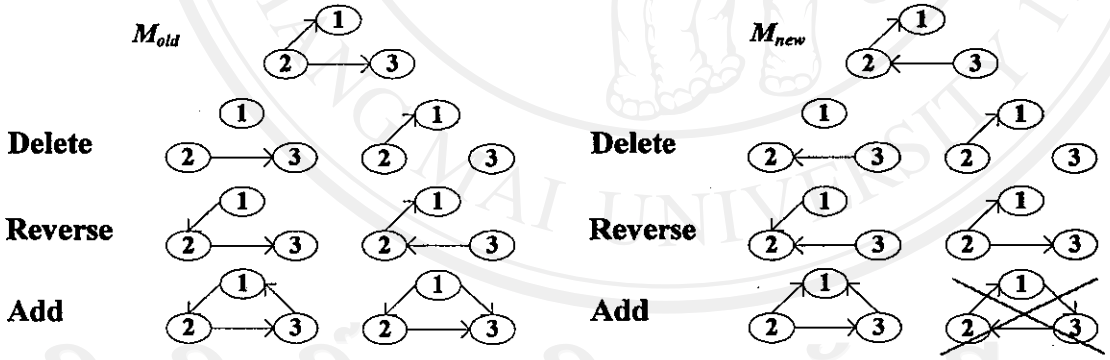
$$edge\_prob[i,j] \sim dirichlet(a) \qquad (18)$$

ซึ่ง a แทนพารามิเตอร์รูปร่าง (Shape parameters) ของการแจกแจง Dirichlet ที่ประกอบด้วยซุปเปอร์-พารามิเตอร์ (Super-parameters) อีกสองตัว คือ $\{alpha_1, alpha_2\}$ ที่ได้มาจากความถี่ในการเกิดเส้นเชื่อมในแต่ละแหน่งของเมทริกซ์เครือข่ายทั้งหมดที่ถูกยอมรับไว้ก่อนหน้า โดยที่ $alpha_1$ จะเป็นพารามิเตอร์สำหรับการสุ่มกรณีมีเส้นเชื่อม (ค่าความสัมพันธ์เป็น 1) และ $alpha_2$ สำหรับการสุ่มกรณีไม่มีเส้นเชื่อม (ค่าความสัมพันธ์เป็น 0)

หลังจากทำการสุ่มเครือข่ายใหม่ M_{new} เรียบร้อยแล้ว จึงเข้าสู่กระบวนการพิจารณาว่าจะยอมรับเครือข่ายใหม่นี้หรือไม่ โดยพิจารณาจากความน่าจะเป็นที่ยอมรับได้ของอัลกอริทึมเมโทรโพลิส – แฮสติงส์ แสดงดังสมการที่ (19)

$$\alpha(M_{old}, M_{new}) = min\left\{1, \frac{P(M_{new} \mid D)}{P(M_{old} \mid D)} * \frac{Q(M_{old} \mid M_{new})}{Q(M_{new} \mid M_{old})}\right\} \qquad (19)$$

โดย D คือข้อมูลการแสดงของกลุ่มยีนในเครือข่ายที่เราสนใจที่ถูกแปลงเป็นข้อมูลชนิดไม่ต่อเนื่องเรียบร้อยแล้ว ในส่วนของการแจกแจงที่คงที่ $P(.|D)$ จะถูกคำนวณโดยอาศัยหลักการของเบย์เซียน ซึ่งจะกล่าวถึงต่อไปในหัวข้อที่ 4.2.2 และในส่วนของการแจกแจงที่นำเสนอ $Q(.|.)$ จะคำนวณมาจากการหาเครือข่ายเนจ์เบอร์ (neighbor network) ของ M_{old} และ M_{new} ในแต่ละรอบของการทดลอง ซึ่งเหมือนกับวิธีการที่ได้เสนอไว้โดยHusmeier (2003) โดยเครือข่ายเนจ์เบอร์นี้ต้องเป็นกราฟชนิด DAG ที่เกิดมาจากการปรับปรุงเครือข่าย (update network) M_{old} และ M_{new} ด้วยวิธีการลบ (deletion) การกลับทิศทาง (reversion) และการเพิ่ม (addition) เส้นเชื่อมของเครือข่าย M_{old} และ M_{new} ทีละหนึ่งเส้นเชื่อม สามารถแสดงตัวอย่างการหาเครือข่ายเนจ์เบอร์ได้ดังรูป 4.2



$$Q(M_{new} \mid M_{old}) = \frac{1}{6} \qquad Q(M_{old} \mid M_{new}) = \frac{1}{5}$$

$$\frac{Q(M_{old} \mid M_{new})}{Q(M_{new} \mid M_{old})} = \frac{1/5}{1/6} = \frac{6}{5} = \frac{N(M_{old})}{N(M_{new})}$$

รูป 4.2 ตัวอย่างการหาเครือข่ายเนจ์เบอร์ของ M_{old} และ M_{new} ที่ได้มาจากการลบ การกลับทิศทาง และการเพิ่มเส้นเชื่อมที่ละหนึ่งเส้นเชื่อม ซึ่งเครือข่ายเนจ์เบอร์จะต้องเป็นกราฟชนิด DAG ดังนั้นเครือข่ายเนจ์เบอร์สุดท้ายของ M_{new} จึงถูกตัดออก ทำให้เครือข่าย M_{old} มีเครือข่ายเนจ์เบอร์ทั้งหมด 6 เครือข่าย และเครือข่าย M_{new} มีเครือข่ายเนจ์เบอร์ทั้งหมด 5 เครือข่าย

จากตัวอย่างการหาเครือข่ายเนจ์เบอร์ในรูป 4.2 เมื่อทำการคำนวณหาค่าความน่าจะเป็นของการแจกแจงที่นำเสนอ $Q(.|.)$ โดยพิจารณาจากเครือข่ายเนจ์เบอร์ของ M_{old} และ M_{new} พบว่า ค่าความน่าจะเป็นของการแจกแจงที่นำเสนอของ M_{old} เมื่อทราบค่าของ M_{new} แทนด้วย $Q(M_{old}|M_{new})$ มีค่าเท่ากับ 1/5 และค่าความน่าจะเป็นของการแจกแจงที่นำเสนอของ M_{new} เมื่อทราบค่าของ M_{old} แทนด้วย $Q(M_{new}|M_{old})$ มีค่าเท่ากับ 1/6 เมื่อพิจารณาค่าของเทอมการแจกแจงที่นำเสนอนี้จะมีค่าเท่ากับอัตราส่วนของจำนวนเครือข่ายเนจ์เบอร์ของ M_{old} และ M_{new} ซึ่งจะถูกเรียกว่าอัตราส่วนแฮสติงส์ (Hastings ratio) ของ M_{old} และ M_{new} ดังนั้นการแจกแจงที่นำเสนอ $Q(.|.)$ ในสมการที่ (19) จะถูกคำนวณด้วยอัตราส่วนในสมการที่ (20) โดย N แทนจำนวนเครือข่ายเนจ์เบอร์ของ M_{old} และ M_{new}

$$\frac{Q(M_{old}|M_{new})}{Q(M_{new}|M_{old})}=\frac{N(M_{old})}{N(M_{new})} \tag{20}$$

สุดท้ายจึงทำการพิจารณาว่าจะยอมรับเครือข่าย M_{new} ที่ถูกสร้างขึ้นมานี้หรือไม่ โดยการเปรียบเทียบค่าความน่าจะเป็นที่ยอมรับได้ซึ่งคำนวณได้มาจากสมการที่ (19) กับค่าความน่าจะไปเป็นที่สุ่มมาได้จากการแจกแจงเอกรูป (Uniform distribution) เมื่อกำหนดค่าพารามิเตอร์ของการแจกแจงดังนี้คือ พารามิเตอร์ค่าน้อยที่สุด (min) เท่ากับ 0 และพารามิเตอร์ค่ามากที่สุด (max) เท่ากับ 1 สามารถแสดงดังอัลกอริทึมของการพิจารณายอมรับเครือข่าย M_{new} ได้ดังต่อไปนี้

U ~ uniform (0, 1)
If (U <= α (M_{old}, M_{new}))
    M_{old} [t+1] = M_{new}[t]
    Network [t] = M_{new}[t]
    t = t + 1

4.2.2 เครือข่ายเบย์เซียน

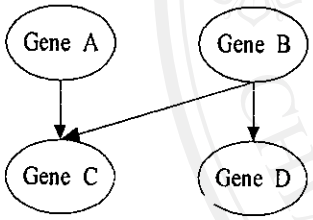
จากที่กล่าวไว้ในหัวข้อ 4.2.1 ว่าหลักการของเบย์เซียนจะถูกนำมาใช้ในส่วนของการแจกแจงที่คงที่ของอัลกอริทึมเมโทรโพลิส – แฮสติงส์ โดยการคำนวณจะอาศัยหลักการทฤษฎีของเบย์ (Baye's theorem) และเครือข่ายความเชื่อของเบย์เซียน (Bayesian belief network) ดังนี้

จากทฤษฎีของเบย์ (Gamerman, 1997) พบว่าเทอมการแจกแจงที่คงที่ $P(.|D)$ ในสมการที่ (17) ก็คือการแจกแจงโพสทีเรียของเครือข่ายที่เราสนใจที่มีเงื่อนไขภายใต้ข้อมูลการแสดงออกของยีนในเครือข่าย D ดังนั้นจากทฤษฎีของเบย์จึงทำให้ได้ว่า

$$\frac{P(M_{new}|D)}{P(M_{old}|D)}=\frac{P(D|M_{new})P(M_{new})}{P(D|M_{old})P(M_{old})} \tag{21}$$

โดยที่ $P(.)$ คือการแจกแจงไพรเออร์ของเมทริกซ์เครือข่าย M และ$P(D|M_i)$ คือค่าไลค์ลิฮูดที่คำนวณมาจากตารางความน่าจะเป็นอย่างมีเงื่อนไขสำหรับแต่ละยีนในเครือข่าย ด้วยวิธีการของเครือข่ายความเชื่อของเบย์เซียน (Han and Kamber, 2001)

เครือข่ายความเชื่อของเบย์เซียนเป็นวิธีที่คำนึงถึงข้อมูลที่เป็นจริงว่า ในแต่ละแถวข้อมูลหรือการทดลองในแต่ละครั้งอาจมีคุณลักษณะบางประการของข้อมูลที่มีความสัมพันธ์ต่อกัน ในเครือข่ายความเชื่อของเบย์เซียนจึงได้จัดให้มีแบบจำลองเชิงกราฟของความเชื่อของความสัมพันธ์ ที่เรียกว่าเครือข่ายความเชื่อ (Belief network) ซึ่งประกอบด้วย 2 ส่วน ส่วนแรกคือกราฟของเครือข่ายความสัมพันธ์ชนิด DAG จากรูป 4.3(ก) เป็นตัวอย่างกราฟ DAG จะเห็นได้ว่า GeneC จะทำงานได้ก็ต่อเมื่อถูกควบคุมโดย GeneA และ GeneB แต่ไม่ขึ้นกับ GeneD ส่วนที่สองคือตารางความน่าจะเป็นอย่างมีเงื่อนไข หรือเรียกง่ายๆ ว่าตารางซีพีที (CPT) สำหรับแต่ละตัวแปร โดยที่ตารางซีพีทีของตัวแปร Z จะเจาะจงการกระจายของเงื่อนไข $P(Z|Parents(Z))$



(ก)

	A↑,B↑	A↑,B−	A↑,B↓	A−,B↑	A−,B−	A−,B↓	A↓,B↑	A↓,B−	A↓,B↓
C↑	0.9	0.2	0.5	0.2	0.15	0.8	0.3	0.25	0.7
C−	0.05	0.7	0.1	0.4	0.35	0.1	0.6	0.45	0.2
C↓	0.05	0.1	0.4	0.4	0.5	0.1	0.1	0.3	0.1

(ข)

รูป 4.3 ส่วนประกอบของเครือข่ายความเชื่อของเบย์เซียน (ก) กราฟของเครือข่ายความสัมพันธ์ชนิด DAG และ (ข) ตารางซีพีที สำหรับ GeneC

จากรูป 4.3 (ข) ซึ่งแสดงตารางซีพีที สำหรับ GeneC ค่าความน่าจะเป็น ที่ GeneC จะทำงาน แต่ละค่าที่ให้มาเป็นการรวมกันของค่าความเป็นไปได้ของค่า ของโหนดพ่อแม่ของ GeneC เช่นจากตารางซีพีที ช่องด้านบนซ้ายสุดและด้านล่างขวาสุด เราจะเห็นได้ว่า

P(GeneC = "↑" | GeneA = "↑", GeneB = "↑") = 0.9
P(GeneC = "↓" | GeneA = "↓", GeneB = "↓") = 0.1

ความน่าจะเป็นร่วมกันของแถวใดๆ $(z_1, z_2, ..., z_n)$ ที่สอดคล้องกับตัวแปรหรือแอททริบิวต์ $Z_1, ..., Z_n$ คำนวณได้จากสมการที่ (22)

$$P(z_1, ..., z_2) = \prod_{i=1}^{n} P(z_i \mid Parents(Z_i)) \qquad (22)$$

โดยที่ค่าสำหรับ $P(z_i \mid Parents(Z_i))$ สอดคล้องกับค่าในตารางซีพีทีสำหรับ Z_i

เราได้ทำการคำนวณค่าในตารางซีพีที ซึ่งก็คือค่าความน่าจะเป็นอย่างมีเงื่อนไขของตัวแปรที่เราสนใจโดยอ้างอิงตามสมการที่ (1) และเรากำหนดตัวแปรดังต่อไปนี้ นั่นคือ ให้ S เป็นเซตของตัวอย่างข้อมูลที่นำมาศึกษา (Training sample) จำนวน s ตัวคือ $X_1, X_2, ..., X_s$ และให้ w_{ijk} เป็นค่าในตารางซีพีทีสำหรับตัวแปร $Y_i = y_{ij}$ โดยมีกลุ่มของตัวแปรพ่อแม่ $U_i = u_{ik}$ โดยที่ i คือตัวแปร, j คือค่าที่เป็นไปได้ของตัวแปร i และ k คือค่าที่เป็นไปได้ของพ่อแม่ของ i ยกตัวอย่างเช่น ถ้า w_{ijk} เป็นค่าในตารางซีพีทีที่อยู่ด้านบนซ้ายสุดของรูป 4.3(ข) ซึ่งมีค่าความน่าจะเป็นเท่ากับ 0.9 จะได้ว่า Y_i เป็น GeneC ที่มีค่าเป็น y_{ij} = "↑" และ U_i เป็นรายการของโหนดพ่อแม่ของ Y_i คือ {GeneA, GeneB} ที่มี u_{ik} แสดงรายการค่าของโหนดพ่อแม่ดังนี้ {"↑", "↑" } เราสามารถมองว่า w_{ijk} เป็นน้ำหนัก (Weight) ของข้อมูล โดยที่ในแต่ละรอบของการทำซ้ำ เราจะปรับค่าน้ำหนักจนกระทั่งเมื่อค่าของน้ำหนักไม่เปลี่ยนแปลง วิธีนี้จะค้นหาค่า w_{ijk} ที่สามารถแสดงข้อมูลได้ดีที่สุด มีเป้าหมายคือต้องการหาค่าที่มากที่สุดของความน่าจะเป็นของกลุ่มตัวอย่างข้อมูลที่นำมาศึกษา ถ้าให้โครงสร้างเครือข่ายมาการทำงานก็จะเกิดขึ้นดังต่อไปนี้

(1) คำนวณค่าความน่าจะเป็น $P(Y_i = y_{ij} \mid U_i = u_{ik})$ สำหรับทุกค่า i, j, k ที่เป็นไปได้ จะได้ค่า w_{ijk} ที่เป็นสมาชิกในตารางซีพีทีสำหรับตัวแปร Y

(2) คำนวณค่าความน่าจะเป็นสำหรับแต่ละ i, j, k โดยพิจารณาในแต่ละกรณีของข้อมูลดังนี้

$$\frac{\partial \ln P_w(S)}{\partial w_{ijk}} = \sum_{d=1}^{s} \frac{P(Y_i = y_{ij}, U_i = u_{ik} \mid X_d)}{w_{ijk}} \tag{23}$$

เมื่อ $P_w(S) = \prod_{d=1}^{s} P_w(X_d)$ และความน่าจะเป็นในด้านขวามือของสมการที่ (23) จะถูกคำนวณสำหรับแต่ละตัวอย่างข้อมูลที่นำมาศึกษา X_d ใน S

(3) ทำการปรับค่าน้ำหนักโดย

$$w_{ijk} \leftarrow w_{ijk} + (l)\frac{\partial \ln P_w(S)}{\partial w_{ijk}} \tag{24}$$

โดยที่ l เป็นอัตราการเรียนรู้ (Learning rate) แทนขนาดของก้าว (Step) ของการเรียนรู้จะถูกกำหนดให้เป็นค่าคงที่ค่าหนึ่งที่มีค่าน้อยมาก และ $\frac{\partial \ln P_w(S)}{\partial w_{ijk}}$ คือค่าที่คำนวณได้จากสมการที่ (23)

(4) ทำการรีนอร์มอลไลซ์น้ำหนัก เนื่องจากน้ำหนัก w_{ijk} เป็นค่าความน่าจะเป็นจึงมีค่าตั้งแต่ศูนย์ถึงหนึ่ง และ $\sum_j w_{ijk}$ จะต้องเท่ากับหนึ่งสำหรับทุกๆ i, k โดยการทำรีนอร์มอลไลซ์ค่าน้ำหนัก หลังจากถูกปรับค่าโดยสมการที่ (24) แล้ว ดังสมการต่อไปนี้

$$w_{ijk} \leftarrow \frac{w_{ijk}}{\sum_{j} w_{ijk}} \qquad (25)$$

และเราจะหยุดการเรียนรู้ค่าในตารางซีพีที เมื่อค่า w_{ijk} ไม่เปลี่ยนแปลง หรือเปลี่ยนแปลงน้อยมากจนเรียกได้ว่าคงที่ และจะได้ตารางซีพีทีที่เป็นตัวแทนของแต่ละตัวแปร บอกค่าความน่าจะเป็นอย่างมีเงื่อนไข (Conditional probability)

(ออมพิไล มโนรัตน์, 2548)

4.2.3 อัลกอริทึมของการทดลอง

จากวิธีการและการทดลองที่ได้กล่าวมาแล้วข้างต้น สามารถสรุปเป็นอัลกอริทึมของการทดลองในครั้งนี้ได้ดังต่อไปนี้

Algorithm MH-based Bayesian Regulatory Network
Input: *ExpData*: Gene expression data
HypoNet: Hypothetical regulatory network
Output: *Network*: A set of all accepted networks by Metropolis – Hastings

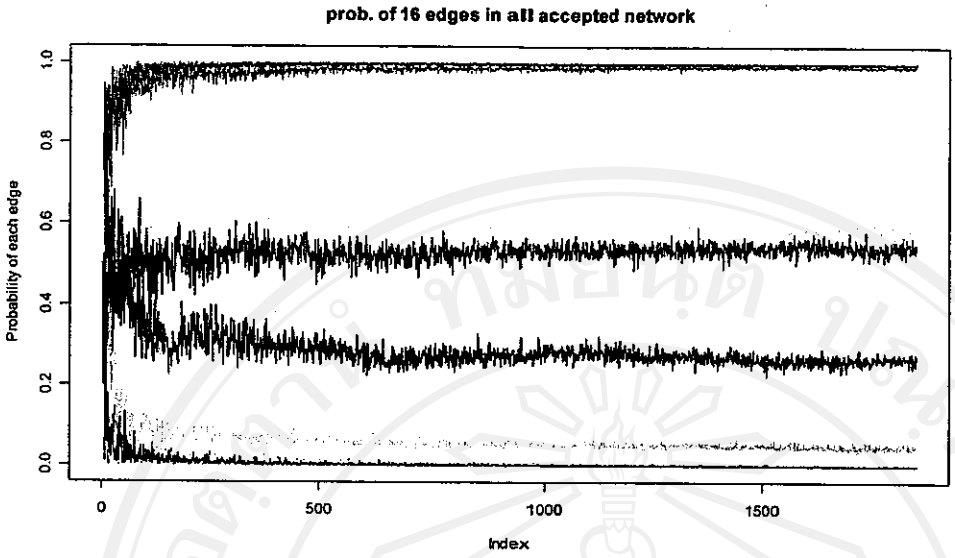
1. Transform *ExpData* to discrete dataset
2. Initialize M_{old} with the hypothetical regulatory network *HypoNet*
3. Repeat step 4 to step 9 for long enough time
4. Sampling a new acyclic network M_{new} by
   If (it is the first sampling) Then
   $M_{new}[i,j] \sim binomial(0.5)$, for each pair of genes (i, j)
   Else
   $edge\_prob[i,j] \sim dirichlet(a)$
   $M_{new}[i,j] \sim binomial(edge\_prob[i,j])$, for each pair of genes (i, j)
5. Find N(M_{old}) the number of acyclic neighbor networks of M_{old}
6. Find $N(M_{new})$ the number of acyclic neighbor networks of M_{new}
7. Compute acceptance probability: $\alpha(M_{old}, M_{new})$
   where $\alpha(M_{old}, M_{new}) = min\left\{1, \frac{P(M_{old}|D)}{P(M_{new}|D)} * \frac{N(M_{old})}{N(M_{new})}\right\}$
8. Random a number U from uniform(0,1)
9. If($U <= \alpha(M_{old}, M_{new})$) Then
   Accept M_{new} and keep it into *Network* set
   Set M_{new} to be M_{old} for the next sampling
   Update Dirichlet parameter $a = \{alpha_1, alpha_2\}$:
   $alpha_1[i,j] = alpha_1[i,j]+2$ If $M_{new}[i,j]=1$
   $alpha_2[i,j] = alpha_2[i,j]+2$ If $M_{new}[i,j]=0$

4.3 ผลการทดลอง

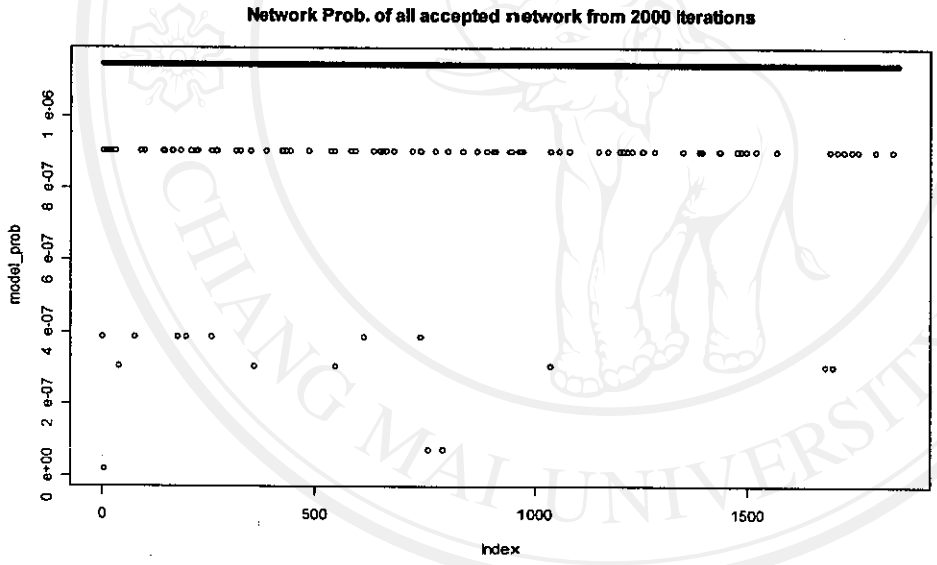
จากการรันโปรแกรมตามอัลกอริทึมในหัวข้อ 4.2.3 แบบทำซ้ำด้วยเครือข่ายที่มีการรายงานไว้แล้วและข้อมูลระดับการแสดงออกของยีน HAP2 HAP3 HAP4 และ CYC1 ทั้งหมด 2,000 รอบ เพื่อทำการสร้าง (reconstructing) เครือข่ายการควบคุมกันระหว่างยีนที่เป็นไปได้ พบว่าด้วยวิธีการของเมโทรโพลิส – แฮสติงส์ จะยอมรับเครือข่ายที่ถูกสร้างขึ้นมาจำนวน 1,853 เครือข่าย โดยค่าของ *edge_prob* ซึ่งเป็นความน่าจะเป็นของเส้นเชื่อมในแต่ละตำแหน่งทั้งหมด 16 เส้นเชื่อมที่เป็นผลจากการแจกแจง Dirichlet ที่นำไปใช้สร้างเครือข่ายที่ยอมรับทั้งหมดนี้จะค่อย ๆ ลู่เข้าจนมีค่าคงที่แสดงดังรูป 4.4 โดยในตอนเริ่มต้นของการทดลอง ค่าความน่าจะเป็นของเส้นเชื่อมทุกตำแหน่งจะถูกกำหนดด้วยค่า 0.5 หลังจากนั้นค่าความน่าจะเป็นที่ได้จากการแจกแจง Dirichlet ของแต่ละเส้นก็จะมีค่ามากขึ้นหรือลดลงจนลู่เข้าสู่ค่าใดค่าหนึ่ง (local optimum)

เมื่อพิจารณาเฉพาะเครือข่ายที่ยอมรับทั้ง 1,853 เครือข่าย พบว่าเครือข่ายที่ได้นั้นมีลักษณะใกล้เคียงกับเครือข่ายจริงของการควบคุมกันระหว่างกลุ่มยีน HAP2 HAP3 HAP4 และ CYC1 นั่นคือในทุกเครือข่ายที่ยอมรับนั้นจะมีความสัมพันธ์ของทั้งสี่ยีนนี้เกิดขึ้นอย่างน้อยหนึ่งความสัมพันธ์ ซึ่งสามารถแสดงค่าความน่าเป็นโพสทีเรียของทุกเครือข่ายที่ยอมรับได้ดังรูป 4.5 พบว่าค่าความน่าจะเป็นของเครือข่ายที่ยอมรับส่วนใหญ่จะมีค่าเท่ากับ 1.143835e-06 ซึ่งเป็นค่าความน่าจะเป็นโพสทีเรียของเครือข่ายที่สูงที่สุด ซึ่งตรงกับแนวคิดที่ว่าเครือข่ายที่ดีที่สุดควรจะมีค่าความน่าจะเป็นของเครือข่ายมากที่สุด ถึงแม้ว่ายังมีบางเครือข่ายที่ยอมรับที่มีค่าความน่าจะเป็นต่ำแต่พบว่าปริมาณของเครือข่ายเหล่านั้นจะมีจำนวนลดลงเมื่อเข้าสู่การทดลองซ้ำในรอบที่มากขึ้น ดังนั้นถ้าทำการเรียนรู้ต่อไปด้วยจำนวนการทดลองที่เพิ่มมากขึ้นอาจจะทำให้ได้เฉพาะผลการทดลองที่เป็นเครือข่ายที่ยอมรับที่มีค่าความน่าจะเป็นของเครือข่ายสูงที่สุดเท่านั้น

จากผลลัพธ์ของการสร้างและพิจารณายอมรับเครือข่ายความสัมพันธ์ของกลุ่มยีนจากการทดลองนี้จนได้เครือข่ายที่ยอมรับทั้งหมดออกมา จากนั้นจึงทำการพิจารณาเลือกเครือข่ายที่ยอมรับทั้งหมด 100 เครือข่ายสุดท้ายออกมา ซึ่งเครือข่ายเหล่านี้ถูกสร้างมาจากความน่าจะเป็นของเส้นเชื่อมที่อยู่ในช่วงที่คงที่แล้ว โดยเครือข่ายที่ยอมรับในช่วงแรกจะเป็นเครือข่ายที่อยู่ในช่วงที่ยังไม่คงที่ (burn - in) ซึ่งจะไม่ถูกนำมาพิจารณา โดยพบว่า เครือข่ายทั้ง 100 เครือข่ายที่ถูกเลือกมานั้นเป็นเครือข่ายที่มีลักษณะความสัมพันธ์คล้ายคลึงกัน โดยมีรูปแบบความสัมพันธ์ที่พบทั้งหมด 10 รูปแบบ แสดงดังรูป 4.6 ซึ่งแต่ละรูปแบบจะประกอบด้วยลักษณะรูปแบบความสัมพันธ์ของเครือข่ายการควบคุมกันระหว่างยีน ค่าความถี่ของการเกิดเครือข่ายที่ยอมรับนั้น (frq.) ค่าความน่าจะเป็นโพสทีเรียของเครือข่าย (prob.) และค่าความน่าจะเป็นของการเกิดขึ้นของเส้นเชื่อมในแต่ละตำแหน่ง (ตัวเลขบนเส้นเชื่อม)



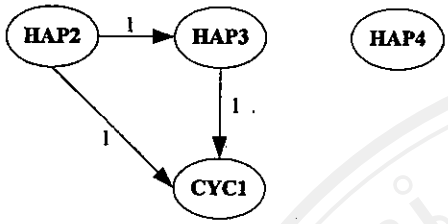
รูป 4.4 การลู่เข้าของค่าความน่าจะเป็นของ *edge_prob* สำหรับทุกเส้นเชื่อม



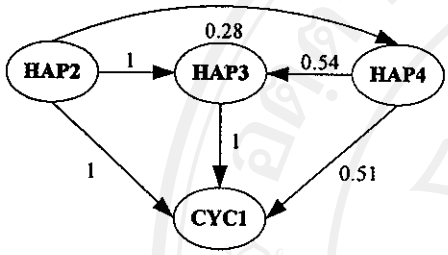
รูป 4.5 ค่าความน่าเป็นของเครือข่ายที่ยอมรับทั้ง 1,853 เครือข่าย

ในรูป 4.6 พบว่าเครือข่ายทั้งสิบรูปแบบนี้มีความสัมพันธ์ของยีน HAP2 และ HAP3 ไปยังยีน CYC1 ปรากฏอยู่ทุกเครือข่าย ซึ่งตรงกับเครือข่ายจริงที่ได้มีการรายงานไว้ว่าทั้งสองยีนนี้ต่างร่วมกันไปควบคุมส่วนโปรโมเตอร์ของยีน CYC1 และเครือข่ายที่ยอมรับในรูปแบบที่ 2, 5, 6, 7, 8 และ 10 นั้นพบว่ามีความสัมพันธ์ระหว่างยีนทั้งสี่นี้เกิดขึ้นตามเครือข่ายจริงทุกความสัมพันธ์ แต่อย่างไรก็ตามยังคงมีบางความสัมพันธ์ที่ไม่ตรงตามเครือข่ายจริงปรากฏอยู่ ซึ่งต้องมีการศึกษาเพิ่มเติมต่อไปทั้งในด้านการคำนวณและด้านชีววิทยาเพื่อเป็นค้นคว้าว่าความสัมพันธ์ระหว่างคู่ยีนอื่นๆ ที่พบในการทดลองครั้งนี้มีความเป็นไปได้มากน้อยเพียงใด

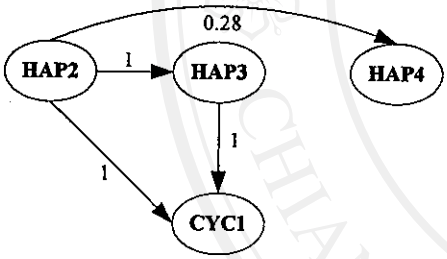
Pattern 1: frq. = 16, prob. = 1.142835e-06



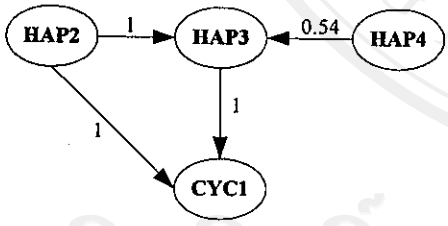
Pattern 2: frq. = 10, prob. = 1.142835e-06



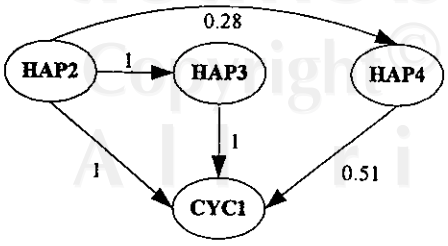
Pattern 3: frq. = 3, prob. = 1.142835e-06



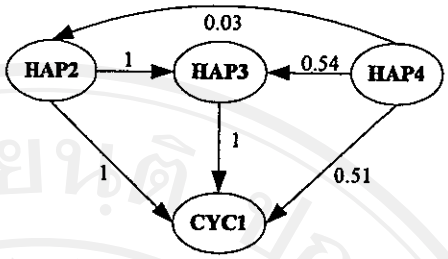
Pattern 4: frq. = 16, prob. = 1.142835e-06



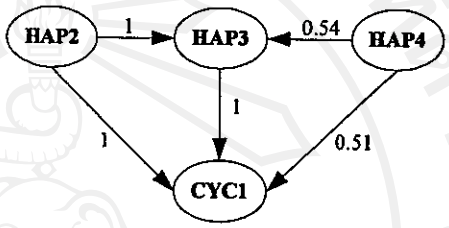
Pattern 5: frq. = 11, prob. = 1.142835e-06



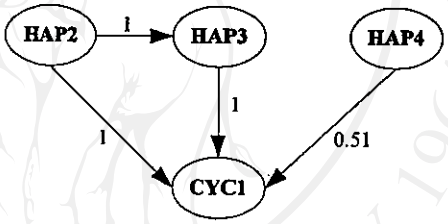
Pattern 6: frq. = 2, prob. = 9.0141e-07



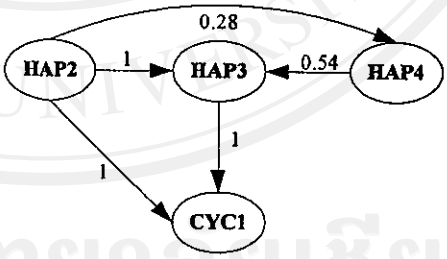
Pattern 7: frq. = 24, prob. = 1.142835e-06



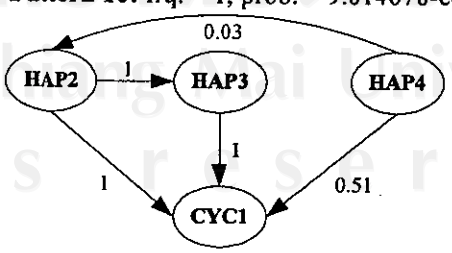
Pattern 8: frq. = 13, prob. = 1.142835e-06



Pattern 9: frq. = 4, prob. = 1.142835e-06



Pattern 10: frq. = 1, prob. = 9.014078-e07



รูป 4.6 รูปแบบความสัมพันธ์ที่พบทั้งหมด 10 รูปแบบ
จากเครือข่ายที่ยอมรับ 100 เครือข่ายที่ถูกเลือก

สุดท้ายเมื่อทำการพิจารณาความน่าจะเป็นของเครือข่ายที่ยอมรับที่เลือกมาทั้ง 100 เครือข่ายนี้ พบว่าเครือข่ายส่วนใหญ่ (97 เครือข่าย) มีค่าความน่าจะเป็นของเครือข่ายเท่ากับค่าความน่าจะเป็นสูงสุดซึ่งเท่ากับ 1.143835e-06 และในทั้ง 97 เครือข่ายนี้ พบว่าเครือข่ายที่มีความถี่ในการยอมรับมากที่สุดเท่ากับ 24 % คือ เครือข่ายรูปแบบที่ 7 ซึ่งมีลักษณะของเครือข่ายจริงปรากฏอยู่ด้วย ส่วนเครือข่ายที่มีค่าความน่าจะเป็นต่ำที่สุดเท่ากับ 9.014078-e07 ทั้งหมด 3 เครือข่าย และเป็นเครือข่ายที่มีความถี่ในการยอมรับน้อยที่สุดเท่ากับ 1 % และ 2 % คือ เครือข่ายในรูปแบบที่ 10 และรูปแบบที่ 6 ตามลำดับ และเมื่อพิจารณาค่าความน่าจะเป็นของการเกิดขึ้นของเส้นเชื่อมในแต่ละตำแหน่งของเครือข่ายที่ถูกเลือกมาทั้ง 100 เครือข่ายนี้พบว่าค่าความน่าจะเป็นนี้มีค่าใกล้เคียงกับค่าความน่าจะเป็นของการสร้างเส้นเชื่อม *(edge_prob)* ดังรูป 4.4 โดยความสัมพันธ์ระหว่างยีนที่เกิดขึ้นเสมอในเครือข่ายทั้งสิบรูปแบบซึ่งมีค่าความน่าจะเป็นของการเกิดของเส้นเชื่อมเป็น 1 คือ ความสัมพันธ์ที่ยีน HAP2 ไปควบคุมการแสดงออกของยีน HAP3 และ CYC1 และความสัมพันธ์ที่ยีน HAP3 ไปควบคุมยีน CYC1 และสำหรับความสัมพันธ์ที่ยีน HAP4 ไปควบคุมการแสดงออกของยีน CYC1 มีค่าความน่าจะเป็นของการเกิดของเส้นเชื่อมเป็น 0.54

4.4 บทสรุป

จากงานวิจัยนี้ได้เสนอวิธีการของมาร์คอฟเชนมอนติคาร์โลเพื่อทำการสร้างเครือข่ายเบย์เซียนจากเครือข่ายการควบคุมกันระหว่างยีนที่ได้มีการรายงานไว้แล้วและข้อมูลการแสดงออกของยีนเหล่านั้นในเครือข่าย ซึ่งจากการทดลองในครั้งนี้ก็แสดงให้เห็นว่ากระบวนการควบคุมกันระหว่างยีนสามารถตรวจสอบได้ด้วยเครือข่ายเบย์เซียนร่วมด้วยวิธีการมาร์คอฟเชนมอนติคาร์โลได้อย่างถูกต้องและมีความน่าเชื่อถือในระดับหนึ่ง โดยในงานวิจัยนี้จะใช้ทั้งสองวิธีการข้างต้นร่วมกันในการสร้างและพิจารณาเครือข่ายการควบคุมกันระหว่างยีน ซึ่งพบว่าผลลัพธ์ที่ได้จากการทดลองในครั้งนี้มีความใกล้เคียงกับเครือข่ายจริงที่ได้มีการรายงานไว้แล้ว แต่ยังมีบางความสัมพันธ์ระหว่างยีนที่ไม่เป็นไปตามเครือข่ายจริง

นอกจากนี้ผลลัพธ์ที่ได้นั้นเป็นผลจากการทดลองกับข้อมูลเพียงชุดเดียว ดังนั้นเพื่อเป็นการยืนยันความถูกต้องของวิธีการตามอัลกอริทึมข้างต้นมากยิ่งขึ้น เราควรทำการทดลองกับข้อมูลชุดอื่นๆ ด้วย และควรทำการพัฒนาโปรแกรมนี้ให้สามารถวิเคราะห์ข้อมูลชนิดต่อเนื่องได้ ซึ่งจะถือได้ว่าเป็นการทำการวิเคราะห์จากข้อมูลการแสดงออกของยีนโดยตรง

แต่อย่างไรก็ดีในการที่จะนำวิธีการนี้ไปใช้เพื่อทำการสร้างหรือทำนายเครือข่ายการควบคุมกันระหว่างยีนที่ยังไม่มีการรายงานไว้ ควรมีการศึกษาเพิ่มเติม หรือ เพิ่มวิธีการบางอย่างเข้าไป

เพื่อให้ได้ผลการทดลองที่มีความถูกต้องและน่าเชื่อถือมากยิ่งขึ้น ซึ่งจะเป็นงานที่จะดำเนินต่อไปในอนาคต