



วารสาร ไทยไทยซีเอ็มพีเอ็น

ปีที่ 2 ฉบับเดือนพฤษภาคม 2548 (หน้า 93-106)

บทความพิเศษวิชาการ สำหรับนักศึกษาต่อเนื่องทางเภสัชศาสตร์



## การศึกษาความสัมพันธ์ระหว่างโครงสร้างและคุณสมบัติของสารเชิงปริมาณ (Quantitative structure property relationship studies)

ภญ.ผศ.ดร.สุชาดา พิริยะประสาธน์

ภาควิชาเทคโนโลยีเภสัชกรรม คณะเภสัชศาสตร์ มหาวิทยาลัยศิลปากร

วิทยาเขตพระราชวังสนามจันทร์ อ.เมือง จ.นครปฐม 73000

รหัส 1-000-SPU-000-0505-01

จำนวน 3.0 หน่วยกิตการศึกษาต่อเนื่อง

วันที่รับรอง : 1 พฤษภาคม พ.ศ. 2548

วันที่หมดอายุ : 1 พฤษภาคม พ.ศ. 2550

### วัตถุประสงค์เชิงพฤติกรรม

1. อธิบายโครงสร้างของ QSPR
2. ทราบหลักเกณฑ์และเทคนิคต่างๆ (โดยสังเขป) ที่ใช้ในการสร้างแบบจำลองความสัมพันธ์
3. ทราบหลักการในการสร้างแบบจำลอง
4. ทราบประโยชน์ของ QSPR ในงานวิจัย

### บทคัดย่อ

การศึกษาความสัมพันธ์ระหว่างโครงสร้างและฤทธิ์ของสารเชิงปริมาณ (Quantitative Structure Activity Relationships, QSAR) ได้ถูกนำมาใช้ในกระบวนการพัฒนายาใหม่มาเป็นเวลานานนับสิบปี ปัจจุบันมีการนำเทคนิคนี้มาใช้ในขั้นตอนการค้นหาสารต้นแบบ (lead) มากขึ้น โดยจะเน้นความสนใจเรื่องคุณสมบัติของสาร (property) เพื่อช่วยในการคัดเลือกสารที่มีคุณสมบัติที่เหมาะสมต่อการนำไปพัฒนาเป็นยา ทำให้สามารถลดจำนวนสารสังเคราะห์ลงได้อย่างมหาศาล บทความนี้จะให้ความรู้เกี่ยวกับการพัฒนาวิธีศึกษาความสัมพันธ์ระหว่างโครงสร้างและคุณสมบัติของสารเชิงปริมาณ (Quantitative Structure Property Relationships, QSPR) การนำไปประยุกต์ใช้ในทางเภสัชกรรม ข้อดีและข้อจำกัดของวิธี รวมถึงชนิดของตัวแปรอิสระ คุณสมบัติของสาร และวิธีทางสถิติที่นิยมใช้ในการศึกษา QSPR

### คำสำคัญ

QSPR, structural descriptor, multivariate statistical technique, property

## บทนำ

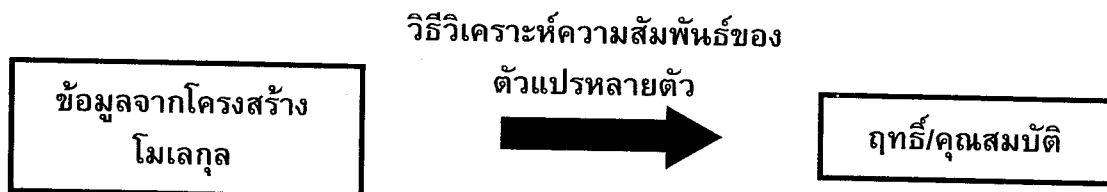
โดยทั่วไปนักเคมีจะใช้วิธีการศึกษาความสัมพันธ์ระหว่างโครงสร้างและฤทธิ์ของสารเชิงปริมาณ (Quantitative Structure Activity Relationships, QSAR) เพื่อศึกษาหาสารที่มีฤทธิ์ทางชีวภาพที่เหมาะสมเพื่อนำไปเป็นสารต้นแบบในการพัฒนา ในปัจจุบันด้วยเทคโนโลยีที่ก้าวหน้าและการใช้เทคนิคทางสถิติและคณิตศาสตร์ ทำให้สามารถค้นหาสารที่จะนำมาศึกษาเพื่อใช้เป็นสารต้นแบบได้เป็นจำนวนมาก

การออกแบบยาโดยอาศัยโครงสร้างของยา (Structure-based design) เป็นวิธีที่ได้รับการยอมรับกันอย่างแพร่หลายในกระบวนการพัฒนาเภสัชภัณฑ์ใหม่ๆ เทคนิคที่นำมาใช้ไม่เพียงแต่เป็นการทำนายฤทธิ์ของสารเท่านั้น แต่ยังใช้ทำนายคุณสมบัติทางเคมีฟิสิกส์และคุณสมบัติอื่นๆ ของสารนั้นๆ ก่อนการสังเคราะห์การออกแบบยาโดยอาศัยโครงสร้างของยา จึงเป็นที่มาของวิธีการศึกษาความสัมพันธ์ระหว่างโครงสร้างและคุณสมบัติของสารเชิงปริมาณ (Quantitative Structure Property Relationships, QSPR)<sup>1</sup> แม้ว่า QSPR จะไม่สามารถแก้ปัญหาความยุ่งยากในเรื่องการใช้เวลาในการค้นหาพารามิเตอร์ที่เหมาะสมสำหรับการสร้างแบบจำลอง QSPR แต่วิธี QSPR มีส่วนช่วยอย่างมากในการลดปริมาณสารที่จะต้องสังเคราะห์ โดยเป็นเครื่องมือเพื่อใช้ในการคัดเลือกสารที่มีคุณสมบัติที่เหมาะสมที่สุดมาสังเคราะห์และทดสอบฤทธิ์ในสัตว์ทดลองและในมนุษย์ต่อไป

โดยหลักการ สูตรโครงสร้างของสารจะมีรหัสข้อมูลในเรื่องคุณสมบัติทางกายภาพ เคมีและชีวภาพของสารนั้นๆ ดังนั้นถ้าสามารถถอดรหัสข้อมูลในเรื่องคุณสมบัติเหล่านี้จากสูตรโครงสร้างทางเคมีได้ย่อมเป็นประโยชน์กับกระบวนการค้นหาใหม่ ในปี ค.ศ.1968 Crum-Brown&Fraser ได้ตีพิมพ์การศึกษาหาความสัมพันธ์ระหว่างผลทางสรีรวิทยากับโครงสร้างของสารเป็นครั้งแรก<sup>2</sup> จากนั้นมีงานวิจัยหลายงานที่ค้นพบความสัมพันธ์ระหว่างคุณสมบัติความชอบละลายในไขมัน (lipophilicity) ซึ่งแสดงโดยค่าการละลายหรือค่าสัมประสิทธิ์การแพร่ผ่านในน้ำกับน้ำมันกับผลทางชีวภาพ<sup>3,4</sup> เช่น ความเป็นพิษหรือฤทธิ์ทางเภสัชวิทยา ในปี ค.ศ.1935 Hammett ได้ทำการศึกษาผลของการเปลี่ยนแปลงสูตรโครงสร้างของสารที่มีต่อกลไกการเกิดปฏิกิริยา โดยสร้างแบบจำลองซึ่งแสดงให้เห็นว่าการเกิดปฏิกิริยาทางเคมีโดยทั่วไปจะเกิดขึ้นได้ยากในส่วนของโมเลกุลที่มี aromatic ring ซึ่งเป็นผลมาจาก resonance effect<sup>5</sup> การศึกษานี้ได้ถูกศึกษาต่อในปี ค.ศ.1956 โดย Taft ซึ่งอธิบายผลที่แตกต่างกันของคุณสมบัติ electrostatic และ steric ของกลุ่มที่มาแทนที่ในสารที่มีโครงสร้างเป็นไฮโดรคาร์บอน (aliphatic hydrocarbon) สายโซ่ยาว<sup>6</sup> วิวัฒนาการในการศึกษาความสัมพันธ์ระหว่างโครงสร้างกับฤทธิ์ทางชีวภาพได้เกิดขึ้นอีกครั้งในช่วงปี ค.ศ.1964 โดย Hansch และคณะ ได้ค้นพบ Linear Free Energy Relationship (LFER) ซึ่งแบบจำลอง (model) ที่สร้างขึ้นตั้งอยู่บนสมมติฐานที่ว่าฤทธิ์ทางชีวภาพมีความสัมพันธ์กับตัวแปรทางเคมีฟิสิกส์ซึ่งแสดงในรูปของ free energy<sup>7</sup> ดังนั้นต่อมาจึงมีการสร้างแบบจำลองมากมาย โดยการใช้เดสคริปเตอร์ซึ่งในที่นี้คือตัวแปรบรรยายโครงสร้าง (structural descriptor) ในรูปแบบต่างๆ เป็นตัวแปรอิสระ ซึ่งกำหนดได้โดยการสังเกต หรือจากการทดลองและการคำนวณ เช่น ตัวแปรทางเคมีฟิสิกส์ ได้แก่ Hammett constant ( $\sigma$ ), Hansch substituent constant ( $\pi$ ), Taft steric parameter ( $E_s$ ) เป็นต้น ในระยะหลังการคำนวณตัวแปรบรรยายโครงสร้างจะได้มาจากการใช้ทฤษฎี ซึ่งจะเน้นการคำนวณค่าตัวแปรจากสูตรโครงสร้างทางโมเลกุลของสารโดยการประยุกต์ใช้เทคนิคทางสถิติและคณิตศาสตร์ และการคำนวณโดยคอมพิวเตอร์ในการสร้างแบบจำลองสำหรับทำนาย ซึ่งทำให้การศึกษาไม่ต้องใช้เวลามากดังเช่นสมัยก่อน

## โครงสร้างของ QSPR

QSPR<sup>8-9</sup> ประกอบด้วยสามส่วนที่สำคัญคือ 1) ข้อมูลจากโครงสร้างโมเลกุลของสารหรือตัวแปรบรรยายโครงสร้าง 2) คุณสมบัติของสารที่ต้องการศึกษาโดยแสดงเป็นค่าที่สามารถวัดได้ และ 3) วิธีหรือเทคนิคที่ใช้ในการสร้างแบบจำลองความสัมพันธ์ เช่น วิธีวิเคราะห์ความสัมพันธ์ของตัวแปรหลายตัว ดังแสดงในรูปที่ 1 ซึ่งจะกล่าวถึงดังต่อไปนี้



รูปที่ 1 แสดงโครงสร้างของ QSPR

### 1. ข้อมูลจากโครงสร้างโมเลกุลของสาร

ในการศึกษา QSPR สิ่งแรกที่จะต้องทำคือการสร้างข้อมูลที่ได้จากส่วนของโครงสร้างของสาร ซึ่งอาจเป็นตัวแปร (parameter) หรือตัวแปรบรรยายโครงสร้าง (structural descriptor) ชนิดต่างๆ ตัวแปร (parameter) จะให้ข้อมูลของคุณสมบัติของทั้งโมเลกุล เช่น มวลโมเลกุล จุดหลอมเหลว ค่าความชอบละลายในไขมัน หรือตัวคาดประมาณ (estimator) ในแบบจำลองความถดถอย (regression model) ส่วนตัวแปรบรรยายโครงสร้าง (structural descriptor) เป็นดัชนี (index) ที่ให้ข้อมูลเกี่ยวกับโครงสร้างของสาร เช่น topological indices ซึ่งคำนวณจากส่วนประกอบทางโครงสร้างโมเลกุลแบบ 2 มิติของสาร โดยไม่จำเป็นต้องใช้โครงสร้างทาง electronic ของโมเลกุล จึงทำให้สามารถคำนวณได้อย่างรวดเร็วจากซอฟต์แวร์เฉพาะเมื่อเปรียบเทียบกับคำนวณจากโครงสร้างโมเลกุลแบบ 3 มิติ ตัวอย่างเช่น Weiner indices, Randic indices, Kier and Hall molecular connectivity indices ตัวแปรเหล่านี้จะมีความสัมพันธ์กับคุณสมบัติทางฟิสิกส์บางประการ เช่น ค่าการละลายในน้ำ ขนาดและพื้นที่ผิวของโมเลกุล เป็นต้น กล่าวโดยสรุปว่าตัวแปรเป็นค่าสังเกต (empirical) ที่ได้จากการทดลอง ในขณะที่ตัวแปรบรรยายโครงสร้างอาจเป็นค่าสังเกตที่ได้จากการทดลอง หรือเป็นค่าที่ได้จากการคำนวณโดยใช้ทฤษฎี (theory) ก็ได้

เป็นที่ทราบกันว่าค่าความชอบละลายในไขมัน เช่น partition coefficient ( $\log P_{o/w}$ ) และ distribution coefficient ( $\log D$ ) เป็นตัวแปรที่สำคัญที่สุดของยาในการศึกษา QSAR ที่เกี่ยวข้องกับคุณสมบัติทางชีวภาพ ทั้งนี้เนื่องจากยาจะต้องมีการดูดซึมผ่านเยื่อเมมเบรนซึ่งอาจมีความหนาของชั้นไขมันมากหรือน้อยตามธรรมชาติในบางครั้งเพื่อให้ประหยัดเวลาในการทำการทดลอง อาจคำนวณค่าดังกล่าวจากการสร้างแบบจำลอง โดยกำหนดให้ partition coefficient หรือ distribution coefficient เป็นคุณสมบัติที่เราสนใจ และใช้ตัวแปรบรรยายโครงสร้างอื่นๆ เพื่อให้ข้อมูลจากโครงสร้างของสาร

### 2. คุณสมบัติของสารที่ต้องการศึกษา

คุณสมบัติของสารที่นิยมใช้ในการศึกษา QSPR ได้แก่คุณสมบัติที่สำคัญต่อการค้นหาและพัฒนาาใหม่ ซึ่งไม่ได้หมายความว่าเพียงแค่อัตราทางชีวภาพเท่านั้น แต่ยังรวมถึงคุณสมบัติทางเคมีฟิสิกส์ต่างๆ ดังแสดงในตารางที่ 1

ตารางที่ 1 แสดงคุณสมบัติทางเคมีฟิสิกส์และฤทธิ์ทางชีวภาพที่ใช้ในการศึกษา QSPR

| Physicochemical property | Biological property                |
|--------------------------|------------------------------------|
| Organoleptic properties  | Activity                           |
| Boiling point            | Acute toxicity (LD <sub>50</sub> ) |
| Dissociation constant    | Alkylating profile (with DNA)      |
| Viscosity                | Bioconcentration                   |
| Meltingpoint             | Biodegradation                     |
| Molar volume             | Carcinogenicity                    |
| Diffusion coefficient    | Chronic toxicity                   |
| Partition coefficient    | Inhibitor constant                 |
| Octanol-water            | Metabolic profile                  |
| Air-water                | Michaelis constant                 |
| Reactivity               | Mutagenesis                        |
| Release characteristic   | Penetration through skin           |
| Solubility               | Pharmacokinetics                   |
| Stability                | Receptor binding                   |
| Transportability         |                                    |

3. เทคนิคหรือวิธีที่ใช้ในการสร้างแบบจำลองความสัมพันธ์เชิงปริมาณ

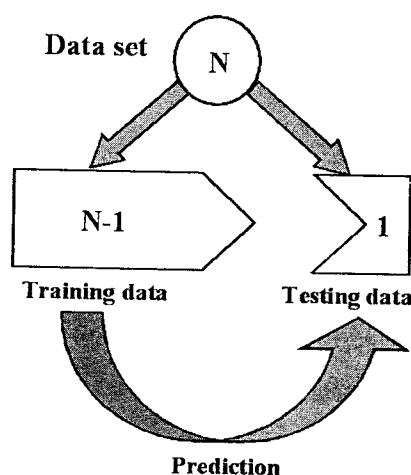
ในการศึกษา QSPR อาจใช้เทคนิคทางสถิติที่ใช้ในการวิเคราะห์ความสัมพันธ์ของตัวแปรหลายตัว (Multivariate Statistical Techniques, MSTs) เพื่อหาความสัมพันธ์ระหว่างคุณสมบัติที่สนใจศึกษา กับตัวแปรบรรยายโครงสร้าง หรือใช้เทคนิคอื่นๆ เช่น comparative molecular field analysis (CoMFA)<sup>10</sup> และ artificial neural network (ANN)<sup>11</sup> เป็นต้น

ในบรรดาเทคนิคทางสถิติที่ใช้ในการวิเคราะห์ความสัมพันธ์ของตัวแปรหลายตัว การวิเคราะห์ความถดถอยเส้นตรงพหุคูณ (Multiple Linear Regression Analysis, MLRA)<sup>12</sup> เป็นวิธีที่นิยมใช้ในการศึกษา QSPR มากที่สุด เทคนิคนี้จะใช้บอกความสัมพันธ์ระหว่างคุณสมบัติของยา (y) กับตัวแปรบรรยายโครงสร้างทั้งหมด (x) ที่ใช้ในการศึกษาได้ในเชิงปริมาณ ซึ่งสามารถแสดงสมการได้ดังนี้

$$y = a_1x_1 + a_2x_2 + a_3x_3 + \dots + a_mx_m + e$$

โดย  $x_1, x_2, x_3, \dots, x_m$  คือค่าตัวแปรที่ต้องการศึกษา  
 $y$  คือค่าของคุณสมบัติที่สนใจ  
 $a_1, a_2, a_3, \dots, a_m$  คือค่าสัมประสิทธิ์ความถดถอย  
 $e$  คือค่าความคลาดเคลื่อน

โดยจะใช้ค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณ (multiple correlation coefficient, R) Fisher's F ratio Student's t-value และส่วนเบี่ยงเบนมาตรฐาน (standard deviation, s) เพื่อหาค่าความถูกต้องของสมการ MLRA และเพื่อตรวจสอบว่ามีความแตกต่างอย่างมีนัยสำคัญของเกณฑ์ที่ใช้หรือไม่ การจำกัดจำนวนตัวแปรอิสระที่ใช้เมื่อเทียบกับจำนวนข้อมูลทั้งหมดจะช่วยลดโอกาสที่กลุ่มตัวแปรอิสระที่ใช้ในการคำนวณจะมีความสัมพันธ์กันสูง (chance correlation) ซึ่งเป็นปัญหาที่มักพบในการวิเคราะห์ความถดถอยเส้นตรงพหุคูณ นอกจากนี้สามารถตรวจสอบแบบจำลองที่สร้างขึ้นโดยใช้วิธี leave-one-out (LOO) cross validation ซึ่งเป็นวิธีที่ได้รับการยอมรับอย่างแพร่หลาย หลักการของวิธี LOO cross validation (รูปที่ 2) อาจกล่าวได้โดยสรุปคือสมมติว่ามีข้อมูลทั้งหมด  $n$  ข้อมูล ข้อมูลหนึ่งตัวจะถูกนำออกไปเพื่อใช้ทดสอบแบบจำลองที่สร้างขึ้นโดยข้อมูลที่เหลือ ( $n-1$ ) ทำการคำนวณซ้ำในลักษณะนี้ จนกระทั่งข้อมูลทุกตัวถูกนำมาใช้ในการทดสอบ จากนั้นใช้ค่า cross validated  $r^2$  ( $r_{cv}^2$ ,  $q^2$ ) เพื่อตรวจสอบความมีนัยสำคัญของความสัมพันธ์ของแบบจำลอง



รูปที่ 2 แสดงขั้นตอนวิธี Leave-one-out (LOO) cross validation โดย  $N$  คือจำนวนข้อมูลทั้งหมด ข้อมูลหนึ่งตัวจะถูกนำออกไปเพื่อใช้ทดสอบแบบจำลองที่สร้างขึ้นโดยข้อมูลที่เหลือ ( $N-1$ )

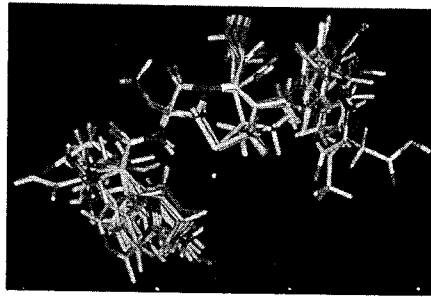
Comparative molecular field analysis (CoMFA)<sup>13</sup> เป็นวิธีหาความสัมพันธ์ระหว่างโครงสร้างกับฤทธิ์ทางชีวภาพหรือคุณสมบัติอื่นๆ แบบ 3 มิติ (3D-QSAR) ซึ่งรวมหลายๆ เทคนิคเข้าด้วยกัน ได้แก่ QSAR การคำนวณพลังงานระหว่างโมเลกุล pharmacophore mapping และเทคนิคการวิเคราะห์ความสัมพันธ์ของตัวแปรหลายตัว (MSTs) CoMFA จะใช้ค่าตัวแปรต่างๆ ได้แก่ พลังงาน steric และ electrostatic รอบๆ โมเลกุลเป็นตัวแปรอิสระที่ให้ข้อมูลของโครงสร้างเพื่ออธิบายถึงฤทธิ์ทางชีวภาพหรือคุณสมบัติของสารในเชิงปริมาณ โดยขึ้นกับลักษณะทางเคมีที่จำเพาะในพื้นที่ที่กำหนด พลังงาน steric และ electrostatic สามารถคำนวณโดยการวาง  $sp^3$  hybridized carbon probe atom ที่ตำแหน่งต่างๆ ในโครงร่างผลึกสามมิติ จากนั้นจึงคำนวณพลังงานอันตรกิริยาแต่ละจุด (หลายร้อยจุด) ภายในโครงร่างผลึกสามมิติที่เกิดขึ้นรอบๆ แต่ละโมเลกุล สถิติที่ใช้ใน CoMFA คือการวิเคราะห์ partial least squares ซึ่งจะอธิบายในช่วงท้ายของบทความ หลักการวิเคราะห์ข้อมูลด้วย CoMFA ประกอบด้วยขั้นตอนดังต่อไปนี้ (รูปที่ 3)

1. ทำการวิเคราะห์ conformation เพื่อเลือก conformation ที่เหมาะสมของสาร เช่น conformation ที่สารนั้นๆ มีพลังงานต่ำที่สุด จากนั้นทำการซ้อนทับ (alignment) สารในกลุ่มข้อมูลที่ต้องการวิเคราะห์ โดยที่สารแต่ละตัวจะต้องมี conformation ตามที่ได้เลือกไว้
2. คำนวณพลังงาน steric (S1, S2, S3,...S999) และ electrostatic (E1, E2, E3,...E999) ในโครงสร้างผลึกสามมิติของสารแต่ละตัว
3. วิเคราะห์ความสัมพันธ์ของพลังงาน steric (S1, S2, S3,...S999) และ electrostatic กับฤทธิ์ทางชีวภาพ หรือคุณสมบัติที่สนใจ และสร้างแบบจำลอง QSAR/QSPR

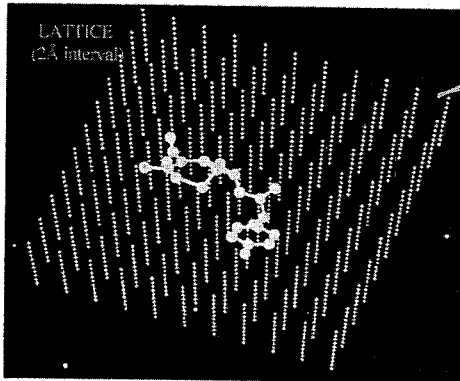
ค่าสัมประสิทธิ์ของพลังงานอันตรกิริยาใน CoMFA จะใช้ในการแสดงรูปร่าง (contour) เพื่อช่วยในการออกแบบยาได้ เช่น Wanchana และคณะ<sup>14</sup> ได้ทำการศึกษาการสร้างแบบจำลองของการ uptake เข้าสู่ CaCO-2 cell ของยาปฏิชีวนะกลุ่มเบต้าแลคแตม 20 ตัว โดยการใช้วิธี CoMFA รูปที่ 4 แสดง contour map ของค่าสัมประสิทธิ์ของพลังงาน steric ซึ่งแสดงให้เห็นว่าบริเวณใดในสูตรโครงสร้างที่ควรมีกกลุ่ม steric และไม่ควรมีกกลุ่ม steric ซึ่งสามารถนำไปเป็นข้อมูลช่วยในการออกแบบยาเพื่อเพิ่มการเข้าสู่เซลล์ได้ นอกจาก CoMFA จะใช้ในการสร้างแบบจำลองของฤทธิ์ทางชีวภาพแล้ว ยังสามารถนำไปใช้สร้างแบบจำลอง 3D-QSAR เพื่อใช้ทำนายคุณสมบัติอื่นๆ เช่น ความเป็นพิษ คุณสมบัติทางเภสัชจลนศาสตร์ การขนส่งยา เป็นต้น

เครือข่ายประสาทเทียม (Artificial neural network, ANN)<sup>11</sup> เป็นโปรแกรมคอมพิวเตอร์ที่ออกแบบขึ้นเพื่อจำลองขบวนการให้และรับข้อมูลของสมองมนุษย์ โดยรวบรวมความรู้จากการค้นพบรูปแบบและความสัมพันธ์ของข้อมูล และการเรียนรู้ (training) ผ่านประสบการณ์ รูปที่ 5 แสดงโครงสร้างพื้นฐานของระบบเครือข่ายประสาทเทียม ANN ที่ถูกสร้างจากหน่วยประสาทเทียมย่อยนับร้อย ซึ่งเรียกว่า artificial neuron หรือ processing element (PE) ต่อเชื่อมกับค่าสัมประสิทธิ์หรือค่าน้ำหนัก (weight) โดยสร้างเป็นเครือข่ายของระบบประสาทและถูกจัดเรียงอย่างเป็นระบบเป็นชั้น ประสิทธิภาพการคำนวณของหน่วยประสาทจะมาจากการเชื่อมต่อกันของตัวหน่วยประสาทในเครือข่าย หน่วยประสาท PE แต่ละตัวจะทำหน้าที่เป็นตัวรับการป้อนข้อมูล (input) จากนั้นจะมีการคำนวณโดยฟังก์ชันการถ่ายโอน (transfer function) และให้ผลลัพธ์ (output) พฤติกรรมของ ANN จะถูกกำหนดโดยฟังก์ชันการถ่ายโอนของหน่วยประสาท ภายใต้กฎของการเรียนรู้ และโครงสร้างของเครือข่าย ANN เป็นระบบที่ขึ้นกับค่าตัวแปร (parameterized system) เนื่องจากค่าน้ำหนักเป็นพารามิเตอร์ที่สามารถปรับค่าได้ในระหว่างการเรียนรู้ ผลรวมของค่าน้ำหนักของตัวรับการป้อนข้อมูลจะทำให้เกิดสัญญาณกระตุ้นระบบประสาทเทียม โดยสัญญาณการกระตุ้นจะถูกส่งผ่านฟังก์ชันการถ่ายโอนเพื่อสร้างผลลัพธ์ การเชื่อมต่อระหว่างหน่วยประสาทจะถูกคำนวณเพื่อหาค่าที่เหมาะสม (optimize)

1. Conformational Analysis and Molecular Alignment



2. Energy Calculation



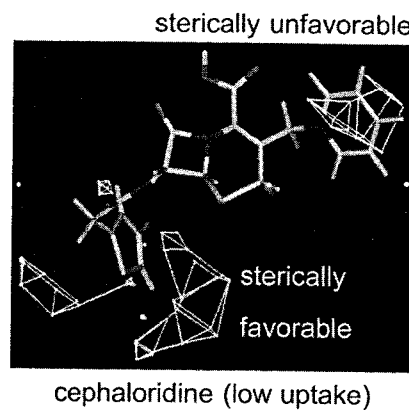
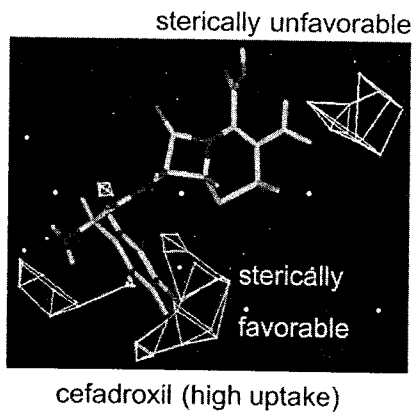
QSAR Table

|       | Uptake | Steric Energy |               |  | Electrostatic Energy |               |  |
|-------|--------|---------------|---------------|--|----------------------|---------------|--|
|       |        | S001          | S002 ... S999 |  | E001                 | E002 ... E999 |  |
| Cpd 1 | 12.3   |               |               |  |                      |               |  |
| Cpd 2 | 0.17   |               |               |  |                      |               |  |
|       |        |               |               |  |                      |               |  |

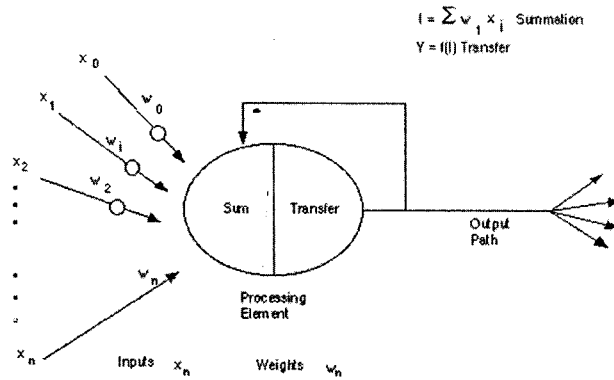
3. QSAR Analysis  
Using PLS\* Method  
\*PLS: Partial Least-Squares

$$Uptake = y + a \times S001 + b \times S002 + \dots + m \times S999 + n \times E001 + \dots + z \times E999$$

รูปที่ 3 แสดงขั้นตอนการวิเคราะห์ข้อมูลด้วยวิธี CoMFA



รูปที่ 4 แสดงโครงสร้างของยา cefadroxil และ cephaloridine ใน contour map ของค่าสัมประสิทธิ์ของพลังงาน steric ของการเข้าสู่ CaCO-2 cell ของยาปฏิชีวนะกลุ่มเบต้าแลคแตม 20 ตัว โดยการใช่วิธี CoMFA โดยสีเขียวแสดงบริเวณที่ควรมีกกลุ่ม steric ในขณะที่สีเหลืองแสดงบริเวณที่ไม่ควรมีกกลุ่ม steric



รูปที่ 5 แสดงโครงสร้างพื้นฐานของระบบเครือข่ายประสาทเทียม

โดยดูจากค่าความแตกต่างระหว่างค่าที่ได้จริงจากการทดลองกับค่าที่ได้จากการคำนวณควรมีค่าน้อยที่สุด และเครือข่ายมีระดับความถูกต้องที่ค่าจำเพาะค่าหนึ่ง ANN มีความสามารถในการทำนายสูง มีพัฒนาการของการเรียนรู้ และสามารถนำไปประยุกต์ใช้ในสถานการณ์ต่างๆ เพื่อสร้างความสัมพันธ์ระหว่างเหตุกับผลได้ จึงทำให้ ANN เป็นที่แพร่หลายในช่วงสิบปีที่ผ่านมา ANN จะให้คำตอบได้ดีโดยเฉพาะอย่างยิ่งในการสร้างแบบจำลองที่มีความสัมพันธ์ที่ไม่เป็นเส้นตรงและค่อนข้างซับซ้อน หรือในสถานะที่มีอันตรกิริยาระหว่างตัวแปรอิสระ อย่างไรก็ตามวิธี ANN มีข้อจำกัดในเรื่องความน่าเชื่อถือ (reliability) และความสามารถในการทำนาย (prediction ability) ซึ่งจะขึ้นกับชนิดและจำนวนของข้อมูลที่ใช้ในการสร้างแบบจำลอง ปัญหาที่มักพบบ่อยในการใช้ ANN คือ overfitting ซึ่งหมายถึงการที่แบบจำลองที่สร้างขึ้น สามารถที่นำมาใช้ในการเรียนรู้ได้ดีแต่นำไปใช้ทำนายข้อมูลอื่นๆ ได้ไม่ดี ซึ่งจะพบเมื่อจำนวนข้อมูลมีน้อย นอกจากนี้ยังมีปัญหาเรื่องการเรียนรู้มากเกินไป (overtraining) ซึ่งทำให้เวลาที่ใช้ในการเรียนรู้เครือข่ายนานขึ้นและนำไปสู่ความไม่คงตัวของระบบ ความสามารถของ ANN จะจำกัดในกรณีที่มีตัวแปรอิสระจำนวนมากในขณะที่จำนวนข้อมูลมีน้อย อย่างไรก็ตามปัญหานี้อาจแก้ไขได้โดยการนำเทคนิค genetic algorithm ซึ่งเป็นขั้นตอนวิธีที่ใช้เลือกตัวแปรอิสระที่เหมาะสมสำหรับการสร้างแบบจำลอง มาใช้ร่วมกับวิธี ANN เป็นต้น

Pattern recognition<sup>10</sup> เป็นวิธีทางคอมพิวเตอร์ที่ใช้ในการหารูปแบบความสัมพันธ์ของกลุ่มข้อมูลที่มีตัวแปรหลายตัวซึ่งไม่เคยทราบรูปแบบความสัมพันธ์มาก่อน โดยการคำนวณหาระยะทางหรือความห่างของหมู่ะตอมในกลุ่มของสารที่ทำการศึกษา และใช้เป็นตัวแปรในการหาความสัมพันธ์กับคุณสมบัติของสารที่สนใจ วิธีนี้จะแตกต่างจากวิธีอื่นๆ คือจะใช้ลักษณะโครงสร้างของโมเลกุล เช่น การปรากฏในโตรเจนอะตอมในโมเลกุล การปรากฏกลุ่มให้และรับไฮโดรเจนอะตอม เป็นต้น โดยไม่เกี่ยวข้องกับการใช้ตัวแปรทางเคมีฟิสิกส์

Factor analysis<sup>15</sup> หรือการวิเคราะห์ปัจจัย เป็นเทคนิคที่ใช้แบ่งกลุ่มตัวแปรหรือรวมตัวแปรที่มีความสัมพันธ์กันไว้ในกลุ่มเดียวกัน อาจทำได้โดยการเปลี่ยนตัวแปรอิสระให้อยู่ในรูปตัวแปรใหม่โดยใช้ส่วนประกอบของตัวแปรดั้งเดิม (original variables) วิธีนี้จึงเน้นที่ความสัมพันธ์ระหว่างตัวแปรที่สร้างขึ้นใหม่ (Z) กับตัวแปรดั้งเดิม (X) ใช้ในการหาความมีมิติภายในหรือมิติอันแท้จริง (intrinsic dimensionality) ของคุณสมบัติทางเคมีที่ได้จากการทดลอง หรืออาจกล่าวอีกนัยหนึ่งว่าเป็นการหาจำนวนปัจจัยพื้นฐานที่มีผลต่อการเปลี่ยนแปลงนั่นเอง Principal component analysis (PCA) ใช้สมมติฐานว่าการเปลี่ยนแปลงของค่าตัวแปรที่ได้จากการทดลองขึ้นอยู่กับค่าตัวแปรหรือคุณสมบัติที่ไม่สามารถสังเกตได้เพียงไม่กี่ชนิด ตัวอย่างเช่น ความแข็งของยาเม็ด (y) ขึ้นกับอุณหภูมิ ( $x_1$ ) และเวลาที่ใช้ในการอบแกรนูล ( $x_2$ ) แรงที่ใช้ในการตอก ( $x_3$ ) ชนิด ( $x_4$ ) และปริมาณสารช่วย ( $x_5$ ) อาจกล่าวได้ว่าความแข็งของเม็ดยา (y) ขึ้นกับ 2 ปัจจัยหลักเท่านั้นคือปัจจัยในเรื่องขบวนการผลิต ( $z_1$ ) และปัจจัยทางด้านตำรับ ( $z_2$ ) เป็นต้น ตามปกติ PCA จะช่วยลดจำนวนมิติของตัวแปรบรรยายโครงสร้างและสร้างตัวแปรบรรยายโครงสร้างใหม่โดยวิธีการหมุนแกนเพื่อเปลี่ยนเมตริกซ์ที่จะนำไปใช้ในการคำนวณ โดยอาจหมุนแกนเพื่อให้ตัวแปรบรรยายโครงสร้างที่ยังคงตั้งฉากกัน (orthogonal) หรือตัวแปรบรรยาย