



วารสาร ไทยไกเซอร์พับลิช
ปีที่ 2 ฉบับเดือนพฤษภาคม 2548 (หน้า 93–106)
เทคโนโลยีและวิชาการ สำหรับการศึกษาต่อเพื่ออาชญาศาสตร์



การศึกษาความสัมพันธ์ระหว่างโครงสร้างและคุณสมบัติของสารเชิงปริมาณ (Quantitative structure property relationship studies)

ภญ. พศ. ดร. สุชาดา พิริยะประสาท

ภาควิชาเทคโนโลยีและวิชาการ 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

- อธิบายโครงสร้างของ QSAR
- ทราบหลักเกณฑ์และเทคนิคต่างๆ (โดยสังเขป) ที่ใช้ในการสร้างแบบจำลองความสัมพันธ์
- ทราบหลักการในการสร้างแบบจำลอง
- ทราบประโยชน์ของ QSAR ในงานวิจัย

บทคัดย่อ

การศึกษาความสัมพันธ์ระหว่างโครงสร้างและคุณสมบัติของสารเชิงปริมาณ (Quantitative Structure Activity Relationships, QSAR) ได้ถูกนำมาใช้ในกระบวนการพัฒนาใหม่มาเป็นเวลานานนับสิบปี 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

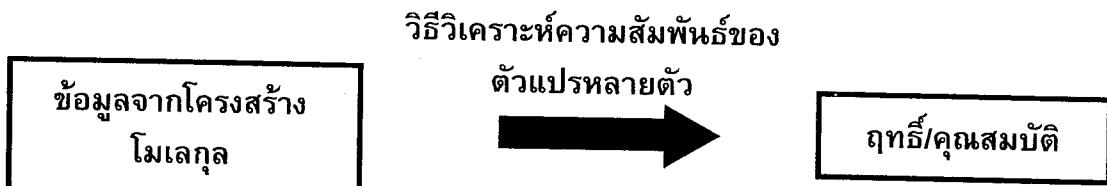
โดยทั่วไปนักเคมีจะใช้วิธีการศึกษาความสัมพันธ์ระหว่างโครงสร้างและฤทธิ์ของสารเชิงปริมาณ (Quantitative Structure Activity Relationships, QSAR) เพื่อศึกษาหาสารที่มีฤทธิ์ทางชีวภาพที่เหมาะสมเพื่อนำไปเป็นสารต้นแบบในการพัฒนา ในปัจจุบันด้วยเทคโนโลยีที่ก้าวหน้าและการใช้เทคนิคทางสถิติและคอมพิวเตอร์ ทำให้สามารถค้นหาสารที่จะนำมาศึกษาเพื่อใช้เป็นสารต้นแบบได้เป็นจำนวนมาก

การออกแบบยาโดยอาศัยโครงสร้างของยา (Structure-based design) เป็นวิธีที่ได้รับการยอมรับกันอย่างแพร่หลายในกระบวนการพัฒนาเภสัชภัณฑ์ใหม่ๆ เทคนิคที่นำมาใช้ไม่เพียงแต่เป็นการทำนายฤทธิ์ของสารเท่านั้น แต่ยังใช้ทำนายคุณสมบัติทางเคมีพิสิกส์และคุณสมบัติอื่นๆ ของสารนั้นๆ ก่อนการสังเคราะห์การออกแบบยาโดยอาศัยโครงสร้างของยา จึงเป็นที่มาของวิธีการศึกษาความสัมพันธ์ระหว่างโครงสร้างและคุณสมบัติของสารเชิงปริมาณ (Quantitative Structure Property Relationships, QSPR)¹ แม้ว่า QSPR จะไม่สามารถแก้ปัญหาความยุ่งยากในเรื่องการใช้เวลาในการค้นหาพารามิเตอร์ที่เหมาะสมสำหรับการสร้างแบบจำลอง QSPR แต่วิธี QSPR มีส่วนช่วยอย่างมากในการลดปริมาณสารที่จะต้องสังเคราะห์ โดยเป็นเครื่องมือเพื่อใช้ในการคัดเลือกสารที่มีคุณสมบัติที่เหมาะสมที่สุดมาสังเคราะห์และทดสอบฤทธิ์ในสัตว์ทดลองและในมนุษย์ต่อไป

โดยหลักการ สูตรโครงสร้างของสารจะมีรหัสข้อมูลในเรื่องคุณสมบัติทางกายภาพ เคมีและชีวภาพของสารนั้นๆ ดังนั้นถ้าสามารถถอดรหัสข้อมูลในเรื่องคุณสมบัติเหล่านี้จากสูตรโครงสร้างทางเคมีได้ย่อมเป็นประโยชน์กับกระบวนการค้นหายาใหม่ ในปี ค.ศ.1968 Crum-Brown&Fraser ได้พิมพ์การศึกษาหาความสัมพันธ์ระหว่างผลทางสรีรวิทยา กับโครงสร้างของสารเป็นครั้งแรก² จากนั้นมีงานวิจัยหลายงานที่ค้นพบความสัมพันธ์ระหว่างคุณสมบัติความชอบ ละลายในไขมัน (lipophilicity) ซึ่งแสดงโดยค่าการละลายหรือค่าสัมประสิทธิ์การแพร่ผ่านในน้ำกับน้ำมันกับผลทางชีวภาพ³⁻⁴ เช่น ความเป็นพิษหรือฤทธิ์ทางเภสัชวิทยา ในปี ค.ศ.1935 Hammett ได้ทำการศึกษาผลของการเปลี่ยนแปลงสูตรโครงสร้างของสารที่มีต่อกลไกการเกิดปฏิกิริยา โดยสร้างแบบจำลองซึ่งแสดงให้เห็นว่าการเกิดปฏิกิริยาทางเคมีโดยทั่วไปจะเกิดขึ้นได้ยากในส่วนของโมเลกุลที่มี aromatic ring ซึ่งเป็นผลมาจากการ resonance effect⁵ การศึกษานี้ได้ถูกศึกษาต่อในปี ค.ศ.1956 โดย Taft ซึ่งอธิบายผลที่แตกต่างกันของคุณสมบัติ electrostatic และ steric ของกลุ่มที่มาแทนที่ในสารที่มีโครงสร้างเป็นไฮdrocarbon (aliphatic hydrocarbon) สายโซ่ยาว⁶ วิวัฒนาการในการศึกษาความสัมพันธ์ระหว่างโครงสร้างกับฤทธิ์ทางชีวภาพได้เกิดขึ้นอีกรอบในช่วงปี ค.ศ.1964 โดย Hansch และคณะได้ค้นพบ Linear Free Energy Relationship (LFER) ซึ่งแบบจำลอง (model) ที่สร้างขึ้นด้วยบันสมมติฐานที่ว่า ฤทธิ์ทางชีวภาพมีความสัมพันธ์กับตัวแปรทางเคมีพิสิกส์ซึ่งแสดงในรูปของ free energy⁷ ดังนั้นต่อมาจึงมีการสร้างแบบจำลองมากมาย โดยการใช้เดสクリปเตอร์ซึ่งในที่นี้คือตัวแปรบรรยายโครงสร้าง (structural descriptor) ในรูปแบบต่างๆ เป็นตัวแปรอิสระ ซึ่งกำหนดได้โดยการสังเกต หรือจากการทดลองและการคำนวณ เช่น ตัวแปรทางเคมีพิสิกส์ ได้แก่ Hammett constant (σ), Hansch substituent constant (π), Taft steric parameter (E_s) เป็นต้น ในระยะหลังการคำนวณตัวแปรบรรยายโครงสร้างจะได้มามากการใช้ทฤษฎี ซึ่งจะเน้นการคำนวณค่าตัวแปรจากสูตรโครงสร้างทางโมเลกุลของสารโดยการประยุกต์ใช้เทคนิคทางสถิติและคอมพิวเตอร์ และการคำนวณโดยคอมพิวเตอร์ในการสร้างแบบจำลองสำหรับคำนวณ ซึ่งทำให้การศึกษาไม่ต้องใช้เวลามากดังเช่นสมัยก่อน

โครงสร้างของ QSPR

QSPR⁸⁻⁹ ประกอบด้วยสามส่วนที่สำคัญคือ 1) ข้อมูลจากโครงสร้างโมเลกุลของสารหรือตัวแปรบรรยายโครงสร้าง 2) คุณสมบัติของสารที่ต้องการศึกษาโดยแสดงเป็นค่าที่สามารถวัดได้ และ 3) วิธีหรือเทคนิคที่ใช้ในการสร้างแบบจำลองความสัมพันธ์ เช่น วิธีวิเคราะห์ความสัมพันธ์ของตัวแปรหลายตัว ดังแสดงในรูปที่ 1 ซึ่งจะกล่าวถึงดังต่อไปนี้



รูปที่ 1 แสดงโครงสร้างของ QSPR

1. ข้อมูลจากโครงสร้างโมเลกุลของสาร

ในการศึกษา QSPR สิ่งแรกที่จะต้องทำการสร้างข้อมูลที่ได้จากส่วนของโครงสร้างของสาร 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 จึงทำให้สามารถคำนวณได้อย่างรวดเร็วจากซอฟต์แวร์เฉพาะเมื่อเปรียบเทียบกับการคำนวณจากโครงสร้างโมเลกุลแบบ 3 มิติ ตัวอย่าง เช่น Weiner indices, Randic indices, Kier and Hall molecular connectivity indices ตัวแปรเหล่านี้จะมีความสัมพันธ์กับคุณสมบัติทางฟิสิกส์บางประการ เช่น ค่าการละลายในน้ำ ขนาดและพื้นที่ผิวของโมเลกุล เป็นต้น กล่าวโดยสรุปว่า ตัวแปรเป็นค่าสังเกต (empirical) ที่ได้จากการทดลอง ในขณะที่ตัวแปรบรรยายโครงสร้างอาจเป็นค่าสังเกตที่ได้จากการทดลอง หรือเป็นค่าที่ได้จากการคำนวณโดยใช้ทฤษฎี (theory) ก็ได้

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

2. คุณสมบัติของสารที่ต้องการศึกษา

คุณสมบัติของสารที่นิยมใช้ในการศึกษา QSPR ได้แก่คุณสมบัติที่สำคัญต่อการค้นหาและพัฒนาใหม่ ซึ่งไม่ได้หมายความเพียงแค่ฤทธิ์ทางชีวภาพเท่านั้น แต่ยังรวมถึงคุณสมบัติทางเคมีฟิสิกส์ต่างๆ ดังแสดงในตารางที่ 1

ตารางที่ 1 แสดงคุณสมบัติทางเคมีฟิสิกส์และฤทธิ์ทางชีวภาพที่ใช้ในการศึกษา QSPR

Physicochemical property	Biological property
Organoleptic properties	Activity
Boiling point	Acute toxicity (LD_{50})
Dissociation constant	Alkylating profile (with DNA)
Viscosity	Bioconcentration
Melting point	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. เทคนิครีวิวที่ใช้ในการสร้างแบบจำลองความสัมพันธ์เชิงปริมาณ

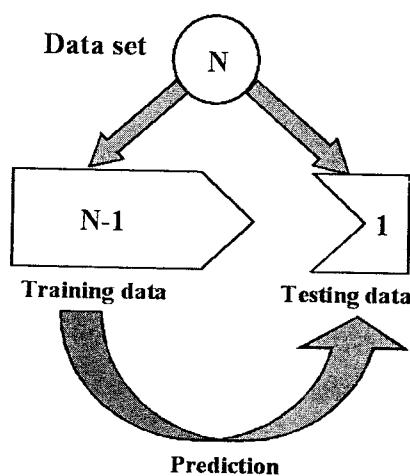
ในการศึกษา QSPR อาจใช้เทคนิคทางสถิติที่ใช้ในการวิเคราะห์ความสัมพันธ์ของตัวแปรหลายตัว (Multivariate Statistical Techniques, MSTs) เพื่อหาความสัมพันธ์ระหว่างคุณสมบัติที่สนใจศึกษา กับตัวแปรบรรยายโครงสร้าง หรือใช้เทคนิคอื่นๆ เช่น comparative molecular field analysis (CoMFA)¹⁰ และ artificial neural network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

ในบรรดาเทคนิคทางสถิติที่ใช้ในการวิเคราะห์ความสัมพันธ์ของตัวแปรหลายตัว การวิเคราะห์ความถดถอยเส้นตรงพหุคุณ (Multiple Linear Regression Analysis, MLRA)¹² เป็นวิธีที่นิยมใช้ในการศึกษา QSPR มากที่สุด เทคนิคนี้จะใช้บวกความสัมพันธ์ระหว่างคุณสมบัติของยา (y) กับตัวแปรบรรยายโครงสร้าง ทั้งหมด (x) ที่ใช้ในการศึกษาได้ในเชิงปริมาณ ซึ่งสามารถแสดงสมการได้ดังนี้

$$y = a_1 x_1 + a_2 x_2 + a_3 x_3 + \dots a_m x_m + e$$

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

โดยจะใช้ค่าสัมประสิทธิ์สหสัมพันธ์พหุคุณ (multiple correlation coefficient, R) Fisher's F ratio Student's t-value และส่วนเบี่ยงเบนมาตรฐาน (standard deviation, s) เพื่อหาค่าความถูกต้องของสมการ MLRA และเพื่อตรวจสอบว่ามีความแตกต่างอย่างมีนัยสำคัญของเกณฑ์ที่ใช้หรือไม่ การจำกัดจำนวนตัวแปรอิสระที่ใช้เมื่อเทียบกับจำนวนข้อมูลทั้งหมดจะช่วยลดโอกาสที่กลุ่มตัวแปรอิสระที่ใช้ในการคำนวณจะมีความสัมพันธ์กันสูง (chance correlation) ซึ่งเป็นปัญหาที่มักพบในการวิเคราะห์ความถูกต้องโดยเส้นตรงพหุคุณ นอกจากนี้สามารถตรวจสอบแบบจำลองที่สร้างขึ้นโดยใช้วิธี leave-one-out (LOO) cross validation ซึ่งเป็นวิธีที่ได้รับการยอมรับอย่างแพร่หลาย หลักการของวิธี LOO cross validation (รูปที่ 2) อาจกล่าวได้โดยสรุปคือสมมติว่ามีข้อมูลทั้งหมด N ข้อมูล ข้อมูลหนึ่งตัวจะถูกนำออกไปเพื่อใช้ทดสอบแบบจำลองที่สร้างขึ้นโดยข้อมูลที่เหลือ (N-1) ทำการคำนวณขั้นตอนลักษณะนี้ 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)

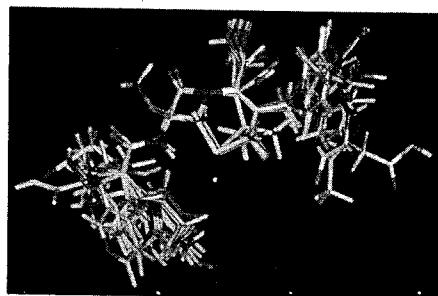
Comparative molecular field analysis (CoMFA)¹³ เป็นวิธีทางความสัมพันธ์ระหว่างโครงสร้างกับฤทธิ์ทางชีวภาพหรือคุณสมบัติอื่นๆ แบบ 3 มิติ (3D-QSAR) ซึ่งรวมหลายๆ เทคนิคเข้าด้วยกัน ได้แก่ QSAR การคำนวณพลังงานระหว่างโมเลกุล pharmacophore mapping และเทคนิคการวิเคราะห์ความสัมพันธ์ของตัวแปรหลายตัว (MSTs) CoMFA จะใช้ค่าตัวแปรต่างๆ ได้แก่ พลังงาน steric และ electrostatic รอบๆ โมเลกุลเป็นตัวแปรอิสระที่ให้ข้อมูลของโครงสร้างเพื่อขอรบกับฤทธิ์ทางชีวภาพหรือคุณสมบัติของสารในเชิงปริมาณ โดยขึ้นกับลักษณะทางเคมีที่จำเพาะในพื้นที่ที่กำหนด พลังงาน steric และ electrostatic สามารถคำนวณโดยการวาง sp^3 hybridized carbon probe atom ที่ตำแหน่งต่างๆ ในโครงสร้างผลึกสามมิติ จากนั้นจึงคำนวณพลังงานอันตรกิริยาแต่ละจุด (หลายร้อยจุด) ภายในโครงสร้างผลึกสามมิติที่เกิดขึ้นรอบๆ แต่ละโมเลกุล สถิติที่ใช้ใน CoMFA คือการวิเคราะห์ partial least squares ซึ่งจะอธิบายในช่วงท้ายของบทความ หลักการวิเคราะห์ข้อมูลด้วย CoMFA ประกอบด้วยขั้นตอนดังต่อไปนี้ (รูปที่ 3)

- ทำการวิเคราะห์ conformation เพื่อเลือก conformation ที่เหมาะสมของสาร เช่น conformation ที่สารนั้นๆ มีพลังงานต่ำที่สุด จากนั้นทำการข้อทับ (alignment) สารในกลุ่มข้อมูลที่ต้องการวิเคราะห์ โดยที่สารแต่ละตัวจะต้องมี conformation ตามที่ได้เลือกไว้
- คำนวณพลังงาน steric (S1, S2, S3,...S999) และ electrostatic (E1, E2, E3,...E999) ในโครงสร้างผลลัพธ์ สามมิติของสารแต่ละตัว
- วิเคราะห์ความสัมพันธ์ของพลังงาน steric (S1, S2, S3,...S999) และ electrostatic กับฤทธิ์ทางชีวภาพ หรือคุณสมบัติที่สนใจ และสร้างแบบจำลอง QSAR/QSPR

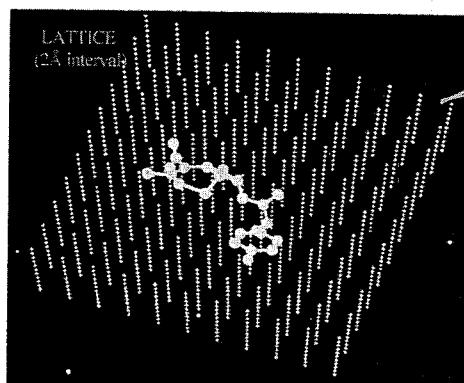
ค่าสัมประสิทธิ์ของพลังงานอันตรกิริยาใน CoMFA จะใช้ในการแสดงรูปร่าง (contour) เพื่อช่วยในการออกแบบยาได้ เช่น Wanchana และคณะ¹⁴ ได้ทำการศึกษาการสร้างแบบจำลองของการ uptake เข้าสู่ CaCO-2 cell ของยาปฏิชีวนะกลุ่มเบต้าแอลด์แอล 20 ตัว โดยการใช้วิธี CoMFA รูปที่ 4 แสดง contour map ของค่าสัมประสิทธิ์ของ พลังงาน steric ซึ่งแสดงให้เห็นว่าบริเวณใดในสูตรโครงสร้างที่ควรเมิกกลุ่ม steric และไม่ควรเมิกกลุ่ม steric ซึ่งสามารถนำไปเป็นข้อมูลช่วยในการออกแบบยาเพื่อเพิ่มการเข้าสู่เซลล์ได้ นอกจาก CoMFA จะใช้ในการสร้างแบบจำลองของ ฤทธิ์ทางชีวภาพแล้ว ยังสามารถนำไปใช้สร้างแบบจำลอง 3D-QSAR เพื่อใช้ทำนายคุณสมบัติอื่นๆ เช่น ความเป็นพิษ คุณสมบัติทางเภสัชศาสตร์ การขนส่งยา เป็นต้น

เครือข่ายประสาทเทียม (Artificial neural network, ANN)¹¹ เป็นโปรแกรมคอมพิวเตอร์ที่ออกแบบขึ้นเพื่อจำลอง ขบวนการให้และรับข้อมูลของสมองมนุษย์ โดยรวมความรู้จากการค้นพบรูปแบบและความสัมพันธ์ของข้อมูล และ การเรียนรู้ (training) ผ่านประสบการณ์ รูปที่ 5 แสดงโครงสร้างพื้นฐานของระบบเครือข่ายประสาทเทียม ANN ที่ถูก สร้างจากหน่วยประสาทเทียมย่อยนับร้อย ซึ่งเรียกว่า artificial neuron หรือ processing element (PE) ต่อเชื่อมกับ ค่าสัมประสิทธิ์หรือค่าน้ำหนัก (weight) โดยสร้างเป็นเครือข่ายของระบบประสาทและถูกจัดเรียงอย่างเป็นระบบเป็นชั้น ประสาทชั้น การคำนวณของหน่วยประสาทจะมาจากการเชื่อมต่อ กันของตัวหน่วยประสาทในเครือข่าย หน่วยประสาท PE แต่ละตัวจะทำหน้าที่เป็นตัวรับการป้อนข้อมูล (input) จากนั้นจะมีการคำนวณโดยฟังก์ชันการถ่ายโอน (transfer function) และให้ผลลัพธ์ (output) พฤติกรรมของ ANN จะถูกกำหนดโดยฟังก์ชันการถ่ายโอนของหน่วยประสาท ภายใต้กฎของการเรียนรู้ และโครงสร้างของเครือข่าย ANN เป็นระบบที่ขึ้นกับค่าตัวแปร (parameterized system) เนื่องจากค่าน้ำหนักเป็นพารามิเตอร์ที่สามารถปรับค่าได้ในระหว่างการเรียนรู้ รวมของค่าน้ำหนักของตัวรับการ ป้อนข้อมูลจะทำให้เกิดสัญญาณกระแสตุ่นระบบประสาทเทียม โดยสัญญาณการกระแสตุ่นจะถูกส่งผ่านฟังก์ชันการถ่ายโอน เพื่อสร้างผลลัพธ์ 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						

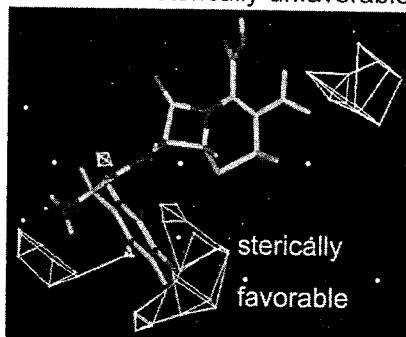
3. QSAR Analysis
Using PLS* Method

*PLS; Partial Least-Squares

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

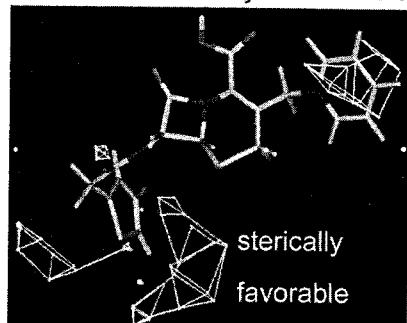
รูปที่ 3 แสดงขั้นตอนการวิเคราะห์ข้อมูลด้วยวิธี CoMFA

sterically unfavorable



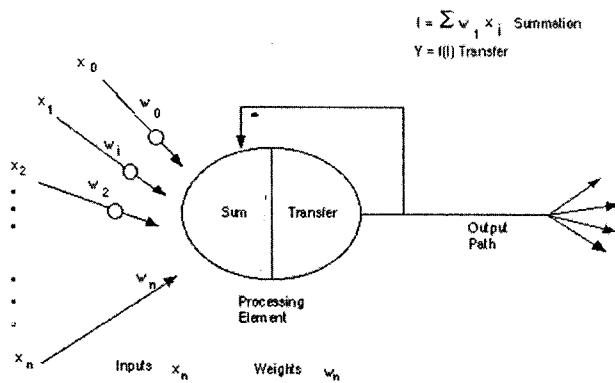
cefadroxil (high uptake)

sterically unfavorable



cephaloridine (low uptake)

รูปที่ 4 แสดงโครงสร้างของยา cefadroxil และ cephaloridine ใน contour map ของค่าสัมประสิทธิ์ของพลังงาน steric ของการเข้าสู่ CaCO-2 cell ของยาปฏิชีวนะกลุ่มเบต้าแลคตาม 20 ตัว โดยการใช้วิธี CoMFA โดยสีเขียวแสดงบริเวณที่ความมีกลุ่ม steric ในขณะที่สีเหลืองแสดงบริเวณที่ไม่ความมีกลุ่ม steric



รูปที่ 5 แสดงโครงสร้างพื้นฐานของระบบเครือข่ายประสาทเทียม

โดยดูจากค่าความแตกต่างระหว่างค่าที่ได้จริงจากการทดลองกับค่าที่ได้จากการคำนวณคอมพิวเตอร์ที่สุด และเครือข่ายมีระดับความถูกต้องที่ค่าจำเพาะค่าหนึ่ง ANN มีความสามารถในการคำนวณสูง มีพัฒนาการของการเรียนรู้ และสามารถนำไปประยุกต์ใช้ในสถานการณ์ต่างๆ เพื่อสร้างความสัมพันธ์ระหว่างเหตุกับผลได้ จึงทำให้ ANN เป็นที่แพร่หลายในช่วงลิบปีที่ผ่านมา ANN จะให้คำตอบได้ดีโดยเฉพาะอย่างยิ่งในการสร้างแบบจำลองที่มีความสัมพันธ์ที่ไม่เป็นเส้นตรงและค่อนข้างซับซ้อน หรือในสภาวะที่มีอันตรายร้ายแรงตัวแปรอิสระ อย่างไรก็ตามวิธี ANN มีข้อจำกัด ในเรื่องความน่าเชื่อถือ (reliability) และความสามารถในการทำงาน (prediction ability) ซึ่งจะขึ้นกับชนิดและจำนวนของข้อมูลที่นำมาใช้ในการสร้างแบบจำลอง ปัญหาที่มักพบบ่อยในการใช้ ANN คือ overfitting ซึ่งหมายถึงการที่แบบจำลองที่สร้างขึ้น สามารถที่นำมาใช้ในการเรียนรู้ได้ดีแต่นำไปใช้ทำนายข้อมูลอื่นๆ ได้ไม่ดี ซึ่งพบเมื่อจำนวนข้อมูลมีน้อย นอกจากนี้ยังมีปัญหาเรื่องการเรียนรู้มากเกินไป (overtraining) ซึ่งทำให้เวลาที่ใช้ในการเรียนรู้เครือข่ายนานขึ้นและนำไปสู่ความไม่คงตัวของระบบ ความสามารถของ ANN จะจำกัดในการแก้ไขที่มีตัวแปรอิสระจำนวนมากในขณะที่จำนวนข้อมูลมีน้อย อย่างไรก็ตามปัญหานี้อาจแก้ไขได้โดยการนำเทคนิค genetic algorithm ซึ่งเป็นขั้นตอนวิธีที่ใช้เลือกตัวแปรอิสระที่เหมาะสมสำหรับการสร้างแบบจำลอง มาใช้ร่วมกับวิธี ANN เป็นต้น

Pattern recognition¹⁰ เป็นวิธีทางคณิตศาสตร์ที่ใช้ในการหารูปแบบความสัมพันธ์ของกลุ่มข้อมูลที่มีตัวแปรหลายตัวซึ่งไม่เคยทราบรูปแบบความสัมพันธ์มาก่อน โดยการคำนวณหาระยะทางหรือความห่างของหมู่อะตอมในกลุ่มของสารที่ทำการศึกษา และใช้เป็นตัวแปรในการหาความสัมพันธ์กับคุณสมบัติของสารที่สนใจ วิธีนี้จะแตกต่างจากวิธีอื่นๆ คือจะใช้ลักษณะโครงสร้างของโมเลกุล เช่น การประภูมิในโตรเจนอะตอมในโมเลกุล การประภูมิกลุ่มให้และรับไฮโดรเจนอะตอม เป็นต้น โดยไม่เกี่ยวข้องกับการใช้ตัวแปรทางเคมีฟิสิกส์

Factor analysis¹⁵ หรือการวิเคราะห์ปัจจัย เป็นเทคนิคที่ใช้แบ่งกลุ่มตัวแปรหรือรวมตัวแปรที่มีความสัมพันธ์กันไว้ในกลุ่มเดียวกัน อาจทำได้โดยการเปลี่ยนตัวแปรอิสระให้อยู่ในรูปตัวแปรใหม่โดยใช้ส่วนประกอบของตัวแปรดั้งเดิม (original variables) วิธีนี้จึงเน้นที่ความสัมพันธ์ระหว่างตัวแปรที่สร้างขึ้นใหม่ (Z) กับตัวแปรดั้งเดิม (X) ใช้ในการหาความมีมิติภายในหรือมิติอันแท้จริง (intrinsic dimensionality) ของคุณสมบัติทางเคมีที่ได้จากการทดลอง หรืออาจกล่าวอีกนัยหนึ่งว่าเป็นการทำให้จำนวนปัจจัยพื้นฐานที่มีผลต่อการเปลี่ยนแปลงนั้นลง Principal component analysis (PCA) ใช้สมมติฐานว่าการเปลี่ยนแปลงของค่าตัวแปรที่ได้จากการทดลองขึ้นอยู่กับค่าตัวแปรหรือคุณสมบัติที่ไม่สามารถสังเกตได้เพียงไม่กี่ชนิด ตัวอย่างเช่น ความแข็งของยาเม็ด (y) ขึ้นกับอุณหภูมิ (x₁) และเวลาที่ใช้ในการอบแห้ง (x₂) แรงที่ใช้ในการตอก (x₃) ชนิด (x₄) และปริมาณสารช่วย (x₅) อาจกล่าวได้ว่าความแข็งของเม็ดยา (y) ขึ้นกับ 2 ปัจจัยหลักเท่านั้น คือปัจจัยในเรื่องของขนาดการผลิต (z₁) และปัจจัยทางด้านตัวรับ (z₂) เป็นต้น ตามปกติ PCA จะช่วยลดจำนวนมิติของตัวแปรบรรยายโดยโครงสร้างและสร้างตัวแปรบรรยายโดยโครงสร้างใหม่โดยวิธีการหมุนแกนเพื่อเปลี่ยนเมตริกซ์ที่จะนำไปใช้ในการคำนวณ โดยอาจหมุนแกนเพื่อให้ตัวแปรบรรยายโดยโครงสร้างที่ยังคงตั้งฉากกัน (orthogonal) หรือตัวแปรบรรยาย

โครงสร้างต่างๆ เป็นอิสระต่อกัน แต่ยังคงมีข้อมูลของตัวแปรบรรยายโครงสร้างดังเดิมอยู่ วิธีนี้ประสบความสำเร็จในการนำไปใช้กับกลุ่มข้อมูลที่มีโครงสร้างคล้ายคลึงกันเมื่อนำไปทดสอบฤทธิ์ทางชีวภาพในระบบต่างๆ

Cluster analysis เป็นเทคนิคที่ใช้ในการแบ่งข้อมูลออกเป็นกลุ่มย่อย โดยให้ข้อมูลที่มีความคล้ายกัน (similarity) อยู่ในกลุ่มเดียวกันและบอกว่ากลุ่มของข้อมูลเหล่านี้เหมือนกับข้อมูลกลุ่มอื่นหรือไม่ นอกจากนี้ยังสามารถแสดงจำนวนกลุ่มที่แตกต่างกันของข้อมูลที่นำมารวิเคราะห์ เช่น ใช้ในการแบ่งกลุ่มข้อมูลที่มีคุณสมบัติทางกายภาพเหมือนกัน หรือสามารถนำมาใช้แยกสารที่ไม่มีฤทธิ์ทางชีวภาพ ออกจากสารที่มีฤทธิ์ทางชีวภาพได้ เทคนิคนี้ไม่จำเป็นต้องทราบจำนวนกลุ่มย่อยมาก่อนว่าควรแบ่งเป็นกี่กลุ่มและแต่ละข้อมูลอยู่ในกลุ่มใด ในขณะที่ Discriminant analysis เป็นเทคนิคการแบ่งกลุ่มข้อมูลโดยใช้หลักเกณฑ์ของการวิเคราะห์แบบคัดถอยเส้นตรงพหุคูณ โดยตัวแปรตามเป็นตัวแปรเชิงกลุ่ม ตัวแปรต้นเป็นตัวแปรเชิงปริมาณ เทคนิคนี้จะต้องทราบมาก่อนว่าข้อมูลอยู่ในกลุ่มใดและมีกี่กลุ่ม เช่น มีข้อมูลยา 3 กลุ่มคือยาที่ละลายน้ำได้มาก กลุ่มยาที่ละลายน้ำได้ปานกลาง และกลุ่มยาที่ละลายน้ำน้อย วัตถุประสงค์ของเทคนิคนี้คือการศึกษาว่าตัวแปรใดที่ทำให้ข้อมูลอยู่ต่างกลุ่มกัน โดยศึกษาหาความสัมพันธ์เชิงเส้น จากนั้น จึงนำสมการเชิงเส้นไปใช้ในการพยากรณ์ว่าข้อมูลใหม่ควรอยู่ในกลุ่มใด

Evolution algorithms เป็นเทคนิคที่นักวิทยาศาสตร์ฐานห้องใช้ความสัมพันธ์เชิงเส้น (polynomial) แต่เป็นเทคนิคที่มีอันดับ polynomial ที่สูงกว่าเพื่อสร้างรูปแบบความสัมพันธ์ที่ดีกว่าการใช้วิธีดั้งเดิม เช่น genetic algorithms (GA)¹⁶ ใช้หลักการของวิวัฒนาการ โดยการเลียนแบบหลักการใช้วิธีการคัดเลือกตามธรรมชาติ (natural selection) ในการแก้ปัญหาการเลือกกลุ่มตัวแปรที่เหมาะสมโดยให้ตัวแปรเปรียบเสมือนยีน วิธีการนี้ประกอบด้วยขั้นตอนที่สำคัญสามขั้นตอน คือ การคัดเลือก การสับเปลี่ยนค่าตัวแปร (crossover) และการกลายพันธุ์ (mutation) โดยกลุ่มตัวแปรที่ใช้ในการทำนายผลได้ดีกว่าจะถูกคัดเลือกไว้ในประชากรรุ่น (generation) ถัดไป ในขณะที่กลุ่มตัวแปรที่ให้ผลในการทำนายไม่ดีจะถูกคัดทิ้ง การวิเคราะห์ด้วยวิธี GA จะเริ่มจากการสร้างประชากร (population) ของกลุ่มตัวแปรอิสระ แต่ละกลุ่มย่อยจะถูกใส่รหัสตัวยับบตัวเลขสองตัวคือ “1” กับ “0” โดยที่ความยาวของรหัสจะเท่ากับจำนวนตัวแปรทั้งหมด ค่า “1” หมายความว่าตัวแปรอิสระนั้นมีความสำคัญ ในขณะที่ค่า “0” หมายความว่าตัวแปรอิสระนั้นไม่สำคัญไม่จำเป็นต้องนำมาใช้ในการสร้างแบบจำลอง จากนั้นกำหนดค่า fitness เพื่อใช้ในการแยกความแตกต่างของแต่ละกลุ่มย่อยว่ากลุ่มย่อยใดมีกลุ่มของตัวแปรที่เหมาะสมกว่า โดยอาจใช้เทคนิควิเคราะห์ตัวแปรหลายตัว เช่น วิธีการวิเคราะห์ความถดถอยเส้นตรงพหุคูณ โดยคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ของค่าที่ทำนายได้ (predictive correlation coefficients, r^2) เพื่อใช้เป็นค่า fitness เป็นต้น ในการคัดเลือกกลุ่มตัวแปรที่เหมาะสมด้วยเทคนิค GA ตัวอย่าง เช่น ถ้ามีตัวแปรทั้งหมด 5 ตัว กลุ่มย่อยของตัวแปรทั้งหมดที่เป็นไปได้มีทั้งหมด 2^5 กลุ่มย่อย ได้แก่ 01110, 00010, 00110 เป็นต้น กลุ่มย่อยของตัวแปรสองกลุ่มจะถูกคัดเลือกขึ้นมาอย่างสุ่มโดยดูจากค่า fitness จากนั้นจะมีการสับเปลี่ยนค่าตัวแปรและทำการกลายพันธุ์ตามสภาวะที่กำหนด จะได้กลุ่มย่อยของตัวแปรรุ่นใหม่เกิดขึ้น ในแต่ละรุ่นจะทำการคำนวณซ้ำจนกระทั่งได้ประชากรครบตามที่กำหนดไว้

Partial least square (PLS) analysis¹⁷ เป็นวิธีที่ใช้ในการสร้างความสัมพันธ์ระหว่างกลุ่มของตัวแปรอิสระกับฤทธิ์เชิงเส้นตรงคล้ายกับวิธี multiple linear regression (MLR) แต่มีข้อแตกต่างคือ PLS จะใช้ PCA มาช่วยในการหาจำนวนปัจจัย (factor) ที่สามารถใช้แทนตัวแปรทั้งหมด เพื่อเป็นการดึงรายละเอียดจากตัวแปรมาไว้ในปัจจัยจากนั้นจึงหาความสัมพันธ์เชิงเส้นตรงระหว่างปัจจัยกับฤทธิ์ ดังนั้น PLS จึงให้ผลลัพธ์ที่วิธี MLR โดยเฉพาะอย่างยิ่งในกรณีที่มีตัวแปรอิสระจำนวนมากในขณะที่ข้อมูลที่ต้องการศึกษามีจำนวนจำกัด เนื่องจากสามารถแก้ปัญหาการที่ตัวแปรอิสระของภาระวิเคราะห์ความถดถอยมีความสัมพันธ์กัน (multicollinearity) ซึ่งเป็นปัญหาที่มักพบในวิธี MLR

หลักการสร้างแบบจำลอง QSPR

หลักการสร้างแบบจำลอง¹⁸ เพื่อทำนายคุณสมบัติต่างๆ มีดังต่อไปนี้

1. ควรใช้ข้อมูลที่มาจากแหล่งเดียวกัน ไม่ควรมีความแตกต่างของแหล่งข้อมูลที่จะนำมาใช้ในการสร้างแบบจำลอง
2. ควรใช้ข้อมูลที่มีจำนวนมากเพียงพอ มีความหลากหลายในโครงสร้างทางโมเลกุลของสาร และครอบคลุมค่าที่ต้องการศึกษาในช่วงกว้าง
3. ควรมีการประยุกต์ใช้เครื่องมือชนิดต่างๆ เช่น การใช้ซอฟต์แวร์ Molconn-Z ในการสร้างตัวแปรบรรยายโครงสร้างในระดับโมเลกุล
4. เลือกใช้วิธีการที่ถูกต้องเหมาะสมในการจำกัดตัวแปรอิสระที่มีความสัมพันธ์กันสูง
5. ในการสร้างความสัมพันธ์ระหว่างโครงสร้างของสารกับฤทธิ์หรือคุณสมบัติที่ต้องการ ควรใช้ขั้นตอนวิธีชนิดต่างๆ ร่วมกัน เช่น การใช้ factor analysis เพื่อลดจำนวนตัวแปรอิสระ ร่วมกับการใช้การวิเคราะห์แบบเครือข่ายประสาทเทียม เป็นต้น วิธีที่เลือกใช้ควรให้แบบจำลองที่สามารถให้ผลลัพธ์ที่สามารถตรวจสอบได้
6. ควรใช้วิธีทางสถิติเพื่อทดสอบ ตรวจสอบความถูกต้อง และเลือกแบบจำลองเพื่อการนำไปใช้
7. สามารถประยุกต์แบบจำลองที่สร้างขึ้นกับอุปกรณ์คอมพิวเตอร์ที่มีจำหน่ายในท้องตลาดได้ เพื่อให้ง่ายต่อการใช้ประโยชน์

ซอฟต์แวร์ที่ใช้ในการศึกษา QSPR

การพัฒนาแบบจำลอง QSPR มุ่งเน้นที่จะหาความสัมพันธ์ระหว่างโครงสร้างและคุณสมบัติของสารเชิงปริมาณ โดยใช้ตัวแปรบรรยายโครงสร้างที่ได้จากโครงสร้างหรือจากคุณสมบัติทางเคมีพิสิกส์หนึ่งตัวหรือมากกว่านั้น แต่ไม่ใช่เรื่องง่ายที่จะสร้างตัวแปรบรรยายโครงสร้างจากโครงสร้างรวมถึงการเลือกตัวแปรบรรยายโครงสร้างที่เหมาะสมในการสร้างความสัมพันธ์ โดยเฉพาะอย่างยิ่งเมื่อมีกลุ่มของตัวแปรบรรยายโครงสร้างจำนวนมาก ในปัจจุบันมีงานวิจัยพยาบาลที่จะพัฒนาโปรแกรมซอฟต์แวร์เพื่อใช้ในการคำนวณค่าตัวแปรอิสระชนิดต่างๆ บางโปรแกรมอาจรวมวิธีวิเคราะห์ความสัมพันธ์ของตัวแปรหลายตัวเพื่อสร้างแบบจำลองทางคณิตศาสตร์ที่จะเชื่อมโยงตัวแปรบรรยายโครงสร้างเข้ากับคุณสมบัติที่สนใจโดยตรง ปัจจุบันมีทั้งโปรแกรมที่มีขายในท้องตลาดและที่กำลังพัฒนา¹⁹ ตัวอย่างเช่น Metabolexpert (Compudrug) META (Multicase) Meteor (Lhasa) บางโปรแกรมสามารถเชื่อมโยงกับการคาดประมาณความเป็นพิษของยาได้ เช่น DEREK (Lhasa) และ MCASE(-ES)/ MC4PC (Multicase) เป็นต้น

การประยุกต์ใช้วิธี QSPR ในงานวิจัย

ตัวอย่างของกลุ่มยาหรือสารเคมีที่มีการศึกษาความสัมพันธ์ระหว่างความชอบละลายในไขมันกับโครงสร้าง ได้แก่ barbiturates (Wong 1988)²⁰ chlorobenzenes และ polychlorinated biphenyl compounds (de Brujin 1990)²¹ teicoplanin antibiotic derivatives (Altomare 1994)²² และ peptides (Akamatsu 1994)²³ เป็นต้น นอกจากนี้ยังมีการศึกษา quantitative structure–pharmacokinetic relationships (QSPKR) ของยา adenosine A1 โดยใช้ค่า Cl, Vd, และ fraction unbound เป็นตัวแปรอิสระ และใช้ค่า log P (PALLAS) เป็นตัวแปรตาม โดยใช้สถิติ CoMFA, PLS ในการสร้างแบบจำลองและตรวจสอบความถูกต้องของความสามารถในการทำนายโดยวิธี cross validation ของ adenosine A1 (van de Graaf 1999)²⁴ และในการวิเคราะห์กลุ่มยาอื่นๆ เช่น barbiturates (Blakey 1997)²⁵ และ cephalosporins (Radouane 1996)²⁶ เป็นต้น

บทสรุป

ในการศึกษา QSPR เมื่อพับความสัมพันธ์ระหว่างโครงสร้างกับคุณสมบัติของสาร เราสามารถที่จะคัดเลือกสารที่มีคุณสมบัติที่เหมาะสม เช่น มีการละลายดี ดูดซึมเข้าสู่ร่างกายได้ดี เพื่อนำไปเป็นสารต้นแบบในการพัฒนายาจากสารจำนวนมาก ทำให้ไม่จำเป็นต้องสังเคราะห์สารทั้งหมดเพื่อมาทดสอบคุณสมบัติที่ต้องการดังกล่าว โดยการใช้คอมพิวเตอร์เลือกโครงสร้างที่มีคุณสมบัติตามที่ต้องการได้อย่างรวดเร็ว นอกจากนี้ยังใช้เป็นเครื่องมือในการวินิจฉัยได้ เช่น ให้ข้อมูลในเรื่องชนิดของแรงที่ทำให้เกิดอันตราย ลักษณะของ receptor และกลไกในการเกิดปฏิกิริยา เป็นต้น

ข้อจำกัดในการใช้วิธี QSPR คือถึงแม้ว่าจะมีข้อมูลจากโครงสร้างโมเลกุลของสารซึ่งนำมาใช้เป็นตัวแปรอิสระในการศึกษาหาความสัมพันธ์จำนวนมาก แต่ข้อมูลเหล่านี้ยังไม่สามารถอธิบายการเกิดอันตรายที่สำคัญบางอย่างได้ เช่น การซึมผ่านของยาเข้าสู่เมมเบรน ความแข็งแรงของการเกิดพันธะไฮโดรเจน หรือการเกิด steric interaction กับอวัยวะเป้าหมาย เป็นต้น ดังนั้นการประยุกต์ใช้วิธี QSPR อาจไม่สามารถอธิบายการทำงานของยาในร่างกายได้ทั้งหมด

เอกสารอ้างอิง

- Waterbeemd VH, Smith DA, Beaumont K, et al. Property-based design: Optimisation of drug absorption and pharmacokinetics. *J. Med. Chem.* 2001; 44: 1313–33.
- Crum-Brown A, Fraser TR. On the connection between chemical constitution and physiological action. Part 1. On the physiological action of the salts of the ammonium bases, derived from Strychnia, Brucia, Thebia, Codeia, Morphia, and Nicotia. *Trans. R. Soc. Edinburgh.* 1968; 25: 151-203.
- Glave WR, Hansch C. Relationship between lipophilic character and anesthetic activity. *J. Pharm. Sci.* 1972; 61(4): 589-91.
- Coats EA, Milstein SR, Pleiss MA, et al. Comparative analysis of the cytotoxicity of substituted [phenylglyoxal bis(4-methyl-3-thiosemicarbazone)] copper(II) chelates. 2. Parabolic correlations and their implications for selective toxicity. *J. Med. Chem.* 1978; 21(8): 804-9.
- Hammett LP. Some relations between reaction rates and equilibrium constants. *Chem. Rev.* 1935; 17: 125-36.
- Taft RW. Separation of polar, steric and resonance effects in reactivity. In: Newman MS (ed) *Steric effect in organic chemistry*. New York: Wiley, 1956: 556-675.
- Hansch C, Fujita T. Rho-sigma-pi analysis: A method for the correlation of biological activity and chemical structure. *J. Am. Chem. Soc.* 1964; 86: 1616-26.
- Grover M, Singh B, Bakshi M, et al. Quantitative structure–property relationships in pharmaceutical research – Part 1. *PSTT.* 2000; 3: 28-35.
- Grover M, Singh B, Bakshi M, et al. Quantitative structure–property relationships in pharmaceutical research – Part 2. *PSTT.* 2000; 3: 50-7.
- Kubinyi H. QSAR and 3D QSAR in drug design Part 1: Methodology. *Drug Discovery Today* 1997; 2: 457–67.
- Agatonovic-Kustrin S, Beresford R. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. *J. Pharm. Biomed. Anal.* 2000; 22(5): 717-27.

12. Dillon WR, Goldstein M. Multiple regression, In: Dillon WR, Goldstein M (ed) *Multivariate analysis: method and applications*. New York: John Wiley& Sons, 1984: 209-50.
13. Cramer RD, Patterson DE, Bunce JD. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* 1988; 110(18): 5959-67.
14. Wanchana S, Yamashita F, Hara H, et al. Two- and three-dimensional QSAR of carrier-mediated transport of beta-lactam antibiotics in Caco-2 cells. *J. Pharm. Sci.* 2004; 93(12): 3057-65.
15. Dillon WR, Goldstein M. Factor analysis, In: Dillon WR, Goldstein M (ed) *Multivariate analysis: method and applications*. New York: John Wiley& Sons, 1984: 53-106.
16. Goldberg DE. A gentle introduction to genetic algorithms, In: Goldberg DE (ed) *Genetic algorithms in search, optimization, and machine learning*. New York: Addison-Wesley, 1989: 1-22.
17. Geladi P, Kowalski BR. Partial least-squares regression: a tutorial. *Anal. Chim. Acta*. 1986; 185:1-17.
18. Ekins S, Boulanger B, Swaan PW, et al. Towards a new age of virtual ADME/TOX and multidimensional drug discovery. *J. Comput.-Aided Mol. Des.* 2002; 16: 381-401.
19. Langowski J, Long A. Computer systems for the prediction of xenobiotic metabolism. *Adv. Drug Deliv. Rev.* 2002; 54: 407-15.
20. Wong O, McKeown RH. Substituent effects on partition coefficients of barbituric acids. *J. Pharm. Sci.* 1988; 77: 926-31.
21. de Brujin J, Hermens J. Relationship between octanol/water partition coefficients and total molecular surface area and total molecular volume of hydrophobic organic chemicals. *Quant. Struct-Act. Relat.* 1990; 9: 11-20.
22. Altomare C, Carotti A, Cellamare S, et al. Lipophilicity of teicoplanin antibiotics as assessed by reversed phase high-performance liquid chromatography: quantitative structure-property and structure-activity relationships. *J. Pharm. Pharmacol.* 1994; 46: 994-9.
23. Akamatsu M, Katayama T, Kishimoto D, et al. Quantitative analyses of the structurehydrophobicity relationship for N-acetyl di- and tripeptide amides. *J. Pharm. Sci.* 1994; 83: 1026-33.
24. Van der Graaf PH, Nilsson J, Van Schaick EA, et al. Multivariate quantitative structurepharmacokinetic relationships (QSPKR) analysis of adenosine A1 receptor agonists in rat. *J. Pharm. Sci.* 1999; 88: 306-12.
25. Blakey GE, Nestorov IA, Arundel PA, et al. Quantitative structure-pharmacokinetics relationships: I. Development of a whole-body physiologically based model to characterize changes in pharmacokinetics across a homologous series of barbiturates in the rat. *J. Pharmacokin. Biopharm.* 1997; 25: 277-312.
26. Radouane A, Pehourcq F, Tramu G, et al. Influence of lipophilicity on the diffusion of cephalosporins into the cerebrospinal fluid. *Fundam. Clin. Pharmacol.* 1996; 10: 309-13.

คำถาม

1. ข้อใดไม่ใช่ส่วนประกอบของ QSAR

1. การซ้อนทับของโมเลกุล
2. คุณสมบัติที่ต้องการศึกษา
3. ข้อมูลที่ได้จากโครงสร้างของสาร
4. วิธีวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรหลายตัว
5. ข้อ 1 และ 4

2. ข้อใดเป็นเทคนิคที่สามารถนำมาใช้ในการสร้างแบบจำลองความสัมพันธ์

1. Multiple linear regression
2. Pattern recognition
3. Partial least squares
4. Artificial neural network
5. ถูกทุกข้อ

3. ข้อใดไม่ถูกต้องเกี่ยวกับ Comparative molecular field analysis

1. เป็นวิธี 3D-QSAR
2. ต้องการการซ้อนทับของโมเลกุลก่อนการคำนวณ
3. ใช้ steric และ electrostatic energy เป็นเดรสคริปเตอร์ระดับโมเลกุล
4. ใช้ Artificial neural network เป็นเครื่องมือในการสร้างแบบจำลอง
5. ข้อ 3 และ 4

4. ข้อใดเป็นตัวแปรที่แสดงถึงความชอบละลายในไขมัน

1. $\log P_{o/w}$
2. $\log D$
3. $\log K_w$
4. เลขพาร์เซ่น 1 และ 2
5. ทั้งข้อ 1, 2 และ 3

5. ตัวแปรในข้อใดที่ไม่ใช่ตัวแปร 3 มิติ

1. Orbital electron densities
2. Randic indices
3. Molecular polarizabilities
4. Atom-atom polarizabilities
5. Quantum molecular energies

6. ข้อใดไม่ถูกต้องเกี่ยวกับการสร้างแบบจำลอง

1. ข้อมูลที่นำมาใช้ในการสร้างแบบจำลองความจากแหล่งเดียวกัน
2. สามารถใช้วิธี leave-one-out cross validation ในการตรวจสอบความถูกต้องของแบบจำลอง
3. ควรเลือกใช้ตัวแปรอิสระที่มีความสัมพันธ์กันสูง
4. ควรมีความหลากหลายในโครงสร้างของสาร
5. มีความครอบคลุมค่าที่ต้องการศึกษาในช่วงกว้าง

7. ข้อใดเป็นข้อดีของการศึกษา QSPR

1. ใช้ในการสร้างแบบจำลองเพื่อใช้ในการทำนายคุณสมบัติของสารที่ต้องการศึกษาได้
2. ช่วยลดปริมาณสารที่ต้องสังเคราะห์-
3. เป็นเครื่องมือในการคัดเลือกสารที่มีคุณสมบัติที่เหมาะสม
4. เฉพาะข้อ 1 และ 3
5. ถูกทุกข้อ

จงใช้ตัวเลือกต่อไปนี้ตอบคำถามข้อ 8-10

1. genetic algorithm
2. discriminant analysis
3. cluster analysis
4. partial least squares
5. principal component analysis

8. ข้อใดใช้ในการคำนวณส่วนประกอบของตัวแปรที่ใช้ในการทำนายเพื่อแยกกลุ่มย่อยจากกลุ่มข้อมูลที่ทราบประเภทของกลุ่มข้อมูลล่วงหน้าอยู่แล้วได้

9. ข้อใดเป็นวิธีที่ใช้ในการคัดเลือกกลุ่มของตัวแปรที่เหมาะสมสำหรับการสร้างแบบจำลอง

10. ข้อใดเป็นเทคนิคที่ใช้ในการซึ่งระหว่างกลุ่มของข้อมูลเหล่านี้เหมือนกับข้อมูลกลุ่มอื่นหรือไม่และสามารถแสดงจำนวนกลุ่มที่แตกต่างกันของข้อมูลที่นำมาวิเคราะห์ได้