

## รายงานฉบับสมบูรณ์

การพัฒนาขั้นตอนวิธีเชิงกราฟและระบบสารสนเทศในการคัดกรองยา:  
กรณีศึกษาในโปรตีนเอสของโคโรนาไวรัส

Development of graph algorithms and informatics system for drug screening:  
a case study of coronavirus proteinase

รหัสโครงการ

SCH-NR2012-121

หัวหน้าโครงการ

นางสาวสมนึก วรวิเศษ

ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น

## รายงานฉบับสมบูรณ์

การพัฒนาขั้นตอนวิธีเชิงกราฟและระบบสารสนเทศในการคัดกรองยา:  
กรณีศึกษาในโปรตีนเอสของโคโรนาไวรัส

Development of graph algorithms and informatics system for drug screening:  
a case study of coronavirus proteinase

### รายนามคณะผู้วิจัย

1. นางสาวสมนึก วรวิเศษ  
หน่วยงาน ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น
2. นายเทวัญ เริ่มสูงเนิน  
หน่วยงาน ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น วิทยาเขต  
หนองคาย

วันเริ่มต้นโครงการ 1 เมษายน 2556

วันสิ้นสุดโครงการ 30 กันยายน 2557

## กิตติกรรมประกาศ

โครงการวิจัยนี้ได้รับการสนับสนุนจาก  
ศูนย์ประสานงานนักเรียนทุนรัฐบาลทางด้านวิทยาศาสตร์และเทคโนโลยี  
กระทรวงวิทยาศาสตร์และเทคโนโลยี  
และสำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ

ปี พ.ศ. ที่ได้รับทุน 2555

## บทคัดย่อ

ปัจจุบันมีสารประกอบเคมีเป็นจำนวนมากมาย ซึ่งถูกเก็บไว้เป็นฐานข้อมูลเพื่อใช้งานกันอย่างกว้างขวาง ซึ่งมีไม่น้อยที่สามารถแสดงศักยภาพเชิงยา หากสามารถคัดกรองและคัดเลือกสารประกอบบางโครงสร้างมาทำการทดลองทดสอบ จะช่วยลดทั้งค่าใช้จ่ายจำนวนมหาศาลและเวลาในการค้นคว้าเพื่อผลิตยารักษาโรค โดยเฉพาะเมื่อเกิดการระบาด ดังนั้นเทคนิควิธีการคัดกรองยาจึงได้รับความสนใจและพัฒนาอยู่ตลอดเวลา โครงสร้างโมเลกุลของสารประกอบเคมี สามารถเขียนแทนได้ด้วยกราฟเชิงโมเลกุล ซึ่งจะสามารถใช้ทฤษฎีและขั้นตอนวิธีทางกราฟ ในการสืบค้นหรือวิเคราะห์โครงสร้าง เช่นการสืบค้นโครงสร้างย่อย การสืบค้นโครงสร้างที่คล้ายกัน การสืบจำแนกโครงสร้างย่อยร่วมกัน เป็นต้น ซึ่งเป็นสารสนเทศสำคัญเพื่อการคัดกรองโมเลกุลสารประกอบที่น่าจะมีศักยภาพทางยา โดยเฉพาะกับโรคที่มีการใช้ยาหรือสารประกอบเคมีบางตัวอยู่แล้ว และโดยที่โครงสร้างโมเลกุลเคมี 3 มิติ และ 2 มิติ สามารถเขียนลดรูปเป็นข้อความ 1 มิติ ในรูปแบบ SMILES (Simplified Molecular-Input Line-Entry System) และสามารถแปลงย้อนกลับได้ ดังนั้นจึงเป็นที่นิยมใช้เป็นรูปแบบในการเก็บข้อมูลโครงสร้างในฐานข้อมูลต่างๆ

ในงานวิจัยนี้จึงได้พัฒนาขั้นตอนวิธีที่จะสามารถสกัดข้อมูลเชิงโครงสร้างของโมเลกุล ด้วยการกำหนดส่วนหรือหมู่สำคัญ (Meaning Fragments, MFs) เพื่อใช้สำหรับลดรูปโครงสร้างโมเลกุล พร้อมกับเสนอวิธีคำนวณค่าความคล้าย RTCS (Refinement Tanimoto Coefficient of Similarity) จากชนิดและจำนวนของ MFs ที่มีอยู่ในโมเลกุลที่นำมาเปรียบเทียบกัน แล้วนำค่า RTCS ไปใช้เป็นดัชนีเพื่อคัดกรองสารประกอบทางเคมีเทียบกับ ค่าความคล้าย TCS (Tanimoto Coefficient of Similarity) ที่นิยมใช้กันอยู่โดยทั่วไป ทั้งนี้ได้เลือกกรณีศึกษาเป็นการคัดกรองยาเพื่อต่อต้านโปรตีนเอสของไวรัสโคโรนา ซึ่งเป็นไวรัสที่ทำให้เกิดโรคทางเดินหายใจในเฉียบพลันรุนแรง (Severe acute respiratory distress syndrome, SARS) ที่มีข้อมูลโครงสร้างยาได้แก่ Ribavirin, Lopinavir, Ritonavir, Calanolide A และ Navirapine กับโปรตีน 3C-LIKE Proteinase และฐานข้อมูลสารประกอบทางเคมีที่เข้าถึงได้โดยไม่มีค่าใช้จ่าย อย่างเช่น PubChem และ DrugBank ผลการศึกษาพบว่าค่า RTCS ที่คำนวณได้โดยเฉลี่ยแล้วจะสูงกว่าค่า TCS ซึ่งเมื่อนำมากำหนดเป็นค่าดัชนีเพื่อคัดกรองจึงได้จำนวนโมเลกุลที่ผ่านเกณฑ์มากกว่า แต่ก็ไม่มากจนเกินไปในการที่จะคำนวณขั้นตอนการเข้าจับเชิงโมเลกุลต่อไป

## Abstract

There is a huge number of chemical compounds that is collected as free access database for worldwide using. Several molecules show potential to use as pharmacological compounds. If we can screen and selected only some compound for experimental testing, the huge of cost and time for investigation could be reduced, especially when the time of epidemic. So several of drug screening techniques are of great interested and developed. The structure of chemical compounds can be represented with molecular graph which theories and algorithms of graph can be applied for structural searching or analyzing, e.g. substructure searching, structure similarity searching, common substructures analysis etc., which are important information for pharmacological drug screening when we have already some used drugs or chemical compounds. The 3D and 2D molecular structures can be reduced and represented by text in one dimension in the form of SMILES (simplified molecular-input line-entry system) format. It can be also reverse transformed back to 2D and 3D, respectively. So it is one format that widely used in molecular databases. In this research we develop the algorithm for extracting structure data from SMILES format. The meaning molecular parts or fragment are defined as "Meaning Fragments", and denoted by MFs for use in the molecular reduced graphs. A similarity score RTCS (Refinement Tanimoto Coefficient of Similarity) is proposed and is obtained based on types and number of MFs as appear in both and each of pair of compared molecules. Then the RTCS is used as the screening criteria in comparison with original TCS (Tanimoto Coefficient of Similarity) that is generally used widely. The screening of drug for SARS coronavirus proteinase is used for case study. The known five drug structures, Ribavirin, Lopinavir, Ritonavir, Calanolide A, Navirapine and 3C-LIKE proteinase are obtained in free accessing databases, e.g. PubChem and DrugBank. The results show that the obtained RTCS values are most higher than the corresponding TCS values, and also yields more hit molecules but not too much to use as input of molecular docking step.

## สารบัญเรื่อง

เนื้อหา	หน้า
บทนำ	9
วัตถุประสงค์ของโครงการวิจัย	11
การทบทวนวรรณกรรม	11
การออกแบบการวิจัย	12
ขอบเขตของการวิจัย	13
ระเบียบวิธีวิจัย	13
ผลการวิจัย	20
สรุปผล	34
ปัญหาและอุปสรรค	34
บรรณานุกรม	35
ภาคผนวก	37

## สารบัญรูปภาพ

รูป	หน้า
รูปที่ 1 แสดงขั้นตอนการนิยามหมู่ย่อยของโมเลกุล	12
รูปที่ 2 แสดงขั้นตอนการนิยามหมู่ย่อยของโมเลกุล	14
รูปที่ 3 เปรียบเทียบค่า TCS และ RTCS ของโมเลกุล Ribavirin กับ โมเลกุลสารประกอบเคมีในฐานข้อมูล DrugBank จำนวน 6797 หมายเลข	22
รูปที่ 4 เปรียบเทียบค่า TCS และ RTCS ของโมเลกุล Ribavirin กับ โมเลกุลสารประกอบเคมีในฐานข้อมูล Pubchem จำนวน 25000 หมายเลขแรก	23
รูปที่ 5 เปรียบเทียบค่า TCS และ RTCS ของโมเลกุล Lopinavir กับ โมเลกุลสารประกอบเคมีในฐานข้อมูล DrugBank จำนวน 6797 หมายเลข	24
รูปที่ 6 เปรียบเทียบค่า TCS และ RTCS ของโมเลกุล Lopinavir กับ โมเลกุลสารประกอบเคมีในฐานข้อมูล Pubchem จำนวน 25000 หมายเลขแรก	25
รูปที่ 7 เปรียบเทียบค่า TCS และ RTCS ของโมเลกุล Ritonavir กับ โมเลกุลสารประกอบเคมีในฐานข้อมูล DrugBank จำนวน 6797 หมายเลข	26
รูปที่ 8 เปรียบเทียบค่า TCS และ RTCS ของโมเลกุล Ritonavir กับ โมเลกุลสารประกอบเคมีในฐานข้อมูล Pubchem จำนวน 25000 หมายเลขแรก	27
รูปที่ 9 เปรียบเทียบค่า TCS และ RTCS ของโมเลกุล Calanolide A กับ โมเลกุลสารประกอบเคมีในฐานข้อมูล DrugBank จำนวน 6797 หมายเลข	28
รูปที่ 10 เปรียบเทียบค่า TCS และ RTCS ของโมเลกุล Calanolide A กับ โมเลกุลสารประกอบเคมีในฐานข้อมูล Pubchem จำนวน 25000 หมายเลขแรก	29
รูปที่ 11 เปรียบเทียบค่า TCS และ RTCS ของโมเลกุล Navirapine กับ โมเลกุลสารประกอบเคมีในฐานข้อมูล DrugBank จำนวน 6797 หมายเลข	30
รูปที่ 12 เปรียบเทียบค่า TCS และ RTCS ของโมเลกุล Navirapine กับ โมเลกุลสารประกอบเคมีในฐานข้อมูล Pubchem จำนวน 25000 หมายเลขแรก	31
รูปที่ 13 เปรียบเทียบค่า RTCS และ TCS ของโมเลกุลหมายเลข 2436 ในฐานข้อมูล DrugBank หมายเลข 21325 ในฐานข้อมูล PubChem ซึ่งมีพลังงานยึดจับเป็น -8.78 และ -7.56 kcal/mol ตามลำดับ	33

## สารบัญตาราง

ตาราง	หน้า
ตารางที่ 1 โครงสร้างโมเลกุล และรูปแบบ SMILES สำหรับโมเลกุลที่น่าจะมีศักยภาพเชิงยา	13
ตารางที่ 2 แสดงการกำหนดและนิยามกลุ่มย่อยในโมเลกุลพร้อมการเขียนในรูปแบบ SMILES	15
ตารางที่ 3 แสดงค่าเฉลี่ยความต่าง (RTCS - TCS) และหมายเลขโมเลกุลที่ผ่านเกณฑ์คะแนน 0.6 ขึ้นไป โดยการเรียงลำดับจากคะแนนสูงไปต่ำด้วยวิธี RTCS และ TCS ตามลำดับ	21
ตารางที่ 4 แสดงผลการคำนวณพลังงานยึดจับ (kcal/mol) ที่ได้จากวิธีเข้าจับเชิงโมเลกุล เฉพาะที่มีค่าพลังงานยึดจับต่ำกว่า -3.0 kcal/mol	32

## บทนำ

ปัจจุบันมีข้อมูลสารประกอบทางเคมีอยู่จำนวนมหาศาล และโดยใช้ประโยชน์จากระบบเทคโนโลยีทางด้านคอมพิวเตอร์และสารสนเทศ ทำให้มีการจัดทำห้องสมุดหรือฐานข้อมูลโครงสร้างโมเลกุลไว้เป็นจำนวนมาก [1] โดยบางฐานข้อมูลก่อกำหนดให้เข้าถึงและใช้งานเพื่อประโยชน์ในด้านต่างๆ ได้อย่างกว้างขวางและไม่มีเสียค่าใช้จ่าย มีโมเลกุลจำนวนไม่น้อยที่ปรากฏอยู่ในฐานข้อมูลเหล่านั้น น่าจะมีศักยภาพเชิงยาที่สามารถนำมาพัฒนาต่อได้ แต่การที่จะต้องทดสอบทดลองกับโมเลกุลจำนวนมากๆ หรือทั้งหมดที่มีอยู่ในฐานข้อมูลจำเป็นจะต้องมีใช้ทรัพยากรที่สูงมาก ทั้งในด้านเงินทุนและด้านเวลาสำหรับการค้นคว้าวิจัย ดังนั้นการสกัดและนำข้อมูลสารสนเทศด้านต่างๆ เช่น องค์ประกอบ โครงสร้าง หรือสมบัติบางประการ เข้ามาใช้ช่วยเป็นเกณฑ์ในกระบวนการสืบค้นคัดกรองก่อน เพื่อให้ได้ ลักษณะและจำนวนโมเลกุลที่เป็นผลลัพธ์ที่เหมาะสมจะช่วยลดต้นทุนได้จำนวนมาก จึงเป็นอีกปัญหาหนึ่งที่ได้รับ ความสนใจศึกษากันอย่างกว้างขวางและมีความท้าทายอย่างยิ่ง โดยเฉพาะประสิทธิภาพในการใช้เพื่อสืบค้นในฐานข้อมูลขนาดใหญ่ที่มีจำนวนโครงสร้างหลายล้านโมเลกุล

โครงสร้างสามมิติของโมเลกุลเคมีต่างๆ สามารถจัดเก็บและเปลี่ยนแปลงกลับไปมาได้หลายรูปแบบ และโดยสมบัติการเชื่อมต่อดัวยพันธะระหว่างอะตอมภายในโมเลกุลที่มีความจำเพาะและมีรูปแบบที่แน่นอน การลดรูปเพื่อเก็บเป็นข้อมูลโครงสร้างสองมิติ หรือแม้แต่ข้อความในหนึ่งมิติ ก็สามารถทำได้โดยไม่สูญเสียข้อมูลไปมากนัก และโดยที่โครงสร้างโมเลกุลคือมีลักษณะการเชื่อมต่อกันของอะตอมด้วยพันธะชนิดต่างๆ ซึ่งคล้ายกับโครงสร้างกราฟในทางคณิตศาสตร์ที่มีองค์ประกอบเป็น จุดและเส้นเชื่อม ดังนั้นจึงสามารถประยุกต์ใช้กราฟเขียนแทนโครงสร้างโมเลกุลรวมถึงการประยุกต์ใช้สมบัติ เทคนิคและขั้นตอนวิธีเชิงกราฟต่างๆ ในการวิเคราะห์และประมวลผลเพื่อดำเนินการต่างๆ เช่นการลดรูปกราฟ การค้นหาและจำแนกโครงสร้างกราฟย่อย การมีโครงสร้างย่อยที่ใหญ่ที่สุดร่วมกันได้โดยตรง ดังนั้นการพัฒนาทฤษฎีและขั้นตอนวิธีเชิงกราฟ การเลือกใช้และปรับปรุงเทคนิคและขั้นตอนวิธีเชิงกราฟที่ดีจะช่วยให้เพิ่มประสิทธิภาพในกระบวนการคัดกรองได้ [2] วิธีการเขียนแทนโครงสร้างโมเลกุลสารประกอบเคมีในรูปแบบ SMILES (Simplified Molecular-Input Line-Entry System) [3-5] ซึ่งเป็นรูปแบบข้อความหนึ่งมิติ ที่ยังคงให้รายละเอียดเชิงโครงสร้างได้อย่างครบถ้วน ซึ่งสามารถแปลงไปเป็นโครงสร้างสองมิติและสามมิติได้ และได้รับยอมรับเป็นอีกมาตรฐานหนึ่งของรูปแบบข้อมูลเข้า เพื่อประมวลผลด้วยโปรแกรมประยุกต์ด้านการสร้าง วิเคราะห์ หรือคำนวณโครงสร้างโมเลกุลต่างๆ เพื่อลดปริมาณพื้นที่ในการจัดเก็บ โดยไม่เสียัยสำคัญของข้อมูลโครงสร้าง ตลอดจนจนถึงความง่ายในการเข้าถึงและประมวลผล SMILES จึงเป็นรูปแบบหนึ่งที่น่าไปใช้ในฐานข้อมูลโครงสร้างโมเลกุลต่าง เช่น DrugBank หรือ PubChem เป็นต้น

เมื่อต้องการค้นโมเลกุลที่คล้ายกันตัวชี้วัดที่สำคัญคือ ค่าความคล้าย (similarity score) ซึ่งได้มีการกำหนดนิยามไว้หลายแบบ แต่ที่นิยมใช้กันอยู่ทั่วไปได้แก่ Tanimoto's coefficient score (TCS) ซึ่งจะทำการแบ่งโมเลกุลที่นำมาเทียบกันทั้งสองออกเป็นส่วนย่อยๆ เรียกว่า บิต (bit) จากนั้นก็คำนวณค่าความคล้ายด้วยสัดส่วนระหว่างจำนวนบิตที่มีเหมือนกันในทั้งสองโมเลกุลและจำนวนบิตที่มีอยู่ร่วมกันทั้งสองโมเลกุล และโดยที่ค่าความเหมือน TCS นี้ได้รับความนิยมนำไปใช้เป็นเกณฑ์เพื่อสืบค้นหาโมเลกุลในฐานข้อมูลโดยใช้โมเลกุลอ้างอิงมีอยู่แล้ว (ซึ่งอาจเป็นโมเลกุลยาหรือสารประกอบที่มีศักยภาพ) การปรับปรุงวิธีการคำนวณค่าความคล้ายให้มีความถูกต้องมากขึ้นจะช่วยให้การคัดกรองสารมีประสิทธิภาพสูงขึ้น นอกจากนี้ในปัจจุบันได้มีการสร้างและเผยแพร่โปรแกรมประยุกต์และเครื่องมือต่างๆ ที่เกี่ยวข้องกับการค้นหาเชิงโมเลกุลเป็นจำนวนมาก แต่อย่างไรก็ตามการใช้งานระบบหรือโปรแกรมเหล่านี้ให้ได้มีประสิทธิภาพ จำเป็นที่ผู้ใช้จะต้องมีความรู้และทักษะด้านคอมพิวเตอร์ที่ดี ทำให้การศึกษาวิชาญและใช้ประโยชน์เครื่องมือต่างๆ เหล่านี้ยังเป็นไปอย่างจำกัด ซึ่งหากมีการศึกษา จัดระเบียบ และสร้างระบบสารสนเทศที่ดี จะช่วยให้การใช้เครื่องมือเหล่านี้ได้อย่างสะดวกและมีประสิทธิภาพ เพิ่มมิติทั้งด้านจำนวนผู้ใช้งานและความหลากหลายในการประยุกต์ใช้ ดังนั้นในการศึกษานี้จึงมุ่งเน้นตอบโจทย์สำคัญสองประการคือ

1) การพัฒนากระบวนการประยุกต์และปรับปรุงวิธีเชิงกราฟเพื่อช่วยในการคัดกรองโมเลกุลที่มีศักยภาพทางยาจากฐานข้อมูลต่างๆ อย่างมีประสิทธิภาพ โดยเฉพาะการนำเสนอแนวทางปรับปรุงเพื่อคำนวณค่า TCS ใหม่และเรียกว่า refinement Tanimoto's coefficient score (RTCS) โดยมีหลักการเบื้องต้นว่า การกำหนดกลุ่มโมเลกุลย่อยที่จะปรากฏหรือไม่ปรากฏในคูโมเลกุลที่นำมาเทียบกันนั้น ควรจะกำหนดให้อย่างน้อยที่สุดเป็นโครงสร้างที่มีความหมาย เช่น ครบวงโมเลกุล หรือครบหมู่ฟังก์ชัน ตามที่ได้กำหนดนิยามไว้ และให้มีการนับที่ครบถ้วนแต่ไม่ซ้ำซ้อน โดยอาศัยกราฟลดรูปมาช่วยในการคำนวณ และ

2) เพื่อพัฒนาสารสนเทศสำหรับการใช้กระบวนการที่พัฒนาขึ้นให้มีความง่ายต่อการใช้งานและเข้าถึงได้ของผู้ต้องการใช้ประโยชน์

โรคทางเดินหายใจเฉียบพลันรุนแรง (SARS) เป็นโรคที่เคยเกิดการระบาดอย่างรุนแรงและผู้เสียชีวิตจำนวนมาก จึงทำให้เป็นโรคหนึ่งที่ได้รับการศึกษาอย่างกว้างขวาง มีตัวยารักษาโรคหรือยาต้านเชื้อที่มีคุณภาพ แต่ด้วยธรรมชาติที่จะเกิดการเปลี่ยนแปลงสายพันธุ์ได้ตลอดเวลา การพัฒนายาจึงจำเป็นต้องทำอย่างต่อเนื่องและทันเวลา การคัดกรองสารด้วยการคำนวณทางคอมพิวเตอร์ [6] จึงมีความสำคัญมากต่อการรับมือกับสถานการณ์ ดังนั้นจึงได้เลือกมาเป็นกรณีศึกษาในโครงการวิจัยนี้

## วัตถุประสงค์ของโครงการวิจัย

1. เพื่อพัฒนากระบวนการประยุกต์หลักการและขั้นตอนวิธีเชิงกราฟในการสืบค้นโมเลกุลสารเคมีที่มีศักยภาพเป็นยาในฐานะข้อมูลต่างๆ
2. เพื่อพัฒนาระบบสารสนเทศในการประมวลผลเพื่อคัดกรองโมเลกุลเคมีที่มีศักยภาพเป็นยารักษาโรค

### การทบทวนวรรณกรรม

การค้นคว้าวิจัยทางยาในปัจจุบันบ่อยครั้งที่จะต้องจัดการกับข้อมูลขนาดใหญ่ เช่นการสังเคราะห์สารด้วยวิธีทางเคมีเชื่อมโยง การคัดกรองสารประสิทธิภาพสูง ที่จะต้องมีผลลัพธ์ที่เกิดขึ้นเป็นปริมาณมากซึ่งถูกจัดเก็บไว้ในฐานข้อมูลสาธารณะ ส่วนบุคคล หรือเชิงธุรกิจ [1] การทำเหมืองข้อมูลเชิงโมเลกุลโดยใช้หลักการและวิธีเชิงกราฟ กำลังได้รับความนิยมศึกษาอย่างกว้างขวางโดยเฉพาะในการสืบค้นด้านยา [2] โดยเฉพาะกับการเขียนแทนโครงสร้างโมเลกุลในรูปแบบ SMILES [3-5] ซึ่งมีลักษณะเป็นข้อความหนึ่งมิติ แต่สามารถแปลงไปเป็นโครงสร้างสองมิติและสามมิติได้โดยไม่เสียนัยสำคัญเชิงโครงสร้าง ซึ่งใช้หลักการของกราฟต้นไม้แผ่ขยาย และการท่องแบบลึกก่อน (depth-first tree traversal) ดังนั้นขั้นตอนวิธีสมมูลฐานกราฟย่อย (subgraph isomorphism algorithm) ถูกประยุกต์ใช้เพื่อค้นหารูปแบบโครงสร้างย่อยของโมเลกุลที่ต้องการจากโครงสร้างโมเลกุลขนาดใหญ่หรือจำนวนมาก [7-8] ซึ่งการประยุกต์ใช้กราฟย่อยนี้จะประโยชน์ทั้งในด้านค้นหาโครงสร้างโมที่พของโปรตีนซึ่งมีประโยชน์ในการทำนายสมบัติของโปรตีน หรือการค้นหาโมเลกุลที่มีอยู่ในฐานข้อมูลขั้นต้น เพื่อเป็นการลดจำนวนโมเลกุลลงให้ได้จำนวนมาก ก่อนที่จะนำไปศึกษาด้วยขั้นตอนวิธีจับคู่กราฟ (graph) หรือนำไปใช้คำนวณทดสอบเคมีทางยาต่อไป การพัฒนาระเบียบวิธีเพื่อสืบค้นข้อมูลเชิงกราฟนี้ยังมีความท้าทายในการเพิ่มประสิทธิภาพในประเด็นต่างๆ เช่น การเก็บและแปลงโมเลกุลให้เป็นกราฟอย่างมีประสิทธิภาพ การนิยามความเหมือนกันของกราฟ และการสร้างขั้นตอนวิธีที่มีประสิทธิภาพในการสืบค้น เปรียบเทียบและจับคู่กราฟ ปัญหาที่ท้าทายอันดับแรกก็คือการจับคู่เปรียบเทียบกราฟซึ่งเป็นปัญหาที่ยาก โดยเฉพาะปัญหาสมมูลฐานกราฟย่อยถูกระบุว่าเป็นปัญหาชนิด NP-complete [9]

บนพื้นฐานของหลักการที่ว่า โมเลกุลที่มีโครงสร้างคล้ายกันจะมีสมบัติส่วนใหญ่ที่เหมือนกันด้วย ดังนั้นการสืบค้นเพื่อหากลุ่มโมเลกุลที่มีโครงสร้างคล้ายกับโมเลกุลอ้างอิงที่มีสมบัติที่ต้องการอยู่แล้ว จึงเป็นแนวทางพื้นฐานหนึ่งในการค้นคว้าการพัฒนายา การพัฒนาตัวชี้วัดที่ใช้ระดับความคล้ายกันของโมเลกุลจึงเป็นอีกหนึ่งหัวข้อที่ได้รับความสนใจศึกษาและพัฒนาในปัจจุบัน และแม้จะมีการเสนอวิธีคำนวณค่าความคล้ายแบบต่างๆ แต่ค่าความคล้าย TCS ก็ยังได้รับการยอมรับและเป็นที่ยอมรับใช้กันอย่างกว้างขวาง [10-12] ในฐานะข้อมูลต่างๆ เช่นฐานข้อมูลโครงสร้างยา DrugBank [13] หรือสารข้อมูลโครงสร้างสารประกอบเคมี PubChem [14] ซึ่งเป็นฐานข้อมูลที่มีโครงสร้างจำนวนมาก มีบริการเครื่องมือสืบค้นสารสนเทศ และสามารถเข้าใช้ได้โดยไม่เสียค่าใช้จ่าย นอกจากนี้ยังสามารถถ่ายโอนดึงเอาข้อมูลโมเลกุลมาเก็บไว้เพื่อใช้งานโดยไม่ต้องเชื่อมต่อระบบเครือข่ายของแหล่งฐานข้อมูล

โรคทางเดินหายใจเฉียบพลันรุนแรง (Severe Acute Respiratory Syndrome, SARS) เป็นโรคที่เคยเกิดการระบาดอย่างรุนแรงและมีผู้เสียชีวิตเป็นจำนวนมาก เป็นตัวอย่างหนึ่งชี้ให้เห็นถึงความจำเป็นต้องมีการพัฒนาอย่างต่อเนื่องและทันเวลา การคัดกรองสารด้วยการคำนวณทางคอมพิวเตอร์ [6] จึงมีความสำคัญมากต่อการรับมือกับสถานการณ์ แม้ว่าจะยังไม่สามารถตกผลึกเพื่อให้ทราบข้อมูลโครงสร้างของโปรตีนสำคัญของไวรัสที่ทำให้เกิดโรค แต่การหาลำดับหมู่อะมิโนก็เป็นสิ่งที่ทำได้รวดเร็ว และสามารถนำมาเป็นข้อมูลเพื่อสืบค้นโครงสร้างโปรตีนที่คล้ายกันในฐานข้อมูลได้ นอกจากนี้ยังสามารถใช้โครงสร้างยาที่ใช้รักษาโรคใช้หวัดอื่นๆ มาเป็น

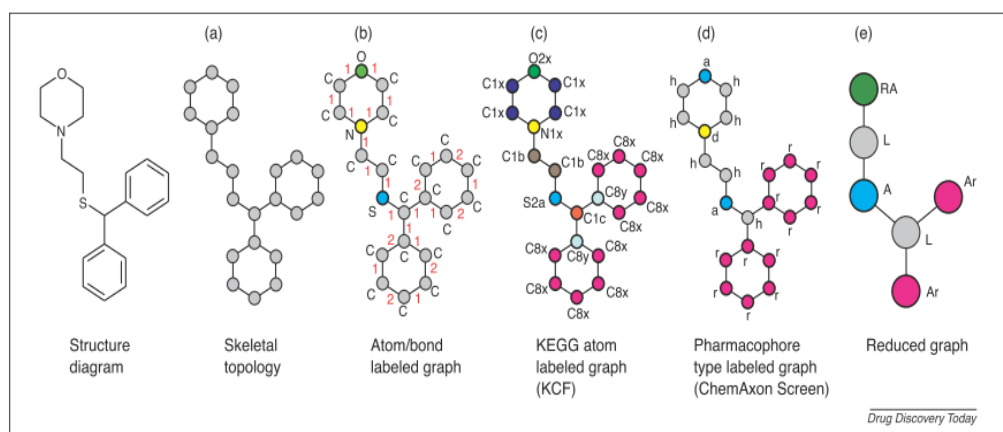
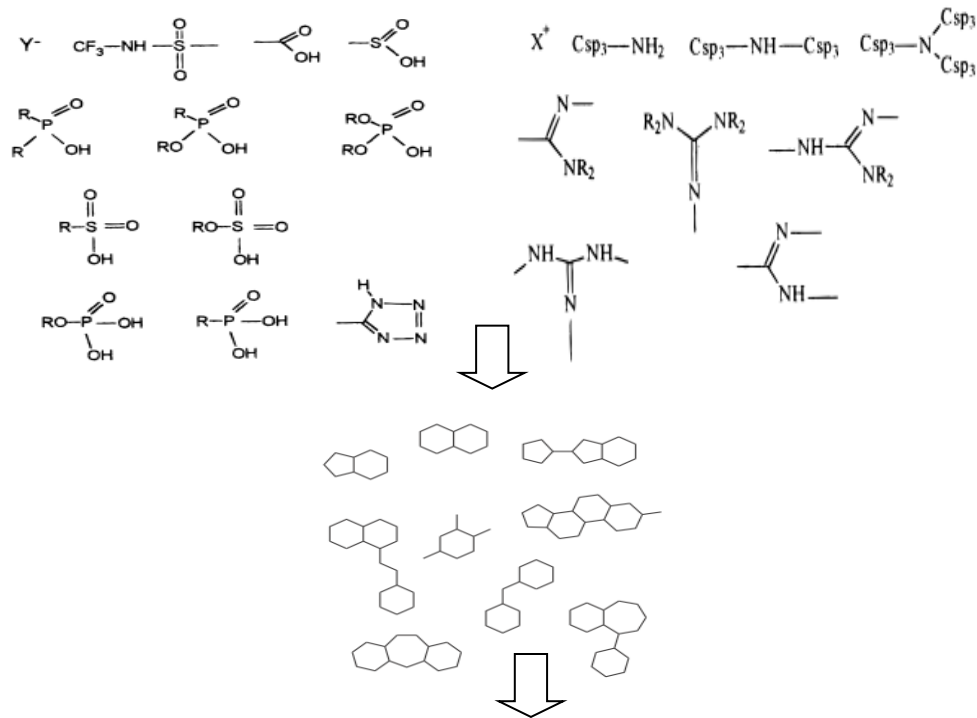
โมเลกุลอ้างอิงเพื่อคัดกรองหาโมเลกุลอื่นๆในฐานข้อมูลได้อีก ซึ่งเป็นเหตุผลที่ผู้วิจัยเลือกมาเป็นกรณีศึกษาในโครงการวิจัยนี้

## การออกแบบการวิจัย

1) ใช้กรณีศึกษาต้านไวรัสโปรตีนเอส เพื่อสร้างข้อมูลโครงสร้างสำหรับสกัดรูปแบบกราฟย่อยที่ต้องการ เพื่อนำไปใช้ในการลดรูปกราฟ (รูปที่ 1) และใช้เป็นต้นแบบอ้างอิงสำหรับสืบค้นหาในฐานข้อมูลเช่น DrugBank หรือ PubChem (ChEBI)

2) ศึกษาวิเคราะห์วิธีการคำนวณค่าความคล้าย TCS พิจารณาหาจุดที่สามารถปรับปรุงได้อีก เสนอขั้นตอนวิธีที่พัฒนาปรับปรุงขึ้นมาใหม่ เพื่อให้การสืบค้นที่มีความแม่นยำ มีประสิทธิภาพและรวดเร็ว

3) นำโครงสร้างโมเลกุลที่ได้จากการสืบค้นมาคำนวณพลังงานในการยึดจับกับโปรตีนเอสของไวรัส เพื่อตรวจสอบคุณภาพระเบียบวิธี เรียบเรียงพัฒนากระบวนการให้มีความรัดกุมแม่นยำก่อนจัดทำระบบสารสนเทศเพื่อให้บริการ



รูปที่ 1 แสดงขั้นตอนการนิยามหมู่ย่อยของโมเลกุลเพื่อนำมาใช้ในการแปลงเพื่อลดรูปโครงสร้างโมเลกุลหรือบางส่วนของโมเลกุลก่อนนำไปเข้ากระบวนการสืบค้นในฐานข้อมูลต่อไป

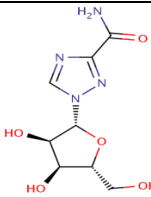
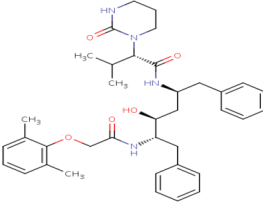
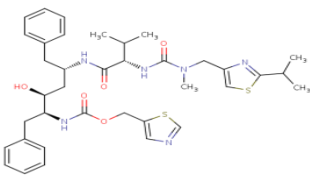
## ขอบเขตของการวิจัย

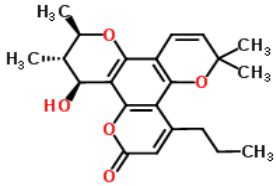
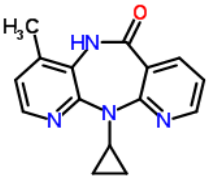
ในการศึกษานี้จะใช้โมเลกุลยาชนิดต่างๆ ที่ใช้ได้ผลในการต้านเชื้อโคโรนาไวรัส สร้างรูปแบบโครงสร้างหรือรูปแบบกราฟที่ต้องการ แล้วนำไปสืบค้นในฐานข้อมูลต่างๆ ด้วยกระบวนการและขั้นตอนวิธีที่พัฒนาขึ้น รวมถึงการสร้างระบบสารสนเทศสำหรับให้บริการแก่ผู้ใช้ทั่วไป

## ระเบียบวิธีวิจัย

1. ดำเนินการสืบค้นหาตัวยาที่มีประสิทธิภาพในการต้านโคโรนาไวรัสโปรตีนเอส เพื่อใช้ข้อมูลโครงสร้างสำหรับสร้างรูปแบบกราฟย่อยที่ต้องการ เพื่อนำไปสืบค้นหาในฐานข้อมูลเช่น DrugBank และ PubChem (ChEBI) โดยการอ้างอิงจากรายงานการวิจัย [6, 15, 16] โดยตัวยาที่ได้รับการรับรองแล้วและมีรายงานการใช้ต่อต้าน SARS-Cov ได้แก่ Ribavirin ซึ่งได้รับความนิยมใช้กับผู้ป่วยในช่วงแรกที่มีการระบาดของโรคและได้ผลพอสมควร (ปานกลาง) ต่อมาพบว่าเมื่อใช้ร่วมกับยาต้านไวรัส HIV บางตัวเช่น Lopinavir และ Ritonavir จะช่วยเพิ่มผลการรักษาให้ดีขึ้น นอกนั้นยังมีรายงานพบว่า Calanolide A และ Navirapine ก็น่าจะมีศักยภาพในการต่อต้านเช่นกัน ดังนั้นในงานวิจัยนี้ จึงได้ใช้ Ribavirin Lopinavir Ritonavir Calanolide A และ Navirapine เป็นโมเลกุลอ้างอิงสำหรับการค้นตัวยาหรือโมเลกุลอื่นๆ ที่มีโครงสร้างคล้ายกันในฐานะโมเลกุลยา โดยมีโครงสร้างโปรตีน 3C-LIKE Proteinase เป็นเป้าหมาย โครงสร้างโมเลกุลของยาและการเขียนแทนโมเลกุลในรูปแบบ SMILES แสดงไว้ในตารางที่ 1

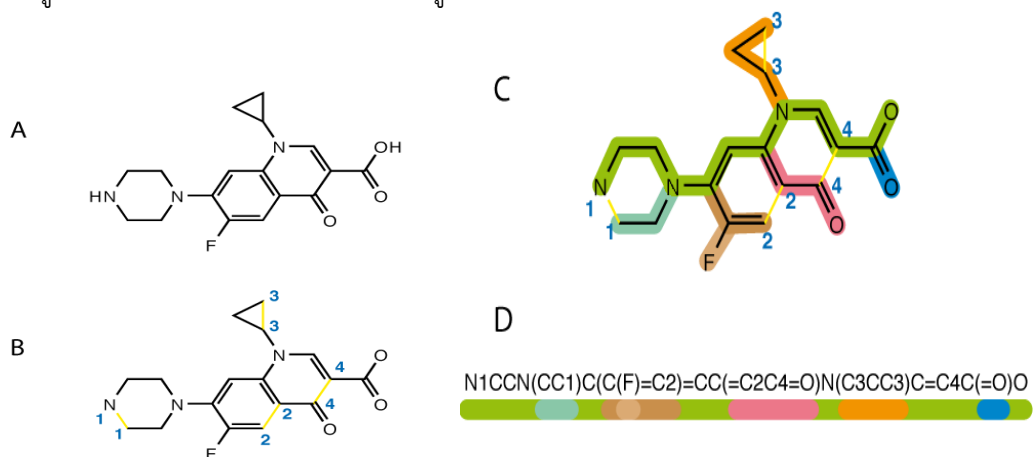
ตารางที่ 1 โครงสร้างโมเลกุล และรูปแบบ SMILES สำหรับโมเลกุลที่น่าจะมีศักยภาพเชิงยา

โมเลกุลยาที่ได้รับการอนุมัติและมีการใช้เพื่อต่อต้าน SARS-Cov แล้ว	
	<p><b>Ribavirin (Virazole)</b> SMILE: <chem>OC[C@H]1O[C@H]([C@@H]([C@H]1O)O)n1cnc(n1)C(=O)N</chem></p>
	<p><b>Lopinavir</b> SMILE : <chem>O=C(N[C@H]([C@H](C[C@H](Cc1ccccc1)NC(=O)[C@@H](N1CCCNC1=O)C(C)C)O)Cc1ccccc1)COc1c(C)cccc1C</chem></p>
	<p><b>Ritonavir</b> SMILE : <chem>O=C(N[C@H]([C@H](C[C@H](Cc1ccccc1)NC(=O)[C@@H](C(C)C)NC(=O)N(Cc1csc(n1)C(C)C)O)Cc1ccccc1)OCc1cncs1</chem></p>

โมเลกุลยาที่ได้รับการอนุมัติและมีการศึกษาวิจัยว่าน่าจะมีศักยภาพในการต่อต้าน SARS-Cov แล้ว	
	<p><b>Calanolide A</b>            SMILE :  <chem>CCCC1=CC(=O)OC2=C1C3=C(C=CC(O3)(C)C)C4=C2[C@H]([C@@H]([C@H](O4)C)C)O</chem></p>
	<p><b>Navirapine</b>            SMILE : <chem>CC1=C2C(=NC=C1)N(C3=C(C=CC=N3)C(=O)N2)C4CC4</chem></p>

2. สร้างขั้นตอนวิธีการแปลงโครงสร้างโมเลกุลให้เป็นโครงสร้างเชิงกราฟ โดย “การนิยามชนิดหมู่โมเลกุล” การตรวจหาพันธะและชนิดพันธะที่เชื่อมต่อ การจัดเก็บข้อมูลด้วยเมตริกซ์ รวมถึง “การลดรูปกราฟ” ให้มีองค์ประกอบน้อยที่สุด ทั้งโครงสร้างต้นแบบ และโครงสร้างโมเลกุลที่อยู่ในฐานข้อมูล โดยพัฒนาขั้นตอนวิธีการสกัดข้อมูลโครงสร้างโมเลกุลจากรูปแบบ SMILES ดังนี้

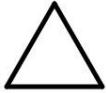

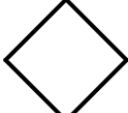
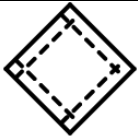

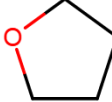
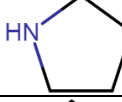
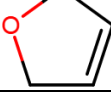
เนื่องด้วยการเขียนแทนโมเลกุลด้วย SMILES จะเริ่มโดยการตัด H อะตอมที่เป็นปลายออก แล้วแตกหมู่ที่เป็นวงออกเพื่อให้ได้กราฟต้นไม้แผ่ขยาย โดยคู่อะตอมที่ถูกตัดเส้นเชื่อมออกจะกำหนดหมายเลขกำกับไว้เหมือนกันเป็นลำดับไป แล้วเขียนลำดับของอะตอมตามลำดับพันธะในกราฟเชิงโมเลกุลนี้ตามแนวการท่องแบบลึกก่อน (depth-first tree traversal) และใช้วงเล็บเพื่อคลุมส่วนต่อพันธะที่แยกออกไปเป็นกิ่ง ขั้นตอนการเขียนแทนโมเลกุลในรูปแบบ SMILES สามารถแสดงได้ตามรูปที่ 2

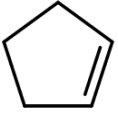
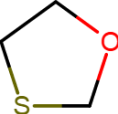
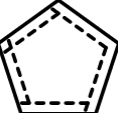
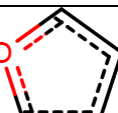
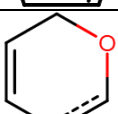
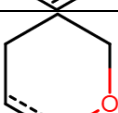
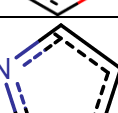

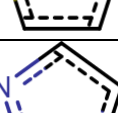

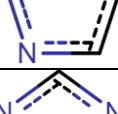


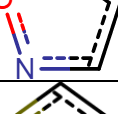


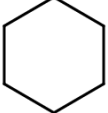
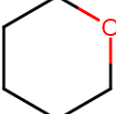
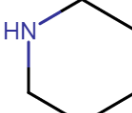
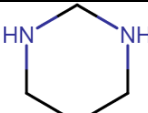
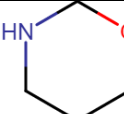
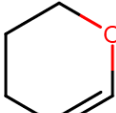

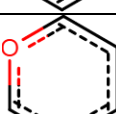
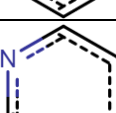
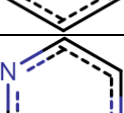

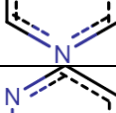
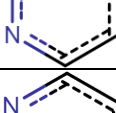
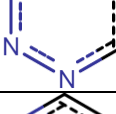
ด้วยวิธีการกำหนดส่วนย่อยให้เป็นส่วนที่มีความหมายเช่น เป็นกลุ่มวงชนิดและขนาดต่างๆ หมู่ฟังก์ชันนอล ซึ่งอาจมีจำนวนมากมาย จึงจำเป็นจะต้องอาศัยเวลาเพื่อการเก็บรวบรวมและจำแนก เพื่อเป็นการลดเวลาดังกล่าว ประกอบกับที่ได้มุ่งเน้นปรับปรุงวิธีการคำนวณค่าความเหมือน ในการศึกษาวิจัยนี้จึงได้กำหนดส่วนย่อยเบื้องต้นเพื่อให้ครอบคลุมส่วนย่อยต่างๆ ที่ปรากฏอยู่ในโมเลกุลยา แล้วจึงค่อยๆ เพิ่มส่วนย่อยอื่นๆ โดยใช้เปรียบเทียบกับค่า RTCS เทียบกับค่า TCS เพื่อให้ค่าที่แตกต่างกันนั้น เป็นผลจากวิธีการคำนวณที่ต่างกัน ไม่ใช่เพราะ


ผลจากการไม่สามารถจำแนกแยกแยะโมเลกุลย่อย เพื่อให้ครอบคลุมส่วนย่อยของโมเลกุลในฐานข้อมูลมากที่สุด จำนวนส่วนย่อยทั้งหมดที่กำหนดได้ในการวิจัยนี้มี 51 หมู่ ตามปรากฏในตารางที่ 2

ตารางที่ 2 แสดงการกำหนดและนิยามกลุ่มย่อยในโมเลกุลพร้อมการเขียนในรูปแบบ SMILES

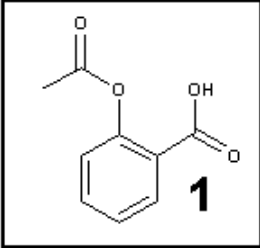
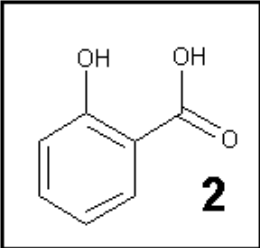
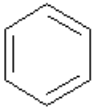
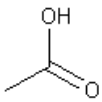

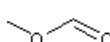

ลำดับกลุ่มย่อย	จำนวนอะตอมที่อยู่ในกลุ่มย่อย	SMILES/SMART	โครงสร้างโมเลกุล
1	1	F	F
2	1	Cl	Cl
3	1	Br	Br
4	1	I	I
5	1	P	P
6	1	[N;R]	-N-
7	1	[N+;!R]	-N <sup>+</sup>
8	1	[O;D2;R;H0]	-O-
9	2	[O;R;H1]	-OH
10	1	[O-;!R]	O <sup>-</sup>
11	1	*=[O]	=O
12	1	[C;R]	-C-
13	1	[S;R]	-S-
14	1	*=[S]	=S
15	3	C1CC1	
16	3	c1cc1	
17	4	C1CCC1	
18	4	c1ccc1	
19	5	C1CCCC1	
20	5	C1CCCO1	
21	5	C1CCCN1	
22	5	C1C=CCO1	

ลำดับกลุ่มย่อย	จำนวนอะตอมที่อยู่ในกลุ่มย่อย	SMILES/SMART	โครงสร้างโมเลกุล
23	5	<chem>C1C=CCC1</chem>	
24	5	<chem>C1OCSC1</chem>	
25	5	<chem>c1cccc1</chem>	
26	5	<chem>c1ccco1</chem>	
27	6	<chem>C1OCC=C1</chem>	
28	6	<chem>C1COCC1</chem>	
29	5	<chem>c1cccn1</chem>	
30	5	<chem>c1cccs1</chem>	
31	5	<chem>c1cncn1</chem>	
32	5	<chem>c1ccnn1</chem>	
33	5	<chem>c1ncnn1</chem>	
34	5	<chem>c1cnco1</chem>	
35	5	<chem>c1ccno1</chem>	
36	5	<chem>c1cncs1</chem>	

ลำดับกลุ่มย่อย	จำนวนอะตอมที่อยู่ในกลุ่มย่อย	SMILES/SMART	โครงสร้างโมเลกุล
37	6	C1CCCCC1	
38	6	C1OCCCC1	
39	6	C1CCCCN1	
40	6	C1NCCCN1	
41	6	C1OCCCN1	
42	6	C1OC=CC1	
43	6	c1ccccc1	
44	6	c1ccco1	
45	6	c1cccn1	
46	6	c1cnccn1	
47	6	c1ccncc1	
48	6	c1ccnnc1	
49	6	c1cnnnc1	
50	6	c1cncnn1	

ลำดับกลุ่มย่อย	จำนวนอะตอมที่อยู่ในกลุ่มย่อย	SMILES/SMART	โครงสร้างโมเลกุล
51	7	c1cnccn1	

3. วิธีการคำนวณค่า Tanimoto's coefficient score (TCS) โดยจะทำการแบ่งโมเลกุลที่นำมาเทียบกันทั้งสอง ออกเป็นส่วนย่อยๆ เรียกว่า บิต (bit) จากนั้นก็คำนวณค่าความคล้ายด้วยสัดส่วนระหว่างจำนวนบิตที่มีเหมือนกัน ในทั้งสองโมเลกุลและจำนวนบิตที่มีอยู่รวมกันทั้งสองโมเลกุล โดยสามารถแสดงตัวอย่างได้ดังนี้

					
					
	S1	S2	S3	S4	S5
1	1	1	1	1	1
2	1	1	1	0	0

เมื่อ A = จำนวนบิตที่ปรากฏ ในทั้งสองโมเลกุล (3)

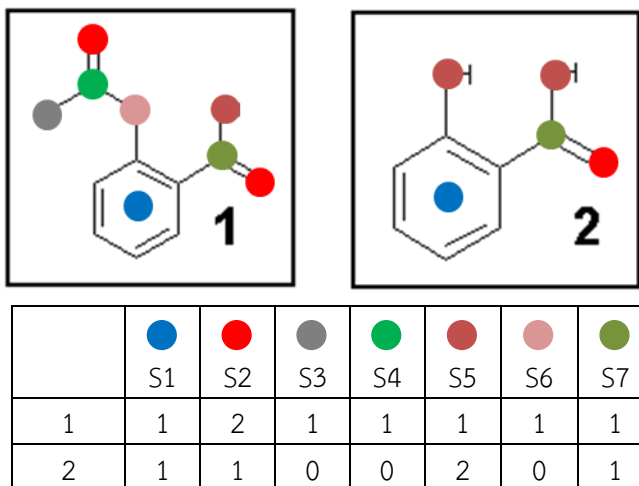
B = จำนวนบิตที่ปรากฏ ในโมเลกุล 1 แต่ไม่ปรากฏใน 2 (2)

C = จำนวนบิตที่ปรากฏ ในโมเลกุล 2 แต่ไม่ปรากฏใน 1 (0)

$$TCS = \frac{A}{A + B + C} = \frac{3}{3 + 2 + 0} = 0.60$$

จะเห็นได้ว่าการแบ่งกลุ่มย่อยแล้วนับด้วยวิธีนี้ อาจเกิดการนับซ้ำเช่นกรณี S2 และ S3 ในโมเลกุล 1 ในขณะที่ กรณี S3 ในโมเลกุล 2 มีปรากฏ 2 ครั้งแต่ก็นับเพียง 1 เท่านั้น

ดังนั้นในงานวิจัยนี้จึงได้เสนอแนวทางปรับปรุงเพื่อคำนวณค่า TCS ใหม่และเรียกว่า refinement Tanimoto's coefficient score (RTCS) โดยมีหลักการเบื้องต้นว่า การกำหนดกลุ่มโมเลกุลย่อยที่จะปรากฏหรือไม่ปรากฏในคู่มอเลกุลที่นำมาเทียบกันนั้น ควรจะกำหนดให้น้อยที่สุดเป็นโครงสร้างที่มีความหมาย เช่น ครอบวงโมเลกุล หรือครบหมู่ฟังก์ชัน ตามที่ได้กำหนดนิยามไว้ และให้มีการนับจำนวนที่ปรากฏอย่างครบถ้วนแต่ไม่ซ้ำซ้อน โดยอาศัยกราฟลดรูปมาช่วยในการคำนวณ



เมื่อ  $k$  = จำนวนส่วนย่อยของโมเลกุลที่กำหนดขึ้น

$N_{S_i}$  = จำนวนอะตอมที่อยู่ในแต่ละส่วนย่อย  $S_k$

$A$  = ผลรวมของ จำนวนบิตที่ปรากฏ ในทั้งสองโมเลกุลคูณจำนวนอะตอมของบิต

$$= \sum_{i=1,k} \min(S_{i1}, S_{i2}) \times N_{S_i}$$

$$= (1 \times 6)_{S1} + (1 \times 1)_{S2} + (0 \times 1)_{S3} + (0 \times 1)_{S4} + (1 \times 1)_{S5} + (0 \times 1)_{S6} + (1 \times 1)_{S7} = 9$$

$B$  = จำนวนบิตที่ปรากฏ ในโมเลกุล 1 แต่ไม่ปรากฏในโมเลกุล 2

$$= \sum_{i=1,k} (S_{i1} - S_{i2}) \times N_{S_i} ; \text{if } (S_{i1} - S_{i2}) \geq 0$$

$$= (0 \times 6)_{S1} + (1 \times 1)_{S2} + (1 \times 1)_{S3} + (1 \times 1)_{S4} + (1 \times 1)_{S6} + (0 \times 1)_{S7} = 4$$

$C$  = จำนวนบิตที่ปรากฏ ในโมเลกุล 2 แต่ไม่ปรากฏในโมเลกุล 1

$$= \sum_{i=1,k} |S_{i1} - S_{i2}| \times N_{S_i} ; \text{if } (S_{i1} - S_{i2}) < 0$$

$$= (1 \times 1)_{S5} = 1$$

$$RTCS = \frac{A}{A + B + C} = \frac{9}{9 + 4 + 1} = 0.64$$

การคำนวณค่า RTCS เปรียบเทียบกับ TCS ของข้อมูลโครงสร้างยาได้แก่ Ribavirin, Lopinavir, Ritonavir, Calanolide A และ Navirapine เปรียบเทียบกับโครงสร้างโมเลกุลยาในฐานข้อมูล DrugBank จำนวน 6979 โมเลกุล และโมเลกุลเคมีในฐานข้อมูล PubChem จำนวน 25000 โมเลกุล โดยค่า TCS คำนวณได้โดยตรงจากโปรแกรม Openbabel

4. พัฒนาปรับปรุงขั้นตอนวิธีนับโครงสร้างย่อย (Substructure Counting Algorithm, SCA) เพื่อนับจำนวนเพื่อการสืบค้นที่มีประสิทธิภาพและรวดเร็ว โดยอาศัยโปรแกรมประยุกต์ OpenBabel [17] ที่มีคำสั่ง obgrep สามารถสกัดโครงสร้างย่อยจากรูปแบบ SMILES ได้โดยอาศัยการเขียนนิพจน์สืบค้นแบบ SMART [18] ทำให้สามารถพัฒนาขั้นตอนวิธีที่สามารถสกัดทั้งชนิดและจำนวนของโครงสร้างย่อยที่อยู่ในโครงสร้างโมเลกุลต่างๆได้ ในการศึกษานี้ได้ใช้ภาษา awk ซึ่งเป็นตัวแปลคำสั่ง (interpreter) ที่มีความสามารถในการรับและประมวลผลข้อมูลที่อยู่ในรูปของข้อความ เพื่ออ่านข้อความ SMILES/SMART ของแต่ละโครงสร้างย่อย แล้วส่งเข้าไปสืบค้นหาจำนวนในแต่ละโครงสร้างโมเลกุลยา และโมเลกุลอื่นๆในฐานข้อมูล เพื่อจะสร้างตารางเมตริกซ์ที่บรรจุ ชนิดของโครงสร้างย่อย และจำนวนของโครงสร้างย่อยที่ปรากฏในแต่ละโมเลกุล ความเร็วในการประมวลผลนี้จะขึ้นอยู่กับจำนวนของโครงสร้างย่อยและจำนวนโมเลกุลที่นำมาคำนวณ อย่างไรก็ตาม ฐานข้อมูลจะมีจำนวนโมเลกุลขนาดใหญ่เช่นมีโมเลกุลหลายล้าน การคำนวณนี้ก็จะดำเนินการแค่ครั้งเดียวในตอนเริ่มต้น แล้วเก็บไว้เป็น

ตารางข้อมูลเชิงตัวเลขไว้ หากมีการปรับปรุงส่วนย่อยเพิ่มเติม หรือมีจำนวนโมเลกุลที่จะต้องคำนวณเพิ่มในภายหลัง ก็จะทำเนิการเฉพาะส่วนที่เพิ่มเท่านั้น ไม่จำเป็นต้องประมวลผลตารางใหม่ทั้งหมด โปรแกรมคำนวณค่า RTCS เขียนโดยภาษา Fortran 90 โดย จะอ่านข้อมูลเข้าเป็นตารางข้อมูลจำนวนของแต่ละส่วนย่อยที่มีในแต่ละโมเลกุลยาหรือโมเลกุลในฐานข้อมูล แล้วคำนวณ ค่า RTCS ของคู่ยาและโมเลกุลที่ต้องการ ผลการคำนวณค่า RTCS เทียบกับ TCS แสดงได้รูปที่ 3 เป็นต้น

5. นำโครงสร้างโมเลกุลที่ได้จากการสืบค้นมาคำนวณพลังงานในการยึดจับกับโปรตีนเอสของโคโรนาไวรัส โดยวิธีการเข้าจับเชิงโมเลกุล โดยใช้โปรแกรม Arguslab 4.0.1 [19] ซึ่งเป็นโปรแกรมอรรถประโยชน์เชิงโมเลกุลที่สามารถที่จะอ่านเข้า ข้อมูลโครงสร้าง เพื่อสร้างหรือแก้ไข การวิเคราะห์ คำนวณด้านโครงสร้างและพลังงานต่างๆ รวมถึงการเชื่อมต่อเพื่อเรียกโปรแกรมภายนอกสำหรับการคำนวณระดับสูง Arguslab มีความโดดเด่นในการคำนวณด้านการเข้าจับเชิงโมเลกุล [20,21] โดยเฉพาะการคำนวณระหว่างเป้าหมายกับฐานข้อมูลโมเลกุล หลักการของการเข้าจับเชิงโมเลกุล คือการนำโมเลกุลเคมี เข้าไปหาตำแหน่งหรือโครงสร้างการยึดจับที่ดีที่สุดซึ่งให้ค่าพลังงานการยึดจับต่ำที่สุด อย่างไรก็ตามโครงสร้างการยึดจับมีความเป็นไปได้นับไม่ถ้วน และไม่สามารถคำนวณได้ครบทุกๆโครงสร้าง เพื่อให้ได้คำตอบที่แม่นยำ การคำนวณส่วนใหญ่จึงเป็นการหาค่าที่เหมาะสมที่สุด ที่ได้จากการสุ่มวางอย่างมีกลยุทธ์ และได้ผลเฉลยเป็นค่าที่ดีที่สุดที่เกิดจากการคำนวณ ใน Arguslab ได้เลือกกำหนดวิธีการหาค่าเหมาะสมด้วย Genetic Algorithm (GA) ซึ่งกำหนดจำนวนรุ่นเป็น 5000 รุ่น จำนวนประชากรเป็น 50 และมีการคำนวณค้นแบบเฉพาะที่อีก 50 ครั้ง ขนาดของกล่องลูกบาศก์ยาว 15 Å (Angstrom) ครอบคลุมบริเวณยึดจับที่ได้จากเอกสารอ้างอิง [6] หมายเลขโมเลกุลทั้งหมดที่ผ่านเกณฑ์ 0.6 โดยพิจารณาทั้ง TCS และ RTCS ได้นำไปคำนวณหาพลังงานในการจับด้วยวิธี การเข้าจับเชิงโมเลกุล ซึ่งได้ทำการแปลงรูป SMILES ของทุกหมายเลขนั้นให้อยู่ในรูป ฐานข้อมูลโครงสร้างสามมิติรูปแบบ SDF มีการคำนวณเพื่อเติม H ให้โมเลกุลมีความสมบูรณ์ ในขณะที่หมายเลขที่ใช้อ้างอิงในงานวิจัยนี้ ถูกกำหนดขึ้นมาใหม่ เพื่อให้ง่ายต่อการประมวลผล ไม่ได้จำเป็นต้องถูกต้องตามหมายเลขอ้างอิงจริงๆ บนฐานข้อมูล ผลการคำนวณการเข้าจับเชิงโมเลกุลแสดงได้ดังตารางที่ 4

6. เรียบเรียง พัฒนาปรับปรุงสร้างกระบวนการสืบค้นจนถึงทดสอบให้มีความเหมาะสมถูกต้อง เพื่อจัดทำระบบสารสนเทศให้มีการเข้าถึงได้ แต่เนื่องด้วย การศึกษาในขั้นตอนที่ 1-4 ได้ใช้เวลามากโดยเฉพาะขั้นตอนที่ 3 จึงทำให้การพัฒนาาระบบสารสนเทศ เพื่อให้บริการไม่สามารถดำเนินการได้ทันในระยะเวลาวิจัย

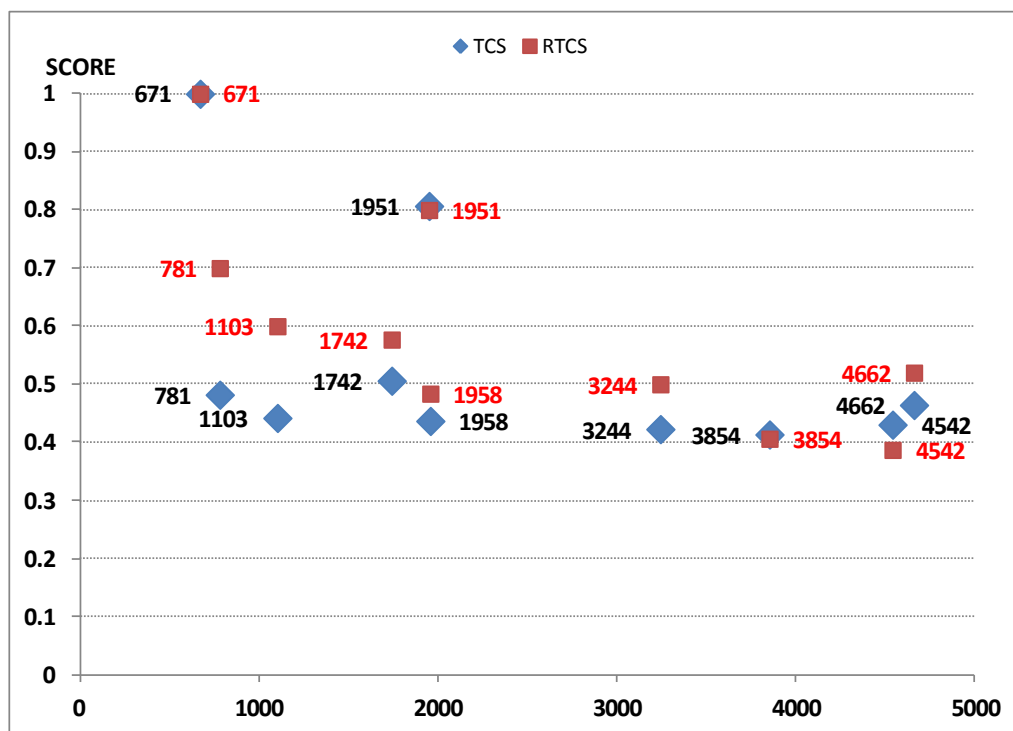
## ผลการวิจัย

1. ผลการคำนวณเปรียบเทียบค่า RTCS และ TCS ของโมเลกุลยา Ribavirin Lopinavir Ritonavir Calanolide A และ Navirapine เทียบกับโมเลกุลยาในฐานข้อมูล DrugBank จำนวน 6797 หมายเลข และในฐานข้อมูล Pubchem จำนวน 25000 หมายเลขแรก โดยโดยค่าเฉลี่ยความต่างของ RTCS กับ TCS และหมายเลขโมเลกุลที่ผ่านเกณฑ์ 0.6 แสดงตามตารางที่ 3

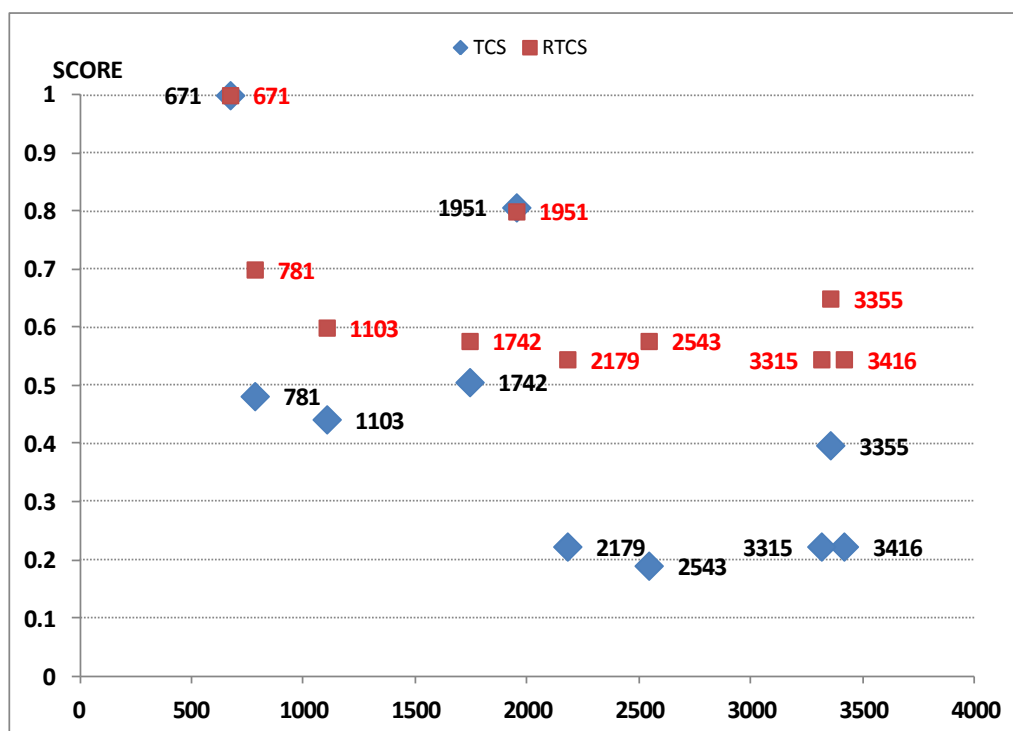
จะเห็นได้ว่าวิธีการคำนวณทั้งสองวิธีนั้นพบโมเลกุลที่มีค่าความคล้าย RTCS และ TCS เป็น 1.0 และให้ผลลัพธ์ที่เป็นเลขเดียวกัน(หมายเลขแรก) ในแต่ละฐานข้อมูล เมื่อพิจารณาค่าเฉลี่ยแล้วจะพบว่ายา Ribavirin Lopinavir Ritonavir และ Calanolide A จะมีผลที่เหมือนกัน คือมีค่าเฉลี่ยความต่างเป็นบวก นั้นหมายความว่าโดยเฉลี่ยแล้วค่า RTCS จะสูงกว่าค่า TCS และสามารถพบจำนวนโมเลกุลผ่านเกณฑ์มากกว่า โดยเฉพาะโมเลกุล Lopinavir ที่ค่า RTCS ให้จำนวนโมเลกุลที่ผ่านเกณฑ์ถึง 73 และ 102 ในฐานข้อมูล DrugBank และ PubChem ตามลำดับ ในขณะที่โมเลกุล Navirapine ให้ผลที่ต่างไป คือมีค่าเฉลี่ย TCS สูงกว่า RTCS และ RTCS ไม่พบโมเลกุลอื่นๆผ่านเกณฑ์เลยในทั้งสองฐานข้อมูล

ตารางที่ 3 แสดงค่าเฉลี่ยความต่าง (RTCS - TCS) และหมายเลขโมเลกุลที่ผ่านเกณฑ์คะแนน 0.6 ขึ้นไป โดยการเรียงลำดับจากคะแนนสูงไปต่ำด้วยวิธี RTCS และ TCS ตามลำดับ

โมเลกุลยา	DrugBank	PubChem
Ribavirin	ค่าเฉลี่ยความต่าง: 0.013 RTCS : 671, 1951, 781, 3355, 1103 TCS: 671,1951	ค่าเฉลี่ยความต่าง: 0.016 RTCS : 5064, 1805, 9444, 1610, 17723, 608, 494, 16886, 952 TCS: 5064
Lopinavir	ค่าเฉลี่ยความต่าง: 0.056 RTCS : 1250,4488, 4483,4403,1941,1978,1489,4372, 2119, 2870, 2070 4705,396, 1643, 4210,2648, 3280, 2137, 4075, 3435, 4529,4465, 3849,4511, 2250, 1980, 4733, 2389, 4795,1391, 4780,4526, 1782, 2652, 1833, 5621, 1397,4432, 6732,3624, 1692, 6167, 6,1937, 1061, 5607, 555,21523239, 2942, 4594, 2721, 5121, 6129, 739, 6774,1892, 2432, 1637, 6644, 6133, 5906, 5069, 924,2060, 5709, 1887, 3732, 4622, 1072, 525, 1740, 5314 TCS: 1250, 4488, 2652, 4483, 642	ค่าเฉลี่ยความต่าง: 0.075 RTCS : 3817, 1644, 1724, 3569, 1519, 1709, 23672, 4319, 3854, 2855, 5489, 1517, 2872, 10080, 23557, 4502, 4621, 1356, 5415, 4318, 10079, 22898, 4785, 5739, 2976, 23225, 5107, 5903, 22502, 5249, 15232, 4863, 10650, 3837, 24392, 6222, 6536, 24721, 14886, 1269, 23066, 24885, 15931, 23570, 6530, 3507, 5377, 5305, 19553, 13986, 4503, 19518, 2745, 2594, 3839, 18764, 13985, 19548, 4297, 22020, 4504, 9045, 10304, 9394, 3204, 3903, 5180, 5005, 2339, 19253, 19154, 22900, 23227, 19551, 21679, 17392, 21325, 5045, 15930, 3497, 5176, 2329, 5160, 5088, 19557, 19555, 21147, 24903, 20168, 19737, 15482, 2884, 6765, 5730, 21718, 9041, 4864, 4666, 2520 TCS: 3817, 1644, 4863, 1724, 1408, 1587, 1358, 4868
Ritonavir	ค่าเฉลี่ยความต่าง: 0.063 RTCS : 375, 4733, 4815, 4622, 1824, 2137, 4019 796, 3160, 2942, 4292, 2652, 1833, 4465, 4642 4809, 4678, 6671, 1391, 1782, 3934, 2884, 1941 2597, 1892, 1248 TCS: 375	ค่าเฉลี่ยความต่าง: 0.089 RTCS: 5076, 4321, 1327, 4621, 4864, 5489, 20697, 3903, 1339, 5966, 2317, 2565, 19561, 9041, 22898, 19559,2745, 5903, 22502, 18305, 4338, 4863,5965, 10650, 19557, 19555, 1269, 9040, 24885, 5170, 23570, 20696, 9043, 2594, 9560, 3839, 4337, 2976, 5167, 10579, 19548, 15284, 19558, 9045, 9394, 5249, 10304, 5777, 1566, 5176, 4912 TCS: 5076
Calanolide A	ค่าเฉลี่ยความต่าง: 0.75 RTCS : 1721, 2436, 342 TCS: 1721	ค่าเฉลี่ยความต่าง: 0.098 RTCS : 1201, 1421, 1422, 1419, 1417, 5254, 1247 10158, 10157, 4871, 1418, 1878, 2977, 2978, 16078, 22462 TCS: 1201, 1417, 1421, 1422, 1419, 5254, 1247, 1418, 1878
Navirapine	ค่าเฉลี่ยความต่าง: -0.109 RTCS : 125 TCS: 125, 4796, 6666, 31	ค่าเฉลี่ยความต่าง: -0.075 RTCS : 4463 TCS: 4463, 1072, 6759, 21500

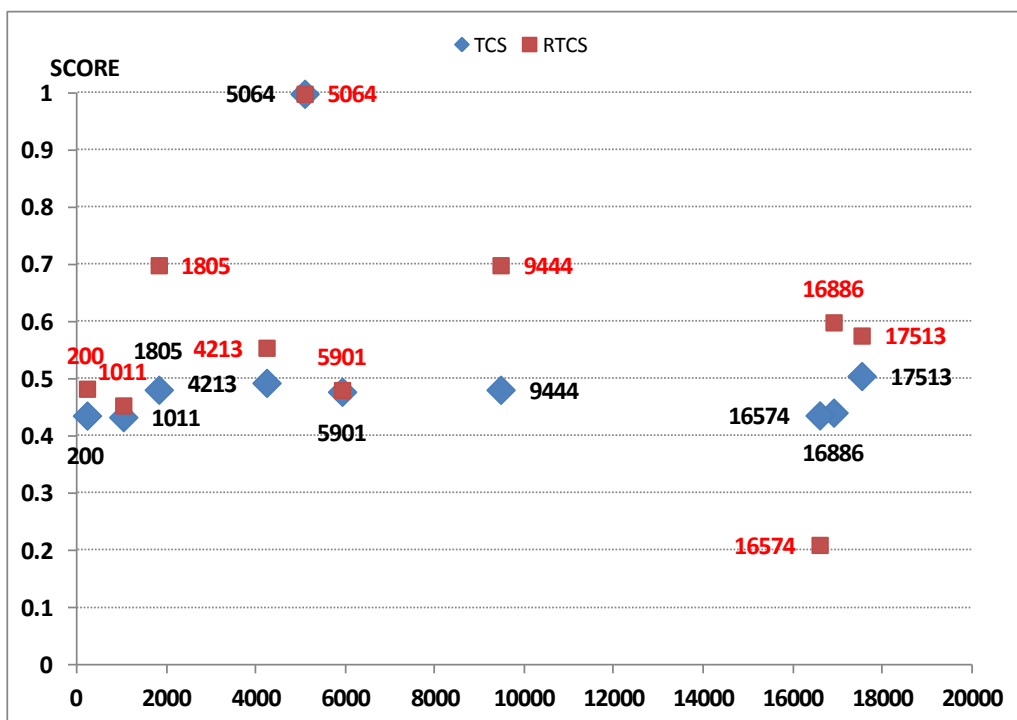


ก) เรียงตามค่าความเหมือน TCS สูงสุด 10 อันดับแรก

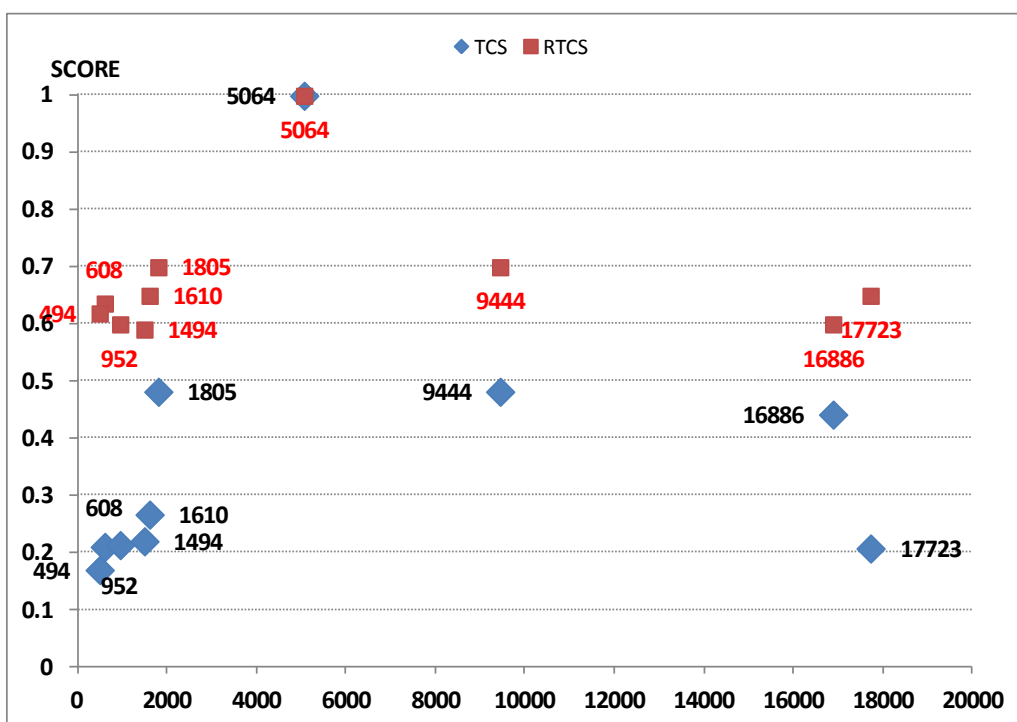


ข) เรียงตามค่าความเหมือน RTCS สูงสุด 10 อันดับแรก

รูปที่ 3 เปรียบเทียบค่า TCS และ RTCS ของโมเลกุล Ribavirin กับ โมเลกุลสารประกอบเคมีในฐานข้อมูล DrugBank จำนวน 6797 หมายเลข

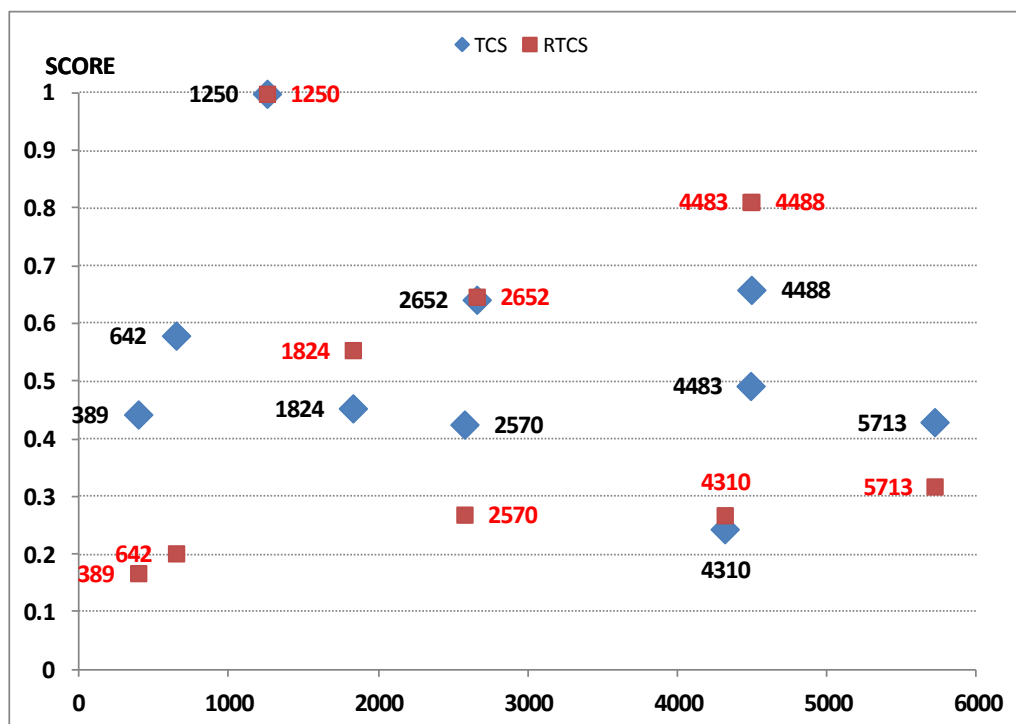


ก) เรียงตามค่าความเหมือน TCS สูงสุด 10 อันดับแรก

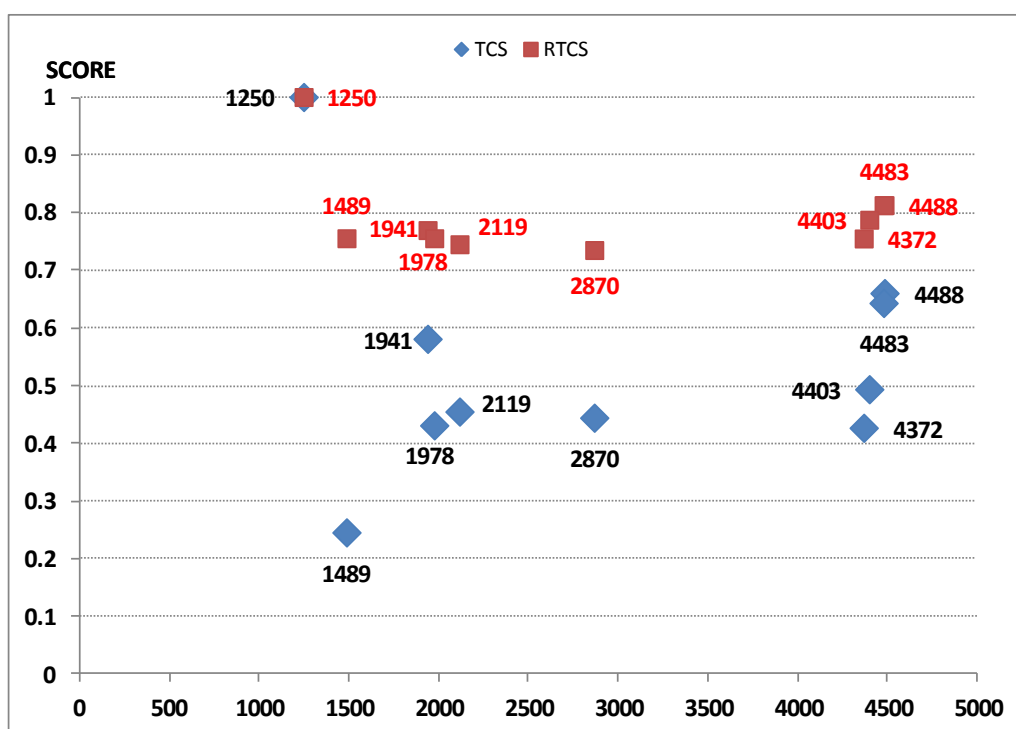


ข) เรียงตามค่าความเหมือน RTCS สูงสุด 10 อันดับแรก

รูปที่ 4 เปรียบเทียบค่า TCS และ RTCS ของโมเลกุล Ribavirin กับ โมเลกุลสารประกอบเคมีในฐานข้อมูล Pubchem จำนวน 25000 หมายเลขแรก

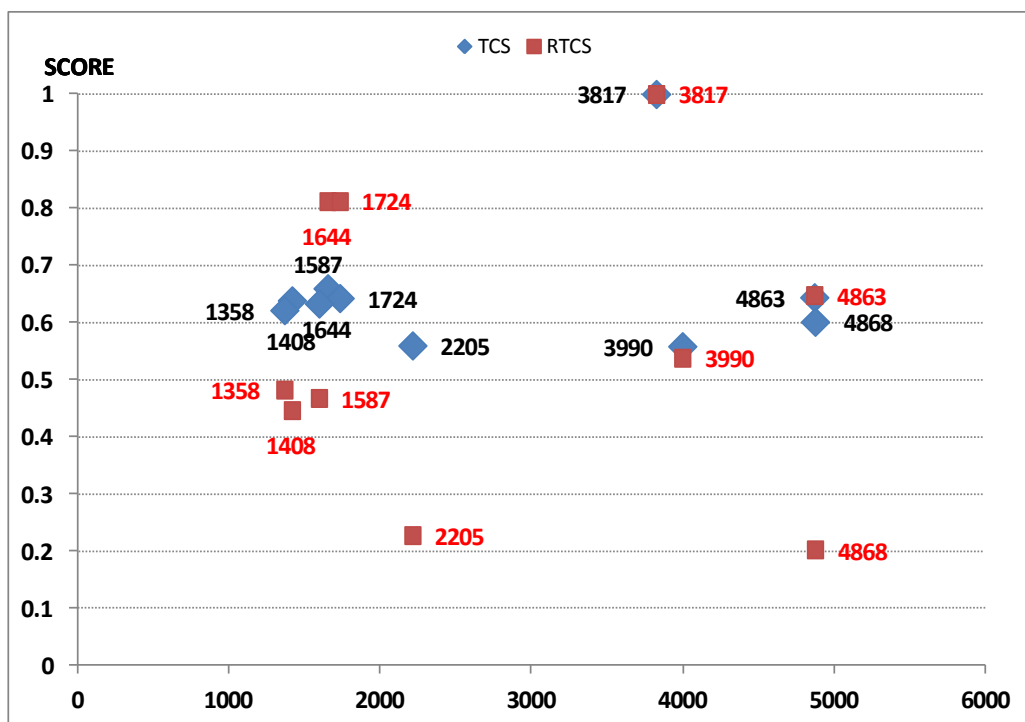


ก) เรียงตามค่าความเหมือน TCS สูงสุด 10 อันดับแรก

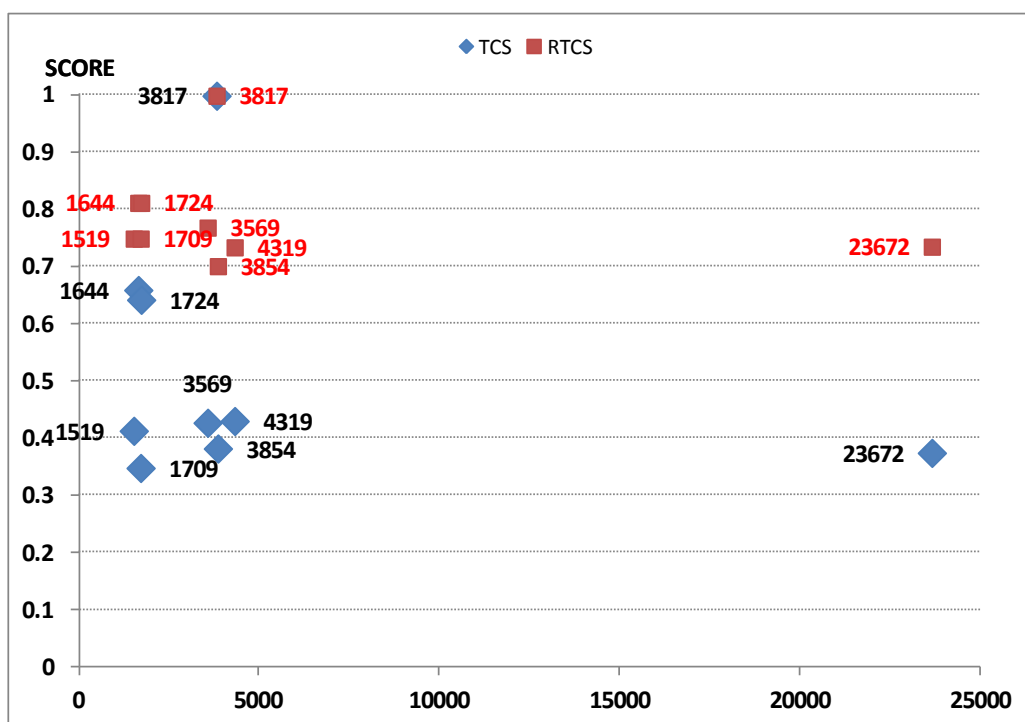


ข) เรียงตามค่าความเหมือน RTCS สูงสุด 10 อันดับแรก

รูปที่ 5 เปรียบเทียบค่า TCS และ RTCS ของโมเลกุล Lopinavir กับ โมเลกุลสารประกอบเคมีในฐานข้อมูล DrugBank จำนวน 6797 หมายเลข

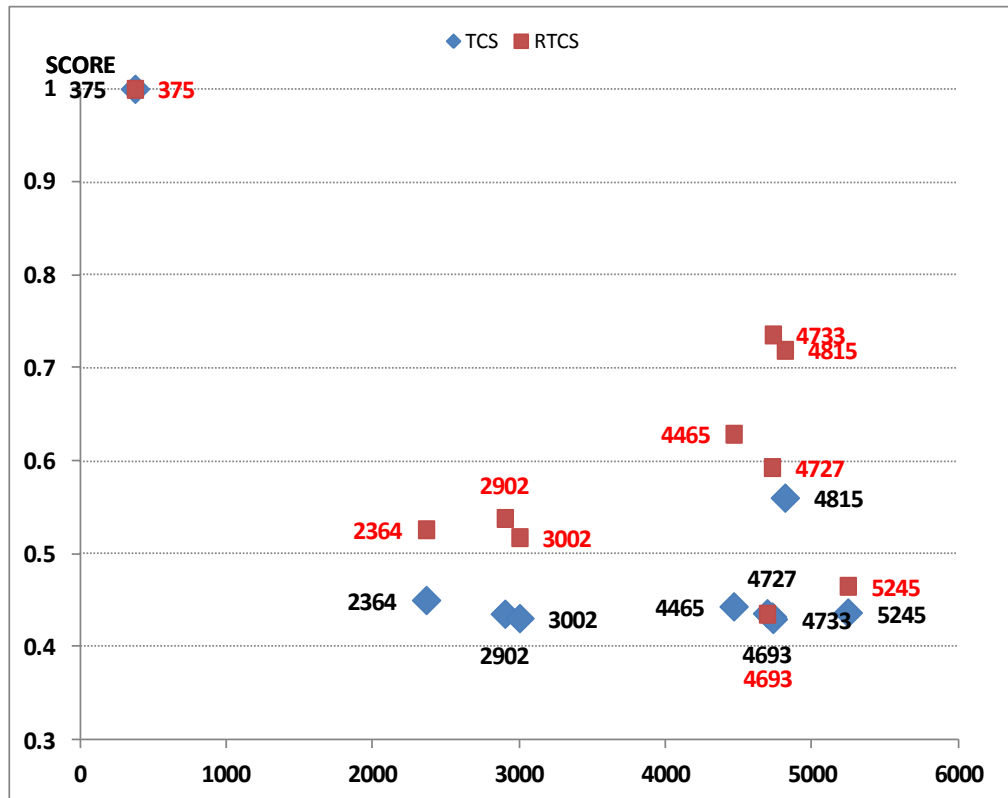


ก) เรียงตามค่าความเหมือน TCS สูงสุด 10 อันดับแรก

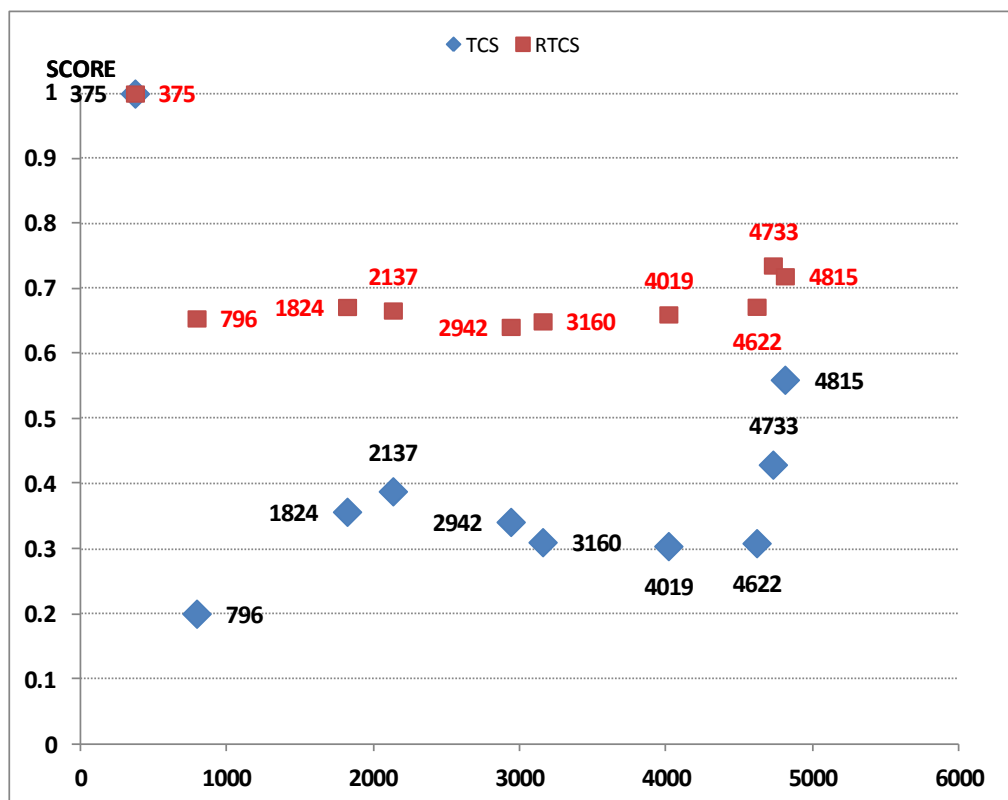


ก) เรียงตามค่าความเหมือน TCS สูงสุด 10 อันดับแรก

รูปที่ 6 เปรียบเทียบค่า TCS และ RTCS ของโมเลกุล Lopinavir กับ โมเลกุลสารประกอบเคมีในฐานข้อมูล Pubchem จำนวน 25000 หมายเลขแรก

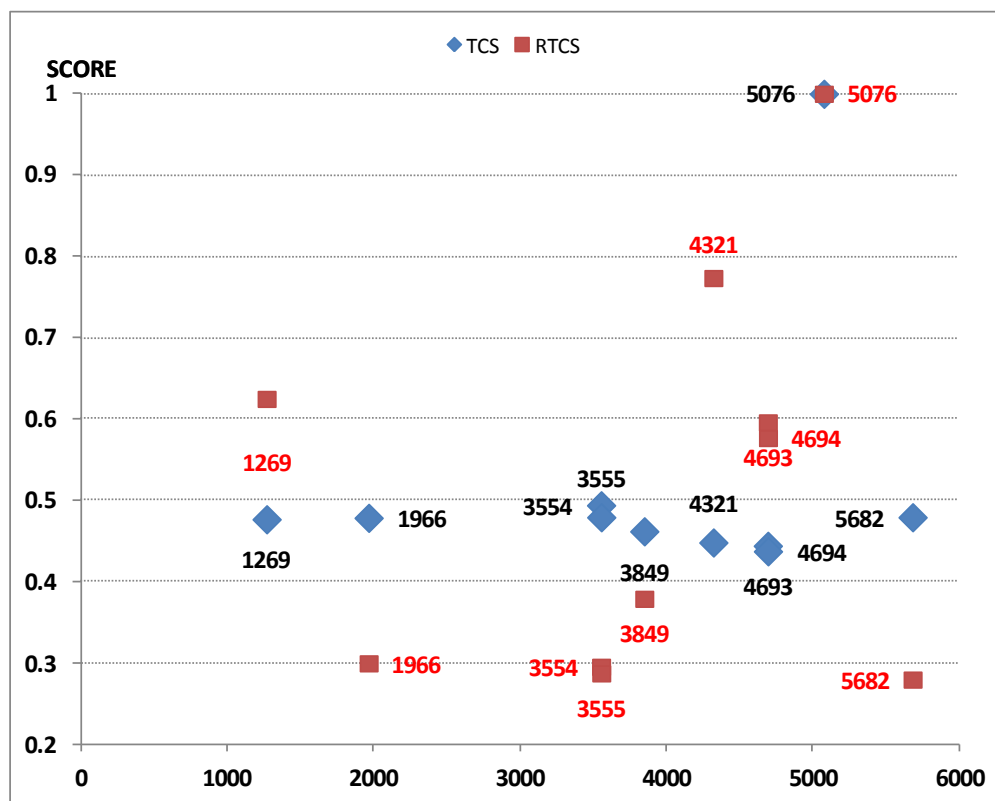


ก) เรียงตามค่าความเหมือน TCS สูงสุด 10 อันดับแรก

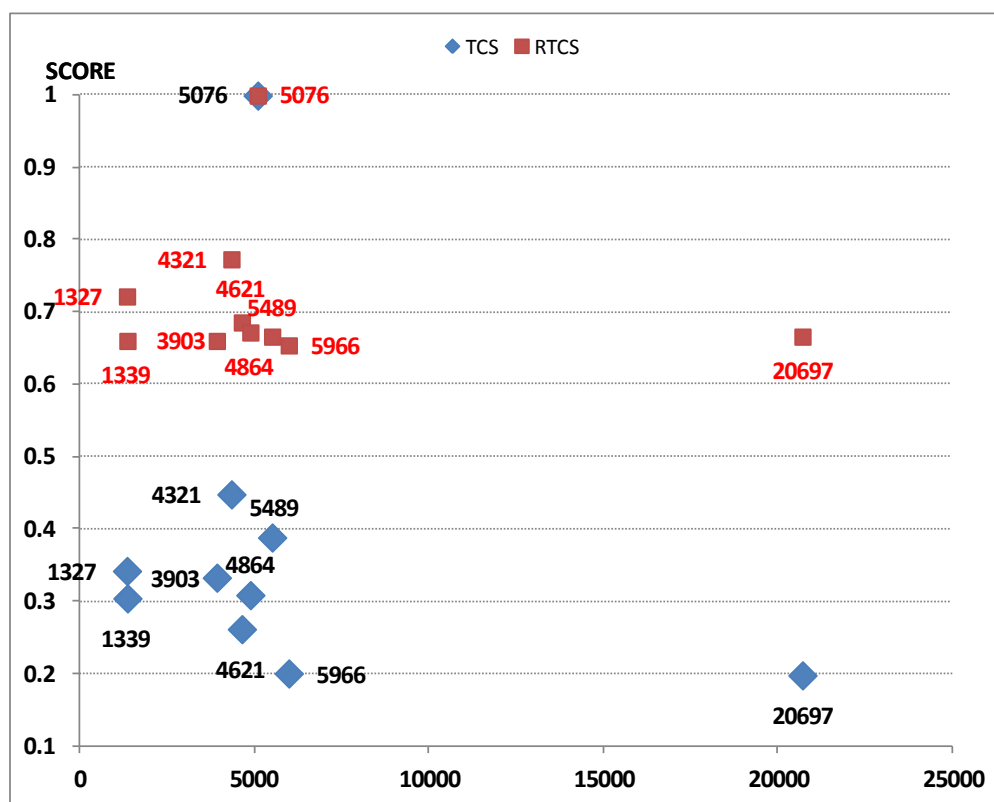


ข) เรียงตามค่าความเหมือน RTCS สูงสุด 10 อันดับแรก

รูปที่ 7 เปรียบเทียบค่า TCS และ RTCS ของโมเลกุล Ritonavir กับ โมเลกุลสารประกอบเคมีในฐานข้อมูล DrugBank จำนวน 6797 หมายเลข

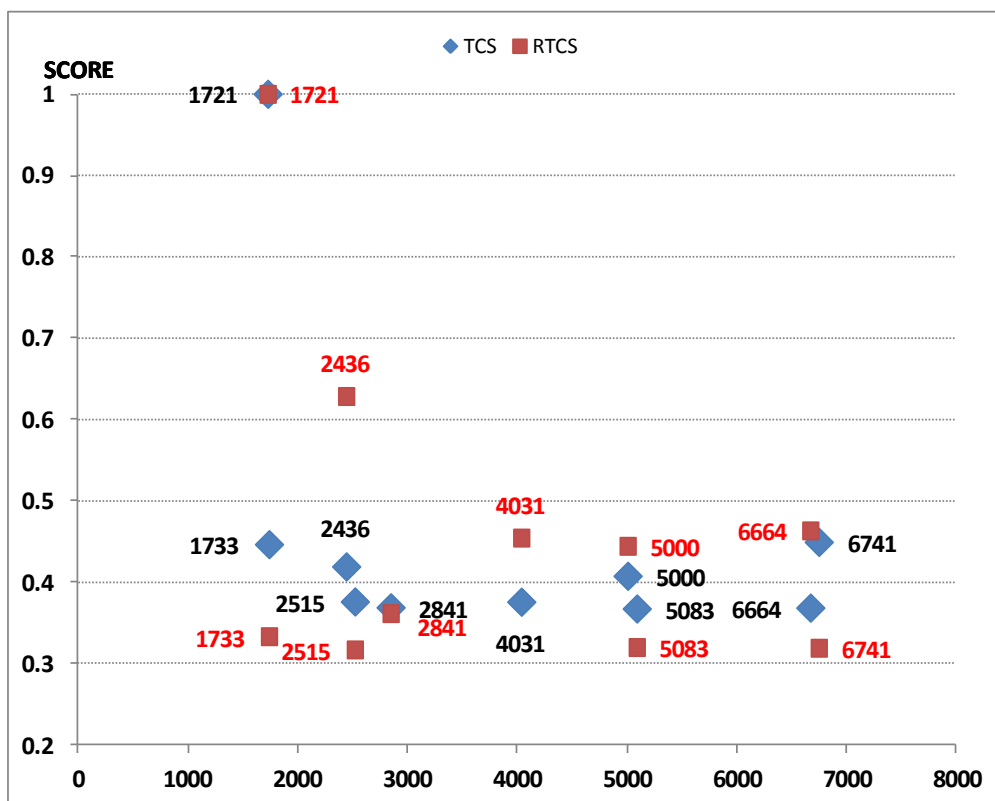


ก) เรียงตามค่าความเหมือน TCS สูงสุด 10 อันดับแรก

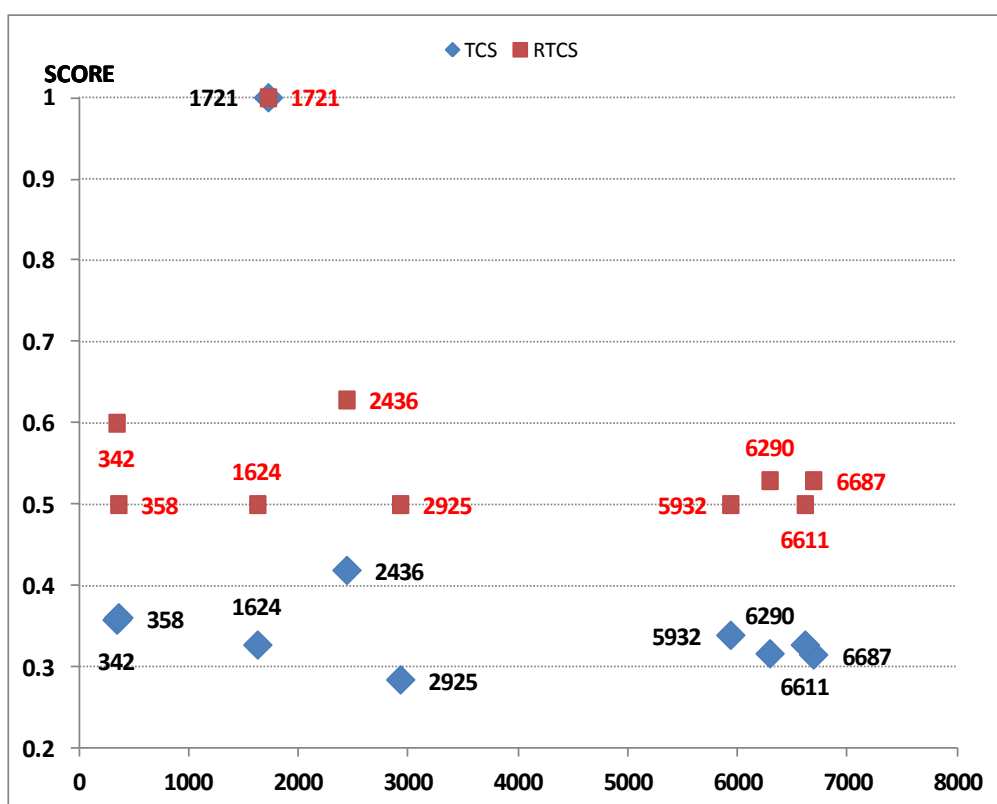


ข) เรียงตามค่าความเหมือน RTCS สูงสุด 10 อันดับแรก

รูปที่ 8 เปรียบเทียบค่า TCS และ RTCS ของโมเลกุล Ritonavir กับ โมเลกุลสารประกอบเคมีในฐานข้อมูล Pubchem จำนวน 25000 หมายเลขแรก

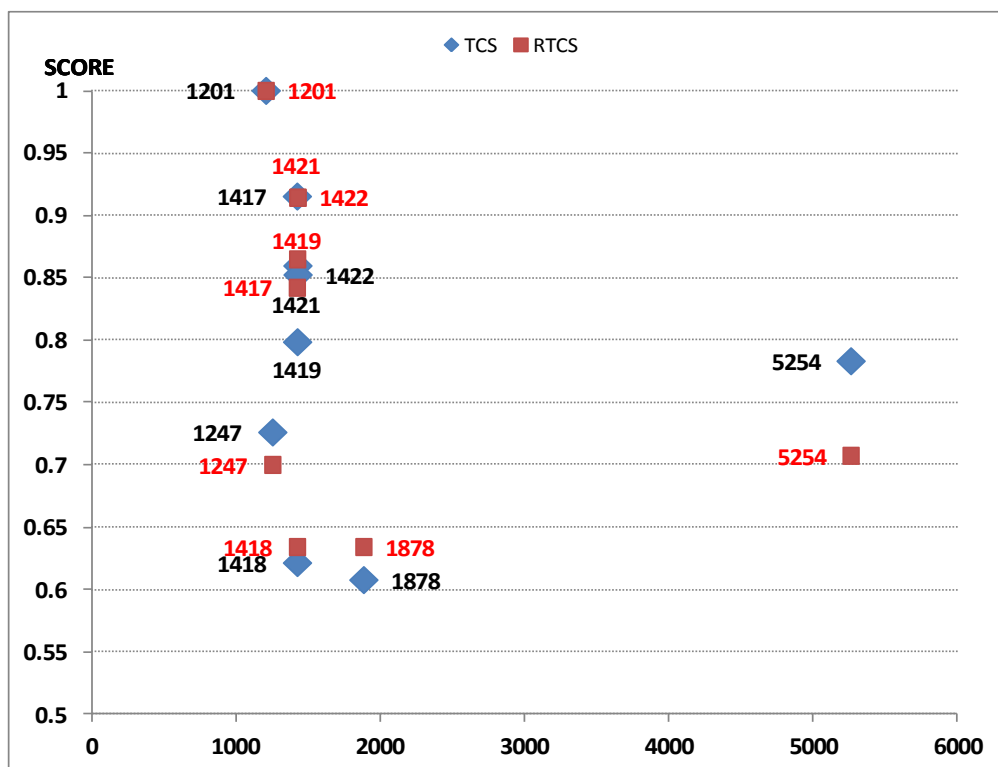


ก) เรียงตามค่าความเหมือน TCS สูงสุด 10 อันดับแรก

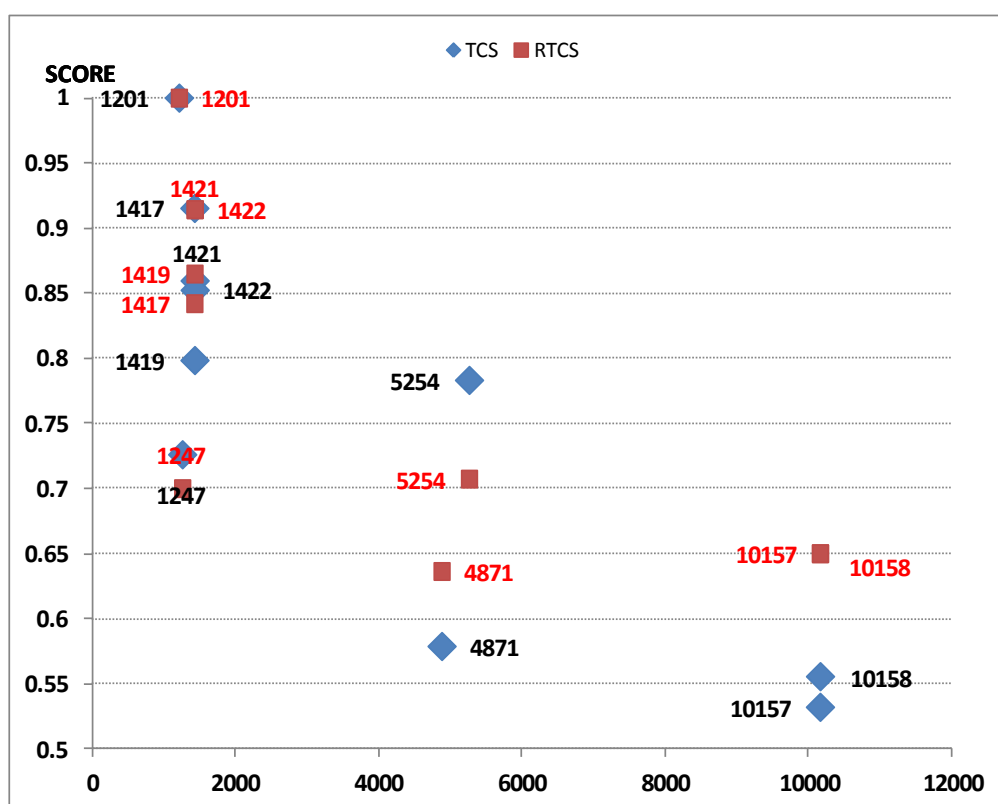


ข) เรียงตามค่าความเหมือน RTCS สูงสุด 10 อันดับแรก

รูปที่ 9 เปรียบเทียบค่า TCS และ RTCS ของโมเลกุล Calanolide A กับ โมเลกุลสารประกอบเคมีในฐานข้อมูล DrugBank จำนวน 6797 หมายเลข

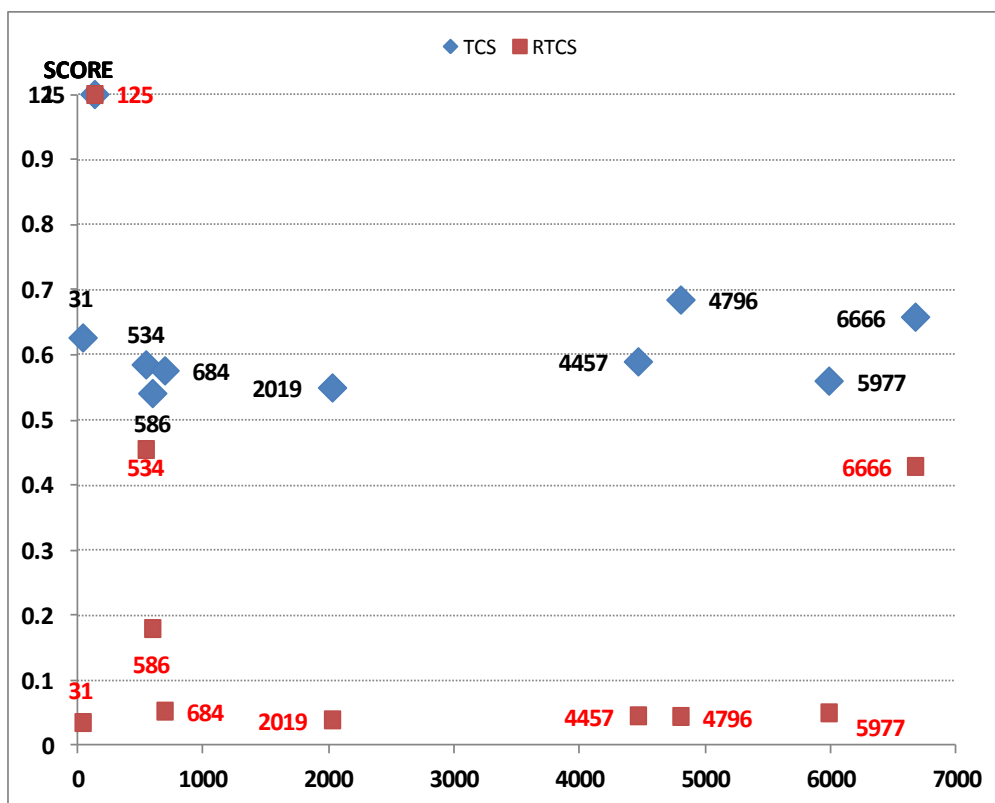


ก) เรียงตามค่าความเหมือน TCS สูงสุด 10 อันดับแรก

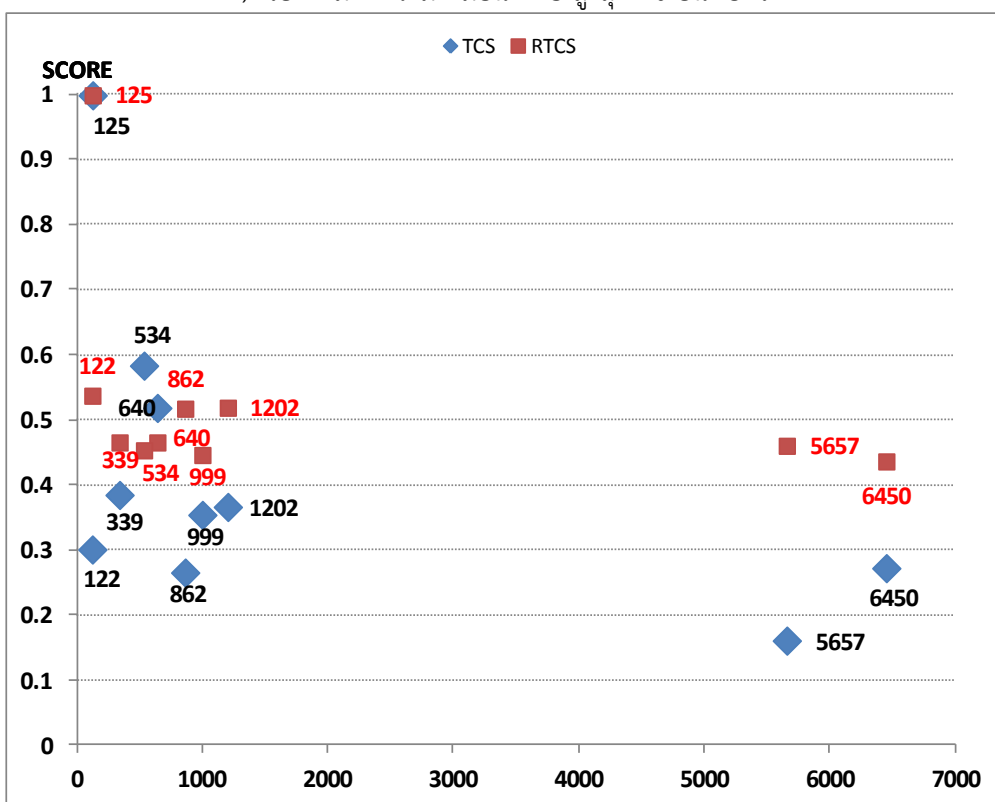


ข) เรียงตามค่าความเหมือน RTCS สูงสุด 10 อันดับแรก

รูปที่ 10 เปรียบเทียบค่า TCS และ RTCS ของโมเลกุล Calanolide A กับ โมเลกุลสารประกอบเคมีในฐานข้อมูล Pubchem จำนวน 25000 หมายเลขแรก

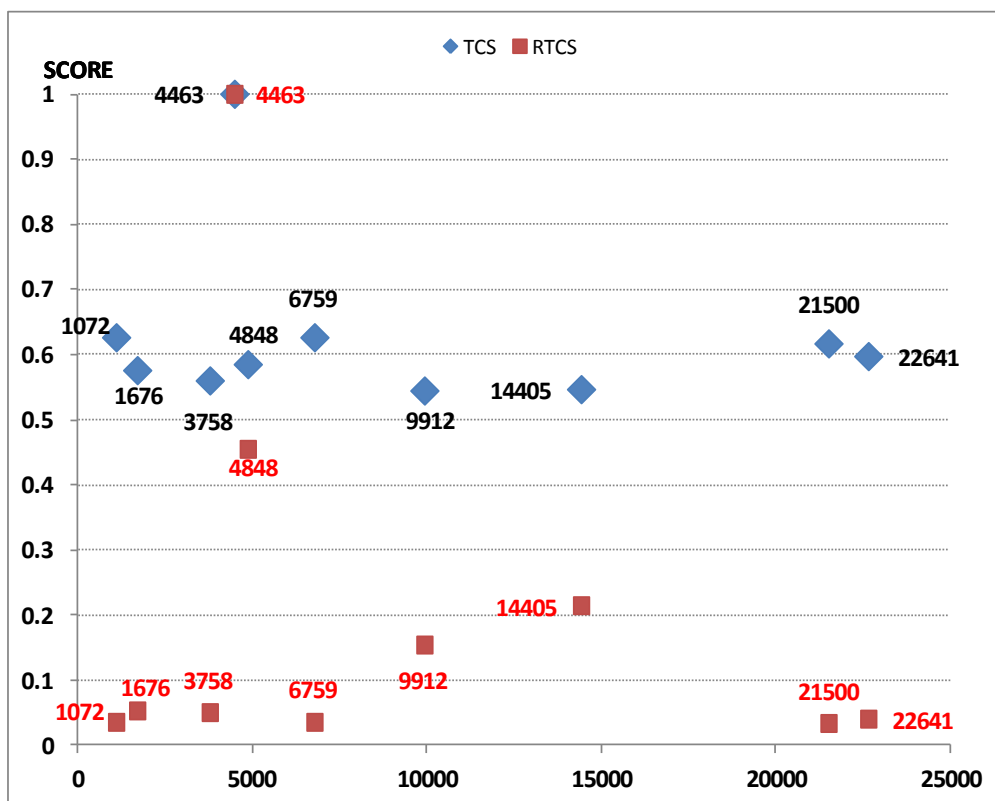


ก) เรียงตามค่าความเหมือน TCS สูงสุด 10 อันดับแรก

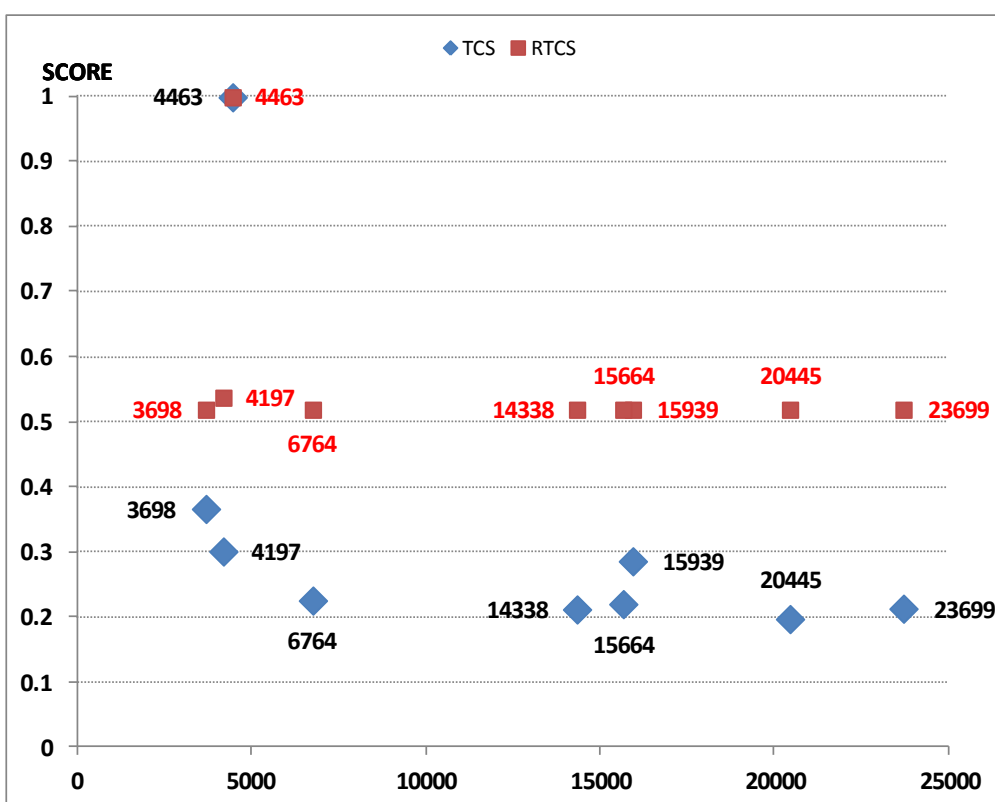


ข) เรียงตามค่าความเหมือน RTCS สูงสุด 10 อันดับแรก

รูปที่ 11 เปรียบเทียบค่า TCS และ RTCS ของโมเลกุล Navirapine กับ โมเลกุลสารประกอบเคมีในฐานข้อมูล DrugBank จำนวน 6797 หมายเลข



ก) เรียงตามค่าความเหมือน TCS สูงสุด 10 อันดับแรก



ข) เรียงตามค่าความเหมือน RTCS สูงสุด 10 อันดับแรก

รูปที่ 12 เปรียบเทียบค่า TCS และ RTCS ของโมเลกุล Navirapine กับ โมเลกุลสารประกอบเคมีในฐานข้อมูล Pubchem จำนวน 25000 หมายเลขแรก

2. จากรูปที่ 4 -12 แสดงเปรียบเทียบค่าความเหมือนสูงสุด 10 อันดับแรกเรียงโดยใช้ TCS และ RTCS จะเห็นได้ชัดเจนว่าโดยเฉลี่ยแล้วค่า RTCS จะให้ค่าที่สูงกว่า TCS และลำดับของหมายเลขเรียงจากมากไปน้อยที่ต่างกัน เช่น สำหรับยา Ribavirin จะพบว่าโมเลกุลที่มีค่า TCS สูงสุดเรียงตามลำดับหมายเลขเป็น 671>1951>1742>781>4662 .. ในขณะที่ถ้าเรียงตามค่าของ RTCS จะได้ลำดับหมายเลขเป็น 671>1951>781>3355>1103 แต่สิ่งที่น่าสนใจคือ หากพิจารณาผลที่ได้จากการเรียงด้วยค่า TCS รูป ก) มากกว่าหรือเท่ากับ 0.5 ขึ้นไป ส่วนใหญ่จะพบว่า RTCS > TCS เช่นกัน ในขณะที่ถ้าค่า TCS น้อยกว่า 0.5 จะพบมีค่า RTCS<TCS

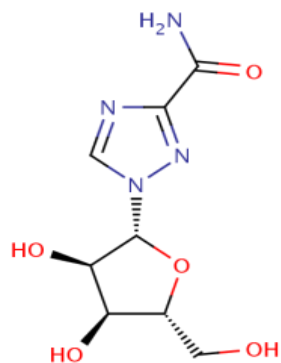
**ตารางที่ 4** แสดงผลการคำนวณพลังงานยึดจับ (kcal/mol) ที่ได้จากวิธีเข้าจับเชิงโมเลกุล เฉพาะที่มีค่าพลังงานยึดจับต่ำกว่า -3.0 kcal/mol

โมเลกุลยา	DrugBank	PubChem
Ribavirin	1. 671* -5.76772 (RTCS, TCS)	1. 494 -6.90791 (RTCS)
	2. 1103 -5.11031 (RTCS)	2. 952 -6.51335 (RTCS)
	3. 781 -4.9425 (RTCS)	3. 608 -6.38317 (RTCS)
	4. 1951 -4.51257 (RTCS,TCS)	4. 1610 -6.27703 (RTCS)
	5. 3355 -3.65696 (RTCS)	5. 1805 -6.04705 (RTCS)
		6. 16886 -4.96632 (RTCS)
		7. 5064* -4.71549 (RTCS, TCS)
Lopinavir	1. 6774 -6.33815 (RTCS)	1. 21325 -7.555046 (RTCS)
	2. 4511 -3.41857 (RTCS)	2. 23225 -6.75864 (RTCS)
		3. 22950 -5.709135 (RTCS)
		4. 5903 -5.27687 (RTCS)
		5. 23570, -5.261719 (RTCS)
		6. 19551 -4.325415 (RTCS)
		7. 5180 -4.308499 (RTCS)
		8. 19557 -3.264581 (RTCS)
Ritonavir	1. 2652 -3.62628 (RTCS)	1. 5903 -5.87177 (RTCS)
		2. 18305 -3.328152 (RTCS)
Calanolide A	1. 2436 -8.78098 (RTCS)	1. 1421 -7.07468 (RTCS, TCS)
		2. 16078 -6.0513 (RTCS)
		3. 2978 -3.74898 (RTCS)
		4. 1421 -3.54537 (RTCS, TCS)
Navirapine	1. 125* -6.73152 (RTCS, TCS)	1. 4463* -7.00563 (RTCS, TCS)
	2. 6666 -6.3402 (TCS)	2. 6759 -4.80091 (TCS)
	1. 4796 -5.21003 (TCS)	

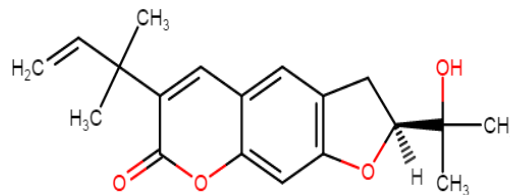
\* หมายเลขโมเลกุลที่มีค่า RTCS และ TCS เท่ากับ 1.0

3. ผลการคำนวณพลังงานยึดจับที่ได้จากวิธีเข้าจับเชิงโมเลกุลจากตารางที่ 4 พบว่า โดยใช้เกณฑ์ RTCS จะให้จำนวนโมเลกุลที่มีความสามารถในการยึดจับโปรตีนเป้าหมายมีมากขึ้น และหลายหมายเลขมีค่าพลังงานยึดจับที่

ดีกว่าตัวยาที่อ้างอิง แต่จำนวนโมเลกุลที่ได้ก็ไม่มากนักเกินไปในการที่จะนำไปคำนวณ วิเคราะห์อื่นๆ โดยเฉพาะเมื่อต้องการความแม่นยำของโครงสร้างและพลังงานยึดจับที่สูงขึ้น โดยโมเลกุลหมายเลข 2436 ในฐานข้อมูล DrugBank และหมายเลข 21325 ในฐานข้อมูล PubChem ซึ่งมีพลังงานยึดจับเป็น -8.78 และ -7.56 kcal/mol ตามลำดับ แสดงโครงสร้างได้ดังรูปที่ 13

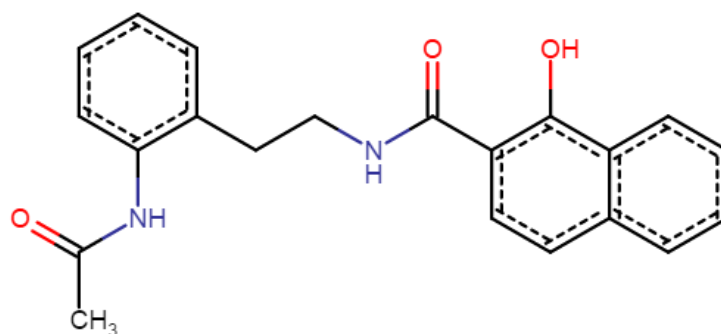
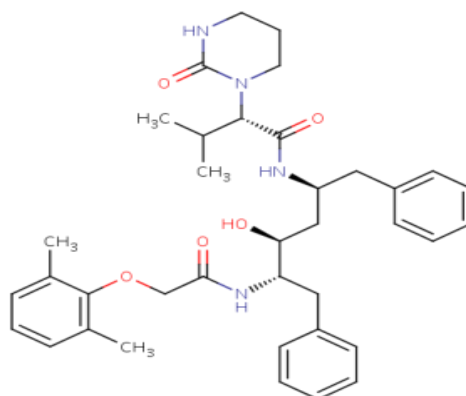


Calanolide A



2436

โดยอ้างอิงกับ Calanolide A มีค่า RTCS = 0.628, TCS = 0.419



Lopinavir

21325

โดยอ้างอิงกับ Lopinavir มีค่า RTCS = 0.617, TCS = 0.260

รูปที่ 13 เปรียบเทียบค่า RTCS และ TCS ของโมเลกุลหมายเลข 2436 ในฐานข้อมูล DrugBank หมายเลข 21325 ในฐานข้อมูล PubChem ซึ่งมีพลังงานยึดจับเป็น -8.78 และ -7.56 kcal/mol ตามลำดับ

## สรุปผล

ในการศึกษานี้ ได้เสนอวิธีการคำนวณค่าความคล้าย RTCS ซึ่งเป็นการปรับปรุงค่า TCS แบบเดิมที่นิยมใช้กันอยู่แล้ว เพื่อนำไปใช้เป็นเกณฑ์สำหรับการคัดกรองด้วยวิธี การเข้าจับเชิงโมเลกุล โดยเลือกด้วยยา 5 ตัว สำหรับใช้ในการรักษาหรือทดลองกับโรค SARS ได้แก่ Ribavirin Lopinavir Ritonavir Calanolide A และ Navirapine โดยมีโครงสร้างโปรตีน 3C-LIKE Proteinase เป็นเป้าหมาย ผลการคำนวณแสดงให้เห็นชัดเจนว่า โดยเฉลี่ยแล้วค่า RTCS จะสูงกว่า TCS โดยเฉพาะบริเวณที่ RTCS และ TCS มีค่าสูงกว่า 0.5 เป็นผลทำให้เมื่อนำไปใช้เป็นเกณฑ์ในการคัดกรองโมเลกุลจากฐานข้อมูล การใช้ RTCS จะให้จำนวนโมเลกุลที่ผ่านเกณฑ์มากกว่า ดังที่เห็นได้ในกรณีของ Ribavirin Lopinavir Ritonavir และ Calanolide A ในขณะที่ผลของ Navirapine ที่ต่างจากโมเลกุลอื่นๆ ก็เป็นปัญหาที่น่าสนใจศึกษาในรายละเอียดเพิ่มขึ้นต่อไป นอกจากนี้ยังพบว่า โมเลกุลหมายเลข 2436 ในฐานข้อมูล DrugBank และโมเลกุลหมายเลข 21325 ในฐานข้อมูล PubChem ที่ใช้ศึกษานี้ น่าจะเป็นโมเลกุลที่มีศักยภาพเชิงยา ซึ่งจะนำไปศึกษาวิเคราะห์ระดับสูงต่อไป

## ปัญหาและอุปสรรค

อนึ่งอีกวัตถุประสงค์ของงานวิจัยนี้คือการสร้างระบบสารสนเทศ เพื่อเป็นแหล่งบริการด้านการคัดกรอง และการคำนวณเชิงโมเลกุลต่างๆ แต่ด้วยเวลาที่มีจำกัด และมีการใช้เวลากับการพัฒนาขั้นตอนวิธีในการระบุชนิด และนับจำนวนที่ปรากฏของส่วนโมเลกุลย่อยนานเกินกว่าที่คาดไว้ ทำให้การพัฒนาาระบบสารสนเทศดังกล่าวยังไม่แล้วเสร็จ แต่คณะผู้วิจัยยังคงดำเนินการเพื่อให้แล้วเสร็จโดยเร็ว และนำผลที่ได้ตีพิมพ์ในวารสารระดับนานาชาติ ที่มี Impact factor (IF) ต่อไป

## บรรณานุกรม

1. Baker M., *Open-access chemistry databases evolving slowly but not surely*, Nat. Rev. Drug Discov, **5(9)** (2006) 707-8.
2. Fischer I. and Meinl T., *Graph Based Molecular Data Mining - An Overview* IEEE International Conference on Systems, Man and Cybernetics, **5** (2004), 4578-4582.
3. Weininger D., *SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules*, Journal of Chemical Information and Modeling, **28** (1988), 31-36. doi:10.1021/ci00057a005.
4. Weininger D., Weininger A., Weininger J.L., *SMILES. 2. Algorithm for generation of unique SMILES notation*, Journal of Chemical Information and Modeling, **29** (1989), 97-101. doi:10.1021/ci00062a008.
5. Weininger D., *SMILES. 3. DEPICT. Graphical depiction of chemical structures*, Journal of Chemical Information and Modeling, **30(3)** (1990) 237-243. doi:10.1021/ci00067a005
6. Lee V.S., Wittayanarakul K., Remsungnen T., Parasuk V., Sompornpisut P., Chantratita W., *Structure and dynamics of SARS coronavirus proteinase: The primary key to the designing and screening for anti-SARS drugs*, Science Asia, **29** (2003), 181-188.
7. Takigawa I. and Mamitsuka H., I. Graph mining: procedure, application to drug discovery and recent advances, Drug Discov Today (2012).  
<http://dx.doi.org/10.1016/j.drudis.2012.07.016>
8. Williams D.W., Huan J., Wang W., Graph Database Indexing Using Structured Graph Decomposition IEEE, 2007.
9. Garey M.R., and Johnson D.S., Computers and Intractability, W. H. Freeman and Company, New York, 1979.
10. Willett P., *Chemical similarity searching*, Journal of Chemical Informatics and Computational Science, **38** (1998), 983-996.
11. Bender A. and Glen R.C., *Molecular similarity: a key technique in molecular informatics*, Organics and Biomolecular Chemistry, **2** (2004), 3204-3218.
12. Wang T. and Zhou J., *3DFS: A New 3D Flexible Searching System for Use in Drug Design*, Journal of Chemical Informatics and Computational Science, **38** (1998), 71-77.
13. Wishart D. S., et al. *DrugBank: a comprehensive resource for in silico drug discovery and exploration*, Nucleic acids research 34.suppl, **1** (2006): D668-72.
14. Wang Y., et al., *An overview of the PubChem BioAssay resource*, Nucleic acids research 38.suppl, **1** (2010): D255-66.
15. <http://emedicine.medscape.com/article/237755-treatment#aw2aab6b6b2>
16. Bacha U. M.; Development of inhibitors against the SARS coronaviral protease 3CLpro, Dissertation; <http://udini.proquest.com/view/development-of-inhibitors-against-goid:304615429/>.
17. OLBoyle N. M., et al., *Open Babel: An open chemical toolbox*, Journal of Cheminformatics, **3** (2011): 33.

18. Giugno R., and Dennis S., *Graphgrep: A fast and universal method for querying graphs*, Pattern Recognition, 2002. Proceedings. 16th International Conference on. Vol. 2. IEEE, 2002.
19. Thompson, Mark A. ArgusLab 4.0. 1; Planaria Software LLC: Seattle, WA, 2004.
20. Joy S., et al., *Detailed comparison of the protein-ligand docking efficiencies of GOLD, a commercial package and ArgusLab, a licensable freeware*, In silico biology, 6(6) (2006), 601-5.
21. Oda A., et al., *Evaluation of docking accuracy and investigations of roles of parameters and each term in scoring functions for protein-ligand docking using ArgusLab software*, Bulletin of the Chemical Society of Japan, **80** (2007), No. 10, 1920-1925.

# ภาคผนวก

**ประวัติย่อ**  
**นางสาวสมนึก วรวิเศษ**  
**หัวหน้าโครงการวิจัย**

1. ชื่อ - นามสกุล (ภาษาไทย) สมนึก วรวิเศษ  
ชื่อ - นามสกุล (ภาษาอังกฤษ) Somnuek Worawiset
2. เลขหมายบัตรประจำตัวประชาชน 3 4503 00404 99 0
3. ตำแหน่งปัจจุบัน อาจารย์
4. e-mail: [wsomnu@kku.ac.th](mailto:wsomnu@kku.ac.th)
5. ประวัติการศึกษา

Dr.rer.nat (Mathematics)	Carl von Ossietzky University of Oldenburg	2554
M.Sc. (Mathematics)	Khon Kaen University	2547
B.Sc. (Mathematics)	Khon Kaen University	2543
6. สาขาวิชาการที่มีความชำนาญพิเศษ (แตกต่างจากวุฒิการศึกษา) ระบุสาขาวิชาการ  
ทฤษฎีกราฟ
7. ประสบการณ์งานวิจัย
  1. Worawiset, S., *On endomorphisms of Clifford semigroups, Semigroups, Acts and Categories with Applications to Graphs*, Edited by V. Laan, S. Bulman-Fleming and R. Kaschek, Proc. Tartu (2007), pp. 143-150.
  2. S. Worawiset and T. Remsungnen, *Molecular similarity searching using meaning-factor reduced graph*, Southeast-Asian Journal of Sciences, Vol. 2, No. 2 (2013), pp. 160-164
  3. S. Chotchaisthait and S. Worawiset, *On the Diophantine equation  $143^x + 143^{2s} n^y = z^2$* , submitted.
  4. Worawiset, S. and Koppitz, J., *Largest coregular subsemigroups of the full transformation Monoid*, submitted.

## ประวัติย่อคณะผู้ร่วมวิจัย

### นายเทวัญ เร่มสูงเนิน

1. ชื่อ - นามสกุล (ภาษาไทย) เทวัญ เร่มสูงเนิน
- ชื่อ - นามสกุล (ภาษาอังกฤษ) Tawun Remsungnen
2. เลขหมายบัตรประจำตัวประชาชน 3 3018 00255 42 1
3. ตำแหน่งปัจจุบัน ผู้ช่วยศาสตราจารย์
4. e-mail: rtawun@kku.ac.th
5. ประวัติการศึกษา

Dr.rer.nat (Computational Chemistry) Innsbruck University 2545

M.Sc. (Computational Science) Chulalongkorn University 2542

B.Sc. (Chemistry) Khon Kaen University 2540

6. สาขาวิชาการที่มีความชำนาญพิเศษ (แตกต่างจากวุฒิการศึกษา) ระบุสาขาวิชาการ  
ตัวแบบเชิงคณิตศาสตร์ ชีวสารสนเทศ

#### 7. ประสบการณ์งานวิจัย

1. Remsungnen T., Rode B.M., *Dynamical properties of the water molecules in the hydration shells of Fe(II) and Fe(III) ions: ab initio QM/MM molecular dynamics simulations*, Chemical Physics Letter, **367** (2003) 586.
2. Remsungnen T., Rode B.M., *QM/MM Molecular Dynamics Simulation of the Structure of Hydrated Fe(II) and Fe(III) Ions*, J. Phys. Chem. A, **107** (2003) 2324.
3. Hannongbua S., Kiselev M., Remsungnen T. and Heinzinger K., *The Structure of Concentrated Li-Ammonia Solutions as Derived from MD Simulations*, Condensed Matter Physics, **6** (2003) 459.
4. Lee V.S., Wittayanarakul K., Remsungnen T., Parasuk V., Sompornpisut P., Chantratita W., Sangma J., Vannarat S., Srichaikul P., Hannongbua S., Saparpakorn P., Treesuwan W., Aruksakulwong O., Pasomsab E., Promsri S., Chuakheaw D., Hannongbua S., *Structure and Dynamics of SARS Coronavirus Proteinase: The Primary Key to the Designing and Screening for Anti-SARS Drugs*, ScienceAsia, **29** (2003) 181.
5. Remsungnen T., Rode B.M., *Molecular dynamics simulation of the hydration of transition metal ions: the role of non-additive effects in the hydration shells of Fe<sup>2+</sup> and Fe<sup>3+</sup> ions*, Chemical Physics Letter, **385** (2003) 491.
6. Siraleartmukul K., Siriwong K., Remsungnen T., Muangsin N., Udomkichdecha W., Hannongbua S., *Solvation structure of glucosamine in aqueous solution as studied by Monte Carlo simulation using ab initio fitted potential*, Chemical Physics Letter, **395** (2004) 233.

7. Saengsawang O., **Remsungnen T.**, Fritzsche S., Haberlandt R., Hannongbua S., *Structure and Energetics of Water-Silanol Binding on the Surface of Silicalite-1: Quantum Chemical Calculations*, J. Phys. Chem. B., **109** (2005) 5684.
8. Saengsawang O., **Remsungnen T.**, Loisuangsin A., Fritzsche S., Haberlandt R., Hannongbua S., *Energy barrier of water and methane molecules due to the silanol groups on the (010) surface of silicalite-1 as studied by quantum chemical calculations*, Stud. Surf. Sci. Cat., **158** (2005) 947.
9. **Remsungnen T.**, Kormilets V., Loisuangsin A., Schuering A. Fritzsche S., Haberlandt R., Hannongbua S., *Optimal Binding Site of a Methane Molecule on the Silanol Covered (010) Surface of Silicalite-1: ONIOM Calculations*, J. Phys. Chem. B., **110** (2006) 11932.
10. Saengsawang O., Schuring A., **Remsungnen T.**, Loisuangsin A., Hannongbua S., Magusin P.C.M.M., Fritzsche S., *Rotational Motion of Pentane in the Flat gamma-Cages of Zeolite KFI*, J. Phys. Chem. C., **112** (2008) 5922.
11. Youngme S., Wannarit N., **Remsungnen T.**, Chaichit N., Albada G.A., Reedijk j., *Hexa- and polynuclear copper(II) coordination compounds based on the carbonato bridge: Synthesis, crystal structures and magnetic properties*, Inorg. Chem. Comm., **11** (2008) 179.
12. Youngme S., Wannarit N., **Remsungnen T.**, Pakawatchai C., Chaichit N., Engkagul C., Albada G.A., Reedijk j., *A novel octanuclear copper(II) compound: Synthesis, crystal structure and magnetic properties*, Inorg. Chem. Comm., **11** (2008) 427.
13. Pianwanit A., Kritayakornupong C., Vongachariya A., Selphusit A., Ploymeerusmee T., **Remsungnen T.**, Nuntasri D., Fritzsche S., Hannongbua S., *The optimal binding site of CH<sub>4</sub> and CO<sub>2</sub> molecules on the metal-organic framework MOF-5: ONIOM calculations*, Chemical Physics, **349** (2008), 77.
14. Seehamart K., Nanok T., Krishna R., van Baten J.M., **Remsungnen T.**, Fritzsche S., *A Molecular Dynamics investigation of the influence of framework flexibility on self-diffusivity of ethane in Zn(tbip) frameworks*, Microporous and Mesoporous Materials, **125** (2009) 97.
15. Thompho S., Chanajaree R., **Remsungnen T.**, Hannongbua S., Bopp P.A., Fritzsche S., *The Permeation of Methane Molecules through Silicalite-1 Surfaces*, J. Phys. Chem. A, **113** (2009) 2004.
16. Saengsawang O., Schuering A., **Remsungnen T.**, Hannongbua S., Newsome D.A, Dammers A.J., Coppens M.O., Fritzsche S., *Diffusion of n-*

- Pentane in the Zeolite ZK5 Studied by High-Temperature Configuration-Space Exploration*, Chem. Phys., **368** (2010) 121.
17. Wetweerapong J., **Remsungnen T.**, *Solving the Rectangular Packing Problem by a New Bottom-Left Placement Method Incorporated with a Multi-Start Local Search*, FCDS, **35** (2010) 127.
  18. Hertäg L., Buxb H., Caro J., Chmelik C., **Remsungnen T.**, Knauth M., Fritzsche S., *Diffusion of CH<sub>4</sub> and H<sub>2</sub> in ZIF-8*, Journal of Membrane Science, **377** (2011) 36.
  19. Arsawang U., Saengsawang O., Rungrotmongkol T., Sornmee P., Wittayanarakul K., **Remsungnen T.**, Hannongbua S., *How do carbon nanotubes serve as carriers for gemcitabine transport in a drug delivery system?*, JMGM, **29** (2011) 591.
  20. Chatuphonprasert W., **Remsungnen T.**, Nemoto N., Jarukamjorn K., *Different AhR binding sites of diterpenoid ligands from Andrographis paniculata caused differential CYP1A1 induction in primary culture in mouse hepatocytes*, Toxicology in Virto, **25** (2011) 1757-1763.
  21. Puphasuk P., **Remsungnen T.**, *Refinement of Molecular Interaction Parameters of AMBER Force Field for CO<sub>2</sub> and 2-Methylimidazolate Complexs*, Journal of Computational and Theoretical Nanoscience., **9** (2012) 889.
  22. Wongsinlatam W., Keawkri S., **Remsungnen T.**, *Parameter Estimation of Intermolecular Interaction for CO<sub>2</sub> and Benzimidazolate [C<sub>7</sub>H<sub>5</sub>N<sub>2</sub>]*, Applied Mathematical Sciences, **6** (2012) 4261.
  23. Saengsawang O., Vchirawongkwin V., **Remsungnen T.**, Wiebcke M., Fritzsche S., Hannongbua S. *Rotational flexibility of bridging ligands in paddle-wheel layer-pillar metal-organic frameworks studied by quantum calculations*, Computational and Theoretical Chemistry, 1001 (2012) 33.
  24. Puphasuk P., **Remsungnen T.**, *Structure and Dynamics of CO<sub>2</sub> Molecules in ZIF-8: Molecular Dynamics Simulations using ab initio Fitted Interactions and Generic Force Fields*, Journal of Computational and Theoretical Nanoscience, **10** (2013) 227.
  25. Chokbunpiam T., Chanajaree R., Saengsawang O., Reimann S., Chmelik C., Fritzsche S., Caro J., **Remsungnen T.**, Hannongbua S., *The importance of lattice flexibility for the migration of ethane in ZIF-8: Molecular dynamics simulations*, Microporous and Mesoporous Materials, **174** (2013) 126.

26. Wongsinlatam W., **Remsungnen T.**, *Molecular dynamics simulations of CO<sub>2</sub> molecules in ZIF-11 using refined AMBER force field*, Journal of Chemistry, (2013), Article ID 415027.
27. Chokbunpiam T., Chanajaree R., Remsungnen T., Saengsawang O., Fritzsche S., Chmelik C., Caro J., Janke W., Hannongbua S., *N<sub>2</sub> in ZIF-8: Sorbate induced structural changes and self-diffusion*, Microporous and Mesoporous Materials, 187, (2014) 1.