



台灣大學 Reaxys 逆合成AI試用活動

活動期間：9/11~10/13

問卷調查：10/2

ELSEVIER Life Science Solutions

黃恪涵 Ryan Huang

r.huang@elsevier.com

2023.09



今日目標

- Reaxys 全方位版 – 與台灣學術授權版差在哪裡？
- Reaxys 逆合成AI – 用AI技術協助你更快取得感興趣的目標分子
- Reaxys 獨特之處 – 與市場競品之比較
- 試用指南 – 指引你應該測試的內容
- 試用後資料收集 (10/2)

Reaxys 全方位版 – 結合生物活性數據與逆合成AI引擎



專業的化學資料庫

- 找文獻
- 找化合物
- 找反應式
- 找實驗方法
- 找光譜



針對未發表的有機分子
「預測」全合成路徑

整理化合物的生物活性數據，幫
化合物找應用，幫你的藥物靶點
蛋白質找化合物

全方位

為什麼聚焦於逆合成AI的開發？

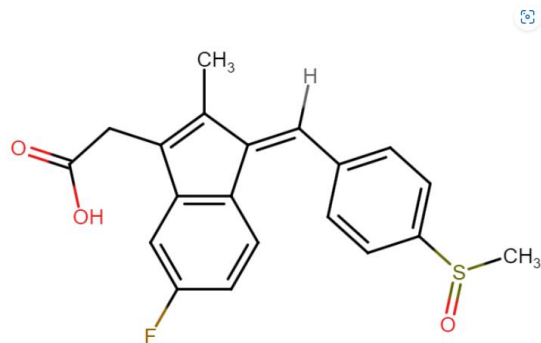
“因為根據統計「合成」仍是研發專案中的瓶頸步驟”

“願意投入有機合成領域的學生越來越少”

“有機分子的應用廣泛”











已知的分子尋找製備方法

示範 published molecule



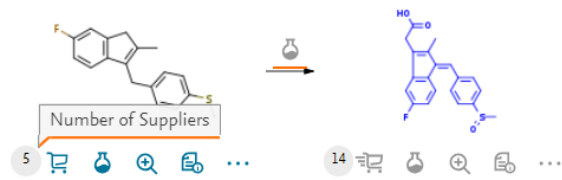
steric

sulindac

	113	Substances	Structure :  as drawn Edit in Query Builder Create Alert
	129	Targets	Structure :  as drawn Edit in Query Builder Create Alert
	3,344	Documents	Structure :  as drawn Edit in Query Builder Create Alert
	117	Substances	Structure :  average similarity; included: taut, solute stereo, additional ring closures allowed, salts, topes, charges, radicals Edit in Query Builder Create Alert
	53	Reactions	Reaction Query :  as drawn Edit in Query Builder Create Alert

已知的分子尋找製備方法

29



1 Conditions Find Similar > Reaction ID: 15546281

Conditions

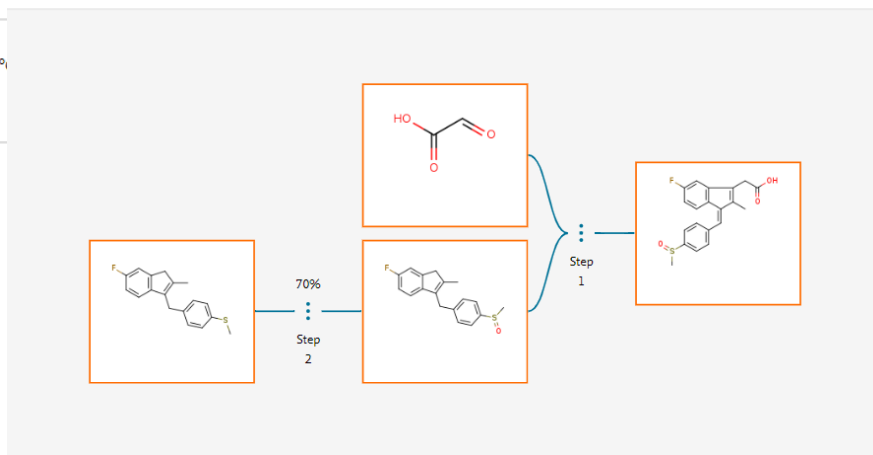
Multi-step reaction with 2 steps

1: 70 percent / m-CPBA / CH₂Cl₂ / 1 h / 0 °C

2: Triton B / methanol / 12 h / 50 °C

[View Scheme >](#)

Multi-step: Reaction ID 15546281



[Transfer to Retrosynthesis >](#)

[Key; O'Connor; Clynes](#)

[4]

[es >](#) [Details >](#) [Abstract >](#)

未發表或創新分子的合成方法

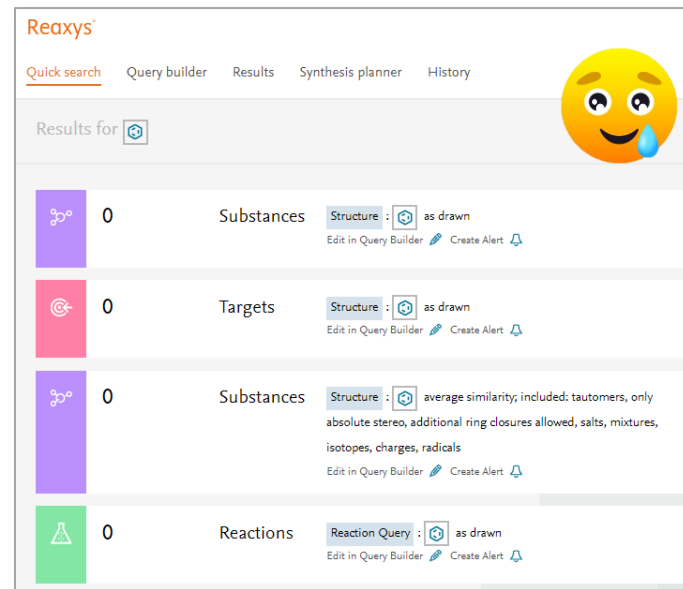
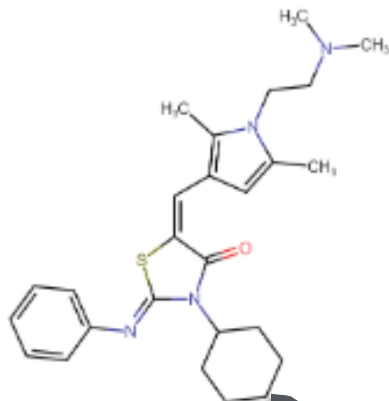
~創新分子不容易找到參考資料~

- 常見的結構斷點?
- 可行性評估?
- 起始材料的價格與時效性?



合成的功力 = 知識 + 時間 + 金錢

藉由AI工具來加速經驗累積的過程，縮短學習曲線。



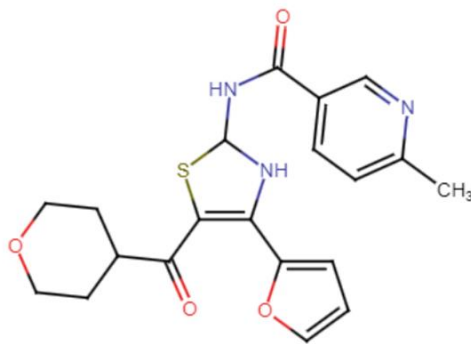
The screenshot shows the Reaxys search interface. At the top, there are navigation tabs: "Quick search", "Query builder", "Results", "Synthesis planner", and "History". A yellow smiley face with a tear is in the top right corner. Below the navigation, it says "Results for" followed by a search icon. The main content area displays search results for four categories: "Substances", "Targets", "Substances", and "Reactions". Each category shows a count of "0" results and options to "Edit in Query Builder" and "Create Alert".

Category	Count	Structure / Query	Options
Substances	0	Structure : as drawn	Edit in Query Builder, Create Alert
Targets	0	Structure : as drawn	Edit in Query Builder, Create Alert
Substances	0	Structure : average similarity; included: tautomers, only absolute stereo, additional ring closures allowed, salts, mixtures, isotopes, charges, radicals	Edit in Query Builder, Create Alert
Reactions	0	Reaction Query : as drawn	Edit in Query Builder, Create Alert

未發表或創新分子的合成方法

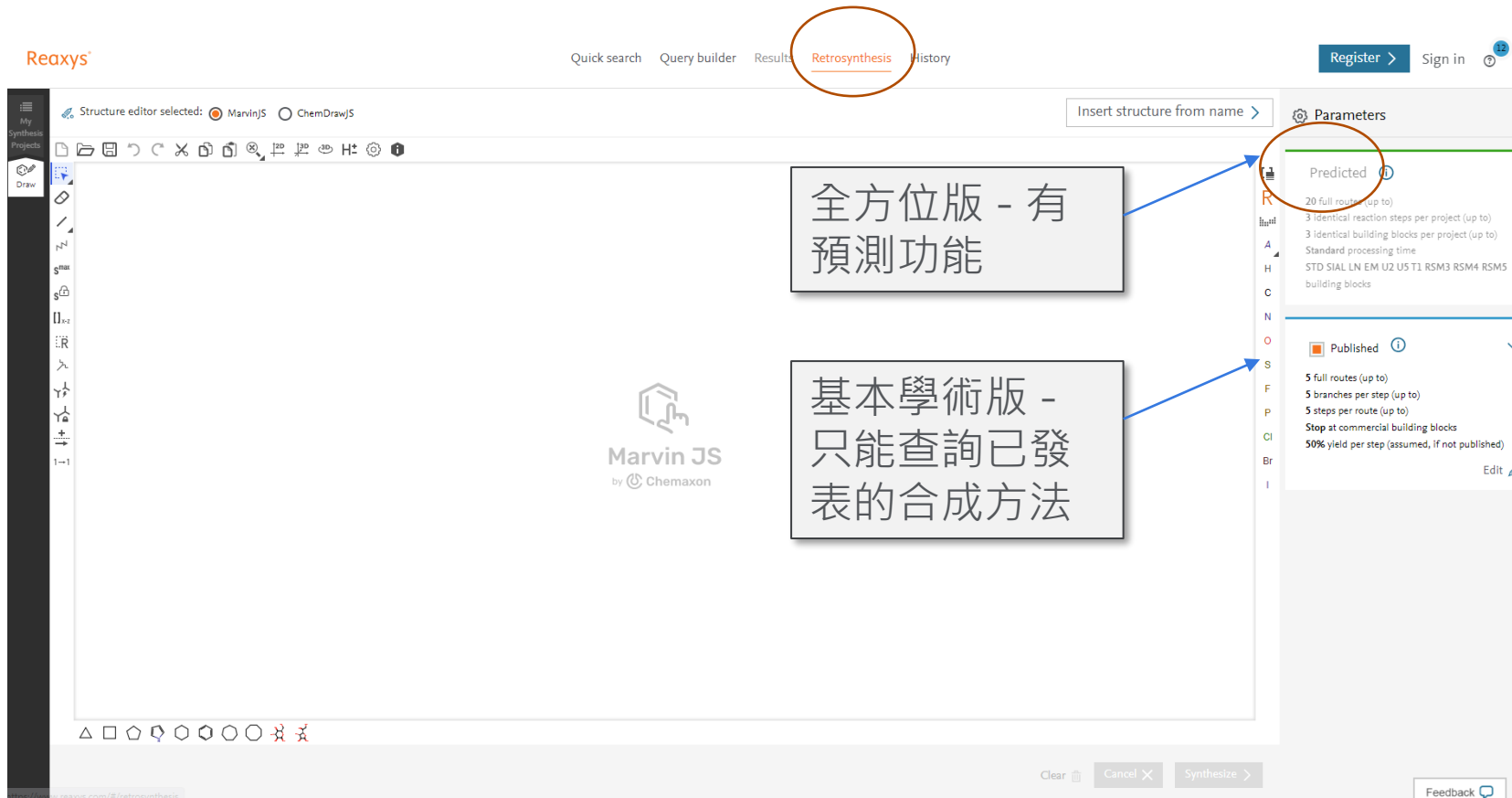
~創新分子不容易找到參考資料~

示範 unpublished molecules



Novel Adenosine A2A receptor antagonist
By Kyowa Kirin Co., Ltd.

Reaxys AI逆合成引擎整合於Retrosynthesis工具



The screenshot shows the Reaxys Retrosynthesis interface. At the top, the navigation bar includes 'Quick search', 'Query builder', 'Results', 'Retrosynthesis' (circled in orange), and 'History'. On the right, there are 'Register' and 'Sign in' buttons. The main workspace is titled 'Structure editor selected: MarvinJS ChemDrawJS' and contains the Marvin JS logo. A 'Parameters' panel on the right is also circled in orange and has two blue arrows pointing to it from text boxes. The 'Predicted' section lists search parameters: '20 full routes (up to)', '3 identical reaction steps per project (up to)', '3 identical building blocks per project (up to)', 'Standard processing time', and 'STD SIAL LN EM U2 U5 T1 RSM3 RSM4 RSM5 building blocks'. The 'Published' section lists: '5 full routes (up to)', '5 branches per step (up to)', '5 steps per route (up to)', 'Stop at commercial building blocks', and '50% yield per step (assumed, if not published)'. The bottom of the interface has 'Clear', 'Cancel', and 'Synthesize' buttons, along with a 'Feedback' button.

Retrosynthesis

Parameters

Predicted

- 20 full routes (up to)
- 3 identical reaction steps per project (up to)
- 3 identical building blocks per project (up to)
- Standard processing time
- STD SIAL LN EM U2 U5 T1 RSM3 RSM4 RSM5 building blocks







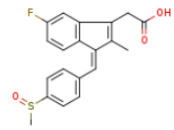







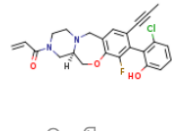







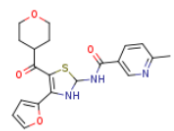






Published

- 5 full routes (up to)
- 5 branches per step (up to)
- 5 steps per route (up to)
- Stop at commercial building blocks
- 50% yield per step (assumed, if not published)

全方位版 - 有預測功能

基本學術版 - 只能查詢已發表的合成方法

Reaxys 逆合成AI不受限發表路徑提供「預測」

0 selected  Delete				
No. 	Date/Time 	Project name 	 Draw new structure	No. of routes
<input type="checkbox"/> 1410201	07 Sep 2023 02:13	Project #1410201 	   ... 	Predicted  Published  
<input type="checkbox"/> 1406540	07 Sep 2023 03:51	Project #1406540 	   ... 	Predicted  Published  
<input type="checkbox"/> 1405529	08 Sep 2023 02:58	Project #1405529 	   ... 	Predicted  Published  

Reaxys 逆合成AI打造給Bench chemists使用

資料科學Data Science流程

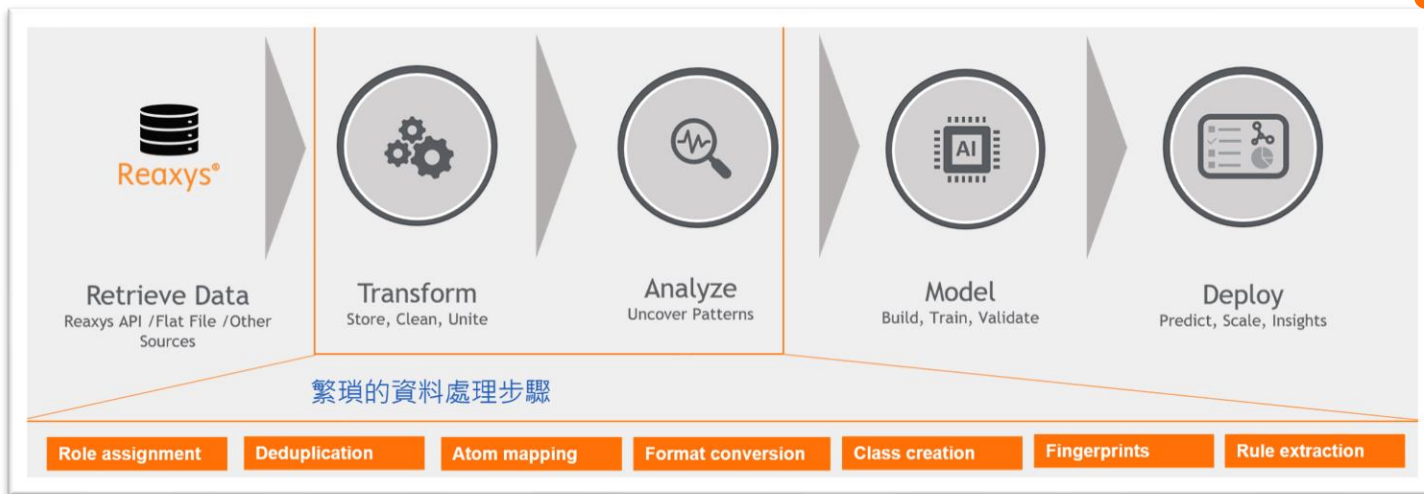
Scientific
Robustness

簡易使用

Reaxys Retrosynthesis Tool



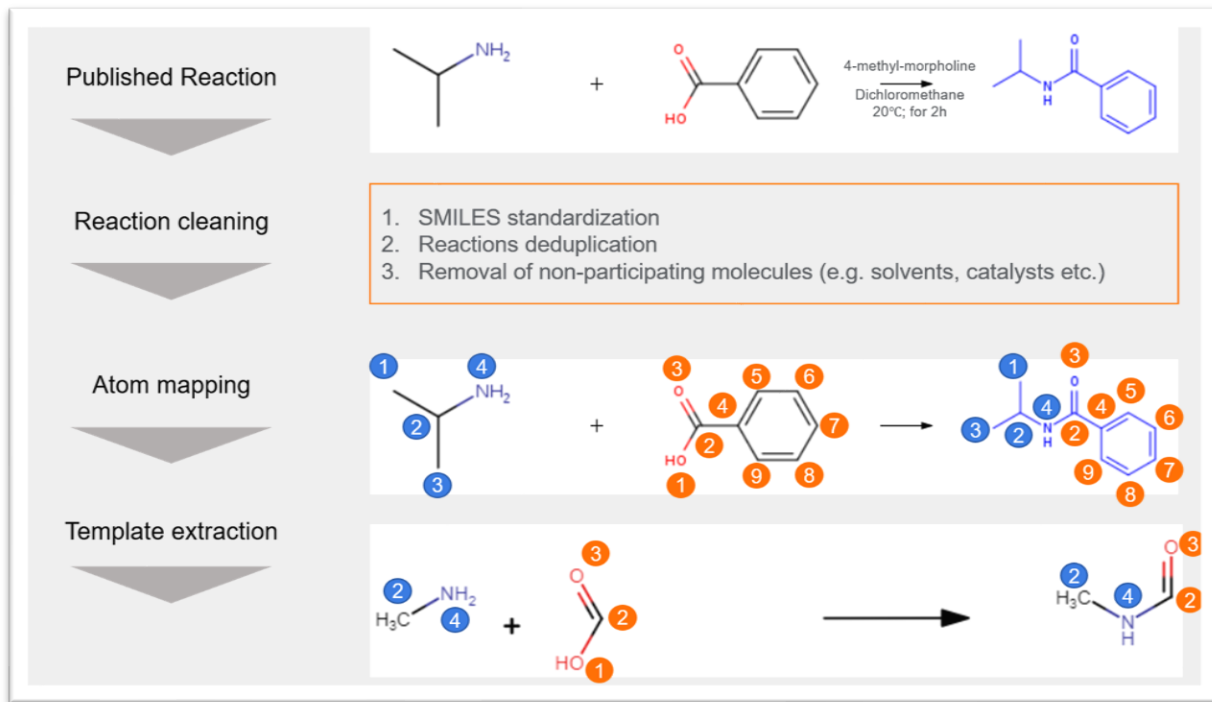
化學家



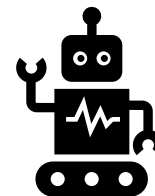
資料科學家



Reaxys-PendingAI 以機器從發表的反應式擷取轉化規則



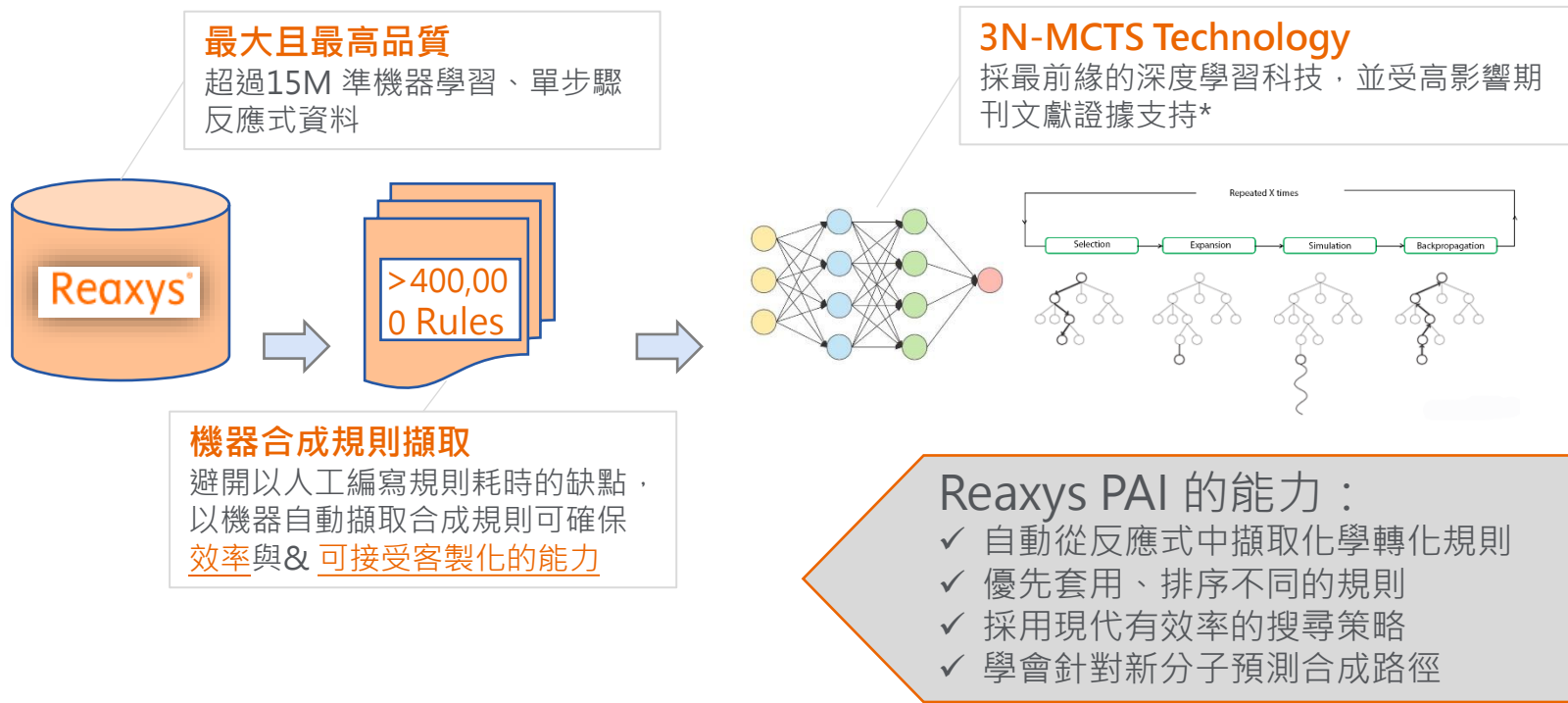
Machine Rule Extraction



- ✓ 新發表的化學研究指數般的成長，機器能跟上發表的**速度**
- ✓ **客製化**的可行性，私人機構累積的大量反應式資料可併入讓機器學習

撥開Reaxys PAI的面紗

結合 最大的Reaxys反應資料庫 與 神經網路技術



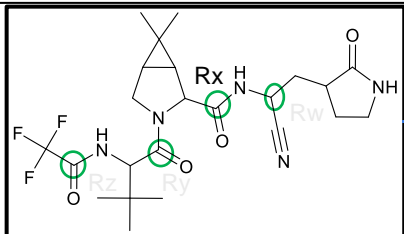
Reaxys PAI 如何推導出新化合物的合成路徑?

Step 1: Identify most probable disconnections

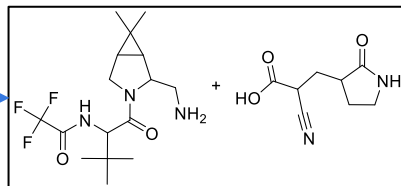
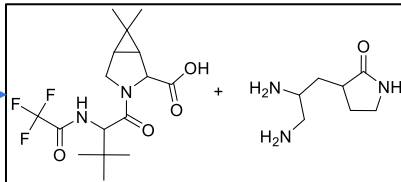
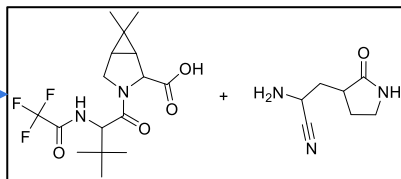
Step 2: Apply retro rule to create starting materials

Retro Rule R_x

找出最有可能的結構斷點



Possible synthons from application of Retro Rule X



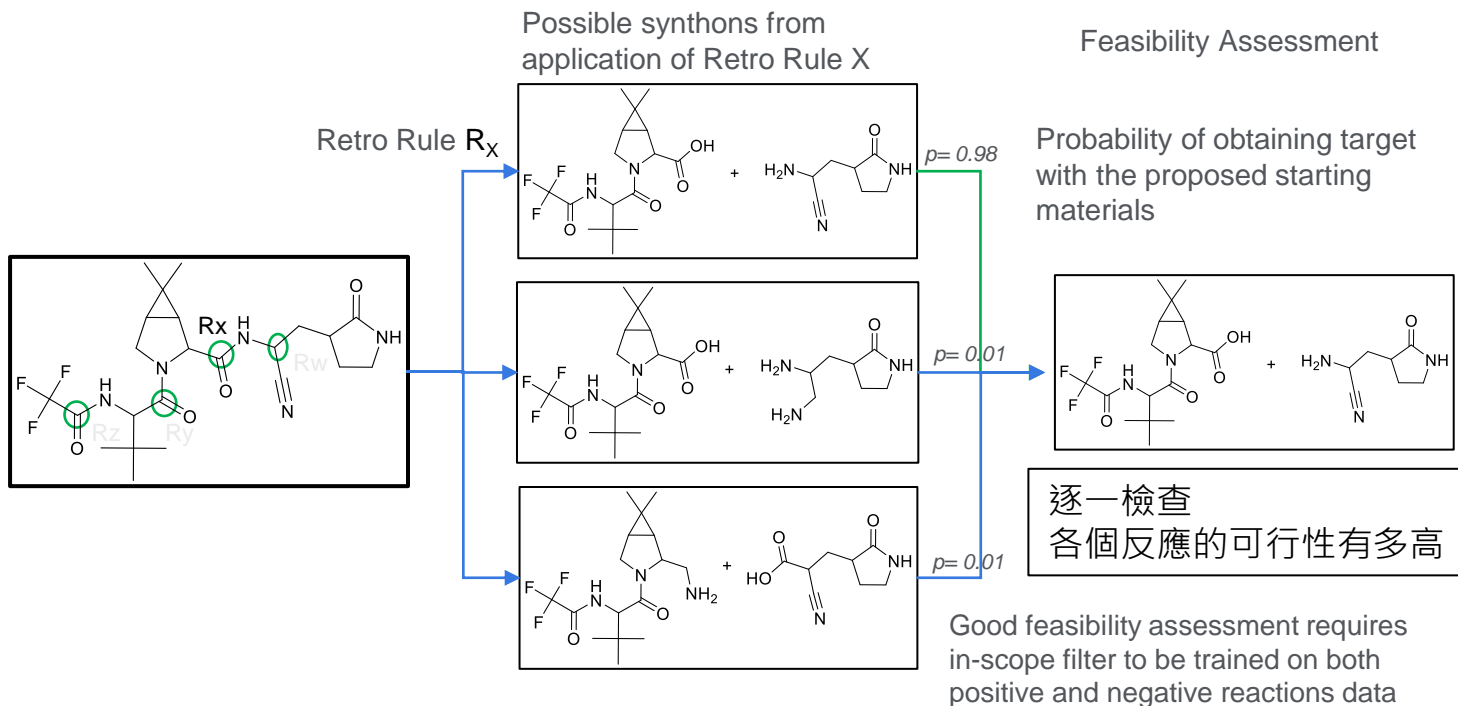
套用rules · 生成可能的組合

Reaxys PAI 如何推導出新化合物的合成路徑?

Step 1: Identify most probable disconnections

Step 2: Apply retro rule to create starting materials

Step 3: For each reaction use in-scope filter



Reaxys PAI 如何推導出新化合物的合成路徑?

Step 1: Identify most probable disconnections

Step 2: Apply retro rule to create starting materials

Step 3: For each reaction use in-scope filter

Step 4: Monte Carlo Tree Search

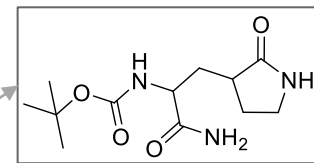
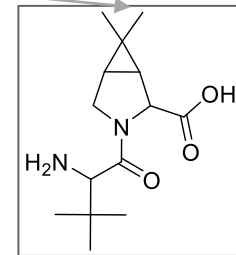
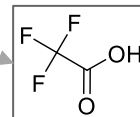
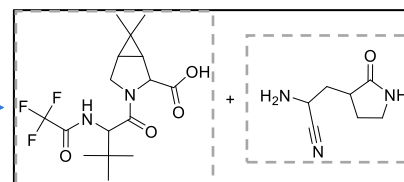
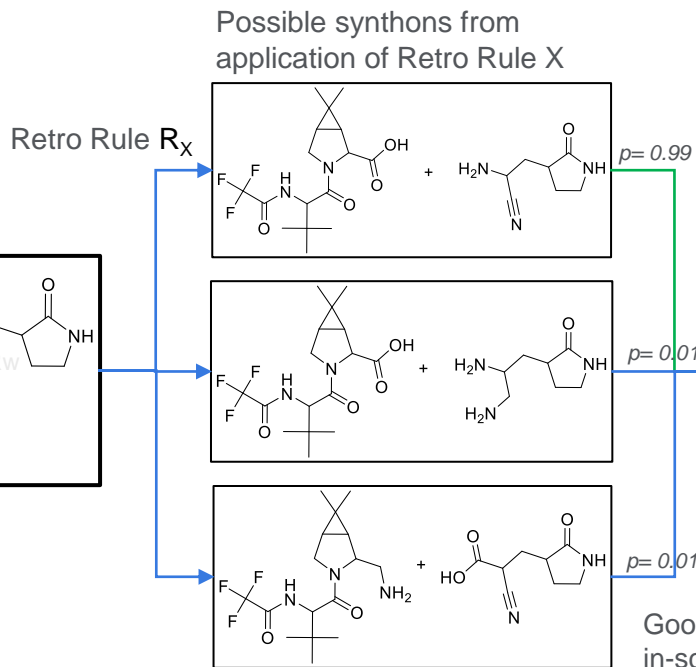
Iterate steps 1, 2 & 3 until commercially available starting materials are identified or **time budget** is reached

時間預算

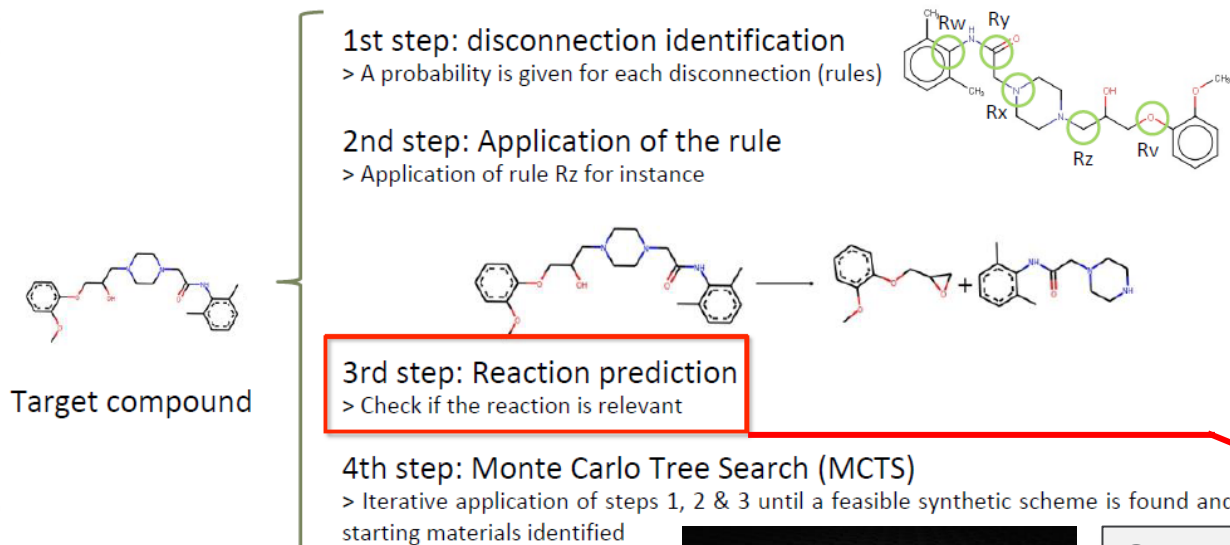
Probability of obtaining target with the proposed starting materials

檢查是否為商用可購買的材料 **Commercial BB**

Good feasibility assessment requires in-scope filter to be trained on both positive and negative reactions data

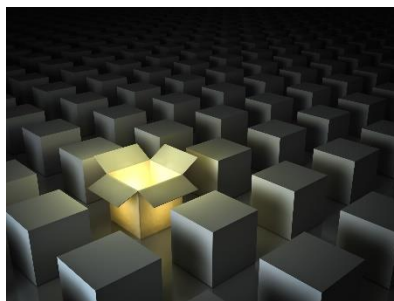


Reaxys PAI 如何給每個預測路徑打分數?



Up to 20 prediction routes
rank by:
Confidence Score
 1(Promising) - 0 (unlikely)

M. H. S. Segler, M. Preuss, M. P. Waller, Nature 555, 604–610 (2018)



Give similar Reaxys examples (Not a black box) **Similarity Score**

- Find conditions
- Validate how evidences are supporting the prediction

Reaxys AI提供合成路徑「預測」



Reaxys

Quick search Query builder Results Retrosynthesis History Alerts

Ryan Huang

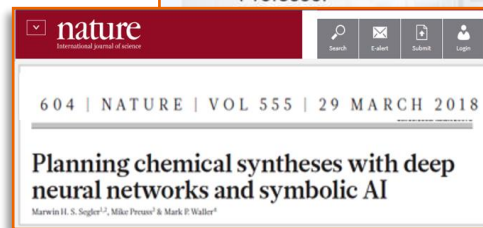
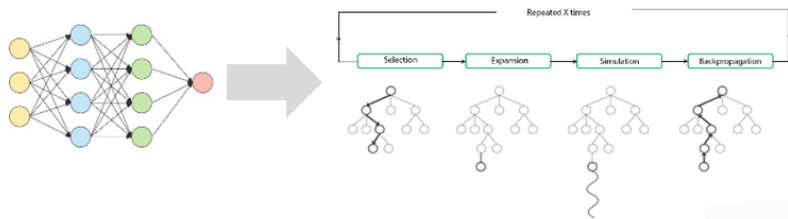
Predicted Route #1		0 selected		Delete	
No.	Date/Time	Project name		Draw new structure	No. of routes
<input type="checkbox"/>	1410201	07 Sep 2023 02:13	Project #1410201 Delete	 + Draw new structure	Predicted 3 Published 5 View >
<input type="checkbox"/>	1406540	07 Sep 2023 03:51	Project #1406540 Delete	 + Draw new structure	Predicted 2 Published 1 View >
<input type="checkbox"/>	1405529	08 Sep 2023 02:58	Project #1405529 Delete	 + Draw new structure	Predicted 2 Published 0 View >

Reaxys的獨特之處①

- >23M高品質的反應式資料作為AI教材 > 15M High quality reaction training data set

Segler and Waller et al. wrote one of **the most cited works in the field of predictive retrosynthesis**, designing a predictive model based on **neural networks** and **Monte-Carlo Tree search**, that:

- ✓ Learns transformation rules from data
- ✓ Learns to prioritize rules
- ✓ Learns to predict reactions
- ✓ Uses modern efficient search methods – 10mins



Reaxys的獨特之處①

- >23M高品質的反應式資料作為AI教材 > 15M High quality reaction training data set

Proprietary sources

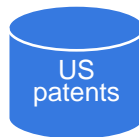
Reaxys®



- > 62 M reactions
- 23 M single step reactions used for ML
- Good quality, ML ready and most often used data for ML

- Typically closed company resource
- Mostly between 1 to 3 M reactions

Public-domain sources



- > 2M reactions
- Open-source extract from US patents
- Questionable quality
- Up to 2016



- > 2M reactions
- Initiative from academia with industry support
- Open data format
- Contains US patents and HTE datasets

Reaxys的獨特之處②

- 科學的穩健性 Scientific robust
- 只有被三篇發表以上再現的反應，會被納入AI教材 (Only reactions replicated in more than three publications will be included in template extraction)

Reaxys Retrosynthesis is tried, tested and valued by top academics

Feedback from Prof. Carreira's Group at ETH Zurich and Prof Rutjes group at Radboud University



Choice of molecules and teams

- 7 teams, 3 to 4 researchers each, at least 1 senior and 1 junior researcher in each team
- **30 molecules selected**
 - Structural diversity
 - Drug like and natural products

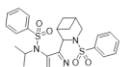
Create synthesis route without predictive retrosynthesis

- In weekly group discussion
- **Assess scientific viability of the route**
 - Time and effort required for optimal route generation

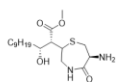
Create synthesis route using predictive retrosynthesis

- In weekly group discussion
- **Compare predicted route to expert generated route**
 - Gather new ideas, measure time and effort required

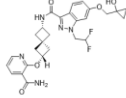
Examples of molecules selected



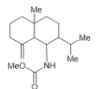
Structure: Novel
Synthesis: Unknown in literature



Structure: Novel
Synthesis: Unknown in literature



Structure: Novel
Synthesis: Unknown in literature



Structure: Known (balichonadin B)
Synthesis: Unknown in literature



Structure: Novel
Synthesis: Unknown in literature



Structure: Novel
Synthesis: Unknown in literature

Feedback from Carreira's Group

1. Interesting approach and **robust route** proposed by retrosynthesis tool
2. The tool is **user friendly and intuitive**. If you are used to Reaxys this tool is easy to use
3. The tool provides **time savings** for designing synthesis routes and getting ideas for conditions that can be used
4. It is **not a black box AI** as it shows literature precedence examples which underpin the predicted routes

ETH

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Radboud Universiteit
SINCE 1923

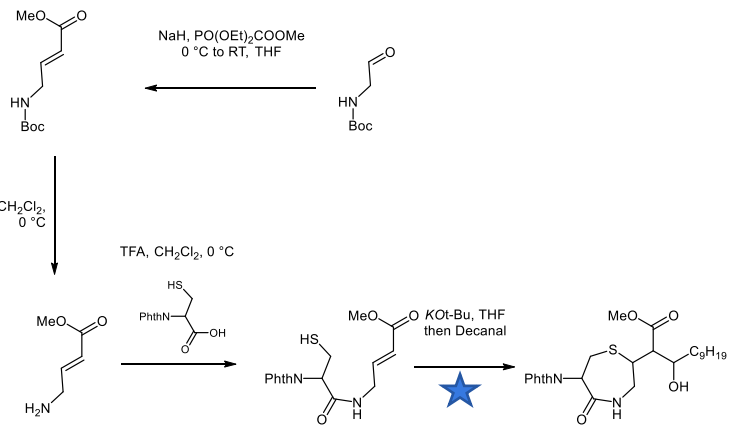


Reaxys retrosynthesis in the hands of chemists from Prof. Carreira's Group at ETH Zurich

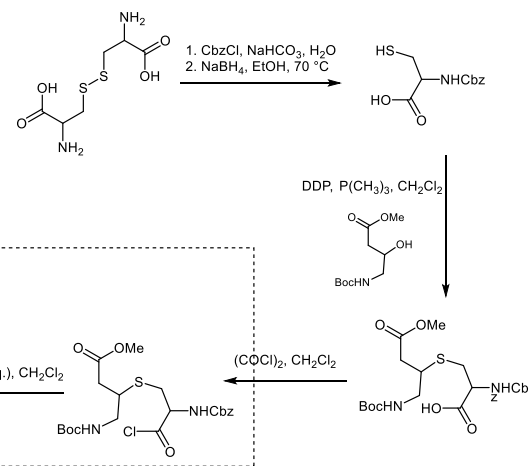
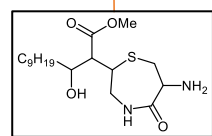
Example of a drug like molecule



Human generated route



Retrosynthesis Tool route



1. The molecule is a fragment from a peptide project in our group and has been synthesised before
2. The molecule doesn't have too much precedence for machine to take inspiration from
3. Robust route proposed by retrosynthesis but chemist needs to review proposals especially with regards to functional group tolerance.

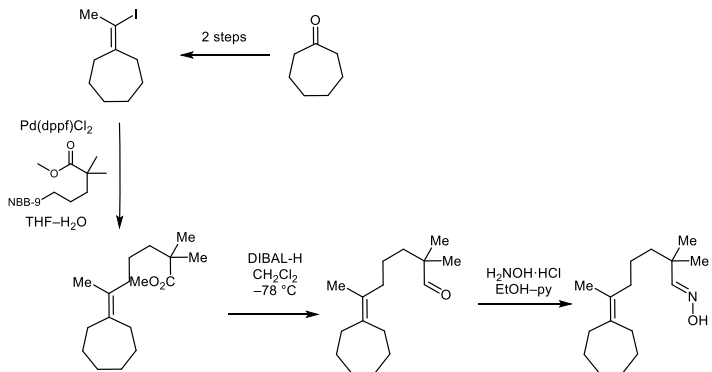
Reaxys retrosynthesis in the hands of chemists from Prof. Carreira's Group at ETH Zurich



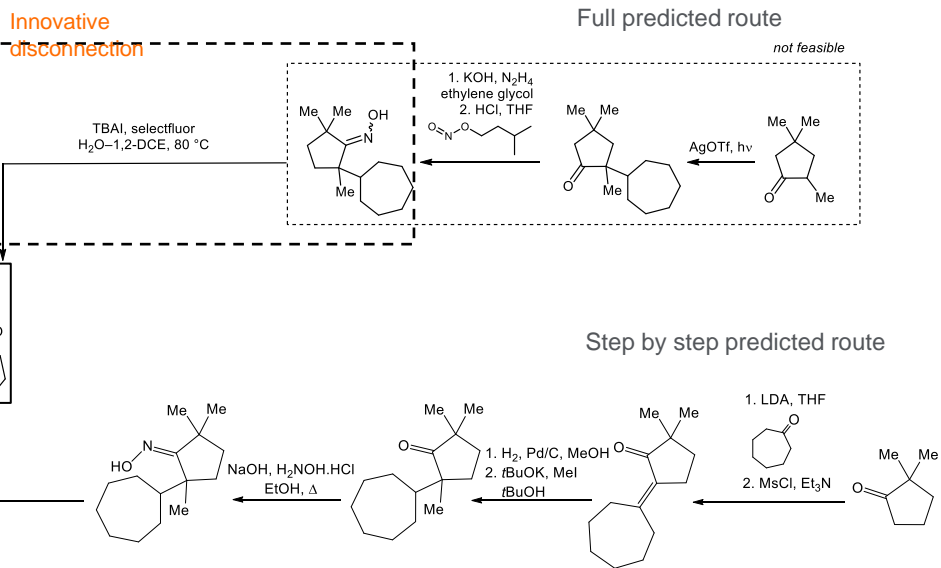
Example of a natural product



Human generated route



Retrosynthesis Tool route

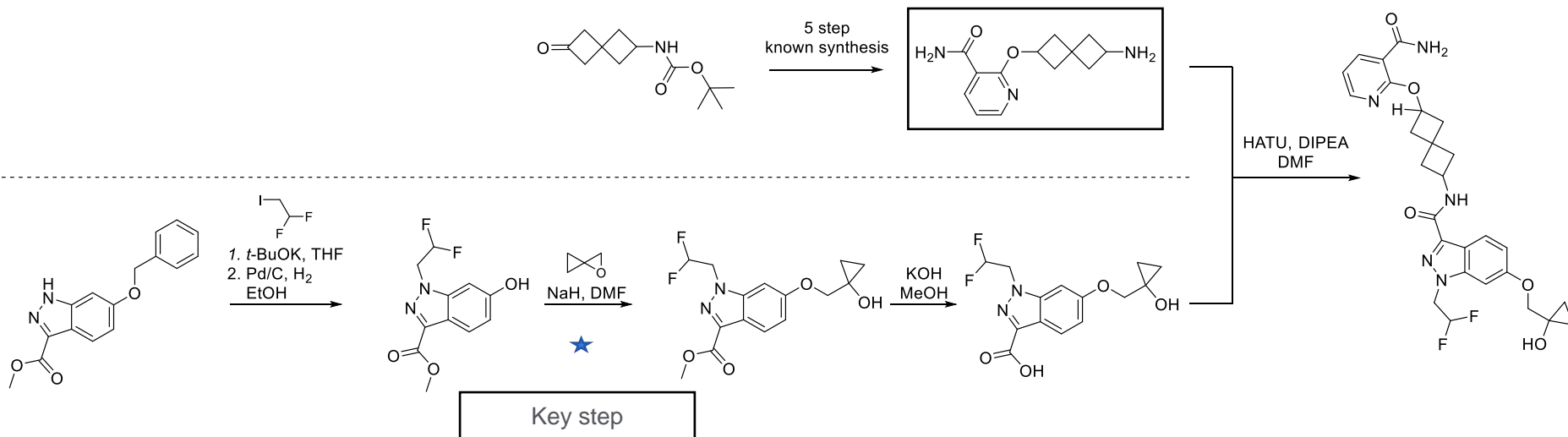


1. Molecule screams "nitrile oxide cyclo-addition"
2. Retrosynthesis route may not be fully feasible but the first disconnection proposed is innovative and something that wouldn't have been obvious.
 - a) Helps at time to remove human bias i.e. explore beyond "nitrile oxide cyclo-addition"
 - b) For molecules of such complexity the ability to custom build route by adding one step at a time is a helpful option

Introducing the cyclopropyl carbinol on the indazole scaffold is a challenge and the predicted route suggests an innovative way to introduce this functionality

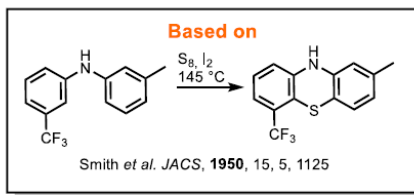
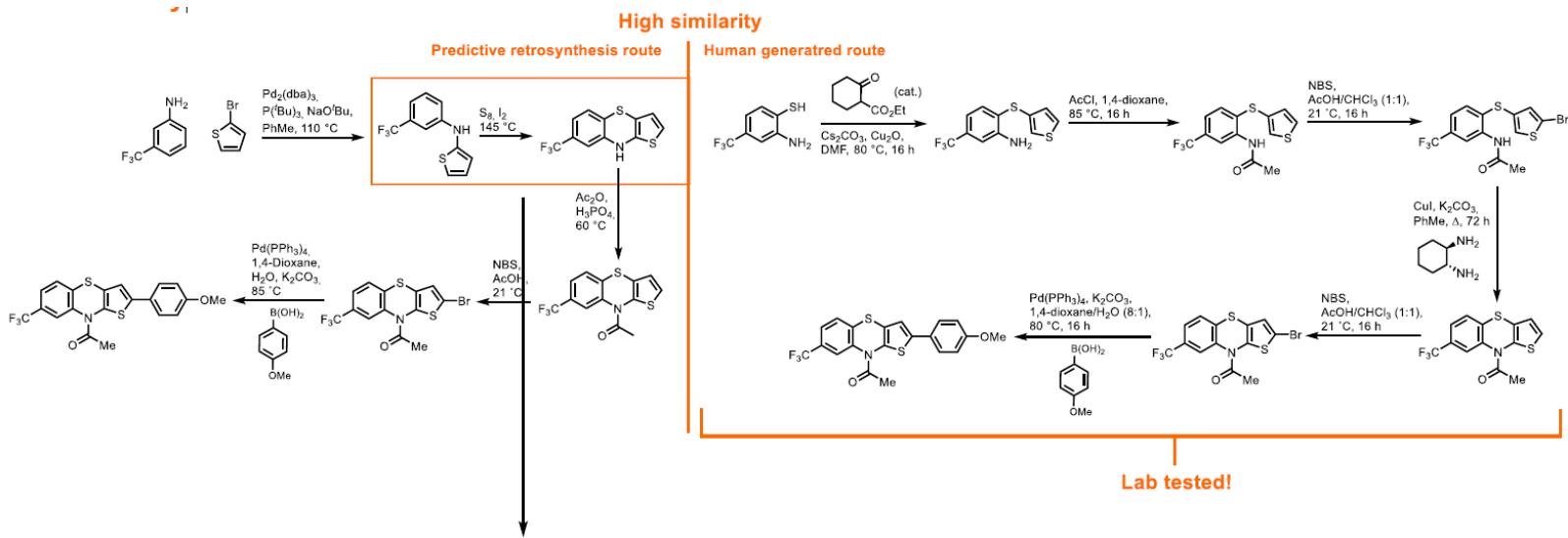


Retrosynthesis Tool route



1. Retrosynthesis gets the key step right – opening of 1-oxaspiro[2.2]pentane to cyclopropyl carbinol. This is an innovative disconnection.
2. Users can extend the synthesis route for selected compounds to go back to more readily available starting materials and combine published steps with predicted steps

Reaxys retrosynthesis in the hands of chemists from the Prof. Rutjes group at Radboud University

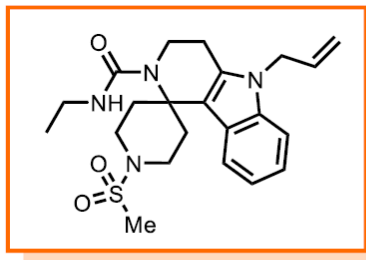


Unique disconnection

- Interesting disconnection that is non-obvious
- Providing alternative for the ring closure



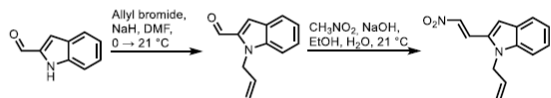
Reaxys retrosynthesis in the hands of chemists from the Prof. Rutjes group at Radboud University



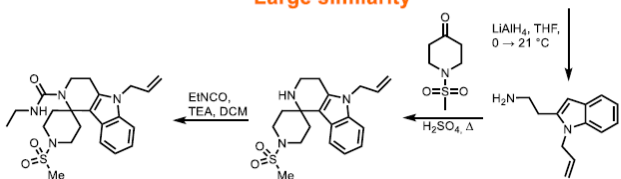
- Large overlap in synthetic output
- More selective
- Starting material commercially available though expensive

Not known in literature

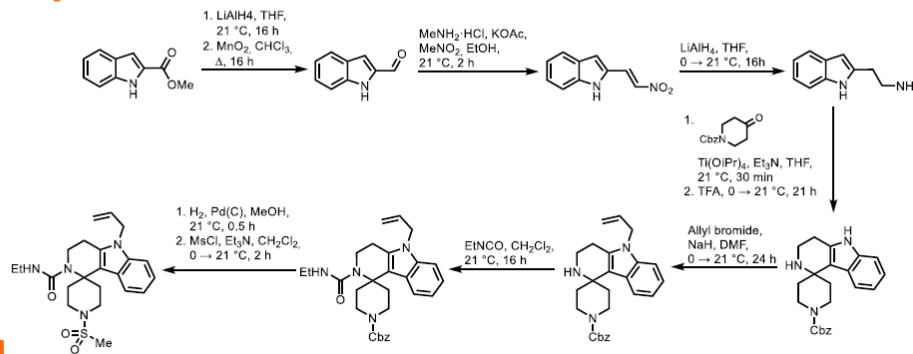
Predictive retrosynthesis route



Large similarity



Human generated route

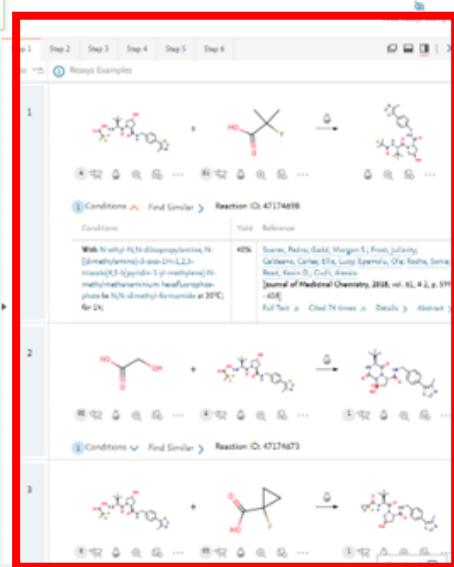
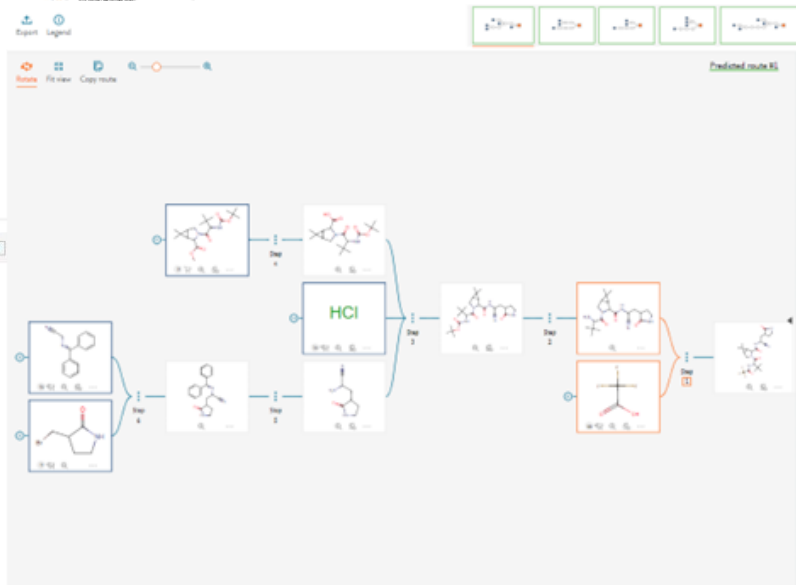
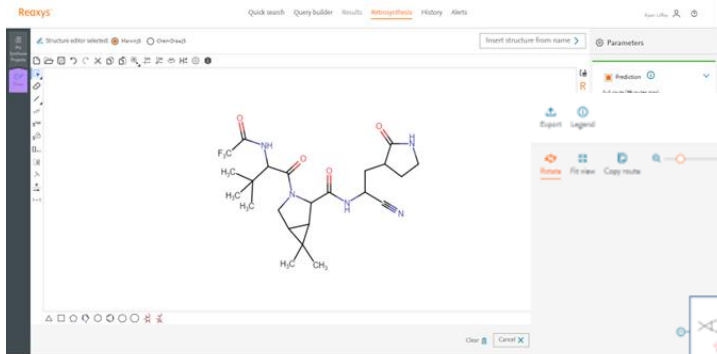


Lab tested!

Reaxys的獨特之處③

- Reaxys AI不是一個黑盒子 Not a black box

試劑、催化劑、文獻
出處一目了然



Reaction Examples window showing three reaction schemes. The first reaction is highlighted with a red box and includes a table with reaction conditions and references.

Conditions	Yield	Reference
With N-ethyl-N-isopropylcarbazole; N-(2-methylpropyl)-2-oxo-2H-1,2,3-benzoxazole (5) (Hydrazide 3-yl-methyl) N-methylmethanamine; benzofuranophene 1a (N-dimethyl formamide at 20°C for 2h)	45%	Saarni, Peltola, Sald, Murgan S., Frost, Juliette, Galimova, Carlos E. A., Lopez Espinosa, Oleg, Nishi, Sohei, Reed, Kevin D., Cullis, Amanda. <i>Journal of Medicinal Chemistry</i> , 2018, vol. 61, # 2, p. 119-126. Full Text » Close 74 times » Details » Abstract »

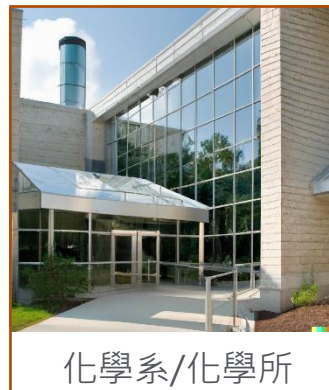
利用Reaxys AI工具來優化學習曲線、節省時間、平衡競爭態勢



- 兩年修業時間極短
- 缺乏有經驗的博士後、實驗室先進進行討論
- 增加學習、實驗的動機



- 專案時間壓力，時間最寶貴
- 整體研發部門的合成能量
- 新進人員的養成需要時間、金錢的累積



- 做研究、發表速度是關鍵因素
- 個別老師研究主題不同，尋求合作研究之不易
- 人力不足，Postdoc之良莠不齊



ELSEVIER

Thank you

歡迎來信留下您的問題與寶貴建議

r.huang@elsevier.com

