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

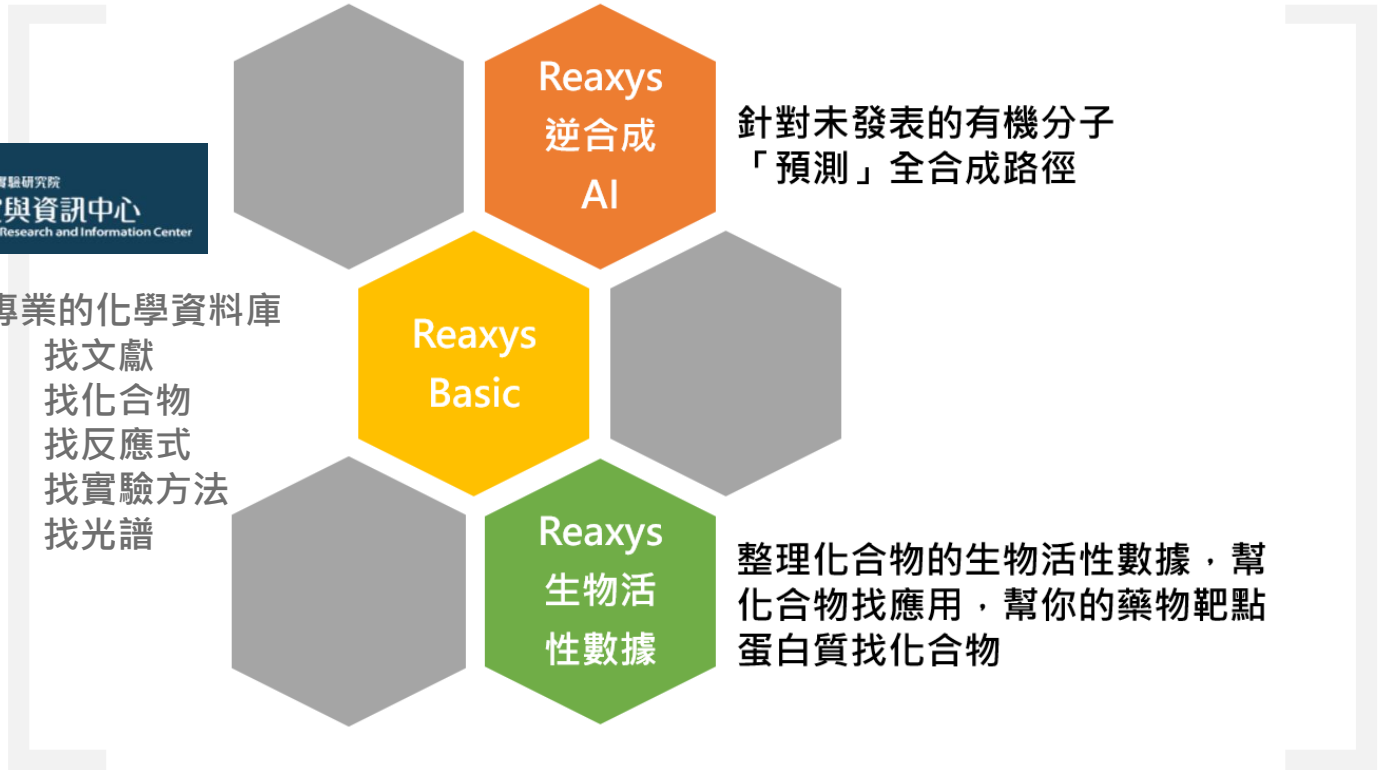
ELSEVIER Life Science Solutions
黃恪涵 Ryan Huang
r.huang@elsevier.com

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



NARLabs 財團法人國家實驗研究院
STPI 科技政策研究與資訊中心
Science & Technology Policy Research and Information Center

為什麼聚焦於逆合成AI的開發?

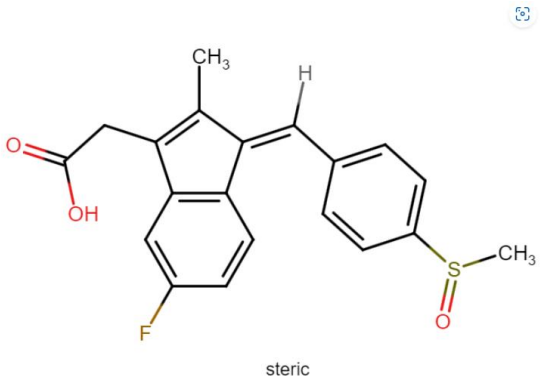
“因為根據統計「合成」仍是研發專案中的瓶頸步驟”

“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

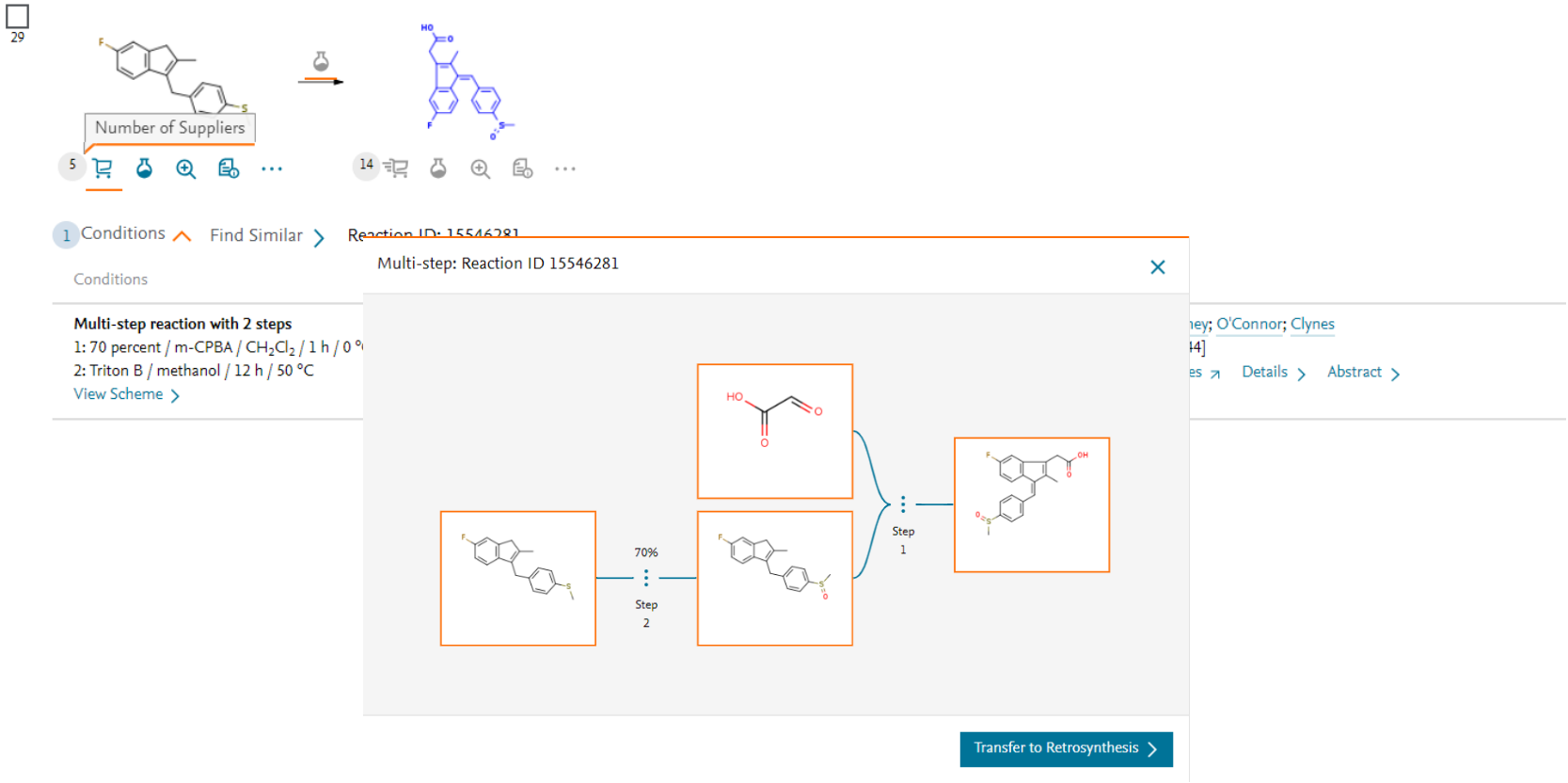
示範 published molecule



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



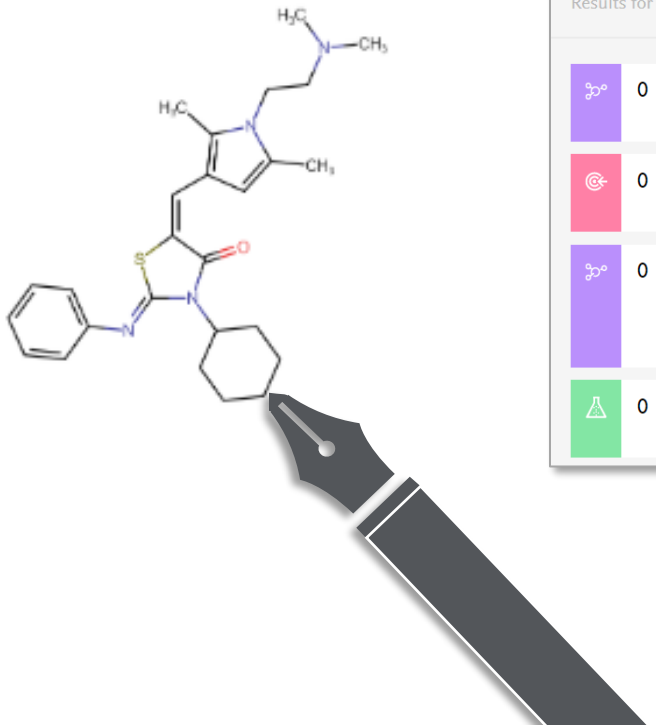
未發表或創新分子的合成方法

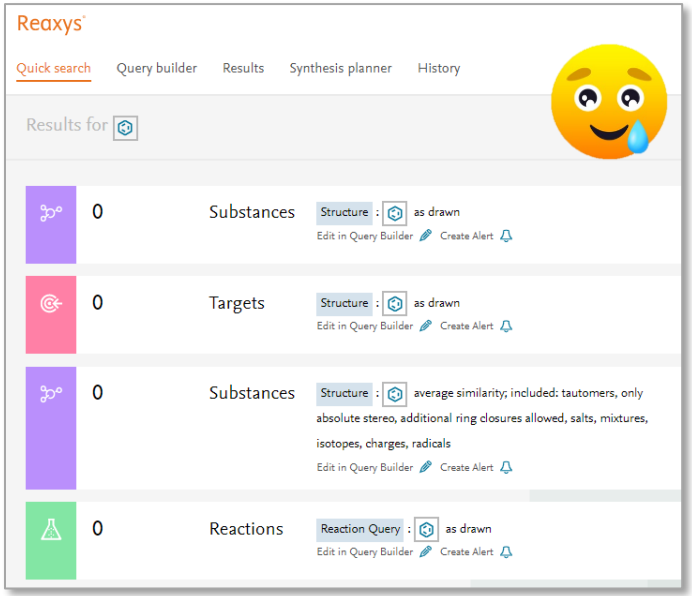
~創新分子不容易找到參考資料~

- 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。

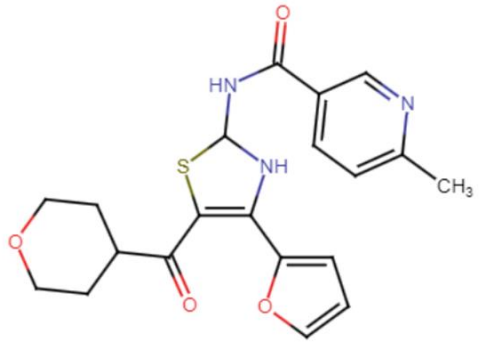




未發表或創新分子的合成方法

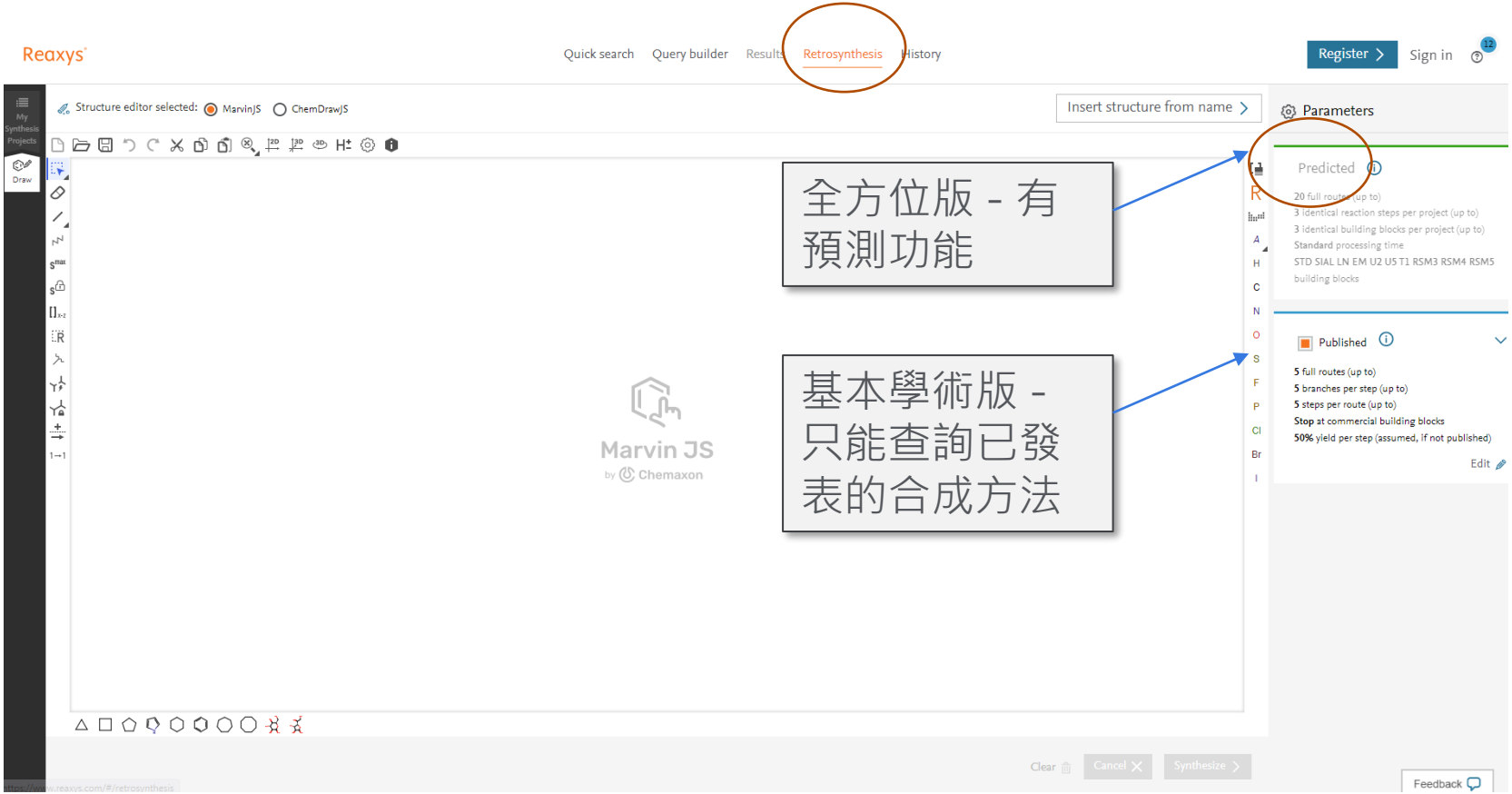
~創新分子不容易找到參考資料~

示範 unpublished molecules



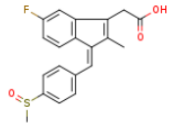
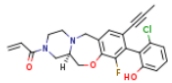
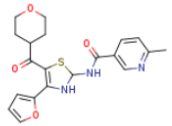
Novel Adenosine A2A receptor antagonist
By Kyowa Kirin Co., Ltd.

Reaxys AI逆合成引擎整合於Retrosynthesis工具



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
1410201	07 Sep 2023 02:13	Project #1410201 Delete	 Edit	Predicted 8 Published 5 View
1406540	07 Sep 2023 03:51	Project #1406540 Delete	 Edit	Predicted 2 Published 1 View
1405529	08 Sep 2023 02:58	Project #1405529 Delete	 Edit	Predicted 2 Published 0 View

Reaxys 逆合成AI打造給Bench chemists使用

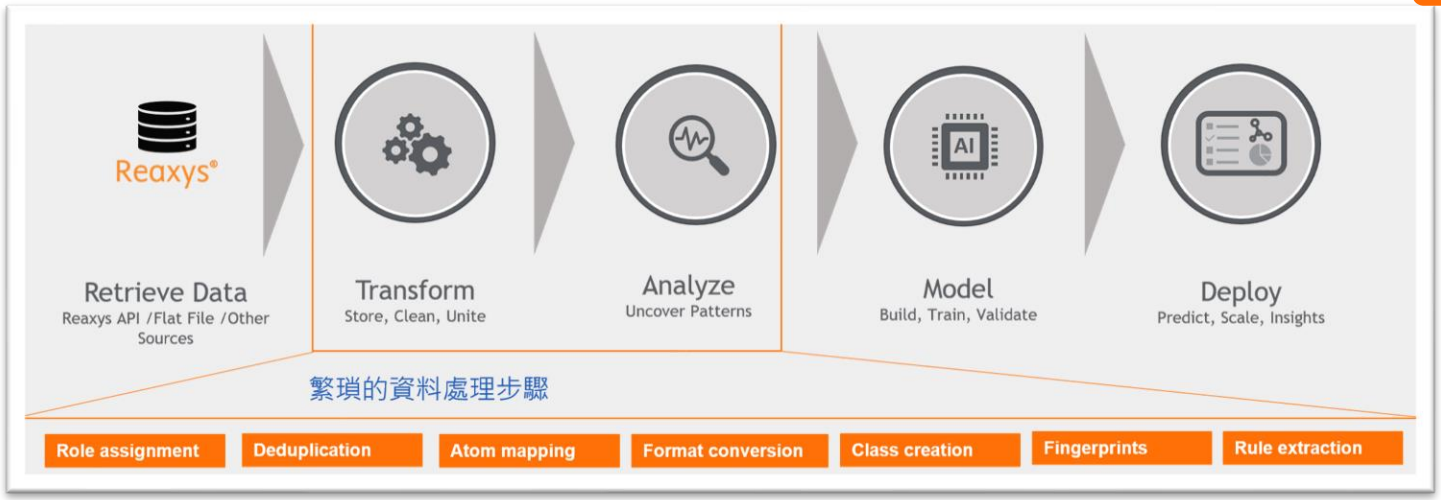
資料科學Data Science流程

Scientific Robustness

簡易使用

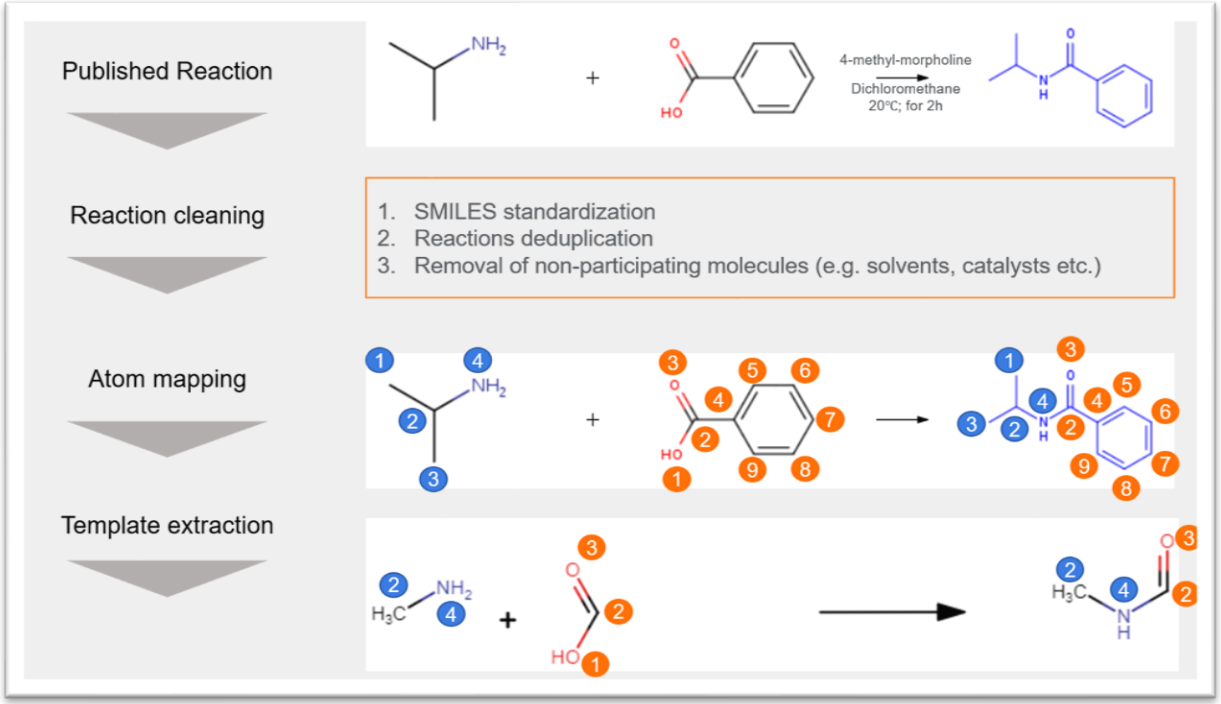
Reaxys Retrosynthesis Tool

化學家

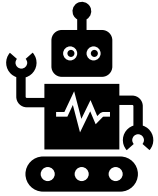


資料科學家

Reaxys-PendingAI 以機器從發表的反應式擷取轉化規則



Machine Rule Extraction

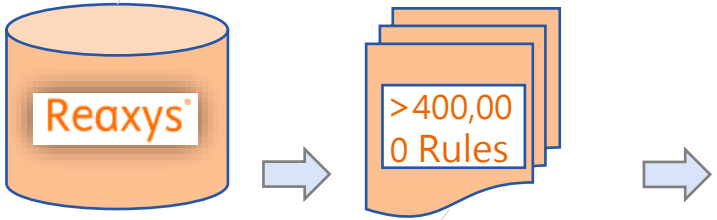


- ✓ 新發表的化學研究指數般的成長，機器能跟上發表的**速度**
- ✓ **客製化**的可行性，私人機構累積的大量反應式資料可併入讓機器學習

撥開Reaxys PAI的面紗

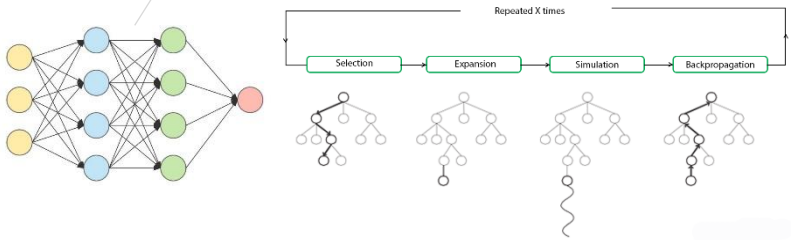
結合最大的Reaxys反應資料庫與神經網路技術

最大且最高品質
超過15M 準機器學習、單步驟反應式資料



機器合成規則擷取
避開以人工編寫規則耗時的缺點，以機器自動擷取合成規則可確保效率與& 可接受客製化的能力

3N-MCTS Technology
採最前緣的深度學習科技，並受高影響期刊文獻證據支持*



Reaxys PAI 的能力：
- ✓ 自動從反應式中擷取化學轉化規則
- ✓ 優先套用、排序不同的規則
- ✓ 採用現代有效率的搜尋策略
- ✓ 學會針對新分子預測合成路徑

Segler, M., Preuss, M. & Waller, M. Planning chemical syntheses with deep neural networks and symbolic AI. Nature 555, 604–610 (2018). https://doi.org/10.1038/nature25978

Reaxys PAI 如何推導出新化合物的合成路徑?

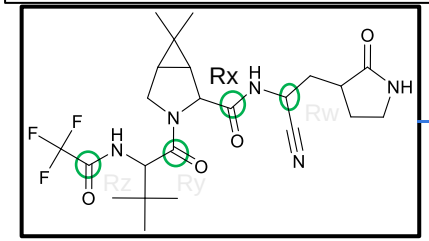
Step 1: Identify most probable disconnections

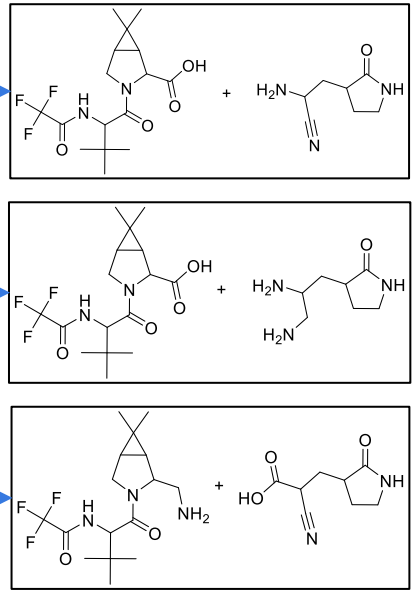
Step 2: Apply retro rule to create starting materials

Possible synthons from application of Retro Rule X

Retro Rule R_X

找出最有可能的結構斷點





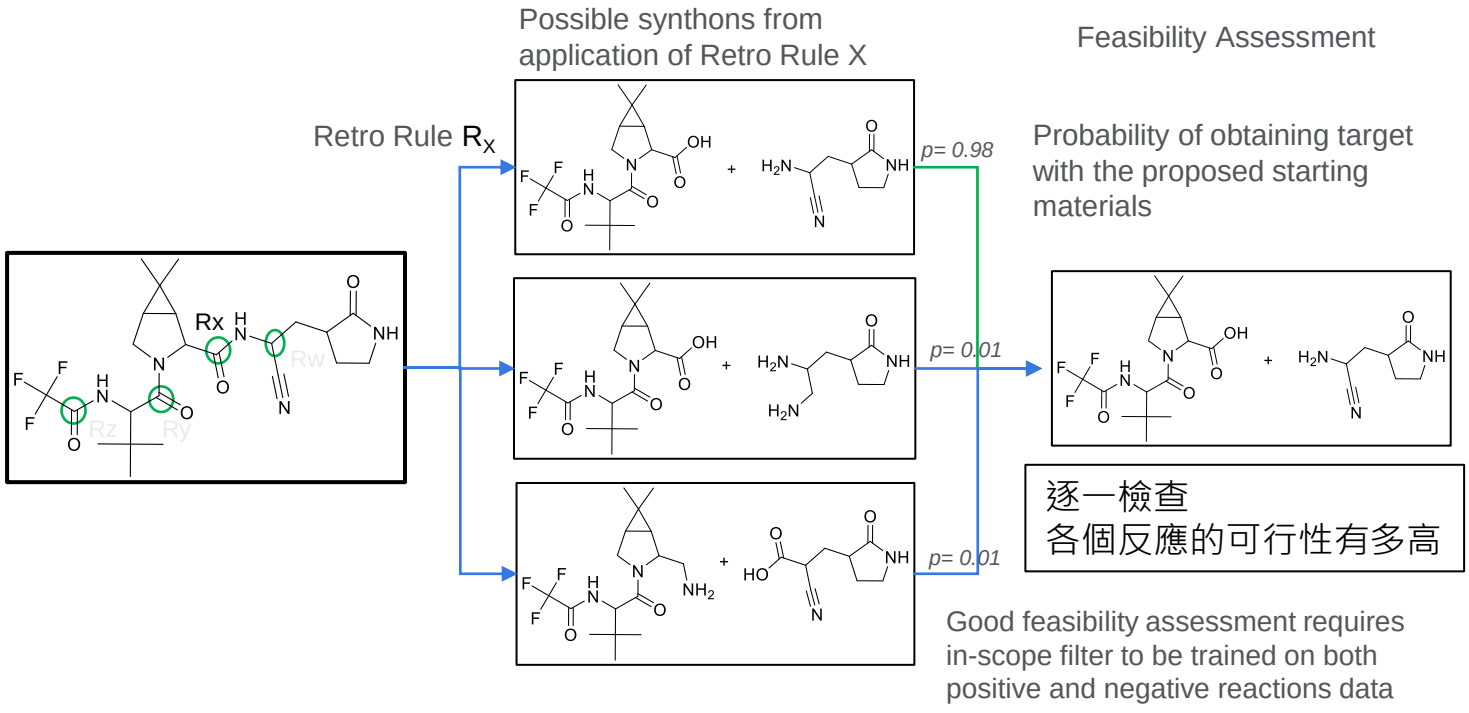
套用rules，生成可能的組合

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?

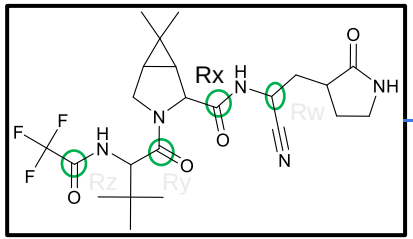
Step 1: Identify most probable disconnections	Step 2: Apply retro rule to create starting materials	Step 3: For each reaction use in-scope filter	Step 4: Monte Carlo Tree Search

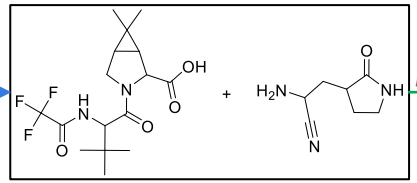
Iterate steps 1, 2 & 3 until commercially available starting materials are identified or time budget is reached

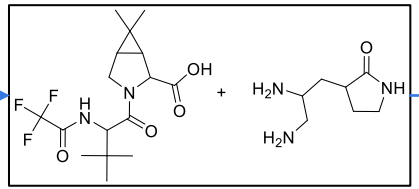
時間預算

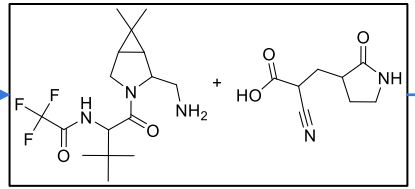
Retro Rule R_X

Possible synthons from application of Retro Rule X







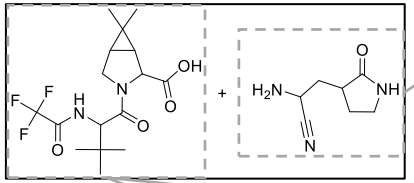


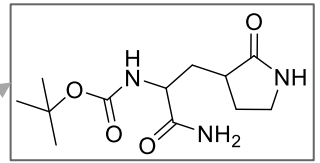
p= 0.99

p= 0.01

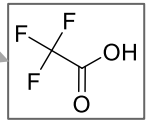
p= 0.01

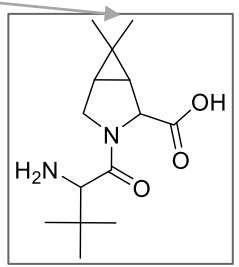
Probability of obtaining target with the proposed starting materials





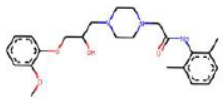
檢查是否為商用可購買的材料Commercial BB





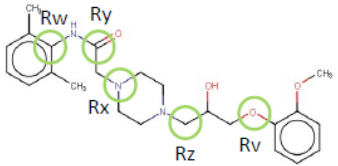
Good feasibility assessment requires in-scope filter to be trained on both positive and negative reactions data

Reaxys PAI 如何給每個預測路徑打分數?

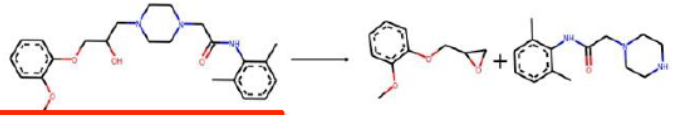


Target compound

1st step: disconnection identification
> A probability is given for each disconnection (rules)



2nd step: Application of the rule
> Application of rule Rz for instance

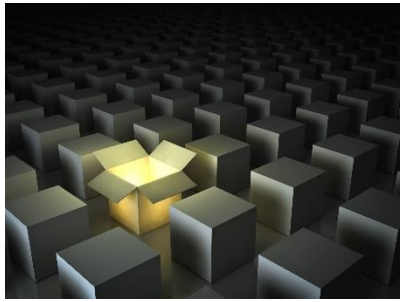


3rd step: Reaction prediction
> Check if the reaction is relevant

4th step: Monte Carlo Tree Search (MCTS)
> Iterative application of steps 1, 2 & 3 until a feasible synthetic scheme is found and starting materials identified

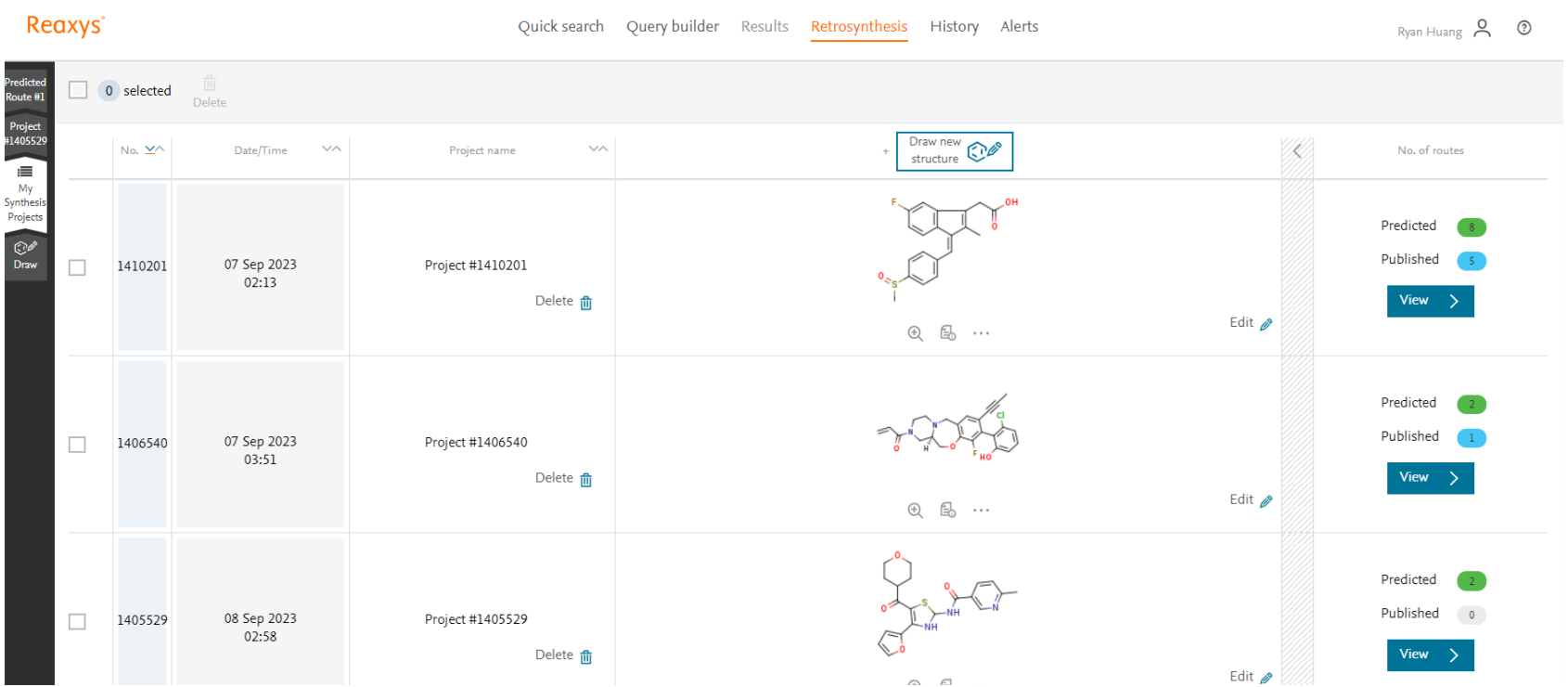
M. H. S. Segler, M. Preuss, M. P. Waller, Nature 555, 604–610 (2018)

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



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」

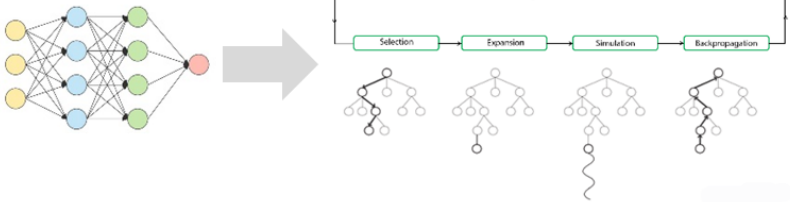


Reaxys的獨特之處①

- >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

ORD

- > 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

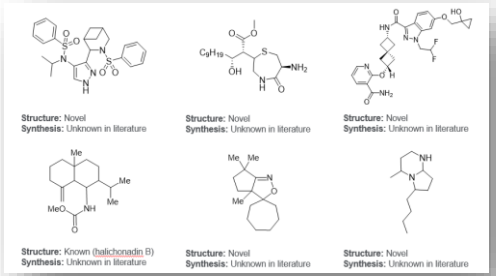
- **Assess scientific viability of the route**
- Time and effort required for optimal route generation

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



Feedback from Carriera'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

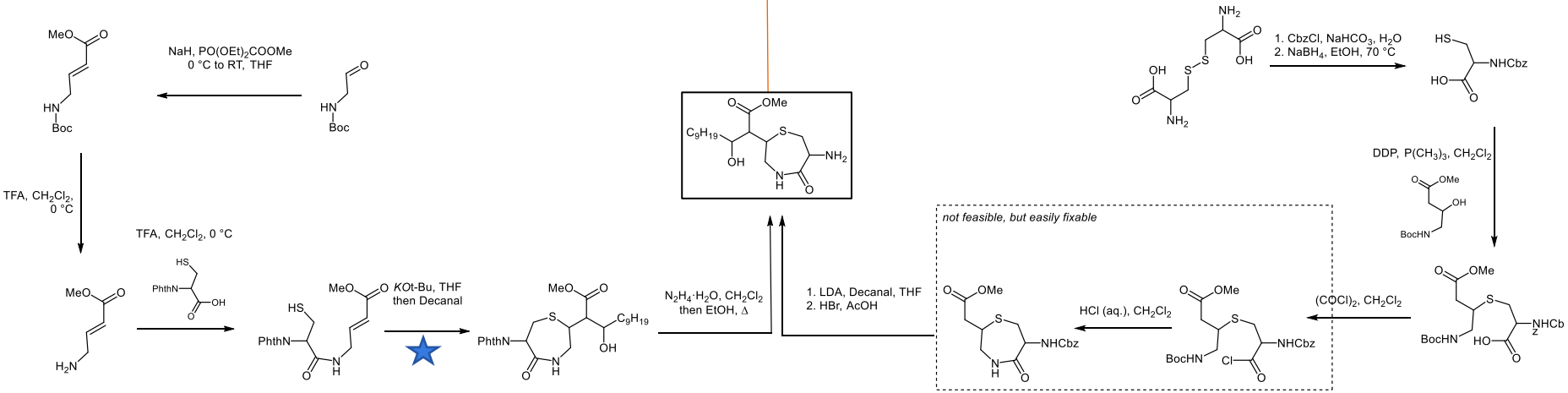
Radboud Universiteit
SINCE 1923

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

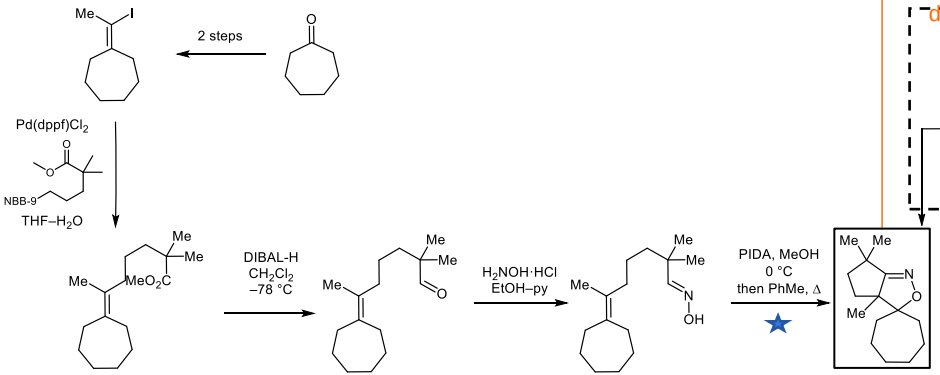


1. The molecule is a fragment from a peptide project in our group and has been synthesised before
2. The molecule doesn't have too much precedence for machine to take inspiration from
3. Robust route proposed by retrosynthesis but chemist needs to review proposals especially with regards to functional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

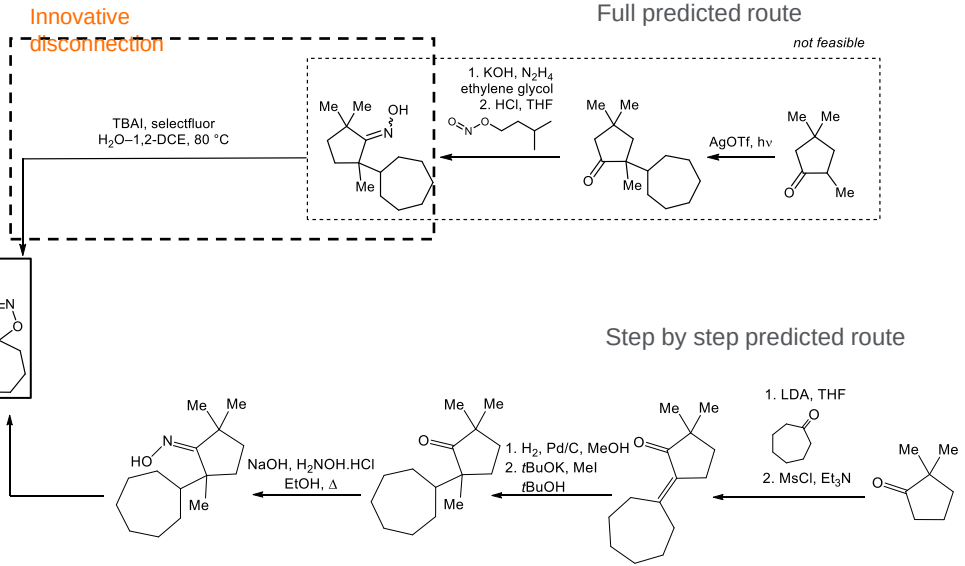
Reaxys retrosynthesis in the hands of chemists from Prof. Carreira's Group at ETH Zurich

Example of a natural product

Human generated route



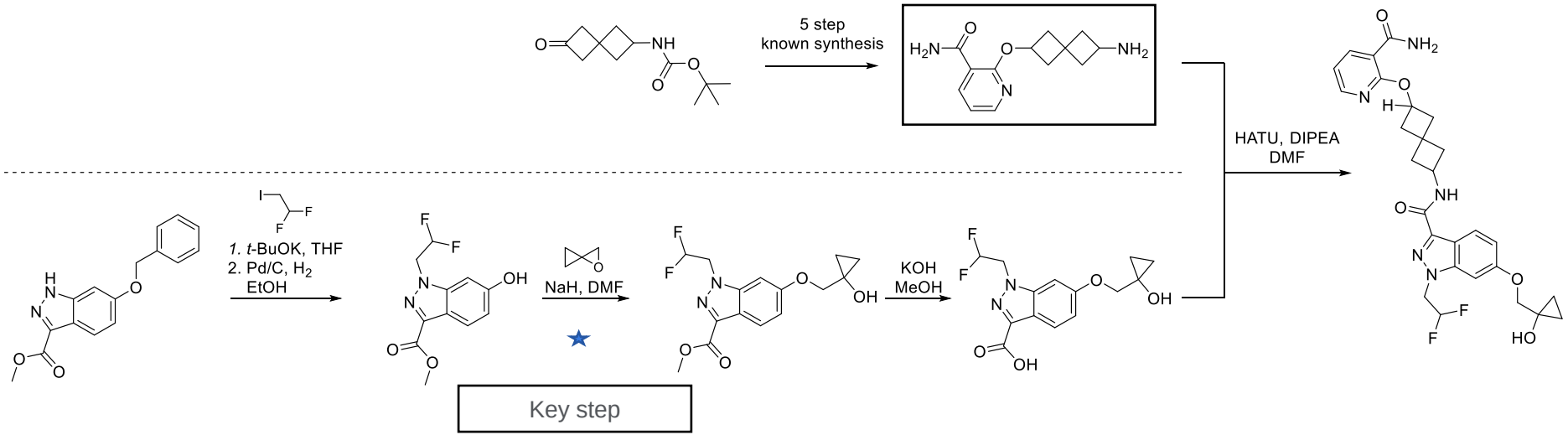
Retrosynthesis Tool route



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

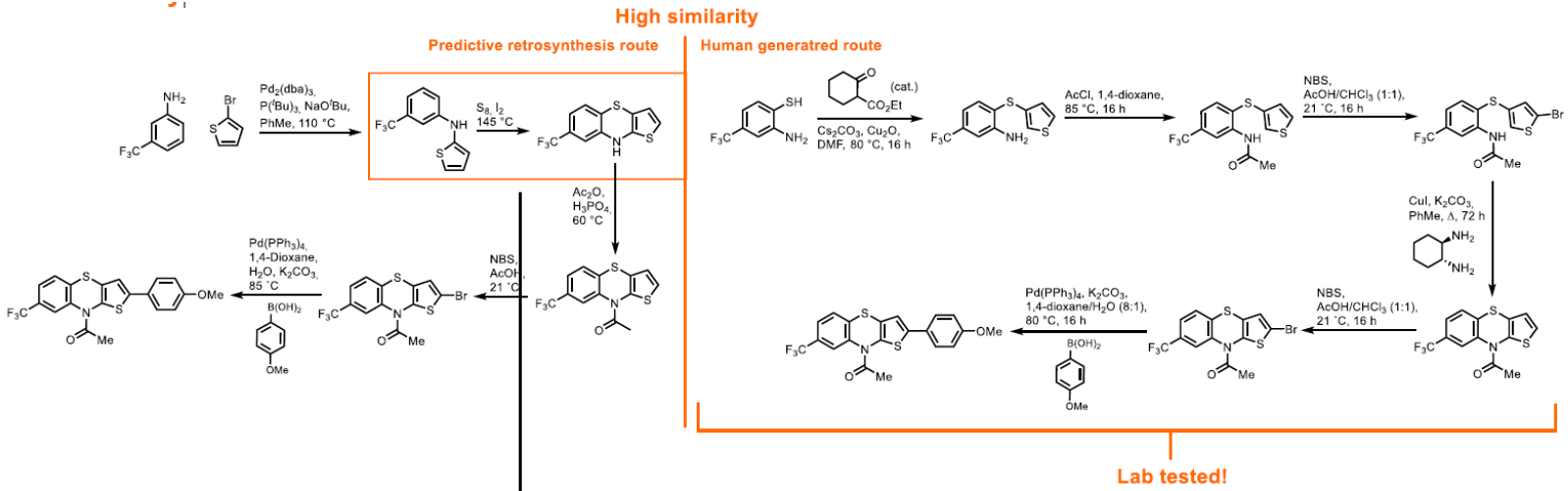
Introducing the cyclopropyl carbinol on the indazole scaffold is a challenge and the predicted route suggests an innovative way to introduce this functionality

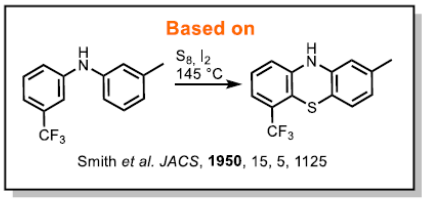
Retrosynthesis Tool route



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

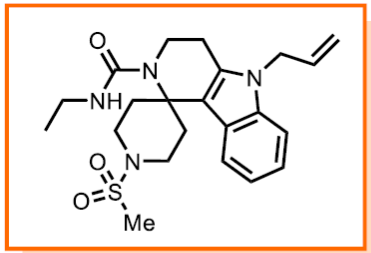




- Interesting disconnection that is non-obvious
- Providing alternative for the ring closure

Radboud Universiteit
SINCE 1923

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

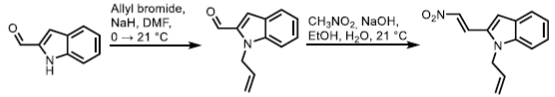


- Large overlap in synthetic output
- More selective

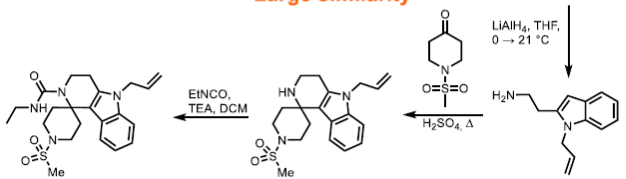
- Starting material commercially available though expensive

Not known in literature

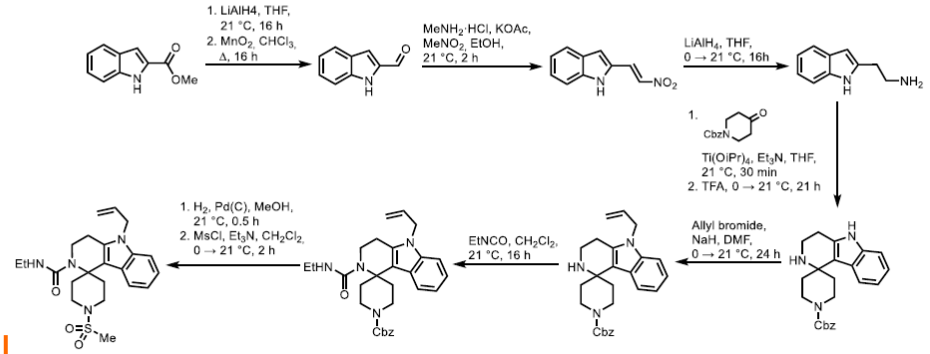
Predictive retrosynthesis route



Large similarity



Human generatred route



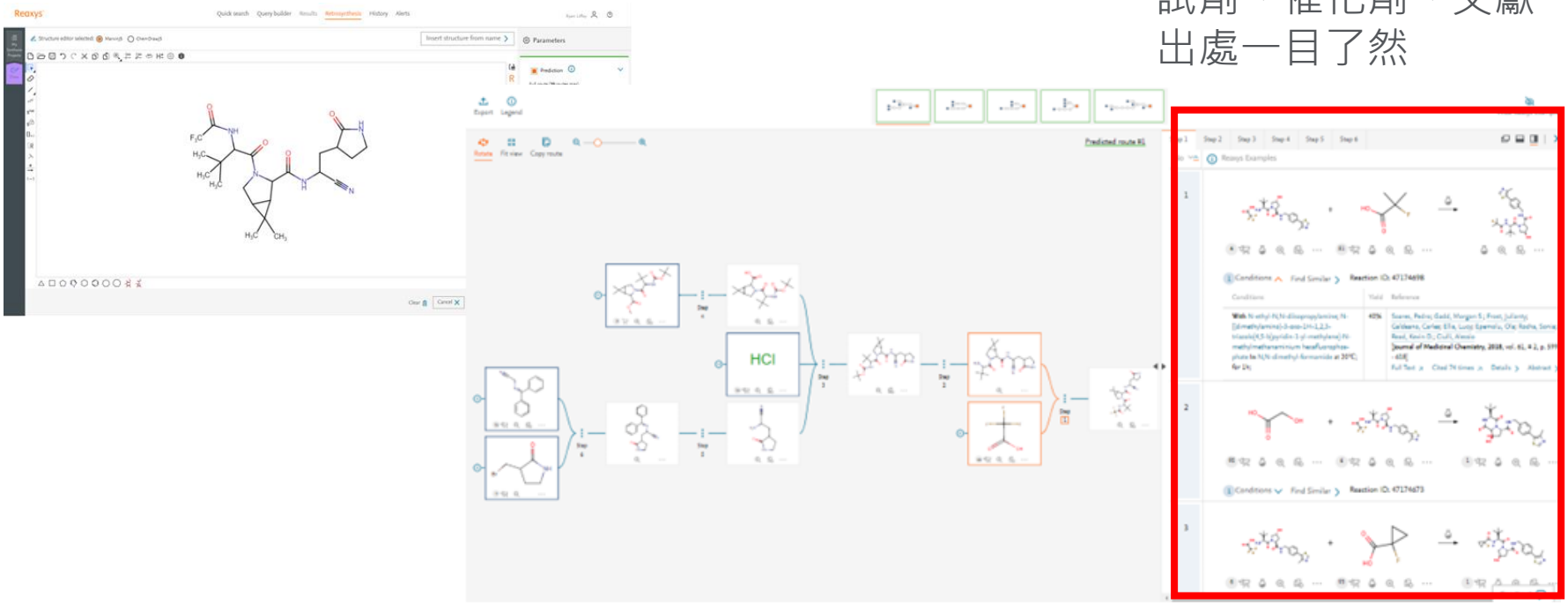
Lab tested!

Radboud Universiteit
SINCE 1923

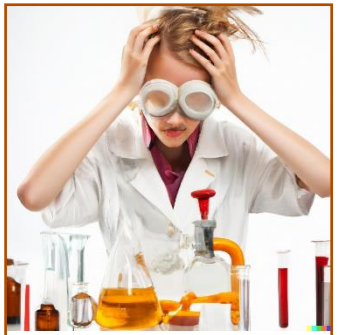
Reaxys的獨特之處③

- 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



利用Reaxys AI工具來優化學習曲線、節省時間、平衡競爭態勢



實驗室學生

- 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

研發團隊

- 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

化學系/化學所

- 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

Thank you

歡迎來信留下您的問題與寶貴建議

r.huang@elsevier.com