

## 研究性文章总结

有机化学学类

# 一个流动合成有机化合物的机器人平台通知

## 按人工智能的规划编制

戴尔, 戴尔 A. 托马斯三世\*, 贾斯汀 A. M. Lummiss\*, 乔纳森·n·贾沃斯基, 克里斯托弗·p·布林, 维克多·舒尔茨, 特拉维斯·哈特, 乔舒亚·菲什曼, 卢克·罗杰斯, 韩愈高, 罗伯特·希克林, 彼得·普莱希尔, 乔舒亚·宾顿, 约

制造功能化合物是至关重要的, 包括小分子药物。尽管在实验室自动化方面取得了进步, 但合成路线的识别和开发仍然是一个手工过程, 实验合成平台必须进行人工配置, 以适应要进行的化学类型, 需要专家化学家的时间和精力投资。理想的自动化合成平台将能够规划自己的合成路线, 并在促进实现生产目标扩大的条件下执行它们。在之前的研究中, 化学开发过程中的各个元素(设计、路线开发、实验配置和执行)已经被简化, 但没有一个提出了计算机辅助合成规划(CASP)、专家精炼化学配方生成和机器人执行化学合成的集成路径。

可扩展的合成方法, 结合了人工智能(AI)技术的规划和机器人技术的执行。数以百万计的先前发表的反应通知合成路线的计算设计; 专家编写的化学配方文件(CRFs)在一个机器人流动化学平台上运行, 以进行可伸缩、可重复的合成。这种开发策略增强了化学家处理面向目标的流合成的能力, 同时大大减少了必要的信息收集和人工工作。

结果: 我们为CASP开发了一个开源软件套件, 从Reaxys数据库的数百万个反应U. S. 专利和商标局。该软件的设计目的是通过学习应用反合成转化, 识别合适的反应条件, 并评估反应是否有可能发生, 将已知的化学反应推广到新的基底上

当通过实验尝试时就成功了。建议的路线部分填充CRF, 这需要化学家用户提供额外的详细信息, 以定义停留时间、化学计量学和与连续流动一致的浓度。为了执行这些合成, 机械臂根据在CRF中定义的期望工艺配置将模块化工艺单元(反应器和分离器)组装成连续的流路径。该机器人还将试剂线和计算机控制的泵连接到反应器入口

在我们的网站上

阅读全文

<http://dx.doi.org/10.1126/science.aax1566>

通过一个流体交换盘。当这完成后, 系统启动线的序数并开始合成。在指定的合成时间后, 系统冲洗这些线

使用清洁溶剂, 机械臂断开试剂线, 并将工艺模块移除到其适当的存储位置。

这种流动化学发展的范例对15个医学相关的小分子得到了证明。为了越来越复杂, 我们连续研究了阿司匹林和西硝唑的合成; 利多卡因和安定连续使用常用原料; (S)-华法林和萨法酰胺以证明规划计划的立体化学意识; 和两个复合文库: 5个ACE抑制剂, 包括奎那普利和四种非甾体抗炎药, 包括塞来昔布。这些目标总共需要8种特定的反合成路线和9种特定的工艺配置。

结论: 本文的软件和平台代表了完全自主化学合成道路上的一个里程碑, 其中路线

仍然需要人类

输入和过程开发。随着时间的推移, 这个和类似的自动化实验平台产生的结果可能会减少我们对历史反应数据的依赖, 特别是与更小规模的流筛选平台相结合。增加反应数据的可用性将进一步使基于人工智能建议的机器人实现合成, 缓解专家化学家的手动任务, 使他们可以专注于新的想法。

作者成员的列表可以在完整的在线文章中找到。

\*这些作者对这项工作的贡献同样大。

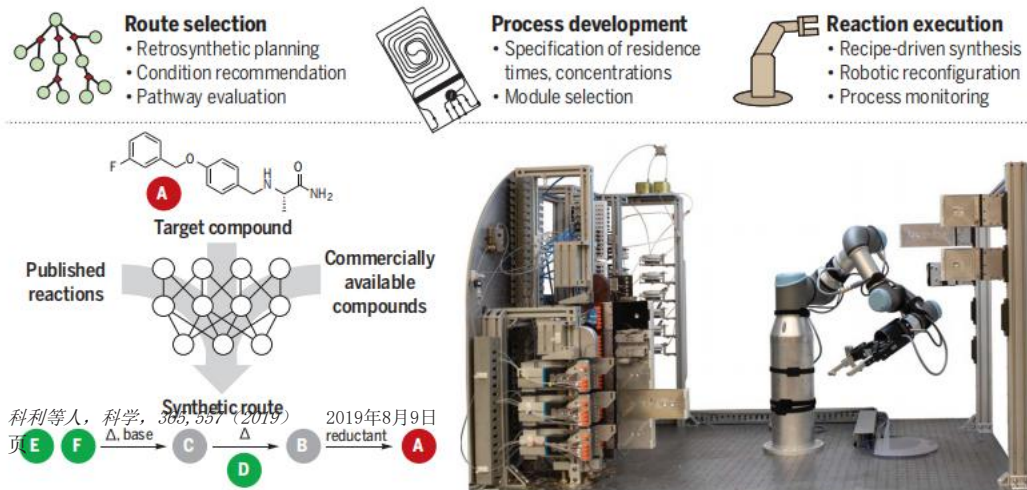
通讯作者: 电子邮件地址:

t.f.j@mit.edu (T. F. J.);

k.f.jensen@mit.edu (K. F. J.)

引用这篇文章为C. W. 科利等人, 科学365, eaax1566 (2019)。

美国能源部: 10.1126/science.aax1566



规划和执行。一个可重构的流动化学平台执行人工智能计划的多步化学合成。

# 一个流动合成有机化合物的机器人平台通知

## 按人工智能的规划编制

康纳, 科利<sup>1\*</sup>, 戴尔A. 托马斯三世<sup>1,2\*</sup>, 贾斯汀·伦米斯<sup>3\*</sup>, 乔纳森N. 贾沃尔斯基<sup>3</sup>, 克里斯托弗P. 布林语<sup>3</sup>, 维克多·舒尔茨<sup>1</sup>, 特拉维斯·哈特<sup>1</sup>, 约书亚·菲什曼<sup>2</sup>, 卢克·罗杰斯<sup>1</sup>, 韩高<sup>1</sup>, 罗伯特·W·希克林<sup>3||</sup>, 彼得的家<sup>1</sup>, 约书亚·宾顿<sup>1</sup>约翰. 皮奥蒂<sup>2</sup>, 威廉先生, 格林先生<sup>1</sup>, A. 约翰. 哈特<sup>2</sup>, 蒂莫西·贾米森<sup>3\*\*</sup>, 克拉夫夫人F. Jensen<sup>1\*\*</sup>

复杂有机分子的合成需要几个阶段, 从构思到执行, 需要专家化学家的时间和精力投资。在此处, 我们报告了一个迈向化学合成范例的步骤, 以减轻化学家的日常任务, 结合人工智能驱动的合成计划和一个机器人控制的实验平台。通过推广数百万种已发表的化学反应, 提出了合成路线, 并在硅胶中进行了验证以最大化他们成功的可能性。额外的实施细节由专家化学家确定, 并记录在可重用的配方文件中, 该文件由一个模块化的连续流平台执行, 该平台由机械臂自动重新配置, 以设置所需的单元操作并执行反应。证明了这种计算机增强化学合成的策略15种药物或类药物的物质。

他按需合成有机化合物的能力有潜力改变分子发现任务。这种典型分子量为50–750g/mol的化合物在一定范围内起着核心作用学科, 包括专业聚合物, 有机光伏, 能量学, 和药物。合成通常是分子药物发现(1)中的一个瓶颈, 其中设计-合成-测试迭代的周期时间为周, 而复合库合成的范围可以确定经验结构-活动关系模型(2)的准确性。材料发现研究人员也面临着类似的限制, 因为他们无法合成各种化合物, 如候选有机光伏, 以及如此快速的(3)。

许多化学家和化学工程师正在追求一种机器的承诺, 一种能够在几乎没有人工干预(4, 5)的情况下合成大量分子的机器。尽管实验室自动化的重大进展减少了进行某些化学反应所需的人工努力(6–8), 但识别和开发新分子的合成途径仍然很大程度上是手工过程, 需要专家化学家的时间投资。此外, 当前的自动合成平台必须首先被配置为适应必要的单元操作序列, 或被限制在其他可访问的化学空间的子集内。与当前自动平台相容的化学反应的范围往往受到反应类型(9)、溶剂和温度(10)或浓度和时间(11)的限制。理想的自动合成平台将与反应条件兼容, 可以直接从小规模的过程开发转化为克或公斤的制造。连续处理方法, 例如插流反应器或连续搅拌罐反应器的合成, 可以提供这种可伸缩性, 并在许多方面被广泛公认为支持技术, 包括工艺质量改进(12)。相对于批合成, 较小的长度尺度增强了热和传质, 并且可以在放大(13)之前进行更精确的量化。此外, 与批量处理平台相比, 流平台提供了更小的足迹

验证通过过程强化(14, 15)获得加速反应速率。许多多步骤合成已经成功地实现了流动合成, 并大大减少了总反应时间(14, 16–18)。

小分子的化学开发过程可分为多个不同的阶段, 包括设计(文献搜索、反合成、条件选择、可行性估计)、路线开发(配方配方)、实验配置(平台重构)和执行(过程执行、可伸缩合成)(图. 1A)。以前的研究已经试图使这个过程的所有方面实现自动化, 但还没有提出完全自动化的路径。可以使用化学疗法的专家编码的反应规则(19, 20)或Segler等人来简化后录合成。但前一种方法的规模很难随着化学文献的发展而扩大, 而且既没有明确提出反应条件, 也没有评估正向反应的可行性。使用预定义指令集的自动化学合成在(6, 8)中得到了很好的证明, 但仅限于批处理, 因此不能提供明确的扩大合成的路径。虽然流程化学平台已开发用于自动筛选、优化和生产(5、7、22–24), 但它们需要手动重新配置到每个流程所需的精确流程路径。

我们的自动化、可伸缩的合成方法结合了人工智能(人工智能)的技术规划和机器人技术执行。

具体地说, 我们描述了一个平台, 可以通过推广数百万个之前发表的反应来设计合成路线(图. 1B), 包括反应条件和工艺变量的部分规范, 然后使用机器人可重构的流动化学平台执行人工精炼的化学配方文件(CRFs)(图. 1C)。对与连续流的兼容性所需的人工智能提出的合成路线的调整被记录在这些可重用的可重用配方中, 以实现可扩展, 可重复的合成。

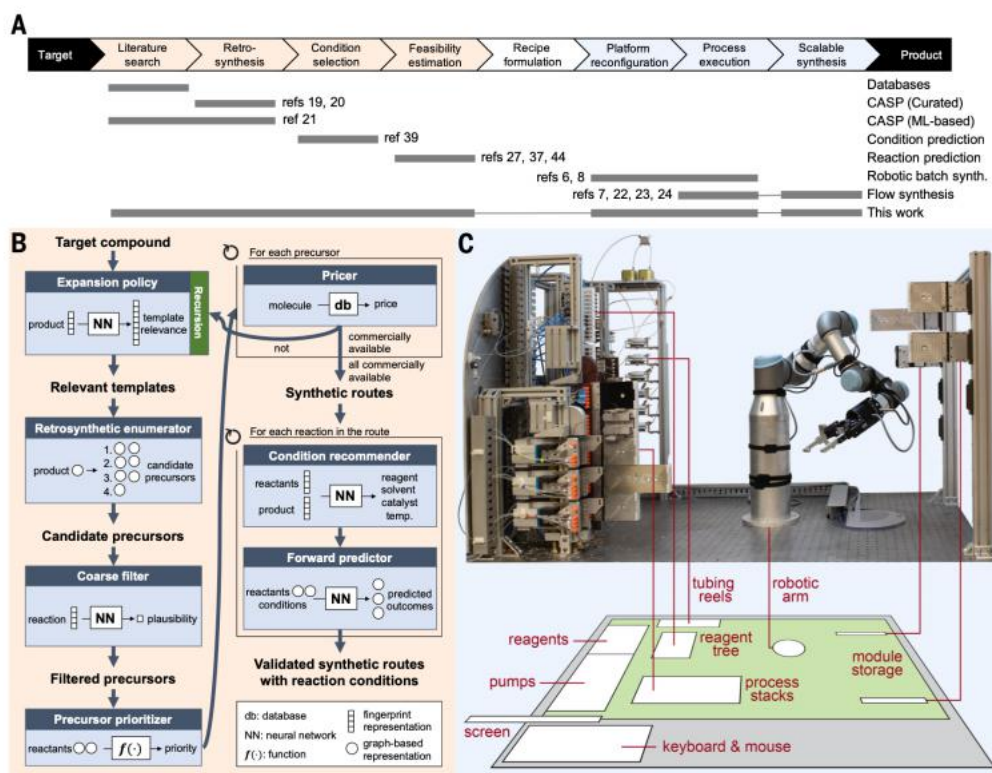
这种开发策略增强了化学家处理面向目标的流合成的能力, 同时大大减少了必要的信息收集和人工工作。我们通过预测和自动化15种药物或药物样分子的合成来说明这种化学发展的范式。

在这个工作流中, 我们依赖于来自化学家用户的专家输入, 最低限度地包括居住时间、等效比和浓度, 以将建议转化为实践。特别是, 预测与流量的相容性需要在目前只有经验才能实现的精确度水平上预测溶解度; 确定伸缩反应序列和放置阶段间分离的机会需要对反应结果和试剂相容性进行定量预测。更实际的是, 描述浓度、等效比和添加顺序的数据没有列在任何可用的反应数据库中, 排除了数据驱动的完整过程规范、流程或其他方法。在中的

<sup>1</sup>剑桥大学, 麻省大道77号, 麻省理工学院化学工程系, 马州02139, 美国。<sup>2</sup>美国麻省理工学院机械工程系, 麻省大道77号, 邮编02139。<sup>3</sup>美国麻省剑桥大道77号麻省理工学院化学系, 邮编02139。  
\*这些作者对这项工作的贡献同样大。  
现地址: Mytide治疗公司, 波士顿, 美国马州02111。  
现地址: 美国新泽西州新不伦瑞克施贵宝, 邮编08901。现地址: 点播制药, 生物技术8, 里士满, 弗吉尼亚州23219。  
||现地址: 辉瑞, 格罗顿, 美国06340。现地址: 比利时根特大学根特大学化学技术实验室。目前地址: L3技术空间系统, 威尔明顿, 马萨诸塞州01887, 美国。  
\*\*通讯作者。电子邮件地址: tfj@mit.edu (T. F. J.); kfjensen@mit.edu (K. F. J.)



图中示。1. 化学开发管道背景下可机器人配置反应规划和执行平台的概述。(A) 按需合成目标产品的工作流程  
有机化合物和 (19, 20) 的有机化合物和代表性工作,  
沃勒 (21), Jensen (27, 37, 39), 施瓦勒 (44), 戈弗雷 (6), 克罗宁 (8), Jamison和Jensen (7, 22, 23) 和Ley (24) 在这个过程中有流线型; 厚的灰色线表示自动步骤。  
(B) 结合化学信息学和机器学习的软件模块, 以设计和验证合成路径。  
(C) 机器人流动化学平台的照片, 投影平面图  
6英尺×4英尺工作台 (灰色背景) 和通风外壳 (绿色背景)。  
CASP, 计算机辅助合成规划; ML, 机器学习。



在“展望”部分, 我们定义了自动生成食谱所必须克服的挑战。

### 综合规划模块

计算机辅助合成规划 (CASP) 最初是一种工具, 帮助化学家在实验室 (25, 26) 中执行路径之前识别路径。反应数据库 (如 Reaxys、SciFinder) 已经简化了搜索已知化合物以寻找已知合成物的过程, 现在已被常规使用。然而, 将合成规划推广到新化合物的项目并没有得到广泛采用, 这可能表明它们目前还很有限的能力和成本。

对计算机建议的合成计划几乎没有明确的验证; 一个例外是最近使用化学程序发现基于专家编码转换规则 (19) 的新路线。数据科学和机器学习的最新进展对 CASP 的兴趣重新导致了最近的几项研究 (27, 28–30), 包括 Segler 等人。的蒙特卡罗树搜索加快推荐过程 (21)。

本部分描述我们的综合规划计划的发展。1B)。它集成了我们以前的努力, 将已知的化学推广到新的衬底上, 通过学习应用反合成转换, 识别合适的反应条件, 并评估反应在尝试实验到一个单一的开源软件框架时是否可能成功, 我们称之为 ASKCOS。这个软件是直接训练在数百万人的反应前

从美国专利商标局 (USPTO) (31) 或 Reaxys 表格。

我们将 CASP 看作是一个推荐问题。对于特定的靶分子, 该程序必须从一系列可行的化学反应物开始提出化学可行的反应步骤。如果目标是只识别已知分子的已知合成物, 那么这可以被视为一个图搜索问题。然而, 为了识别新的合成物, 必须动态生成候选反应步骤来定义搜索空间。反应模板是子图匹配规则, 可以从文献先例反应中算法提取, 并应用于新的基底来识别适合反向合成分离的结构基题。从 Reaxys 数据库中发布的 1250 万个单步反应中, 我们准备了所有 10 次观察  $\geq$  的规则和所有 5 次观察特定立体化学  $\geq$  的规则, 共 163, 723 条规则。立体化学的转变本质上更具体, 预计会出现较少, 但必须包括允许程序预测手性分子的合成。该程序使用 RDKit 和 RDChiral 应用转换并强制处理立体化学反应, (32, 33)。

我们训练了一个前馈神经网络模型来预测 163, 723 条变换规则中的哪一个最适用于基于其分子结构的目标分子 [ECFP4 (34)] (28)。仅应用被认为最相关的模板有两个角色: (i) 减少详尽的模板子图匹配的计算成本和 (ii) 增加的可能性

提出了一种在化学上可行的反应。我们把大部分精力集中在后一点上: 确保反应建议不是假阳性 (即在实验室中不起作用的建议)。这对于最大化实验成功的概率是至关重要的。

两个模块评估一个逆转录合成建议的质量。第一个是一个二进制分类器, 旨在只删除基于 Segler 等人的最低质量的建议的“范围内过滤器” (21)。该神经网络模型由 1500 万发表的正反应例和 1.15 亿人工负反应例应用算法提取的正向反应模板生成, 完全遵循 (27) 中的协议。二进制分类器是为了回答这个问题: 这些反应物是否有任何一套条件, 使这些反应物将形成这个产品? 使用用户可调阈值通过此过滤器的反应将被添加到不断增长的搜索树中。

在返回使用适当起始材料的所有路径之前, 在预定的时间内 (下面描述的所有示例不超过 30 秒) 或达到指定的深度。该程序使用根并行蒙特卡罗树搜索 (特别是一种上置信界方法) 来平衡被认为很有前途的分支的开发和访问较少的分支 (35, 36, ) 的探索。我们的可购买化学品数据库包括大约 107, 000 种化合物, 低于 \$100/ 格的分子和西格玛-奥尔德里奇, 尽管每个扩展的最高价格是每个扩展和用户可调参数

还有额外的停止标准（请参见第部分 1.5. 可在补充文本中使用）。

由二元分类器认为合理的反应步骤组成的路径可以通过更严格的正向预测模型来评估，该模型根据提供的反应物和反应条件（37, 38）明确生成产物分子；使用的精确模型在（37）中详细描述。反应条件由一个神经网络模型提供，训练提出一个优先列表的试剂，溶剂，催化剂，和温度最适合的转变，如先前报道的（39）。如果根据正向预测器的可能结果与预期结果匹配，那么我们可以更有信心，该反应在化学上是可行的。前向预测器模型还提供了副产品的预测，有助于工艺开发过程中的杂质识别。由于局限性，立体选择性反应还不可能得到明确的验证

在三维分子表示和不一致的数据报告中。

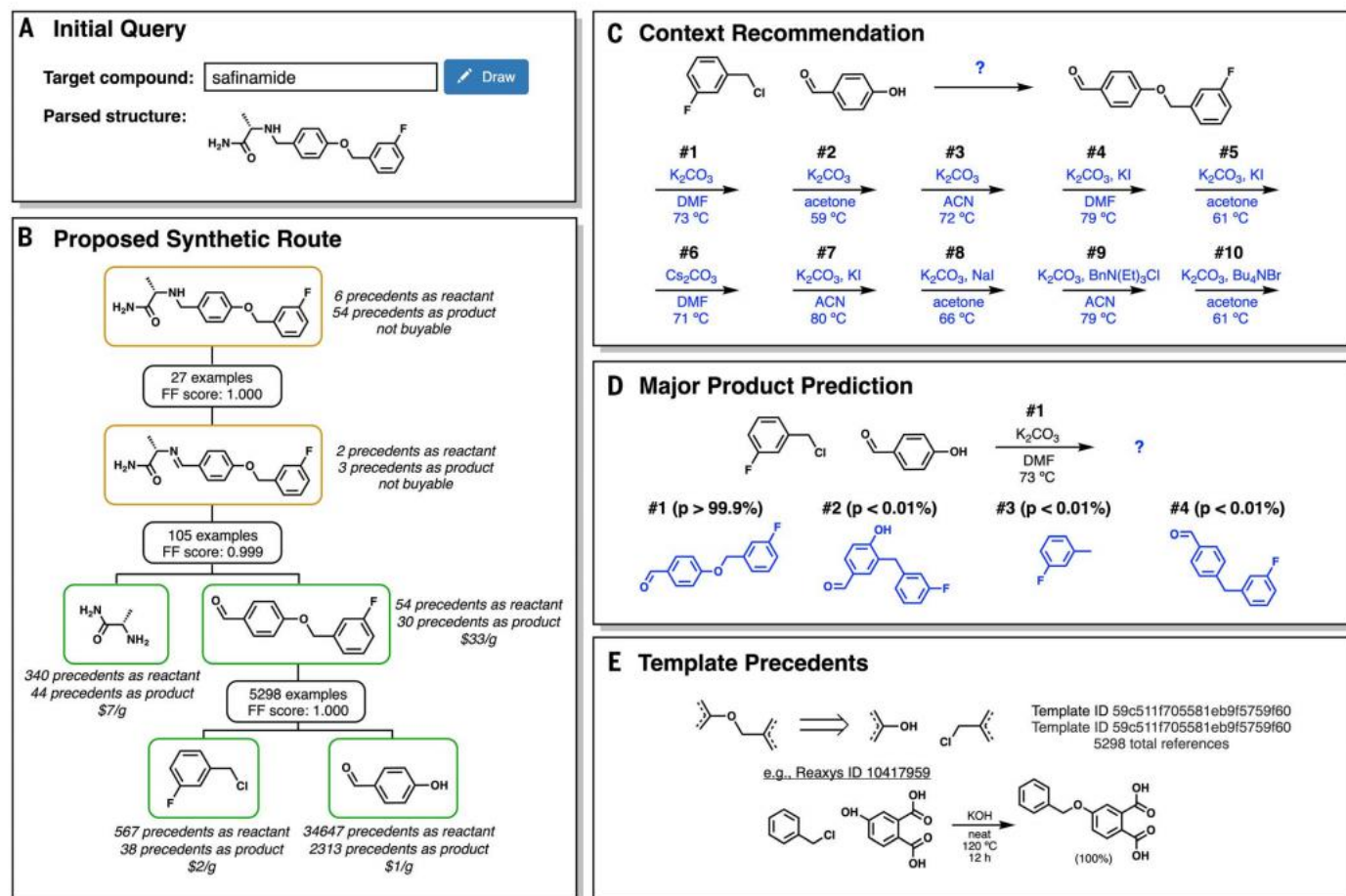
即使在这些验证和过滤阶段之后，途径级别的关注也可能仍然存在。除了上述流动化学考虑，用户还可以偏向那些经常大规模实现的路径（例如，过程应用），或使中间体多样化的路径（例如，发现应用）。我们的代码框架允许扩展到这样的考虑因素。图所示目标代码输出的详细说明。2.

我们可以计算规划的合成路线的复杂性超过了在伸缩流合成中实现的意义。此外，CASP项目可以预测到它以前从未见过的目标的路线，如最近报道的现代活性药物成分（APIs）和全新的分子优化-

目标（见补充文本第1.10节）。

## 机器人技术流动化学平台

我们考虑了设计自可配置系统的概念，基于之前关于手动即插即用单元操作（如连续合成（22, 40）的反应器和分离器）的几个概念。作为一种灵活、高效的过程模块设计，我们选择实现一个六轴机器人机械臂，从存储位置选择过程模块，然后按照特定合成所需的顺序排列模块（图. 1C和无花果. S12至S14）。这种安排还使现有过程模块的直接升级以及添加新的模块成为可能，所有这些都使用一个通用的流体接口。或者，我们考虑用双向阀实现机组操作组；然而，这种安排会产生冗余和更大的复杂性，



物；  
(B) 向用户展示的众多合成路线之一，附加关于商业可用性和可用的先例信息；(C) 以体化步骤提出的十大反应条件；(D) 顶部主要产品的预测

出S1到S34包含结果的精确显示格式。FF，快速滤波器，用于反应可行性的二元分类；DMF、二甲基甲酰胺；ACN、乙腈。



特别是具有在每次操作前添加流体（反应物、试剂或溶剂）的综合能力。例如，装备这种替代设计来进行两种反应（可以选择五种不同的反应堆类型和尺寸），一种有三种不同配置的分流，7种不同的流体需要77个阀门，每个阀门都出现由于泄漏和堵塞而潜在的故障点（图。S11）

在我们的方法中，机器人机械手通过根据需组装所需的单元操作和试剂线来配置合成设备。工艺堆自动装载工艺模块（如反应器、分离器或包装床），然后工艺槽和连接被气动夹紧，密封流体界面，形成连续的流动路径。试剂通过机器人操作的“交换机”垂直到工艺堆，类似于电话交换机，但有柔性管而不是电线。通过这种方式，我们根据合成方案的要求，将流体泵、出口和废水流连接到每个工艺模块上（图。3B和图。），因此不是

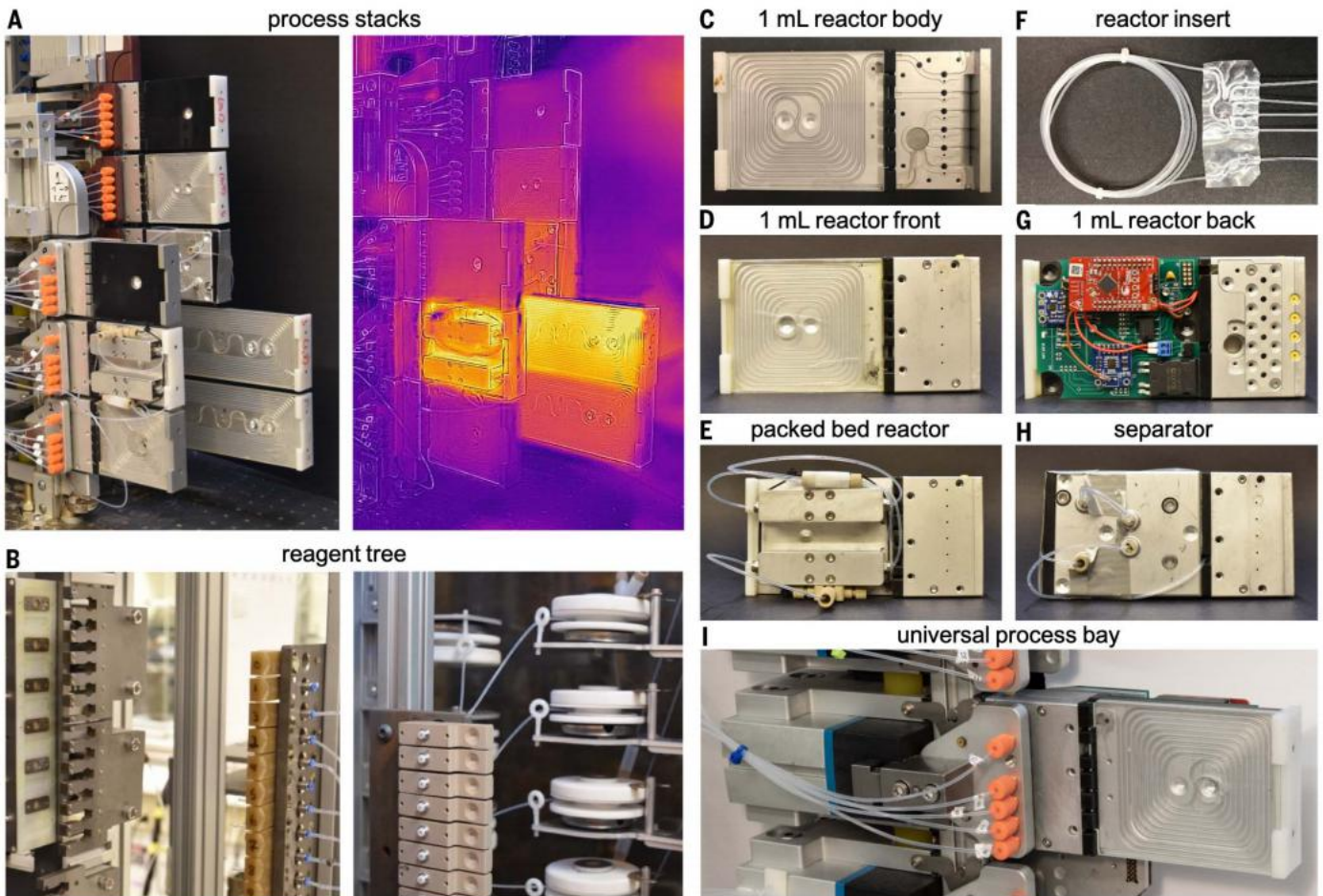
需要复杂的阀门组（有关演示，请参见电影S1到S3）。

机器人将每个流体管道的末端使用磁预加载的运动学接头可逆地耦合到过程模块，机械控制的执行器密封流路径，允许高压操作[每平方英寸计250磅(psig)]。为了避免缠绕，使用安装在试剂树上的动力弹簧预加载的油管卷轴在每个油管上施加恒定的张力。3B和图。S16)。对于流体输送，系统会自动控制连续流动的活塞泵的组合。为了进行库合成，两个泵配备了选择阀，以便从最多24个预装原料中进行选择。

为了支持各种化学转换，我们开发了一个过程模块库，包括不同体积的层流反应器（100毫升到3毫升，图。填充床反应器（1到2毫升，图。3E）能够在环境温度至200° C，压力高达250psig下工作，以及液液萃取的膜分离器单元（41）（图。3H）。

工艺装置放置在由通用工艺槽 (UPB) 组成的两个工艺堆中的一个（图。3I）。UPB为流体、电气和气动过程连接提供密封和对准机构。合成过程中相邻单元之间所需的流体连接是通过按照要求的顺序垂直叠加在塔中，当所有单元到位时，气动滑动它们，以密封线性流体路径。这两个堆栈被设计为允许多达八步线性或五步收敛合成。

为了创建多功能的、一次性的和化学兼容的流体路径，我们开发了一种吹形成工艺，该工艺将百万流体阀和通道集成到全氟烷氧烷(PFA)膜中(图。3F和无花果。S22和S23)。嵌入式允许最多四个流程流连接到每个单元。反应器还具有一个可选的辅助端口，可用于在包装床模块中引入气体或在液液分离器模块中作为二次出口。为了使高温和压力操作升高，膜反应器被an



图中示。3. A在机器人流化学平台上实现的流程和子模块。(A) 模块与UPB接口形成连续流路径（左），显示热反应器的热图像（右）；(B) 反应树和反应歧管前视图；(C) 反应器体前视图图像；(D) 1.0ml反应器工艺模块；(E) 两列包装床反应堆工艺模块；(F) 一次性PFA反应堆插件；(G) 1.0ml反应器背面的集成电子设备；(H) 在线膜分离器；(I) 容纳1.0ml反应器的UPB的特写。

铝制外壳（图中。3C）。每个过程模块都配备了允许进行温度控制的集成电子设备。3G），以及增加新的工艺模块，而不改变或接线新的组件到中央控制系统。

化学合成翻译到机器人平台是通过创建CRF。每个CRF指定必须构建的流体路径；库存解决方案的位置；必须从存储器移动到流程堆栈的流程模块的顺序；以及启动、稳态和关闭流量。根据化学家操作员的输入进行操作之前进行操作和优化。要运行CRF，平台将加载其指定位置的试剂进行合成。然后，用户按下流程开始按钮以执行配方。该程序执行模块存储和工艺堆栈、试剂树和试剂歧管之间的所有路径规划，以组装感兴趣的流程路径。然后，系统遵循CRF启动，设定流量，设定压力和温度，并清洗和拆卸过程。

### 15个小分子的预测和自动合成

该平台的主要应用是按需合成目标小分子有机化合物。在这种背景下，一类重要的化合物是APIs，它在结构和合成的复杂性上差异很大。因此，我们选择了一套15个医学相关的小分子，它们最终需要8种特定的追溯合成路线和9种特定的工艺配置。

虽然所有15个目标都有文献先例，但合成计划程序不能仅仅从记忆中回忆任何合成路线作为精确的匹配；所有的路径都需要通过抽象的转换规则和学习化学反应模式来重新发现。讨论软件功能和更复杂目标的例子可以在补充文本的第1.10节中找到。下面所示示例的原始程序输出也可在补充材料中获得（网站打印输出S1至S34）。每个分子的CRFs包括实现所议转换所需的浓度、流速和过程栈配置的额外规范。这些信息可以在提供足够详细的先例反应数据库，或由专家制作的启发式方法定义时，以数据驱动的方式预测，但考虑到当前数据库中的信息不足，它在这项工作中作为手动任务保存。为了增加的复杂性，我们研究了阿司匹林（1）和外消旋西硝唑（4）的合成；利多卡因（7）和安定（12）连续使用常用原料；（S）-华法林（15）和萨芬酰胺（18）来演示计划的立体化学意识（图。4）。我们还包括在补充材料a

作为软件计划的综合方法，但经过专家评估和筛选后，发现它是翻译为连续流的不良候选品（见补充文本第2.13.1节）。

另外，我们还计划并执行了两个复合物的合成：一个代表血管紧张素转换酶（ACE）抑制剂，包括喹纳普利（24a）、依纳普利（24c）和一个代表非甾体抗炎药（NSAId），包括塞来昔布（28aa）和3种类物（28ab、28ba和28bb）（图。5）。以下各节在其说明的平台功能的上下文中描述这些目标。详细的操作员说明、C关系F、照片、红外照片和每个工艺配置的原理图可以在补充材料中找到。

### 阿司匹林和西硝唑

该平台的自可重构性使它对背靠背的合成特别有吸引力，我们首先使用阿司匹林（1）和塞硝唑（4）进行了演示。

该项目提出了通过将水杨酸（2）与醋酸酐（3）、乙酰氯酸或乙酸乙酰化来一步合成阿司匹林（1）。由于模板相关性神经网络（即，因为它更常用于类似的基质），以及正向预测器更相信酐将产生所需的产品，该程序偏向于醋酸酐。最推荐的反应条件在108°C下使用整洁的反应剂（无溶剂），不需要其他试剂或催化剂，或在较温和的温度下使用硫酸。虽然反应可以并且经常批量运行，为了降低堵塞风险，我们制备2乙酸乙酯（EtOAc），这也使下游分离产生91%（1.72g/h的阿司匹林）。

经过几次稳态运行后，将供给流改道为纯溶剂，降低反应器的温度设定点，通过净化背压调节器对流路进行减压。然后，机械臂用塞硝唑所需的工艺模块取代了阿司匹林的工艺模块，修复了该系统，并开始运行。

### 塞硝唑的一步合成方法

（4）是在47°C下通过2-甲基-5-硝基咪唑（5）催化打开环氧丙烷（6）；排名较低的建议包括1-溴-2-丙烷醇或类似条件下的氯烷基化等价物。环氧化物开口在140°C[选择二甲基甲酰胺（DMF）和过量环氧化物（未优化摩尔比3:1以增加反应速率）使用三乙胺（TEA），作为连续合成氢氧化钠的标准替代品，提供95%产量的塞硝唑

（792mg/h）使用一个3.0ml的反应器和一个分离器模块。

### 背靠背合成利多卡因和安定

我们利用利多卡因（7）和安定（12）等机器人的可重构性，继续探索利多卡因的背靠背合成技术。在提议的路线中，有两个带有优先的反应步骤，如图所示。4A。虽然程序在迈向利多卡因的第一步对氯乙酰氯非常偏好，但我们选择使用溴乙酰氯（11）

来利用常见的试剂流（aj，对应于整洁11），因为正向预测器对溴基完成闭环步骤的可能性更有信心。该程序提出了当使用等效的氯13时，从13到12的闭环胺的合适条件（氢），但提出了一个不太合理的钠或氮化锂本身，这可能需要额外的还原步骤。商用的化合物10和11容易反应形成中间体9，在基、热和二乙胺（8）存在的情况下提供7/77%的产率（1.09g/h）。操作6小时后，将所有试剂流切换为纯N-甲基吡咯烷酮（NMP），并将温度设定点设置为零。将NMP的反应溶剂作为MeOH、二氯甲烷（DCM）和丙腈（ACN）之间的妥协反应步骤，作为多用途高沸点的极性无质子溶剂。在重建工艺堆栈之前，机器人将系统减压和分解，将第二个反应堆从3.0换到1.0ml大小。使用相同的流体流aj和ak，配置了安定的流路径。分离器后恢复75%（638mg/小时）。

### （S）-华法林和沙芬酰胺的立体选择性和立体反应合成

为了增加我们的目标的复杂性，然后我们研究了（S）-华法林（15），它可以通过乙酰蒽醌（16）和4-羟基香豆素（17）通过不对称的迈克尔添加在一个步骤中产生。该程序成功地识别了这种断开，认识到这个迈克尔添加可以立体选择性地执行。背景推荐人提出了手性胺催化剂（8S、9S）-6'-甲氧基金鸡南-9胺。尽管根据之前的文献，该催化剂可能合适，但我们替换了更容易的（1S，1S，2S）-（+）-1,2-二苯乙二胺[（S，S）-DPEN]，因为其成本超过\$500/g；虽然在计划期间考虑起始材料的成本，但催化剂的成本并非因为它们通常用于亚化学计量量。在迈克尔添加之前，使用16和混合丙酮（S，S）-DPEN，只有在手动筛选后才能通过经验确定；这两个步骤都是在高于推荐的温度下运行的（50°C对20到30°C），以降低



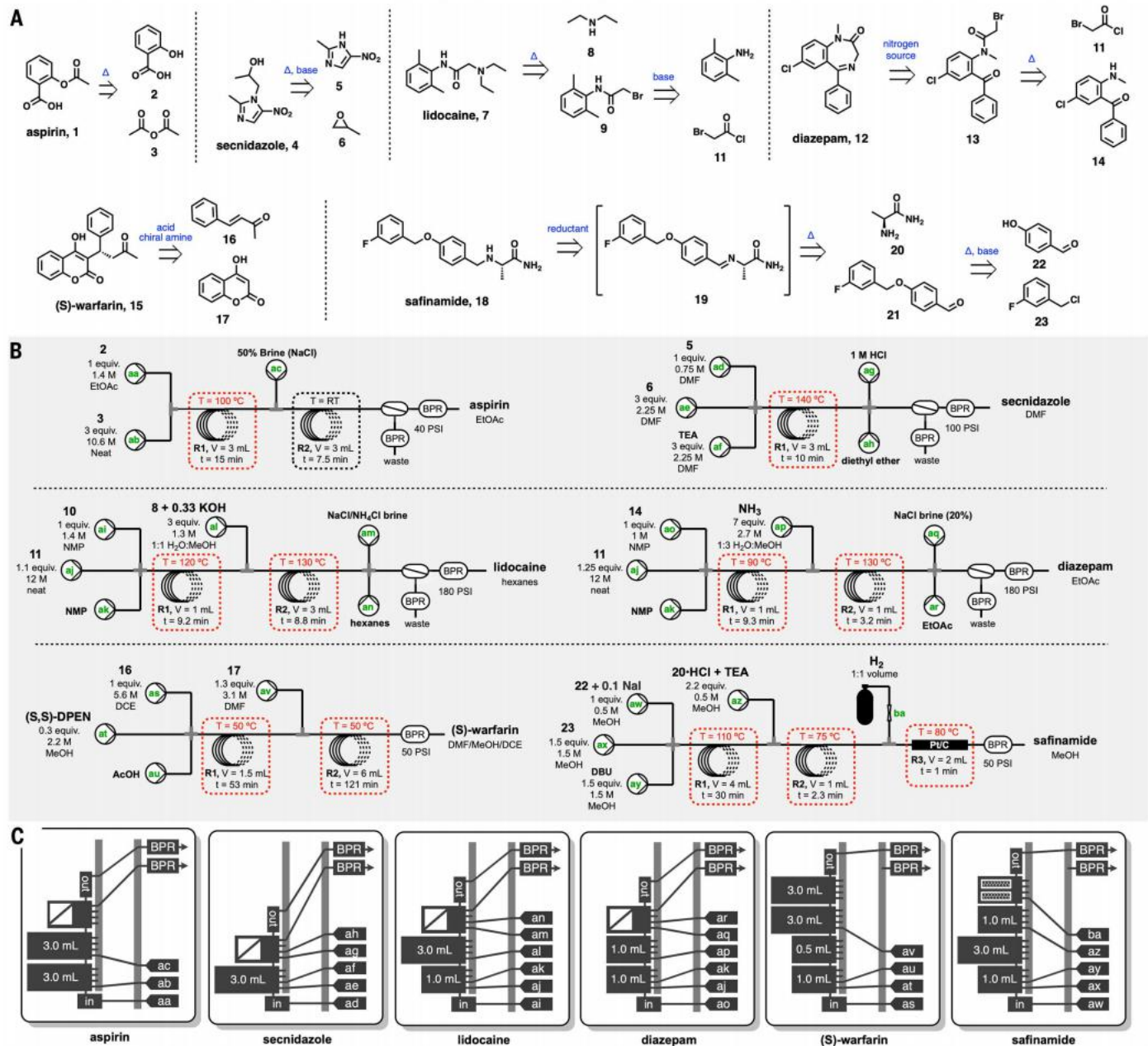


图4. 六种实例药物物质的合成规划和执行：阿司匹林4、秒硝唑4、利多卡因7、安定12、(S)-华法林15和沙芬酰胺18。(A) 综合规划程序提出的合成路线，包括正向反应的条件（蓝色）；(B) 拟定路线的连续流实现；(C) 机器人配置的流路径和装置在模块化流化学平台上执行这六种合成的操作。塞硝唑和阿司匹林被用回

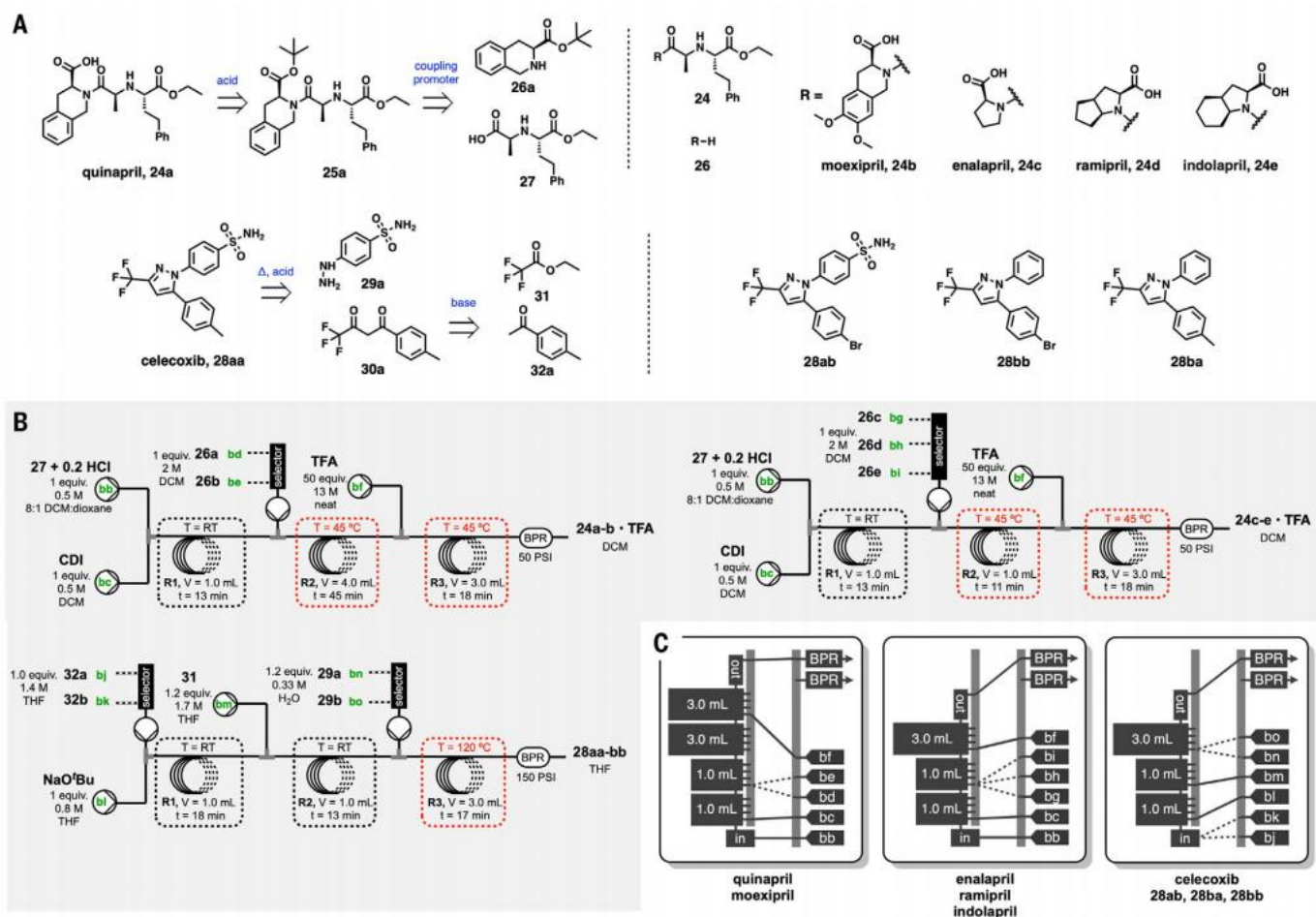
返回。利多卡因和安定同样运行，进一步利用共同前体11和相应的试剂流aj(S)-华法林和萨芬酰胺，以证明成功识别立体选择和立体反应路线。特定的流体流将按字母顺序标记（绿色）。DMF，二甲基甲酰胺；TEA，三乙胺；NMP，N-甲基吡咯烷酮；(S,S)-DPEN、(1S,2S)-(+)-1,2-二苯乙二胺；DBU，1,8-二氮双酚[5.4.0]。

以牺牲对映体选择性的停留时间为代价。最后一条路线的78%的产量为730毫克/小时，4.1:1对映体比。

通过利用1-丙氨酸(20)的合成规划程序，确定了手性药物物质亚苯酰胺(18)的立体反应途径。提出了醛21与20的还原胺化的单步反应或

等价的两步反应与对应的亚胺19作为显式中间体。虽然21在我们的可购买化合物\$33/g，我们施加了\$10/g的最高价格，以迫使递归膨胀进一步扩展到氯化物23和苯酚22作为更简单、更便宜的起始材料。这种以代化和随后的还原性的建议条件

[DBU(1,8-二氮双酚[5.4.0]undec-7-ene)是一种常见的有机、可溶性的K替代品。它不会产生气态的副产品]。在甲醇中完成亚氨酸还原<sub>2</sub>在一个加热的Pt/C包装床反应堆模块(排名第六的条件，而不是使用硼氢化钠或氰化钠作为他们



图中示。5. **A**基于奎那普利（24a）和塞来昔布（28aa）的两个复合文库的合成规划和执行。（A）综合计划程序提出的合成路线，包括正向反应条件（蓝色）；（B）使用选择阀动态改变试剂流实现连续流；（C）在模块流化学平台上执行这些合成的机器人配置的流路和单元操作。特定的流体流将按字母顺序标记（绿色）。DCM，二氯甲烷；CDI，1,1'-碳化二咪唑；TFA，三氟乙酸；THF，四氢呋喃。

高度吸湿，容易淬火），产生18/32%的产量（265毫克/小时）。

#### ACE抑制剂文库的合成规划与执行

流体连接的自动重新路线和两个24路选择阀允许使用几十种不同的原料。为了展示这一能力，我们准备了5个基于奎那普利（24a）的ACE抑制剂库和通过次级胺反应物26的变异可获得的衍生物。

奎那普利（24a）具有三个四面体中心，该程序在可购买的前体中识别。激活1,1'-碳化二咪唑（CDI），使氨基酸衍生物27与26a耦合产生‘推荐布保护奎那普利（25a）、酞磷酸二乙酯和N,N'-二甲基碳二亚胺（DCC）作为偶联试剂，但由于担心急性毒性而避免。根据程序的建议，在三氟乙酸的情况下很容易裂解

（tfa）。该项目还提出了一种与相当于26a的羧酸的单步酰胺耦合；然而，这一途径被正向预测器标记为潜在的问题，因为预测二次胺的相互竞争，尽管预期产物占主导地位。

初步筛查显示，奎那普利和莫西普利（24b）有利于延长停留时间，达到26至去保护前27之间的初始耦合；依那普利（24c）、雷米普利（24d）和內多拉普利（24e）进展更快，允许使用较小的反应器。

我们准备了两个CRFs，包括到这一对和三个目标的路线。在室温下安装了一个1.0ml的小反应器，允许足够的时间用CDI激活1.0ml反应器，然后将26引入随后的1.0或4.0ml加热反应模块。通过24端口选择阀在前体之间切换。收集粗奎那普利、莫西普利、依那普利、雷米普利和因多拉普利作为他们的

TFA盐和线下分离物，分别为70%、50%、58%、58%、59%和66%的产量（459、369、342、390和420）分别为mg/h）。整个图书馆总共运行了68个小时。

#### NSAID库的综合规划和执行

接下来，我们准备了一个基于塞来昔布（28aa）的二维非甾体抗炎药化合物库。塞来昔布的特点是一种内部核心，该软件提出通过由亲丙酮32a和三氟乙酸乙酯（31）制备的肼29a和二酮30a的缩合来合成。建议使用碱加酸和加热与文献先例是一致的。通过将肼与苯肼（29b）和/或4-甲基乙丙苯与4-溴乙酞苯（32b）交换，制备另外三种塞来昔布类似物（28ab、28bb、28ba）。

塞来昔布的CRF被写成使用两个室温反应器（每个1.0ml），以实现适当的添加顺序（即，生成



乙酰从32a开始前与31进行反应)。通过将试剂线从32a切换到32b, 然后从29a切换到29b和32b切换回32a, 按顺序合成这四个靶点。塞来昔布和28ab、28bb和28ba的产量分别为91%、59%、79%和86%(572、417478和分别为432mg/h)。整个图书馆总共运行了28个小时。

### 《市场展望》

通过集成CASP和机器人流动化学平台, 我们实现了人工智能计划的合成路线到小分子目标。CRF作为两者之间的中间体, 记录专家化学家指定的完整工艺细节。

批处理合成的近似条件可以根据文献生成, 正如我们在本研究所做的那样, 但它们在流中的直接实现是具有挑战性的。渴望过程强化(例如, 减少反应时间), 需要减少固体形成以避免堵塞, 以及伸缩多个单元操作的重要性, 需要偏离批次条件, 以及流动化学尚未实现的预测的置信度。对非水溶剂和非环境温度下的溶度的计算预测仍然难以计算。预测合适的纯化程序是一个普遍的挑战, 而不仅仅是对流动化学, 特别是在使用非色谱方法时。要开发到新目标分子的途径, 粗产品中物种的识别和量化也需要自动化。如果给有一组合理的起始条件, 对于单步反应和双步反应将非常简单, 就像之前使用其他流化学平台(7, 23)所演示的那样。然而, 由于参数从早期步骤到后期反应的传播变化, 多步骤反应的优化将会更加复杂。虽然这些路由目前已从脱机过程开发中获益, 但这在很大程度上是对公共领域存在的数据的限制, 以及我们对利用流化学技术的好处的兴趣。随着时间的推移, 这个和类似的自动化实验平台产生的结果可能会消除我们对历史反应数据的依赖, 特别是与更小规模的流筛选平台(11, 42)相结合。增加反应数据的可用性将进一步使基于人工智能建议的机器人实现的合成成为可能, 缓解专家化学家的手动任务, 使他们可以专注于新的想法。

### 方法总结

在补充材料中提供了对合成规划软件、机器人平台、库存溶液制备和化学分析的更详细的描述。以下是用户工作流程的摘要。

### 配方的生成

从基于网络的图中得到了建议的合成路径和反应条件

针对每个目标分子的ASKCOS程序的主要用户界面。我们优先考虑较短的合成路线, 因为需要许多反应步骤才能使流程中的实现复杂化。反应条件通过非编码的流化学直觉, 适应了蝙蝠产生的建议。本文中描述了每个目标的变化, 通常包括使用可溶性有机碱而不是无机碱、普通溶剂和温度升高来提高反应速率。手动筛选条件, 以确定适当的浓度和近似的居住时间; 未来的工作应该是自动化这一过程。填充电子表格文件(补充材料中的“配方计划文件S1”), 并通过编程方式扩展为每个目标的完整CRF(“配方计划脚本S1”(补充材料中的“配方S1到S7”)。

### 平台操作

操作员根据CRF定义的浓度制备库存溶液, 并按指定的试剂线号加载到平台上。试剂溶液和纯溶剂副本(作为冲洗溶液)放置于5psig氩大气下。CRF被加载到平台控制图形用户界面中并执行。机械臂将所需的单元操作放置在工艺堆栈中, 并在试剂树中进行必要的流体连接。在使用冲洗溶液自动启动和加压过程中, 检查系统作为安全措施是否泄漏。流体管线自动从冲洗溶液改为试剂溶液, 达到温度设定点, 开始分数采集。在运行CRF中规定的持续时间后或产品流的数量显示合理的转换和/或产量后, 流体管换回冲洗溶液, 使系统开始清洗和减压。经过一个清洗循环后, 机械臂自动返回模块到存储中, 并断开试剂树上的流体线。

### 参考文献和注释

1. D. A. C. 布莱克莫尔等人, 有机合成改变药物发现提供了机会。不。化学材料。10, 383–394 (2018)。部门: 10.1038/s41557-018-0021-z; 页码: 29568051
2. G. 施耐德, 自动药物发现。不。修订版。药物迪斯科舞厅。17, 97–113 (2018)。产品目录: 10.1038/nrd.2017.232; 页码: 29242609
3. D. A. P. Tabor等人, 加快在智能自动化时代发现清洁能源材料。不。修订版。脱线。3, 5–20 (2018)。部门: 10.1038/s41578-018-0005-z
4. M. 《有机合成物: 机器人化学家》。自然》512, 20–22 (2014)。多部门: 10.1038/512020a; pmid: 25100466
5. S. V. 李伊, 医学博士。E. A. 菲茨帕特里克, R. J. 英格汉姆, R. M. 迈尔斯, 有机合成: 机器人的三月。使用的角度。化学材料。内部。Ed. 发动机总成。54, 3449–3464 (2015)。美国国防部长: 10.1002/anie.201410744; 页码: 25586940
6. A. A. G. A. 戈弗雷先生。马斯奎林, H. 海默尔, 远程控制自适应医学化学实验室: 21世纪药物发现的创新方法。药物迪斯科舞厅。今天。18795–802 (2013)。国防部: 10.1016/j.drudis.2013.03.001; pmid: 23523957
7. A. A. C. A. Bedard等人, 可重构系统, 自动优化的各种化学反应。科学361, 1220–1225 (2018)。部门: 10.1126/science.aat0650; 页码: 30237351

8. S. 斯坦纳等人, 由化学编程语言驱动模块化机器人系统中的有机合成。科学363, (2019)。部门: 10.1126/science.aav2211
9. J. 李等人, 用一种自动化工艺合成许多不同类型的有机小分子。科学347, 1221–1226 (2015)。部门: 10.1126/science.aaa5414; 型号: 25766227
10. A. A. 桑塔尼拉等人, 有机化学。纳米脂钙高通量化学的合成复杂分子。科学347, 49–53 (2015)。剂量剂量: 10.1126/科学.1259203; 邮编: 25554781
11. D. A. Perera等, 一个自动化纳米摩尔级反应筛选和微摩尔级合成的平台。科学359, 429–434 (2018)。多部门: 10.1126/science.aap9112; pmid: 29371464
12. 食品和药物管理局, 预防战略计划和减轻药物短缺 (2013); <https://www.fda.gov/downloads/Drugs/DrugSafety/DrugShortages/UCM372566.pdf>.
13. R. L. 哈特曼, J. P. 麦克马伦, K. F. A. 延森, 决定是否采用流: 评估流反应器的优点。使用的角度。化学材料。内部。Ed. 发动机总成。50, 7502–7519 (2011)。美国国防部长: 10.1002/anie.201004637; 页码: 21710673
14. M. B. A. 普鲁沙克, B. 皮伯尔, K. 吉尔摩, P. H. 西伯格, 厨房的流动化学指南。化学材料。修订版。117, 11796–11893 (2017)。多部门: 10.1021/acs.chemrev.7b00183; pmid: 28570059
15. K. F. A. 流动化学-微反应技术来了的年龄。附件J。63, 858–869 (2017)。美国国防部长: 10.1002/aic.15642
16. B. A. 古特曼博士, 医学博士。康蒂, 罗。O. 连续流动技术: 一种安全制造活性药物成分的工具。使用的角度。化学材料。内部。Ed. 发动机总成。54, 6688–6728 (2015)。美国国防部长: 10.1002/anie.201409318; 页码: 25989203
17. C. A. 怀尔斯, P. 瓦茨, 连续流反应器: 一个视角。绿色的奶酪。14, 38–54 (2012)。部门: 10.1039/C1GC16022B
18. D. R. Snead, T. F. Jamison, 三分钟布洛芬的合成和净化: 推动连续流处理的极限。使用的角度。化学材料。内部。Ed. 发动机总成。54, 983–987 (2015)。美国国防部长: 10.1002/anie.201409093; 页码: 25470733
19. T. Klucznik等人, 多样性的有效综合物, 医学上的由计算机计划和在实验室中执行的相关目标。奶酪4, 522–532 (2018)。使用剂量: 10.1016/j.chempr.2018.02.002
20. S. Szymkuc等人, 计算机辅助综合规划: 开始结束。使用的角度。化学材料。内部。Ed. 发动机总成。55, 5904–5937 (2016)。美国国防部长: 10.1002/anie.201506101; 页码: 27062365
21. M. H. S. 塞格勒, 美国石油天然气股份有限公司, M. 前, M. P. 用深度神经网络和符号人工智能来规划化学合成。自然555, 604–610 (2018)。部门: 10.1038/nature25978; 页码: 29595767
22. A. A. Adamo等人, 在一个紧凑、可重构的系统中按需连续流生产药品。科学352, 61–67 (2016)。部门: 10.1126/science.aaf1337; 页码: 27034366
23. B. A. J. 雷兹曼, K. F. A. 延森, 关于加速反应开发的流动反馈。执行。化学材料。平均分率。49, 1786–1796 (2016)。多部门: 10.1021/acs.accounts.6b00261; pmid: 27525813
24. D. A. E. A. 菲茨帕特里克, C. 巴蒂洛奇奥, S. V. 一种新型的基于互联网的化学合成反应监测、控制和自动优化平台。组织结构。工艺记录。开发。20, 386–394 (2016)。部门: 10.1021/acs.oprd.5b00313
25. E. J. Corey, W. L. 乔根森, 计算机辅助合成设计: 基于附属物的合成策略和重连接变换的使用。J. 上午。化学材料。大豆含量。98, 189–203 (1976)。部门: 10.1021/ja00417a030
26. A. A. 库克等人, 计算机辅助合成设计: 40年过去了。威利的相互作用。修订版。投入。月。科学研究。2, 79–107 (2012)。美国国防部长: 10.1002/wcms.61
27. C. W. Coley, R. 巴齐莱, T. S. Jaakkola, W. H. 绿色, K. F. Jensen, 使用机器学习预测有机反应结果。ACS百美分。科学研究。3, 434–443 (2017)。美国国防部长: 10.1021/acscentsci.7b00064; pmid: 28573205
28. M. H. S. 塞格勒, 美国石油天然气股份有限公司, M. P. 沃勒, 神经网络机器学习的回顾合成和反应预测。化学, 23, 5966–5971 (2017)。doi: 10.1002/化学

产品。201605499；页密德：28134452

29. B. 刘等，神经回顾反应预测  
序列对序列的模型。ACS百美分。科学研究。  
3,1103 - 1113 (2017)。美国国防部长：  
10.1021/acscentsci.7b00303；页码：29104927

30. C. 科利，西州。H. 绿色，K.。F. 计算机辅助  
合成规划中的机器学习，教授。执行。化学材料。  
平均分辨率。51,1281 - 1289 (2018)。多部门：  
10.1021/acs.accounts.8b00087；pmid: 29715002