

合成化学中的逻辑推演理论：科里的
逆向思维与计算机辅助*朱 晶¹ 姜雪峰^{2**}

(1. 华东师范大学哲学系 上海 200241;

2. 华东师范大学化学与分子工程学院 上海 200062)

摘要 理论、实验和计算机技术之间的关系，是科学哲学研究的核心主题之一，也是科学教育研究的关注焦点。已有的研究多从物理学或生物学中的案例来思考三者之间的关系，少有从化学实践出发进行的专门考察。以 1990 年诺贝尔化学奖得主科里 (Elias James Corey) 的研究与教学工作为例，分析了科里在合成研究过程中使用的逆向思维，考察了他提出的计算机辅助合成技术尚未受到和逆合成分析同等关注的原因。本文提出，化学家借助计算机技术将合成规则和方法抽象化、概念化的基础，是化学家在教学和实践过程中对合成规则的可操作性和合理性的经验与实验检验，也是合成化学不同于其他科学领域的独特研究风格。计算技术如何与化学家的经验和合成理论深度交互，做出基于机理的因果解释和有效推断，是交叉科学将要进一步努力的方向。

关键词 化学哲学 逆合成分析 人工智能 交叉科学 化学思维

DOI: 10.13884/j.1003-3807hxjy.2022110122

20 世纪中后期，化学家合成了大量结构复杂的分子。这些合成有规律可循吗？有机合成领域独有的理论和方法是如何发展出来的？合成化学是否有区别于其他学科的思维方式？计算机辅助技术在化学研究中的位置是什么？这些问题吸引着科学家、科学教育研究者、智能技术研究者的兴趣，也是科学史与科学哲学领域的学者关注的主题。已有的研究多从物理学或生物学中的案例出发，来考察理论、实验和计算机技术三者之间的关系，较少从具体的化学研究和教育实践来对这些问题进行专门考察^[1]。1990 年诺贝尔化学奖得主科里 (Elias James Corey, 1928—) 的研究与教学，对化学哲学和化学教育的研究和实践均具有重要的启发意义。

瑞典皇家科学院在给科里的诺贝尔化学奖授奖词中这样说：“可能没有哪一位化学家能够像科里那样开创如此广泛而多样的合成方法，这些方法显示出独创性和简洁性，在有机合成实验室里被普遍使用。”^[2] 纵观科里的研究生涯，他不仅在合成天然产物方面做出了重要贡献，突破了多类结构复杂的分子在合成上的限制，包括萜类、甾体、生物碱、大环内酯等，比如赤霉素和银杏内酯 B。更重要的

是，科里研究小组创建了独特的有机合成理论——逆合成分析 (retrosynthetic analysis)，令有机合成方案系统化且逻辑可推，使得种类繁多具有高度生物活性的复杂天然产物合成变为可能，并成为教学与实践合成设计的一种标准方法^[3]。在此基础上，他编制了第一个计算机辅助有机合成路线的设计程序 LHASA (Logic and Heuristics Applied to Synthetic Analyses)。值得注意的是，LHASA 并没有和逆合成分析方法一样被广泛用于合成，但是逆合成分析背后的推理规则却开拓了此后计算机辅助有机合成的新领域。当计算模拟成为除了理论和实验之外、科学家做研究的第三种方式的今天，笔者从科学史上的具体案例出发，用科学哲学的理论框架分析科里提出的逆合成分析与计算机辅助有机合成方法，有益于我们深入思考化学思维方式，以及如何促进理论、实验和计算机技术三者之间的深度交互。

1 逆合成分析中的逆向思维

复杂有机物的合成常常被认为是一门艺术抑或如下棋^[4]，合成工作不但要精心设计合成路线，还需要对试剂及反应条件进行缜密选择。早期的化学家合成复杂有机物一般从直觉出发，对反应初始原

* 国家自然科学基金 (22125103, 21971065); 上海市基础研究领域项目 (22JC1401000, 20XD1421500); 华东师范大学人文社会科学青年跨学科学术创新团队项目 (2022QKT003)

** 通信联系人, E-mail: xfjiang@chem.ecnu.edu.cn

料、路线的选择没有特定的原理可循,而合成过程中会不断遇到新问题,化学家会不断改变合成路径。这意味着,复杂有机物的合成需要化学家有很强的灵活应用能力和创造性。

科里为了突破复杂化合物在合成方法上的困境,开始探索具有概括性、更能有效解决合成问题的思维方式。在他看来,面对具有复杂结构的分子,化学家可以同时考虑演绎相关(inductive associative)的思想或者逻辑引导(logic-guided)的思想^[5],随着合成化学遇到的问题难度逐渐增加,相比早期较为简单的合成程序,逻辑引导的思维方式显得更加重要。特别是,20世纪50年代,化学家解析出大量新颖而独特的天然产物结构,而合成这些天然产物却很难寻找到合适的起始原料或者合成模块^[6]。科里意识到解决化学问题是来自于一种自动化的“知其所以然”(know how)的过程,而不是刻意使用已经设计好的程序^[3],也就是说,应该在理解相应规则的前提下进行合成。1957年秋,他提出了一种截然不同的化学合成路线设计方法——逆合成分析原理,并提出了合成子(synthon)和切断(disconnection)2个基本概念。

逆合成分析的基本原则是:合成应该是多步的^[5],需要首先确定如何将目标分子按照可再结合的原则进行切断,将化学键进行切断时所产生的分子碎片称为合成子或合成元。切断时,确定如何将目标分子分割成更小的部分以及应当打破哪一个化学键,这个步骤极其重要。合成子应该是合理的、较简单的、较易得到的、较小的起始底物,可以是市面上可获得的简单化合物或者是可通过参考已有文献来合成的化合物。化学家通过“逆合成分析”

进行合成设计,可以得到多种可供选择的合成方法,然后根据自己的判断选择路线并展开合成。在这个过程中,化学家从目标产物逆向推导到更简单的结构,这种形式化的逻辑推导过程是与化学反应过程相反的。这种反复迭代的过程,使得有机合成不再只凭化学家的直觉,而是有据可循,变得有规则。通过这种逆向推理,化学家可以寻找到消除目标产物复杂性的方法,使合成工作变得更具有操作性(图1)。

然而,将这个看起来简单的原则应用于实际的合成之中,并非易事。逆合成分析这种方法的难度在于2个方面。首先,合成子是一个抽象化的概念,并不等同于实际存在的自由基、离子和分子,它可能是不稳定的瞬时的活性中间体碳正离子或者自由基,也可能并不存在。其次,切断可以有多种方式,选择哪一种需要依赖于化学家已有的合成经验。比如,单一的一个化学键A—B至少有3种切断方式(图2)。按照科里的原则,逆合成分析要求化学家一开始就应考虑目标化合物的所有特征,包括直链、环、支链、官能团、不对称中心以及每个基团的反应特异性、敏感性及稳定性。此外,在考虑合成目标化合物的前体时,必须降低分子的复杂性,包括断链、开环、移去支链、消除官能团和不对称中心,还要修饰化合物中具有高反应性或或不稳定性的基团。也就是说,仅仅有这种反向推理的逆向思维还不够,还需要有可供参考的具体的推理规则以及如何按照这种规则进行合成的成功经验来证实其合理性。接下来,笔者将从科里的教学与研究实践出发,分析化学家的经验是如何推动逆合成分析被普遍使用并成为一种可操作的原则。

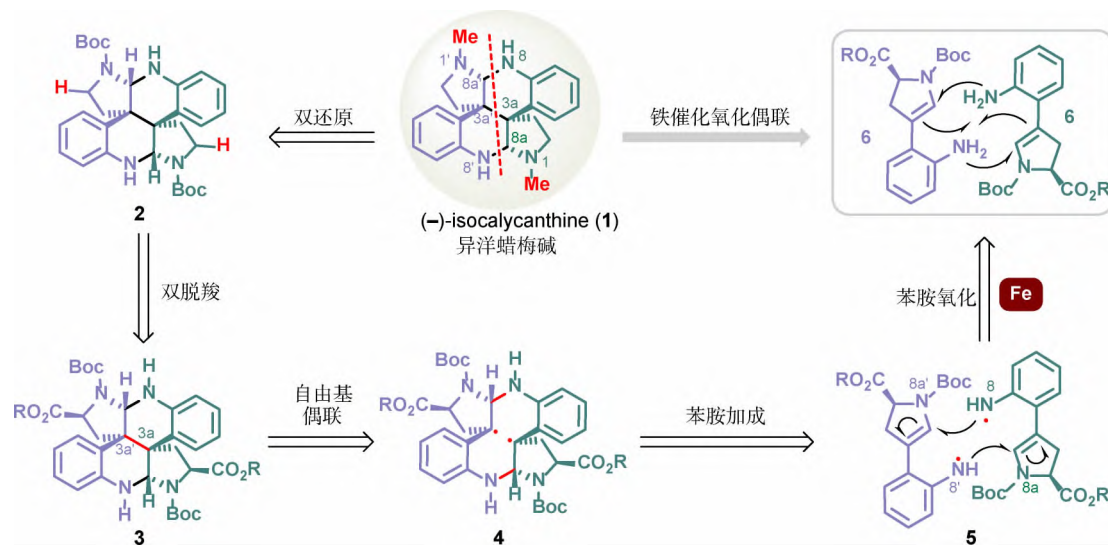
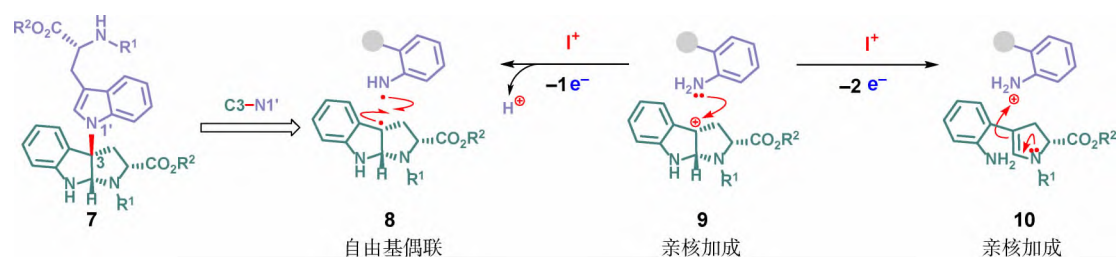


Fig 1 Total synthesis of (-)-isocalycanthine: an example to show the logic and operational process of retrosynthetic analysis^[7]

图1 以异洋蜡梅碱为例揭示逆合成的逻辑性和可操作性^[7]

Fig. 2 Three ways for disconnecting the C—N bonds^[8]图 2 以 C—N 键为例展示 3 种切断方式^[8]

2 合成原则的可操作性与经验基础

新近的科学哲学研究认为,科学家要实现或者通达对彼此研究目标的理解,需要相应的技能^[9]。杰出的化学家在发展出一系列新的抽象原则用于研究时,他们也会考虑原则的可理解性,会思考新的科学理论或者原则提出后,如何通达其他科学家的理解。这种做法可以是将抽象的理论形象化、简单化,也可以是通过教学和自己的研究成果来帮助其他人理解。比如,化学家鲍林(Linus Pauling)的共振论提出之初,化学家对量子力学还较为陌生,他通过化学家熟悉的概念和方式来描述和表征新概念,并通过教学、公开演讲和出版物来促进共振论被其他科学家理解和使用^[10]。

在科里之前,有许多化学家的合成方法符合逆合成分析的规则,但是他们没有将实际操作经验总结为符合逻辑推理的原理,并加以推广。科里不仅提出了逆合成分析方法,还在哈佛大学开设了一门新的合成课程,革新了合成有机化学的讲授方式。在此之前,大学有机合成课程的讲授往往是通过呈现一系列真实的合成案例,而这些案例之间并没有联系。化学研究人员通过学习这些单个的分散案例来了解有机合成方法^[5],结果是,他们在解决有机合成问题时只能靠直觉,而缺乏普遍可应用的规则。意识到这一点,科里在 1960 年秋季学期为研究生开设了新的合成课程。据他回忆,这门课程一开始就收到非常好的效果。这门课程的目标被延续下来,即教授学生设计复杂分子合成路线的基本技巧、在课程中评述当时最前沿和最重要的合成。科里开设的课程逐渐形成了 3 个主题:解决合成问题的基本策略、化学转化的分类、目标导向合成的文献案例。在这种训练之下,课程培养学生形成一种分类化、可拓展的合成工具箱,从而可以在合成研究中将逆合成分析策略运用于解决合成的问题^[11]。

除了教学,科里团队自身的工作证实了逆合成分析策略的有效性。利用逆合成分析法,科里团队合成了 100 余种具有生物活性的天然产物,将多次实验总结出的合成方法系统化,使之成为可以供借

鉴和应用的合理方法。比如切断的具体原则包括,一般在目标分子中有官能团的地方进行,或者在有支链的地方进行切断等。他强调逆合成分析是决定合成路线的关键,它关系到整个合成的策略、成败和评价,包括合成过程中各种官能团的转变、引入等一系列系统修饰分子的原则和方法,如转换策略、拓扑策略、立体化学策略等^[3]。在获得诺贝尔奖之后,科里等人又合成了一系列手性催化剂,并将其应用于不对称合成,取得了成功。在不对称催化领域,科里设计的化学酶在药物和天然产物合成中显示出强大的实力,给不对称化学提供了极为有用的模型。他系统应用各种类型的有机金属试剂,并采用多种类似酶的催化剂,使得合成反应中只生成目标产物的特定光学异构体。1991 年,科里报道了一类用于 D-A 反应的化学酶,研究发现,反应过程中的过渡态配合物内,静电吸引与空间排斥同时存在,共同作用^[12]。1996 年,科里等发展了另一类型的化学酶,扩展了 D-A 反应化学酶的应用范围,使反应活性很低的 1,3-环己二烯、丁二烯、取代丁二烯也能用于不对称合成^[13]。这些工作在他研究生涯中发表的一千多篇科学论文和综述性文章^[14]中得到了清晰的呈现。

可见,一个合理的合成策略要启发后续的合成工作,除了要对目标产物进行结构分析,识别出目标产物的特征,判断哪些合成子是容易合成的,进而识别出一系列可能得到这些难度大的特征化合物的反应,还需要在这些方法中做出选择,整个过程实际上依然是相当复杂的。因为每一步依然会涉及到对合成步骤的有效性的评估,而此时有机化学家必须使用他所知道反应机理和已有的化合物的结构,才能做出这种推测。这意味着,化学家的合成经验与理论有机化学其实是交织在一起的。接下来笔者将进一步分析,这种极大地依赖于经验的复杂推理过程,是否能够被计算机技术所推动。

3 “失败”的计算机辅助合成与模型概念化

仅有“逆合成”的逆向思维,还不足以让其成为有效的甚至是标准的合成思路,如前文所述,将

复杂的目标产物切断为合成子,可以有多种方法。将一个化学键 A—B 切断,至少有 3 种方法,但对于有多个化学键的目标产物而言(图 3),这种切断组合的可能性就变得非常多,单凭经验等很难全面而系统地推出各种可能的组合。科里开始尝试将具有严格逻辑性的逆合成分析法与计算机技术相结合,是首位将人工智能应用于设计化学合成路线的科学家。

1967 年,科里等人编写了第一个有机化学合成模拟程序 OCSS (Organic Chemical Simulation of Synthesis)。根据该程序,研究者可以直接得出复杂目标分子合成的可能途径,而不必输入起始反应原料或中间体。计算机设计合成路线与实验室的实验过程正好相反:实验室合成是从原料到产品,而计算机设计则是从产品寻找原料。此后,用计算机来设计有机合成路线开始迅速发展起来。科里还研制成功了 LHASA (逻辑与启发应用于合成分析) 计算机程序,利用它可以得出前体结构树,通过输入图形和辨认,可以获得合成策略,并对反应进行选择 and 评价。虽然科里提出的程序在今天看来远不成熟,但科里程序化设计合成的思想,已经可以和今天的化学家取得的合成成就的逻辑性相匹敌了。

值得我们进一步思考的是,逆合成分析在今天

已经成为合成化学家的常规分析方法,而计算机辅助人工合成,却没有像逆合成分析方法一样,成为广泛用于合成实践的工具。换言之,LHASA 计算程序设计背后的规则和原理,即逆合成分析方法,被化学家广泛使用,而 LHASA 计算程序却并未受到广泛好评。这个现象需要我们重新思考计算机辅助设计与化学家人工合成工作之间的关系,以及化学家使用逆向推理逻辑规则的能力和机器推演预测的能力之间的关系。科里和他的团队在编写 LHASA 计算程序的过程中,程序编写的要求促使他们更加清晰地表达和阐释已有的合成工作,反过来促进了他们对有机合成方法和原理的理解^[15]。科里研发的 LHASA 计算程序虽然没有获得广泛使用,但是编写这些程序需要将底层逻辑和理论概念化与抽象化,这个要求反过来推进了科里对合成设计与如何真正达到合成目标之间的关系进行深入思考。

另外,科里团队开发 LHASA 计算程序所处的时代是 20 世纪中期,无论是计算能力还是存储能力,远没有达到今天的高度,更谈不上模拟人类思维的能力。从科学史的脉络来看,科里团队的观点无疑在当时是相当具有前瞻性的。逻辑分析针对的是在真实的合成化学实践中呈现出的关系,不仅仅是简单的专家与机器智能的高下问题,而是不同

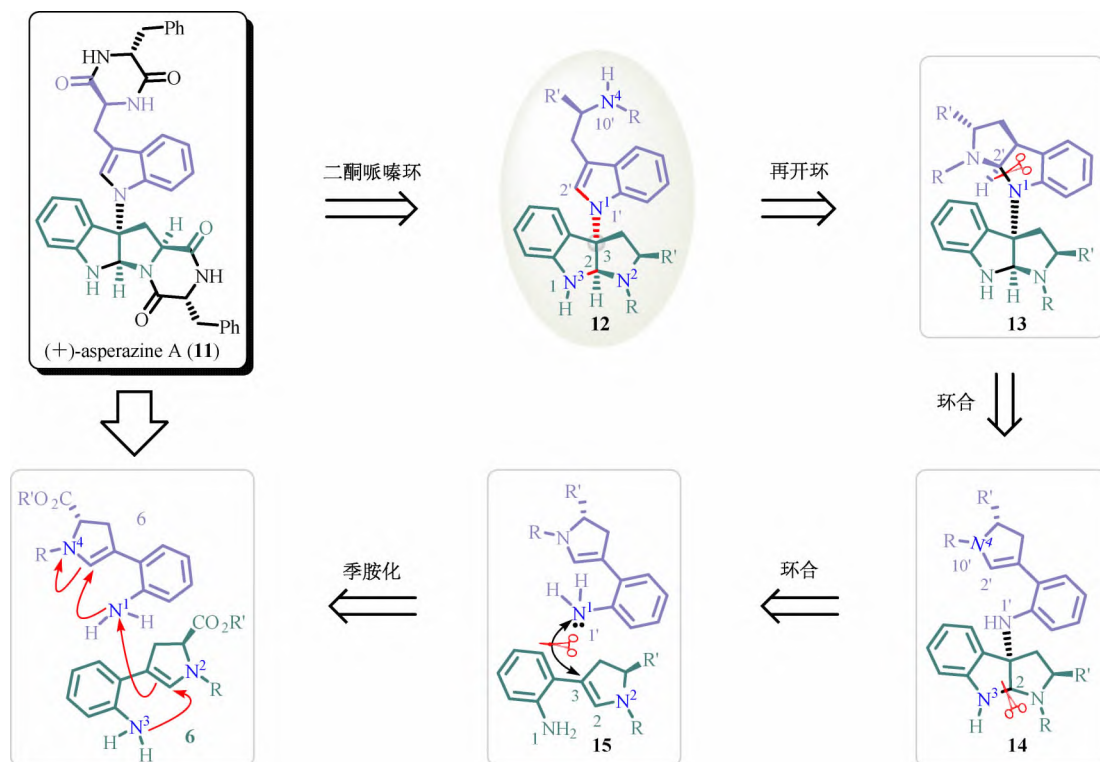


Fig 3 Total synthesis of (+)-asperazine A: an example to show the art of the disconnection approach in retrosynthetic analysis^[8]

图 3 以 asperazine A 为例揭示逆合成切断的艺术^[8]

领域的理论和技术在各自发展过程中如何进行深度交互、互为条件的过程。科里的前瞻性想法,因为有了新的计算机科学与智能技术作为条件,计算机辅助设计在今天逐渐得到了实现。不仅如此,基于量子化学理论和计算技术对分子结构进行计算和预测的计算化学,在 20 世纪 90 年代还发展成了一个专门的学科领域。虽然计算化学和科里提出的通过计算机模拟化学家进行逆合成分析的策略在框架上有所不同,但 2 者的实现均需要计算技术硬件和软件的同步发展,还需要将研究思路概念化和抽象化。小型计算机与网络化设备、新的计算模型概念、密度泛函理论这 3 者结合起来,才共同促进了计算化学在 20 世纪 90 年代的真正实现^[16]。即便计算化学在今天已经发展成为一种方便可用的“桌面建模”(desktop modeling),在制药业获得了大量应用,但是模型的参数化问题依然是化学家、计算机科学家和软件工程师在认识上的模糊地带^[17]。

科学史上化学研究与计算机技术的相遇“失败”与后续融合这一现象,启发我们去思考今天实现学科深度交叉所需要的条件:学科交叉需要多学科的协同发展,多学科的理论、技术和方法共同发展,才可能共同推动来解决具体的科学问题。对于化学研究而言,化学家如何将可用于计算和模拟的模型概念化和参数化,比如科里将他提出的逆合成原则概念化、转换成可供计算机计算和模拟的参数,在计算机辅助设计中扮演着至关重要的角色,而与此同时,不同领域的研究者也需要通力合作。

4 经验、合成策略的理论化与计算模拟

在科里之后,对化学键先进行切断、再进行重排,或者是先分解再整合,成为了有机化学研究的标准策略。如何能够让计算机在合成中更好地发挥作用,我们需要进一步思考,在使用这些标准策略时,化学理论与合成操作之间的关系。科学研究中理论和实验之间的关系一直是科学哲学的核心主题之一,理论和实验之间的互相促进关系,或者理论和实验谁在先谁在后的关系,是科学哲学重点关注的对象。已有的研究要不认为是理论在先,要不认为是实验在先。但是合成化学中,谁在先并不重要,合成化学的许多工作不仅仅是为了探讨理论本身,而是为了拓展合成的范围或者实现合成的目标,比如科里和夏普勒斯(K. Barry Sharpless)。

科里提出逆合成分析的策略,与 20 世纪 50 年代要合成新发现的具有复杂结构的化合物这一目标有密切关系。但是,当时化学家对复杂化合物的结构、性质和变化之间的关系缺乏理解,相应的反应

机理、结构和功能之间的关系是什么,在当时并不清楚。化学家如何在缺乏这种理解或者反应机理指导下,实现对物质进行操控,比如合成已有的物质和制备新物质,该问题已经不再是简单的理论与实验之间的关系问题。虽然实验在物理学等其他学科中也发挥着重要作用,但是化学作为制备性科学(而不仅仅是实验科学)的独特之处在于,化学家对化学反应理论和合成化学的操作之间的相互作用,并不是简单的理论在先或者实验在先的问题。科里团队对有机合成实验的探索,不仅是为了对有机合成的某个理论进行检验,而是为了提炼有关合成的知识、扩大合成的范围,这个过程为探索反应机理创造了机会。逆合成策略就是如此,它既是有关如何合成的理论,又在这个过程中促使化学家开始思考有机合成自身的机制。

将合成策略而不是理论作为研究目标,正是合成化学不同于其他学科领域的独特之处。科里通过逆合成分析和发明的各种全新合成反应,成功创造出了对生命体极其重要的物质,他将有机合成从百变的艺术转变为具有逻辑规律的科学,揭示了复杂合成背后的逻辑分析法则,他将化学家的注意力从单个分子转向了系列的复杂问题。正如霍夫曼(Roald Hoffmann)所言,逆合成分析原理的提出和发展,为化学家与自然界一争高低提供了合理的研究方式,而不仅仅是理解自然、提出科学理论。比如 2 次诺贝尔化学奖得主夏普勒斯在 20 世纪中期的研究,最开始也是从扩大合成的范围逐渐转向对反应机理的探索,在 20 世纪末期提出了“点击化学”的合成策略,而不是科学理论本身^[18]。科里和夏普勒斯等科学家的工作启示我们,化学研究中对有关如何合成的规则和方法的探索,它本身也是化学研究的重要目标。这种不同于其他领域的关系,正是作为制备性科学的合成化学本身所独有的研究方式。

需要指出的是,这并不说明对合成机理的研究不重要,而是说合成机理是化学家探究世界或者追求真理的一种工具,而不是探究目标。化学家通过逆向推理得到了多种可能的合成路线,要对这些路线的可能性做出精准评估,有机化学家必须使用他所知道的反应机理和已有的化合物的结构,才能做出这种推测。而背后的这种规则,就是化学反应的机理或者是科学哲学中讨论的机制性解释。化学家在将机理作为工具时,有两层含义。第一层是,机理作为一种薄的概念(thin conception),可以看作是化学家将化学变化的过程分解成符合标准的简

单步骤, 每一步都对应着特定的物质和它的反应活性。为此, 在科里等化学家并不是特别清楚具体的反应机制时, 依然可以凭借经验和探索性实验来成功地合成目标产物, 化学家对化学反应中可观察到的特征做出解释。LHASA 计算程序也可以将化学家背后的经验规则变得可计算, 优势是模拟出多种合成路线的可能性。第二层是, 机理作为厚概念(thick conception), 是将化学反应的理论模型和计算应用于有机合成, 做出因果解释和精准预测。为此, 在未来的人机交互中, 将基于化学家的个体经验和合成经验的薄概念型机理, 迈向深层次的基于有机化学理论的厚概念型机理, 才能将化学家的个体经验与合成机理转化为计算模拟做出因果解释的基础, 与计算模拟进行深度交互, 最终实现对合成步骤做出更加精准的推断。

参 考 文 献

- [1] Woody A I, Hendry R F, Needham P. *Philosophy of chemistry*. Elsevier, 2021
- [2] Gronowitz S. Award ceremony speech [EB/OL]. (2023-04-12) [2023-04-12]. <https://www.nobelprize.org/prizes/chemistry/1990/ceremony-speech/>
- [3] Corey E J. *Angew. Chem. Int. Ed.*, 1991, 30 (5): 455—465
- [4] Pool R. *Science, New Series*, 1990, 250 (4980): 510—511
- [5] Corey E J, Chen X. *The logic of chemical synthesis*. John Wiley & Sons, 1989
- [6] Corey E J. *J. Org. Chem.*, 2004, 69 (9): 2917—2919
- [7] Bai L, Ma Y, Jiang X. *J. Am. Chem. Soc.*, 2021, 143: 20609—20615
- [8] Bai L, Li J, Jiang X. *Chem*, 2022, DOI: 10.1016/j.chempr.2022.10.021
- [9] De Regt H W. *Understanding scientific understanding*. New York: Oxford University Press, 2017
- [10] 朱晶, 叶青. *自然辩证法研究*, 2016, 32 (6): 70—75
- [11] Burke M D, Lalic G. *Chem. Biol.*, 2002, 9 (5): 535—541
- [12] Corey E J, Loh T. *J. Am. Chem. Soc.*, 1991, 113 (23): 8966—8967
- [13] Hayashi Y, Rohde J J, Corey E J. *J. Am. Chem. Soc.*, 1996, 118 (23): 5502—5503
- [14] Corey E J. *Angew. Chem. Int. Ed.*, 2022, 41 (10): 1650—1667
- [15] Hepler-Smith E. *Hist. Stud. Nat. Sci.*, 2018, 48 (3): 300—337
- [16] Lenhard J. *Stud. Hist. Philos. Sci.*, 2014, 48: 89—96
- [17] Wieber F, Hocquet A. *Perspective Sci.*, 2020, 28 (5): 610—629
- [18] Zhu J, Chen M, Liang H, et al. *Chinese Chem. Lett.*, 2023, 34 (1), DOI: 10.1016/j.cclet.2022.107936

Logic-Guided Theory in Synthetic Chemistry: The Way of Thinking Backward and Computer Program of E. J. Corey

ZHU Jing¹ JIANG Xue-Feng^{2**}

(1. Department of Philosophy, East China Normal University, Shanghai 200241, China; 2. School of Chemistry and Molecular Engineering, East China Normal University, Shanghai 200062, China)

Abstract The relationship between scientific theory, experiment and computational technology is the main issue in the research fields of philosophy of science and science education. The previous studies focused more on physics and biology to investigate the relationship, while the scientific practice in chemistry was overlooked. Taking Elias James Corey's research and teaching practice as an example, this article analysis his scientific thinking and the reason why his computer program was not accepted as widely as his method of retrosynthetic analysis. It reveals that Corey made the methods and rules for synthesis understandable and reasonable to scientific community from his teaching and experimental practice. Based on his experimental work and thinking backward, Corey conceptualized and abstracted the rules for retrosynthetic analysis. However, computer-aid or artificial intelligent design can help to solve synthetic problems only when hardware and software interact deeply with theoretical organic chemistry and chemists' experience.

Keywords philosophy of chemistry; retrosynthetic analysis; artificial intelligence; interdisciplinary science; chemical thinking