

· 专题:交叉科学前沿与发展 ·

## 多学科研究的三种形态

吴家睿<sup>1,2\*</sup>

1. 中国科学院大学 杭州高等研究院生命与健康科学学院,杭州 310024
2. 中国科学院 分子细胞科学卓越创新中心,上海 200031

**[摘要]** 多学科研究在现代科学的发展中扮演着重要角色,呈现出三种主要形态。第一种形态是交叉学科研究,往往通过不同学科之间的高度融合而形成全新的学科,如物理学、化学与生物学交叉形成了分子生物学。第二种形态是会聚研究,即不同学科的专业知识和技术进行集成和协作,用以解决涉及多个领域之间的界面上重大科学问题或者社会挑战。第三种形态是当前大数据时代所特有的——数据驱动型研究,它利用大数据及其相应的技术打通或者跨越各门学科的边界,进而可以开展跨领域的研究。

**[关键词]** 多学科研究;交叉学科研究;会聚研究;数据驱动型研究

中国科教领域最近迎来了一个发展多学科研究的高潮。2020年8月,国务院学位委员会会议决定,将“交叉学科”定为中国教育体系中的第14个学科门类;教育部网站随后公布了160所高校提交的549个交叉学科备案名单。同年11月,国家自然科学基金委员会正式成立“交叉科学部”,负责统筹交叉科学领域基金资助等相关工作。2021年1月,国务院学位委员会、教育部发布《关于设置“交叉学科”门类、“集成电路科学与工程”和“国家安全学”一级学科的通知》,除了以法规的形式正式确立了“交叉学科”之地位,还首次批准设立了属于该门类的两个一级学科。

然而,对许多人而言,多学科研究的具体内涵并不一定很清楚,往往简单地望文生义,理解为只要涉及到两个或多个学科的研究就可以称为交叉学科研究。从教育部公布的交叉学科备案名单就可以看到这个问题的存在。首先,许多高校提出了同样的交叉学科名称,但涉及到的一级学科却各不一样,如“生物信息学”作为一门交叉学科被3所高校列入其交叉学科名单中,但是,3所高校在其涉及到的一级学科却有着不同的内容:A学校标出“生物学、数学”;B学校给出“生物学、生物医学工程,基础医学”;



**吴家睿** 中国科学院分子细胞科学卓越创新中心研究员;现任国科大杭州高等研究院生命与健康科学学院执行院长,中国科学院系统生物学重点实验室主任。担任 *Journal of Molecular Cell Biology* 主编, *BMC Systems Biology* 和 *Frontiers in Systems Physiology* 副主编,《医学与哲学》和《生命的化学》副主编,中国生物化学与分子生物学会分子系统生物学专业委员会主任委员。国家杰出青年科学基金获得者,中国科学院“百人计划”入选者;上海市领军人才入选者。发表学术论文100多篇。

C学校则提出“数学、生物学、计算机科学与技术”。此外,另有一所学校在“医学生物信息学”的名称下列入3个一级学科:数学、控制科学与工程、基础医学;还有一所学校在“生物信息技术”的名称下给出这样3个一级学科:控制科学与工程、计算机科学与技术、生物医学工程。其次,有些显然不应该属于交叉学科范畴,如一所医学院校把“包虫病学”列入“交叉学科”,涉及到基础医学、临床医学、公共卫生与预防医学、药理学、公共管理等5个一级学科。按照这种逻辑,任何一种疾病的研究都可以称之为“交叉学科”,因为至少要涉及到临床医学和药理学两个一级学科。让人更为担心的是,有的单位对交叉学科的命名比较随意,“生物功能分子学”是研究什么?从

收稿日期:2021-02-04;修回日期:2021-03-13

\* 通信作者,Email: wujr@sibs.ac.cn

本文受到国家自然科学基金重大研究计划项目(91753000)和中国科学院先导专项项目(XDB38020000)的资助。

字面上看,与生物学有关,但在学校列出的相关一级学科里,包括了化学、药学、基础医学、食品科学与工程,唯独没有生物学。

从术语称谓来看,“多学科研究”一词的英文为 Multidisciplinary Research,常常与“交叉学科研究”(Interdisciplinary Research)混用;而“交叉学科”则往往有一个非规范的英文名词“Interdiscipline”,并有时等同于“交叉科学”(Interdisciplinary Science);近年来新兴的“会聚科学”(Convergence Science)也常常被用来代表“交叉科学”。

从科学发展史来看,多学科研究是 20 世纪推动科学和技术发展的主要动力;进入 21 世纪,多学科研究的重要性只增不减。通过对生命科学发展历程的梳理,笔者分析了多学科研究具有的主要形态及其相应的特征,进而揭示出多学科研究在科学进程中的角色多样性以及对科研范式影响的复杂性。

## 1 高度整合的交叉学科研究

随着化学的发展,尤其是 19 世纪初期化学的一个分支学科——有机化学在欧洲的形成,研究者开始把化学与生物学整合在一起,如当时的德国已经出现了 *Zeitschrift fur Physiologische Chemie* (《生理化学杂志》)这样明显具有交叉学科色彩的科学期刊。1905 年的美国也诞生了一本相似的期刊——*Journal of Biological Chemistry* (《生物化学杂志》);在该杂志的第一届编委会成员中,一位来自伯克利大学的科学家娄伯(Loeb J)在给主编的信中表达了他对学科交叉的信念:“生物学的未来建立在从化学的角度来研究生物学问题”<sup>[1]</sup>。

这种交叉学科首先体现在对生命本质的研究之上。早期的研究者认为,生物体的构成材料与非生物体的有着根本的区别,前者称为有机物,后者称为无机物,且前者只能来自生物体而不能人为地用后者合成。1828 年,德国化学家维勒(Wohler F)首次在实验室用氰酸和氨水两种无机分子合成了一种有机分子“尿素”,从而证明了构成生命的物质和非生

命的之间并没有一个不可逾越的界限。研究者由此开始了对生物体的核酸、蛋白质、糖和脂类等各种生物分子的研究;随后诞生了一门化学与生物学高度整合的交叉学科——生物化学(Biochemistry);它被《大英百科全书》定义为:涉及发生在植物、动物和微生物的化学物质和过程的科学研究领域。也就是说,在研究者的眼里,生命只不过是化学研究中一种特定的对象,它的存在及其功能都建立在其构成材料以及构成方式之上。

生命科学领域中最重要交叉学科是诞生于 20 世纪中叶的分子生物学。20 世纪 40 年代,奥地利物理学家薛定谔(Schrödinger E)发表了《生命是什么》一书;他在书中明确提出,生命和非生命一样,“在它内部发生的事件必须遵循严格的物理学定律”<sup>[2]</sup>;因此,研究者可以通过物理和化学的技术和方法来研究生物体的属性或特征。最有代表性的是美国化学家鲍林(Pauling L)对蛋白质空间结构的研究;他应用量子化学和 X 射线晶体学技术分析了蛋白质上连接氨基酸残基的肽键(C—N 键),认为肽键具有部分双键的性质而不能自由旋转,导致连接肽键两端 6 个原子的空间位置处在一个相对接近的平面上,称为“肽键平面”;他在此基础上进一步提出,蛋白质内的各个肽键平面之间可以通过连接它们的  $\alpha$ -碳原子进行旋转,从而可以形成  $\alpha$  螺旋等二级结构。此外,著名的 DNA 双螺旋结构同样是在物理学家、化学家和生物学家的通力合作下得以阐明;而遗传密码的主要提出人则是著名的物理学家伽莫夫(Gamow G)。由此可见,分子生物学是高度整合了物理学、化学和生物学的一门交叉学科(图 1A)。

分子生物学的成功使得还原论思维成为了现代生命科学的主流方法论。在分子生物学家建构的生命科学大厦里,所有研究对象无论是简单的原核生物还是复杂的动植物,都不过是一架按照物理和化学规律运转的“分子机器”;研究者的任务就是从分子层面去认识和揭示这种规律,即寻找分子机器运

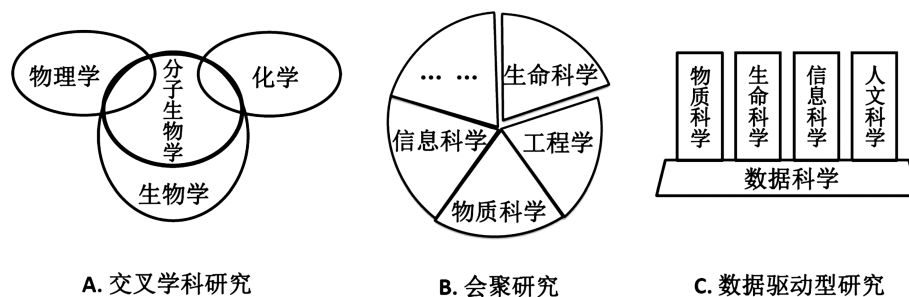


图 1 多学科研究的三种形态

行的因果关系。对还原论者而言,生命科学的主要目标通常就是去研究有关某种生理或病理活动的因果关系之科学假设。因此,建立在分子生物学基础上的现代生命科学采用的就正是这种“假设驱动”的研究范式。美国著名肿瘤生物学家温伯格(Weinberg R)在一篇题为“假设优先”的文章中做过一个很好的总结:“在20世纪,生物学从传统的描述性科学转变成为一门假设驱动的实验科学。与此紧密联系的是还原论占据了统治地位,即对复杂生命系统的理解可以通过将其拆解为组成的零部件并逐个地拿出来进行研究”<sup>[3]</sup>。

## 2 紧密协同的会聚科学研究

尽管多学科研究推动了生物化学和分子生物学等交叉学科的诞生,进而使得生命科学在20世纪下半叶取得了巨大的发展,但研究者还希望要进一步提升多学科研究的能力,以满足维护人类健康和防止环境污染等重大社会需求。2001年末,美国的国家科学基金会(NSF)等政府部门围绕着“会聚四大技术,提升人类能力”的主题,举行了一次科学家与政府官员等各界人物参与的研讨会,在会议上首次提出了“会聚技术”(Converging Technologies)的概念,并特别强调了来自四个科学技术领域协同作用的“会聚技术”——纳米技术、生物技术、信息技术和认知科学(其简化英文的联式为Nano-Bio-Info-Cogno,缩写为NBIC)。NSF基于这次会议编写了一份报告:《提升人类能力的会聚技术》。

会聚科学研究的目标与20世纪生物学领域的交叉学科研究的目标有明显的区别。后者是要揭示生命的活动规律,属于基础研究领域;而前者则是要提高社会的创新能力或满足社会重大需求,属于应用研究领域。例如,美国麻省理工学院(MIT)在2016年发布了一份报告,题目就是“会聚:健康研究领域的未来”(Convergence: The Future of Health)。在美国科学家夏普(Sharp PA)等人看来,分子生物学和基因组学是过去生物医学领域的两次革命,而会聚研究则代表第三次生物医学革命,“工程师和物质科学研究者将与生物学家和医生一道去解答众多新的医学挑战”<sup>[4]</sup>。他们还特别通过“组织工程”(Tissue-Engineering)为例讨论了会聚研究与交叉学科研究的不同:“这不同于典型的交叉学科形态(Interdisciplinary Situation)——把一种确定的细胞类型给工程师或者让工程师能够确定在某种生物系统中什么样的材料是有用的;相反,这种(组织工程)从一开始就要进行多学科合作(Multidisciplinary Collaboration),所有的参与者都

要有共同的参考点和语言。这一领域如果没有会聚方法(Convergence Approach)是不会存在的”<sup>[4]</sup>。

会聚研究的一个重要特点是对工程学的强调,这与会聚研究偏重应用的目标是高度一致的。例如,美国科学院研究理事会2014年发表的战略报告题目就是“会聚观:推动跨学科融合——生命科学与物质科学和工程学等学科的跨界”。工程学的介入不仅推动研究工作进入应用领域,而且能够产生具有工程特色的成果,如美国科学家夏普等人在评论文章“实现会聚在卫生保健中的价值:整合物质科学、工程学和生物医学”中指出:工程学在生物相容性材料和纳米技术领域发展了全新的策略,这种策略在促进卫生保健方面具有前所未有的潜力<sup>[5]</sup>。

会聚研究的另外一个重要特点是对技术的强调,一方面是高度重视技术在应用层面的价值,如“NBIC”会聚技术的提出,体现出围绕着会聚研究目标的科学与技术外在的一体化;另一方面还强调了不同学科的技术在推进交叉研究本身的价值——包括实验仪器和材料、分析方法和技术等,即注重研究过程中科学与技术内在的一体化,从实验对象到实验操作再到实验结果。由此可以看到,新兴的交叉学科如“化学生物学”(Chemical Biology)和“物理生物学”(Physical Biology)等与传统的交叉学科如“生物化学”(Biochemistry)和“生物物理学”(Biophysics)有着巨大的差别;前者强调研究者应用化学和物理的有关技术来开展生物学研究,如化学生物学的一位创始人曾这样说过:我们的目标是为每一个基因找到相应的小分子化合物,用它们来分析细胞和有机体的功能; *Nature Chemical Biology* (《自然·化学生物学》)创刊时发表的社论是这样说的:“化学生物学作为一个新领域,是植根于化学家和生物学家紧密合作带来的技术进步之上”<sup>[6]</sup>。

会聚科学研究的这些特点提示我们,这显然是不同于交叉学科研究形态的另一种多学科研究形态。前者往往通过不同学科之间的高度整合而形成全新的交叉学科,如生物化学或分子生物学;而后者则表现为不同学科之间的相互协作,如化学生物学或物理生物学。会聚科学研究强调的不是学科间的“交叉”,而是不同学科的“会聚”(比较图1A和1B)。换句话说,各门学科代表了不同的专业化分工,而会聚研究则是把这些专业中相关的技术和理论进行集成,以便针对维护健康或开发新能源等国家/社会的重大需求去建立或发展新的技术方法。正如美国科学院关于会聚研究的战略报告所指出的那样:“会聚观体现了一种交叉学科研究的扩展形式,专业知识构成了研究活动的‘宏观’模块,而各个‘宏观’模块

又组合形成一个更大的整体”<sup>[7]</sup>。

### 3 跨越边界的数据驱动型研究

21 世纪之初人类基因组计划的实施推动生命科学进入了大数据时代。人的基因组拥有 30 亿个碱基对,即相当于 3 GB 的数据;目前国际上储存的个体基因组序列的数据量已超过百万人级;数据科学家估计到 2030 年时,每年世界范围内产生的基因组数据将为 25 PB(1 PB=10<sup>15</sup> Byte)<sup>[8]</sup>。大数据在医疗健康领域则有更明显的增长,如美国国立肿瘤研究所的基因组数据平台在 2016 年成立之后的第一年内,就收到了研究者提交的 4.5 PB 的数据<sup>[8]</sup>。据统计,世界范围内产生的医疗健康数据在 2013 年大约为 153 EB(1 EB=10<sup>18</sup> Byte),而在 2020 年估计将增长到 2 314 EB<sup>[8]</sup>。如果说有什么是 21 世纪的生命科学乃至整个科学比 20 世纪进步的标志,那就是数据的获取和数据的利用。

生物医学大数据不仅为人类认识和改造世界提供了巨大的资源,而且改变了生命科学和医学的多学科研究范式,其典型代表正是人类基因组计划。该项计划也属于多学科交叉研究,其实施过程需要依靠新型测序仪器和试剂的研发,以及海量数据的处理与分析等,涉及到物理学、化学、生物学、信息科学和数学等多个学科。但与假设驱动的学科交叉研究不一样的是,该项计划属于“发现的科学”(Discovery Science),是要获得有关人体细胞基因组的全部核苷酸序列的数据。也就是说,这项研究表现出了一种新的研究范式——数据驱动的研究范式,开展研究的出发点不是科学假设,而是高通量的数据采集能力;而研究的目标也不是去解决某个具体的科学问题,而是要去获取海量的数据。

数据驱动的研究范式有一个重要的特征:“迭代”(Iterate),即每一次研究工作可以是一种不完备的阶段性工作,然后在前期研究结果的基础上反复地进行完善,通过多次研究逐渐逼近预定的总体目标。人类基因组计划明显具有这种“迭代”特征,尽管其终极目标是揭示人类基因组的所有核酸序列,但在 2001 年 2 月发表的人类基因组测序成果只不过是一个覆盖了基因组 90% 核酸序列的“草图”;2004 年 10 月在 *Nature* 杂志上发表了人类基因组全图的论文,也只给出了常染色质区域内大约 99% 核酸序列的测定结果。事实上,人类基因组的核酸序列中至今仍然存在许多高度重复序列区域(如中心粒)没有被测定;2020 年 9 月,研究者终于在《自然》周刊上发表了人类第一条完整的、没有测序“缺口”的染色体的核酸序列,但仍然还有 22 条人类染

色体上的序列“缺口”待研究者去补测。不久前,一个比“人类基因组计划”更为宏大的“人类细胞图谱”(Human Cell Atlas, HCA)研究计划正式启动,其基本目标是,通过各种单细胞分析技术来鉴定人体拥有的 40 到 60 万亿个体细胞中的所有细胞类型;其采用的主要研究策略也明显具有“迭代”特征<sup>[9]</sup>。

数据驱动的研究范式的另一个重要特征是研究的“开放性”,即不受已有理论框架的限制。20 世纪生命科学的主流是假设驱动的研究范式,以解决具体的科学问题为主要目标;其研究工作的开展是依赖于既有的理论框架,从而使得研究者在事实的选择和理解方面容易受限于指导研究的理论框架。而数据驱动的研究范式则不仅能够让研究者避免理论框架带来的主观偏见,而且可以帮助其发现在假设或者现有理论范围之外的全新知识。正如美国生物学家戈卢伯(Golub T)在一篇题为“数据优先”的文章中所指出的:“如果没有获得全面的肿瘤基因组数据,将难以区分信号和噪音。尽管假设驱动的实验科学依然处于研究领域的中心位置,但不带偏好的肿瘤全基因组测序将提供前所未有的机会去催生新的想法”<sup>[10]</sup>。

数据驱动的研究范式的“开放性”特征,使得研究工作从追求事物之间的因果关系转变为寻找事物之间的相关性。这就导致了过去看上去彼此没有关系的学科产生了全新的联系,如谷歌的工程师开发了一种算法,可以根据每天汇总的谷歌搜索数据近乎实时地对流感疫情进行预测;其预报流感爆发的地域性和时效性比美国疾控中心报的更好<sup>[11]</sup>。即使是复杂的人类行为,也可以通过大数据的分析和计算来进行研究;为此 10 年前已经出现了一门相应的交叉学科——“计算社会科学”(Computational Social Science)<sup>[12]</sup>。这种基于大数据的跨学科研究形成了多学科交叉研究的第三种形态,利用数据科学来打通或者跨越各门学科的边界(图 1C)。国际著名的生命科学期刊 *Cell* 在 2020 年创立了一个以数据科学为纽带的多学科研究的子刊 *Patterns*, 其主编在发刊词里这样写到:“*Patterns* 将通过数据科学技术的共享来打破各门学科的边界,这些数据科学技术可以用来解决跨领域的问题”<sup>[13]</sup>。

更重要的是,数据驱动的研究范式作为一种超越理论框架的“开放式科学”,其研究产生的资源和数据可以被广泛地用于研究各种各样的科学问题;如收集了 50 万英国人的生物样本和健康医学数据的“英国生物资源库”(UK Biobank),自 2012 年建成至今,已经用于支持过世界各国数万名研究人员的研究工作,仅 2018 年度利用该库开展研究的科学

家就有 4 000 多人,发表了涉及到健康医学各个方面近 300 篇研究论文。换句话说,这种“开放性”使得数据驱动的研究范式之组织模式明显不同于假设驱动的研究范式,前者强调众多研究者之间的合作、交流和共享;而后者则更注重以项目负责人(Principle Investigator, PI)为主导的“个人英雄主义”式的研究;正如最近的一篇评论文章所指出的:传统的卓越观是崇拜天才而与社会环境无关;这种对卓越的狭隘看法导致了资源集中到已获认可的科学家手中,进而限制了科学的进步和新思想的产生以及多学科研究领域的发展<sup>[14]</sup>。2020 年 12 月,美国科学院发表了一份战略报告《无止境的前沿——科学的未来 75 年》,在其结论中特别指出:科学激励制度应该是一种“共赢游戏”(Positive-Sum Game)而非“零和游戏”(Zero-Sum Game)<sup>[15]</sup>。显然,数据驱动的研究范式正在推动科学研究的生态环境发生革命性的改变。

### 参 考 文 献

- [1] Edsall JT. The Journal of Biological Chemistry after seventy-five years. *Journal Biological Chemistry*, 1980, 255 (19): 8939—8951.
- [2] 埃尔温·薛定谔. 生命是什么. 罗来鸥, 罗辽复, 译. 长沙: 湖南科学技术出版社, 2003: 8.
- [3] Weinberg R. Point: Hypotheses first. *Nature*, 2010, 464 (7289): 678.
- [4] Sharp PA, Langer R. Promoting convergence in biomedical science. *Science*, 2011, 333(6042): 527.
- [5] Sharp P, Jacks T, Hockfield S. Capitalizing on convergence for health care: Integrate physical sciences, engineering, and biomedicine. *Science*, 2016, 352(6209): 1522—1523.
- [6] Editorials. A community of chemists and biologists. *Nature Chemical Biology*, 2005, 1(1): 3.
- [7] 美国科学院研究理事会. 会聚观: 推动跨学科融合——生命科学与物质科学和工程学等学科的跨界. 王小理, 熊燕, 于建荣, 译. 北京: 科学出版社, 2015: 13.
- [8] Banks MA. Sizing up big data. *Nature Medicine*, 2020, 26 (1): 5—6.
- [9] Regev A, Teichmann SA, Lander ES, et al. The human cell atlas. *eLife*, 2017, 6: e27041.
- [10] Golub T. Counterpoint: data first. *Nature*, 2010, 464 (7289): 679.
- [11] Ginsberg J, Mohebbi MH, Patel RS, et al. Detecting influenza epidemics using search engine query data. *Nature*, 2009, 457(7232): 1012—1014.
- [12] Lazer DMJ, Pentland A, Watts DJ, et al. Computational social science: Obstacles and opportunities. *Science*, 2020, 369(6507): 1060—1062.
- [13] Editorial. The first piece of the pattern. *Patterns*, 2020, 1: 100020.
- [14] Urbina-Blanco CA, Jilani SZ, Speight IR, et al. A diverse view of science to catalyse change. *Nature Chemistry*, 2020, 12(9): 773—776.
- [15] Olson S. The endless frontier: The next 75 years in science. Washington, DC: The National Academies Press, 2020.

## Three Forms of Multidisciplinary Research

Wu Jiarui<sup>1,2\*</sup>

1. School of Life and Health Sciences, Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Hangzhou 310024

2. Center for Excellence in Molecular Cell Science, Chinese Academy of Sciences, Shanghai 200031

**Abstract** Multidisciplinary researches play important role in promoting modern science, which presents three forms. The first form is called interdisciplinary research, which means that various disciplines are highly integrated and then form a newly interdisciplinary, for example, the integration between physics, chemistry and biology results in molecular biology. The second form is called convergence research, which means that the knowledges and techniques of various disciplines are converged to collaboratively tackle scientific and social challenges at the interfaces of multiple fields. The third form is called data-driven research, which means a transdisciplinary research to solve the problems spanning various fields by breaking down those disciplinary boundaries based on sharing big data and related techniques.

**Keywords** multidisciplinary research; interdisciplinary research; convergence research; data-driven research

(责任编辑 刘敏)

\* Corresponding Author, Email: wujr@sibs.ac.cn