

2021



易凯资本健康产业团队

“AI+” 系列报告之 1: AI+新药研发

2021.9

重要声明

分析师声明

每位主要负责编写本研究报告全部或部分内容的研究分析师在此声明：（1）本报告中所述的任何观点均准确地反映了其个人对于行业及公司的看法；（2）分析师无论在过去、现在及将来，均不会因本报告中所表达的观点或建议收取任何报酬。

一般性声明

本研究报告仅供我们的客户与合作伙伴使用。除了与易凯资本相关的披露，本研究报告是基于我们认为可靠的目前已公开的信息，但我们不保证该信息的准确性和完整性，客户也不应该依赖该信息是准确和完整的。我们在可能的情况下会适时地更新我们的研究。除了一些定期出版的行业报告之外，绝大多数报告是在分析师认为适当的时候不定期地出版。

我们的投行人员、销售人员和其他专业人员可能会向我们的客户提供与本研究报告中的观点截然相反的口头或书面市场评论或交易战略。我们的资产管理部门可能会做出与本报告的建议或表达的意见不一致的投资决策。

本报告不构成明示或暗示的投资建议，也没有考虑到个别客户特殊的投资目标、财务状况或需求。客户应考虑本报告中的任何意见或建议是否符合其特定状况，以及(若有必要)寻求专家的意见，包括税务意见。

我们的部分研究报告会以电子出版物的形式刊登在我们的内部客户网或微信阅读号上并向所有客户和合作伙伴同步提供。并非所有研究内容都转发给我们的客户或者向第三方整合者提供，我们也并不对由第三方整合者转发的我们研究报告承担任何责任。

易凯资本不承诺不在过去、现在或未来与报告中提及的公司发生业务关系或寻求与它们有关的业务机会。

本报告版权为易凯资本所有。未经易凯资本公司事先书面同意，本材料的任何部分均不得(i)以任何方式制作成任何形式的拷贝、复印件或复制品，或(ii)以盈利为目的再次分发。

易凯资本健康产业团队“AI+”系列报告之1：AI+新药研发

2020年以来，国内AI计算驱动的创新药物研发领域（下称“AI+药物研发”）热度明显攀升，出现多起大额融资：

晶泰科技

2020年9月 C轮 3.188亿美元

（当时全球领域内最高纪录）

2021年8月 D轮 4亿美元

英矽智能

2021年6月 C轮 2.55亿美元

百图生科

2021年8月 A轮 数亿美元

望石智慧

2021年4月 B及B+轮 1亿美元

齐鲁锐格

2021年2月 B轮 9000万美元

星药科技

2020年12月 A轮 数千万美元

2021年8月 战略融资

.....

根据公开信息整理，2021年截至目前，国内已完成31起融资交易，总融资额超过110亿元人民币，已超过2020年全年融资交易数量与融资额。

针对大众普遍关注的12个问题，易凯资本给出了我们的答案。

01 为什么这一两年 AI+药物研发这么热？

近年来深度学习算法的突破，算力大幅提升，以及生物学技术发展和成本下降带来的生物医药大数据加速积累为 AI 在生物医药领域的应用提供了客观条件。

海外 AI 驱动研发的候选药物进入临床，以 Schrodinger 和 Relay 为代表的 AI/计算驱动制药公司上市，以及多个跨国药企和互联网巨头的涉足成为该领域热度攀升的催化剂。

新冠疫情对行业与资金面的影响不可忽略。疫情驱动市场对生物医药行业有了前所未有的关注，流动性宽松的背景下资金向风险资产倾斜，而 AI+药物研发作为市场想象空间巨大的交叉领域，同时受到 TMT 和医疗这两个大行业资本的追捧。

对国内企业来说，国家对于创新药的扶持是 AI+药物研发热的重要驱动因素。以 2015 年药品审评审批制度改革为起点，国家从注册审批、医保支付、资本市场上市规则、人才政策、专利保护等各个角度全面鼓励创新药的发展，me-too、me-worse 等“伪创新药”和仿制药的价值空间被极大压缩，对创新的追逐以及降本增效的需求驱使医药企业加速拥抱 AI。

根据易凯资本的扫描，截至 2021 年 8 月底，国内共计有 57 家 AI+药物研发公司。其中最早一批公司成立于 2012 年-2014 年，多数公司成立于 2018 年后，就 AI 的应用场景及 AI 驱动的研发成果而言，目前行业整体还处于起步阶段。

02 AI 能从根本上颠覆创新药市场格局吗？

就目前的应用效果而言，AI 在药物研发领域只是扮演赋能工具的角色，谈 AI 对市场格局的颠覆还为时尚早。我们预期 AI 带来的研发质量与效率的提升会成就一批采用创新研发模式的“新生代”biotech，但 big pharma 多年积累构筑的临床开发与商业化壁垒非一朝一夕可以建成，行业主流格局在短期内难以被撼动。但我们不排除随着数据积累、模型创新与迭代、新的监管与商业模式的出现，AI 将来有可能重塑创新药行业格局。

03 AI 在创新药研发领域有哪些应用？

AI 在靶点发现与验证、分子发现与优化、临床前研究、临床试验设计与患者筛选入组、药物上市后研究等新药研发的各个环节均有大量应用。新药研发的链条很长，不同环节所需要的技术能力与经验背景差异很大，因此多数 AI+药物研发公司选择从其中一两个环节切入，且主要集中在靶点发现与验证和分子发现与优

化阶段,这是因为生物学机制理解和候选分子性质是决定临床试验成败的最关键因素。

04 AI 辅助药物研发 (AIDD) 目前主要应用的算法有哪些? 有何局限性?

目前应用在创新药研发行业的 AI 技术主要指的基于数据驱动的机器学习 (ML) 算法,实际上 ML 乃是 AI 大范畴下的一个主要任务。而以深度神经网络 DNN 代表的深度学习技术则是目前最前沿的 ML 方向,在这次 AIDD 热潮中发挥了至关重要的作用。AIDD 当前主要应用的深度学习算法模型包括卷积神经网络 CNN、循环神经网络 RNN (LSTM)、生成对抗网络 GAN、Transformer、图神经网络 GNN 及其变种和组合等。除此之外,模型中还经常有许多机器学习中的经典算法来配合使用,如决策树、支持向量机、蒙特卡洛树搜索、MCMC 等。

对于深度神经网络相关算法来说,其形成的端到端的系统结构对于输入输出而言类似一个黑盒,因此模型参数的可解释性相对较差也就成为其主要的局限性,如何用现有的生物医药知识有效地指导模型搭建和效果优化也是当前的一个主要研究方向。

05 创新药的研发主要依托于有机化学、生物化学、临床医学等一系列实验科学,而偏重理论的计算科学一直以来对药物研发过程指导甚微。将 AI 技术引入药物研发后能否引发学科革命,使得计算科学能够切实指导药物研发?

计算机辅助药物设计 (CADD) 在小分子药物研发领域已有很长的应用历史, NASDAQ 上市公司 Schrodinger 提供的解决方案就是 CADD 的应用典范。CADD 基于物理规则,利用计算机来模拟化学事件并精确计算原子和分子的物理化学性质,对算力要求高,通量低,主要应用于已有苗头化合物后的分子优化。

AI 则是基于数据的统计学习方法, AI 的引入可以加速数据处理过程,显著提升 CADD 的效率,但 AIDD 绝不能简单理解为 CADD 的迭代与延伸, AI 技术可以被有机整合至新药研发的各环节,将计算科学在药物研发中的应用从 CADD 聚焦的分子发现和优化阶段扩展到药物研发的整个生命周期,为药物研发带来更彻底、广泛的变革。

06 对于 AI+药物研发公司来说,算法重要还是数据重要? 如何理解算法的优势以及判断数据质量?

都很重要,引用一句在业界广泛流传的话:数据与特征决定了机器学习的上限,模型和算法只是无限逼近这个上限。

算法模型的关键是理解其底层生物学意义,对 domain knowledge 的理解是决定模型质量的关键。算法模型的优势可以体现在多个维度,比如结果的精准度、计

算速度、模型体量、泛化性能等，不同算法模型可能有不同的侧重方向，因此优势也会不尽相同，在特定任务需求和应用场景下合理选择具有相应优势的算法模型至关重要。

数据质量主要体现在数据本身的统计特征(如单因素分析与多因素分析的统计学显著性)、不同来源和不同批次数据间的一致性等方面。另外，原始数据的结构化程度，噪声含量，以及与真实应用场景下的数据分布偏差程度，也都是判断数据质量的重要标准。

07 AI+药物研发公司的主要数据来源有哪些？

公共数据是很重要的数据来源，如 TCGA、PDB、Drugbank、GenBank 等大量公共数据库及丰富的文献专利为 AI+药物研发企业提供了海量珍贵数据。

通过委托 CRO、客户合作及公司自有实验平台定向产生有价值的高质量数据。

数据的质量和数量对 AI+药物研发公司来说至关重要，但对数据的挖掘、理解以及对关键问题的定义的重要性不亚于数据本身，这需要基于公司团队对于生物学与药物研发的深入理解。

08 AIDD 主流算法模型需要大量数据驱动，而创新药行业内真实数据依然相对匮乏，如何解决这一问题？

目前主流的深度学习都是依靠大量训练数据，通过自动学习的算法得到包含相关知识的模型参数，从而实现模型功能。面对训练数据不足的情况，目前主要从“开源”和“增效”两个角度来解决。

“开源”即是增加训练数据量，可以利用传统的数据处理思路，如机器预标注、数据增强等方式来减少数据准备成本从而扩增数据量；也可以通过行业整合的方式，联合多方数据源头搭建数据共享平台，利用区块链和联邦学习等技术在保证数据隐私和所有权的情况下，实现行业数据的合理利用与有效整合。

“增效”即是通过创新和改良深度学习方法，提高对有限训练数据的使用效果，比如利用迁移学习、小样本学习等方法，利用少量训练数据对现有预训练模型进行微调，从而实现模型在新数据集下的相应功能（预训练模型一般来自相同任务下现有的通用模型，或者可通过低成本大体量的数据集训练获得）。

09 目前绝大多数 AI+药物研发企业集中在小分子药物研发领域，AI 是否也能应用到大分子药物研发领域？

AI 在大分子药物领域的应用的确远远落后于小分子领域，主要原因是大分子结

构与功能的复杂性以及大分子数据的缺乏。

在蛋白药物研发领域，AI 辅助的抗体定向进化已在工业界崭露头角。蛋白定向进化通过构建突变文库、高效表达与筛选蛋白、对筛到的突变体进行再次突变与筛选，多轮迭代后得到性质满意的蛋白，这一手段在酶工程领域已有成熟应用，近年来有不少公司将这一思想应用到抗体药物筛选中，如国外的 BigHat、Absci、Ginkgo 和国内少数从事 AI+大分子药物开发的公司。AI 的应用可以提高文库构建的效率与质量，减少迭代筛选次数，使得抗体定向进化更为高效。



然而在巨大的潜在蛋白质序列空间中，定向进化方法只能覆盖很小一部分，且即使有 AI 辅助，抗体定向进化过程依然需要多轮实验迭代，繁琐耗时。尽管在小分子领域虚拟筛选已得到广泛应用，但大分子的虚拟筛选一直是学术与工业界的难题。最近我们欣喜地看到国内的信华生物已实现基于抗体虚拟筛选的重大突破，该公司开发的 Vibrantfold 平台可以高通量精确地预测抗体结构和构建虚拟抗体突变文库，实现不依赖于实体突变文库的、不受库容限制的高效抗体虚拟筛选。

当前的抗体定向进化与虚拟筛选均需要基于已有的天然抗体，在大分子药物领域的终极目标是实现蛋白质的从头设计。蛋白从头设计不依赖现有的天然蛋白，而是基于生物物理与生物化学原理从头搭建、设计具有全新结构和全新功能的蛋白质。华盛顿大学 David Baker 教授是蛋白质设计领域的领军人物，他领导的实验室已经在这一领域取得了一系列奠基性的成果，我们期待在不久的将来能看到蛋

白质设计的前沿成果在工业界中得到应用。

除了蛋白药物以外，我们也在基因治疗、mRNA、新抗原等多个生物药领域看到AI的应用，例如Dyno利用机器学习和高通量实验改造病毒衣壳；mRNA公司将AI用于mRNA密码子优化、mRNA结构优化和LNP筛选；而机器学习在肿瘤新抗原的筛选过程中几乎成为“标配”，国内华大吉诺因已经将AI驱动研发的新抗原疗法推向临床。

10 什么样的AI+药物研发公司团队称得上是一个优秀的团队？

对于AI+药物研发这样的前沿交叉领域而言，创始人需要具备持续学习、快速迭代及合作共赢的企业家精神。

决策层具有计算和药物研发的双重背景是很大的加分项，复合背景的管理层可以引导计算团队和药物研发团队高效深度融合。

11 AI+药物研发公司如何进行商业变现？

AI+药物研发公司有两种典型商业模式，一种是以服务为主的SaaS供应商/CRO模式，一种是自研管线为主的Biotech模式。SaaS供应商/CRO模式优势是可以更快实现收入，并通过广泛合作积累更多数据以支持算法迭代，但因药物研发里程碑实现周期长、本土药企付费意愿不强，因而天花板比较明显；Biotech模式通过自有管线推进能更快验证算法平台能力，且创新药资产可以给公司带来更大的价值爆发潜力，我们更看好拥有创新自研管线的AI+药物研发公司。

12 2022年AI+药物研发融资趋势如何？

随着二级市场估值回调及美元对国内投资的收紧，预计整个市场热度将出现一定程度的回落，但二八分化效应明显，具有顶尖计算与药物研发背景团队、顶级基金和MNC背书的AI+制药公司仍将受到资本追捧。

融资整体向中后期转移。根据公开信息披露，获投项目数量占比超85%，而获投项目中近半仍处于早期阶段，预测未来一年资金将向中后期市场转移。

AI+大分子热度提升。目前，多数AI药物研发公司的主要研究方向集中在小分子领域，随着算法模型和研发思路的不断创新发展，AI在大分子领域开始落地并将迎来更多关注。

产业资本数量不会大幅增长，仍旧以财务投资为主。以药企为主的产业主体更倾向于与AI创新公司合作的模式。主要原因是行业整体还处于早期阶段，而药物研发的周期太长，无法短时间内看到验证结果，多数本土药企对于AI的价值仍

处于观望态度。

总结与展望

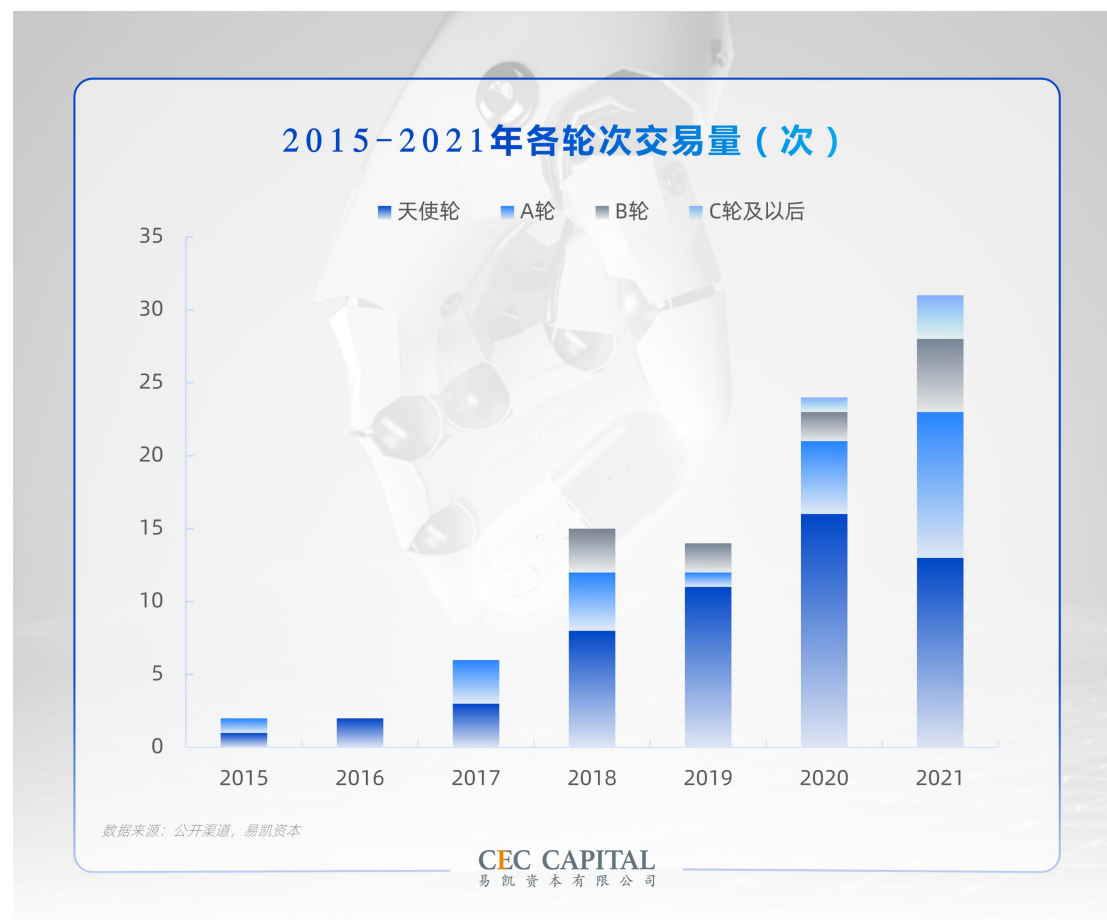
生物医药研发长期处于“作坊式”状态，数字化与自动化程度低，依靠大量人力与专家经验，不同研发环节之间存在明显脱节。而创新药平均10亿美金级的研发成本、平均超过10年的研发周期、小于10%的临床开发成功率和药企逐年走低的资本回报率导致行业变革的需求迫在眉睫，AI无疑让大家看到了希望。

但AI在药物研发领域的应用还存在很多现实挑战，比如高质量数据的缺乏、模型可解释性不足、人才的稀缺、传统药企对AI的谨慎态度等。尽管AI在短期内还无法彻底“重塑”创新药研发流程，但随着数据积累、模型创新与迭代、新的监管与商业模式的出现，我们期待未来能看到AI将与药物研发流程深度融合，并对创新药行业带来更彻底的变革。

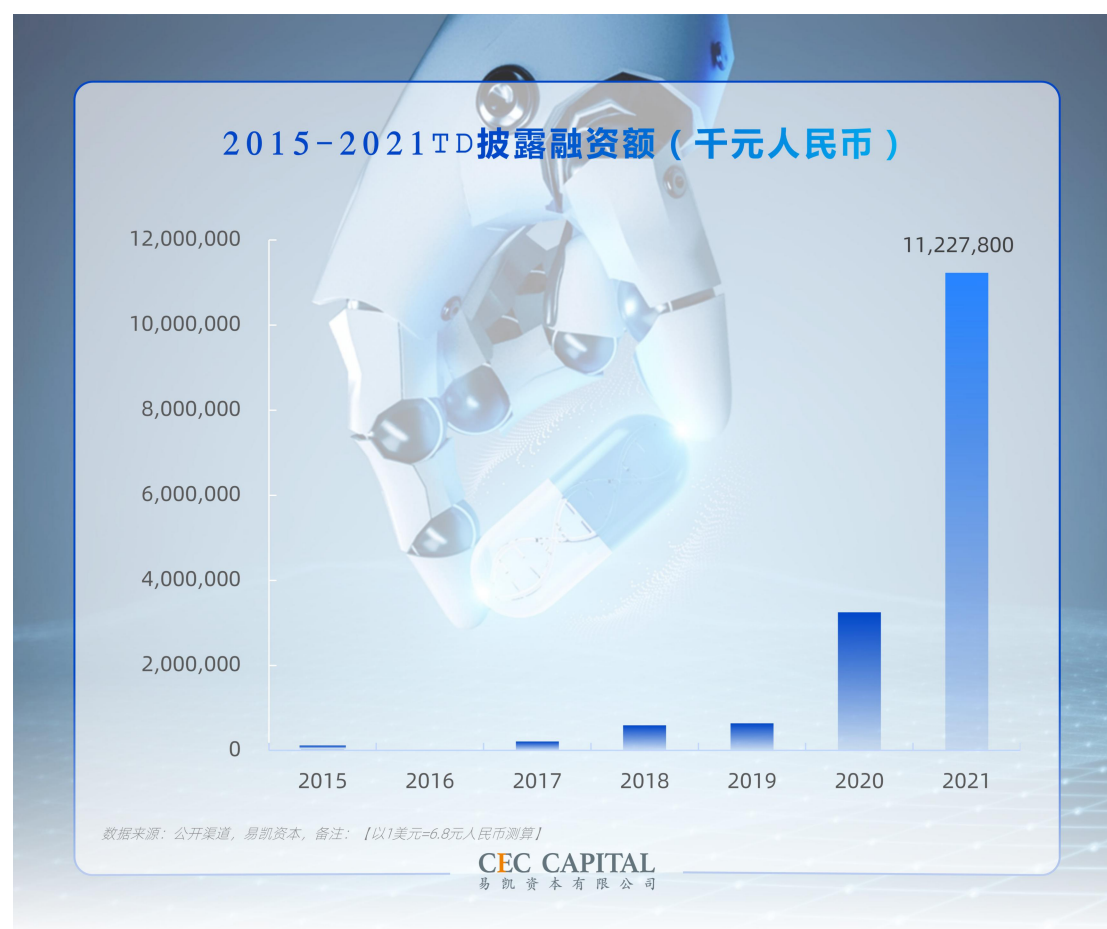
附录 AI+药物研发资本市场现状

公开数据显示，截至2021年8月17日，全国共计46个AI+药物研发项目获得融资，产生交易99起，参投机构（VC、基金、产业资本）190家，累计融资总额近160亿元人民币。

融资偏向早期，逐步向中后期发力。融资行为自2015年开启，每年皆集中在早期阶段，交易量占比超过50%，中后期占比逐步提升。与AI行业整体融资高峰期主要集中在2018年之前相反，AI+药物研发交易主要集中在2018年后，尤其2021年截至当前交易量已超2020全年的24次。



整体而言，因大笔融资额的影响，2021年公开披露完成逾110亿元人民币。刨除大额交易，21年截至当前融资额已超2020全年32亿元人民币的总额。此外，受中后期交易增多影响，2020年融资额同比增长超400%。



资本活跃，以财务投资为主。参投机构（VC、基金、产业资本）共计 190 家，二次追加及以上者占比近 30%，其中参与 5 次交易以上的有 8 家机构，分别是红杉资本、五源资本（原晨兴资本）、峰瑞资本、高榕资本、源码资本、百度、真格基金、BAI 资本。其中产业资本 24 家，药企近半，共计参与 40 起交易，以百度、晶泰科技、腾讯、依图科技、药明康德为代表。此外，与其他产业资本偏好选择与财务资本合投不同，药企更偏好独立投资。



作者

易凯资本医药与生物科技组 张骁 卜佩璇
易凯资本健康产业研究院 李海存

制图

李鑫 赵毅