

一颗药的诞生

新药为什么要十五年、几十亿美元，以及 AI 到底改变了什么

费曼式写法 · 由 AI 代理撰写与互相审校

导读

新药为什么要十五年、几十亿美元，以及 AI 到底改变了什么

先别问 AI，先想一个问题：**一万个候选分子进实验室，为什么只有一个能走进药店？**

如果新药研发像造手机——零件确定、图纸确定、剩下的只是工程——那它不该这么慢，也不该这么贵。但它平均要花十到十五年、十几亿美元，而且九成的努力最后走进死胡同。慢和贵都不是管理不善，它们是某种更本质的东西的账单：我们对生命的无知。

这本书要带你完整走一遍这颗药的旅程：先找到身体里那个「坏开关」（靶点），再在 10 的 60 次方的化学宇宙里捞一根针，把它雕成药，让它闯过动物实验、三期临床、药监审批，最后面对定价和专利的现实。每一关你都会看到同一个模式：**每一道关卡的存在，都是因为历史上在那道关卡死过人、或烧掉过几十亿美元。**

然后，我们会正面回答那个这几年被问了无数遍的问题：AI 到底能不能让这一切变快？

答案不是一句「能」或「不能」。AI 在蛋白质结构预测上做到了人类五十年没做到的事；在虚拟筛选和分子生成上是真实可用的加速器；但它对「为什么九成临床试验失败」这个最贵的问题，至今几乎无能为力——因为失败的根源不在算力，而在数据里没有的生物学。这本书会讲清楚每一条边界在哪里，以及为什么边界恰好在那里。

读完它，下次再看到「AI 将让新药研发从十五年变成五年」这样的标题，你应该能自己拆出：哪几分是已经发生的事实，哪几分是有依据的期望，哪几分是融资 PPT。

旅程从第一章那个残酷的漏斗开始。

第一章 · 一万个里活一个

先说一个数字：一个全新的小分子药，从科学家在白板上画下第一个候选分子算起，到它摆上药店的货架，平均要花 **十到十五年**，直接成本按行业最常被引用的估算在 **十亿到二十多亿美元** 这个量级。这还不是最贵的部分——最贵的是，绝大多数项目根本走不到终点。

行业里有一句顺口溜：进实验室的一万个候选分子，最后只有一个能变成药。这个数字不精确，但它传达的比例感是对的。新药研发的本质不是「制造」，而是**淘汰**——它是一台设计得极其残酷的过滤器，每一层都要筛掉百分之八十以上的选手。

这台过滤器长什么样

把整个过程想象成一个漏斗，从上往下越来越窄：

- **发现阶段 (discovery)**：先找到一个值得打的「靶子」（身体里某个和疾病有关的蛋白质或基因，下一章细讲），再找出能影响这个靶子的候选分子。这一步可能有上万个分子被测试，绝大多数直接出局。
- **临床前 (preclinical)**：剩下的几十个分子在细胞和动物身上做毒性、代谢测试。这一步过后，通常只剩个位数。
- **临床试验 (clinical trial)**：在人身上做，分三期，每期规模更大。这是漏斗最陡的一段——临床失败率高到什么程度？进入一期临床的候选药，最终能上市的只有大约 **十分之一**。
- **审批与上市**：监管机构（比如美国的 FDA，中国的国家药监局的药审中心）审查全部数据，决定给不给上市许可。

大白话

大白话：研发新药不像造一台手机——造手机是把确定的零件组装起来；研发新药更像从一万个报名者里选一个宇航员，每轮体检都刷掉大部分人，而且后面每一轮都比前面贵得多。

钱到底花在哪了

很多人以为药贵是因为药片本身贵。其实一片药的生产成本往往只有几毛钱。真正的成本在三个地方：

第一，时间。十几年里，一个几十到几百人的团队全职只做这一件事。人员的工资、实验室的设备、临床中心的费用，是按年烧的。

第二，越往后越贵。一期临床可能几十个健康志愿者，花几千万；三期临床动辄几千个病人、几十个国家、几百个医院同时做，一项三期试验烧掉几亿美元是常事。这就像拍电影的预算结构：前期筹备不贵，实拍阶段最贵——而研发偏偏是「实拍」阶段失败率最高。

第三，也是最容易被忽略的：失败也要付账。一家大药厂同时跑着几十个项目，九成会死掉。死掉的项目花的钱不会消失，全部摊进了活下来的那个药的定价里。你买一颗药时付的钱，大头其实是在替它的九千九百九十九个「阵亡兄弟」买单。

为什么非这么残酷不可

可以放宽一点吗？让筛子松一点，成本不就下来了？不能，因为这台过滤器的每一层都对应着一条真实的生物学风险：

- 分子在试管里有效，不代表能活着到达病灶——它可能被肝脏分解、被肾脏排掉、被细胞膜挡住；
- 在动物身上安全，不代表在人身上安全——物种差异坑过无数项目；
- 在三百个病人身上有效，不代表在三万个病人身上有效——罕见的副作用要到大样本才现形。

历史告诉我们，任何一层放松，代价都会以「出人命」的方式追回来。1937 年美国一款磺胺糖浆用二甘醇（一种有毒的工业溶剂）做溶剂，一百多人死亡，直接催生了现代药品审批制度。这台残酷的过滤器，是用事故一点点加固出来的。

这本书要回答的问题

了解完这台漏斗，自然会有两个问题：每一层具体在筛什么、为什么非筛不可？以及——这几年被吹得神乎其神的 **AI**，到底能钻进漏斗的哪一层、把哪一段缩短？

这本书先带你走完前八章：一颗药从靶点到货架的完整旅程。然后第九章开始正面回答 AI 的问题——它在哪里真的改变了游戏规则（剧透：蛋白质结构预测是一个），在哪里只是杯水车薪（剧透：临床失败率的根因不在算力）。读到最后，你应该能自己对一句常见的宣传做出判断：「AI 将让新药研发从十五年变成五年」——这句话，哪些部分可信，哪些部分是在讲故事。

第二章·找到那个坏开关

要造一把钥匙，先得知道锁长什么样。新药研发的第一步，不是找分子，而是回答一个生物学问题：**这个病，到底是身体里哪个零件出了问题？**这个「零件」就是靶点（target）——大白话说，就是药物要去拨动的那个开关，通常是一个蛋白质，偶尔是一段基因。

这一步听起来最不像「研发」，没有试管没有白大褂，更多是看文献、分析数据、做遗传学研究。但它是整本书最重要的一步：**选错靶点，后面十几年、几十亿美元全部归零**，而且你会走得很远才知道自己错了。

先有假设，才有药

每一个药物项目的起点，都是一句生物学假设，形如：「如果把蛋白质 X 的活性关掉，这个病就会好转。」注意，这是一句假设——它可能来自动物实验，来自病人的基因数据，来自对一个天然分子的观察，但它始终是未经验证的推测。验证它的唯一办法，是造出药、给人吃、看效果——也就是说，**要验证一句假设，代价是走完整个漏斗**。

这就解释了为什么制药公司对靶点的选择近乎偏执。2015 年《自然·遗传学》上有一项影响很大的分析：回看上万个药物项目，发现**如果靶点有人类遗传学证据支持**（比如天生携带某个基因突变的人恰好更少得这个病），项目最终获批的概率大约是没这类证据的两倍。两倍听起来不多，但在一个成功率只有 10% 的行业里，这是足以改变投资组合的差异。

最好的老师：天不生病的幸运儿

怎么找到有人类遗传学证据的靶点？一个绝妙的思路是：**去找那些基因「坏掉」却反而更健康的人**。大自然替我们做完了一个不可能做的实验——你不可能故意敲掉一个人的某个基因看看会怎样，但基因突变的彩票在几十亿人身上随机开过无数次，只要找到中奖者就行。

教科书级的例子是 PCSK9。2003 年，研究者发现这个基因的功能异常增强会导致家族性高胆固醇血症——说明它和胆固醇水平强相关。随后几年，更关键的证据出现了：一些人天生携带 PCSK9 的失活突变，开关从出生起就是关着的，他们的低密度脂蛋白胆固醇（俗称「坏胆固醇」）水平极低，心脏病风险大幅下降，而且——关键来了——**他们没有任何健康**

问题。这句话翻译成药厂的语言就是：「把 PCSK9 这个开关关掉，收益巨大，代价为零。」于是针对 PCSK9 的抗体药被开发出来（2015 年获批），从靶点确认到上市只用了十二年，算快的。整个项目最大的底气不是实验室数据，而是那些天生胆固醇极低的健康人——他们替所有病人提前活了一遍「这个药起效后的人生」。

大白话

大白话：选靶点就像选赌桌——人类遗传学证据是告诉你哪张桌子历史上赔率高。赌还是要赌，但聪明的玩家只坐有数据的那几桌。

靶点选错的坟场

反例同样触目惊心。阿尔茨海默病（俗称老年痴呆）的淀粉样蛋白假说认为，大脑里堆积的 β -淀粉样蛋白斑块是病因，清掉斑块就能治病。这个假说统治了这个领域二十多年，几百个临床试验围绕它展开。结果是一连串的惨败：多个清除淀粉样蛋白的药物在三期临床中要么无效，要么只带来微小且有争议的获益，而清除斑块本身确认是成功的——药确实把斑块清掉了，病人的认知却没有好转。

这就是靶点选错最残酷的样子：**药是好药，锁找错了**。斑块可能只是疾病的「结果」或「伴随现象」，而不是「原因」——就像火灾现场的灰烬不是纵火犯。几千亿研发经费烧进去，换来的是整个领域在 2020 年代开始痛苦的转向。为什么非要烧这么多才发现？因为验证靶点的唯一办法是走完漏斗，而漏斗的出口在人体试验——这是上一章说的，这台过滤器的最陡一段。

还有一个隐藏关卡：这个开关「摸得到」吗

选对了开关，还要看它能不能被药够到。多数小分子药只能结合蛋白质表面特定形状的「口袋」，而人体内很多致病的蛋白质表面平平无奇，没有口袋可钻——这类靶点被叫作 undruggable（不可成药），大白话就是「锁孔太浅，钥匙插不进去」。著名的致癌蛋白 RAS 被称为「不可成药之王」，三十年里无数项目折戟，直到 2021 年才有第一个直接针对它的药获批（利用的是一个此前没注意到的、只在突变型上出现的隐蔽口袋）。

所以「找到一个坏开关」其实是三个问题的合取：它真是病因吗？关掉它会不会殃及别处？它有没有能让药钻进去的形状？三个都答「是」，才轮到化学家登场——下一章，去几十万

个分子里捞那根针。

第三章·几十万分子里捞一根针

靶点定了，假设写在白板上：「关掉这个开关」。现在需要一把真正能拨动它的钥匙——一个能和这个蛋白质结合的小分子。这一步叫苗头化合物发现 (hit discovery)，大白话就是：在茫茫化学宇宙里，先捞出一根「虽然粗糙但确实有点用」的针。

先感受一下这个宇宙有多大。一个勉强能当药的小分子，大概是几十个原子按某种方式连起来。这样的分子理论上有多少种？常用的估算是 **10 的 60 次方** 量级——比地球上所有沙子的数量再乘上一万亿还多。而全球所有药厂的化合物库加起来，存量也就十亿 (10^9) 上下。**化学宇宙我们只探索过微不足道的一小角**，这就是找药难的第一层原因。

笨办法的巅峰：高通量筛选

最直接的想法：既然不知道哪个分子有用，那就挨个试。这就是高通量筛选 (HTS, high-throughput screening) ——把公司积攒的几十万到几百万个化合物，用机器人一个个加进装着靶点蛋白质的微孔板里，看哪个能让蛋白质「变色」（实际测的是荧光、吸光之类的信号）。

这套系统在 1990 年代被寄予厚望，药厂为此建了巨大的自动化工厂，一天能筛几十万个反应。结果呢？令人失望。命中率低得可怜，常常几十万分子里只捞出一两个靠谱的；而且捞出来的「针」往往有隐藏缺陷（比如它让信号变色不是因为结合了靶点，而是自己聚集成一团「污染」了实验）。钱花了，自动化程度拉满了，成功率并没有质的飞跃。这是行业学到的第一课：**在 10^{60} 的宇宙里， 10^6 的穷举依然约等于瞎猜。**

聪明人的三条近路

既然穷举不够，就得耍聪明。过去三十年，化学家发展出几种思路，每一种都是对「怎么缩小搜索空间」的不同回答：

第一，跟着大自然走。微生物和植物花了几亿年进化出各种化学武器（抗生素、毒素），这些分子天生就是「能和蛋白质打交道」的。历史上相当比例的药来自天然产物或其改造版

——青霉素来自霉菌，他汀类降胆固醇药来自真菌，抗癌药紫杉醇来自红豆杉。缺点是天然产物结构复杂，改造成本高。

第二，用小碎片拼。与其试完整的大分子，不如只试只有十几个原子的「化学碎片」。碎片太小，结合力弱，传统方法测不出来，但用核磁共振、X 射线晶体学能看到它们结合在蛋白质的哪个角落。找到两三个各自贴在不同角落的碎片，再把它们「焊接」成一个大分子——基于片段的药物设计（fragment-based design）。这就像不直接找一把完整的钥匙，而是先找几小段能卡进锁芯的金属丝，再拼成钥匙。2011 年获批的抗癌药维莫非尼（Zelboraf）就是这条路线的成名作。

第三，在计算机里先筛一遍。既然物理筛选又贵又慢，能不能先在电脑里算一下哪个分子可能结合？这就是虚拟筛选（virtual screening）——用物理模拟或经验公式估算每个分子和靶点口袋的「匹配度」，只把得分最高的几千个送去实验室真测。它不能取代实验（计算模型的误差很大），但能把「十万选一」压缩成「百选一」，省钱省时间。这条路，正是第九章 AI 登场的入口。

大白话

大白话：找起点分子像找对象——高通量筛选是相亲大会（量大、成功率低），天然产物是「相信丈母娘的眼光」，片段拼接是自己攒一个，虚拟筛选是先刷一轮简历再见面。没有哪条路单独够用，实际项目通常几条路同时走。

捞到针之后，针还很钝

假设运气好，几条路线交上来几个「苗头」：它们确实能影响靶点，但离药还差十万八千里。典型的苗头化合物是这样的：效力不够强（要很高的浓度才起效）、可能也结合别的蛋白质（副作用的种子）、可能一进血液就被分解（根本活不到病灶）、可能溶不进水（没法做成药片）。

把这样一根钝针打磨成药，是下一章的活——药物化学家的主场，整个研发流程里最像「手艺」的一段。在那里你会看到一个反直觉的事实：这段工作至今几乎无法自动化，它靠的是化学家对分子的直觉。记住这一点，到第九章谈 AI 时它会变得重要。

第四章 · 把一块糙石头雕成药

上一章结尾，我们捞到了一根「钝针」：能碰到靶点，但浑身是毛病。把它打磨成药的过程叫先导优化（lead optimization）——大白话说，就是围绕这根针，系统地改动它的化学结构，一次改一小步，每改一次就测一遍所有指标，直到它在几十个指标上同时过关。这是药物化学家的主场，也是整个研发流程里「手艺」含量最高的一段。

不是爬山，是在钢丝上挑扁担

先破除一个直觉：优化一个分子，不是「越改越好」的爬山过程，而是一个**多目标平衡**。一个候选药要同时满足一大堆互相打架的要求，行业里最常念叨的一串指标缩写是 ADMET——吸收（Absorption，吃进去能进血液吗）、分布（Distribution，能到达病灶吗）、代谢（Metabolism，会不会太快被肝脏拆掉）、排泄（Excretion，会不会赖在体内太久累积毒性）、毒性（Toxicity，会不会误伤别的器官）。

这些指标为什么打架？因为它们是同一个分子的不同侧面。举两个具体的例子：

- 想让分子更容易穿过细胞膜（吸收好），通常要让它更「油」一点；但太油了溶不进水，药片做不成，而且容易卡在脂肪组织里赖着不走（排泄差）。
- 想增强对靶点的结合力，可以加一些化学基团让分子「贴」得更牢；但那些基团可能恰好是肝脏代谢酶的最爱，分子一进体内半小时就被拆光了（代谢太快）。

所以先导优化的日常是：化学家盯着一串指标的「仪表盘」，改一个原子，效力上去了半格、溶解度掉了两格，再换个地方改，找回溶解度、效力又退一步……像挑着扁担走钢丝，任何一端都不能掉。一个典型的优化项目要合成和测试 **几百到几千个分子**，花 **三到五年**。

大白话

大白话：先导优化像装修一套学区房——既要地段好（效力）、又要面积大（溶解度）、还要总价低（毒性小）、还得学区不掉（代谢稳定）。每项单独都能做到，全都要就只能一点点磨。

化学家的「手感」从哪来

这一步有量化工具辅助——比如把「结构改动 → 指标变化」的数据喂给统计模型，预测下一个分子该怎么改，这叫定量构效关系（QSAR）。但实战中，决定性的一步往往来自化学家的直觉：看一个分子的结构式，就像老木工看一块木料，能「感觉」到哪个部位别扭。

这种直觉不是玄学，它是对大量隐性知识的压缩：哪些基团容易惹毒性、哪些结构在做大规模合成时会爆炸性地贵、哪些改动历史上改善过代谢稳定性。一个资深药物化学家脑子里装着几千个「上次这么改发生了什么」的案例。这也是为什么这个环节几十年来自动化进展缓慢——它的知识散落在无数失败项目里，大部分从未被系统地记录下来。第九章我们谈 AI 时会回到这一点：**没有数字化过的知识，AI 就学不到。**

优化到一半，靶点的锅又冒出来了

先导优化还有一个外人很少知道的坑：有时候分子本身越改越好，**指标全绿，但你逐渐发现这个靶点选错了**。比如细胞实验里关掉开关能杀死癌细胞，可换成更接近人体的模型后效果消失了。这时团队面临最痛苦的决定：止损，还是再赌一把？沉没成本在这里是天文数字，许多项目就是在这种「再试一年说不定有救」的心态里多烧掉几年的。

活过这一关的分子，会拿到一个类似「准生证」的身份：临床前候选化合物（preclinical candidate）。它在试管和细胞里证明了自己，仪表盘全绿。但下一关的规则完全不同——它要离开试管，进入真正的生物体。第五章去看它在动物身上的遭遇。

第五章·动物替人先趟一遍

候选分子在试管里再风光，也只是面对一盘细胞。现在它要进入完整的生命体了。这一步叫临床前研究（preclinical studies）：在人身上碰它之前，先在动物身上把「会不会毒死、会不会致癌、剂量怎么定」这些问题尽量问完。它有一道硬性的法律门槛——不给监管机构看过合格的动物数据，任何药都不许进入人体。

动物实验在问三个问题

第一，毒在哪里？团队会不断提高剂量，直到动物出现不良反应，找到那条「安全线」——专业上叫未见不良反应剂量（NOAEL），大白话就是「吃这么多还没事的最高量」。第一次给人用的剂量，就是从这个量再往下打一个大折扣（通常除以十以上）推出来的。

第二，药在体内经历了什么？这就是上一章那串 ADME 指标的活体版：药进血液后浓度怎么随时间变化、多久被肝脏代谢、从尿还是胆汁排出去。这些数据画出来的「浓度-时间曲线」，决定了病人将来是一天吃一片还是一天吃三片。

第三，长期吃会怎样？短期毒性之外，还有几个月到两年的长期试验：会不会致癌、会不会影响生育、会不会损伤胎儿。这也是为什么研发快不起来——有些问题只能用时间回答，没有捷径。

还有一个你可能想知道的规则：毒性试验通常要在**两种动物**上做，一种啮齿类（大鼠或小鼠）加一种非啮齿类（犬或猴）。原因很简单：单一种族可能恰好对某种毒不敏感，两种完全不同的物种能兜住更多意外。

动物到底多像人？答案：大约七成

现在来到本章的核心问题，也是整个研发漏斗里最深的裂缝之一：**动物实验能多大程度预测人体？**

2000 年有一项被引用至今的系统性研究（Olson 等人，汇总了多家药厂一百多个化合物的数据）：动物毒理试验和人体毒性的吻合度大约是 **70%**。也就是说，三成的时候，动物说

「没事」，人有事；或者动物说「有毒」，人其实没事——而后者意味着一些本来能救人的药被错杀了。

为什么会差这三成？因为「物种差异」不是细节差异，而是系统差异。人的肝脏代谢酶谱和鼠不同——同一个分子在鼠肝里三小时分解，在人肝里可能十分钟就没了，也可能反过来。免疫系统差异更大：2006 年英国著名的 [TGN1412](#) 事件中，一种在猴身上以五百倍剂量都安全的抗体药，给六名健康志愿者注射后，几小时内全部因免疫系统剧烈失控（细胞因子风暴）进了重症监护室。猴子的免疫系统和人的，在那一个受体上恰好长得不一样。

大白话

大白话：用动物预测人，像用其他城市的天气预报预测本市——大方向通常对，但台风路径会偏。而且这里的『台风』是出人命级别的偏差。

效力和毒性，是两道不同的坎

还要分清两件事。动物实验预测**毒性**，还有上面说的七成把握；预测**疗效**，把握低得多。因为大多数动物疾病模型是「人造」的——让小鼠得「类似阿尔茨海默的病」要靠基因改造硬造出来，它和人类自然演化了几十年的疾病差得很远。在癌症领域，移植了人类肿瘤的小鼠模型无数次让药物「治愈小鼠、对人无效」，多到行业里流传着一句自嘲：「我们治愈小鼠癌症的次数，比治愈人多得多。」

所以临床前阶段的真实定位是：**它是一张昂贵的入场券，不是预测，而是排雷**——把「显然会毒死人」和「显然进不了血液」的分子拦在外面。排完雷，剩下的分子终于走向唯一的终极考场：人体。下一章，进入研发史上最贵、失败率最高的三期临床。

第六章 · 人体才是终极考场

所有的实验室数据、动物数据，到这一章全部变成背景音。药物研发的终极裁判只有一个：**生病的人体**。临床试验（clinical trial）就是在人身上验证药物的过程——这是漏斗最陡的一段，一半以上的项目死在这里，烧掉的钱也最多。它分三期，一期比一期人多、一期比一期贵。

三期各问一个问题

一期：安全吗？几十个健康志愿者（抗癌药等剧毒药物例外，直接在病人身上做），从极小的剂量开始往上加，看人体能耐受多少、血药浓度怎么走。一期不问疗效，只问「这药在人身上的脾气」。

二期：有苗头吗？几百个真正的病人，开始看疗效信号——肿瘤有没有缩小、血压有没有降。同时继续盯安全性。二期是整个流程里死亡率最高的一期：大约**三分之二的候选药死在这里**，死因几乎都是一个：动物和细胞里的疗效，在人身上没复现。

三期：真的比现有疗法强吗？几千甚至几万个病人，遍布多个国家的几十上百家医院，通常要跑好几年。三期的结果将直接决定监管机构批不批。一项三期花掉几亿美元很常见——这就是为什么二期末期的「继续还是止损」是药厂最难的决策：往前一步是天文数字，往后一步是前功尽弃。

临床试验的心脏：一个极其精巧的「对照实验」

程序员都懂 A/B 测试。临床试验本质上就是 A/B 测试，但难度高到离谱，因为你的「用户」是病人，而人这个系统充满了噪声。三期试验的标准设计叫随机双盲对照（randomized double-blind controlled trial），拆开看每一层都是在堵一个具体的噪声源：

- **随机分组**：病人被随机分到吃药组或对照组。不随机行不行？不行——如果让医生挑，他多半会把病情轻的分给新药，结果新药「显得」有效。随机化是为了让两组的病人除了吃没吃药之外，其他所有方面都统计上一样。

- **双盲**：病人不知道自己吃的是什么，*医生也不知道*。为什么要蒙住医生？因为医生的期望会漏进他的观察里——「我觉得这个病人好转了」是一种主观判断，期望会扭曲它。
- **对照**：对照组吃的要么是现有最好的疗法，要么是安慰剂——外观一模一样、不含药的面粉片。为什么不能「吃 vs 不吃」？因为有安慰剂效应：光是「我在接受治疗」这个信念，就能让相当一部分病人的症状真实改善——这不是幻觉，信念能调动真实的生理变化。在疼痛、抑郁这类指标主观的病上，安慰剂效应能解释掉三到六成的「改善」。不扣掉它，你测的根本不是药效。

大白话

大白话：临床试验是给『生命这个史上最复杂的黑箱』做控制变量。随机化堵选择偏差，双盲堵期望偏差，安慰剂堵信念本身——三层装甲全是因为被坑过才焊上去的。

为什么失败集中在疗效上

行业内有个统计：临床失败的原因里，「疗效不足」占一半左右，远超「安全性问题」。这很反直觉——前面筛了那么多轮，怎么疗效还会不够？

答案藏在第二章里：**靶点假设从头到尾都没被真正验证过**。分子在细胞里关掉开关有效，是「分子层面的正确」；但「关掉这个开关能不能治好这个病」是「生物系统层面的问题」——人体有冗余通路、有代偿机制，关掉一个开关，系统可能绕路走。这个问题只有真人试验能回答。所以二期的高死亡率不是设计失误，而是漏斗在正常工作：这里是「靶点假设」第一次接受真人检验的地方，之前所有环节都只是在确认「分子配得上这个假设」。

理解了这一点，你就会明白为什么行业对「用 AI 提高临床成功率」寄予厚望又屡屡受挫——失败的根源在生物学认知，不在计算能力。这个话题第十一章展开。先看活过三期的幸运儿下一关要面对什么：药监部门的签字。

第七章·药监签字才算数

三期临床数据齐了，药厂把几万页的资料装订成一份申请（在美国叫 NDA，新药上市申请；在中国递交给国家药监局的药品审评中心），交到监管机构手里。很多人以为这是「走过场」，其实这里是又一道真实的关卡——监管机构拥有一票否决权，而它的视角和药厂完全不同：药厂想证明「这药有效」，监管机构的职责是确认「**给几百万陌生人用的收益，大于风险**」。

审评员在看什么

审评团队里有医生、统计学家、药理学家、化学家，他们看的不是药厂的总结报告，而是**原始数据**：每个受试者的病历级记录。他们会问的问题包括：

- 统计显著性是真的吗？试验设计有没有中途偷改终点？（临床试验开始前必须注册并「锁定」分析方案，就是防这个的。）
- 疗效大到「有临床意义」吗？统计学上显著的改善，可能只是血压多降了 2 毫米汞柱——数字上赢了，病人感受上无感。
- 安全信号有没有被平均数稀释？比如总体肝损伤率不高，但某一类基因型的患者风险很高。
- 工厂能稳定地生产出和临床试验用药一模一样的药吗？（这叫 CMC，生产工艺审查——很多人不知道，「怎么造」和「有没有效」是同等重要的审批内容。）

这个过程通常要十个月上下。历史上审批要慢得多，1980 年代艾滋病危机中，病人们围着 FDA 抗议「药在审评桌上，人在坟里」，直接推动了加速审批通道的诞生——对危重疾病，可以用「替代终点」（比如病毒载量下降，而不是生存期延长）先批，上市后再补验证性试验。这是一个深思熟虑的权衡：**用一部分确定性，换时间**。

批了不等于安全了：四期，永不结束的试验

上市不是终点，而是最大规模试验的开始。前三期临床加起来也就几千个病人，而上市后用 药的是几百万人——**发生率万分之一的严重副作用，要十万人用药才大概率见到一次**，临床阶段根本看不见。这就是四期临床（上市后监测）存在的理由：监管机构持续收集不良反应报告，发现信号就采取行动。

行动的梯度从轻到重：更新说明书、加黑框警告（印在说明书最上方的粗黑框，是最高级别的警示）、限制适用人群，直到撤市。撤市不是传说：2004 年，关节炎药罗非昔布（Vioxx）因被发现长期服用显著增加心脏病风险而全球撤市——那时它已经上市五年、被八千万人用过。这个风险在上市前的几万人试验里没有统计学意义，是上市后真实世界的数据让它现形的。

大白话

大白话：临床试验像驾考——考过了说明你会开，不代表你开一辈子不出事。四期监测是交警加行车记录仪，开的人够多、时间够长，罕见的风险才现形。

一道永远在移动的线

回过头看，审批制度的核心难题是一个没有标准答案的权衡：批得太严，病人等不到救命药；批得太松，病人成了试验品。磺胺糖浆事件催生了审批制度，艾滋病人的抗议催生了加速通道，Vioxx 的撤市强化了上市后监测——每一道规则背后都有名字和数字。这条线还在移动：罕见病基因疗法只有几十个病人可做试验，传统的三期范式根本塞不下，监管机构正在学习用真实世界数据、用单臂试验加外部对照来做判断。

至此，这颗药走完了「科学」的旅程。但还有一个问题你没在任何一章里看到答案，而它决定了这颗药能不能到达你手里——钱。下一章算账。

第八章·为什么药这么贵

药为什么贵？最常见的回答是「药厂贪婪」，最学术的回答是「研发成本高」。两个都对了一点，但都没抓到结构。这一章把账算清楚——算完之后你会发现，药的定价逻辑和别的商品根本不同，它更像一场「限时垄断的生意」。

第一颗药片十亿美元，第二颗五毛

行业里流传一句话：「第一颗药片的成本是十亿美元，第二颗的成本是五毛。」这句话几乎就是药价之谜的全部答案。制药业的经济结构是**极致的「高固定成本 + 低边际成本」**：研发（固定成本）是天文数字，生产一片药（边际成本）几乎可以忽略。同样的结构你在软件业见过——Windows 第一张光盘贵，复制一张几乎免费。

但软件和药有一个关键差别：软件的成本大头是「造出来」，药的成本大头是「**失败掉**」。回到第一章的漏斗：一家公司花十年烧掉几十亿，烧掉的绝大部分钱属于那些死在三期临床的项目。活下来的那一个药的定价，必须把这些尸体全部背在身上——药厂在定价时计算的从来不是「这颗药花了多少」，而是「我们整个投资组合要多少年才能回本」。

专利悬崖：一门会到期的生意

新药受专利保护，专利期二十年——但注意，**专利钟是在实验室阶段就开始走的**，等药批下来上市，往往只剩七到十二年。这段时间是药厂唯一回本的窗口。窗口一关，仿制药（generic，别人照着你公开的配方造的同款药）蜂拥而入，价格通常在几年内跌掉八九成。原研药的销量曲线像从悬崖上跳下去，行业里干脆叫它专利悬崖（patent cliff）。史上最著名的悬崖是降胆固醇药立普妥（Lipitor）：巅峰时年销一百三十亿美元，2011 年专利到期后两年内销售额掉了大半。

这个结构解释了制药业很多外人看不懂的行为：

- **为什么罕见病药敢定价一针几百万美元？** 因为病人只有几千个。总账要平，单价就得离谱——这不是为贪婪辩护，这是这门「限时垄断 + 高固定成本」的算术在极端人口下的结果。

- **为什么没人愿意研发抗生素？** 新抗生素会被「留着备用」以防耐药，销量天然低，专利窗口内收不回成本。这是市场机制的真空地带，需要政府额外补贴。
- **为什么药厂拼命延长专利？** 窗口每多一年就是几十亿。围绕「晶型专利」「用途专利」的博弈，是律师和科学家并肩作战的第二战场。

大白话

大白话：新药定价不是『成本加利润』，是『限时独家生意的回本计划』。你觉得药片贵，是因为你在替整个漏斗买单，而且必须在专利钟走完之前买完。

所以，AI 能撼动这张账单吗

现在你能精确地回答这个问题了。如果 AI 能把「发现阶段」从五年压缩到一年——了不起，但发现阶段只占总成本的一小部分，大头在临床。如果 AI 能把临床失败率从九成降到五成——那才是真正的地震，因为**药价的大头是失败摊销，失败率减半意味着回本压力减半**。

可惜，正如第六章说的，临床失败的根源是靶点假设错——是生物学认知问题，不是计算问题。AI 到底有没有本事撬动那里？别急，我们先把 AI 真正擅长的事情讲清楚。下一章，AI 进场。

第九章 · AI 进场：先学会看分子

铺垫了八章，AI 终于可以登场了。现在你对研发流程已经有了完整的地图，我们可以精确地问：**这张地图上，哪几个格子是「计算瓶颈」，哪几个格子是「认知瓶颈」？** AI 只能在计算瓶颈的格子里发威。这一章先看它真正发力的地方——研发的前端，也就是第三、四章讲过的「找分子」和「雕分子」。

AI 在这里干的活，本质上是「学会了化学家的手感」

还记得第四章说的 QSAR 吗——用统计模型预测「结构改动会怎样改变性质」。AI（具体说，机器学习）做的就是这件事的加强版：不再依赖人手挑选的特征，而是让模型直接从海量「分子结构 → 实测性质」的数据里自己学规律。学出来的模型能干三件实事：

第一，虚拟筛选从「粗筛」变「精筛」。第三章说传统虚拟筛选能把「十万选一」压缩到「百选一」，靠的是简化的物理公式。机器学习模型更便宜更快，现在可以在几天内对**十亿级**的分子库打分。2020 年 MIT 的团队用这种方法从超过一亿个分子中筛出了抗生素 halicin——一个结构上和现有抗生素都不像、对多种耐药菌有效的分子。它原本是作为糖尿病药研发的，躺在库里几十年没人想到它能抗菌。这种「跨领域联想」正是机器学习的老本行：它不带人类的偏见。

第二，提前预测 ADMET，让失败前移。一个分子会不会被肝酶快速代谢、会不会堵住心脏的某个离子通道（hERG 毒性，一类著名的猝死风险），过去要合成出来、做了实验才知道。现在模型可以只在屏幕上就给出概率预测。预测不准没关系——只要准到能把「明显会死」的分子提前枪毙，就是在省钱：让分子死在一行代码里，而不是死在半年后的一次动物实验里。行业里管这个思路叫 fail fast（早点失败，便宜地失败）。

第三，生成式模型：从「筛选」到「设计」。这是最新的一步：不只在已有分子里挑，而是让模型直接「画」出满足条件的新分子——给定靶点口袋的形状，生成一批结构上能钻进去、性质上也达标的候选。第三章那个 10^{60} 的化学宇宙，第一次有了一个「生成器」而不仅是「筛子」。AI 制药公司英矽智能（Insilico Medicine）的抗纤维化药物 ISM001-055 是这条路线的标志性案例：靶点由 AI 从组学数据里提名，分子由生成模型设计，从立项到进入临床只用了约三十个月，而传统路径通常要四到六年。

大白话：AI 在研发前端的角色，是把『先试再知道』变成『先算再试』——筛选范围扩大了一千倍，失败的成本从『半年加一只猴子』降到『一小时加一度电』。

但要看清这三件事的共同前提

三个故事都很漂亮，但它们共享一个隐藏前提：**都有大量干净的「结构 → 结果」数据可以学**。分子的结构是数字化的，溶解度、结合力、代谢速率这些性质是几十年来上百万次实验积累下来的。机器学习在这里如鱼得水，因为这是一个「输入输出都有数字、样本量管够」的世界。

换句话说，AI 在药物前端的成功，恰恰是因为前端是整条链里**最像计算机科学问题**的一段：搜索空间巨大但规则明确，评价函数清晰可算。这和围棋、和蛋白质折叠（下一章的主角）是同一类问题。而链条往后走——动物、人体、疾病机制——数据变得稀疏、脏乱、充满没有记录的失败，规则的边界开始模糊。AI 的脚步到这里就慢下来了。这个边界第十一章详细拆。

不过在那之前，先看 AI 在制药领域最辉煌的一场胜仗——它没有发生在筛选或生成里，而是发生在「看懂蛋白质」这件人类卡了五十年的难事上。

第十章 · AlphaFold：一个真被解决的难题

如果要选一件事代表「AI 改变了生物学」，几乎所有人都会选 [AlphaFold](#)。2024 年诺贝尔化学奖一半颁给了它的两位主创（Demis Hassabis 和 John Jumper）。这一章讲清楚三件事：它解决的问题为什么难、它为什么能赢、以及——同样重要的——它赢了之后，制药业哪些事变了、哪些没变。

一个卡了五十年的问题

蛋白质是细胞里干活的机器，而机器的本事取决于形状。蛋白质出厂时是一条氨基酸「珠子链」，在细胞里它会自动折叠成特定的三维结构——结构决定功能：一个「口袋」的形状决定了什么样的小分子能钻进去，也就是决定了能不能为这颗蛋白质设计药。

问题在于：知道珠子链的顺序（基因测序便宜又成熟），**不等于知道它折成什么样**。1969 年 Cyrus Levinthal 算过一笔著名的账：一条普通的蛋白质链，理论上的折叠方式多到天文数字，就算每微秒试一种，试完也要比宇宙年龄还久——可细胞里的蛋白质几毫秒就折好了。这个悖论说明折叠不是穷举，而是被物理规律引导的。那么人能算出这个规律的结果吗？这就是 [蛋白质结构预测问题](#)：给一串珠子，算出三维形状。

五十年里，测结构主要靠实验：X 射线晶体学、冷冻电镜，一个结构动辄几个月几年、几十万美元。到 2020 年，人类已知的蛋白质序列有上亿条，实验测出的结构只有十几万个。计算预测做了几十年，精度始终差一口气。

AlphaFold 怎么赢的

2020 年，在两年一度的结构预测竞赛 [CASP](#) 上，DeepMind 的 AlphaFold2 把预测精度推到了和实验方法难分伯仲的水平——上一届它的最好成绩还不够用，这一届直接「问题基本解决」。评委的原话大意是：这个五十年的难题，可以认为被攻克了。

它赢在一套组合拳：用进化信息（同一蛋白质在不同物种里的序列差异，藏着「哪些珠子必须保持接触」的线索）、用注意力机制让模型自己找出链上相距遥远但折叠后贴在一起的珠子、再用物理约束收尾。两年后，DeepMind 和欧洲生物信息研究所干脆把预测好的结构免

费开放——覆盖几乎所有已知蛋白质，**两亿多个结构**。从此任何研究者查一个蛋白质的结构，像查字典一样。

对制药改变了什么

最直接的变化发生在第二章和第四章的交界处：过去很多靶点蛋白质「没有结构数据」，基于结构的药物设计无从下手，现在默认有了一份草稿。原本卡在「等晶体」上的项目可以马上开工；一些「不可成药」靶点的口袋，可以先在预测结构上打量一番。它把结构生物学的瓶颈从「测不出来」变成了「够不够用」。

以及，改变不了什么

这里是大多数新闻标题停下来的地方，但我们要继续问：AlphaFold 能直接给药吗？不能，而且差距比想象中大：

- **它给的是一张照片，不是一部电影。**蛋白质在细胞里是动态的，口袋会开会合、会变形。药的结合常常依赖那些转瞬即逝的「瞬态口袋」，静态快照抓不住。
- **它算的是「长什么样」，不是「贴得多牢」。**知道锁的形状，离知道哪把钥匙转得动、转多顺，还隔着一整套结合自由能的物理计算——那是另一个更难的问题。
- **它对「不成形」的部分无能为力。**人体里相当比例的蛋白质区域天生没有固定形状（内在无序区），而它们恰恰深度参与疾病信号传导。
- **最根本的：结构正确，不等于靶点正确。**就算把每个蛋白质的形状都印成图册，第二章那个核心问题——「关掉这个开关能治病吗」——一个字都没有被回答。锁的图册不能告诉你哪把锁锁着这个病。

大白话

大白话：AlphaFold 是把人体这台机器的零件图册从『缺 99% 页』补到『基本全了』——了不起。但图册是静态的，而机器是活的；看懂零件和修好机器之间，隔着一整本动力学和一整本病理学。

AlphaFold 是真正的胜利，但它的胜利边界恰好画出了 AI 在制药里的能力边界：**它攻克的是一个规则明确的物理问题，而新药研发最贵的失败，全在规则不明的生物学问题里。**下一

章，正面拆这条边界。

第十一章 · AI 帮不上忙的地方

第九章讲了 AI 的胜利，第十章讲了 AI 最辉煌的胜利。这一章必须讲硬币的另一面——不是因为要泼冷水，而是因为**看不清边界的乐观，和看不见价值的悲观，是同一种错误。**

边界一：AI 是数据的镜子，而生物数据又脏又偏

机器学习的第一定律：模型只能学到数据里有的东西。药物前端的数据很好——分子结构是精确数字，性质是标准化实验。但越往人体走，数据越糟糕：

- **阴性结果被吃掉了。** 科学界只发表成功的实验。一个分子「对某靶点无效」的数据，十次里有九次躺在某个实验室的抽屉里。可训练一个「预测什么会失败」的模型，恰恰最需要这些失败样本。用只记录成功的数据训练模型，就像用只记录胜局的棋谱学棋——学到的全是幸存者偏差。
- **实验不可重复。** 生物医学有一个公开的伤疤：多项大规模复现项目发现，已发表的明星研究里，能在独立实验室重复出来的比例可能不到一半。喂给 AI 的训练集里混着大量「事实上的假数据」，而 AI 没有火眼金睛。
- **临床数据量小得可怜。** 训练图像模型动辄几亿张图，而人类历史上所有三期临床试验加起来才几万个，每个还不可比（不同病、不同设计、不同终点）。想用机器学习「预测临床成败」，样本量先天不够。

边界二：最贵的问题不在算力，在认知

第六章说过：临床失败的一半死因是疗效不足，而疗效不足的根源是靶点假设错——我们对疾病的理解错了。这类错误的特点是：**它不在任何数据集里。**「 β -淀粉样蛋白可能不是阿尔茨海默的病因」这个认识，无法从既有数据中「训练」出来，因为既有数据全都是在这个假设下产生的。AI 擅长在规则明确的空间里搜索，而这里连「问题的形状」都还没画对。

这就是为什么 AI 制药公司喊了十年「成功率将大幅提升」，而行业的整体临床成功率并没有显著变化。一些被寄予厚望的「AI 发现药物」在临床阶段一样会死：Exscientia 和日本住友合作的 DSP-1181 曾以「全球首个 AI 设计进入临床的分子」刷屏（设计周期据称不到一

年)，后来一期之后悄悄停止开发——它证明了 AI 能大幅压缩「发现阶段的时间」，但对「进入临床之后能不能活」这个最贵的问题，没有给出额外保证。

边界三：伊鲁姆定律——越算越慢的制药业

半导体行业有摩尔定律：芯片性能每两年翻倍。制药业有一个名字故意反着写的伊鲁姆定律（Eroom's law，把 Moore 倒过来拼）：大约每过九年，「每十亿美元研发投入产出的新药数」就减半。过去几十年，计算能力涨了上万倍，DNA 测序成本掉了一百万倍，而新药研发的单位成本不降反升。

这个反直觉的趋势告诉我们一件事：**计算的进步从来没有自动传导成药的进步**。原因不难理解——研发瓶颈一直在迁移：筛选快了，瓶颈挪到优化；优化快了，瓶颈挪到毒理；毒理快了，瓶颈挪到临床；而临床的瓶颈是「人体的复杂度」和「疾病的异质性」，这两个都不服从摩尔定律。AI 会继续压低前端的成本，这真实且有价值；但只要我们没能更好地理解疾病本身，漏斗最陡的那一段就依然陡峭。

大白话

大白话：AI 目前解决的是『找到好钥匙』的速度问题；而制药业最贵的失败是『找错了门』。前者是工程，后者是科学。工程进步快，科学进步慢——这不是 AI 不行，是问题的种类不同。

那么，诚实的结论是什么

把九章到这一章合起来：AI 把「从靶点到候选分子」这段原本要四到六年的路，在最好的案例里压到了两三年；它让十亿级分子筛选成为日常；它解开了蛋白质结构。这些都是真金白银的进步。但它没有改变（至少还没有）临床成功率，没有回答「这个靶点对不对」，没有让伊鲁姆定律掉头。下一章——也是最后一章——我们把目光放到未来十年：哪些环节真的会被缩短、会以什么方式缩短，以及人类科学家在这个新分工里站在哪。

第十二章 · 接下来十年

最后一章，我们把前十一张拼图拼成一幅完整的画：未来十年，新药研发会变成什么样？AI 会站在哪里，人会站在哪里？回答的依据很简单——沿着这本书走下来的逻辑推：**每个环节按「它是计算瓶颈还是认知瓶颈」分类，前者 AI 会吃掉，后者 AI 只能帮衬。**

会明显变快的：前端闭环

第三、四章讲的「找分子、雕分子」正在连成一条自动化流水线，行业里叫闭环实验室或者「设计-合成-测试-分析」循环：AI 设计一批分子，机器人自动合成，自动化平台测性质，结果回灌给模型，模型据此设计下一批。一圈过去，模型比一圈前更聪明。这件事没有魔法——它就是第四章化学家手工迭代的机械化，但迭代速度从「每月一轮」变成「每周几轮」，而且每一次实验（包括失败的）都被完整记录下来，反哺给 AI。还记得第四章那个伏笔吗——「没有数字化的知识，AI 就学不到」？闭环实验室正是整个行业在补这门课：把化学家脑子里和抽屉里的隐性知识，变成模型能吃的数据。这个环节的时间从四五年压到两三年，是正在进行的事实，不是预测。

会慢慢变好的：靶点，靠「人」的数据

第六章和第十一章说，最贵的失败是靶点假设错，而 AI 无能为力，因为认知不在数据集里。但注意第二章埋的线索：人类遗传学证据能让成功率翻倍。过去二十年，生物银行（biobank，大规模人群基因 + 健康档案的数据库，如英国生物银行的五十万志愿者数据）正在成熟，单细胞测序让「疾病时每个细胞里发生了什么」变成可测量的数据。这意味着靶点发现正在从「生物学家的直觉 + 零散文献」变成「在全人群数据上做假设检验」——这是 AI 的主场。

这是未来十年最值得期待的结合点：**不是 AI 变聪明了，而是「人的数据」变多了**。更多带着真实结局的人群数据，会把「靶点假设」这个纯认知问题，一部分转化为可计算问题。别指望它让成功率翻倍再翻倍，但把二期死亡率从三分之二往下磨，是现实的期望。

大概率不会变的：人体的考场

三期临床的巨兽会继续存在。监管的铁律——「终点在人体验证」——是七十年事故史堆出来的，不会因为算法进步而松绑。「数字病人」「数字孪生」这类概念会有所帮助（比如用模型优化试验设计、挑选更可能获益的病人入组、用外部数据替代一部分对照组），但它们是**让考场更有效率，不是取消考场**。罕见病和基因疗法这类人群极小的领域，可能是例外：那里本来就没有传统三期的空间，监管正在学着接受新证据形式。

所以，那句宣传语该怎么拆

回到第一章结尾的问题：「AI 将让新药研发从十五年变成五年」，可信吗？

- **发现阶段的四到六年压到两三年**：可信，已在发生。
- **临床前的一两年压缩一部分**：部分可信，毒理预测和 fail-fast 在省钱。
- **临床阶段的六到八年大幅缩短**：不可信。人体的试验时间由生物学节奏（要等肿瘤进展、要等心血管事件发生）和监管要求决定，不由算力决定。

诚实的版本是：**十五年变成十到十二年，是大概率事件；变成五年，需要的不只是更好的 AI，而是对疾病的理解出现范式级的进步**。前者是工程趋势，后者是科学机遇，把两者混为一谈的宣传，就是在卖融资 PPT。

人站在哪

这本书里，人被机器替代的地方都是同一类：穷举、打分、重复迭代——把「在明确规则里搜索」的工作交给了机器。人留下的位置也清晰了：**提出假设、定义问题、为选择负责**。选哪个靶点，赌哪个适应症，二期数据暧昧时继续还是止损，面对「统计显著但病人无感」的结果怎么判断——这些决定没有评价函数可算，它们的依据是经验、品味和勇气，而责任终究要落在签名的人身上。

一颗药的诞生，过去是、未来仍然是人类好奇心和耐心跑赢生命复杂度的故事。AI 没有改变这个故事的主角，它只是——终于——给了主角一双好得多的眼睛。而从一万个候选里活下来的那一个，走到药店货架上时，身上背的不再只是九千九百九十九个兄弟的尸体，还有成千上万次代码里的「早点失败」。这，就是 AI 在这门古老行业里真实的价值。

一颗药的诞生