

# 基于专利视角下人工智能在合成生物学中的应用

相强宇<sup>1</sup>, 马丽丽<sup>2</sup>, 高婉莹<sup>1</sup>, 吴宗震<sup>1</sup>, 左锟澜<sup>1</sup>, 张璐<sup>1</sup>, 陈泽欣<sup>3</sup>, 李骏<sup>3</sup>, 刘欢<sup>1\*</sup>

1 中国科学技术大学, 安徽 合肥

2 中国科学院武汉文献情报中心, 湖北 武汉

3 国家知识产权局知识产权发展研究中心, 北京

相强宇, 马丽丽, 高婉莹, 吴宗震, 左锟澜, 张璐, 陈泽欣, 李骏, 刘欢. 基于专利视角下人工智能在合成生物学中的应用[J]. 微生物学报, 2025, 65(2): 828-845.

XIANG Qiangyu, MA Lili, GAO Wanying, WU Zongzhen, ZUO Kunlan, ZHANG Lu, CHEN Zexin, LI Jun, LIU Huan. Application of artificial intelligence in synthetic biology from the patent perspective[J]. *Acta Microbiologica Sinica*, 2025, 65(2): 828-845.

**摘要:**【目的】通过专利分析挖掘人工智能在合成生物学中的应用趋势与分布特征, 为该领域的技术创新、研发方向以及产业布局提供实践参考和理论支持。【方法】本研究概述了合成生物学的研究内容和方法, 借助专利网络进行深入挖掘, 并结合文献分析的角度, 探究人工智能在合成生物学领域的发展态势。【结果】检索并分析相关专利数据, 揭示了人工智能在合成生物学领域的申请和公开趋势, 以及主要国家和申请人的分布情况。进一步地, 从专利角度分析了生物合成基因簇、蛋白质结构分析以及转录因子结合位点等3个技术分支的发展情况。此外, 本研究还讨论了人工智能在合成生物学应用中面临的挑战, 并针对这些挑战提出了一些应对建议。【结论】本研究结果有助于深入了解人工智能在合成生物学领域的技术发展脉络, 为相关企业和科研机构的研究决策提供参考, 强调了人工智能技术在合成生物学领域的重要性, 以及其在推动该领域发展中的关键作用。同时, 本研究还针对进一步加强人工智能在合成生物学领域的研究和应用, 提出了一系列建议, 旨在为全球科技竞争态势下的国家科技情报体系建设, 特别是合成生物学的发展提供更多创新性的思路和技术支撑。

**关键词:** 合成生物学; 人工智能; 专利分析; 技术态势; 科技情报

资助项目: 合成生物学国家重点研发计划(2024YFA0917200); 中国科学院高质量数据池和数据产品服务体系建设项目(2019WQZX012); 国家知识产权局合成生物学关键核心技术专利分析研究项目(FX202309)

This work was supported by Synthetic Biology National Key Research and Development Program (2024YFA0917200), the Construction Project of High-quality Data Pool and Data Product Service System of Chinese Academy of Sciences (2019WQZX012), and the China National Intellectual Property Administration Synthetic Biology Program (FX202309).

\*Corresponding author. E-mail: liuhuan520@ustc.edu.cn

Received: 2024-09-17; Accepted: 2024-11-25; Published online: 2024-12-26

# Application of artificial intelligence in synthetic biology from the patent perspective

XIANG Qiangyu<sup>1</sup>, MA Lili<sup>2</sup>, GAO Wanying<sup>1</sup>, WU Zongzhen<sup>1</sup>, ZUO Kunlan<sup>1</sup>, ZHANG Lu<sup>1</sup>, CHEN Zexin<sup>3</sup>, LI Jun<sup>3</sup>, LIU Huan<sup>1\*</sup>

1 University of Science and Technology of China, Hefei, Anhui, China

2 National Science Library (Wuhan), Chinese Academy of Sciences, Wuhan, Hubei, China

3 Intellectual Property Development and Research Center, China National Intellectual Property Administration, Beijing, China

**Abstract:** [Objective] To analyze the application trends and distribution of artificial intelligence in synthetic biology from the patent perspective, providing practical insights and theoretical support for technological innovation, research and development (R&D) direction, and industrial layout of this field. [Methods] The paper presents a comprehensive overview of the research contents and methodologies of synthetic biology, and delves into the evolving landscape of artificial intelligence in synthetic biology by extensive patent network mining and literature analysis. [Results] Through thorough examination of pertinent patent data, this study unveils the application patterns and disclosure trends of artificial intelligence in synthetic biology alongside the major countries involved and key applicants. Furthermore, it analyzes the advancements in biosynthetic gene clusters, protein structure analysis, and transcription factor binding sites from the patent perspective. Additionally, this paper expounds the challenges confronting the integration of artificial intelligence into synthetic biology while offering recommendations to address them. [Conclusion] The findings presented herein offer valuable insights into understanding the technical developmental context surrounding artificial intelligence in synthetic biology while serving as a reference for relevant enterprises and research institutions when making R&D decisions. Moreover, this paper underscores the pivotal role played by artificial intelligence in advancing development of synthetic biology while emphasizing its significance. Simultaneously, it provides suggestions to further bolster research efforts on integrating artificial intelligence into synthetic biology with an aim to generate innovative ideas and technical support for constructing national science and technology information systems in global science and technology competition scenarios—particularly concerning advances in synthetic biology.

**Keywords:** synthetic biology; artificial intelligence; patent analysis; technical situation; science and technology information

生命是否可以通过研究其组成的每一个部分来理解? 如果科学家完全掌握了细胞内的酶如何催化化学反应, DNA 和 RNA 如何记录并传递遗传信息, 以及细胞的结构蛋白如何构建

细胞的结构等关于细胞的科学原理, 是否可以预测细胞的行为或者人工合成一个细胞<sup>[1]</sup>? 以机械论知识为指导的人工合成生命, 实际上就是有机生命的机械制造, 但机械论知识并不能涵

盖“生命密码本”的复杂性、生命系统的有机性以及生命形成的自创性，只有在机械论知识与有机体知识、生态知识、进化知识相结合条件下，人工合成生命的实践才能走上合理的道路<sup>[2-3]</sup>。

2016年3月，Venter等发表相关研究论文，设计并制造出了最简单的人工合成生命体<sup>[4-5]</sup>。自然科学各学科的研究对象在时空尺度上的差别与相连，既决定了学科间的差别，也决定了学科间的交叉。生物学在发展到以机理研究为主的“生命科学”阶段后，得益于学科交叉和技术创新，特别是在20世纪“分子生物学”和“基因组学”革命的基础上，通过引入工程学理念，形成了“合成生物学”，并迅速形成会聚研究的第三次革命<sup>[6]</sup>。

2015年之后，合成生物学进入发展的第4个阶段，即“设计-构建-测试-学习”(design-build-test-learn, DBTL)循环，通过反复使用来实现预期结果<sup>[7-13]</sup>，生物技术与信息技术的融合趋势愈加突出。合成生物学DBTL循环当中的学习(learn)即对应人工智能在合成生物学中的应用。随着人工智能技术的迅猛发展，其在合成生物学中的应用也日益广泛和深入。人工智能通过大数据分析、机器学习和高效计算等手段，极大地提升了合成生物学的研究效率 and 创新能力。2008年，Faulon等通过机器学习的方法，使用分子特征描述符来预测酶的活性，并基于原子结构预测蛋白质间的相互作用<sup>[14]</sup>；在调控元件领域，随着机器学习的进步，特别是数据集规模的增加，现在可以成功地构建从调控序列预测基因表达水平的模型<sup>[15]</sup>。

人工智能技术凭借其处理海量数据的持续学习能力和在未知领域中的智能探索能力，完美契合了当前合成生物学中工程化试错平台的需求。生命体的复杂性和未完全理解的特性给

研究带来了挑战，但人工智能可在许多方面找到突破口，显著提升合成生物学工程的效率。然而，合成生物学实验通常具有时间跨度长、成本高、DBTL循环次数有限等特点，这导致训练预测模型的数据极为有限，也为人工智能技术带来了新的挑战。

## 1 检索策略

以关键技术和专利领域分析方法，通过研讨、深度访谈、综合问卷对合成生物学发展情况做出综合判断，基于检索数据库对合成生物学领域相关关键技术梳理，将人工智能合成生物学的检索式制定为((((“合成生物学”OR“合成生物学技术”OR“合成生物系统”OR“合成生命”OR“合成组织”OR“合成细胞”OR“DNA人工合成”OR“DNA折纸术”OR“合成基因网络”OR“基因开关”OR“基因振荡器”OR“人工细胞”OR“人工基因线路”OR“DNA编码”OR“人工生物系统”OR“合成生物学基因电路”OR“菌种改造”OR“蛋白质工程”OR“代谢工程”OR“元件工程”OR“线路工程”OR“基因组工程”OR“细胞工程”)OR(((“synthetic biolog\*” OR “synthetic dna” OR “synthetic genom\*” OR “synthetic nucleotide” OR “synthetic promoter” OR “synthetic gene\* cluster”) NOT (“photosynthe\*”) OR (“synthetic mammalian gene\*” AND “mammalian cell”) NOT “photosynthe\*”) OR (“synthetic gene\*” NOT (“synthetic gener\*” OR “photosynthe\*”)) OR (“artificial gene\* network” OR (“artificial gene\* circuit\*” AND “biological system”)) NOT “gener\*”) OR (“artificial cell”) NOT (“cell\* telephone” OR “cell\* phone” OR “cell\* culture” OR “logic cell\*” OR “fuel cell\*” OR “battery cell\*” OR “load-cell\*” OR “geo-synthetic cell\*” OR “memory cell\*” OR “cellular network” OR “ram

cell\*" OR "rom cell\*" OR "maximum cell\*" OR "electrochemical cell\*" OR "solar cell\*")) OR (("synthetic cell") NOT ("cell\* telephone" OR "cell\* phone" OR "cell\* culture" OR "logic cell\*" OR "fuel cell\*" OR "battery cell\*" OR "load-cell\*" OR "geo-synthetic cell\*" OR "memory cell\*" OR "cellular network" OR "ram cell\*" OR "rom cell\*" OR "maximum cell\*" OR "electrochemical cell\*" OR "solar cell\*" OR "photosynthe\*")) OR (("artificial nucleic acid\*" OR "artificial nucleotide")) OR (("bio brick" OR "biobrick" OR "bio-brick")))) OR("合成生物学基因电路"OR"机器学习催化元件设计"OR"人工智能调控元件设计"OR"合成元件工程"OR"改造中间代谢途径"OR"基因组合成"OR"合成生物学元件设计"OR"生物元件优化"OR"DNA 元件设计"OR"合成基因线路"OR"蛋白质设计"OR"蛋白质结构预测"OR"蛋白质功能预测"OR"AlphaFold"OR"基因组挖掘"OR"全细胞模型"OR"合成生物传感器" OR"合成生物控制器"OR"转录因子识别"OR"结合位点预测"OR"基因注释"OR"基因线路预测" OR"基因线路设计"OR"代谢网络预测"OR"复杂微生物群落表征"OR"复合体亲和力预测"OR"生物合成路径"OR"生物合成途径优化"OR"多细胞系统构建"OR"生物数据建模"OR"生物合成基因簇"OR"生物元件定向进化"OR"定向进化蛋白质序列优化"OR"Cello 系统"OR"酿酒酵母启动子预测"OR"单点突变预测"OR"预测增强子活性"OR"基因组调控元件预测"OR"酶功能预测"OR"酶反应性质预测" OR "synthetic biological gene circuit" OR "catalytic element" OR "regulatory element" OR "synthetic element engineering" OR "modification of intermediate metabolic pathways" OR "Genome synthesis" OR "synthetic biology element design" OR "biological element optimization" OR "DNA

element design" OR "synthetic gene circuitry" OR "protein design" OR "protein structure prediction" OR "protein function prediction" OR "AlphaFold" OR "genome mining" OR "whole cell model" OR "synthetic biosensor" OR "synthetic biological controller" OR "transcription factor recognition" OR "binding site prediction" OR "gene annotation" OR "gene circuit prediction" OR "gene circuit design" OR "metabolic network prediction" OR "Complex microbial community Characterization" OR "Complex affinity prediction" OR "Biosynthetic pathways" OR "biosynthetic pathway optimization" OR "multicellular system construction" OR "biological data modeling" OR "biosynthetic gene clusters" OR "Directed evolution of biological elements" OR "Directed evolutionary protein sequence optimization" OR "Cello system" OR "Prediction of the promoter of *Saccharomyces cerevisiae*" OR "Single point mutation prediction" OR "Prediction of enhancer activity" OR "Prediction of genomic regulatory elements" OR "Enzyme function prediction " OR "Prediction of enzyme reaction properties" OR "EMOPEC" OR "BioAutoMata")) AND ("人工智能"OR"机器学习"OR"深度学习"OR"生成式模型"OR"计算模型"OR"生成对抗网络"OR"神经网络"OR"卷积神经网络"OR"集成学习"OR"循环神经网络"OR"算法系统"OR"合成生物学开放语言"OR"大语言模型"OR"图神经网络"OR"鲁棒性分析集成建模"OR"强化学习"OR"规则学习"OR"监督学习"OR"非监督学习"OR"变换模型"OR"逻辑回归模型"OR"决策树模型"OR"贝叶斯概率模型"OR"支持向量机"OR"隐马尔可夫模型" OR "Artificial intelligence" OR "machine learning" OR "deep learning" OR "generative model" OR

"computational model" OR "generative adversarial network" OR "neural network" OR "convolutional neural network" OR "ensemble learning" OR "recurrent neural network" OR "algorithmic system" OR "Synthetic Biology Open Language" OR "Large Language Model" OR "biosynthetic gene cluster" OR "Directed evolution of biological elements" OR "Ensemble Modeling For Robustness Analysis" OR "family-wide hallucination" OR "Transformer" OR "Logistic regression model" OR "decision tree model" OR "Bayesian probability model" OR "support vector machine" OR "Hidden Markov Model" OR "Fuzzy logic" OR "Logic programming" OR "Description logics" OR "Expert systems" OR "Logic programming" OR "Machine learning" OR "Bio-inspired approaches" OR "Classification regression trees" OR "Deep Learning" OR "Instance-based learning" OR "Latent representations" OR "Logical relational learning" OR "Machine learning" OR "Multi-task learning" OR "Neural networks" OR "Probabilistic graphical models" OR "Reinforcement learning" OR

"Rule learning" OR "Supervised learning" OR "Support vector machines" OR "Unsupervised learning" OR "Ontology engineering" OR "Probabilistic reasoning" OR "Search methods" OR "Computer vision" OR "Control methods" OR "Distributed artificial intelligence" OR "Knowledge representation reasoning" OR "Natural language processing" OR "Planning scheduling" OR "Predictive analytics" OR "Robotics" OR "Speech processing"), 检索数据库为 incoapat, 检索时间截至 2024 年 7 月。

## 2 专利态势分析

### 2.1 申请趋势

图 1 展示的人工智能合成生物学领域中相关专利申请量的发展趋势。通过申请趋势可以从宏观层面把握分析对象在各时期的专利申请热度变化。申请数量的统计范围是目前已公开的专利, 专利申请年份分布如图 1 所示。

技术起步期(2005–2011 年): 人工智能在合成生物学中应用专利的申请起始于 21 世纪早期, 但早期申请数量较少, 维持在百件以下。

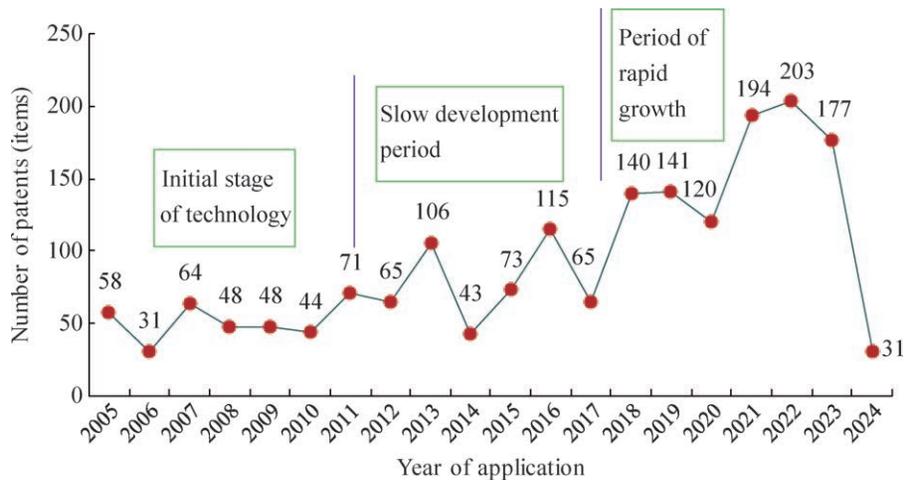


图1 专利申请趋势

Figure 1 Patent application trend.

缓慢发展期(2012–2017年): 21世纪被广泛认为是生物学的元世纪, 新兴技术的相互融合推动了人工智能与合成生物学交叉领域的快速发展, 相关专利申请数量呈现出显著增长的趋势。

快速增长期(2018–): 近年来, 随着新兴生物技术和计算机技术的快速发展, 人工智能与合成生物学交叉领域相关专利申请量有了大幅增加, 整体呈逐年上升趋势, 2022年人工智能在合成生物学中应用的相关技术专利申请量超200余件。

人工智能在合成生物学领域的专利申请量发展趋势, 反映了各时期专利申请热度的变化。在2019年之后, 人工智能技术迎来了一次巨大的发展浪潮, 这与快速增长期的合成生物学领域相关专利申请量大幅增加密切相关。这一趋势表明, 新兴生物技术和计算机技术的迅速发展, 尤其是人工智能技术的大暴发, 推动了人工智能与合成生物学交叉领域的研究和创新, 促进了专利申请数量的增长。

## 2.2 公开趋势

图2展示的人工智能在合成生物学领域中

应用的相关专利公开量的发展趋势。申请趋势分析和公开趋势分析密切相关。申请趋势反映了技术研发的活跃程度, 而公开趋势则显示了这些技术在市场中的接受度和应用情况。申请数量的增加可能表明相关技术正在快速发展, 而公开数量的变化则有助于评估这些技术的成熟度及其市场动态。通过结合这2种分析, 能够更全面地理解技术的发展脉络和未来的市场前景。

如图2所示, 专利公开年份分布图可知: 从2005年到2018年, 专利公开数量整体呈现缓慢上升趋势。根据SynBioBeta统计, 合成生物学公司融资规模, 从2009年的不足1亿美金到2020年相关投资资金规模逼近80亿美金, 可见合成生物学相关领域吸引了越来越多的投资。

2018年之后, 专利公开数量的增长更为显著, 在2023年相关专利公开快速增长至近300件。一般发明专利在申请后3–18个月公开, 实用新型专利和外观设计专利在申请后6个月左右公开, 因此2023年和2024年的相关专利情况仅作参考。

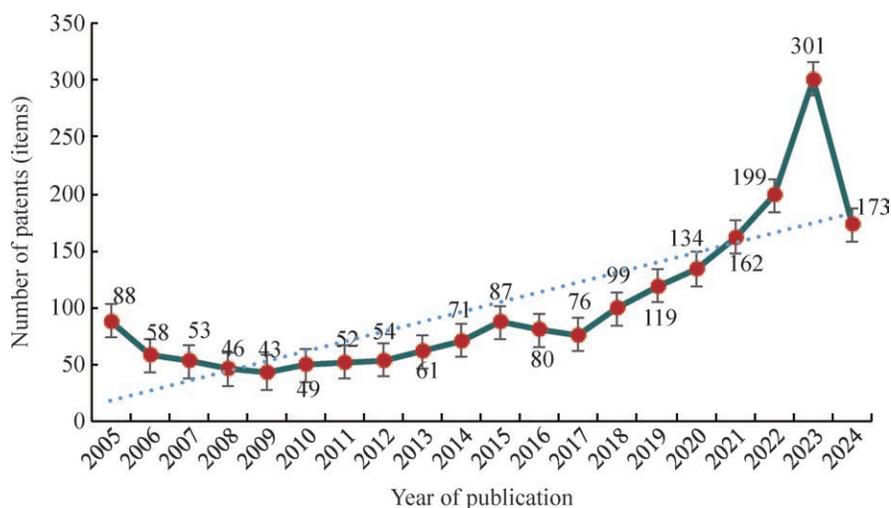


图2 专利公开趋势

Figure 2 Patent disclosure trend.

### 2.3 重要国家分布

对于在不同国家/组织受理专利数量的统计分析，不仅可以揭示专利技术的全球布局和趋势，还能为企业的市场策略、技术发展方向以及知识产权保护提供重要参考。这种统计分析有助于企业了解哪些国家或地区对其技术更感兴趣，从而有针对性地制定市场拓展计划。同时，还能帮助企业识别潜在的技术竞争对手，为技术研发和专利申请提供方向。

如图 3 所示，中国是受理人工智能在合成生物学中应用相关专利最多的国家，前 10 位除中国外依次是世界知识产权组织、欧洲专利局、美国、日本、加拿大、韩国、澳大利亚、印度以及法国。

其中具体的情况，中国占比 32.40%，居于首位；世界知识产权组织占比 21.40%，欧洲专利局占比 12.80%，两大区域性专利组织占比超过 34.00%，人工智能在合成生物学中的应用专利公开方面占据多数；美国以 12.70% 的占比位居第四。后六位分别为日本占比 7.00%、加拿大占比 4.80%、韩国占比 3.20%、澳大利亚占比 2.80%、印度占比 1.70%、法国占比 1.20%。

### 2.4 重要申请人分布

对重要专利申请人的统计分析是了解技术趋势、评估市场地位、识别合作伙伴与对手，

以及优化资源配置的关键手段，帮助企业洞察行业技术发展动态，评估自身在竞争中的位置，发现潜在的合作伙伴和竞争对手，从而制定更有针对性的技术战略和专利布局，推动技术创新和产业升级，同时降低创新风险。

如表 1 所示，中国有关人工智能在合成生物学中应用的重要专利申请人，分别为上海交通大学、中国科学院上海有机化学研究所、中国科学院南海海洋研究所、山东大学以及武汉大学。各个重要申请人/所属单位的研究领域有所区分，比如上海交通大学技术分布于基于深度学习的生物合成基因簇，蛋白质主链骨架设计与蛋白质结构生成方法以及基于多任务深度学习的 RBP 结合位点预测方法；中国科学院上海有机化学研究所则注重于智能预测微生物次级代谢产物分子结构的方法和系统，以及特定合成生物基因簇及其应用；中国科学院南海海洋研究所专利主要布局在特定基因簇序列合成及其预测；山东大学侧重于控制脂肽脂链长度改变的关键氨基酸位点及其突变体和应用，以及预测蛋白质-多肽结合位点的方法及系统；武汉大学专利主要集中在基于 Transformer-Encoder 和多尺度卷积神经网络的转录因子识别方法。

中国有关人工智能在合成生物学中的应用的前 10 位主要申请人，除了上述 5 位之外，依

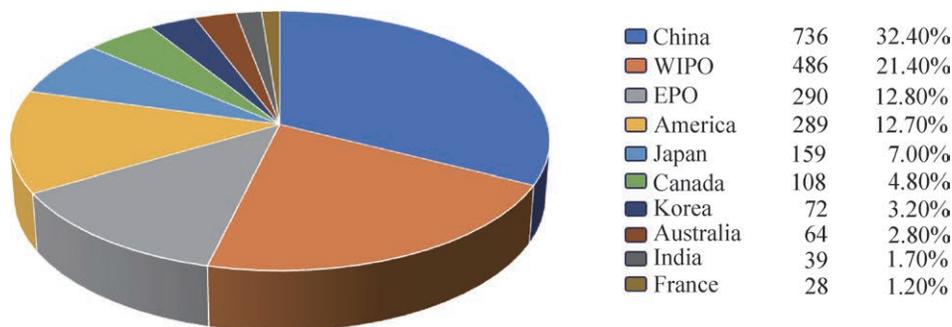
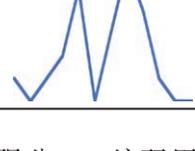


图3 专利受理国家/组织分布

Figure 3 Distribution of patent accepting countries/organizations.

表1 中国专利重要申请人

Table 1 Important applicant for Chinese patent

Applicant	Number of patents	Year of application	Application trend	Last five years patents	Technical distribution
Shanghai Jiao Tong University	54	2003–2023		24	Biosynthetic gene cluster based on deep learning; Protein backbone design and protein structure generation method; RBP binding site prediction method based on multi-task deep learning
Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences	41	2005–2024		3	Intelligent prediction of microbial secondary metabolites; Synthetic biological gene clusters and their applications
South China Sea Institute of Oceanology, Chinese Academy of Sciences	39	2010–2024		10	Sequence synthesis and prediction of specific gene clusters
Shandong University	23	2017–2024		16	Amino acid sites and their mutants and applications; Prediction of protein polypeptide binding sites
Wuhan University	23	2013–2024		14	Transcription factor recognition method based on Transformer-Encoder and multi-scale convolutional neural network

次分别是浙江大学、北京百度网讯科技有限公司、中国医学科学院医药生物技术研究所、浙江工业大学以及南京理工大学。

如表 2 所示, 美国有关人工智能在合成生物学中的应用, 前 5 位重要专利申请人分别为 DeepMind、IBM、明尼苏达大学、哈佛大学以及斯坦福大学。各个重要申请人/所属单位的研究领域有所区分, 比如 DeepMind 公司技术分布于基于神经网络的蛋白质结构预测及多序列对比; IBM 公司则注重于机器学习的蛋白质结构预测, 以及应用深度学习神经网络预测生物表型的概率分布; 明尼苏达大学专利主要布局在

编码甲基霉素的 DNA 合成基因簇编码的产物; 哈佛大学侧重于使用神经网络来自氨基酸序列的蛋白质结构以及鉴定推定的调节基因的调节元件; 斯坦福大学偏向用于鉴定和表达基因簇的系统和方法。

美国有关人工智能在合成生物学中的应用的前 10 位主要申请人, 除了上述 5 位之外, 依次分别是 MIN HE、Zymergen 公司、BEN SHEN、Jose Luis Riechmann 以及阿卜杜拉国王科技大学。

## 2.5 应用领域

如图 4 所示, 人工智能在合成生物学中应

表2 美国专利重要申请人

Table 2 Important applicant for US patent

Applicant	Number of patents	Year of application	Application trend	Last five years patents	Technical distribution
DeepMind	11	2019–2024		11	Protein structure prediction and multi-sequence comparison based on neural network
IBM	8	2015–2024		7	Protein structure prediction by machine learning; Applying deep learning neural networks to predict the probability distribution of biological phenotypes
University of Minnesota	6	2010–2024		3	The product of the DNA synthesis gene cluster encoding methylmycin and picomycin
Harvard University	5	2017–2024		4	Using neural networks to derive protein structures from amino acid sequences; Identification of regulatory elements of presumed regulatory genes
Stanford University	5	2017–2024		3	Systems and methods for identifying and expressing gene clusters

用前 10 的主要技术领域包括：C12N (微生物或酶)、C12P (合成目标化合物)、C07K (获得肽的基因工程方法)、G16B (生物信息学，特别适用于计算分子生物学中的遗传或蛋白质相关数据处理的信息与通信技术)、C12R (涉及微生物的 C12C 至 C12Q 小类)、A61K (医用配制品)、G06N (基于生物学模型的计算机系统)、C12Q (在微生物学方法或酶学方法中的条件反应控制)、C07H (有关基因工程的 DNA 或 RNA 的制备与纯化)、A61P (化合物或药物制剂的特定治疗活性)等领域。

就各应用领域的具体情况而言，其中中国在 C12N (微生物或酶)、C12P (合成目标化合

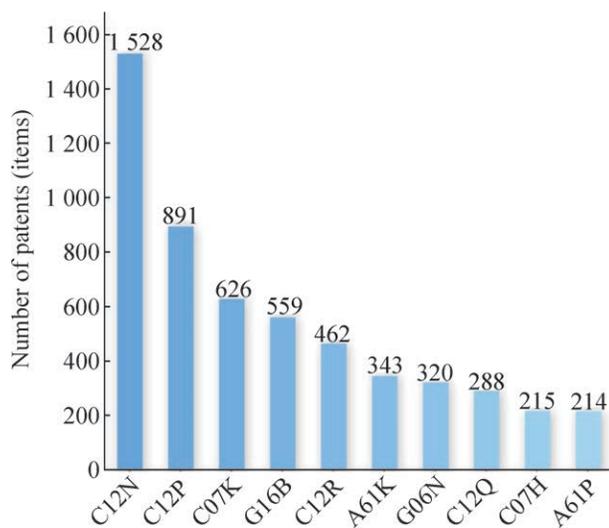


图4 专利应用领域

Figure 4 Patent application areas.

物)、C07K (获得肽的基因工程方法)、G16B (生物信息学, 特别适用于计算分子生物学中的遗传或蛋白质相关数据处理的信息与通信技术)等领域专利申请方面显著靠前, 美国则在 C12N (微生物或酶)、C12R (涉及微生物的 C12C 至 C12Q 小类)、G16B (生物信息学, 例如特别适用于计算分子生物学中的遗传或蛋白质相关数据处理的信息与通信技术)以及 C12Q (在微生物学方法或酶学方法中的条件反应控制)等领域专利布局较多。从中美两国不同的技术领域专利申请数量来看, 二者选择的技术分支存在差别。

## 2.6 公开类型

专利公开类型分析有助于评估技术成熟度、判断市场潜力以及监测竞争态势。通过分析不同类型专利的公开情况, 研究者和企业能够识别技术发展趋势、商业化潜力和市场需求, 从而优化研发与投资策略。

人工智能在合成生物学中应用的相关专利类型如图 5 所示。

发明申请占比 69.07%, 发明申请是指对新的、有技术水平的技术方案提出的专利申请。它要求所申请的发明在技术上具有创造性和实用性, 能够解决实际问题并产生技术效益。发明申请需要经过严格的审查过程, 包括初审和实审, 对创造性的要求较高, 因此更难获得授

权, 这表明合成生物学人工智能相关专利的申请具有高度的创新性; 发明授权占比 24.19%, 发明授权则是指经过审查并满足授权条件的发明申请被正式授予专利权的状态。这意味着发明人对其发明享有独占权, 可以阻止他人在专利保护范围内进行制造、使用、销售、进口等行为。

## 3 关键专利技术动向

根据人工智能在合成生物学中应用相关专利申请人来看, 中美两国是专利申请数量最多的 2 个国家, 从二者的专利申请技术分支来看有所区分。基于现有的专利内容, 通过手工标引聚类, 人工智能在合成生物学中应用的专利主要集中在生物合成基因簇、蛋白质结构分析及功能预测、转录因子识别及结合位点预测等三大类, 在下文中将结合专利内容以及文献角度对三大主题进行深入解读。

### 3.1 生物合成基因簇

天然产物(natural product, NP), 作为生物体在生长和适应环境中产生的非必需代谢物, 在医药、农业和食品行业中展现出显著的多功能性<sup>[16-17]</sup>。随着科技的飞速发展, 已能够详细解析这些化合物的生物合成机制。更重要的是, 生物信息学工具如今能够高度准确地预测基因

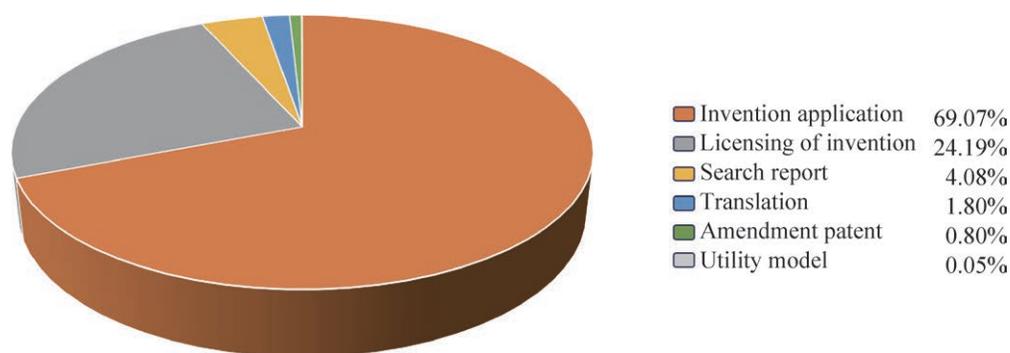


图5 专利公开类型

Figure 5 Types of patent disclosure.

与生物合成路径以及最终分子结构之间的关系，这为深入研究和利用天然产物提供了强大的支持。

对相关专利进行二次手工聚类发现，生物合成基因簇主要集中在生物合成基因、聚酮化合物、生物合成途径、链霉菌属、狭霉素以及次级代谢产物方面。如图 6 所示，生物合成基因簇相关专利申请公开起始于 2005 年，专利公开相较于专利申请有所滞后且数量也有所减少，因授权比政策性控制等环境影响，专利申请数量增长变缓后呈现上升并快速发展。

如图 7 所示，生物合成基因簇专利申请主要集中于中国，前 10 位主要申请人中，中国占据 6 位，分别是中国科学院南海海洋研究所、上海交通大学、中国科学院上海有机化学研究所、武汉大学、沈阳福洋医药科技有限公司以及山东大学；其余申请人分别为明治精工制药株式会社、美国惠氏公司、明尼苏达大学以及诺瓦克塔生物系统有限公司。

在生物合成基因方面，上海交通大学曾申请过一项专利号为 CN102516374A，用于四霉素生物合成的蛋白系统及其编码的基因簇，该发明涉及一种生物技术领域中用于四霉素生物合成的蛋白系统及其编码的基因簇，该发明提供的用于四霉素生物合成的蛋白系统及其编码的基因簇，能够实现生物合成四霉素，用于医药、工业、农业等诸多领域；杭州力文所生物科技有限公司(简称“力文所”)曾申请过一项热稳定性提升的 FDH 突变体及其参与的辅酶再生系统的专利，申请号为 CN116904410B，该发明利用力文所开发的基于深度学习算法的蛋白质设计平台对野生型甲酸脱氢酶的序列进行改进，并进行生物实验验证，人工智能计算结果和生物实验结果表明该反应体系的催化效率可以明显提高，适于工业推广。

自 20 世纪中叶以来，尽管天然产物的发现率有所下降，但基因组测序结果表明，绝大多数由基因编码的天然产物依然未知<sup>[18]</sup>。生物合

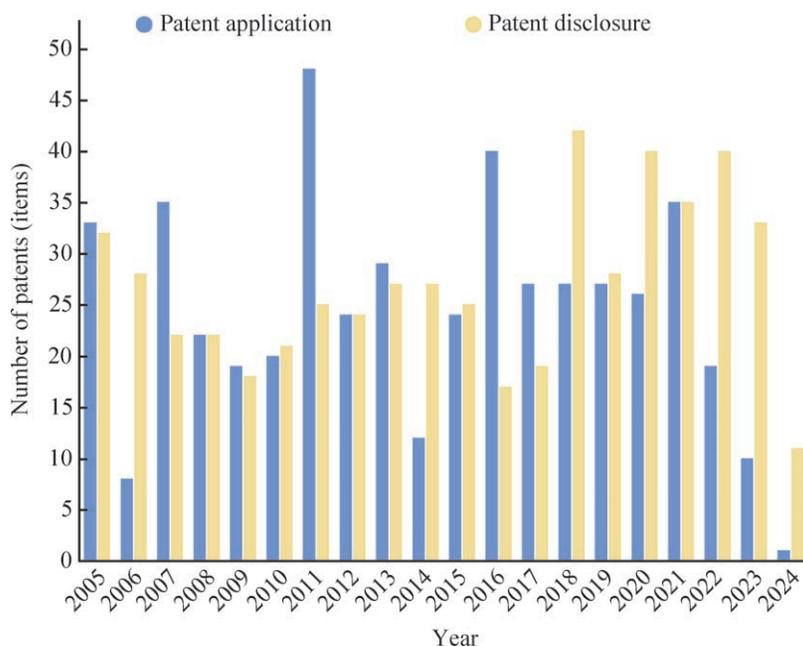


图6 生物合成基因簇申请公开趋势

Figure 6 Application trend for disclosure of biosynthetic gene clusters.

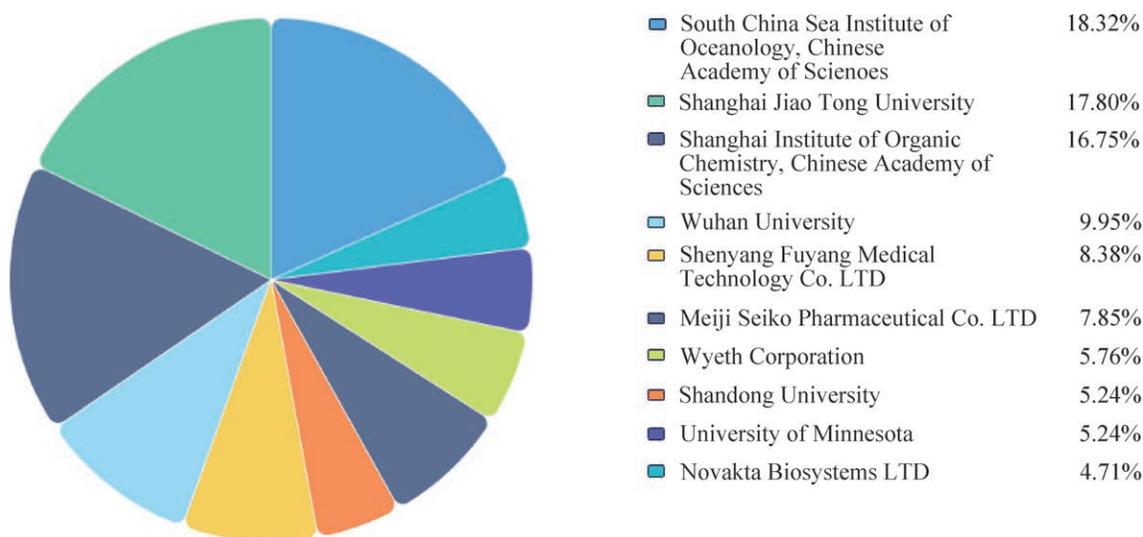


图7 生物合成基因簇专利主要申请人

Figure 7 Main applicants for biosynthetic gene cluster patents.

成基因簇是指基因组中紧密排列的基因，这些基因协同编码一系列酶和蛋白质，负责特定次级代谢产物(如抗生素、色素、毒素等)的合成<sup>[19]</sup>。生物合成基因簇通过协调基因依次表达，催化前体分子转化为最终产物。作为天然产物合成的重要工具，为合成生物学提供了丰富的元件库，在设计和构建生物合成系统中具有巨大潜力<sup>[20]</sup>。

### 3.2 蛋白质结构分析及功能预测

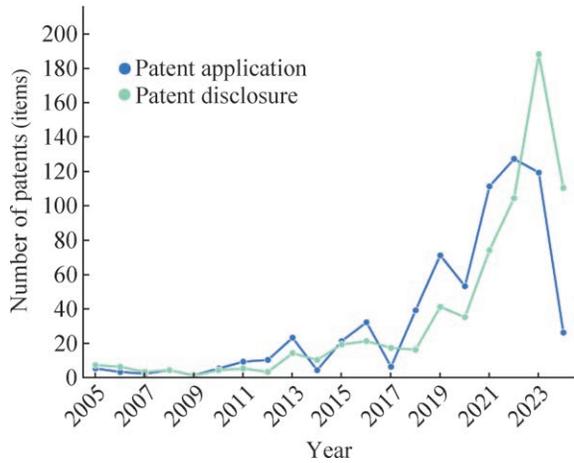
蛋白质是由氨基酸通过肽键连接而成的高分子有机化合物，作为生命体内的重要生物大分子，执行诸如代谢、信号传导、免疫等关键功能<sup>[21]</sup>。其三维结构决定了蛋白质的催化活性和与其他分子的相互作用，通过研究蛋白质结构，可以深入理解其功能机制，并用于设计具有特定功能的蛋白质<sup>[22-23]</sup>。蛋白质的三维结构可以为基因工程和蛋白质工程提供重要的信息，有助于在药物研发、治疗手段开发等领域取得突破。对本次检索标引的蛋白质结构与功能的相关专利进行二次聚类发现，其主要集中在蛋白质设计、蛋白质结构域、氨基酸序列、估计

距离以及结构参数等方面。

如图 8 所示，蛋白质结构分析及功能预测相关专利起始阶段申请与公开量较小，并且一般专利公开相较于专利申请有所滞后，且数量也有所减少。然而，在 2023 年，其相关专利的公开量超过了申请量。到 2022 年，这些专利逐渐进入公开阶段，而新的申请量可能暂时有所放缓。

如图 9 所示，蛋白质结构分析及功能预测相关专利申请人中，DeepMind 公司以绝对优势占据首位，其次是 ILLUMINA 公司；来自中国的申请人占据 7 位，分别是北京百度网讯科技有限公司、浙江工业大学、上海交通大学、南京理工大学、渊慧科技有限公司、四川大学以及中山大学。

DeepMind 公司在蛋白质结构及其功能预测方面表现突出，通常前沿技术的影响主要以学术论文为主，专利形式的成果影响具有滞后性，但其作为一种知识产权的保护方式，分析其利布局，有利于洞察其研发思路。DeepMind 公司专利申请的公开国家或组织，包括世界知识产



**图8** 蛋白质结构分析及功能预测相关专利申请公开趋势

Figure 8 Disclosure trend of patent applications related to protein structure analysis and function prediction.

权组织、美国、印度、加拿大和欧洲专利局。这些专利主要集中在使用各种机器学习方法，特别是神经网络技术，来预测或设计蛋白质的结构，包括使用氨基酸分子分布预测蛋白质突变的致病性、使用基于蛋白质结构嵌入的生成模型预测氨基酸序列、使用质量分数梯度的迭代蛋白质结构预测等。

基于 DeepMind 公司申请的专利，围绕其技术布局，构建其技术路线发展概况。

### 3.2.1 蛋白质结构预测

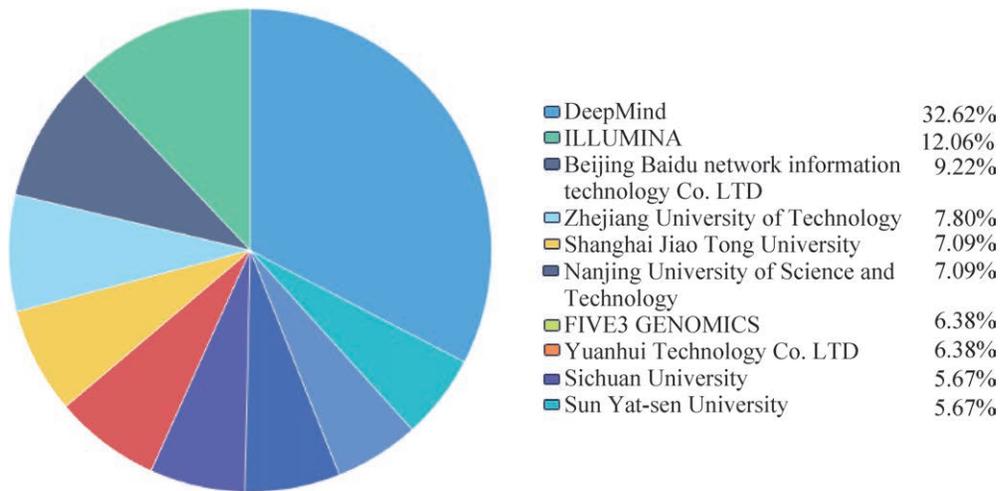
使用质量分数梯度的迭代蛋白质结构预测如专利号为 EP3821433B1：该方法通过迭代优化过程预测蛋白质的三维结构。蛋白质结构的准确预测是合成生物学的核心，可以帮助设计新的蛋白质功能，改进现有蛋白质的性能，以及理解蛋白质在细胞中的作用。

### 3.2.2 蛋白质突变致病性预测

使用氨基酸分布及蛋白质突变的致病性预测，专利号为 WO2024079204A1：这种方法利用机器学习模型来预测特定突变对蛋白质功能的影响，特别是其致病性。这对于合成生物学中设计安全且有效的基因编辑工具至关重要，有助于减少非目标效应和提高基因治疗的安全性。

### 3.2.3 生成模型预测氨基酸序列

基于蛋白质结构嵌入的生成模型预测蛋白质氨基酸序列，申请号为 US20240120022A1，以及针对蛋白质结构包埋条件的生成模型预测蛋白质氨基酸序列，申请号为 JP2024506535A。



**图9** 蛋白质结构分析及功能预测相关专利主要申请人

Figure 9 Main applicants for patents related to protein structure analysis and function prediction.

这些专利利用生成模型, 基于蛋白质结构信息来预测新的氨基酸序列, 推动了合成生物学中合成新的蛋白质和酶的能力。生成模型能够帮助设计具有特定功能的蛋白质, 从而满足不同的生物技术应用需求。

DeepMind 公司推出的 AlphaGo 标志着其技术布局的重要里程碑。AlphaGo 的成功展示了深度学习在复杂博弈中的潜力; AlphaFold 利用深度学习技术解决了蛋白质折叠预测的难题, 在专利方面的布局从 2019 年基于机器学习的蛋白质结构到 2020 年使用自注意神经网络从氨基酸序列预测蛋白质结构; 从 2021 年使用循环预测多次迭代的蛋白质结构到 2024 年生成模型预测氨基酸序列, 从蛋白质结构预测到突变致病性分析再到氨基酸序列生成的逐步演进, 形成了一个完整的技术链条。每一步技术的进展不仅推动了科学研究的深入, 还为合成生物学的应用提供了强有力的支持。

近年来, 随着生物信息学的发展, 神经网络和深度学习等人工智能技术在蛋白质结构分析和功能预测方面取得了显著进展。深度神经网络凭借强大的并行计算能力, 能够高效处理和解析大规模蛋白质序列与结构数据, 显著提升了预测速度和准确性。AlphaFold 通过训练数百万蛋白质序列及其已知结构, 展示了深度神经网络在处理海量数据方面的优势<sup>[24]</sup>。相比传统方法需要多个步骤和模型组合, 深度神经网络通过端到端学习直接从蛋白质序列预测三维结构<sup>[25]</sup>, 无需复杂的特征工程或规则设计, 简化了预测流程并提升了模型的一致性和统一性<sup>[26]</sup>。

### 3.3 转录因子识别及结合位点预测

转录因子是一类结合特定 DNA 序列以调控基因转录的蛋白质, 能够激活或抑制基因表达, 进而影响细胞功能和发育<sup>[27-28]</sup>。结合位点是转

录因子在 DNA 上的特定位点, 通常位于启动子或增强子区域, 通过与这些序列结合来调控 RNA 聚合酶的转录起始效率<sup>[29-30]</sup>。如图 10 所示, 转录因子识别及结合位点预测相关专利申请公开起始于 2005 年, 2013 年之前相关专利申请数量较少, 2018 年相关专利申请数量达到峰值, 2023 年相关专利公开数量达到历史最高水平, 可能因授权比政策性控制等环境影响, 专利申请数量快速增长后可能会呈现短暂下落。

如图 11 所示, 转录因子识别及结合位点预测相关专利主要申请人有美国先锋国际基因公司、明治精工制药株式会社、LifeMine Therapeutics、瑞士先正达公司、Zymergen、杜邦公司、SINVENT AS、斯坦福大学、加州大学以及奥本大学。美国先锋国际基因有限公司以 27.74% 的专利占据首位, 对其所属相关专利进行人工分析, 发现主要围绕在植物基因组编辑, 如专利申请号为 WO2024006802A1, 人工智能介导的基因组编辑方法和系统。

在转录因子识别及结合位点预测方面, LifeMin Therapeutics 公司曾申请过一项名为用于发现生物合成基因簇中嵌入的靶基因的方法和系统, 专利号为 AU2022383192A1, 该发明涉及基于计算机的方法和系统, 其用于使用比较基因组学技术和机器学习模型来鉴定与生物合成基因簇相关联的基因, 包括作为潜在治疗靶标的同源物的嵌入靶基因; 西安理工大学曾申请过一项基于深度学习区别编码和非编码核糖核酸的方法, 专利号为 CN113808671B, 该发明公开了一种基于深度学习区别编码和非编码核糖核酸的方法, 将其输入构建的卷积神经网络模型中进行训练得到预测模型, 将待区分的核酸序列输入到预测模型中得到区分结果, 解决了现有技术中存在的会受到不良基因注释和需要消耗大量的计算时间的问题。

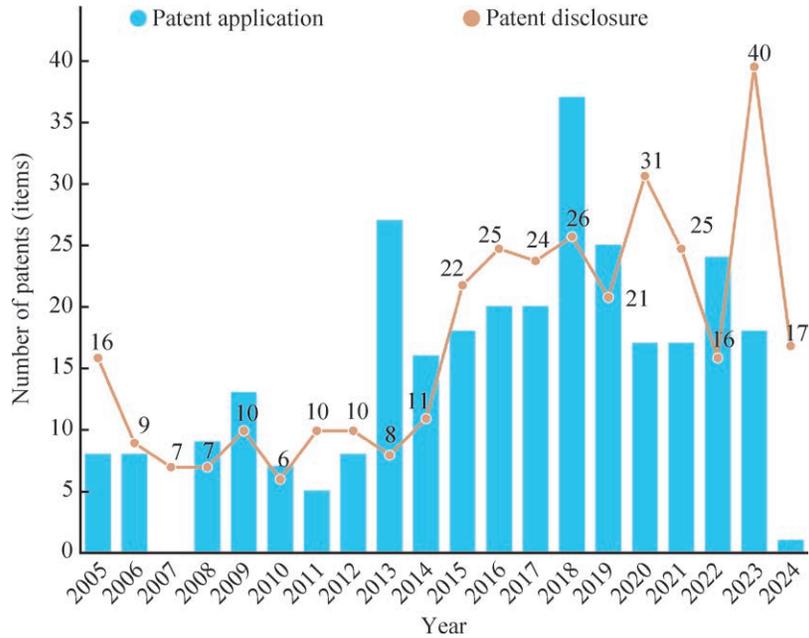


图10 转录因子识别及结合位点预测相关专利申请公开趋势

Figure 10 Transcription factor recognition and binding sites predict the disclosure trend of related patent applications.

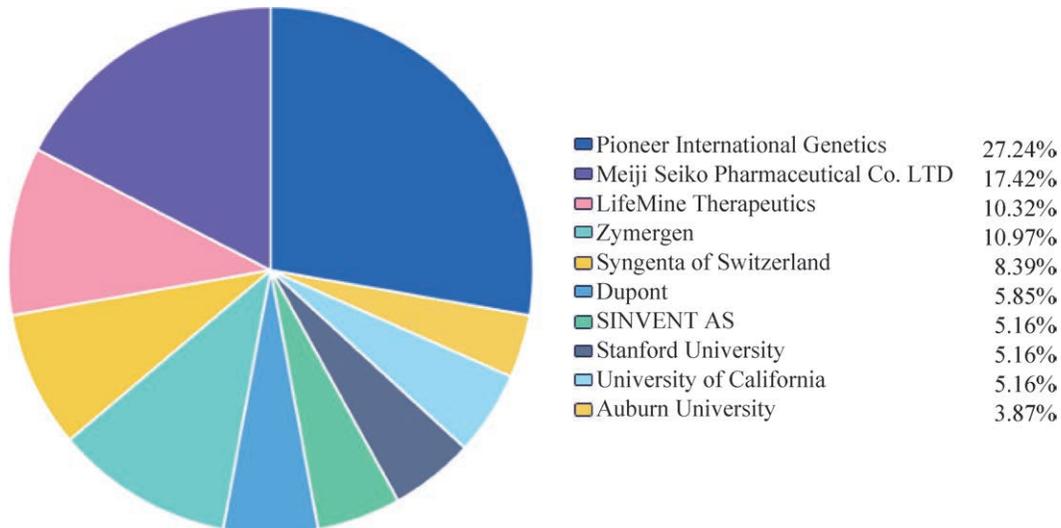


图11 转录因子识别及结合位点预测相关专利主要申请人

Figure 11 Main patent applicants for transcription factor recognition and binding site prediction.

转录因子识别及结合位点预测在基因调控研究、疾病研究、基因工程和合成生物学以及药物开发中具有重要意义，通过识别转录因子及其结合位点，可以深入了解基因表达调控机制，发现

与疾病相关的调控异常，为疾病的诊断和治疗提供新靶点，同时在基因工程中设计合成基因回路，实现精确的基因表达控制，并筛选和设计新药物，调控特定基因的表达以治疗相关疾病。

## 4 总结与展望

### 4.1 潜在风险及与挑战

人工智能与合成生物学同为前沿科学研究, 将二者进行交叉研究更是处于发展始端, 存在诸多潜在风险与挑战。在合成生物学中, 人工智能可能会被用来设计和优化具有特定功能的生物系统, 如产生药物、清洁环境的微生物等。这些生物体如果意外释放到环境中, 其可能与当地的生物种群竞争资源, 导致生态系统的失衡<sup>[31]</sup>。这可能导致原生物种的减少或灭绝, 以及对生态系统功能的影响; 并且合成生物体的基因可能会与野生生物的基因相互作用, 通过水平基因转移等方式传播到自然环境中, 产生未知的生态和健康影响。

人工智能在蛋白质结构解析和功能分析中的应用, 例如通过算法预测蛋白质的三维结构或其相互作用, 已经在生物学研究和药物开发领域展现出巨大的潜力。然而, 这样的应用也伴随着一些安全隐患: 蛋白质结构和功能的研究需要处理大量的敏感生物数据, 这可能包括独特的生物标志物和个人健康信息。若这些数据未能得到妥善保护, 可能会泄露个人健康信息, 引起隐私泄露和数据滥用问题; 结合人工智能和基因编辑技术, 可以实现对蛋白质结构和功能的精确编辑和调控。然而, 若这些技术被滥用, 可能导致创造出具有潜在危害的蛋白质, 例如抗药性病原体或毒素, 这些技术来创造新的生物制剂或制造有害的生物材料, 对公共健康和生态系统构成威胁; 通过人工智能设计的蛋白质可能具有突变或变异的趋势, 可能会逃逸实验室环境, 在自然界中引发物种生存风险或生态系统威胁; 人工智能算法可能存在误判蛋白质功能的情况, 导致设计或修改的蛋白质具有意外的生物学效应。这种误判可能会

导致不可预见的后果, 包括对人类或环境的不良影响。

### 4.2 建议参考

鉴于人工智能与合成生物学的交叉应用, 需要对涉及到的技术创新进行管理。从基础的生物元件、生物线路设计到人工智能在生物数据分析、模型预测等方面的应用, 完善对技术创新进行及时、全面的专利申请和保护。制定前瞻性专利布局策略, 考虑到人工智能与合成生物学融合应用的特点和发展趋势, 包括对关键技术要素、整体技术流程以及跨领域创新的专利保护策略。针对人工智能与合成生物学融合应用的技术创新, 建议定期进行技术监测和评估。

及时了解领域内的新进展和竞争态势, 为知识产权管理提供及时的参考和指导。(1) 建立统一的数据标准和共享机制, 使来自不同实验和研究的数据能够无缝整合。这包括制定数据格式标准、创建共享数据库以及开发能够自动化转换和整合数据的工具。通过提高数据的一致性和可用性, 可以显著提升人工智能模型的训练效果和预测精度。(2) 投资和发展智能试错技术, 通过机器学习和优化算法来减少实验次数和成本。这可以通过开发新的算法来更好地指导实验设计, 以及利用现有的数据和知识来预测和优化实验结果, 从而提高实验效率和成功率。(3) 加强对实验自动化设备的研发, 特别是针对合成生物学实验的特定需求, 开发更智能、更灵活的自动化系统。这些系统应该能够处理更复杂的实验流程, 并且具备更高的适应性和灵活性, 以应对不同类型的生物实验。

## 致谢

感谢中国科学技术大学知识产权研究院、TISC 中心、图书馆, 以及中国科学院武汉文献情报中心为本研究提供的数据服务。

## 作者贡献声明

相强宇：文稿撰写、方法论的设计以及数据的整理，包括形式分析与概念化、可视化等；马丽丽：主要参与方法论审查、验证以及前期的调研；高婉莹：主要参与审编、辅助分析以及前期的调研；吴宗震：主要参与审编以及前期的调研；左锷澜：主要参与审编以及前期的调研；张璐：主要参与论文返修的审编；陈泽欣：主要参与该文的背景分析及结果讨论；李骏：主要参与该文的背景分析及结果讨论；刘欢：主要参与包括完成概念的定义、方法论讨论与审编等工作。

## 作者利益冲突公开声明

作者声明不存在任何可能会影响本文所报告工作的已知经济利益或个人关系。

## 参考文献

- [1] WONG F, deLa FUENTE-NUNEZ C, COLLINS JJ. Leveraging artificial intelligence in the fight against infectious diseases[J]. *Science*, 2023, 381(6654): 164-170.
- [2] TANG TC, AN BL, HUANG YY, VASIKARAN S, WANG YY, JIANG XY, LU TK, ZHONG C. Materials design by synthetic biology[J]. *Nature Reviews Materials*, 2021, 6: 332-350.
- [3] RADIVOJEVIĆ T, COSTELLO Z, WORKMAN K, GARCIA MARTIN H. A machine learning automated recommendation tool for synthetic biology[J]. *Nature Communications*, 2020, 11(1): 4879.
- [4] HUTCHISON III CA, CHUANG RY, NOSKOV VN, ASSAD GN, DEERINCK TJ, ELLISMAN MH, GILL J, KANNAN K, KARAS BJ, MA L, PELLETIER JF, QI ZQ, RICHTER A, STRYCHALSKI EA, SUN LJ, SUZUKI Y, TSVETANOVA B, WISE KS, SMITH HO, GLASS JI, et al. Design and synthesis of a minimal bacterial genome[J]. *Science*, 2016, 351(6280): aad6253.
- [5] KANNAN K, TSVETANOVA B, CHUANG RY, NOSKOV VN, GARCIA NA, MA L, HUTCHISON III CA, HAMILTON OS, GLASS JI, MERRYMAN C, VENTER JC, GIBSON DG. One step engineering of the small-subunit ribosomal RNA using CRISPR/Cas9[J]. *Scientific Reports*, 2016, 6(1): 30714.
- [6] 赵国屏. 合成生物学: 开启生命科学“会聚”研究新时代[J]. *中国科学院院刊*, 2018, 33(11): 1135-1149. ZHAO GP. Synthetic biology: unsealing the convergence era of life science research[J]. *Bulletin of Chinese Academy of Sciences*, 2018, 33(11): 1135-1149 (in Chinese).
- [7] 陈大明, 周光明, 刘晓, 范月蕾, 王跃, 毛开云, 张学博, 熊燕. 从全球专利分析看合成生物学技术发展趋势[J]. *合成生物学*, 2020, 1(3): 372-384. CHEN DM, ZHOU GM, LIU X, FAN YL, WANG Y, MAO KY, ZHANG XB, XIONG Y. Analysis of global patents for the trend of synthetic biology inventions[J]. *Synthetic Biology Journal*, 2020, 1(3): 372-384 (in Chinese).
- [8] CARBONELL P, RADIVOJEVIC T, GARCÍA MARTÍN H. Opportunities at the intersection of synthetic biology, machine learning, and automation[J]. *ACS Synthetic Biology*, 2019, 8(7): 1474-1477.
- [9] 李敏, 林子杰, 廖文斌, 陈廷柏, 李坚强, 陈杰, 肖敏凤. 人工智能在合成生物学的应用[J]. *集成技术*, 2021, 10(5): 43-56. LI M, LIN ZJ, LIAO WB, CHEN TB, LI JQ, CHEN J, XIAO MF. Application of artificial intelligence in synthetic biology: a review[J]. *Journal of Integration Technology*, 2021, 10(5): 43-56 (in Chinese).
- [10] DIXON TA, CURACH NC, PRETORIUS IS. Bio-informational futures: the convergence of artificial intelligence and synthetic biology[J]. *EMBO Reports*, 2020, 21(3): e50036.
- [11] STANO P, KURUMA Y, DAMIANO L. Synthetic biology and (embodied) artificial intelligence: opportunities and challenges[J]. *Adaptive Behavior*, 2018, 26(1): 41-44.
- [12] ESLAMI M, ADLER A, CACERES RS, DUNN JG, KELLEY-LOUGHNANE N, VARALJAY VA, MARTIN HG. Artificial intelligence for synthetic biology[J]. *Communications of the ACM*, 2022, 65(5): 88-97.
- [13] 熊燕, 马雪晴, 陈大明, 刘晓, 赵国屏. 合成生物学赋能: 从学科发展到产业转化[J]. *中国科学院院刊*, 2024, 39(5): 851-861. XIONG Y, MA XQ, CHEN DM, LIU X, ZHAO GP. Synthetic biology enablement: from academic development to industrial transformation[J]. *Bulletin of Chinese Academy of Sciences*, 2024, 39(5): 851-861 (in Chinese).
- [14] FAULON JL, MISRA M, MARTIN S, KEN SL, SAPRA R. Genome scale enzyme-metabolite and drug-target interaction predictions using the signature molecular descriptor[J]. *Bioinformatics*, 2008, 24(2): 225-233.
- [15] de JONGH RPH, van DIJK ADJ, JIJSING MK, SCHAAP PJ, de RIDDER D. Designing eukaryotic gene expression regulation using machine learning[J]. *Trends in Biotechnology*, 2020, 38(2): 191-201.
- [16] HENG E, TAN L, TAY DWP, LIM YH, YANG LK, SEOW DCS, LEONG CY, NG V, NG SB, KANAGASUNDARAM Y, WONG FT, KODURU L. Cost-effective hybrid long-short read assembly delineates alternative GC-rich *Streptomyces* hosts for natural product discovery[J]. *Synthetic and Systems Biotechnology*, 2023, 8(2): 253-261.
- [17] BLIN K, SHAW S, STEINKE K, VILBRO R, ZIEMERT N, LEE SY, MEDEMA MH, WEBER T. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline[J]. *Nucleic Acids Research*,

- 2019, 47(W1): W81-W87.
- [18] NETT M, IKEDA H, MOORE BS. Genomic basis for natural product biosynthetic diversity in the actinomycetes[J]. *Natural Product Reports*, 2009, 26(11): 1362-1384.
- [19] ZHAO HM, MEDEMA MH. Standardization for natural product synthetic biology[J]. *Natural Product Reports*, 2016, 33(8): 920-924.
- [20] 刘海燕. 人工智能: 开启生物体系计算设计的新篇章[J]. *合成生物学*, 2023, 4(3): 419-421.  
LIU HY. Artificial intelligence: opening a new chapter in computational design of biological systems[J]. *Synthetic Biology Journal*, 2023, 4(3): 419-421 (in Chinese).
- [21] Widłak W. Protein structure and function[M]//*Molecular Biology: Not Only for Bioinformaticians*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013: 15-29.
- [22] MAKAM P, YAMIJALA SSRKC, TAO K, SHIMON LJW, EISENBERG DS, SAWAYA MR, WONG BM, GAZIT E. Non-proteinaceous hydrolase comprised of a phenylalanine metallo-supramolecular amyloid-like structure[J]. *Nature Catalysis*, 2019, 2(11): 977-985.
- [23] PEARCE EL, POFFENBERGER MC, CHANG CH, JONES RG. Fueling immunity: insights into metabolism and lymphocyte function[J]. *Science*, 2013, 342(6155): 1242-1245.
- [24] JUMPER J, EVANS R, PRITZEL A, GREEN T, FIGURNOV M, RONNEBERGER O, TUNYASUVUNAKOOL K, BATES R, ŽÍDEK A, POTAPENKO A, BRIDGLAND A, MEYER C, KOHL SAA, BALLARD AJ, COWIE A, ROMERA-PAREDES B, NIKOLOV S, JAIN R, ADLER J, BACK T, et al. Highly accurate protein structure prediction with AlphaFold[J]. *Nature*, 2021, 596(7873): 583-589.
- [25] TOWNSHEND RJL, BEDI R, SURIANA PA, DROR RO. End-to-end learning on 3D protein structure for interface prediction[EB/OL]. [2019-12-26]. <https://arxiv.org/abs/1807.01297v5>.
- [26] 陈志航, 季梦麟, 戚逸飞. 人工智能蛋白质结构设计算法研究进展[J]. *合成生物学*, 2023, 4(3): 464-487.  
CHEN ZH, JI ML, QI YF. Research progress of artificial intelligence in designing protein structures[J]. *Synthetic Biology Journal*, 2023, 4(3): 464-487 (in Chinese).
- [27] LAMBERT SA, JOLMA A, CAMPITELLI LF, DAS PK, YIN YM, ALBU M, CHEN XT, TAIPALE J, HUGHES TR, WEIRAUCH MT. The human transcription factors[J]. *Cell*, 2018, 172(4): 650-665.
- [28] TRUJILLO-OCHOA JL, KAZEMIAN M, AFZALI B. The role of transcription factors in shaping regulatory T cell identity[J]. *Nature Reviews Immunology*, 2023, 23(12): 842-856.
- [29] KUMAR DK, JONAS F, JANA T, BRODSKY S, CARMİ M, BARKAI N. Complementary strategies for directing *in vivo* transcription factor binding through DNA binding domains and intrinsically disordered regions[J]. *Molecular Cell*, 2023, 83(9): 1462-1473.e5.
- [30] MALIK S, ROEDER RG. Regulation of the RNA polymerase II pre-initiation complex by its associated coactivators[J]. *Nature Reviews Genetics*, 2023, 24(11): 767-782.
- [31] LIU H, GONG X, LIU Z, ZUO K, CHENG H. Medicine has no borders, health unites us all: Henry Norman Bethune as a pioneer, medical scientist, and internationalist[J]. *hLife*, 2024, 2(3): 97-99.