



## 治疗性抗体可开发性评估研究进展

Anthony Mackitz DZISOO<sup>1</sup>, 任丽萍<sup>2</sup>, 谢诗扬<sup>1</sup>, 周雨薇<sup>1</sup>, 黄健<sup>1\*</sup>

(1. 电子科技大学生命科学与技术学院 成都 610054; 2. 成都东软学院健康医疗科技学院 成都 611844)

**【摘要】**抗体在疾病防治中起着关键作用。其中,治疗性抗体已形成千亿美元的市场,适应症涵盖肿瘤、自身免疫及感染性疾病等,但也存在研发成本高、成功率低的困境。如何全面、合理、快速地对海量候选抗体进行可开发性评估,降低研发后期失败率,是治疗性抗体开发领域亟待解决的关键科技问题。该文综述了抗体可开发性评估实验研究及相关生物信息学研究的进展,总结了现有研究存在的问题,有助于相关领域学者未来研究的开展。

**关键词** 生物信息学; 可开发性; 机器学习; 治疗性抗体

**中图分类号** TP39; Q81 **文献标志码** A **doi**:10.12178/1001-0548.2021060

## Progress in Research on Evaluation of Developability of Therapeutic Antibody

Anthony Mackitz DZISOO<sup>1</sup>, REN Li-ping<sup>2</sup>, XIE Shi-yang<sup>1</sup>, ZHOU Yu-wei<sup>1</sup>, and HUANG Jian<sup>1\*</sup>

(1. School of Life Science and Technology, University of Electronic Science and Technology of China Chengdu 610054;

2. School of Healthcare Technology, Chengdu Neusoft University Chengdu 611844)

**Abstract** Antibody plays a key role in disease prevention, diagnosis, and treatment. Therapeutic antibodies have a market over 100 billion dollars with indications for various diseases such as cancers, autoimmune diseases, and infectious diseases. However, the trade suffers from high costs and low success rate. To assess the developability of thousands or even millions of antibody candidates comprehensively, reasonably, and quickly so as to decrease the late stage failure of antibody development is the key problem need to be solved in the field of therapeutic antibody development. This paper reviewed the advances in experimental and bioinformatics studies on evaluation of the developability of therapeutic antibodies. The problems of existing researches were also summarized. Therefore, it would benefit the relevant scholars in their further studies.

**Key words** bioinformatics; developability; machine learning; therapeutic antibody

抗体在重大疾病的预防、诊断与治疗中起着至关重要的作用。抗体相关基础与应用研究,既具有重大科学意义,又与国民经济和社会发展息息相关。科学家对抗体的研究已先后 7 次获得了诺贝尔奖。例如,1901 年,首枚诺贝尔生理学及医学奖,授予了德国科学家冯·贝林,表彰他开创了血清疗法,尤其是在治疗急性呼吸道传染病白喉中的成功应用。当前,新冠肺炎肆虐全球,康复患者血浆疗法临危受命,其实质是冯·贝林开创的抗体过继被动免疫疗法。又比如,已广泛使用的新冠病毒 IgM/IgG 抗体胶体金法快速检测试剂盒,临床实验中的托珠单抗治疗新冠肺炎重症患者,均依赖

于 1984 年荣获诺贝尔奖的杂交瘤单克隆抗体技术。

由于抗体既是生物医学科学研究不可或缺的工具,又是疾病防治的利器,所以抗体产业迅速发展壮大。尤其是抗体药物,已经给人类健康与生物医药产业带来了革命性变革,不少化学小分子不能作用的蛋白成为抗体治疗的高效药靶。据统计,美国 FDA 批准上市的抗体药迄今已多达 92 种,适应症范围覆盖了各类肿瘤、多种自身免疫性疾病、眼科及一些罕见病等多个方面,年销售额约 1000 亿美元。近十年来,全球最畅销药物前十名中,抗体药物占据半壁以上江山,近年更是屡拔头筹。例如,治疗多种自身免疫病的阿达木单抗,单个品种

收稿日期: 2020-12-24; 修回日期: 2021-03-20

基金项目: 国家自然科学基金(62071099)

作者简介: Anthony Mackitz Dzisoo(1986-), 男, 博士生, 主要从事生物信息学方面的研究。

通信作者: 黄健, E-mail: hj@uestc.edu.cn

的年销售额就已接近 200 亿美元; 淘选该单抗所用到的噬菌体展示技术也获得 2018 年诺贝尔奖。

目前, 全球正在进行 I、II 期临床试验的抗体药物超过 550 种, 另有 79 种已进入开发的最后阶段, 再创新高<sup>[1]</sup>。但是, 即便是人源或人源化抗体, 即便已进入到临床试验阶段, 最终能够成功开发上市的只有 15% 左右<sup>[2]</sup>。如何提高抗体开发的成功率, 降低开发后期失败带来的人力、物力、财力的浪费及时间上的耽搁是抗体产业界想要解决的重大难题。由于二代测序与噬菌体展示抗体技术的广泛应用<sup>[3-5]</sup>, 全球研究机构与制药公司临床前阶段的候选抗体数以万计。只有从中找出具有理想药效、安全性和药代动力学特性, 并且具有理想的理化特性, 满足生产、制剂工艺各项技术要求的先导抗体, 才能提高后期开发的成功率。上述过程就是抗体的可开发性 (developability) 评估<sup>[6]</sup>。目前, 化学小分子的成药性可用简单成熟的“里宾斯基 5 规则”来快速评估<sup>[7]</sup>, 但抗体大分子还没有类似评价标准。因此, 如何全面、合理、快速地对海量候选抗体进行可开发性评估, 是抗体药物开发领域亟待解决的关键科技问题。本文结合国内外抗体可开发性评估的研究现状, 聚焦抗体生物信息学尤其是可开发性预测研究的进展, 总结存在的问题, 提出可能的解决方案。

## 1 抗体可开发性的实验研究

抗体药物开发是资金与技术密集型行业, 充满挑战。一个抗体药从临床前到批准, 开发费用超过 10 亿美元<sup>[8]</sup>。为了降低后期研发失败的风险与损失, 提高成功率, 临床前阶段就要对候选抗体进行可开发性评估。广义上, 可开发性评估包括: 有效性与安全性、可生产性 (manufacturability)、系列理化特性等 3 个部分<sup>[6]</sup>。传统上, 开发者用于筛选抗体的首要标准是抗体与相应抗原的亲和力, 可用酶联免疫吸附、石英晶体微天平、表面等离子体共振、功能测试等实验进行检测<sup>[9]</sup>; 其次是抗体在动物实验中的药效、药代动力学特性及安全性。可生产性相关的影响因素主要包括抗体生产细胞系稳定性、表达水平、纯化回收率、放大生产性能、制剂稳定性、生产成本等, 是涉及细胞工程、抗体工程、发酵工程、药剂学等相关大量实验技术的系统工程。其中, HEK 滴定可测定抗体表达水平, 常用来反映可生产性<sup>[10]</sup>。抗体的有效性、安全性与可生产性主要取决于抗体本身的生物物理与生物化学

特性。因此, 对抗体生物物理与生物化学特性进行测试评估近年来迅速成为研究热点。相关实验方法主要有: 差示扫描荧光实验 (TmDSF), 反映抗体三维结构的稳定性; 亲和捕获自相互作用纳米颗粒光谱 (AC-SINS)、克隆自相互作用生物膜干涉 (CSI-BLI)、疏水相互作用色谱 (HIC)、直立单层色谱 (SMAC)、体积排除色谱 (SEC), 反映了抗体的粘度、聚集倾向、溶解度等; 交叉作用色谱 (CIC)、与常见抗原或杆状病毒颗粒 (BVP) 的酶联免疫吸附试验、多特异性试剂结合试验 (PSR), 反映了抗体结合特异性<sup>[9-10]</sup>。当前, 采用多种实验方法对抗体进行综合测试已成为一种趋势。例如, 文献 [10] 对 FDA 批准上市或已进入二期或三期临床试验的 137 个抗体进行了 12 种实验测定。他们给每项实验中表现最差的 10% 的抗体标记一项缺陷, 汇总结果显示, 65% 的已获批抗体无缺陷, 而二期临床试验中的抗体大多有一项或多项缺陷, 提示大规模综合实验测试有助于筛选到可开发性更高的候选抗体。然而, 实验测试费力、费时、费钱。制药巨头默克公司的最新研究结果认为, 要筛选到一个最佳先导抗体, 至少要按此实验流程测试 100~1000 个抗体分子<sup>[9]</sup>。因此, 要降低抗体药物研发成本, 加快研发速度, 就需要更多、更好、更全面的生物信息方法用于抗体可开发性评估。

## 2 抗体生物信息学研究进展

在生物信息学研究中, 构建专业的数据库和高质量的数据集是关键的基础性工作。没有相应的专业数据支持, 后续研究工作很难开展, 抗体生物信息学研究也不例外。1970 年, 免疫学家 Kabat 开风气之先, 构建了世界上第一个抗体数据库。目前, 全世界商业的、学术的抗体数据库已有几十个。根据抗体数据库收录范围, 可将其分为综合数据库与专业数据库。前者如 IMGT<sup>[11]</sup>, 既包括治疗性抗体, 又包括其他抗体, 甚至还包括 T 细胞受体、HLA 等其他免疫分子。后者如东南大学开发的 iCAN, 专注纳米抗体<sup>[12]</sup>; 牛津大学的 TheraSAbDab, 只收治疗性抗体<sup>[13]</sup>。抗体数据库研究的一个趋势是更加专业化。根据抗体数据库的主要内容, 可分为 3 类<sup>[14]</sup>。第一类是以抗体基因或蛋白质序列为主的数据库, 如 Kabat、iReceptor<sup>[15]</sup> 等。第二类是以抗体结构为主的数据库, 如 AbDb 等<sup>[16]</sup>。第三类是以抗体实验数据为主的数据库, 如 AB-BIND<sup>[17]</sup> 等。当前, 抗体数据库研究存在的最大问

题就是序列、结构、实验数据之间脱节。因此，数据整合是抗体数据库研究的另一发展趋势。例如，ABCD数据库尝试整合抗原与抗体的信息<sup>[18]</sup>，IMGT、abYsis<sup>[19]</sup>整合了序列与结构信息，SAbDab整合了结构与亲和力实验数据信息<sup>[20]</sup>。尽管如此，抗体领域数据整合的尝试仍不充分，缺乏对序列、结构、实验数据等3类数据的有机整合，特别是仍然只关注抗原抗体结合的实验数据，而忽略了抗体可开发性相关实验得到的抗体生物物理与生物化学特性的实验数据。

有了数据支撑，抗体生物信息分析软件纷纷问世，相关软件或网络服务已有近百种，广泛用于抗体理化性质计算、翻译后修饰位点及化学降解位点预测、序列比对分析、抗体三维空间结构建模、表位与对位预测、抗体设计、免疫原性预测、可开发性预测、抗体库分析与设计等。最近，文献[14]对此做了系统的总结与综述，这里不再枚举。值得指出的是，一些通用的生物信息预测分析软件，如分子对接软件、翻译后修饰位点预测软件等，也在抗体生物信息学研究中广泛应用，我国不少研究组对此皆有贡献。

### 3 抗体可开发性预测研究进展

抗体可开发性预测已经成为抗体生物信息学研究的重要组成部分。2000年，百时美施贵宝公司科学家 Lipper 等首次将可开发性这一术语用于化学药物开发中的先导化合物选择与优化。直到2012年，麻省理工学院的 Trout 课题组才将这一术语引入抗体药物研发领域<sup>[21]</sup>。他们与诺华公司合作，花了两年时间测定了12种抗体的长期稳定性数据。抗体的稳定性很大程度上取决于是否容易聚集，而这又主要取决于抗体的疏水相互作用及静电相互作用。所以，他们根据全长抗体的结构，计算了抗体的净电荷；又根据决定簇互补区的空间结构，计算了抗体的空间聚集属性 (spatial aggregation propensity, SAP)；然后用这二者定义了可开发性指数 (developability index, DI)。基于该指数，通过回归模型可定量预测抗体的聚集倾向与稳定性。该方法已集成到工业级药物设计分子模拟计算平台 Discovery Studio<sup>TM</sup> 中，成为业界进行可开发性评估的常用生物信息学方法。遗憾的是，这似乎进一步增加了抗体研发成本，因为购买该平台大概需要100万元人民币左右。虽然如此昂贵，但该方法基于抗体晶体结构或同源建模形成的理论结构进行计

算，其消耗计算资源、速度慢、准确率受理论预测模型精度的影响大，所以难以对筛选后的抗体文库进行高通量评估。因此，抗体可开发性预测研究的一个发展趋势是根据抗体序列而非空间结构来进行预测。

龙沙生物的 Stallwood 课题组采用了寡聚体检测与高效液相凝胶色谱两种实验方法检测了500多种抗体；经过主成分分析后选取了9种与亲水性、疏水性、静电、大小、氢键、侧链表面积等相关的氨基酸属性量表，采用了偏最小二乘、随机森林、高斯过程、AdaBoost 集成学习等方法建立模型将抗体分成高聚集或低聚集两类；最终 AdaBoost 方法获得了准确率84%的最佳预测模型<sup>[22]</sup>。该方法的优点是只需要抗体的氨基酸序列，速度快；缺点是不能定量，只能定性，且软件归属龙沙公司，收费使用。2017年，Adimab公司在902个抗体晶体结构数据基础上，采用随机森林方法建立了根据抗体序列直接预测其可变区每个氨基酸残基液相可及表面积 (SASA) 的模型；他们进一步研究了抗体疏水相互作用色谱滞留时间与 SASA 之间的关系，使用逻辑回归实现了只根据抗体氨基酸序列就能预测与抗体可开发性紧密相关的疏水相互作用<sup>[23]</sup>。

本课题组将文献[10]对137个抗体进行的12种实验测定结果作为金标准，采用支持向量机训练了仅根据抗体序列就能够预测其是否存在交叉或自身相互作用缺陷的免费网络服务 CISI，五折交叉检验中准确率达到88.2%，可快速高通量地从一个方面反映抗体的可开发性<sup>[24]</sup>。文献[25]进一步根据直立单层色谱 (SMAC)、体积排除色谱 (SEC)、疏水相互作用色谱 (HIC) 等反映抗体疏水相互作用倾向的实验数据，构建了基于抗体序列预测其是否存在疏水相互作用缺陷的免费网络服务 SSH，余一法检验中准确率达到91.2%。抗体疏水相互作用与其溶解度、聚集倾向等成药性质密切相关，因此 SSH 可快速高通量地从另一个方面反映抗体的可开发性。牛津大学 Deane 课题组对比分析了242种临床I期后抗体与大量二代测序人天然抗体的序列与结构，提出了治疗性抗体可开发性评估的5项指导意见：1) 决定簇互补区 (CDR) 的总长度不能太长或太短；2) CDR 附近表面疏水区不能过大或过小；3) CDR 附近正电区不能过大；4) CDR 附近负电区不能过大；5) 重链和轻链的净电荷要对称。他们给出了相应阈值，开发了免费的网络服务 TAP<sup>[26]</sup>。使用该服务虽然只需要输入待预测抗体

重链与轻链的可变区序列, 但 TAP 在后台会进行分子建模与多种结构生物信息学的计算分析, 需要花费较长的时间, 不大适用于大规模的抗体可开发性评估。

最近, 来自全球十余家制药公司的近二十位学者共同总结了治疗性抗体结构、异质性及可开发性评估的流程<sup>[6]</sup>。他们把抗体可开发性评估分为 3 大步骤。首先是序列分析, 发现有各种问题的或不必要的氨基酸残基或序列特征; 其次是广泛检测抗体的热稳定性、可溶性、粘度、疏水性等各种理化性质; 最后是强制降解实验<sup>[6]</sup>。

事实上, 各种生物信息学方法已经主导了上述评估流程的第一步。其后两个步骤, 虽然目前仍然以实验为主, 但也提示抗体可开发性预测还有很多工作需要进一步开展与深入。目前, 可开发性预测已经成为抗体生物信息学研究的前沿与热点, 但仍存在不少问题。例如: 1) 较多依赖结构, 影响计算速度, 难以胜任二代测序时代抗体库级可开发性筛选需求; 2) 集中在抗体个别的生物物理与生物化学属性, 如聚集属性, 缺乏对更多实验方法、更多理化特性及其相互关系的研究; 3) 用于学习与训练的数据有限, 评价中较多使用交叉验证或个案说明, 独立数据集或实验验证不足, 真实应用场景中的性能与效果有待进一步验证。

总之, 在抗体可开发性评估领域, 急需开展数据整合与专业数据库构建, 并在此基础上进行机器学习辅助的可开发性预测研究并实验验证。这些工作的开展, 不仅可能发现抗体可开发性背后从序列到结构到功能的科学规律, 其专业数据库及可开发性预测工具等平台也将有助于降低研发成本, 提高抗体药物的研发效率, 有助于包括新冠肺炎在内的各种重大新发传染病抗体药物的开发<sup>[27]</sup>, 具有重要的社会经济意义。

### 参 考 文 献

- [1] KAPLON H, MURALIDHARAN M, SCHNEIDER Z, et al. Antibodies to watch in 2020[J]. *MABs*, 2020, 12(1): 1703531.
- [2] CARTER P J, LAZAR G A. Next generation antibody drugs: Pursuit of the 'high-hanging fruit'[J]. *Nature Reviews Drug Discovery*, 2018, 17(3): 197-223.
- [3] GEORGIU G, IPPOLITO G C, BEAUSANG J, et al. The promise and challenge of high-throughput sequencing of the antibody repertoire[J]. *Nature Biotechnology*, 2014, 32(2): 158-168.
- [4] SOTO C, BOMBARDI R G, BRANCHIZIO A, et al. High frequency of shared clonotypes in human B cell receptor repertoires[J]. *Nature*, 2019, 566(7744): 398-402.
- [5] WINTER G. Harnessing evolution to make medicines (nobel lecture)[J]. *Angewandte Chemie International Edition*, 2019, 58(41): 14438-14445.
- [6] XU Ying-da, WANG Dong-dong, MASON B, et al. Structure, heterogeneity and developability assessment of therapeutic antibodies[J]. *MABs*, 2019, 11(2): 239-264.
- [7] LIPINSKI C A. Rule of five in 2015 and beyond: Target and ligand structural limitations, ligand chemistry structure and drug discovery project decisions[J]. *Advanced Drug Delivery Reviews*, 2016, 101: 34-41.
- [8] ELGUNDI Z, RESLAN M, CRUZ E, et al. The state-of-play and future of antibody therapeutics[J]. *Advanced Drug Delivery Reviews*, 2017, 122: 2-19.
- [9] BAILLY M, MIECZKOWSKI C, JUAN V, et al. Predicting antibody developability profiles through early stage discovery screening[J]. *MABs*, 2020, 12(1): 1743053.
- [10] JAIN T, SUN Ting-wan, DURAND S, et al. Biophysical properties of the clinical-stage antibody landscape[J]. *PNAS*, 2017, 114(5): 944-949.
- [11] LEFRANC M P, LEFRANC G. IMGT® and 30 years of immunoinformatics insight in antibody V and C domain structure and function[J]. *Antibodies*, 2019, 8(2): E29.
- [12] ZUO Jing, LI Jian, ZHANG Rong-xin, et al. Institute collection and analysis of Nanobodies (iCAN): A comprehensive database and analysis platform for nanobodies[J]. *BMC Genomics*, 2017, 18(1): 797.
- [13] RAYBOULD M I J, MARKS C, LEWIS A P, et al. TheraSAbDab: The therapeutic structural antibody database[J]. *Nucleic Acids Research*, 2020, 48(D1): D383-D388.
- [14] NORMAN R A, AMBROSETTI F, BONVIN A M J J, et al. Computational approaches to therapeutic antibody design: Established methods and emerging trends[J]. *Briefings in Bioinformatics*, 2020, 21(5): 1549-1567.
- [15] CORRIE B D, MARTHANDAN N, ZIMONJA B, et al. iReceptor: A platform for querying and analyzing antibody/B-cell and T-cell receptor repertoire data across federated repositories[J]. *Immunological Reviews*, 2018, 284(1): 24-41.
- [16] FERDOUS S, ACR M. AbDb: Antibody structure database-a database of PDB-derived antibody structures[J]. *Database*, 2018, 2018: 1-9.
- [17] SIRIN S, APGAR J R, BENNETT E M, et al. AB-Bind: Antibody binding mutational database for computational affinity predictions[J]. *Protein Science*, 2016, 25(2): 393-409.
- [18] LIMA W C, GASTEIGER E, MARCATILI P, et al. The ABCD database: A repository for chemically defined antibodies[J]. *Nucleic Acids Research*, 2020, 48(D1): 261-264.
- [19] SWINDELLS M B, PORTER C T, COUCH M, et al. abYsis: Integrated antibody sequence and structure management, analysis, and prediction[J]. *Journal of Molecular Biology*, 2017, 429(3): 356-364.
- [20] DUNBAR J, KRAWCZYK K, LEEM J, et al. SAbDab: The structural antibody database[J]. *Nucleic Acids*

- [Research](#), 2014, 42(D1): 1140-1146.
- [21] LAUER T M, AGRAWAL N J, CHENNAMSETTY N, et al. Developability index: A rapid in silico tool for the screening of antibody aggregation propensity[J]. [Journal of Pharmaceutical Sciences](#), 2012, 101(1): 102-115.
- [22] OBREZANOVA O, ARNELL A, CUESTA R G D L, et al. Aggregation risk prediction for antibodies and its application to biotherapeutic development[J]. [MAbs](#), 2015, 7(2): 352-363.
- [23] JAIN T, BOLAND T, LILOV A, et al. Prediction of delayed retention of antibodies in hydrophobic interaction chromatography from sequence using machine learning[J]. [Bioinformatics](#), 2017, 33(23): 3758-3766.
- [24] DZISOO A M, HE Bi-fang, KARIKARI R, et al. CISI: A tool for predicting cross-interaction or self-interaction of monoclonal antibodies using sequences[J]. [Interdisciplinary Sciences](#), 2019, 11(4): 691-697.
- [25] DZISOO A M, KANG Juan-juan, YAO Peng-cheng, et al. SSH: A tool for predicting hydrophobic interaction of monoclonal antibodies using sequences[J]. [BioMed Research International](#), 2020, 2020(2): 1-6.
- [26] RAYBOULD M I J, MARKS C, KRAWCZYK K, et al. Five computational developability guidelines for therapeutic antibody profiling[J]. [PNAS](#), 2019, 116(10): 4025-4030.
- [27] NING Lin, ABAGNA H B, JIANG Qian-hu, et al. Development and application of therapeutic antibodies against COVID-19[J]. [International Journal of Biological Sciences](#), 2021, 17(6): 1486-1497.

编辑 刘飞阳