

• 新药论坛 •

单克隆抗体药物序列分析的审评思考

齐连权, 白玉, 罗建辉*

(国家药品监督管理局药品审评中心, 北京 100022)

摘要: 改造用于重组表达生物技术药物宿主的遗传信息, 包括将外源基因转入宿主细胞之后, 对其 DNA、mRNA 和蛋白水平上的序列加以分析是产品/工艺开发和质量研究的重要部分。本文以单克隆抗体药物为例, 介绍了近年来新一代测序和液质联用等新技术在序列分析方面的应用。这些技术在成本、耗时和专业能力要求方面各不相同, 应按照逐步递进的原则, 在药物生命周期的不同阶段合理选择使用, 以提高研发的成功率并保证产品质量, 进而保证临床使用的疗效和安全性。

关键词: 序列分析; 单克隆抗体; 质谱; 新一代测序

中图分类号: Q816 文献标识码: A 文章编号: 0513-4870(2019)11-2113-05

Sequence analysis of therapeutic monoclonal antibody

QI Lian-quan, BAI Yu, LUO Jian-hui*

(Center for Drug Evaluation, National Medical Products Administration, Beijing 100022, China)

Abstract: Sequence analysis of DNA, mRNA and protein is an essential component of biologics or bioprocess development. Analysis of sequences at the DNA, mRNA, and protein levels after the transfer of the gene of interest into a host cell is an important part of quality control. This article reviews the application of new technologies such as next-generation sequencing and LC-MS/MS in biological drug development such as monoclonal antibodies. These techniques have different requirements in term of cost, handling time and expertise. Selecting an appropriate technique with a sound rationale at different stages of drug development will add to the success rate of research and development, and ensure product quality, thus ensuring the clinical efficacy and safety.

Key words: sequence analysis; monoclonal antibody; mass spectrometry; next generation sequencing

从大肠杆菌到酿酒酵母, 从中国仓鼠卵巢 (CHO) 细胞到转基因动物, 多种宿主已被成功用于生物技术药物生产。在用于药物生产前, 一般通过基因工程手段改造宿主的遗传信息以适于高效的药物生产, 对改造的遗传信息和药物序列分析是产品/工艺开发和质量研究的重要部分。近年来, 新一代测序和液质联用等新技术的应用在加快药物研发的同时, 也给技术审评带来新的挑战。本文以单克隆抗体药物为例, 介绍对重组 CHO 细胞的序列分析方面的审评思考, 为药物研发提供参考, 也为其他类型的生物技术药物, 如 Fc

融合蛋白、抗体药物偶联物 (ADC)、细胞因子和酶类等研究开发提供借鉴。

通常将抗体分子的重链和轻链基因克隆到带有筛选标记的载体上, 经转染 CHO 细胞后筛选得到目的基因稳定整合到 CHO 基因组上的克隆, 可用于稳定表达以生产单抗药物。经过多天的连续传代扩增和大规模培养, 分离、纯化分泌表达在培养液中的单抗分子, 可制成制剂用于医学预防或治疗。近年来新的单抗药物, 如敲除核心岩藻糖以改善抗体依赖性细胞介导的细胞毒性 (ADCC) 作用或采用含有阻遏子的启动子以实现诱导关闭表达等, 通常还涉及对宿主细胞的遗传改造, 以进一步提高产品的产量、安全性和疗效等。这类改造后的遗传稳定性也需要分析确认。

收稿日期: 2019-04-08; 修回日期: 2019-05-11.

*通讯作者 Tel: 86-10-85243072, E-mail: luojh@cde.org.cn

DOI: 10.16438/j.0513-4870.2019-0255

在申报临床和上市申请的药学资料中,药品研发机构通常能够提交在DNA、mRNA转录和蛋白表达水平上序列信息,包括拷贝数和整合位点等支持性数据。但对于多拷贝基因的突变情况,转录本的序列一致性以及低丰度的、表达水平的序列变异体等方面仍鲜有深入研究,期待这些新技术的广泛应用能够提高研发的效率和成功率,并进而有效保证产品质量。

1 序列分析的科学基础和技术要求现状

转染宿主细胞前,应通过测序、分析比对以确定构建好的重组载体的核酸序列和理论DNA序列一致^[1]。筛选得到的细胞株在传代过程中可能发生遗传突变,这将导致细胞群的微观异质性,而且突变株还可能获得选择优势而产生克隆替换。在包括疫苗、基因治疗和治疗性蛋白等生物技术药物的生产过程中,这些自发性突变对最终产品的质量构成风险。传代稳定性研究是用于确认细胞株适用于重组蛋白药物生产并能够保持产品质量一致性的重要内容。很多监管当局都发布了生物制品生产时需确保蛋白表达系统遗传稳定性的指导原则。例如《中华人民共和国药典》(2015年版,三部)“生物制品生产检定用动物细胞基质制备及检定规程”中对重组细胞基质的稳定性提出了明确要求,包括重组细胞的遗传稳定性、目的基因表达稳定性和目的产品持续生产的稳定性等,ICH Q5D、FDA^[2]、欧洲药典及WHO^[3]也有类似的指导原则。遗传稳定性中的一项关键评估是分析确认目标区域基因组的大小和核酸序列,包括确认启动子和其他功能元件。此类评估至少应该在主细胞库和生产终末细胞库上开展^[4]。

对于生物类似药,《生物类似药研发与评价技术指导原则》(试行)也明确要求,候选药的氨基酸序列原则上应与参照药相同,并要求在药学研究与评价时,“对于氨基酸序列测定的比对试验研究,可以与已知的参照药序列直接进行比对”。

通常借助测序分析技术手段,从DNA、RNA和氨基酸序列3个不同水平上确认遗传信息和表达产物的正确性以及连续生产培养过程中的稳定性。在分子水平,mRNA样品中发现的突变可能来自DNA模板、转录错误、剪切错误或转录后修饰^[5]。未合理优化密码子偏性的外源基因、tRNA错载或培养基中部分种类的氨基酸耗尽都可能导致在遗传信息正确的情况下蛋白合成时氨基酸的错误掺入^[6],因此也需要采用蛋白测序或质谱等技术手段在氨基酸水平进行序列分析以评估序列变异体对产品质量的影响。

在细胞株开发初期直至商业化生产过程中上游培养工艺的变更,都需要谨慎细致地研究,包括定性和定

量的研究遗传信息突变及氨基酸错误掺入导致的序列变异体,关注并降低其潜在的稳定性、生物学功能和免疫原性等方面的质量风险^[7]。

2 DNA水平的序列分析和技术考量

DNA水平上的遗传稳定性包括目的基因的拷贝数、整合位点以及序列信息等。常可采用qPCR方法检测外源基因拷贝数来评估遗传稳定性,但该方法有包括参比品和待检测样品的扩增效率差异等技术局限^[8]。另一项数字PCR(digital PCR, dPCR)方法也可用于确认拷贝数^[9]。通过分隔出单个模板分子的PCR扩增,直接读取阳性和阴性信号计数,通过泊松分布转换得到目标基因的绝对拷贝数,数字PCR可实现不依赖于标准曲线和参照样本的准确、绝对定量。荧光原位杂交技术(FISH)可用于确认目的基因的整合位点信息^[10]。

DNA印迹(Southern blot)、限制性片段长度多态性(RFLP)和Sanger测序能够进一步提供目的基因序列信息。但和新一代测序(next generation sequencing, NGS^[11]),也称为高通量测序,high throughput sequencing, HTS)相比,其灵敏度和信息量则稍显不足。Sanger法基于被扩增的混合物测序,一般能检测到低至15%~20%的变异体^[12],而NGS则可以检测到低至0.1%的变异体,甚至更低^[13]。而且NGS还可以提供除目的基因外更大范围的序列信息,甚至全基因组测序,可用于评估更多的遗传信息改变。虽然目前监管当局尚未要求采用NGS开展序列分析,但随着其更广泛的应用及显示出的巨大优势,相信未来会被越来越多的研发机构采用^[14]。当然,伴随着NGS提供的更高灵敏度,在确认变异体时也应关注其潜在的高序列错误率^[15],同时该方法也应该经过适当的方法学验证。

有研究团队采用称为“靶向位点扩增”(targeted locus amplification, TLA)和NGS技术快速完成了克隆筛选,用比传统手段更短的时间排除了出现点突变、基因重排及非单克隆来源的细胞株^[16]。另一项研究在细胞株开发早期使用NGS对扩增子测序,发现并排除了4个临床项目中的序列变异克隆,安全地加快了研发进度^[17]。

3 RNA水平的序列分析和技术考量

部分多拷贝基因并非以一致的效率转录成mRNA,而且在转录过程中也有可能发生序列变异,因此同样有必要在RNA水平上进行单抗药物的序列分析。通常可以采用RNA印迹(Northern blot)、或将mRNA反转录成cDNA测序研究其序列变化情况。RNA-seq,即mRNA水平上的NGS,可以实现对低水平基因突变的研究。

在一项对长期培养的、表达单抗的CHO细胞突变发生情况的研究中,作者分析了包括重链、轻链和二氢叶酸还原酶(DHFR)3个外源基因和看家基因GAPDH的突变情况,结果发现随着传代次数增加及存在氨甲蝶呤(methotrexate, MTX)的情形下,外源基因的突变率明显增加,含80 nmol·L⁻¹ MTX的100PDLs样品的突变率高于0.5%^[18]。而应用NGS的核糖体足迹画像(ribosome footprint profiling, RiboSeq)方法则通过提供全基因组范围的翻译信息,为利用CHO细胞生产抗体药物在工艺开发和细胞株工程方面提供更多解决方案^[19]。而另一项针对经多代次培养的、稳定表达重组蛋白的克隆和不稳定表达克隆的转录组分析发现,与稳定表达细胞株相比,不稳定细胞株的早期代次和晚期代次有更多的基因表达水平出现较大差异。同时,不稳定细胞株的启动子甲基化和基因拷贝数则未发生显著变化^[20]。

4 氨基酸水平的序列分析和技术考量

即使在mRNA序列正确的情况下,由于培养基中氨基酸耗尽等原因,氨基酸错误掺入也可产生序列变异体^[21]。因此一级氨基酸序列分析是结构确证的重要资料之一,通常可采用液相色谱(LC)结合质谱(MS)或串联质谱(MS/MS)完成样品的分离和分析,即LC-MS或LC-MS/MS,利用不同的样品前处理方式,可以获得包括抗体蛋白完整分子量、脱糖分子量、还原分子量、还原脱糖分子量、肽段分子量、N/C末端异质性、糖基化、糖化、脱酰胺、蛋氨酸氧化、二硫键/三硫键和游离巯基及氨基酸排列顺序和序列变异体等多层次的全面结构信息^[22]。另一种离子化质谱技术-基质辅助激光解吸电离(MALDI)可有效电离分子量大于500 kDa的分子,得到单电荷离子,可用于完整分子量、翻译后修饰和修饰位点等研究。当然,序列变异体通常比例较低,且理化性质和正常序列抗体分子差异不大,需要采用高分辨率和高精度质量的液质联用技术开展定性定量分析,同时相应方法应通过适当的性能确认(performance qualification)。

5 蛋白结构的全面表征分析

在氨基酸序列分析的基础上,从翻译后修饰(糖基化、糖化、硫化、磷酸化)、变异体(分子大小、电荷异质)、聚合体/降解物、氨基酸的转化物(氧化、脱酰胺、环化等)等方面,研发结构应对抗体蛋白结构进行全方位的解析研究。部分翻译后修饰如糖化、N-寡糖糖型以及脱酰胺、氧化等虽可采用相同技术途径加以分析,且也和抗体分子结构、活性关系密切,但其更多和上下游生产工艺及储存条件有关。

对于分子大小变异体如聚集体、片段和非糖基

化重链等可采用传统的SEC-HPLC、SDS-PAGE或CE-SDS进行定性和定量分析;对于电荷变异体可以采用IEX-HPLC、cIEF或CZE实现分离、分析;对于抗体和其降解片段在亲疏水性方面的差异,可以尝试HILIC、HIC-HPLC或RP-HPLC实现分离、分析。序列变异体和正常抗体分子由于只有1个或几个氨基酸的差异,需采用更高分辨率的多属性方法(multi-attribute method, MAM)检测分析^[23],通常需结合适当的样品前处理或预先富集^[24]。在一项采用多种分析技术头对头研究雷珠单抗的生物类似药的研究中,SEC-HPLC、CE-SDS和活性分析均未发现差异,而CEX-HPLC和CZE发现其酸性峰和原研药存在差异,完整分子和亚单位的质谱数据分析显示,有约6%~9%样品的轻链分子量增加27 Da,进一步的二级质谱确认了轻链上多个位点发生了丝氨酸替换成了天冬酰胺的序列变异^[25]。雷珠单抗是由大肠杆菌表达产生的,而在CHO和NS0细胞中类似氨基酸替换也已被发现^[26]。

研发机构可以使用自上而下(top down)的方法对完整抗体分子进行表征,还可以富集制备糖化、氧化、电荷变异体和分子大小变异体等各个组分,解析其结构和生物活性的关系。可以采用强制降解样品以更高效制备并确定和评价这些产品相关物质和/或杂质与结构/功能关系,也可以为设定产品的长期保存条件提供依据。这些结果可用于确定药物分子的关键质量属性(CQA)并开发适当的工艺和分析控制策略。直接片段化后分析可以进一步提供一级结构、糖基化和异构体等信息。

多种技术手段能够降低产品的复杂性,实现自中而下(middle down)或亚单位的抗体分子表征^[27]。例如可以用PNGase F切除N-寡糖,以获得去除糖基化的抗体分子,测定其精确分子量,并推断非糖基化重链的比例;还可以同时利用还原剂分离得到单独的轻链及重链分子,测定其精确分子量,以发现低丰度的序列变异体;抗体蛋白可以被木瓜蛋白酶和IdeS等特异性的肽链内切酶降解成片段,结合酶解处理的亚单位水平的LC-MS可以得到100%覆盖率的肽指纹图谱。

采用自下而上(bottom up)的分析策略虽耗时较长,但可以提供更丰富、精细的结构信息。经变性、还原和烷基化后的蛋白样品,通过胰蛋白酶、Lys-C或Asp-N等肽链内切酶酶切后采用LC-MS/MS进行一级结构表征,RP-HPLC肽图分析联合质谱(MS)、串联质谱(MS/MS)分析确定抗体药物由cDNA推测的氨基酸序列。常规的技术路线是首先分析所有未修饰肽段的质量并与MS分析预测的质量相比较,之后再通过肽段的二级MS/MS碎片进一步确认序列信息。肽图

分析联合MS和MS/MS分析可用于肽段含量确定,通过MS提取离子色谱图的峰面积积分定量测定修饰氨基酸的水平,其结果可以表示为所有该肽段峰面积的百分比。

通常可能需要分析多种酶切肽图以获得完整的序列覆盖,并鉴别所有氨基酸修饰。赖氨酸残基糖化与蛋白质的Lys-C酶切相互干扰,需要其他类型的蛋白酶提供互补数据。研发机构应分别在一级MS和二级MS/MS水平上确认氨基酸序列,分析测定的序列覆盖率,同时给出b和y碎片离子信息。可将氨基酸序列与根据cDNA预测推导得到的序列进行匹配分析,审评建议对于一级质谱应得到100%的序列覆盖率,同时应对二级质谱信号进行序列分析,以获取更多信息。

可通过比较峰保留时间偏移和/或肽质量的测量值与预测值之间的差异等方法检测确认各种序列变异体、翻译后修饰和确定二硫键连接方式,并通过MS/MS分析确定具体位点。通过提取离子色谱图的峰面积积分定量测定修饰氨基酸的水平,结合不同的样品处理或富集手段及其他质量属性可以推断其对抗体药物生物学活性的影响程度。常见的氨基酸修饰包括抗体蛋白N端谷氨酰胺、谷氨酸环化为焦谷氨酸;天冬酰胺(谷氨酰胺、天冬氨酸和谷氨酸)脱酰胺和异构化/环化,产生天冬氨酸、异天冬氨酸和稳定的环状琥珀酰亚胺;蛋氨酸和色氨酸等氨基酸氧化为蛋氨酸亚砷和犬尿氨酸等;赖氨酸糖化;N-糖基化和O-糖基化;重链C末端赖氨酸剪切丢失等多种形式。这些不同的氨基酸修饰对抗体分子的结构、稳定性、生物学活性和免疫原性分别有着不同程度的影响,而且即使相同的氨基酸修饰,也可能会因为处于抗体分子蛋白序列中的不同位置而影响程度完全不同。

6 结语

确认目的基因在DNA、mRNA转录和蛋白表达水平上序列的正确性,是保证药品质量和患者安全的基础。包括NGS和液质联用等多种技术可用于序列分析,这些技术在成本、耗时和专业能力要求方面各不相同,研发机构应按照逐步递进的原则选择这些技术,在生物技术药物生命周期的不同阶段合理使用。近年来应用多属性方法检测生物技术药物纯度以提高整体控制策略得到更多重视^[28],也为药物生命周期管理提供了新的选择。这些技术通常都能产生大量数据,但如何优化数据分析策略以高效检测和鉴定低丰度的序列变异体是技术提供商和药物研发机构的重要课题^[29]。技术审评将依据现有法规和指导原则,基于目前科学认知,对序列分析和确认情况进行整体评价,以保证产品质量、临床使用的疗效和安全性。

References

- [1] Qian Y, Chen Z, Huang X, et al. Early identification of unusually clustered mutations and root causes in therapeutic antibody development [J]. *Biotechnol Bioeng*, 2018, 115: 2377-2382.
- [2] No authors listed. Supplement to the points to consider in the production and testing of new drugs and biologicals produced by recombinant DNA technology: nucleic acid characterization and genetic stability [J]. *Biologicals*, 1993, 21: 81-83.
- [3] WHO. WHO expert committee on biological standardization, guidelines on evaluation of similar biotherapeutic products (SBPs), ECBS [EB/OL]. 2009 [2019-5-10]. https://www.who.int/biologicals/areas/biological_therapeutics/BIOTHERAPEUTICS_FOR_WEB_22APRIL2010.pdf.
- [4] ICH Q5B. Quality of biotechnological products: analysis of the expression construct in cell lines used for production of rDNA derived protein products [EB/OL]. 1996 [2019-5-10]. https://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q5B/Step4/Q5B_Guideline.pdf.
- [5] Peng Z, Cheng Y, Tan BC, et al. Comprehensive analysis of RNA-Seq data reveals extensive RNA editing in a human transcriptome [J]. *Nat Biotechnol*, 2012, 30: 253-260.
- [6] Wong HE, Huang CJ, Zhang Z. Amino acid misincorporation in recombinant proteins [J]. *Biotechnol Adv*, 2018, 36: 168-181.
- [7] Lin TJ, Beal KM, Brown PW, et al. Evolution of a comprehensive, orthogonal approach to sequence variant analysis for biotherapeutics [J]. *MAbs*, 2018, 11: 1-12.
- [8] Svec D, Tichopad A, Novosadova V, et al. How good is a PCR efficiency estimate: recommendations for precise and robust qPCR efficiency assessments [J]. *Biomol Detect Quantif*, 2015, 3: 9-16.
- [9] Azizi A, Aidoo F, Gisonni-Lex L, et al. Determination of HSV-1 UL5 and UL29 gene copy numbers in an HSV complementing Vero cell line [J]. *J Biotechnol*, 2013, 168: 382-387.
- [10] Betts Z, Dickson AJ. Improved CHO cell line stability and recombinant protein expression during long-term culture [J]. *Methods Mol Biol*, 2017, 1603: 119-141.
- [11] Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies [J]. *Nat Rev Genet*, 2016, 17: 333-351.
- [12] Tsiatis AC, Norris-Kirby A, Rich RG, et al. Comparison of Sanger sequencing, pyrosequencing, and melting curve analysis for the detection of KRAS mutations: diagnostic and clinical implications [J]. *J Mol Diagn*, 2010, 12: 425-432.
- [13] Cartwright JF, Anderson K, Longworth J, et al. Highly sensitive detection of mutations in CHO cell recombinant DNA using multi-parallel single molecule real-time DNA sequencing [J]. *Biotechnol Bioeng*, 2018, 115: 1485-1498.
- [14] Treangen TJ, Salzberg SL. Repetitive DNA and next-generation sequencing: computational challenges and solutions [J]. *Nat Rev*

- Genet, 2011, 13: 36-46.
- [15] Ng SH, Azizi A, Edamura K, et al. Preliminary evaluation of next-generation sequencing performance relative to qPCR and in vitro cell culture tests for human cytomegalovirus [J]. *PDA J Pharm Sci Technol*, 2014, 68: 563-571.
- [16] Aeschlimann SH, Graf C, Dmytro M, et al. Enhanced CHO clone screening: applying targeted locus amplification and next-generation sequencing technologies during cell line development [J]. *Biotechnol J*, 2019. DOI:10.1002/biot.201800371.
- [17] Wright C, Groot J, Swahn S, et al. Genetic mutation analysis at early stages of cell line development using next generation sequencing [J]. *Biotechnol Prog*, 2016, 32: 813-817.
- [18] Zhang S, Hughes JD, Murgolo N, et al. Mutation detection in an antibody-producing Chinese hamster ovary cell line by targeted RNA sequencing [J]. *Biomed Res Int*, 2016, 2016: 8356435.
- [19] Tzani I, Monger C, Kelly P, et al. Understanding biopharmaceutical production at single nucleotide resolution using ribosome footprint profiling [J]. *Curr Opin Biotechnol*, 2018, 53: 182-190.
- [20] Li H, Chen K, Wang Z et al. Genetic analysis of the clonal stability of Chinese hamster ovary cells for recombinant protein production [J]. *Mol Biosyst*, 2016, 12: 102-109.
- [21] Wong HE, Huang CJ, Zhang Z. Amino acid misincorporation propensities revealed through systematic amino acid starvation [J]. *Biochemistry*, 2018, 57: 6767-6779.
- [22] Rathore D, Faustino A, Schiel J, et al. The role of mass spectrometry in the characterization of biologic protein products [J]. *Expert Rev Proteomics*, 2018, 15: 431-449.
- [23] Rogers RS, Nightlinger NS, Livingston B, et al. Development of a quantitative mass spectrometry multi-attribute method for characterization, quality control testing and disposition of biologics [J]. *mAbs*, 2015, 7: 881-890.
- [24] Zhao Y, Sun L, Knierman MD, et al. Fast separation and analysis of reduced monoclonal antibodies with capillary zone electrophoresis coupled to mass spectrometry [J]. *Talanta*, 2016, 148: 529-533.
- [25] Griaud F, Winter A, Denefeld B, et al. Identification of multiple serine to asparagine sequence variation sites in an intended copy product of LUCENTIS® by mass spectrometry [J]. *MABs*, 2017, 9: 1337-1348.
- [26] Wen D, Vecchi MM, Gu S, et al. Discovery and investigation of misincorporation of serine at asparagine positions in recombinant proteins expressed in Chinese hamster ovary cells [J]. *J Biol Chem*, 2009, 284: 32686-32694.
- [27] Fornelli L, Ayoub D, Aizikov K, et al. Middle-down analysis of monoclonal antibodies with electron transfer dissociation orbitrap Fourier transform mass spectrometry [J]. *Anal Chem*, 2014, 86: 3005-3012.
- [28] Rogers RS, Abernathy M, Richardson DD, et al. A view on the importance of "multi-attribute method" for measuring purity of biopharmaceuticals and improving overall control strategy [J]. *AAPS J*, 2018, 20: 7.
- [29] Brady LJ, Scott RA, Balland A. An optimized approach to the rapid assessment and detection of sequence variants in recombinant protein products [J]. *Anal Bioanal Chem*, 2015, 407: 3851-3860.