



基于质谱的抗体测序技术*

刘生梅^{1,2)} 薛 鹏^{2)**} 王晓润^{1)**}

(¹⁾ 延安大学生命科学学院, 延安 716000; ²⁾ 广州国家实验室, 广州 510005)

摘要 本文系统综述了质谱技术在抗体组学研究中的应用进展, 探讨了其技术优势、应用场景与当前挑战。抗体的多样性不仅源于基因重组与体细胞超突变等机制, 更体现在蛋白质层面的翻译后修饰及结构异质性。近年来, B细胞受体 (BCR) 测序虽能追踪抗体基因的重组路径, 但难以反映实际分泌抗体的功能状态。因此, 质谱技术凭借高灵敏度和高分辨率, 能在蛋白质层面进行直接解析, 已成为抗体研究的核心分析手段。基于质谱的抗体蛋白质组学测序 (Ab-Seq) 技术, 主要包括依赖参考数据库的自下而上 (bottom-up)、自中而下 (middle-down) 和自上而下 (top-down) 方法, 以及不依赖参考数据库的从头 (de novo) 测序技术, 实现了从肽段、结构域到完整抗体分子的多层次、高精度表征。本文系统梳理了各类质谱分析策略的技术原理、应用场景与发展动态, 并剖析了各自的优势与局限。进一步提出整合 BCR-seq 与 Ab-seq 的协同分析策略, 通过构建基因-蛋白质关联图谱, 为抗体组的研究、抗体药物的结构解析与疫苗效力评估提供有力的研究工具。

关键词 抗体, 质谱分析, B细胞受体测序, 蛋白质抗体测序

中图分类号 Q5, Q7

DOI: 10.3724/j.pibb.2025.0446

CSTR: 32369.14.pibb.20250446

抗体作为适应性免疫系统的关键效应分子, 通过基因重组、体细胞超突变和克隆选择等机制可产生超过 10^{12} 种抗体^[1], 这些数量惊人的抗体构成了生物体应对病原体入侵的特异性防御屏障, 这种多样性不仅体现在抗原结合位点的结构变异上, 更反映在抗体分子糖基化、磷酸化等翻译后修饰 (post-translational modifications, PTMs) 的动态调控上^[2-5]。

近年来, 随着测序技术的发展, B细胞受体 (B cell receptor, BCR) 测序已实现对 BCR 基因组的系统性解码, 能够追踪抗体基因的 V(D)J 重组路径、体细胞超突变模式及克隆演化路径等^[6-8]。然而, 抗体分为膜型抗体和分泌型抗体, 膜型抗体表达在 B 细胞表面, 参与抗原识别和 B 细胞激活, 分泌型抗体是 B 细胞激活后分泌的可溶性蛋白质, 主要存在于血清、组织液和黏膜分泌液中^[9], 仅凭基因序列难以完全反映抗体的真实功能状态, 基因测序技术仅能捕获抗体多样性的“基因蓝图”, 无法揭示其 PTMs、空间构象稳定性及抗原结合活

性等关键功能属性, 这制约了抗体组库研究的临床应用转化。

质谱技术凭借其高灵敏度、高分辨率以及强大的定性和定量能力, 已成为抗体研究不可或缺的核心工具^[10]。随着离子化技术的持续发展与轨道阱 (Orbitrap) 等高精度质量分析器的广泛应用, 结合不断完善的生物信息学算法, 高分辨质谱技术显著提升了对血清抗体的解析能力, 已成为抗体药物结构特异性表征的关键支撑手段。该技术能够直接在蛋白质层面进行分析, 从而精准鉴定抗体蛋白的完整氨基酸序列, 特别是其关键的可变区 (variable region, V 区) 特征, 这对于理解抗体多样性及特异性结合机制至关重要。其次, 质谱技术能全面解析抗体上的多种 PTMs, 如糖基化、氧化、磷酸化

* 广州国家实验室自立项目 (GZNL2023A03007) 和陕西省科技计划 (S2025-YF-YBNY-1080) 资助。

** 通讯联系人。

薛鹏 Tel: 13219387757, E-mail: xue_peng@gzlab.ac.cn

王晓润 Tel: 15891680366, E-mail: dooble@163.com

收稿日期: 2025-10-12, 接受日期: 2026-03-06

和泛素化等, 准确识别这些修饰的位点及丰度分布, 揭示其对蛋白质功能、稳定性及相互作用的调控作用, 为抗体功能研究提供直接的分子证据, 有效弥补了基因序列信息与实际蛋白质状态之间的偏差^[11-13]。

通过同步获取BCR测序的基因重组特征、质谱解析的蛋白质修饰谱组数据, 能够系统阐明抗体多样性产生的分子机制, 这种基因-蛋白质-功能的多维关联分析, 为抗体组研究、抗体药物的结构解析与疫苗效力评估提供了有力的研究工具。近年来, 抗体序列结构及功能等方面已取得诸多进展, 随着质谱技术的持续发展, 抗体的研发技术不断提升, 已经发展出依赖参考数据库的自下而上(bottom-up)、自中而下(middle-down)和自上而下(top-down), 以及不依赖参考数据库的从头(*de novo*)测序等研究方法。本文针对这些质谱技术对抗体组学研究应用及最新进展进行分析和总结, 探讨其技术优势、应用场景及当前面临的挑战, 以期为未来的抗体研究提供指导和参考。

1 抗体的基本结构

抗体的精密结构是解析免疫机制的关键, 通常抗体是Y型结构(图1), 其Y形结构通过二硫键将两条相同的重链(heavy chain, H链)和轻链(light chain, L链)连接起来, 每条链分为V区和恒定区(constant region, C区)。

1.1 可变区(V区): 抗原识别的分子基础与表征难点

V区位于L链靠近N端的1/2(约含108~111个氨基酸残基)和H链靠近N端1/5或1/4(约含118个氨基酸残基), 包含4个框架区(framework region, FR)和3个互补决定区(complementarity determining region, CDR), V区L链(VL)的CDR在24~34、50~56、89~97氨基酸位置, V区H链(VH)的CDR在31~35、50~56、95~102氨基酸位置。V区的多样性由V(D)J基因重排、体细胞高频突变(somatic hypermutation, SHM)和N区插入等机制产生, 其中CDR3的多样性最高, 直接决定抗原结合的特异性^[14-15]。这种特性使得传统结构生物学方法(如X射线晶体学)难以高通量、全

面地捕获其天然溶液状态下的构象分布及细微的化学修饰(如CDR区发生的氧化), 从而构成了对其功能关键区域进行精准结构表征的首要挑战。

1.2 恒定区(C区): 效应功能的调控中心与修饰复杂性

C区位于L链靠近C端的1/2(约含105个氨基酸残基)和H链靠近C端的3/4区域或4/5区域(约从119位氨基酸至C端), 决定抗体的效应功能。根据重链C区的不同, 抗体可分为IgA、IgG、IgM、IgD和IgE, 对应 α 、 γ 、 μ 、 δ 、 ϵ 五种不同类型H链, 各类别在免疫应答中发挥不同作用。其中IgG因高丰度、长半衰期及强中和能力成为治疗性抗体的主要形式, IgG进一步分为4个亚类(IgG1、IgG2、IgG3、IgG4), 其铰链区结构差异影响功能特异性^[16]。尤为重要的是, Fc段存在多种PTMs(如糖基化、氧化、脱酰胺等), 不同类别和亚类之间, 甚至同一亚类内的修饰模式都可能存在差异, 直接影响抗体的稳定性、半衰期和效应功能。这种PTMs的复杂异质性, 对分析技术的分辨率、灵敏度及定量准确性提出了极致要求, 是抗体质控与功能研究中的又一核心表征难题。

1.3 铰链区: 结构域模块化分析与酶切工具

连接Fab与Fc的铰链区(hinge region)构象十分灵活, 使抗原结合区域能够在空间上自由移动, 更加有效地结合不同位置的抗原。铰链区可被木瓜蛋白酶、胃蛋白酶这两种非特异性酶切割形成Fab段(抗原结合区)和Fc段(效应功能区), Fab段通过CDRs识别抗原表位, 而Fc段则激活补体系统或介导抗体依赖性细胞毒性^[17-18]。除木瓜蛋白酶和胃蛋白酶外, 还有不少特异性酶能把抗体分成Fab段和Fc段^[19]。它们的主要差别在于, 不同的酶在抗体铰链区的切割位置不一样, 这会影响到Fab、F(ab')₂和Fc片段的大小(表1)^[20-26]。这些酶切工具不仅为研究抗体各结构域的独立功能提供了手段, 更重要的是, 它们为质谱的自中而下(分析~25 ku片段)和自上而下(分析完整分子)策略提供了关键的样品预处理方法, 使得针对完整抗体或其大片段的结构与修饰进行直接分析成为可能^[24, 27-29]。

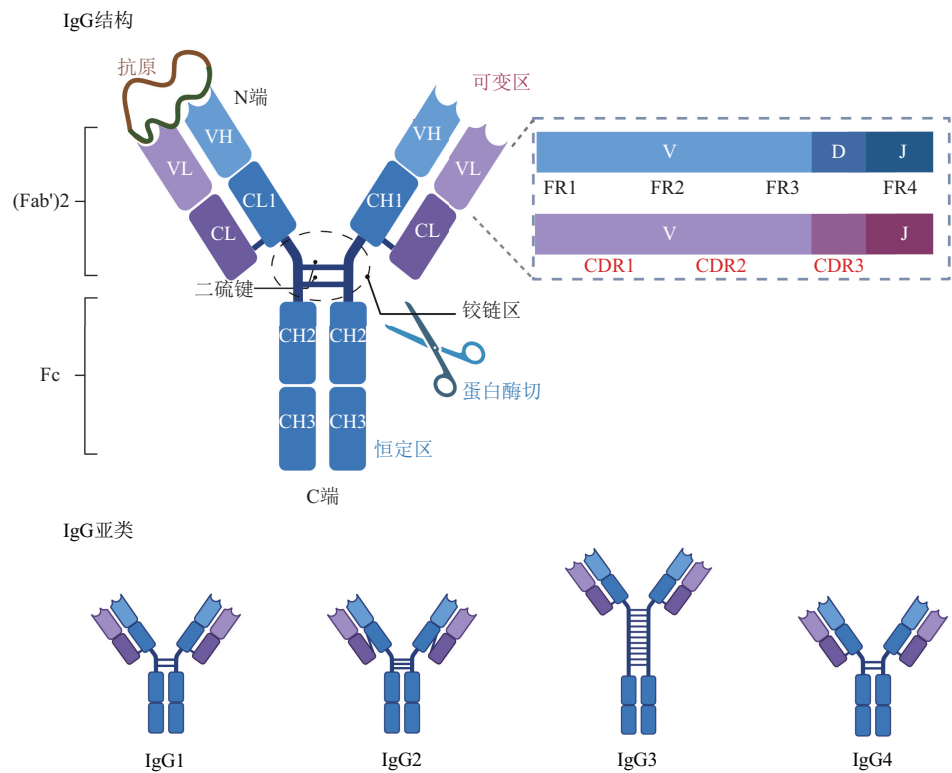


Fig. 1 Schematic diagram of the structure of IgG antibodies and their subclasses

图1 IgG抗体结构及其亚类示意图

IgG抗体结构示意图及其4种亚类。上：IgG分子由两条重链（蓝）和两条轻链（紫）组成Y形结构，经蛋白酶切割后产生两个Fab段（抗原结合区）和一个Fc段（效应功能区）；可变区（VH和VL）包含框架区（FR1~FR4）和互补决定区（CDR1~CDR3），负责抗原结合。下：人IgG1~IgG4四种亚类，主要差异在于铰链区长度。

Table 1 Antibody hinge region cleaving enzyme

表1 抗体铰链区裂解酶

蛋白酶名称	靶抗体类型	切割位点	来源	参考文献
木瓜蛋白酶	IgG	铰链区	番木瓜（ <i>Carica papaya</i> ）	[19]
胃蛋白酶	IgG	铰链区	胃黏膜（ <i>Gastric mucosa</i> ）	[19]
IdeS (FabRICATOR®)	IgG	铰链区下方（CPPCPAPELLG/GPSV）	酿脓链球菌（ <i>Streptococcus pyogenes</i> ）	[20]
IdeZ (FabRICATOR®Z)	IgG	铰链区下方（CPPCPAPELLG/GPSV）	马链球菌兽疫亚种（ <i>Streptococcus equi</i> SSP. zooepidemicus）	[20]
SpeB (FabULOUS™)	IgG	铰链区上方（KTHT/CPPCPAPE）	酿脓链球菌（ <i>Streptococcus pyogenes</i> ）	[22]
OgpA (OpeRATOR®)	IgA1	O-糖基化铰链区裂解	肺炎链球菌（ <i>Streptococcus pneumoniae</i> ）	[23]
IgdE (FabALACTICA®)	IgG1	铰链区上方（KSCDKT/HTCPPC）	无乳链球菌（ <i>Streptococcus agalactiae</i> ）	[24]
BdpK (FabDELLO)	IgG1	铰链区上方（KSCDK/THTCPPCP）	蛭弧菌（ <i>Bdellovibrio bacteriovorus</i> ）	[24]
Kgp (GingisKHAN®)	IgG1	铰链区上方（KSCDK/THTCPPC）	牙龈卟啉单胞菌（ <i>Porphyromonas gingivalis</i> ）	[25]
IgMBRAZOR™	IgM	IgM铰链区	NA	[26]

2 基于质谱的抗体组分析技术

抗体组的高效分析在很大程度上取决于样品前

处理的质量，尤其是目标抗体的有效分离与富集。由于血清、血浆或黏膜分泌物等生物样本中的抗体库组成复杂，包含多种同种型（如 IgG、IgA、

IgM) 及其亚类, 因此在质谱分析前需采用高特异性的亲和纯化方法以降低背景干扰并获得高质量数据。样品前处理不仅影响后续质谱检测的灵敏度和准确性, 更是实现精准定量的关键。目前常用的策略包括利用 Protein A/G 与抗体 Fc 区结合, 或使用 Protein L 特异性识别 kappa 链的 Fab 区, 从而实现对不同类别或结构域抗体的选择性富集^[30-31] (表 2), 此外, Protein G 除可用于完整 IgG 的富集外, 在某些条件下也可用于 Fab 片段的捕获。值得注意的是, 不同物种及抗体亚类与各类亲和配体的结合特性存在显著差异, 例如: 人类 IgG1、IgG2 和 IgG4 与 Protein G 结合较强, 而 IgG3 则较弱; 小鼠总 IgG 及 IgG1、IgG2a、IgG2b 等多通过 Protein A 进行有效捕获。这些方法的选择需综合考虑抗体物

种来源、亚类分布及研究目标, 以确保在降低非特异性吸附的同时, 最大程度保留目标抗体的结构完整性与生物活性。

在获得高纯度抗体样品后, 可进一步通过酶切等处理将其转化为肽段混合物, 并通过液相色谱-串联质谱 (liquid chromatography-tandem mass spectrometry, LC-MS/MS) 进行系统表征。基于质谱的抗体组分析技术能够实现抗体 V 区与 C 区序列的高通量解析, 并同时鉴定糖基化、氧化等多种 PTMs 的位点与程度。通过整合定量信息, 该技术可全面揭示不同抗体亚类的分布、特异性抗体的动态变化以及结构异质性对抗体功能的影响, 从而在系统水平上阐明抗体组的组成与功能关联。

Table 2 The affinity of Protein A, Protein G, and Protein L for antibodies from different species^[30-31]
表2 Protein A, Protein G, Protein L对不同种属抗体的亲和能力^[30-31]

种属	抗体类型	Protein A	Protein G	Protein L (kappa)链
人类	总IgG	+++	+++	+++
	IgG1	+++	+++	+++
	IgG2	+++	+++	+++
	IgG3	-	+++	+++
	IgG4	+++	+++	+++
	IgM	+	-	+++
	IgD	-	-	+++
	IgA	+	-	+++
小鼠	总IgG	+++	+++	+++
	IgG1	+	++	+++
	IgG2a、2b	+++	+++	+++
	IgG3	+++	+++	+++
	IgM	+	+	+++
大鼠	总IgG	+	++	+++
	IgG1	+	++	+++
	IgG2a	-	+++	+++
	IgG2b	-	+	+++
	IgG2c	+++	+++	+++
山羊	总IgG	+	+++	-
	IgG1	+	+++	-
	IgG2	+	+++	-
绵羊	总IgG	+	+++	-
	IgG1	+	+++	-
	IgG2	+++	+++	-
兔子	总IgG	+++	+++	+
马	总IgG	+	+++	NA
牛	总IgG	+	+++	-
豚鼠	总IgG	+++	-	NA
猪	总IgG	+++	+	+++

续表2

种属	抗体类型	Protein A	Protein G	Protein L(kappa)链
狗	总IgG	+++	-	NA
猫	总IgG	+++	-	NA
鸡	总IgY	-	-	-

+结合，-不结合，NA无数据。

现有两种策略。一种是依赖参考数据库的策略：包括自下而上、自中而下和自上而下这三种方式（图2）；二是不依赖参考数据库的策略：从头

测序技术，直接从抗体蛋白的氨基酸序列出发，尤其适用于未知抗体或物种缺乏基因组数据的场景^[32-35]。

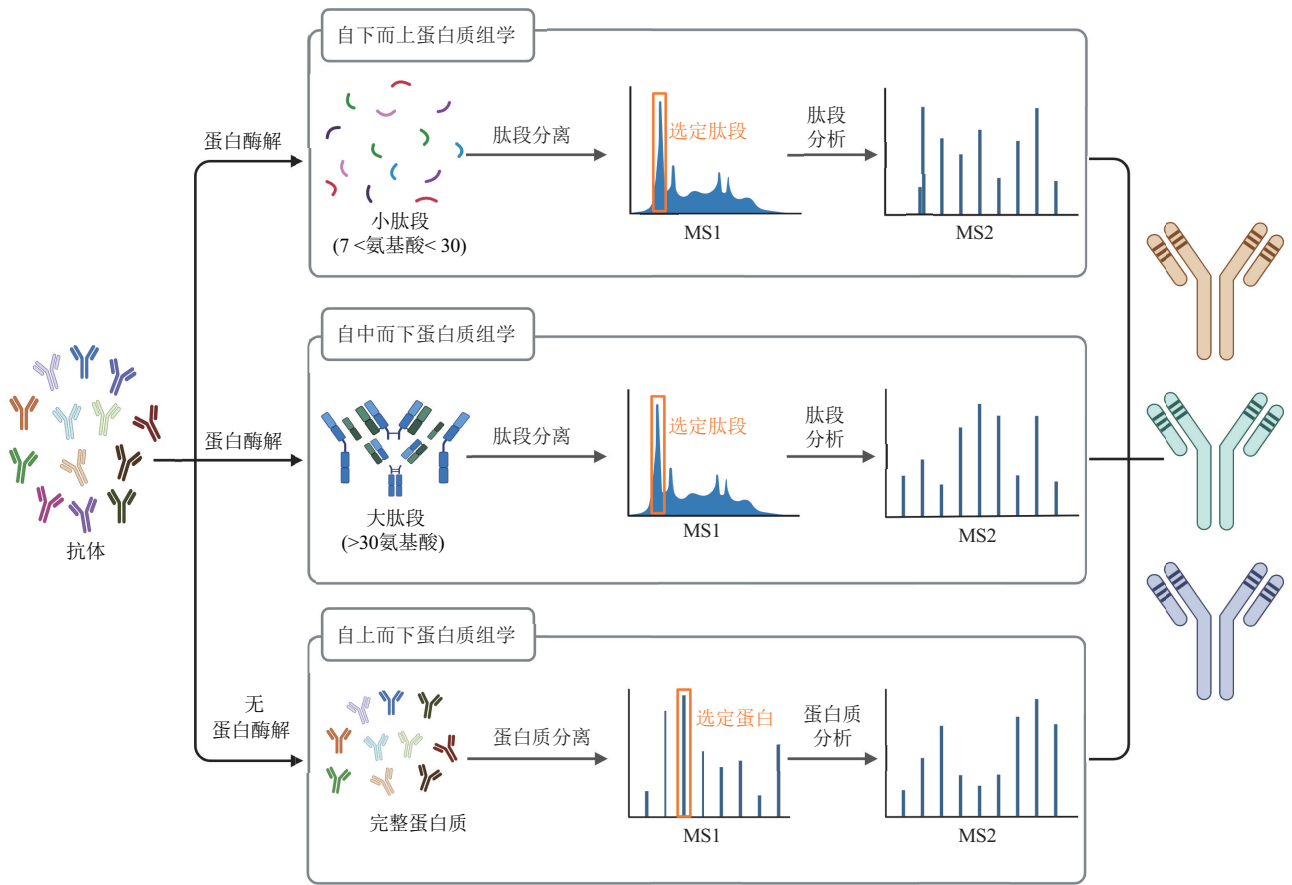


Fig. 2 Different analysis strategies for the antibody panel based on mass spectrometry
图2 基于质谱的抗体组的不同分析策略

三种基于质谱的蛋白质组学分析策略的工作流程：自下而上蛋白质组学分析将蛋白质酶解为小肽段（7~30个氨基酸），经色谱分离后进行一级质谱（MS1）和二级质谱（MS2）分析，通过鉴定肽段来推断蛋白质信息；自中而下蛋白质组学分析则产生较大的肽段（>30个氨基酸），在保持更多序列信息的同时进行质谱分析，有助于识别翻译后修饰和序列变异；自上而下蛋白质组学直接分析完整蛋白质，不经酶解，可全面表征蛋白质的完整分子质量及修饰组合。

2.1 自下而上策略

自下而上是目前最成熟、应用最多的蛋白质分析质谱技术，将纯化的抗体在变性条件下进行还原和烷基化后，使用质谱特异性蛋白酶进行蛋白质水解，将完整抗体分子酶解为小分子肽段，然后通

过LC-MS/MS对这些肽段进行分析鉴定^[36-37]。该技术凭借成熟的操作流程、优异的灵敏度以及出色的高通量性能，已成为大规模抗体组库研究的首选方案。其在抗体研究中的核心优势在于，能够对复杂样本（如血清、细胞培养上清）实现高通量、高

灵敏度的深度覆盖, 尤其擅长鉴定高丰度抗体及常见的PTMs, 为抗体组库的全面筛查与定量比较提供了强大工具。

然而, 其应用于抗体研究时也存在鲜明的固有局限与挑战。由于抗体序列 (尤其是CDR) 的多样性, 标准胰蛋白酶切位点可能分布不均, 导致关键功能区域的肽段覆盖不全或产生过长、过短的肽段, 影响离子化与检测效率。更重要的是, 该策略将完整蛋白质结构解构为肽段, 不可避免地导致更高层次结构信息的丢失, 例如CDR的空间构象、不同结构域上修饰之间的关系等。这使得全长序列 (尤其是V区) 的从头组装面临挑战, 对糖基化等复杂修饰的精细表征能力受限。此外, 该策略高度依赖已知序列数据库, 在很大程度上限制了对全新抗体或罕见变异体的发现能力^[38]。

针对上述挑战, 该领域已发展出多项标志性的方法与优化策略, 显著提升了自下而上在抗体关键区域解析上的能力。在提升序列覆盖方面, 使用胰蛋白酶 (trypsin)、赖氨酰精氨酰内肽酶 (又称为胰蛋白酶镜像酶, LysargiNase)、赖氨酸特异性蛋白酶 (Lys-C)、赖氨酸特异性蛋白酶 (Lys-N)、胰凝乳蛋白酶 (chymotrypsin)、谷氨酸特异性蛋白酶 (Glu-C)、精氨酸特异性蛋白酶 (Arg-C) 和天冬氨酸特异性蛋白酶 (Asp-N) 等多种蛋白酶进行联合酶解, 增加抗体序列的覆盖度^[39-40], 已经是抗体解析的常用工具。例如, Peng等^[41]通过使用9种互补蛋白酶并结合高能碰撞解离 (high-energy collision dissociation, HCD) 与电子转移/高能碰撞解离 (electron-transfer/higher-energy collision dissociation, EThcD) 双重碎裂模式, 实现了对治疗性单克隆抗体赫赛汀 (曲妥珠单抗) 的近乎全序列覆盖, 在可变区的序列准确率达到99%。Yang等^[42]还开发了乙酰化赖氨酰精氨酰内肽酶 (又称为乙酰化胰蛋白酶镜像酶, acetylated LysargiNase,

Ac-LysargiNase), 比LysargiNase具有更高的活性和稳定性, 和胰蛋白酶一起创建互补离子系列, 分别从C端和N端进行酶切, 使两端的离子覆盖率能够达到100%^[42-43]。

在质谱分析中, 为了提高抗体CDR的覆盖率, 通常会采用多种二级碎裂技术的组合。碰撞诱导解离 (collision-induced dissociation, CID) 与HCD可高效产生b/y型离子, 提供基础序列信息; 而电子转移解离 (electron transfer dissociation, ETD)、电子捕获解离 (electron capture dissociation, ECD)、电子活化解离 (electron activated dissociation, EAD) 等技术则能产生c/z型离子, 更好地保留不稳定的PTMs, EThcD可同时提供b/y和c/z离子, 显著提升序列覆盖与修饰定位的可靠性, 更重要的是这几种模式利用亮氨酸和异亮氨酸侧链的差异性中性丢失, 可以成功区分这对同分异构体, 使大分子质量蛋白质及修饰位点鉴定成为可能; 紫外光解离 (ultraviolet photodissociation, UVPD) 则能产生更均匀的碎片离子分布, 有利于大肽段的解析, 从而显著提升CDR的覆盖率^[37, 44-45] (表3)。除上述核心策略外, 质子转移电荷还原 (proton transfer charge reduction, PTCR) 作为补充手段, 通过电荷状态调控, 可以将序列覆盖率推至87%^[27]。

尽管如此, 必须指出自下而上策略的局限性是结构性的。多重酶解与组合碎裂在提升信息量的同时, 也加剧了信息的“碎片化”。这种片段化分析模式, 导致完整糖型在单个抗体分子上的精确组合, 以及不同修饰在空间上的关联性信息永久丢失。因此, 尽管自下而上策略在序列鉴定与修饰筛查方面不可替代, 但其在揭示抗体更高层次的结构和功能关联方面存在局限, 需要自中而下与自上而下策略进行补充。

Table 3 Main mass spectrometry fragmentation patterns^[37, 44-45]

表3 主要质谱碎裂模式^[37, 44-45]

碎裂模式	能量类型	主要离子	特点
碰撞诱导解离 (CID)	碰撞能量	b/y	常规肽段鉴定; 能量可控
高能碰撞解离 (HCD)	高能碰撞	b/y	碰撞能量更高; 产生的碎片更丰富
电子转移解离 (ETD)	电子转移	c/z	保留修饰, 适合大分子
电子捕获解离 (ECD)	电子捕获	c/z	保留修饰, 适合大分子
电子活化解离 (EAD)	电子轰击	c/z	保留修饰, 适合大分子
电子转移/高能碰撞解离 (EThcD)	电子转移+高能碰撞	c/z/b/y	综合ETD和HCD优势, 能区分L/I
紫外光解离 (UVPD)	紫外激光	a/b/y/x	可调节光波长; 需要紫外激光源

2.2 自中而下和自上而下策略

为克服自下而上策略因酶解导致的结构信息碎片化问题,自中而下与自上而下两种质谱分析方法应运而生。它们避免了将完整抗体彻底消化为短肽,从而在更高层次上保留了抗体的结构完整性,与自下而上策略相互补充,共同构成了从肽段、结构域到完整分子的多尺度抗体质谱分析体系。

自上而下技术直接对完整抗体分子(~150 ku)进行质谱分析,从根本上避免酶解过程带来的信息损失,能够全面保留抗体的天然结构和翻译后修饰信息。然而,该技术对质谱平台与分析流程均提出了更高要求:仪器需具备高质量分辨率、宽质量范围以及高效碎裂能力,以应对完整抗体分子(~150 ku)产生的多电荷、宽分布离子信号;同时,数据解析软件需能处理高度复杂的碎片谱图,并实现对大质量范围内多种修饰的准确归属。虽然该方法会产生高度带电的碎片离子,数据解析较为复杂,但在完整抗体表征方面具有不可替代的优势^[46]。

尽管门槛极高,但自上而下策略已在完整抗体表征方面取得标志性突破。例如,Peng等^[41]利用自上而下技术从患者血清IgG1中检测到Fab糖基化的M蛋白,这可能与单克隆丙种球蛋白血症或多发性骨髓瘤等疾病有关,这种直接从外周血样本中确定M蛋白序列的能力为理解这些疾病的分子机制提供了机会。在技术层面,结合EAD等高效碎裂技术,可实现对完整抗体及其亚单位的高分辨率表征^[47]。此外还可以结合ECD等技术,解析接近1 Mu的IgM五聚体和IgG基寡聚体的序列标签,从重糖基化的IgA1抗体中获得清晰的抗原结合区序列信息^[48-49]。

自中而下技术作为连接自下而上与自上而下的桥梁,通常利用IdeS、IgdE等特异性蛋白酶在抗体铰链区进行切割,产生约25~50 ku的Fab、F(ab')₂或Fc片段,在保持结构域完整性的同时显著降低了质谱分析的复杂性。该策略在保留关键结构信息(如CDR完整性、结构域内二硫键与局部修饰)与实现较高序列覆盖之间取得了较好平衡,尤其适用于靶向分析抗体的功能区域^[29, 50]。因此,自中而下策略尤其适用于对抗体关键功能区域进行靶向深度表征。

然而,25 ku左右的多电荷前体离子会导致二级谱图高度复杂与碎片离子信号重叠,严重阻碍对关键修饰位点的精确解析,针对这一挑战,PTCR

与离子淌度(ion mobility spectrometry, IMS)等辅助技术展现出关键价值。PTCR通过气相离子-离子反应降低离子电荷态有效简化谱图,从而有效“解卷积”高度重叠的谱图,大幅度提升序列覆盖的深度^[51];而IMS则依据碰撞截面积(collision cross section, CCS)的差异,在质谱检测前对具有细微构象或修饰差异的抗体片段进行分离。例如,Fiala等^[12]整合timsTOF系统的IMS功能,成功对低至纳克级的IgG1 Fab片段进行了高质量分析,并区分了其不同的微异构。这些技术极大地增强了对结构域水平微观异质性的深度解析能力,为抗体尤其是双特异性抗体、抗体-药物偶联物(antibody-drug conjugate, ADC)等复杂结构抗体的精准表征提供了强大工具^[52-53]。

综上所述,这两种策略与自下而上方法形成技术互补,自中而下在结构域水平提供平衡的分辨率和完整性,自上而下则专注于完整分子的全面表征,三者共同构成了抗体质谱分析的完整技术体系。最新研究趋势表明,这些方法的结合(如自中而下预分离后进行自上而下分析)正在成为深入解析抗体结构与功能关系的重要途径。这种多尺度整合策略不仅拓展了质谱技术在抗体研究中的应用范围,也为抗体药物的开发和优化提供了更全面的分析手段。

2.3 从头测序策略

从头测序(*de novo sequencing*)技术是一种不依赖蛋白质数据库的质谱解析技术,该技术通过直接解析质谱数据中的肽段序列信息,能够有效克服传统质谱蛋白质组学方法对参考数据库的依赖性,为新型抗体和稀有突变体的鉴定提供了新的解决方案^[54]。近年来,该领域已从依赖固定规则的算法,演变为一个多层次计算模型驱动的快速领域,其技术演进的核心在于如何更有效地从高噪声、高复杂度的质谱数据中提取并组装序列信息(表4)^[55-67]。

基于传统机器学习的算法,如Novor采用谱图(spectrum graph)结合决策树(decision tree)的方法^[55],迟浩团队开发的pNovo系列算法,持续推动了该领域的发展,其中pNovo3引入支持向量机(support vector machine, SVM)来提升分类性能^[56]。随着深度学习技术的发展,卷积神经网络(convolutional neural network, CNN)和循环神经网络(recurrent neural network, RNN)的组合架构逐渐成为主流^[57-62],其中CNN通过局部感受野捕

获质谱图中相邻碎片离子的特征模式, RNN 则利用序列记忆机制逐步推断氨基酸序列。值得指出的是, Bioinformatics Solutions Inc. (BSI) 团队开发的 DeepNovo 及其面向 DIA 数据的变体 DeepNovo-DIA 是这一范式下具有广泛影响力的实现, 其采用双模型迭代解码框架, 通过离子特征卷积与序列循环建模协同工作, 显著提升了测序的准确性与适用性^[57-58]。

近年来, 更先进的模型架构不断推动该领域向前突破。图神经网络 (graph neural network, GNN) 通过将质谱峰关系建模为图结构, 有效捕获了离子间的全局拓扑特征^[63]; Transformer 则凭借其自注意力机制实现了对谱图上下文的整体建模, 催生了如 Casanovo、 π -HelixNovo 等高性能变体^[64-65]。最新进展如 InstaNovo 引入扩散编码器增强鲁棒性, π -PrimeNovo 采用非自回归 Transformer 大幅提升解码效率, 显示该领域在精度与速度两方面持续优化^[66-67]。

在算法性能的实证比较方面, 已有众多研究进行了深入探讨。Beslic 等^[68] 在同一大型人类谱图库上重新训练多个主流模型 (包括 Novor、pNovo3、DeepNovo、SMSNet、PointNovo 和 Casanovo), 以公平评估其在抗体数据上的表现, 结果显示 Casanovo 和 PointNovo 在不同酶切条件下均展现出更高的肽段召回率^[68-69]。Tran 等^[70] 直接测试了多个工具 (包括 PEAKS、PointNovo、Casanovo 和 GraphNovo) 在跨物种及模拟数据上的泛化能力。研究发现, 尽管所有工具在与其训练数

据相近的人类样本上表现最佳, 但当面对亲缘关系较远的物种数据时, 性能均出现下降, 这揭示了当前算法在跨域泛化方面仍存在局限性。综合来看, 以 Casonovo 和 GraphNovo 为代表的基于先进架构 (如 Transformer 与图神经网络) 的模型在多项评估中表现出相对更优且稳健的性能。

结合多种蛋白酶酶解技术, 并通过 ECD 或 HCD/ET_hcD 等多重碎裂技术, 与软件算法相结合, 可以实现抗体序列的高效覆盖和准确组装^[34-35, 71-72]。但将从头测序成功应用于抗体研究仍面临一系列整合性挑战。实验上, 对起始样品的纯度与丰度有较高要求, 高度依赖高质量的样品制备与富集, 并需结合多种蛋白酶与互补碎裂技术以获取高信息量的谱图。在数据分析层面, 现有模型在面对低信噪比谱图、混合克隆背景以及存在多种 PTMs 共存的复杂肽段时, 其识别与解析能力仍显不足。此外, 算法的跨数据集泛化能力也有待提高, 如在远源物种数据上性能往往下降, 提示当前方法在普适性上仍存在局限。

抗体从头测序的发展需进一步融合实验设计与智能计算。一方面应发展更精准的样本处理与多维质谱采集策略, 从源头降低数据复杂性; 另一方面需构建能够处理大规模克隆共存、整合多维度谱图信息, 并直接输出完整修饰谱的新一代算法平台。通过实验科学与计算科学的深度协同, 从头测序有望从一项序列解析技术, 演进为能够系统阐明抗体组多样性、动态变化及功能关联的核心分析体系, 为抗体药物研发与免疫精准诊疗提供坚实基础。

Table 4 De novo sequencing analysis algorithm

表4 从头测序解析算法

软件/模型名称	算法/模型类型	参考文献
Novor	Spectrum graph+Decision tree	[55]
pNovo	Spectrum graph+SVM	[56]
DeepNovo DeepNovo-DIA	CNN+RNN	[57-58]
SMSNet	CNN+RNN	[59]
PointNovo	PointNet+RNN	[60]
GraphNovo	GNN	[63]
PepNet	CNN	[61]
Spectralis	CNN	[62]
Casanovo	Transformer	[64]
π -HelixNovo	Transformer	[65]
π -PrimeNovo	Non-autoregressive Transformer	[66]
InstaNovo	Transformer+Diffusion encoder	[67]

3 BCR测序与质谱抗体蛋白质组学测序联用

抗体功能的最终实现, 依赖于其基因在蛋白质层面的表达水平。因此, 整合BCR-seq的基因信息与质谱抗体蛋白质组学测序 (mass spectrometry -based antibody repertoire sequencing, Ab-seq) 的蛋白质实体信息, 成为系统解析抗体组、贯通“基因型-表型”的必然方向。BCR-seq 主要包括批量 (bulk) 与单细胞 (single cell, sc) 两种策略, 前者高效捕获群体多样性但丢失轻重链配对信息, 后者能获取天然配对的完整受体序列但通量有限, 二者互为补充, 共同形成了从序列到功能的BCR解析体系^[73-84]。

尽管BCR-seq技术在抗体分析中具有显著优势, 但它主要侧重于B细胞基因组的克隆演化信息, 无法直接解析血清或黏膜分泌物中实际存在的抗体蛋白质丰度与功能特征, 且血清中BCR表达水平与其对应抗体产生效率之间存在显著差异, 二者的转化关系尚未完全阐明^[39]。一项关于破伤风疫苗免疫应答的研究揭示了BCR-seq与血清抗体检测结果的重要差异, 接种破伤风疫苗后, BCR-seq可检测到近100种抗体克隆型, 但血清中实际存在的功能性IgG抗体仅占40%, 更值得注意的是, 9个月后血清中持续存在的记忆抗体, 其对应的克

隆型不足初始BCR检测结果的5%^[85], 这一发现凸显单纯依赖BCR-seq的局限性, 证实只有整合BCR-seq与Ab-seq分析, 才能准确评估功能性抗体水平。

在技术整合层面, BCR-seq与Ab-seq的整合使用展现出显著优势 (图3), 其核心在于构建“个人定制化”或“疾病特异性”的抗体序列数据库。通过将质谱获得的肽谱数据与BCR-seq构建的参考数据库进行比对^[86], 将基因组与蛋白质组信息整合。一方面, BCR-seq提供深度克隆型参考序列, 能显著提升质谱对低丰度、高突变抗体肽段的鉴定效率和可信度, 克服了传统通用参考数据库的匹配瓶颈; 另一方面, Ab-seq的结果直接验证了哪些BCR序列被实际翻译、分泌成抗体蛋白, 并同时提供其精确的翻译后修饰谱 (如糖型、氧化) 和相对丰度, 这是单纯的基因测序无法获得的。

Lê Quý等^[32]结合bulkBCR-seq、scBCR-seq和Ab-seq三种技术, 对健康人外周血B细胞的BCR和血清抗体的多样性进行全面分析, 个体内bulkBCR-seq和scBCR-seq之间的VH基因库特征高度一致, 将Ab-seq数据与BCR-seq参考库相结合, 成功从血清肽段数据中重建出完整的、配对的抗体克隆型序列。这种整合策略在疾病研究中也取得重要突破, 如重症肌无力研究揭示了IgG V区N-糖基化异常与SHM的关联, 而多发性骨髓瘤研

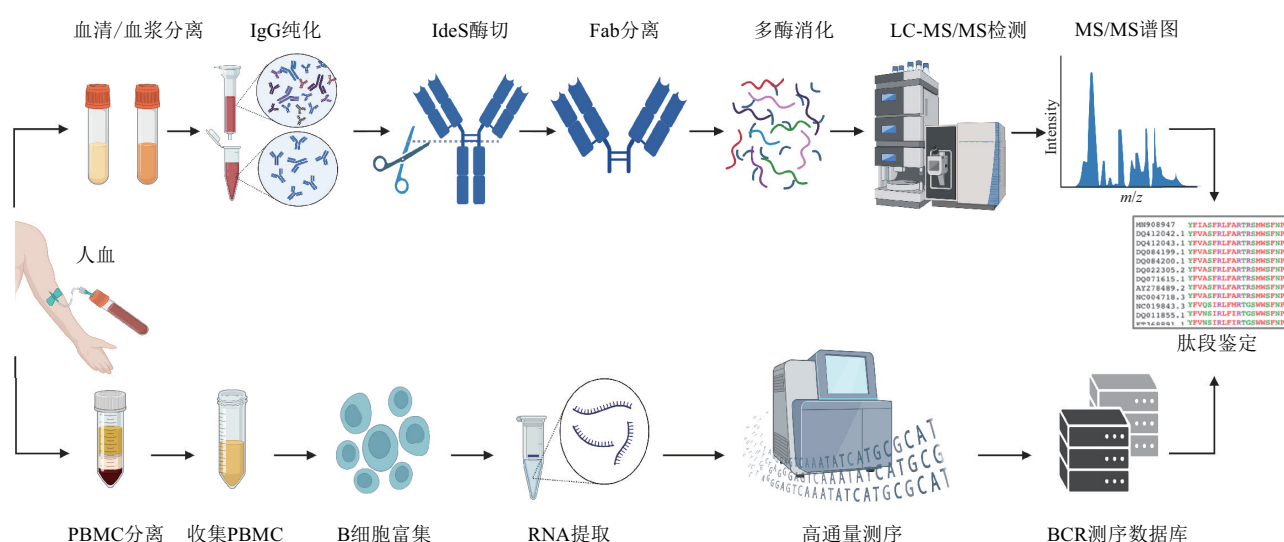


Fig. 3 Schematic diagram of the overall process of antibody repertoire research

图3 抗体库研究总流程示意图

通过蛋白质组学与B细胞受体 (BCR) 测序两条互补技术路线进行系统性抗体分析的完整实验流程。上方路线为基于LC-MS/MS的抗体蛋白质组学分析: 经血清/血浆分离、IgG纯化、IdeS酶切、Fab片段分离及多酶消化后, 通过液相色谱-串联质谱检测获得MS/MS谱图, 实现肽段鉴定与抗体蛋白质组解析。下方路线为基于高通量测序的BCR组库分析: 经外周血单个核细胞 (PBMC) 分离、B细胞富集、RNA提取后进行高通量测序, 构建BCR测序数据库。

究则发现了免疫球蛋白L链的新型PTMs(如辅酶M和N-糖基化)^[27, 87-88]。值得注意的是, 将BCR-seq与自上而下、自中而下、自下而上和从头测序联用的质谱策略可全面表征抗体的结构异质性, 为治疗性抗体的质量评估提供了新方法^[51, 89-90]。这些研究共同表明, BCR-seq与Ab-seq的协同分析不仅能克服单一技术的局限性, 更能发现传统方法难以捕捉的抗体特征变异, 为精准免疫研究和抗体药物开发提供了强有力的技术支撑。

BCR-seq与Ab-seq的协同应用虽然在抗体组学研究中展现出互补优势, 但其实际应用中仍面临一系列技术和生物学层面的关键挑战^[91]。有报道指出, 仅有少数Ab-seq肽段(约4.8%)能与BCR-seq参考序列匹配, 多映射到低丰度克隆型, 导致数据整合效率低^[32]。BCR-seq需要合理设计实验, 选择合适的样本类型和采集时间点, 以确保能够全面反映BCR的多样性和动态变化, 并且数据库的构建可能包含潜在的错误, 包括测序错误、测序深度不均匀等, Ab-seq对质谱参考数据库的依赖会降低对突变抗体的识别, 如何构建和更新准确的抗体序列数据库, 以及如何减少对参考数据库的依赖, 是提高Ab-seq分析质量的关键问题。

4 总结与展望

基于质谱的抗体组学研究已实现对单克隆抗体的深度表征, 但在系统解析整个抗体组方面仍面临诸多挑战。采用BCR-seq与Ab-seq联合分析的方法, 虽能通过BCR-seq为抗体蛋白质组研究提供基础序列信息, 但BCR-seq主要是对B细胞进行测序, 其数据无法完全反映循环系统中实际存在的抗体, 从而导致BCR数据库的覆盖范围不全面。因此, 仍需大力发展不依赖参考数据库的质谱方法, 以推动抗体系统的全面解析。

质谱设备的发展为抗体解析提供新的手段。当前, 高性能质谱平台大多支持两种碎裂模式, 产生b/y、c/z离子, 多维多级离子阱(Ominitrap)进一步将CID、EAD和UVPD等多种碎裂模式结合^[92-93], 可以覆盖所有离子(a/x、b/y、c/z), 实现对肽段氨基酸序列的完整解析, 这一技术进展显著提升从头测序的准确性与可靠性, 对于抗体可变区这类复杂序列的解析具有重要价值。与此同时, 软件算法的不断优化也为抗体解析提供了重要支持, 基于深度学习的从头测序工具, 采用卷积-Transformer混合网络架构, 能够从碎片化质谱数

据中有效重建CDR序列, 显著降低未知抗体序列解析的错误率, 使单个抗体的解析逐步摆脱对参考数据库的依赖。

然而, 面对人体内真实存在的、高度复杂的多克隆抗体系统, 当前技术体系仍面临严峻挑战。首先, 在样品制备方面, 尽管目前已有成熟的富集方法用于血清中含量较高的IgG和IgA等抗体, 但对于血清中含量极低的IgE抗体, 仍缺乏有效的富集策略, 这需要开发更具针对性的分离技术, 以提高这些抗体的检测灵敏度和覆盖率。其次, 在测序方法学上, 目前尚未有将自下而上、自中而下和自上而下3种策略整合应用于抗体分析的成熟方案, 未来的研究应致力于开发多策略协同的分析平台, 以充分发挥各策略的优势, 实现从肽段到完整分子的多尺度、高覆盖抗体组解析。最后, 在从头测序方面, 关键在于破解数据解卷积的困局, 面对多克隆抗体谱图中海量、重叠的肽段信号, 必须深度融合先进的生物信息学算法与人工智能技术, 通过发展高通量谱图解析、高精度序列组装以及智能化克隆家族归类的计算平台, 才能从复杂质谱数据中准确还原出高度多样化的抗体序列全景。

综上所述, 克服这些技术瓶颈将是实现全面、精准抗体组解析的关键。随着高分辨率质谱技术、多维整合分析策略与人工智能算法的协同发展, 有望逐步揭开人体内完整抗体组库的复杂面纱。这不仅将深化对抗体免疫应答规律的基础理解, 更将为新一代抗体药物发现、疫苗设计与精准免疫诊疗提供更有力的支持。

参 考 文 献

- [1] Mhanna V, Bashour H, Lê Quý K, *et al.* Adaptive immune receptor repertoire analysis. *Nat Rev Meth Primers*, 2024, **4**: 6
- [2] Townsend D R, Towers D M, Lavinder J J, *et al.* Innovations and trends in antibody repertoire analysis. *Curr Opin Biotechnol*, 2024, **86**: 103082
- [3] Yang Y, Chen X, Pan J, *et al.* Pan-cancer single-cell dissection reveals phenotypically distinct B cell subtypes. *Cell*, 2024, **187**(17): 4790-4811.e22
- [4] Cao Y, Su B, Guo X, *et al.* Potent neutralizing antibodies against SARS-CoV-2 identified by high-throughput single-cell sequencing of convalescent patients' B cells. *Cell*, 2020, **182**(1): 73-84.e16
- [5] Yu S, Lin Y, Li Y, *et al.* Systemic immune profiling of Omicron-infected subjects inoculated with different doses of inactivated virus vaccine. *Cell*, 2023, **186**(21): 4615-4631.e16
- [6] Playoust E, Remark R, Vivier E, *et al.* Germinal center-dependent

- and-independent immune responses of tumor-infiltrating B cells in human cancers. *Cell Mol Immunol*, 2023, **20**(9): 1040-1050
- [7] Nicholas C A, Smith M J. Application of single-cell RNA sequencing methods to develop B cell targeted treatments for autoimmunity. *Front Immunol*, 2023, **14**: 1103690
- [8] Koonin E V, Krupovic M. Evolution of adaptive immunity from transposable elements combined with innate immune systems. *Nat Rev Genet*, 2015, **16**(3): 184-192
- [9] Yang J, Luo Y, Shibu M A, *et al.* Cell-penetrating peptides: efficient vectors for vaccine delivery. *Curr Drug Deliv*, 2019, **16**(5): 430-443
- [10] Kumar D, Gautam N, Alnouti Y. Analyte recovery in LC-MS/MS bioanalysis: an old issue revisited. *Anal Chim Acta*, 2022, **1198**: 339512
- [11] Kulyyassov A, Makhsatova S, Kurmanbay A. Protein identification improvement in complex samples using higher frequency MS acquisition and PEAKS software. *Appl Sci*, 2025, **15**(2): 666
- [12] Fiala J, Schuster D, Ollivier S, *et al.* Protein-centric analysis of personalized antibody repertoires using LC-MS-based fab-profiling on a timsTOF. *J Am Soc Mass Spectrom*, 2024, **35**(6): 1292-1300
- [13] Schneck N A, Khatri K, Maust M D, *et al.* Separation techniques for intact antibody analysis by mass spectrometry. *J Liq Chromatogr Relat Technol*, 2022, **45**(17/18/19/20): 271-283
- [14] Akbar R, Robert P A, Pavlović M, *et al.* A compact vocabulary of paratope-epitope interactions enables predictability of antibody-antigen binding. *Cell Rep*, 2021, **34**(11): 108856
- [15] Rodriguez O L, Safonova Y, Silver C A, *et al.* Genetic variation in the immunoglobulin heavy chain locus shapes the human antibody repertoire. *Nat Commun*, 2023, **14**(1): 4419
- [16] Schroeder H W, Cavacini L. Structure and function of immunoglobulins. *J Allergy Clin Immunol*, 2010, **125**(2): S41-S52
- [17] Gogesch P, Dudek S, van Zandbergen G, *et al.* The role of fc receptors on the effectiveness of therapeutic monoclonal antibodies. *Int J Mol Sci*, 2021, **22**(16): 8947
- [18] Ye S, Zhang D, Wu M Z, *et al.* Abstract 1363: Blocking inhibitory receptor ILT3 by an antibody with enhanced Fc function promoted adaptive immunity against hematopoietic malignancy independent of ADCC. *Cancer Res*, 2024, **84**(6_Supplement): 1363
- [19] Akita E M, Nakai S. Production and purification of Fab' fragments from chicken egg yolk immunoglobulin Y (IgY). *J Immunol Methods*, 1993, **162**(2): 155-164
- [20] Boksha I S, Lunin V G, Danilova T A, *et al.* Recombinant endopeptidases IdeS and IdeZ and their potential application. *Biochemistry*, 2023, **88**(6): 731-740
- [21] van Rijswijk D M H, Bondt A, Raafat D, *et al.* Persistent IgG1 clones dominate and personalize the plasma antibody repertoire. *Sci Adv*, 2025, **11**(16): eadt7746
- [22] Vassen V, Tanaka E, Moll K, *et al.* Group A streptococcal SpeB modifies IgA through targeting regions other than the hinge. *Microbiol Spectr*, 2025, **13**(5): e0245024
- [23] Peris-Díaz M D, Deslignière E, Jager S, *et al.* Asymmetric N-glycosylation in the tailpiece of recombinant IgA1. *J Am Chem Soc*, 2024, **146**(50): 34720-34732
- [24] van Rijswijk D M H, Bondt A, de Kat N, *et al.* Direct comparison of the hinge-cleaving proteases IgD and BdpK for LC-MS-based IgG1 clonal profiling. *Anal Chem*, 2024, **96**(1): 23-27
- [25] Moelleken J, Endesfelder M, Gassner C, *et al.* GingisKHAN™ protease cleavage allows a high-throughput antibody to Fab conversion enabling direct functional assessment during lead identification of human monoclonal and bispecific IgG1 antibodies. *MAbs*, 2017, **9**(7): 1076-1087
- [26] Li Y, Reusch S, Van Tol B D M, *et al.* Fc profiling of polyclonal IgG, IgA and IgM by light chain capturing coupled with nanoRP-LC-MS. *JACS Au*, 2026, **6**(1): 644-652
- [27] Srzentić K, Fornelli L, Tsybin Y O, *et al.* Interlaboratory study for characterizing monoclonal antibodies by top-down and middle-down mass spectrometry. *J Am Soc Mass Spectrom*, 2020, **31**(9): 1783-1802
- [28] Yin V, Deslignière E, Mokiem N, *et al.* Not all arms of IgM are equal: following hinge-directed cleavage by online native SEC-orbitrap-based CDMS. *J Am Soc Mass Spectrom*, 2024, **35**(6): 1320-1329
- [29] Shi D, Zhang Y, Xu Z, *et al.* Noninvasive evaluation of EGFR expression of digestive tumors using ^{99m}Tc-MAG₃-cet-F(ab')₂-based SPECT/CT imaging. *Mol Imaging*, 2022, **2022**: 3748315
- [30] Fishman J B, Berg E A. Protein a and protein G purification of antibodies. *Cold Spring Harb Protoc*, 2019, **2019**(1): pdb. prot099143
- [31] Sviatenko O V. Application of immunoglobulin-binding proteins A, G, L in the affinity chromatography. *Biotechnol Acta*, 2014, **7**(2): 34-45
- [32] Lê Quý K, Chernigovskaya M, Stensland M, *et al.* Benchmarking and integrating human B-cell receptor genomic and antibody proteomic profiling. *NPJ Syst Biol Appl*, 2024, **10**(1): 73
- [33] Yu M, Zhu Z, Wang Y, *et al.* A new strategy: identification of specific antibodies for neutralizing epitope on SARS-CoV-2 S protein by LC-MS/MS combined with immune repertoire. *Mol Biomed*, 2022, **3**(1): 20
- [34] Le Bihan T, Nunez de Villavicencio Diaz T, Reitzel C, *et al.* De novo protein sequencing of antibodies for identification of neutralizing antibodies in human plasma post SARS-CoV-2 vaccination. *Nat Commun*, 2024, **15**(1): 8790
- [35] He C, Wong C C L. Novel proteoform discovery by precise semi-De novo sequencing of novel junction peptides. *Anal Chem*, 2023, **95**(28): 10610-10617
- [36] Shuken S R. An introduction to mass spectrometry-based proteomics. *J Proteome Res*, 2023, **22**(7): 2151-2171
- [37] Jiang Y, Rex D A B, Schuster D, *et al.* Comprehensive overview of bottom-up proteomics using mass spectrometry. *ACS Meas Sci Au*, 2024, **4**(4): 338-417
- [38] Dupree E J, Jayathirtha M, Yorkey H, *et al.* A critical review of

- bottom-up proteomics: the good, the bad, and the future of this field. *Proteomes*, 2020, **8**(3): 14
- [39] Bonissone S R, Lima T, Harris K, *et al.* Serum proteomics expands on identifying more high-affinity antibodies in immunized rabbits than deep b-cell repertoire sequencing alone. *BioRxiv*, 2019. DOI: 10.1101/833871
- [40] Sato S, Beausoleil S A, Popova L, *et al.* Proteomics-directed cloning of circulating antiviral human monoclonal antibodies. *Nat Biotechnol*, 2012, **30**(11): 1039-1043
- [41] Peng W, den Boer M A, Tamara S, *et al.* Direct mass spectrometry-based detection and antibody sequencing of monoclonal gammopathy of undetermined significance from patient serum: a case study. *J Proteome Res*, 2023, **22**(9): 3022-3028
- [42] Yang H, Li Y C, Zhao M Z, *et al.* Precision *de novo* peptide sequencing using mirror proteases of Ac-LysargiNase and trypsin for large-scale proteomics. *Mol Cell Proteomics*, 2019, **18**(4): 773-785
- [43] Zhang J, Peng X, Wang F, *et al.* Development of LysargiNase, a mirror trypsin and its application in proteomics. *Chin J Biotechnol*, 2019, **35**(5): 741-748
- [44] Frese C K, Maarten Altelaar A F, Hennrich M L, *et al.* Improved peptide identification by targeted fragmentation using CID, HCD and ETD on an LTQ-Orbitrap Velos. *J Proteome Res*, 2011, **10**(5): 2377-2388
- [45] Guthals A, Bandeira N. Peptide identification by tandem mass spectrometry with alternate fragmentation modes. *Mol Cell Proteomics*, 2012, **11**(9): 550-557
- [46] Yates J R III, Kelleher N L. Top down proteomics. *Anal Chem*, 2013, **85**(13): 6151
- [47] Searfoss R M, Zahn E, Lin Z, *et al.* Establishing a top-down proteomics platform on a time-of-flight instrument with electron-activated dissociation. *J Proteome Res*, 2025, **24**(3): 1230-1240
- [48] Greisch J F, den Boer M A, Lai S H, *et al.* Extending native top-down electron capture dissociation to MDa immunoglobulin complexes provides useful sequence tags covering their critical variable complementarity-determining regions. *Anal Chem*, 2021, **93**(48): 16068-16075
- [49] Greisch J F, den Boer M A, Beurskens F, *et al.* Generating informative sequence tags from antigen-binding regions of heavily glycosylated IgA1 antibodies by native top-down electron capture dissociation. *J Am Soc Mass Spectrom*, 2021, **32**(6): 1326-1335
- [50] Mesonzhnik N, Belushenko A, Novikova P, *et al.* Enhanced N-glycan profiling of therapeutic monoclonal antibodies through the application of upper-hinge middle-up level LC-HRMS analysis. *Antibodies (Basel)*, 2024, **13**(3): 66
- [51] Oates R N, Lieu L B, Kline J T, *et al.* Towards a universal method for middle-down analysis of antibodies *via* proton transfer charge reduction-Orbitrap mass spectrometry. *Anal Bioanal Chem*, 2024, **416**(28): 6463-6472
- [52] Tian Y, Han L, Buckner A C, *et al.* Collision induced unfolding of intact antibodies: rapid characterization of disulfide bonding patterns, glycosylation, and structures. *Anal Chem*, 2015, **87**(22): 11509-11515
- [53] Villafuerte-Vega R C, Li H W, Slaney T R, *et al.* Ion mobility-mass spectrometry and collision-induced unfolding of designed bispecific antibody therapeutics. *Anal Chem*, 2023, **95**(17): 6962-6970
- [54] den Boer M A, Greisch J F, Tamara S, *et al.* Selectivity over coverage in *de novo* sequencing of IgGs. *Chem Sci*, 2020, **11**(43): 11886-11896
- [55] Ma B. Novor: real-time peptide *de novo* sequencing software. *J Am Soc Mass Spectrom*, 2015, **26**(11): 1885-1894
- [56] Yang H, Chi H, Zeng W F, *et al.* pNovo 3: precise *de novo* peptide sequencing using a learning-to-rank framework. *Bioinformatics*, 2019, **35**(14): i183-i190
- [57] Tran N H, Zhang X, Xin L, *et al.* *De novo* peptide sequencing by deep learning. *Proc Natl Acad Sci USA*, 2017, **114**(31): 8247-8252
- [58] Tran N H, Qiao R, Xin L, *et al.* Deep learning enables *de novo* peptide sequencing from data-independent-acquisition mass spectrometry. *Nat Methods*, 2019, **16**(1): 63-66
- [59] Karunratanakul K, Tang H Y, Speicher D W, *et al.* Uncovering thousands of new peptides with sequence-mask-search hybrid *de novo* peptide sequencing framework. *Mol Cell Proteomics*, 2019, **18**(12): 2478-2491
- [60] Qiao R, Tran N H, Xin L, *et al.* Computationally instrument-resolution-independent *de novo* peptide sequencing for high-resolution devices. *Nat Mach Intell*, 2021, **3**(5): 420-425
- [61] Liu K, Ye Y, Li S, *et al.* Accurate *de novo* peptide sequencing using fully convolutional neural networks. *Nat Commun*, 2023, **14**(1): 7974
- [62] Klaproth-Andrade D, Hingerl J, Bruns Y, *et al.* Deep learning-driven fragment ion series classification enables highly precise and sensitive *de novo* peptide sequencing. *Nat Commun*, 2024, **15**(1): 151
- [63] Mao Z, Zhang R, Xin L, *et al.* Mitigating the missing-fragmentation problem in *de novo* peptide sequencing with a two-stage graph-based deep learning model. *Nat Mach Intell*, 2023, **5**(11): 1250-1260
- [64] Yilmaz M, Fondrie W E, Bittremieux W, *et al.* Sequence-to-sequence translation from mass spectra to peptides with a transformer model. *Nat Commun*, 2024, **15**(1): 6427
- [65] Yang T, Ling T, Sun B, *et al.* Introducing π -HelixNovo for practical large-scale *de novo* peptide sequencing. *Brief Bioinform*, 2024, **25**(2): bbae021
- [66] Zhang X, Ling T, Jin Z, *et al.* π -PrimeNovo: an accurate and efficient non-autoregressive deep learning model for *de novo* peptide sequencing. *Nat Commun*, 2025, **16**(1): 267
- [67] Eloff K, Kalogeropoulos K, Mabona A, *et al.* InstaNovo enables diffusion-powered *de novo* peptide sequencing in large-scale proteomics experiments. *Nat Mach Intell*, 2025, **7**(4): 565-579
- [68] Beslic D, Tscheuschner G, Renard B Y, *et al.* Comprehensive evaluation of peptide *de novo* sequencing tools for monoclonal antibody assembly. *Brief Bioinform*, 2023, **24**(1): bbac542
- [69] Wang M, Wang J, Carver J, *et al.* Assembling the community-scale

- discoverable human proteome. *Cell Syst*, 2018, **7**(4): 412-421.e5
- [70] Tran N H, Qiao R, Mao Z, *et al.* NovoBoard: a comprehensive framework for evaluating the false discovery rate and accuracy of *de novo* peptide sequencing. *Mol Cell Proteomics*, 2025, **24**(6): 100939
- [71] Peng W, Pronker M F, Snijder J. Mass spectrometry-based *de novo* sequencing of monoclonal antibodies using multiple proteases and a dual fragmentation scheme. *J Proteome Res*, 2021, **20**(7): 3559-3566
- [72] Gadush M V, Sautto G A, Chandrasekaran H, *et al.* Template-assisted *de novo* sequencing of SARS-CoV-2 and influenza monoclonal antibodies by mass spectrometry. *J Proteome Res*, 2022, **21**(7): 1616-1627
- [73] Halper-Stromberg E, McCall C M, Haley L M, *et al.* CloneRetriever: an automated algorithm to identify clonal B and T cell gene rearrangements by next-generation sequencing for the diagnosis of lymphoid malignancies. *Clin Chem*, 2021, **67**(11): 1524-1533
- [74] Chaudhary P, Chaudhary S, Patel F, *et al.* Validation of a novel NGS based BCR: ABL1 kinase domain mutation detection assay in Indian cohort. *Sci Rep*, 2024, **14**(1): 15745
- [75] Wen G P, He L, Tang Z M, *et al.* Quantitative evaluation of protective antibody response induced by hepatitis E vaccine in humans. *Nat Commun*, 2020, **11**(1): 3971
- [76] Bachmann Salvy M, Santuari L, Schmid-Siegert E, *et al.* Seq2scFv: a toolkit for the comprehensive analysis of display libraries from long-read sequencing platforms. *MAbs*, 2024, **16**(1): 2408344
- [77] Zhang P, Liu B, Zhang S, *et al.* Clinical application of targeted next-generation sequencing in severe pneumonia: a retrospective review. *Crit Care*, 2024, **28**(1): 225
- [78] Chen X, Mohapatra A, Nguyen H T V, *et al.* The presence of broadly neutralizing anti-SARS-CoV-2 RBD antibodies elicited by primary series and booster dose of COVID-19 vaccine. *PLoS Pathog*, 2024, **20**(6): e1012246
- [79] Murugan R, Scally S W, Costa G, *et al.* Evolution of protective human antibodies against *Plasmodium falciparum* circumsporozoite protein repeat motifs. *Nat Med*, 2020, **26**(7): 1135-1145
- [80] Corrado M, Pearce E L. Targeting memory T cell metabolism to improve immunity. *J Clin Invest*, 2022, **132**(1): e148546
- [81] Jian F, Wang J, Yisimayi A, *et al.* Evolving antibody response to SARS-CoV-2 antigenic shift from XBB to JN. 1. *Nature*, 2025, **637**(8047): 921-929
- [82] MacLean A J, Deimel L P, Zhou P, *et al.* Affinity maturation of antibody responses is mediated by differential plasma cell proliferation. *Science*, 2025, **387**(6732): 413-420
- [83] Zaslavsky M E, Craig E, Michuda J K, *et al.* Disease diagnostics using machine learning of B cell and T cell receptor sequences. *Science*, 2025, **387**(6736): eadp2407
- [84] Ye Y, Chen Z, Jiang S, *et al.* Single-cell profiling reveals distinct adaptive immune hallmarks in MDA5+ dermatomyositis with therapeutic implications. *Nat Commun*, 2022, **13**(1): 6458
- [85] Lavinder J J, Wine Y, Giesecke C, *et al.* Identification and characterization of the constituent human serum antibodies elicited by vaccination. *Proc Natl Acad Sci USA*, 2014, **111**(6): 2259-2264
- [86] Nesvizhskii A I. Proteogenomics: concepts, applications and computational strategies. *Nat Methods*, 2014, **11**(11): 1114-1125
- [87] Dupré M, Duchateau M, Sternke-Hoffmann R, *et al.* *De novo* sequencing of antibody light chain proteoforms from patients with multiple myeloma. *Anal Chem*, 2021, **93**(30): 10627-10634
- [88] Mandel-Brehm C, Fichtner M L, Jiang R, *et al.* Elevated N-linked glycosylation of IgG V regions in myasthenia gravis disease subtypes. *J Immunol*, 2021, **207**(8): 2005-2014
- [89] Bondt A, Hoek M, Dingess K, *et al.* Into the dark serum proteome: personalized features of IgG1 and IgA1 repertoires in severe COVID-19 patients. *Mol Cell Proteomics*, 2024, **23**(1): 100690
- [90] Schulte D, Peng W, Snijder J. Template-based assembly of proteomic short reads for *de novo* antibody sequencing and repertoire profiling. *Anal Chem*, 2022, **94**(29): 10391-10399
- [91] Snapkov I, Chernigovskaya M, Sinitcyn P, *et al.* Progress and challenges in mass spectrometry-based analysis of antibody repertoires. *Trends Biotechnol*, 2022, **40**(4): 463-481
- [92] Papanastasiou D, Kounadis D, Lekkas A, *et al.* The omnitrapp platform: a versatile segmented linear ion trap for multidimensional multiple-stage tandem mass spectrometry. *J Am Soc Mass Spectrom*, 2022, **33**(10): 1990-2007
- [93] Smyrnakis A, Levin N, Kosmopoulou M, *et al.* Characterization of an omnitrapp-orbitrap platform equipped with infrared multiphoton dissociation, ultraviolet photodissociation, and electron capture dissociation for the analysis of peptides and proteins. *Anal Chem*, 2023, **95**(32): 12039-12046

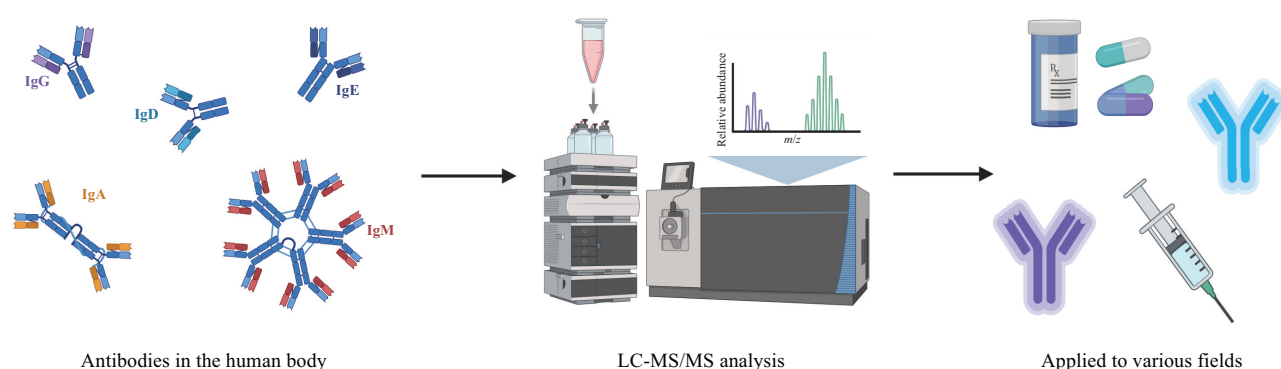
Mass Spectrometry-based Antibody Sequencing Technologies*

LIU Sheng-Mei^{1,2)}, XUE Peng^{2)**}, WANG Xiao-Jian^{1)**}

¹⁾College of Life Sciences, Yan'an University, Yan'an 716000, China;

²⁾Guangzhou National Laboratory, Guangzhou 510005, China)

Graphical abstract



Abstract Antibodies play a critical role in adaptive immune responses and serve as key components in disease diagnosis and treatment. These molecules exhibit dynamic post-translational modifications (PTMs), such as glycosylation and phosphorylation, which regulate their effector functions. To date, nearly all of our knowledge about antibody repertoires has come from B cell receptor (BCR) sequencing (BCR-seq), which facilitates the profiling of clonal composition and the tracing of maturation trajectories within B-cell repertoires. However, circulating antibodies found in bodily fluids—such as serum, saliva, milk, mucosal secretions, and cerebrospinal fluid—exhibit diversities and specificities beyond what BCR-seq alone can predict. Therefore, identifying and quantifying antibody clonotypes at the protein level could enhance diagnosis, prognosis, and treatment strategies in personalized medicine. The critical gap between genotype and phenotype necessitates complementary methodologies that enable the direct characterization of antibody proteins in their native functional states. Mass spectrometry (MS)-based antibody repertoire sequencing (Ab-seq) is currently the only feasible approach for this task and primarily includes database-dependent methods—such as bottom-up, middle-down, and top-down approaches—as well as database-independent *de novo* sequencing technology. These strategies enable multi-level, high-precision characterization ranging from peptides and domains to intact antibody molecules. Unlike the shotgun strategy commonly used in routine proteomics, obtaining full sequences of all antibodies presents unique

* This work was supported by grants from the Project of Guangzhou National Laboratory (GZNL2023A03007) and Science and Technology Project of Shaanxi Province (S2025-YF-YBNY-1080).

** Corresponding author.

XUE Peng. Tel: 86-13219387757, E-mail: xue_peng@gzlab.ac.cn

WANG Xiao-Jian. Tel: 86-15891680366, E-mail: dooble@163.com

Received: October 12, 2025 Accepted: March 6, 2026

challenges. It requires specialized methodological adaptations to address issues related to dynamic range, sequence variation, and sample complexity. This review introduces the technical principles, methodological workflows, and recent applications of various mass spectrometry-based antibody repertoire sequencing (Ab-seq) strategies, with a focus on approaches designed to improve sequence coverage and identification accuracy. These include multi-enzyme digestion, hybrid fragmentation methods, and artificial intelligence-assisted *de novo* sequencing. By systematically comparing database-dependent techniques—such as bottom-up, middle-down, and top-down approaches—with database-independent *de novo* sequencing, this review outlines their respective advantages and limitations in terms of sample throughput, sequence coverage, post-translational modification characterization, and data analysis complexity. In addition, this review discusses emerging technological trends, including the integration of ion mobility separation, native mass spectrometry, and artificial intelligence-driven data interpretation, which are expected to enhance the depth and accuracy of antibody characterization. Although current methods continue to face challenges related to sample complexity, dynamic range, and unambiguous sequence variant assignment, we emphasize the importance of integrating BCR-seq and Ab-seq data to construct gene-protein association maps. These maps help validate sequence accuracy and facilitate epitope discovery. This dual-platform strategy helps bridge the gap between genotype and phenotype, thereby enhancing both the resolution and scope of antibody repertoire studies. Such an integrative approach also offers a valuable tool for therapeutic antibody development, structure-function analysis, and precise evaluation of vaccine efficacy.

Key words antibody, mass spectrometry analysis, BCR-seq, Ab-seq

DOI: 10.3724/j.pibb.2025.0446

CSTR: 32369.14.pibb.20250446