



## 血浆低分子量蛋白与多肽的研究进展\*

李娜<sup>1,2)</sup> 蔡潭溪<sup>1,2)</sup> 杨福全<sup>1,2)\*\*</sup><sup>(1)</sup> 中国科学院生物物理研究所, 北京 100101; <sup>(2)</sup> 中国科学院大学, 北京 100049

**摘要** 血浆是临床上使用频率最高的体液样本, 具有采样无创、使用便捷以及内含物丰富等特点, 是挖掘疾病相关潜在生物标志物的重要来源。血浆中低分子量蛋白与多肽 (low molecular weight proteins and peptides, LMWPs) 包括细胞因子、生长因子、多肽类激素和蛋白质降解产物等, 其以分子质量较小、序列可表征、来源广泛等优势, 得到了研究人员的青睐。近年来, 随着高分辨率、高灵敏度质谱仪与计算科学的不断发展, 新技术新方法的融入, LMWPs 相关研究也有了新的发展。本文结合当前本领域的研究热点, 介绍了血浆 LMWPs 的分类, 包括完整小蛋白、蛋白质降解物、神经肽、免疫多肽、载体蛋白结合的 LMWPs 和小开放阅读框编码的多肽, 同时总结了血浆 LMWPs 在富集方法、研究策略、功能研究以及疾病相关生物标志物方面的研究进展, 并对血浆 LMWPs 面临的挑战和未来的研究方向进行了讨论。

**关键词** 血浆低分子量蛋白与多肽, 富集方法, 质谱, 生物标志物, 小开放阅读框编码的多肽

**中图分类号** Q51, Q6-33, Q592

**DOI:** 10.16476/j.pibb.2021.0241

低分子量蛋白与多肽 (low molecular weight proteins and peptides, LMWPs) 通常指复杂生物样本中分子质量  $\leq 30$  ku 的所有蛋白质及多肽, 也被称作多肽组。LMWPs 是继基因组、转录组及蛋白质组之后, 更复杂的一种分子组合, 造成其复杂程度如此高的原因主要有: a. 体内具有多种蛋白酶 (超过 550 种)<sup>[1]</sup>; b. 底物多样性, 即同种酶有多种底物; c. 体内存在多种蛋白酶抑制剂和激活剂。在庞大的蛋白酶调节网络中, 体内上百万的蛋白质变体 (proteoforms) 可产生数量巨大的酶解产物<sup>[2]</sup>。血浆作为循环系统的一部分, 不仅可以协助细胞、组织或器官之间进行物质交换, 还可以携带大量来源不同的多肽、蛋白质及其降解物等, 而 LMWPs 以其分子质量较小的优势, 可通过血脑屏障、肿瘤等从组织或病灶进入血浆。与血浆蛋白相比, LMWPs 的组成或丰度差异与疾病的相关性更强<sup>[3]</sup>。因此, 血浆 LMWPs 可用于病理生理的探究、生物标志物的挖掘、治疗靶点的发现等。此外, 通过血浆 LMWPs 也可以发现新的分子并进一步探究其功能, 例如新型活性肽<sup>[4]</sup> 以及小开放阅读框编码的多肽 (sORF-encoded peptides, SEPs)。本文立足于当前的研究热点, 综述 LMWPs 的分类 (图 1) 及其应用, 以期对血浆 LMWPs 的研究与发

展提供参考。

## 1 LMWPs 的分类

### 1.1 完整小蛋白

血浆中常见的完整小蛋白包括激素、白介素、趋化因子、生长因子和干扰素等, 它们可参与调控细胞生长、增殖、分化、炎症、组织修复和免疫反应等多种重要生物学过程, 在肿瘤治疗<sup>[5]</sup>、造血功能障碍<sup>[6]</sup>、自身免疫性疾病<sup>[7]</sup> 等方面具有广阔的应用前景。近年来, 完整小蛋白的研究热点是表征样本中的蛋白质变体, 即来自同一基因所有形式的蛋白质分子的集合, 包括由遗传变异、RNA 转录本选择性剪接和翻译后修饰 (post-translational modifications, PTMs) 引起的各类变化<sup>[8]</sup>。不过, 随着分子质量增加, 蛋白质的离子化效率降低、质谱信号抑制作用与扩散作用增强, 使得谱图解析更加困难, 从而阻碍了完整蛋白质的检测<sup>[9]</sup>。因此, 在组学水平表征血浆中小蛋白的完整序列是一件充满挑战性的工作。

\* 国家重点研发计划 (2018YFA0507801) 资助项目。

\*\* 通讯联系人。

Tel: 010-64888581, E-mail: fqyang@ibp.ac.cn

收稿日期: 2021-07-27, 接受日期: 2021-10-13

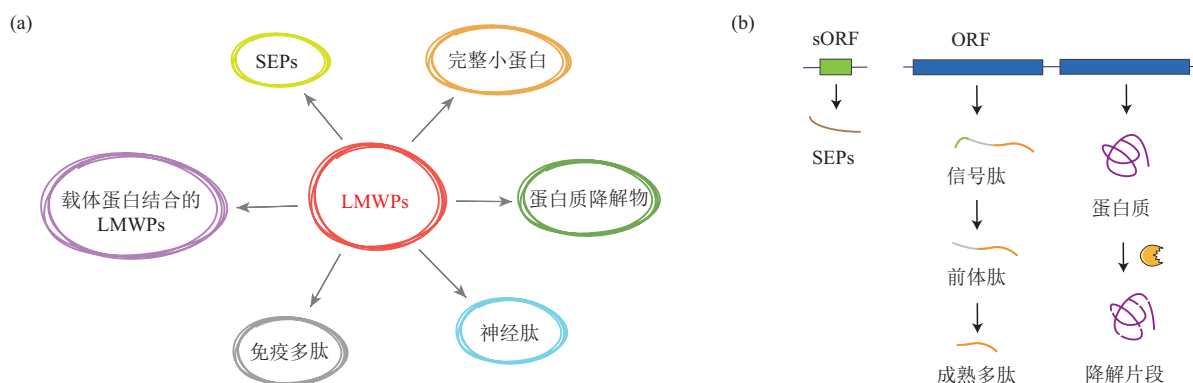


Fig. 1 The classify of plasma LMWPs

图1 血浆LMWPs的组成

(a) LMWPs主要分为完整小蛋白、蛋白质降解物、神经肽、免疫多肽、载体蛋白结合的LMWPs及SEPs几类。(b) SEPs由小开放阅读框(sORF)直接编码产生, 而传统多肽由前体蛋白经酶解产生。

## 1.2 蛋白质降解物

大部分 LMWPs 是由蛋白质经酶解作用产生的多肽组成。根据催化机理, 蛋白酶可分为 5 大类, 包括天冬氨酸类、半胱氨酸类、丝氨酸类、金属蛋白酶类和其他类。在哺乳动物体内, 蛋白酶可选择性地切割特定的蛋白质底物, 以影响细胞的行为、存活和死亡<sup>[10]</sup>。研究发现, 它们在肿瘤中的功能更加复杂与多样<sup>[11-12]</sup>。转移性癌症的多种特征是蛋白酶活性发生变化直接导致的结果。蛋白酶通过介导蛋白质、细胞外基质降解过程来影响肿瘤细胞侵入间质、血管生成和转移; 级联蛋白水解相互作用控制着细胞凋亡, 癌细胞在其发展为恶性肿瘤的过程中必须逃脱该调控过程; 蛋白酶通过特异性裂解前体蛋白来激活其他蛋白酶或信号分子, 如激酶和细胞表面受体, 控制信号在肿瘤细胞与其他细胞之间进行传导。肿瘤中失调的蛋白酶可改变 LMWPs 的分子组成, 而这些 LMWPs 不同于高分子质量蛋白质, 易分泌至细胞外间质, 通过血管屏障最终释放到循环系统中。Li 等<sup>[3]</sup> 研究发现, 与健康对照相比, 羧肽酶 N (carboxypeptidase N, CPN) 在乳腺肿瘤组织中的表达水平上调, 但在血浆中的含量没有显著变化。不过其催化产物 C3f<sub>R1310</sub>-L<sub>1319</sub> 在血浆中的含量显著增加, 且该增加主要与早期乳腺癌相关, 表明肿瘤微环境中失调的蛋白酶催化而产生的多肽片段可在血浆中被检测到, 且含量与肿瘤早期的病理变化有相关性, 提示蛋白降解物可作为疾病早期的诊断标志物。

## 1.3 神经肽

在神经元之间传递信号的一类多肽被称为神经

肽。神经肽来源于前体蛋白质, 以一种受调控的方式释放, 并通过与细胞表面的 G 蛋白偶联受体或酶偶联受体 (如胰岛素受体) 结合, 在中枢和周围神经系统中发挥作用。它们作用的靶点包括神经元、胶质细胞或非神经元靶细胞, 如腺体或肌细胞<sup>[13]</sup>。不同于经典的神经递质, 这些神经肽生理功能的调节不是通过特定的突触再摄取机制实现, 而是通过降解机制实现。它们不仅在神经元中表达, 也在大脑其他类型细胞 (如胶质细胞) 和外周组织细胞 (如内分泌细胞) 中表达<sup>[14]</sup>, 影响着广泛的生理过程, 包括进食、体重调节、昼夜节律、睡眠、疼痛、恐惧、焦虑、学习和记忆, 以及奖励机制等<sup>[15]</sup>。中国肥胖人群队列研究显示, 神经肽 Y (neuropeptide Y, NPY) 在血清中的含量增加与代谢异常型肥胖 (metabolically unhealthy obesity, MUO) 表型存在正相关性<sup>[16]</sup>, 且 NPY 过表达可导致转基因小鼠出现肥胖表型和胰岛素抵抗<sup>[17]</sup>, 表明 NPY 与肥胖相关, 提示神经肽在血液循环中的含量变化可作为评价生理功能失调程度的指标。此外, 值得注意的是, 根据人类基因组分析, 100 多个 G 蛋白偶联受体的配体未知<sup>[18]</sup>, 因此识别特定细胞、组织或体液中与 G 蛋白偶联受体结合的配体, 可用于新型神经肽的发现<sup>[19]</sup>。

## 1.4 免疫多肽

免疫多肽组指由主要组织相容性复合体 (MHC) 分子 (在人类中称为人类白细胞抗原, HLA) 呈递到细胞表面的所有多肽。其中, 部分呈递到细胞表面的 HLA 分子及其结合的多肽被释放到血浆中, 形成血浆可溶性 HLA 多肽组

(soluble HLA-peptidome, sHLA)。sHLA与呈递到细胞表面的膜多肽组(membranal peptidomes, mHLA)类似,均由细胞蛋白质降解产生的多肽组成<sup>[20]</sup>。当处于病理状态时,细胞产生并降解疾病相关的蛋白质,然后将降解产生的多肽呈递出来,因此HLA多肽组可作为癌症免疫治疗靶点候选物来源之一<sup>[21]</sup>。此外,正常情况下,HLA多肽组在人体内保持相对稳定,但在特定疾病影响下,细胞会向血浆中释放比正常状态下更多的sHLA分子,所以血浆sHLA与疾病严重程度相关,可用于疾病潜在生物标志物的挖掘<sup>[22]</sup>。Shraibman等<sup>[23]</sup>通过对142份多形性胶质母细胞瘤(glioblastoma multiforme, GBM)患者血浆样本的sHLA和其中10例肿瘤样本的mHLA进行大规模质谱鉴定分析,发现无论是分子组成还是含量sHLA和mHLA均有较强的相关性。此外,经过数据库筛选,作者也提供了多个GBM早期诊断的潜在生物标志物(如仅在GBM中检测到的来自SOX11的多肽AHSASEQQL和NFSDLVFTY)以及免疫治疗的潜在靶点(如来自ETV5的6条多肽)。这强有力地说明,在癌症生物标志物和免疫抑制剂候选物筛选中,血浆sHLA研究意义重大。总之,系统表征HLA多肽组有助于诊断、预防和治疗感染性疾病、自身免疫病和癌症等。

相较于其他类型的多肽,sHLA多肽在富集技术上更具有优势,它们主要是利用HLA-A/B/C单克隆抗体(W6/32)进行富集纯化,该方法简单且高效。

### 1.5 载体蛋白结合的LMWPs

血浆中蛋白质浓度的高动态范围是限制血浆蛋白质组及LMWPs研究的重要因素之一,例如白蛋白(albumin)是血浆中丰度最高的蛋白质,含量高达50 g/L,远大于细胞因子等低丰度蛋白质的含量(ng/L),严重阻碍了LMWPs的发现与检测。因此,在富集和提取血浆LMWPs时,去除被视为“污染蛋白”的高丰度蛋白质是实验方法中的首要步骤。不过,白蛋白具有多种功能,包括维持血浆渗透压,作为脂蛋白、小分子和药物的载体等,同时,因分子质量高于肾脏过滤和清除功能的界限,白蛋白成为LMWPs常见的天然循环载体。Mehta等<sup>[24]</sup>在其研究中提出了以下几个发现与观点:  
a. LMWPs在循环载体蛋白上的累积极大地放大了潜在生物标志物在血浆中的总浓度;  
b. 血浆LMWPs的浓度主要由载体蛋白而非其本身清除速

率决定;  
c. 与特定载体蛋白结合的LMWPs可能包含重要的诊断信息。随后,许多研究人员将焦点转向载体蛋白结合的LMWPs,试图分析载体蛋白结合的LMWPs与各类疾病之间的关系,包括卵巢癌<sup>[25]</sup>、子宫内膜癌<sup>[26]</sup>、阿尔茨海默病<sup>[27]</sup>、脑卒中<sup>[28]</sup>、酒精滥用<sup>[29]</sup>等。例如,Kikuchi等<sup>[26]</sup>在研究子宫内膜癌患者和健康对照的血清白蛋白结合的多肽时,发现3种多肽( $m/z$  4769、6254和11792)的质谱信号强度在两组之间存在显著性差异。在这些研究中,质谱技术分析载体蛋白结合的LMWPs已被证明是一种识别疾病潜在生物标记物的有效手段。

### 1.6 小开放阅读框(sORF)编码的多肽

一直以来,大多数寻找开放阅读框(ORF)的算法都将300个核苷酸或100个氨基酸作为最小检测限,以降低假阳性编码区被检测的可能性。然而,这种过滤方法直接导致一些小的具有蛋白质编码潜力的转录本被错误地归类为更大类别的非编码RNA(ncRNAs)。近年来,核糖体图谱结合质谱技术加速了这类SEPs或小蛋白(microproteins)的发现<sup>[30]</sup>。与经典的多肽类激素或神经肽不同,SEPs不是由前体蛋白经酶解产生,而是由sORF直接翻译而来,也属于LMWPs。

迄今为止,许多研究发现了SEPs的存在,并表明这类多肽在生长发育<sup>[31]</sup>、机体调节<sup>[32]</sup>,以及疾病发生中(如癌症<sup>[33]</sup>)具有重要作用。例如,由pri基因编码的SEP,可以引发转录因子Svb的截短,使之从阻遏物转变为激活物,影响果蝇的表皮分化过程<sup>[34]</sup>;在D'lima等<sup>[35]</sup>研究中,lncRNA翻译而成的NoBody被证实位于P-bodies中,与mRNA脱帽蛋白(尤其是EDC4)存在相互作用,可调节P-body的数量。该研究揭示,NoBody是mRNA脱帽复合物中的一个新成分,为SEPs作为调节因子参与重要生物过程提供了新的证据;Pauli等<sup>[36]</sup>发现,在斑马鱼原肠胚形成过程中,lncRNA编码的多肽Toddler,通过与APJ/Apelin受体结合激活G蛋白偶联信号,促进细胞运动;Humanin是Hashimoto等在筛选阻止阿尔茨海默病患者神经元死亡的相关因子时发现的,它可以影响Bax定位,进而阻止细胞死亡<sup>[37-38]</sup>。目前,SEPs的存在已被广泛接受并引起了越来越多的关注。

虽然生物信息学结合核糖体图谱技术在多种物种中预测发现了大量SEPs<sup>[39-40]</sup>,但是证实这些SEPs的表达较难。质谱技术作为一种高通量的发

现及验证手段,能够检测蛋白质/多肽的分子质量和氨基酸序列,目前已广泛应用于SEPs表征和验证中。不过,SEPs分子质量小且浓度低,它们在高丰度蛋白质的掩盖下很难被有效地富集和鉴定,所以如何有效地富集SEPs至关重要。为此, Ma等<sup>[41]</sup>比较了几种富集K562细胞中SEPs的方法,包括酸沉淀、30 ku分子截断超滤法及反相C8固相萃取法,结合其他参数的优化,建立了一套高效表征细胞中SEPs的方法。截至目前,细胞、组织、体液等样本中均有大规模SEPs的鉴定报告。Zhang等<sup>[42]</sup>系统开展了SEPs富集和鉴定方法学研究,从人源和鼠源细胞或组织中共鉴定到762条SEPs。Li等<sup>[43]</sup>首次报道了在血浆中鉴定到19条SEPs,而Cai等<sup>[44]</sup>也首次报道了在血浆外泌体中鉴定到48条SEPs。

## 2 LMWPs研究进展及发展趋势

### 2.1 LMWPs富集方法进展

相较于蛋白质组学持续、蓬勃的发展,LMWPs相关研究处于相对滞后状态,主要是因为血浆样本具有蛋白质丰度的高动态范围和组成的高度复杂性。长期以来,研究血浆LMWPs受限于富集与分析方法。不过,近年来,随着新方法、新技术的应运而生,LMWPs的研究也取得可喜进展。

在富集血浆LMWPs的策略中,去除高丰度蛋白质是关键的一步,常用的方法有:有机溶剂沉淀法、分子截断超滤法、固相萃取法、亲和法和凝胶

分离法(图2a)。虽然这些方法在一定程度上可降低血浆样本的复杂性和动态范围,但高丰度蛋白质去除的效果有限,质谱定性分析的再现性差,例如,固相萃取法三次技术重复的结果仅有40%重叠<sup>[45]</sup>。所以,LMWPs相关研究迫切需要一种更高效、稳定、便捷的富集或分离方法,以进一步降低样品的复杂性,便于利用组学手段研究其组成和变化情况。2017年,悉尼大学Parker等<sup>[46]</sup>运用三氯乙酸(trichloroacetic acid, TCA)沉淀法开展了运动相关的血浆多肽组研究。在这项研究中,该团队首先将TCA沉淀法与乙腈沉淀法、丙酮沉淀法和分子截断超滤法进行了比较,发现TCA法鉴定的LMWPs数目最多(近550个)。虽然TCA沉淀法展现出较好的鉴定结果,但是血浆样本基质的复杂性,例如大量脂质小分子存在,依然干扰着LMWPs的富集与鉴定。2019年,Harney等<sup>[47]</sup>发表了SPEA (small-protein enrichment assay)方法学研究工作。该方法的卓越之处是在单针非靶向进样分析的前提下,共定性到100多种低分子质量蛋白质,有效地提高了小蛋白的鉴定率。虽然该方法结合分子排阻色谱法实现了小蛋白与载体蛋白的分离、高丰度蛋白质和脂质的去除,但整个实验流程过于繁琐,难以应用于大规模临床样本制备中。Li等<sup>[43]</sup>提出了一种新的富集血浆LMWPs的方法——顺序沉淀脱脂法(sequential precipitation and delipidation, SPD),该方法的再现性及定量稳定性显著高于经典的乙腈沉淀法与近期报道的TCA沉淀法(图2b)。

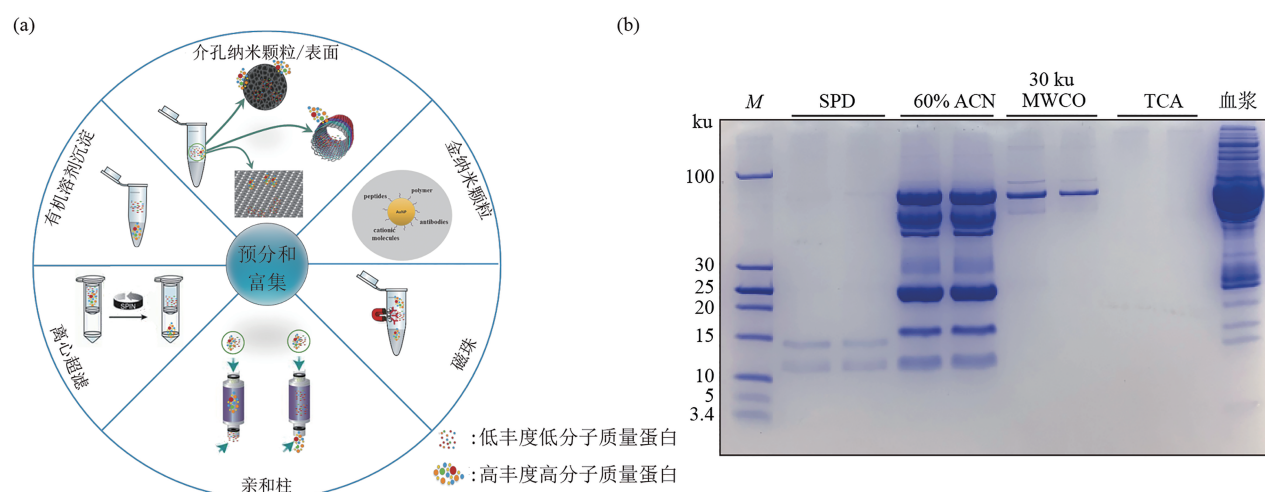


Fig. 2 Strategies for enrichment of plasma LMWPs<sup>[43, 48]</sup>

图2 血浆中LMWPs的富集策略<sup>[43, 48]</sup>

(a) 富集血浆LMWPs常用的方法。(b) Tricine-SDS-PAGE展示SPD法、60%乙腈(ACN)法、TCA沉淀法与30 ku分子截断超滤法(30 ku MWCO)富集的血浆LMWPs。

## 2.2 LMWPs鉴定策略进展

除了富集方法上的进步, 质谱鉴定策略也趋于多样化, 不断更新。基质辅助激光解吸电离飞行时间质谱 (matrix-assisted laser desorption/ionization time-of-flight mass spectrometry, MALDI-TOF-MS) 是表征 LMWPs 常用的质谱技术, 具有灵敏度高、检测快速、通量高和成本低等优势, 已广泛应用于癌症<sup>[49-51]</sup>、慢性肾病<sup>[52]</sup>、感染性疾病<sup>[53-54]</sup>、认知障碍<sup>[55]</sup>, 甚至是当前全球流行的 COVID-19<sup>[56]</sup> 等疾病潜在生物标志物的挖掘中。不过, 该技术无法在较宽的质量范围内实现多肽组可靠的定性定量分析。随着高分辨、高灵敏度质谱仪器与相应软件的不发展, 多肽组学研究也变得越来越全面和系统。

LMWPs 鉴定策略的进步主要包括: a. 蛋白质组学研究策略的进步。在自下而上蛋白质组学 (bottom-up proteomics, BUP) 技术中, 质谱采集速度的大幅提升显著增加了蛋白质组的鉴定深度, 推动了该技术的广泛应用; 作为表征完整小蛋白的理想手段之一, 自上而下蛋白质组学 (top-down proteomics, TDP) 的技术策略是结合多种质谱碎裂模式, 直接使样本中的 LMWPs 实现有效碎裂并最终完成氨基酸序列的完整解析<sup>[57]</sup>。相较于 BUP, TDP 的技术优势是保留了 LMWPs 的原始序列信息, 但是由于完整 LMWPs 的分子质量较大, 该技术难以产生信息丰富的串联质谱数据, 不过新一代 Orbitrap 仪器的开发为 TDP 技术的快速发展奠定了坚实的基础。b. 质谱碎裂模式的多样化应用。例如 Parker 等<sup>[46]</sup> 利用具有高效分离能力的纳流速超高效毛细管液相色谱与新一代超高分辨三合一质谱联用技术, 在 EThcD 和 HCD 两种质谱碎裂模式下分别对样本进行数据采集, 极大地提高了多肽的鉴定数目。c. 质谱数据采集方法的进步。近年来, 随着数据非依赖采集 (data-independent acquisition, DIA) 技术的蓬勃发展, LMWPs 的相关研究也得到了进一步延伸。Lin 等<sup>[58]</sup> 首次运用 DIA 技术开展了肾透明细胞癌相关的血清 LMWPs 研究, 在该临床研究中, 研究人员将 DIA 技术与传统的数据依赖采集 (data-dependent acquisition, DDA) 技术进行了对比, 虽然两种技术在 3 次重复实验中鉴定的多肽数目相差无几, 但 DIA 技术定性结果的重现性与定量结果的稳定性均显著高于 DDA, 具有巨大的发展潜力。d. 各类数据检索软件及算法的进步。例如, 具有 *de novo* 未知序列解序功能的

PEAKS 软件<sup>[59]</sup>、第一个 TDP 质谱鉴定算法 ProSight<sup>[60]</sup>、高谱图解析率的新一代开放式搜索引擎 Open-pFind<sup>[61]</sup> 以及 DIA 数据解析软件 Spectronaut<sup>[62]</sup> 等。

## 2.3 LMWPs功能研究进展

从葡萄糖调节 (胰岛素) 到免疫识别 (MHC 分子), 以各类生物活性肽形式存在的 LMWPs, 构成了众多生理稳态调节机制。在细胞和组织中, 蛋白酶/肽酶密切参与了这些多肽的产生、降解和信号转导, 调控着 LMWPs 的组成、浓度及生物活性。探究体内 LMWPs, 一方面可发现具有生物活性的新分子; 另一方面可揭示内源性多肽代谢的途径以及体内调节生物活性肽水平和信号转导的新靶点。

Taguchi 等<sup>[4]</sup> 利用肽组学技术鉴定到几千个多肽, 通过化学合成多肽序列及功能研究, 他们发现了 3 种来自表皮分化标记蛋白 suprabasin 的新型生物活性肽。当向小鼠腹腔注射 SBSN\_HUMAN [279-295] 时, 该多肽可有效地抑制食物/水的摄入, 并诱导小鼠运动, 而 SBSN\_HUMAN [225-237] 和 SBSN\_HUMAN [243-259] 通过激活血管细胞内 NF- $\kappa$ B 信号, 刺激炎症前细胞因子的表达。该研究基于从头测序的数据库检索方式生成多肽库, 结合功能研究, 为发现血浆中未知生物活性肽提供了新的途径。

研究蛋白酶-多肽的关系可以揭示蛋白酶调节生理功能的机制以及疾病潜在的治疗靶点。二肽基肽酶 IV (dipeptidyl peptidase IV, DPPIV 或 CD26) 参与了多种激素和趋化因子生物活性的调节, 包括胰高血糖素样肽 1 (glucagon-like peptide 1, GLP-1) 和葡萄糖依赖性促胰岛素多肽 (glucose-dependent insulinotropic polypeptide, GIP)<sup>[63-64]</sup>。尽管完整的 GLP-1-(7-36)-酰胺和 GIP-(1-42) 可以增强葡萄糖刺激的胰岛素分泌, 但是当它们 N 端二肽在体外被 DPPIV 剪切后即失去这种活性<sup>[65]</sup>。N 端截短的 GLP-1 和 GIP 可以在血液循环中被检测到<sup>[64, 66]</sup>, 提示蛋白质降解可能是调节它们生理活性的机制。因此, Marguet 等<sup>[67]</sup> 利用 CD26 基因靶向失活的小鼠探究了该酶在血糖控制中的作用, 发现在葡萄糖的刺激下, CD26<sup>-/-</sup>小鼠血液中的胰岛素与完整形式的 GLP-1 含量增加, 同时 CD26 酶活性特异性抑制剂可以提高野生型小鼠的葡萄糖耐受性, 表明 CD26 很可能通过调节 GLP-1 或其他底物的活性控制血糖, 提示 CD26 可作为 II 型糖尿病的潜在治疗

靶点。

成纤维细胞生长因子 21 (fibroblast growth factor 21, FGF21) 是一种参与调节葡萄糖稳态、胰岛素敏感性、脂质代谢和体重的激素, 是治疗代谢紊乱的潜在靶点。但是 FGF21 在血液循环中的快速水解降低了其潜在的生物活性。为探究 FGF21 的调控机制, Zhen 等<sup>[68]</sup> 利用质谱技术系统地研究了人 FGF21 的加工过程, 证明了 DPPIV 是负责其在 Pro-2 和 Pro-4 残基后 N 端裂解的主要蛋白酶, 以及成纤维细胞活化蛋白 (fibroblast activation protein, FAP) 是负责 FGF21 在 Pro-171 后 C 端裂解的蛋白酶。FAP 可使 FGF21 失活, 该功能的发现不仅扩展了 FGF21 的信号通路, 也在治疗上为增强 FGF21 生物活性提供了新的方法。与之类似, Sierra 等<sup>[69]</sup> 发现, 细胞表达、分泌的完整趋化因子基质衍生因子 1 $\alpha$  (stromal cell-derived factor 1 $\alpha$ , SDF-1 $\alpha$ ) 经血清处理后以截短体的形式存在, 同时, 与其结构对应的生物功能随之降低, 揭示了蛋白质降解是调节 SDF-1 $\alpha$  生物功能的重要机制。该团队还进一步从人血清中分离、纯化和鉴定了这种特异性剪切 SDF-1 $\alpha$  C 端赖氨酸的蛋白酶, 即羧肽酶 N<sup>[70]</sup>。SDF-1 $\alpha$  是造血、淋巴细胞归家、B 细胞生长和血管生成的重要调节因子, 该发现阐明了 SDF-1 $\alpha$  功能活性精细调控的重要分子机制。

## 2.4 疾病相关生物标志物研究进展

临床上, 癌症等重大疾病的高发病率及高死亡率威胁着全球人口的健康, 而当前老龄化社会的发展也凸显了各类慢性病对人们生活质量的影响。因此, 目前迫切需要发展各类疾病早期诊断的生物标志物。Anderson<sup>[71]</sup> 分析了 2008 年美国食品和药物管理局批准的所有蛋白质类的实验项目, 结果显示在这些实验中来自血浆的蛋白质共有 109 个, 其中第一类高丰度蛋白质约占 50%, 由组织分泌到血浆中的蛋白质约占 25%, 剩余则是一些异常分泌的蛋白质或者激素、受体等。在回顾利用血浆蛋白质组学技术挖掘生物标志物的研究时, Geyer 等<sup>[72]</sup> 提出血浆蛋白质组鉴定深度达 1 500 个蛋白质时才能覆盖由组织分泌出来的蛋白质, 在此鉴定深度下, 将蛋白质按丰度高低进行排序, 依照前 300 个蛋白质中约 23% 的蛋白质为生物标志物这一比例进行推算, 在后续 1 200 个蛋白质中应该还存在许多生物标志物。

一般认为, 源自组织的蛋白质分子质量较大, 无法以被动运输的方式通过内皮扩散到循环系统

中, 而 LMWPs 以其分子质量较小的优势, 可作为细胞间信息交流的媒介。同时, LMWPs 富含更多可用于临床诊断的信息, 包括多肽数目、序列信息、蛋白质归属、酶切位点信息等。因此, 来源广泛的 LMWPs 中尚有大量生物标志物有待挖掘, 具有巨大发展潜力。

在肿瘤微环境中, 蛋白质水解异常可促进肿瘤生长、细胞外和组织的重塑、炎症、细胞侵袭和转移<sup>[73]</sup>。因此, 与癌症生物学相关的蛋白酶催化而产生的蛋白质降解物可以用作指示癌症发生、发展及治疗情况的生物标志物。据 Petricoin 等<sup>[74-75]</sup> 报道, 在卵巢癌与前列腺癌中, 血浆/血清中多肽组合模式在癌症及其早期诊断中展现的临床特异性与灵敏度均显著高于临床上已有的生物标志物——分别是癌症抗原 (cancer antigen 125, CA125)<sup>[74]</sup> 与前列腺特异性抗原 (prostate-specific antigen, PSA)<sup>[75]</sup>。随后, Zhang 等<sup>[76]</sup> 利用交叉验证和独立的多中心样本鉴定出 3 个 LMWPs, 这些潜在生物标志物结合 CA125 提高了早期卵巢癌检测的灵敏度与特异性。此外, Bedin 等<sup>[51]</sup> 探索了与结肠癌进展相关的多肽组, 通过研究发现, 蛋白质降解模式在结肠癌的不同阶段各不相同。相较于健康对照, 患者的蛋白酶活性更高, 揭示了疾病进展与 LMWPs 变化的相关性。癌症的早期检测是治疗恶性肿瘤的关键, 因此, 早期生物标志物的发现将极大地帮助患者进行疾病控制, 显著提高患者的生存率。

除了癌症之外, 基于质谱技术探究 LMWPs 已被广泛应用于各类疾病潜在生物标志物的发现中, 例如代谢综合征<sup>[77-78]</sup>、精神类疾病<sup>[79-80]</sup> 和感染性疾病<sup>[81]</sup> 等。不过, 随着研究的逐渐增多与深入, 越来越多的结果表明, 血浆中与不同疾病相关的差异多肽大部分源自高丰度蛋白质, 这些蛋白质似乎没有特异性, 因为它们在不同类型的疾病中也有差异表达。这可能是由于高丰度蛋白质确实与多种疾病之间存在显著性关联<sup>[82]</sup>。另外, 在复杂生物体中, 多种因素调控着多肽的产生, 因此, 与疾病生理病理直接相关的或许不是差异多肽, 而是与多肽产生相关的蛋白酶或其调控因子, 丰度发生改变的多肽是疾病间接相关标志物; LMWPs 富集方法的局限性也是一个重要原因, 血浆样本中高丰度蛋白质及其衍生的多肽片段的覆盖极大地掩盖了低丰度 LMWPs 的鉴定, 现有的检测深度降低了其他可能具有诊断作用的低丰度多肽的检出率, 这也再次强

调了推动方法学及研究策略进展的必要性。不过, 从多种研究结果来看, 比起单一的生物标志物, 多肽组合或指纹图谱有望满足更高的临床敏感性和特异性的要求<sup>[73-75]</sup>。因此, 无论是什么原因导致了差异多肽的产生, LMWPs在生物标志物的挖掘方面仍具有无限的潜力。

### 3 总结与展望

随着样本处理技术、质谱技术以及计算机技术的高速发展, 大量关于血浆 LMWPs 的质谱数据已产生, 如何从多肽数量、含量和序列组成等信息中寻找与临床应用、生理病理现象相关的分子, 需要生物信息学等多种技术的支持。其次, LMWPs 的研究策略仍需进一步优化。事实上, 受限于疾病的异质性、研究方法的再现性以及研究队列的有限性与延伸性, 研究发现的差异多肽并没有得到很好的队列或功能验证, 因此, 研究策略的转变与升级将是促进 LMWPs 研究进展的重要环节。实践证明, 多中心、多样本、多疾病的研究策略将有助于挖掘临床有效的差异分子。最后, 探究差异多肽的来源、功能作用将有助于疾病病理学的理解, 并推动其向临床转化的速度。未来, 血浆 LMWPs 全面而深入的研究将为多种疾病的早期诊断、病理分型、预后判定和精准治疗带来新的希望。

### 参 考 文 献

- [1] Perez-Silva J G, Espanol Y, Velasco G, *et al.* The degradome database: expanding roles of mammalian proteases in life and disease. *Nucleic Acids Res*, 2016, **44**(D1): D351-D355
- [2] Cai W, Tucholski T M, Gregorich Z R, *et al.* Top-down proteomics: technology advancements and applications to heart diseases. *Expert Rev Proteomics*, 2016, **13**(8): 717-730
- [3] Li Y, Li Y, Chen T, *et al.* Circulating proteolytic products of carboxypeptidase N for early detection of breast cancer. *Clin Chem*, 2014, **60**(1): 233-242
- [4] Taguchi T, Kodera Y, Oba K, *et al.* Suprabasin-derived bioactive peptides identified by plasma peptidomics. *Sci Rep*, 2021, **11**(1): 1047
- [5] Waldmann T A. Cytokines in cancer immunotherapy. *Cold Spring Harb Perspect Biol*, 2018, **10**(12): a028472
- [6] Brittain J E, Parise L V. Cytokines and plasma factors in sickle cell disease. *Curr Opin Hematol*, 2007, **14**(5): 438-443
- [7] Mateen S, Zafar A, Moin S, *et al.* Understanding the role of cytokines in the pathogenesis of rheumatoid arthritis. *Clin Chim Acta*, 2016, **455**: 161-171
- [8] Smith L M, Kelleher N L, Consortium for Top Down Proteomics. Proteoform: a single term describing protein complexity. *Nat Methods*, 2013, **10**(3): 186-187
- [9] Donnelly D P, Rawlins C M, Dehart C J, *et al.* Best practices and benchmarks for intact protein analysis for top-down mass spectrometry. *Nat Methods*, 2019, **16**(7): 587-594
- [10] Lopez-Otin C, Overall C M. Protease degradomics: a new challenge for proteomics. *Nat Rev Mol Cell Biol*, 2002, **3**(7): 509-519
- [11] Mason S D, Joyce J A. Proteolytic networks in cancer. *Trends Cell Biol*, 2011, **21**(4): 228-237
- [12] Quesnel A, Karagiannis G S, Filippou P S. Extracellular proteolysis in glioblastoma progression and therapeutics. *Biochim Biophys Acta Rev Cancer*, 2020, **1874**(2): 188428
- [13] Burbach J P. Neuropeptides from concept to online database www.neuropeptides.nl. *Eur J Pharmacol*, 2010, **626**(1): 27-48
- [14] Hallberg M. Neuropeptides: metabolism to bioactive fragments and the pharmacology of their receptors. *Med Res Rev*, 2015, **35**(3): 464-519
- [15] Fricker L D, Lim J, Pan H, *et al.* Peptidomics: identification and quantification of endogenous peptides in neuroendocrine tissues. *Mass Spectrom Rev*, 2006, **25**(2): 327-344
- [16] Tang H N, Xiao F, Chen Y R, *et al.* Higher serum neuropeptide Y levels are associated with metabolically unhealthy obesity in obese Chinese adults: a cross-sectional study. *Mediators Inflamm*, 2020, **2020**: 7903140
- [17] Vahatalo L H, Ruohonen S T, Makela S, *et al.* Neuropeptide Y in the noradrenergic neurones induces obesity and inhibits sympathetic tone in mice. *Acta Physiol (Oxf)*, 2015, **213**(4): 902-919
- [18] Robas N M, Fidock M D. Identification of orphan G protein-coupled receptor ligands using FLIPR assays. *Methods Mol Biol*, 2005, **306**: 17-26
- [19] Tanaka H, Yoshida T, Miyamoto N, *et al.* Characterization of a family of endogenous neuropeptide ligands for the G protein-coupled receptors GPR7 and GPR8. *Proc Natl Acad Sci USA*, 2003, **100**(10): 6251-6256
- [20] Admon A, Bassani-Sternberg M. The human immunopeptidome project, a suggestion for yet another postgenome next big thing. *Mol Cell Proteomics*, 2011, **10**(10): O111.011833
- [21] Rammensee H G. Peptides made to order. *Immunity*, 2006, **25**(5): 693-695
- [22] Bassani-Sternberg M, Barnea E, Beer I, *et al.* Soluble plasma HLA peptidome as a potential source for cancer biomarkers. *Proc Natl Acad Sci USA*, 2010, **107**(44): 18769-18776
- [23] Shraibman B, Barnea E, Kadosh D M, *et al.* Identification of tumor antigens among the HLA peptidomes of glioblastoma tumors and plasma. *Mol Cell Proteomics*, 2019, **18**(6): 1255-1268
- [24] Mehta A I, Ross S, Lowenthal M S, *et al.* Biomarker amplification by serum carrier protein binding. *Dis Markers*, 2003, **19**(1): 1-10
- [25] Lowenthal M S, Mehta A I, Frogale K, *et al.* Analysis of albumin-associated peptides and proteins from ovarian cancer patients. *Clin Chem*, 2005, **51**(10): 1933-1945
- [26] Kikuchi S, Honda K, Handa Y, *et al.* Serum albumin-associated

- peptides of patients with uterine endometrial cancer. *Cancer Sci*, 2007, **98**(6): 822-829
- [27] Lopez M F, Mikulskis A, Kuzdzal S, *et al.* High-resolution serum proteomic profiling of Alzheimer disease samples reveals disease-specific, carrier-protein-bound mass signatures. *Clin Chem*, 2005, **51**(10): 1946-1954
- [28] Lopez M F, Krastins B, Sarracino D A, *et al.* Proteomic signatures of serum albumin-bound proteins from stroke patients with and without endovascular closure of PFO are significantly different and suggest a novel mechanism for cholesterol efflux. *Clin Proteomics*, 2015, **12**(1): 2
- [29] Lai X, Liangpunsakul S, Crabb D W, *et al.* A proteomic workflow for discovery of serum carrier protein-bound biomarker candidates of alcohol abuse using LC-MS/MS. *Electrophoresis*, 2009, **30**(12): 2207-2214
- [30] Makarewich C A, Olson E N. Mining for micropeptides. *Trends Cell Biol*, 2017, **27**(9): 685-696
- [31] Anderson D M, Anderson K M, Chang C L, *et al.* A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. *Cell*, 2015, **160**(4): 595-606
- [32] Nelson B R, Makarewich C A, Anderson D M, *et al.* A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle. *Science*, 2016, **351**(6270): 271-275
- [33] Huang J Z, Chen M, Chen, *et al.* A peptide encoded by a putative lncRNA HOXB-AS3 suppresses colon cancer growth. *Mol Cell*, 2017, **68**(1): 171-184
- [34] Kondo T, Plaza S, Zanet J, *et al.* Small peptides switch the transcriptional activity of Shavenbaby during *Drosophila* embryogenesis. *Science*, 2010, **329**(5989): 336-339
- [35] D'lima N G, Ma J, Winkler L, *et al.* A human microprotein that interacts with the mRNA decapping complex. *Nat Chem Biol*, 2017, **13**(2): 174-180
- [36] Pauli A, Norris M L, Valen E, *et al.* Toddler: an embryonic signal that promotes cell movement *via* Apelin receptors. *Science*, 2014, **343**(6172): 1248636
- [37] Guo B, Zhai D, Cabezas E, *et al.* Humanin peptide suppresses apoptosis by interfering with Bax activation. *Nature*, 2003, **423**(6938): 456-461
- [38] Hashimoto Y, Niikura T, Tajima H, *et al.* A rescue factor abolishing neuronal cell death by a wide spectrum of familial Alzheimer's disease genes and A $\beta$ . *Proc Natl Acad Sci USA*, 2001, **98**(11): 6336-6341
- [39] Aspden J L, Eyre-Walker Y C, Phillips R J, *et al.* Extensive translation of small open reading frames revealed by Poly-Ribo-Seq. *Elife*, 2014, **3**: e03528
- [40] Hanada K, Akiyama K, Sakurai T, *et al.* sORF finder: a program package to identify small open reading frames with high coding potential. *Bioinformatics*, 2010, **26**(3): 399-400
- [41] Ma J, Diedrich J K, Jungreis I, *et al.* Improved identification and analysis of small open reading frame encoded polypeptides. *Anal Chem*, 2016, **88**(7): 3967-3975
- [42] Zhang Q, Wu E, Tang Y, *et al.* Deeply mining a universe of peptides encoded by long noncoding RNAs. *Mol Cell Proteomics*, 2021, **20**: 100109
- [43] Li N, Zhou Y, Wang J, *et al.* Sequential precipitation and delipidation enables efficient enrichment of low-molecular weight proteins and peptides from human plasma. *J Proteome Res*, 2020, **19**(8): 3340-3351
- [44] Cai T, Zhang Q, Wu B, *et al.* LncRNA-encoded microproteins: a new form of cargo in cell culture-derived and circulating extracellular vesicles. *J Extracell Vesicles*, 2021, **10**(9): e12123
- [45] Mahboob S, Mohamedali A, Ahn S B, *et al.* Is isolation of comprehensive human plasma peptidomes an achievable quest?. *J Proteomics*, 2015, **127**(Pt B): 300-309
- [46] Parker B L, Burchfield J G, Clayton D, *et al.* Multiplexed temporal quantification of the exercise-regulated plasma peptidome. *Mol Cell Proteomics*, 2017, **16**(12): 2055-2068
- [47] Harney D J, Hutchison A T, Su Z, *et al.* Small-protein enrichment assay enables the rapid, unbiased analysis of over 100 low abundance factors from human plasma. *Mol Cell Proteomics*, 2019, **18**(9): 1899-1915
- [48] Cai T, Yang F. Strategies for characterization of low-abundant intact or truncated low-molecular-weight proteins from human plasma. *Enzymes*, 2017, **42**: 105-123
- [49] Klupeczynska A, Swiatly A, Hajduk J, *et al.* Identification of serum peptidome signatures of non-small cell lung cancer. *Int J Mol Sci*, 2016, **17**(4): 410
- [50] Yang J, Xiong X, Liu S, *et al.* Identification of novel serum peptides biomarkers for female breast cancer patients in Western China. *Proteomics*, 2016, **16**(6): 925-934
- [51] Bedin C, Crotti S, Ragazzi E, *et al.* Alterations of the plasma peptidome profiling in colorectal cancer progression. *J Cell Physiol*, 2016, **231**(4): 915-925
- [52] Gajjala P R, Bruck H, Noels H, *et al.* Novel plasma peptide markers involved in the pathology of CKD identified using mass spectrometric approach. *J Mol Med (Berl)*, 2019, **97**(10): 1451-1463
- [53] Zhou C X, Xie S C, Li M Y, *et al.* Analysis of the serum peptidome associated with *Toxoplasma gondii* infection. *J Proteomics*, 2020, **222**: 103805
- [54] Bauza-Martinez J, Aletti F, Pinto B B, *et al.* Proteolysis in septic shock patients: plasma peptidomic patterns are associated with mortality. *Br J Anaesth*, 2018, **121**(5): 1065-1074
- [55] Abe K, Shang J, Shi X, *et al.* A new serum biomarker set to detect mild cognitive impairment and Alzheimer's disease by peptidome technology. *J Alzheimers Dis*, 2020, **73**(1): 217-227
- [56] Yan L, Yi J, Huang C, *et al.* Rapid detection of COVID-19 using MALDI-TOF-based serum peptidome profiling. *Anal Chem*, 2021, **93**(11): 4782-4787
- [57] Toby T K, Fornelli L, Kelleher N L. Progress in top-down proteomics and the analysis of proteoforms. *Annu Rev Anal Chem (Palo Alto Calif)*, 2016, **9**(1): 499-519
- [58] Lin L, Zheng J, Zheng F, *et al.* Advancing serum peptidomic

- profiling by data-independent acquisition for clear-cell renal cell carcinoma detection and biomarker discovery. *J Proteomics*, 2020, **215**: 103671
- [59] Zhang J, Xin L, Shan B, *et al.* PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. *Mol Cell Proteomics*, 2012, **11**(4): M111.010587
- [60] Taylor G K, Kim Y B, Forbes A J, *et al.* Web and database software for identification of intact proteins using "top down" mass spectrometry. *Anal Chem*, 2003, **75**(16): 4081-4086
- [61] Chi H, Liu C, Yang H, *et al.* Comprehensive identification of peptides in tandem mass spectra using an efficient open search engine. *Nat Biotechnol*, 2018, **36**(11): 1059-1061
- [62] Bruderer R, Bernhardt O M, Gandhi T, *et al.* Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. *Mol Cell Proteomics*, 2015, **14**(5): 1400-1410
- [63] Mentlein R, Gallwitz B, Schmidt W E. Dipeptidyl-peptidase IV hydrolyses gastric inhibitory polypeptide, glucagon-like peptide-1 (7-36)amide, peptide histidine methionine and is responsible for their degradation in human serum. *Eur J Biochem*, 1993, **214**(3): 829-835
- [64] Kieffer T J, McIntosh C H, Pederson R A. Degradation of glucose-dependent insulintropic polypeptide and truncated glucagon-like peptide 1 *in vitro* and *in vivo* by dipeptidyl peptidase IV. *Endocrinology*, 1995, **136**(8): 3585-3596
- [65] Drucker D J. Glucagon-like peptides. *Diabetes*, 1998, **47**(2): 159-169
- [66] Deacon C F, Nauck M A, Toft-Nielsen M, *et al.* Both subcutaneously and intravenously administered glucagon-like peptide I are rapidly degraded from the NH<sub>2</sub>-terminus in type II diabetic patients and in healthy subjects. *Diabetes*, 1995, **44**(9): 1126-1131
- [67] Marguet D, Baggio L, Kobayashi T, *et al.* Enhanced insulin secretion and improved glucose tolerance in mice lacking CD26. *Proc Natl Acad Sci USA*, 2000, **97**(12): 6874-6879
- [68] Zhen E Y, Jin Z, Ackermann B L, *et al.* Circulating FGF21 proteolytic processing mediated by fibroblast activation protein. *Biochem J*, 2016, **473**(5): 605-614
- [69] Sierra M D L L, Yang F, Narazaki M, *et al.* Differential processing of stromal-derived factor-1alpha and stromal-derived factor-1beta explains functional diversity. *Blood*, 2004, **103**(7): 2452-2459
- [70] Davis D A, Singer K E, Sierra M D L L, *et al.* Identification of carboxypeptidase N as an enzyme responsible for C-terminal cleavage of stromal cell-derived factor-1alpha in the circulation. *Blood*, 2005, **105**(12): 4561-4568
- [71] Anderson N L. The clinical plasma proteome: a survey of clinical assays for proteins in plasma and serum. *Clin Chem*, 2010, **56**(2): 177-185
- [72] Geyer P E, Holdt L M, Teupser D, *et al.* Revisiting biomarker discovery by plasma proteomics. *Mol Syst Biol*, 2017, **13**(9): 942
- [73] Murphy G. The ADAMs: signalling scissors in the tumour microenvironment. *Nat Rev Cancer*, 2008, **8**(12): 929-941
- [74] Petricoin E F, Ardekani A M, Hitt B A, *et al.* Use of proteomic patterns in serum to identify ovarian cancer. *Lancet*, 2002, **359**(9306): 572-577
- [75] Petricoin E F, 3rd, Ornstein D K, Paweletz C P, *et al.* Serum proteomic patterns for detection of prostate cancer. *J Natl Cancer Inst*, 2002, **94**(20): 1576-1578
- [76] Zhang Z, Bast R C, Jr., Yu Y, *et al.* Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer. *Cancer Res*, 2004, **64**(16): 5882-5890
- [77] Meng Q, Ge S, Yan W, *et al.* Screening for potential serum-based proteomic biomarkers for human type 2 diabetes mellitus using MALDI-TOF MS. *Proteomics Clin Appl*, 2017, **11**(3-4): 1600079
- [78] Agostini M, Bedin C, Enzo M V, *et al.* Multivariate analysis approach to the serum peptide profile of morbidly obese patients. *Dis Markers*, 2013, **34**(4): 269-278
- [79] Nakamura A, Kaneko N, Villemagne V L, *et al.* High performance plasma amyloid-beta biomarkers for Alzheimer's disease. *Nature*, 2018, **554**(7691): 249-254
- [80] Fu Y, Zhou N, Yu Y, *et al.* Profiling of schizophrenia-associated serum peptides by MALDI-TOF-MS. *J Neural Transm (Vienna)*, 2020, **127**(1): 95-101
- [81] Han E C, Lee Y S, Liu Y C, *et al.* Mass spectrometry-based plasma peptide profiling of acute exacerbation in HBeAg-positive chronic hepatitis B. *Clin Chim Acta*, 2011, **412**(23-24): 2174-2182
- [82] Rutkowski M J, Sughrue M E, Kane A J, *et al.* Cancer and the complement cascade. *Mol Cancer Res*, 2010, **8**(11): 1453-1465

## Advances in Plasma Low Molecular Weight Proteins and Peptides\*

LI Na<sup>1,2)</sup>, CAI Tan-Xi<sup>1,2)</sup>, YANG Fu-Quan<sup>1,2)\*\*</sup>

<sup>(1)</sup>Key Laboratory of Protein and Peptide Pharmaceuticals & Laboratory of Proteomics, Institute of Biophysics,  
Chinese Academy of Sciences, Beijing 100101, China;

<sup>2)</sup>University of Chinese Academy of Sciences, Beijing 100049, China)

**Abstract** Plasma is the most frequently clinically used body fluid sample, characterized by noninvasive sampling, convenient to use, and rich contents, and is an important source sample for the excavation of potential biomarkers related to the disease. Low molecular weight proteins and peptides (LMWPs) in plasma include cytokines, growth factors, peptide hormones, and protein degradation products, which are favored by researchers because of their small molecular weight, measurable sequence information, and wide sources. In recent years, the studies on LMWPs have made new progress with the rapid development of high-resolution, high-sensitivity mass spectrometers and computational science, and the integration of new technologies and new methods. This review introduces the classification of plasma LMWPs, including small intact proteins, protein degradation products, neuropeptides, immunopeptides, carrier protein-bound peptides, and sORF-encoded peptides, based on the current research hotspots in this field. In addition, it also summarizes the research progresses on enrichment strategies, functional studies, and disease-related biomarkers. Finally, the challenges and future research direction of plasma LMWPs are discussed.

**Key words** plasma low molecular weight proteins and peptides, enrichment methods, mass spectrometry, biomarker, sORF-encoded peptides

**DOI:** 10.16476/j.pibb.2021.0241

---

\* This work was supported by a grant from the National Key R&D Program of China (2018YFA0507801).

\*\* Corresponding author.

Tel: 86-10-64888581, E-mail: fqyang@ibp.ac.cn

Received: July 27, 2021 Accepted: October 13, 2021