CLE家族植物小肽激素的质谱碎裂特征研究

袁薪茹[1,2],闫吉军[2],范金石[1],褚金芳[2,3]
(1.青岛科技大学海洋科学与生物工程学院,山东 青岛 266042;
2.中国科学院遗传与发育生物学研究所,国家植物基因研究中心(北京),北京 100101;
3.中国科学院大学现代农业科学学院,北京 100039)

摘要:CLE(CLAVATA3/Embryo surrounding region)小肽是植物小肽激素中最大的家族,在调控植物分生组织干细胞的分化和分裂中发挥着关键性作用。本工作采用超高效液相色谱-四极杆飞行时间质谱法(UPLC-QTOF MS)研究5种典型CLE小肽的质谱特征,并总结质谱碎裂行为的规律,即CLE小肽的最强准分子离子峰带有2~3个电荷,碰撞诱导解离(CID)主要产生b型和y型离子;在15~20 V最佳碰撞电压下,最容易从脯/羟脯氨酸的氨基、甘氨酸的氨基和羧基端发生断裂形成高丰度的b_3和y_4等碎片离子。对合成的CLE类似肽(CLEL)的质谱碎裂行为进行预测和实验验证,结果表明,CLEL的质谱碎裂行为与预期完全相符,表明该碎裂规律对CLE家族的小肽具有较强的适用性,可为CLE小肽激素的定性和定量分析提供重要的数据参考。CLE家族结构具有高度保守性,通过对拟南芥中的CLE十二肽基序的质谱碎裂方式和碎片离子进行预测,可根据这些数据信息构建非数据依赖的CLE小肽筛选和鉴定方法,有助于生物学家探寻新的CLE小肽激素。

关键词:CLE小肽;植物小肽激素;液相色谱-质谱(LC-MS);碎裂规律
中图分类号:O657.63 **文献标志码**:A **文章编号**:1004-2997(2022)01-0056-11
doi:10.7538/zpxb.2021.0023

Mass Spectrometric Characteristics for CLE Family of Plant Peptide Hormones

YUAN Xin-ru[1,2], YAN Ji-jun[2], FAN Jin-shi[1], CHU Jin-fang[2,3]
(1. *College of Marine Science and Biological Engineering, Qingdao University of Science and Technology, Qingdao* 266042, *China*;
2. *National Centre for Plant Gene Research (Beijing), Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing* 100101, *China*;
3. *College of Advanced Agricultural Sciences, University of Chinese Academy of Sciences, Beijing* 100039, *China*)

Abstract: CLAVATA3/Embryo surrounding region (CLE) peptides, the largest family of plant peptide hormones, is critical in regulating cell differentiation and division in meristematic tissues in plants. In this study, ultra-high performance liquid chromatography-quadrupole time-of-flight mass spectrometry (UPLC-QTOF MS) equipped with an electrospray ionization source (ESI) was used to investigate the fragmentation mech-

国家自然科学基金"青年基金"项目(31800303);中国科学院战略性先导科技专项(A类)(XDA24040202)
本文通信作者范金石,褚金芳

anism of five typical CLE peptides. The results showed that the strongest parent ion of CLE peptides was two or three fold protonated ion. Under the optimal collision energy (15-20 V), the collision-induced dissociation (CID) of CLE peptides mainly generate b-type and y-type fragment ions due to the cleavage of polypeptide bonds. Based on the MS/MS spectra of peptides standards, it was concluded that highly abundant product ions such as b_3 and y_4 ion were easily formed by the cleavage at the Xxx-Pro (Xxx denotes any other amino acid), Xxx-Hyp, Xxx-Gly or Gly-Xxx amide bond. In order to verify above fragment mechanism, a synthesized CLE-like peptide (CLEL) was tested and the mass spectrometry fragmentation behavior of experiment was completely in line with that of predictions. It could be seen that the fragmentation mechanism mentioned above had strong applicability to CLE family peptides. According to the structure-activity relationship revealed by plant physiologist, the CLE residues R1/H1, P4, P7, P9, G6 and H12/N12 are critical for their function. Combined with the cleavage patterns of CLE peptides and the conserved amino acid residues mentioned above, y_1, y_4, b_5, b_6, y_6 and y_7 ion are valuable for qualitative analysis. So the mass spectrum fragmentation mode and fragment ions (y_1, b_3, y_4, and y_7 ion) of the CLE dodecapeptide motif in Arabidopsis were predicted, which would provide the necessary scientific basis for the quantitative and qualitative analysis of CLE peptides. Considering the highly conservative structure of the CLE family structure, a data-independent CLE peptides screening and identification strategy based on these mass spectrometry fragmentation behaviors was put forward. Herein, CLEL was used as an unknown peptide to initially verify the feasibility of above screening strategy. The deduced structure by MassLynx software was almost identical with real peptide sequence of CLEL but minor difference in order of the last four residues, which could be easily corrected based on the conserve structure of CLE peptide family. This work will be very helpful for biologists to search those hidden predicted CLE peptides in plant kingdom. Furthermore, the high-abundance product ions of CLE peptides, b_3, y_4 and b_6 ion, are valuable for quantitative analysis.

Key words: CLAVATA3/Embryo surrounding region (CLE) peptides; plant peptide hormones; liquid chromatography-mass spectrometry (LC-MS); fragmentation mechanism

植物小肽激素是植物激素研究领域的一个新兴分支,是一类天然氨基酸以不同组成和排列方式构成的活性生物分子,通常可以与周边细胞膜表面的受体激酶(主要是富含亮氨酸的类受体蛋白激酶)相互作用,在植物细胞间信息交流中发挥着关键作用[1-2]。随着技术手段的快速发展和研究的不断深入,人们陆续在番茄及其他植物中发现了与植物免疫、细胞分化、生殖与发育等植物生理活动密切相关的系统素、植物磺肽素、CLE (CLAVATA3/Embryo surrounding region)和快速碱化因子等数十种小肽激素[1-2],其中CLE作为维持干细胞增殖与分化平衡的关键调控因子[3],在低等植物和高等植物(包括单子叶植物和双子叶植物)中均有分布。CLE小肽结构具有一定的保守性,其保守序列由14个氨基酸组成,其中12个氨基酸序列(十二肽基序)足以使CLE发挥其生理作用[4]。根据生物信息学预测,在拟南芥中有32个CLE成员[5],在水稻中有47个CLE成员[6],其他物种中也含有类CLE同源基因[7]。在CLE家族中,目前研究较为透彻的只有CLV3、导管分化抑制因子(tracheary elements differentiation inhibitory factor, TDIF)等少数成员,还有大量的预测小肽尚未得到证实。植

物体内潜在的 CLE 小肽及其类似物的数量十分庞大，因此 CLE 小肽的筛选与结构鉴定极具挑战性。质谱技术具有准确性高、灵敏度高、选择性强等优点，在 CLE 小肽的发现、结构鉴定和定量分析中发挥着关键作用[8-9]，小肽的质谱碎裂规律和碎片信息是小肽定性、定量分析的重要依据。

本工作拟采用超高效液相色谱-四极杆飞行时间质谱法(UPLC-QTOF MS)系统性地研究 CLV3、CLE19、CLE40、TDIF 和 GmRIC1 (Glycine max Rhizobia-induced CLE1)5 种典型的 CLE 小肽激素的质谱特征，并利用合成的 1 个 CLE 类似小肽(CLEL)对碎裂规律进行验证。通过研究 CLE 小肽的碎裂规律和特点，提出利用质谱手段对未知 CLE 小肽进行筛选与结构鉴定的可能性思路，希望可以帮助生物学家准确定量分析 CLE 小肽以及寻找和鉴定新的 CLE 小肽。

1 实验部分

1.1 主要仪器与试剂

Acquity™ UPLC-Synapt G2 QTOF-MS 联用仪：美国 Waters 公司产品，配有电喷雾离子源(ESI)及 MassLynx4.1 数据处理系统。

CLV3、CLE19、CLE40、TDIF、GmRIC1：纯度＞95%，均由杭州中肽生化有限公司合成；CLEL：由中国科学院遗传与发育生物学研究所许操研究员实验室提供，各小肽标品的氨基酸序列等信息列于表 1；乙腈、甲醇：色谱纯，美国 Thermo Fisher 公司产品；甲酸：色谱纯，美国 Sigma-Aldrich 公司产品；其他化学试剂均为分析纯或更高纯度级别；去离子水(电阻率≥18.2 MΩ)：由 Milli-Q 纯化水净化系统制备。

采用去离子水配制小肽标准储备液(1 g/L)，放于－80 ℃冰箱中保存；用含 0.1%甲酸的 5%乙腈酸性水溶液逐级稀释储备液，配制工作溶液。

1.2 实验条件

1.2.1 色谱条件 色谱柱：Waters Peptide CSH C18 柱(2.1 mm×100 mm×1.7 μm)；流动相：A 为 0.1%甲酸-水溶液，B 为 0.1%甲酸-乙腈溶液；梯度洗脱程序：0～1 min(99%A)，1～10 min(99%～43%A)，10～13 min(43%～0%A)，13～14 min(0%A)，14～16 min(0%～99%A)；流速 0.3 mL/min；柱温为室温；进样量 5 μL。

表 1 CLE 小肽及 CLEL 的氨基酸序列和一级质谱特征准分子离子

Table 1 Amino acid sequences and typical quasi-molecular ions of five CLE peptides and CLEL

小肽 Peptide	氨基酸序列 Amino acid sequence	准分子离子 Quasi-molecular ion (m/z)
CLV3	RTVPSGPDPLHH	656.8420$[M+2H]^{2+}$
		438.2306$[M+3H]^{3+}$
		328.9249$[M+4H]^{4+}$
CLE19	RVIPTGPNPLHN	657.8680$[M+2H]^{2+}$
		438.9146$[M+3H]^{3+}$
CLE40	RQVPTGSDPLHH	672.3449$[M+2H]^{2+}$
		448.5659$[M+3H]^{3+}$
		336.6764$[M+4H]^{4+}$
TDIF	HEVOSGONPISN	**640.2998$[M+2H]^{2+}$**
		427.2025$[M+3H]^{3+}$
GmRIC1	RLAPEGPDPHHN	670.3293$[M+2H]^{2+}$
		447.2221$[M+3H]^{3+}$
		335.6685$[M+4H]^{4+}$
CLEL	RGVPAGODPLHH	634.8289$[M+2H]^{2+}$
		423.5552$[M+3H]^{3+}$
		317.9184$[M+4H]^{4+}$

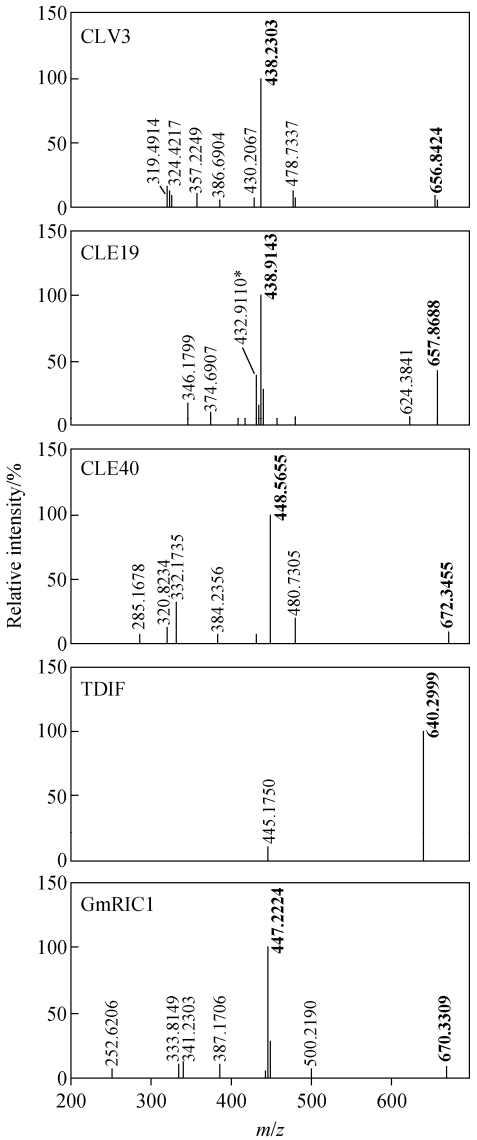
注：1) O 表示羟脯氨酸；2) 加粗表示响应最高的准分子离子

1.2.2 质谱条件 电喷雾电离源正离子模式(ESI^+，m/z 50～1 500)，毛细管电压3.0 kV，样品锥孔电压 30 V，离子源温度 100 ℃，脱溶剂气温度 350 ℃，样品载气流量 30 L/h，脱溶剂气流量 800 L/h，采用亮氨酸脑啡肽为质量校准溶液(浓度 1 mg/L，流速 5 μL/min)。二级质谱实验采用氩气作为碰撞气，碰撞诱导解离(CID)模式，除了设定碰撞电压外，其他实验条件与一级质谱参数一致。

2 结果与讨论

2.1 CLE 家族小肽激素的一级质谱图分析

选取的 5 种典型 CLE 小肽激素均可检测到多种多电荷离子，示于图 1，TDIF 的最强准分子离子为双电荷，而其他4种小肽的最强准



注:加粗表示准分子离子;* 表示脱水离子

图 1　5 种 CLE 小肽的一级质谱图

Fig. 1　MS spectra of five CLE peptides

分子离子均为三电荷。结合小肽序列结构可知(表 1),上述 5 种典型 CLE 小肽的最强准分子离子携带电荷数与其碱性氨基酸残基的数目呈正相关,小肽的质子化位点可能是空间上远离的碱性氨基酸侧链及端氨基,这一规律可能同样适用于其他 CLE 小肽。

从检测灵敏度的角度出发,最强准分子离子占所有准分子离子的比例及其相关 CID 碎片离子是小肽定量分析的关键。除 CLE19 外,其余小肽的最强准分子离子峰比例大部分达到 90%以上,这有利于谱图解析和定量分析,示于图 2a。谱图分析过程中,除 TDIF 外,其余小肽均发生了脱水现象,脱水离子峰的产生可能主要来源于氨基酸侧链(O、S、T、Y)上的羟基。脱水离子的比例随着电荷数目的增加而增加,CLE40 和 CLV3 的四电荷准分子离子甚至接近 100%,中性丢失形成了$[M+4H-H_2O]^{4+}$,示于图 2b。随着样品锥孔电压的增加,脱水离子所占比例越来越大(结果未展示),可见一级质谱的脱水离子是源内碎裂的结果。因此,在对 CLE 家族进行分析尤其是定量分析时,需要特别关注含有存在羟基的氨基酸残基的小肽,防止源内碎裂给分析灵敏度带来不利的影响。另一方面,该结果表明不同电荷的小肽离子稳定性具有较大差异,电荷数越多越容易碎裂。不同电荷的小肽准分子离子峰的碎裂特征将在下文继续探讨。

小肽多电荷离子的分布情况是影响检测灵敏度及谱图解析难度的重要因素。除小肽自身的结构特征外,流动相组成及工作溶剂的选择

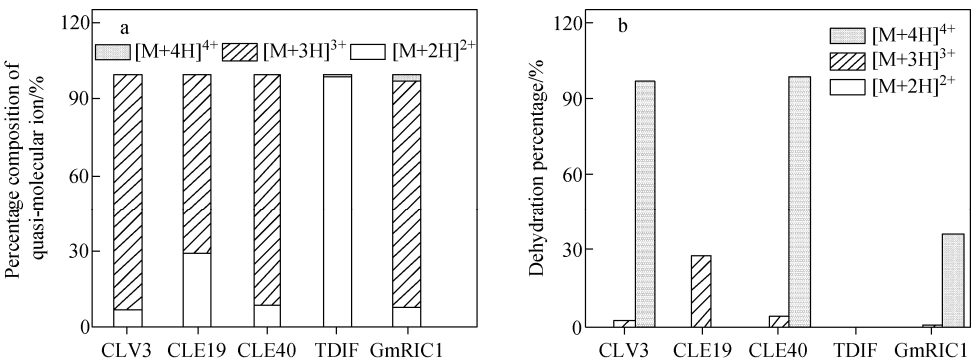


图 2　携带电荷数目对 CLE 小肽准分子离子强度(a)及脱水效率(b)的影响

Fig. 2　Influence of the charge state of CLE peptides parent ions on corresponding ion abundance (a) and dehydration efficiency (b)

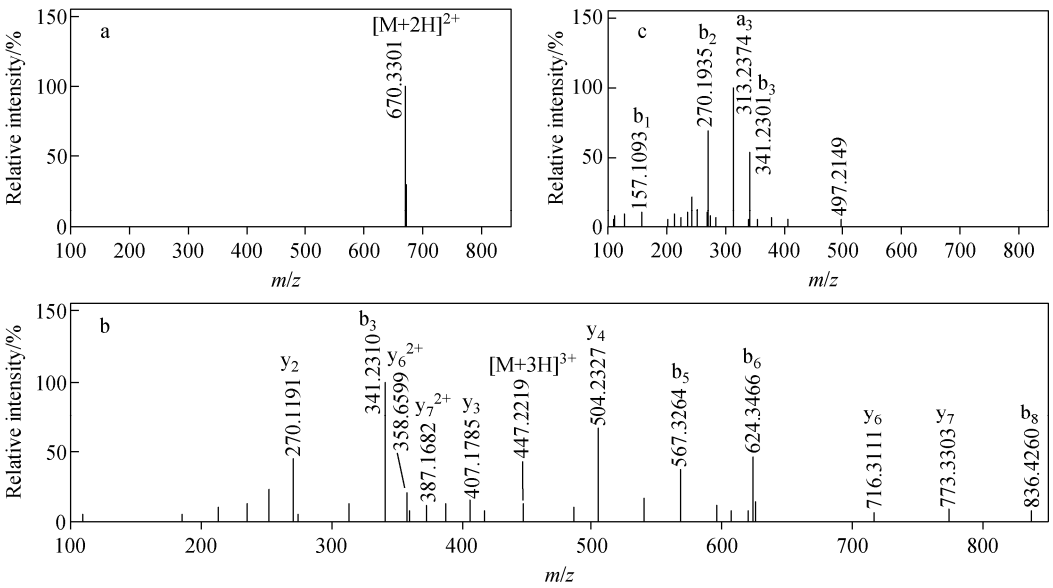
也非常关键。Kebarle 等[10]研究发现，小肽离子化后的电荷状态与工作溶剂密切相关。气相状态下具有高质子亲和势的溶剂（乙腈、丙酮等）会使小肽的电荷数减少，而具有低质子亲和势的溶剂（如甲醇和水）对小肽离子的电荷状态没有影响。为此，本研究将流动相 B 由乙腈换为甲醇，以期减少低电荷离子丰度，提升最强准分子离子峰的比例。然而，实验结果表明，甲醇体系下各小肽最强准分子离子峰的比例变化不显著，而且出现色谱峰变宽、峰分叉及灵敏度下降等问题。故后续实验仍然采用乙腈-水溶液作为流动相。

2.2 CLE 家族小肽激素的二级质谱图分析

在相同的质谱条件下，GmRIC1 的双电荷、三电荷和四电荷准分子离子的二级质谱图示于图 3。可以看出，在 20 V 碰撞电压下，$[M+2H]^{2+}$ 基本未碎裂，$[M+3H]^{3+}$ 大部分被裂解，而 $[M+4H]^{4+}$ 几乎完全碎裂成子离子，特征碎片离子峰较少。可见，小肽准分子离子所带电荷越多，越容易发生碎裂，此结果与一级质谱研究中的脱水程度与所带电荷数的关系相符，也与文献[11]报道一致。进一步分析表明，最强准分子离子 $[M+3H]^{3+}$ 产生的碎片离子最丰富且强度较高，更适合用于小肽结构鉴定。另一方面，小肽的定量分析时，为了获得最佳的分析效果，一般选择最强的准分子离子为母离子，选择丰度高的特征碎片离子为子离子。因此，本实验选择每个 CLE 小肽最强的准分子离子探索 CID 碎裂规律。

在 CID 碎裂模式下，碰撞能量是影响二级质谱图质量的主要因素。在低碰撞电压（10 V/15 V）下，小肽 CLV3 无法有效裂解，可获取的子离子信息较少；当碰撞电压升高（18 V/20 V），小肽产生的碎片离子峰较多，强度高且稳定，适合对其进行定性定量分析；继续升高碰撞电压至 25 V 时，特征碎片离子峰减少且强度降低，示于图 4。综合对比，CLV3、CLE19、CLE40、TDIF 以及 GmRIC1 的最佳碰撞能量分别为 18、15、18、20、18 V。双电荷 TDIF 需要的碰撞能量比三电荷的其他小肽高，这进一步表明，小肽准分子离子碎裂程度与其所带电荷的数目正相关。鉴于 CLE 家族小肽的高度结构保守性，其他 CLE 小肽的最佳碰撞能量可能也在 15～20 V 之间，实际应用时需要考虑最强准分子离子的电荷数，电荷数越多，能量越低。

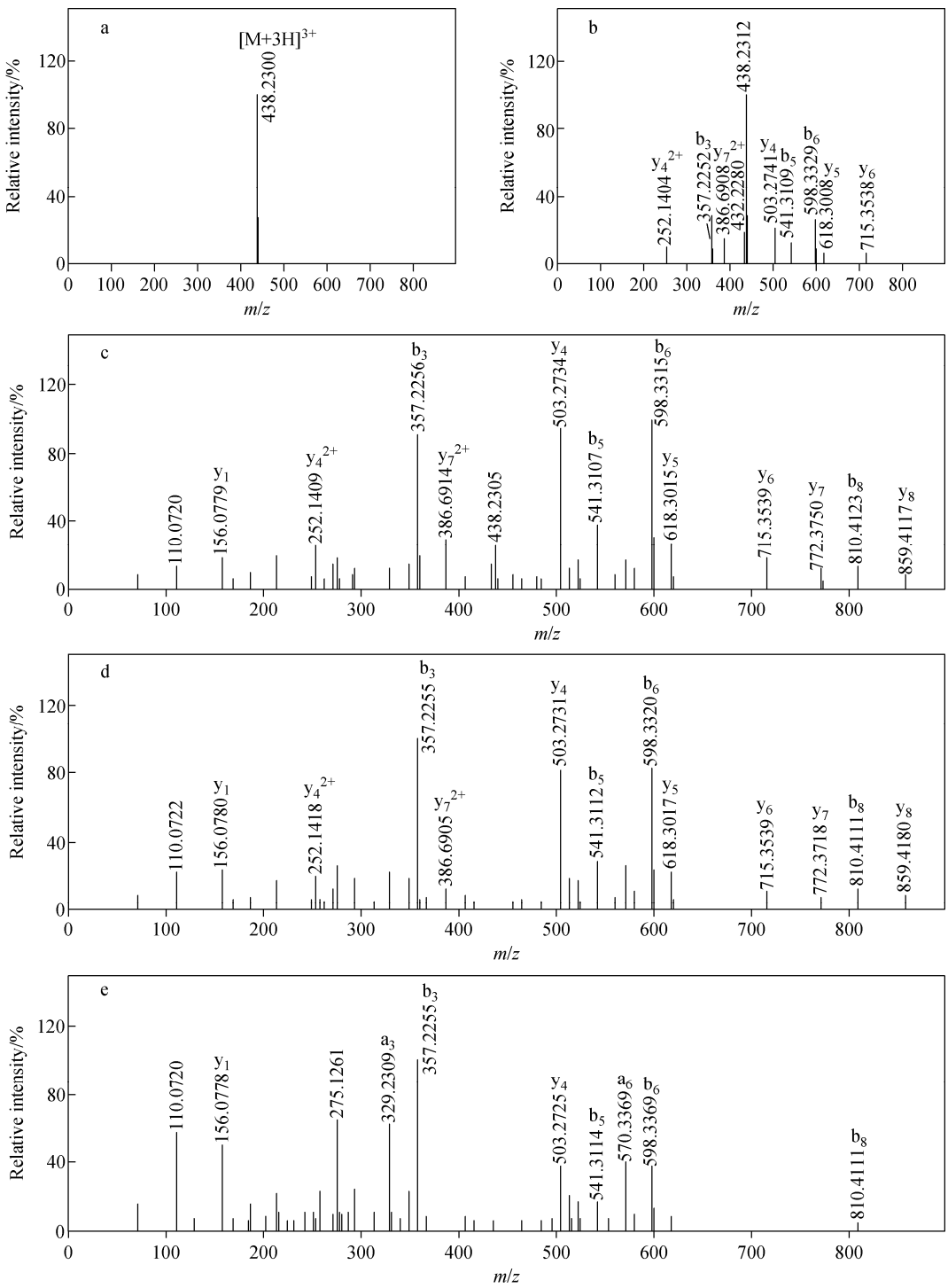
在最佳碰撞能量下，CLE 家族 5 种小肽激素的二级质谱图示于图 5。可以看出，在 CID 模式下，CLE 小肽主要产生 b 型和 y 型碎片离子，还有少量的 a 型离子。CLV3 的碎片离子



注：a. $[M+2H]^{2+}$；b. $[M+3H]^{3+}$；c. $[M+4H]^{4+}$

图 3 在 20 V 碰撞电压下，GmRIC1 不同电荷准分子离子的二级质谱图

Fig. 3 MS/MS spectra of different parent ions of GmRIC1 at 20 V collision voltage

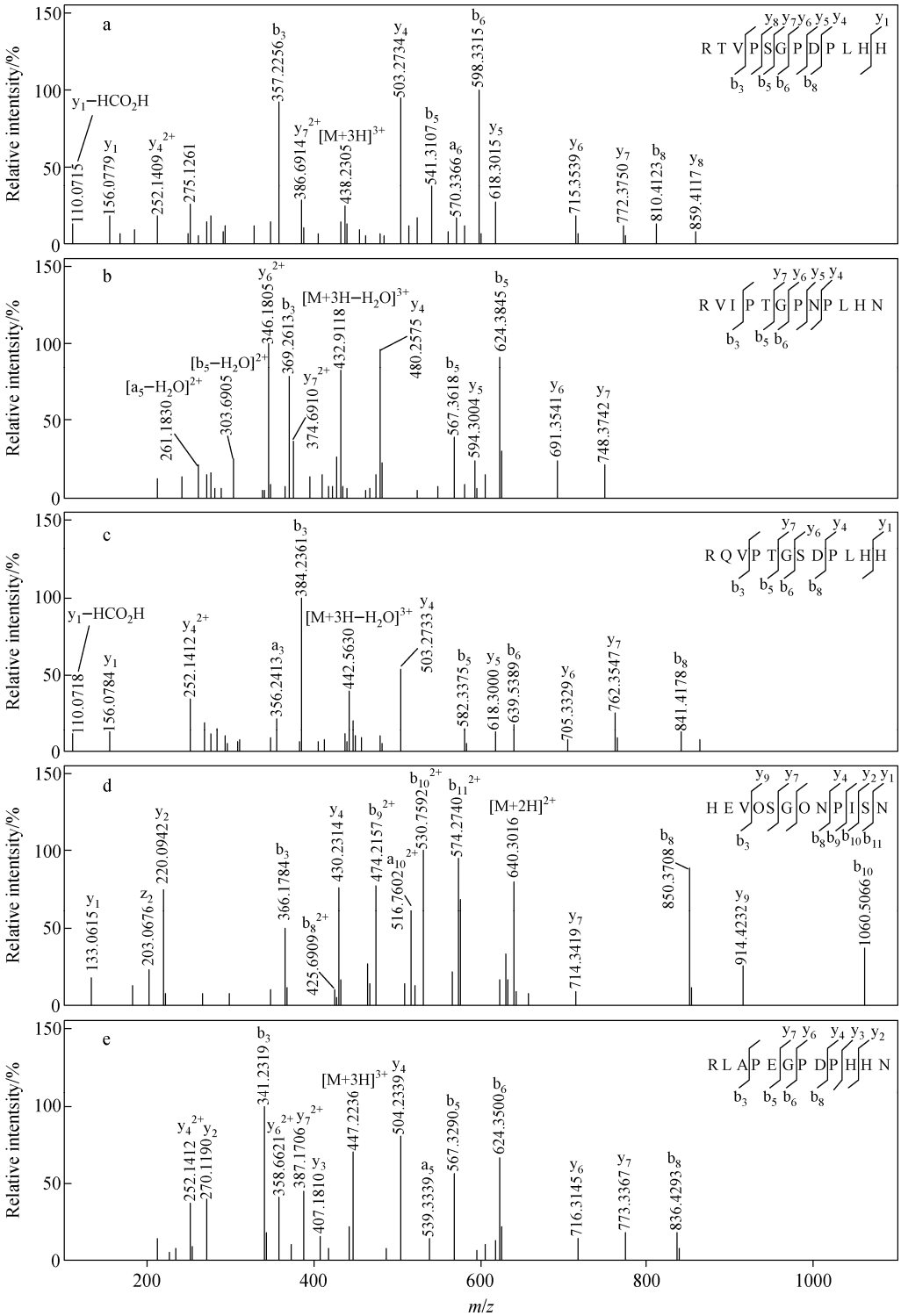


注:a. 10 V;b. 15 V;c. 18 V;d. 20 V;e. 25 V

图 4　在 10～25 V 碰撞电压下,CLV3 准分子离子[M+3H]$^{3+}$的二级质谱图

Fig. 4　MS/MS spectra of [M+3H]$^{3+}$ of CLV3 at 10-25 V collision voltage

中,b_3、b_6、y_4 的相对丰度最高,b_5、y_6、y_7 次之,表明 3 位缬氨酸和 4 位脯氨酸、6 位甘氨酸和 7 位脯氨酸以及 8 位天冬氨酸和 9 位脯氨酸之间的肽键最容易断裂,5 位丝氨酸和 6 位甘氨酸之间的酰胺键次之。CLE19 的碎片离子中,b_3、y_6 的相对强度最高,b_5、y_4、y_7 次之,表明 3 位异亮氨酸和 4 位脯氨酸、6 位甘氨酸和 7 位脯氨酸的肽键最容易断裂,5 位苏氨酸和 6 位



注：a. CLV3；b. CLE19；c. CLE40；d. TDIF；e. GmRIC1

图 5 在最优碰撞电压下，5 种 CLE 小肽的特征碎片离子图

Fig. 5 Characteristic fragment ions of five CLE peptides under the optimized collision energy

甘氨酸以及 8 位天冬酰胺和 9 位脯氨酸之间的肽键次之。CLE40 的碎片离子中，b_3、y_4 的相对强度最高，b_5、b_6、y_7 次之，说明 3 位缬氨酸和 4 位脯氨酸、8 位天冬氨酸和 9 位脯氨酸之间的肽键最容易断裂，5 位苏氨酸和 6 位甘氨酸以及 6 位甘氨酸和 7 位苏氨酸之间的肽键次之。

GmRIC1的碎片离子中相对强度最高的为b_3、b_5、b_6、y_4,说明酰胺键的断裂容易发生在3位丙氨酸和4位脯氨酸、5位谷氨酸和6位甘氨酸、6位甘氨酸和7位脯氨酸以及8位天冬氨酸和9位脯氨酸。TDIF的碎片离子中,b_8、b_9、b_{10}、b_{11}、y_2的相对强度最高,b_3、y_4次之,说明末端5个氨基酸之间的肽键均容易断裂,3位缬氨酸与4位羟脯氨酸之间的肽键次之。综合对比分析,CLE小肽N端的精氨酸或组氨酸的羧基位置很难碎裂,均未检测到b_1离子,而y_1、y_4、y_7和b_3离子则普遍存在于各CLE标品的二级质谱图中,尤其以y_4和b_3离子丰度较高。在CLE小肽结构中,脯/羟脯氨酸的氨基(特别是氨端起4位和9位脯/羟脯氨酸)以及甘氨酸的氨基和羧基端最容易发生断裂。根据自由移动质子引发的电荷诱导碎裂模型,b/y离子形成的主要驱动力是电荷在肽段上的转移。在外界能量激发下,电荷向热力学更稳定的位置移动,当电荷移动到肽键上的氮原子时,酰胺键的电子云密度分布发生改变,从而使肽键键能减弱而更容易碎裂[12]。因此,肽键碎裂难易的关键因素是肽键中氮原子的质子亲和势。脯氨酸作为唯一的环状氨基酸,其氮原子的pKa(10.6)大于其他氨基酸,更容易结合质子。在这种情况下,小肽链中脯氨酸的氨基侧肽键比其他位置更容易发生断裂,从而产生相应的高丰度b/y离子。文献[11]报道表明,烟草系统素也容易在脯氨酸的氨基端发生断裂。可见,脯氨酸氨基侧肽键容易碎裂形成相应的y离子及互补b离子这一规律可能具有普适性,可用于其他活性小肽的定性定量分析。目标小肽中都存在相对强度较高的b_5、y_7以及b_6、y_6互补离子对,对于CLE家族其他小肽激素的定量分析具有重要的参考意义。

CLE家族的小肽激素具有特定的保守结构域,这些保守结构提供了受体结合和信号转导的关键作用位点,是该家族成员间具有显著功能冗余性的重要原因。6位甘氨酸在中心区域扭折结构的形成中发挥着关键作用。4、7、9位脯氨酸构成了典型的聚脯氨酸结构域,通常形成刚性的螺旋结构,能够用作蛋白-蛋白作用域或组装多肽链。1位精氨酸或组氨酸对前体蛋白剪切定位及成熟小肽的形成十分重要,12位组氨酸或天冬酰胺则普遍存在于拟南芥、水稻等物种的CLE结构中。这些保守结构的存在与否可以作为CLE小肽激素筛选的重要判断依据。结合前述质谱碎裂特点,CLE小肽激素定性分析的关键碎片离子为y_1(12位组氨酸或天冬酰胺)、b_1(1位精氨酸或组氨酸)、b_4(包含4位脯氨酸)、y_3(包含9位脯氨酸)。在现有条件下,b_1离子难以检出,而b_4和y_3离子的质荷比可能性太多(包含可变氨基酸),使用价值有限。综合判断,y_1离子是CLE分析的关键离子之一。6位甘氨酸位于肽链中间,无法直接通过b和y型离子判断其存在与否。而甘氨酸两端的肽键容易碎裂,可以通过中性丢失分析,间接推测6位甘氨酸的存在(b_7-b_6或y_8-y_7)。同样的,脯氨酸氨基侧的肽链容易碎裂,也可通过类似的思路间接解析。因此,上述保守离子可作为小肽筛选的依据,构建合适的质谱筛选方法(数据依赖或非数据依赖),帮助研究人员探寻新的CLE小肽激素。

2.3　CLE家族小肽质谱碎裂规律验证

为了检验上述总结的规律,实验合成了1个与CLE结构十分相似的小肽CLEL,其基本信息列于表1。根据上文研究,CLE小肽的质谱特征为:1)最强准分子离子峰携带的电荷数目为空间分散的碱性氨基酸个数加1;2)最强准分子离子峰的最佳碰撞能量在15～20 V之间,双电荷需要20 V左右,三电荷需要18 V左右,四电荷的最佳碰撞能量更低;3)在CID模式下,主要产生y_1、y_4、y_7和b_3离子,其中b_3和y_4为高丰度离子。据此,推测CLEL的最强准分子离子为三电荷离子,最佳碰撞能量为15～20 V,容易在脯氨酸氨基侧及甘氨酸两侧断裂形成b_3、y_9、y_4、y_8、b_5、y_7、b_6和y_6离子。CLEL的最强准分子离子为$[M+3H]^{3+}$,最佳碰撞电压为18 V,碎片离子中b_3、b_5、b_6、y_4、y_6的丰度最高,示于图6,该结果与推测结果完全契合。上述质谱碎裂规律对CLE小肽的相关研究具有较强的实用性。

2.4　CLE家族小肽质谱碎裂规律的应用

在拟南芥的CLE十二肽基序中[13],N端为组氨酸或精氨酸,中间大多含有PxxPxP(P表示脯氨酸,x表示可变氨基酸残基)片段,

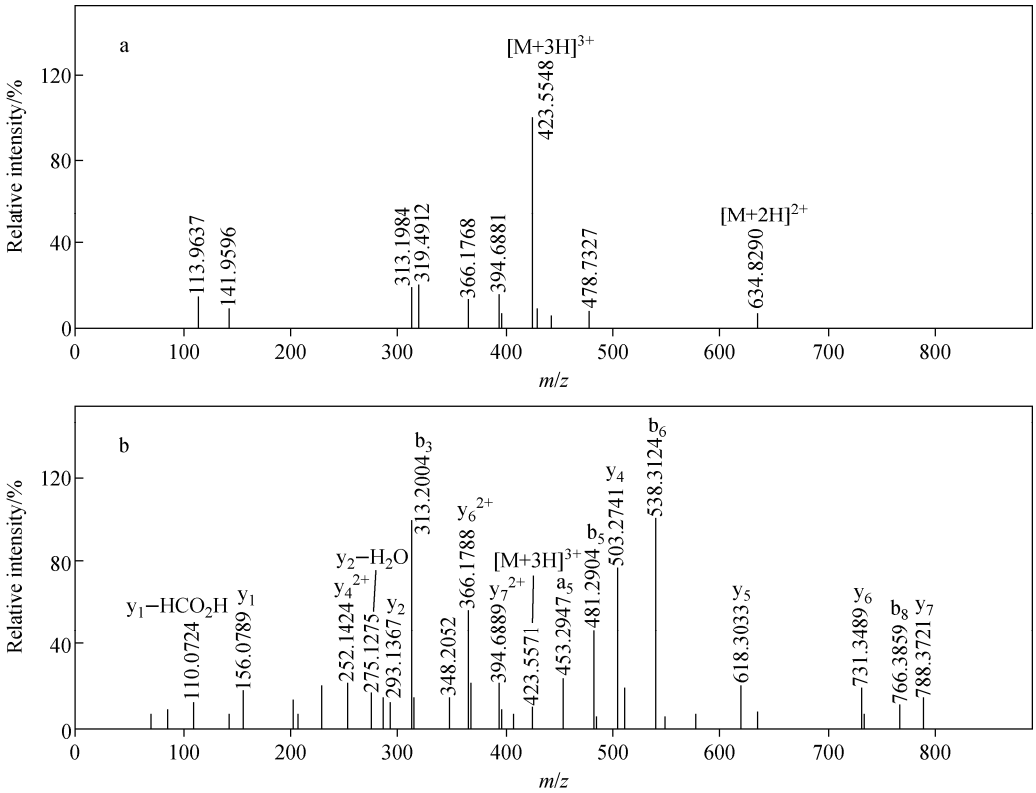


图 6 在最优碰撞电压(18 V)下,CLEL 的一级质谱图(a)和准分子离子[M+3H]$^{3+}$的二级质谱图(b)

Fig. 6 MS spectrum of [M+3H]$^{3+}$ (a) and MS/MS spectrum of CLEL (b) under the optimized collision energy (18 V)

C 端为天冬酰胺或组氨酸,6 号位置为甘氨酸。根据出现频次,最普遍的保守序列为 R-x-x-P-x-G-P-D/N-P-L-H-H/N。可见,对未知 CLE 小肽筛选与鉴定最有价值的是保守性极强的 b_1、y_1、y_2、y_3、y_4 离子。结合 CLE 小肽的质谱碎裂特点和碎片离子的丰度分布,适合未知小肽筛选标准的子离子为 y_1 和 y_4。基于上述依据,可对拟南芥 26 个十二肽基序以及其他物种中未知 CLE 小肽的质谱行为进行预测,结果列于表 2。可以看出,由于结构的保守性,CLE 小肽的子离子分布具有很大的相似性,因此可以提出 1 种基于质谱的未知 CLE 小肽的筛选策略。以 Synapt G2 QTOF-MS 的 MS^E(一种非数据依赖采集模式)为例,同时以低能量和高能量进行质谱扫描,典型条件如下:1) 低能量扫描,正离子扫描模式(ESI$^+$,m/z 100~800),碰撞能量 0 V,可采集一级质谱图;2) 高能量扫描,正离子扫描模式(ESI$^+$,m/z 100~1 500),碰撞能量 18 V,可采集二级质谱图。数据采集后,在高能量总离子流图(TIC)中采用表中预测的 y_1 (133.061/156.077) 和 y_4(406.220/383.204/421.195/303.167/333.177/449.237)离子提取色谱图(EIC),筛选出保留时间相同的谱峰,即同时具有 y_1 和 y_4 子离子的化合物,然后运用 MassLynx 软件的谱图合并(combine spectrum)功能分别提取这些化合物的一级和二级质谱图,分析母离子和子离子分布,推测小肽序列。为了考察这一策略的可行性,本文假定 CLEL 是一种未知 CLE 小肽,利用 MS^E 方法进行搜寻和序列鉴定,最终软件预测该化合物的小肽序列为 RGVPAGODHPLH。CLEL 的真实序列为 RGVPAGODPLHH,预测的氨基酸组成与实际氨基酸组成基本相符,仅 C 端 4 个氨基酸的排列顺序有差异。可见,该筛选策略对未知 CLE 小肽的筛选和结构鉴定具有潜在的应用价值。

由于 CLE 家族的小肽激素内源含量低、数量庞大,而且具有一定的物种差异性,其筛选和鉴定是一项极富挑战性的工作。本文提供的筛选策略可用于CLE家族小肽的初步筛选,后期

表2　拟南芥中已知CLE十二肽基序的质谱行为预测

Table 2　Predicted mass spectrometric characteristics of CLE dodecapeptide motif in Arabidopsis

小肽 Peptide	序列 Sequence	准分子离子 Quasi-molecular ion (m/z)[a]	电荷数 Charge number	典型碎片离子质荷比 Typical fragment ion mass (m/z)			
				y_1	y_4	y_7	b_3
CLE13	RLVPSGPNPLHH	441.915	3	156.077	406.220	714.369	369.261
CLE11	RVVPSGPNPLHH	437.243	3	156.077	406.220	714.369	355.246
MCLV3	RTVPSGPDPLHH	438.231	3	156.077	406.220	715.353	357.225
CLE8	RRVPTGPNPLHH	460.926	3	156.077	406.220	714.369	412.278
CLE12	RRVPSGPNPLHH	456.254	3	156.077	406.220	714.369	412.278
CLE40	RQVPTGSDPLHH	448.566	3	156.077	406.220	705.332	384.236
CLE45	RRVRRGSDPIHH	372.212	4	156.077	406.220	705.332	412.278
CLE7	RFSPGGPDPQHH	444.547	3	156.077	421.195	730.327	391.209
CLE5/6	RVSPGGPDPQHH	428.547	3	156.077	421.195	730.327	343.209
CLE2	RLSPGGPDPQHH	433.219	3	156.077	421.195	730.327	357.225
CLE1/3/4	RLSPGGPDPRHH	332.177	4	156.077	449.237	758.370	357.225
CLE16	RLVHTGPNPLHN	339.440	4	133.061	383.204	691.353	369.261
CLE22	RRVFTGPNPLHN	469.926	3	133.061	383.204	691.353	412.278
CLE14	RLVPKGPNPLHN	447.930	3	133.061	383.204	691.353	369.261
CLE9/10	RLVPSGPNPLHN	434.243	3	133.061	383.204	691.353	369.261
CLE19	RVIPTGPNPLHN	438.915	3	133.061	383.204	691.353	369.261
CLE21	RSIPTGPNPLHN	434.903	3	133.061	383.204	691.353	357.225
CLE17	RVVHTGPNPLHN	335.936	4	133.061	383.204	691.353	355.246
CLE20	RKVKTGSNPLHN	338.446	4	133.061	383.204	681.332	384.272
CLE25	RKVPNGPDPIHN	448.578	3	133.061	383.204	692.337	384.272
CLE26	RKVPRGPDPIHN	347.200	4	133.061	383.204	692.337	384.272
CLE18	RQIPTGPDPLHN	448.906	3	133.061	383.204	692.337	398.252
CLE27	RIVPSCPDPLHN	449.900	3	133.061	383.204	692.337	369.261
CLE46	HKHPSGPNPIGN	418.883	3	133.061	303.167	611.315	403.221
TDIF[b]	HEVPSGPNPISN	624.305	2	133.061	333.177	641.326	366.178
CLE42	HGVPSGPNPISN	588.294	2	133.061	333.177	641.326	294.157

注:a 表示最强准分子离子的质荷比;b 表示此处的TDIF序列为拟南芥十二肽基序,未经过翻译后修饰,如无特殊说明,正文中的TDIF均为羟基化修饰的成熟小肽,其序列结构列于表1

的工作可以从以下几个方面进行深入研究。首先,可根据需要进一步挖掘CLE小肽的保守性质谱特征,提高筛选的靶向性。通过提高碰撞能量、改变惰性气体、综合运用多种碎裂方式(如主要产生c型和z型子离子的电子转移碎裂,改善低质量碎片丢失效应的高能诱导裂解等)等方法获得CLE保守氨基酸残基(6位甘氨酸,4、7、9位脯氨酸,1位精氨酸/组氨酸)相关的高丰度碎片离子。其次,可利用高灵敏的多反应监测(MRM)模式作为辅助提高筛选效率。通过质谱碎裂规律和生物信息学预测的CLE结构可以推算出典型的子离子,从而给出MRM检测的离子对,设置合适的信号阈值触发二级质谱扫描,从而实现对可能真实存在的CLE小肽的检测和结构确定。最后,与生物学家合作,结合遗传学研究,发现和鉴定不同物种中的CLE小肽激素。

3　结论

本工作采用UPLC-QTOF MS联用技术研究了5种典型CLE小肽激素的一级质谱特征和最强准分子离子的CID碎裂规律,发现

CLE 小肽的质谱行为具有较高的相似性。在 ESI 离子化方式下，5 种 CLE 小肽的最强准分子离子携带的电荷均等于空间分散的碱性氨基酸残基数目加 1。CID 模式下进行二级质谱碎裂，最佳碰撞能量与母离子的电荷数相关，所带电荷数越高，需要的最佳碰撞能量越低，双电荷与三电荷的最佳碰撞能量分别约为 20 V、15～18 V。最强准分子离子主要产生 b 型和 y 型碎片，也会产生少量的 a 型碎片。从碎裂位置看，脯/羟脯氨酸的氨基（特别是氨端起 4 位和 9 位脯/羟脯氨酸）以及甘氨酸的氨基和羧基端最容易发生断裂，从而形成高丰度的 b_3 和 y_4 离子。人工合成的 CLE 类似肽的质谱碎裂与上述规律完美契合，可为 CLE 小肽激素的定性定量分析提供科学依据。结合 CLE 的结构特点和质谱碎裂规律对拟南芥十二肽基序的质谱行为进行预测，提出一种非数据依赖的 CLE 筛选思路，有可能应用于新 CLE 小肽激素的发现与结构鉴定中。

致谢：感谢中国科学院遗传与发育生物学研究所许操研究员实验室提供的人工合成 CLE 类似肽标准品（CLEL），感谢本实验室辛培勇博士对实验和数据分析提供的帮助和建议。

参考文献：

[1] MATSUBAYASHI Y, SAKAGAMI Y. Peptide hormones in plants[J]. Annual Review of Plant Biology, 2006, 57: 649-674.

[2] MOTOMITSU A, SAWA S, ISHIDA T. Plant peptide hormone signalling[J]. Essays in Biochemistry, 2015: 115-131.

[3] HIRAKAWA Y, SAWA S. Diverse function of plant peptide hormones in local signaling and development[J]. Current Opinion in Plant Biology, 2019, 51: 81-87.

[4] GAO X, GUO Y. CLE peptides in plants: proteolytic processing, structure-activity relationship, and ligand-receptor interaction[J]. Journal of Integrative Plant Biology, 2012, 54(10): 738-745.

[5] COCK J M, MCCORMICK S. A large family of genes that share homology with CLAVATA 3[J]. Plant Physiology, 2001, 126(3): 939-942.

[6] SAWA S, KINOSHITA A, BETSUYAKU S, FUKUDA H. A large family of genes that share homology with CLE domain in Arabidopsis and rice[J]. Plant Signaling & Behavior, 2008, 3(5): 337-339.

[7] OELKERS K, GOFFARD N, WEILLER G F, GRESSHOFF P M, MATHESIUS U, FRICKEY T. Bioinformatic analysis of the CLE signaling peptide family[J]. Bmc Plant Biology, 2008, 8: 1.

[8] OHYAMA K, SHINOHARA H, OGAWA-OHNISHI M, MATSUBAYASHI Y. A glycopeptide regulating stem cell fate in Arabidopsis thaliana[J]. Nature Chemical Biology, 2009, 5(8): 578-580.

[9] KONDO T, SAWA S, KINOSHITA A, MIZUNO S, KAKIMOTO T, FUKUDA H, SAKAGAMI Y. A plant peptide encoded by CLV3 identified by in situ MALDI-TOF MS analysis[J]. Science, 2006, 313(5 788): 845-848.

[10] HEWAVITHARANA A K, HERATH H M D R, SHAW P N, CABOT P J, KEBARLE P. Effect of solvent and electrospray mass spectrometer parameters on the charge state distribution of peptides-a case study using liquid chromatography/mass spectrometry method development for beta-endorphin assay[J]. Rapid Commun Mass Spectrometry, 2010, 24(24): 3 510-3 514.

[11] 孙林，郭琳，秦群，郑弦，阮贵华，杜甫佑. 烟草系统素多肽的质谱裂解规律研究[J]. 质谱学报，2015，36(5)：394-402.
SUN Lin, GUO Lin, QIN Qun, ZHENG Xian, RUAN Guihua, DU Fuyou. Study on mass spectrometric fragmentation mechanism of tobacco systemin polypeptides[J]. Journal of Chinese Mass Spectrometry Society, 2015, 36(5): 394-402(in Chinese).

[12] PAIZS B, SUHAI S. Fragmentation pathways of protonated peptides[J]. Mass Spectrometry Reviews, 2005, 4(4): 508-548.

[13] 高丽. 植物多肽信号分子 CLE 家族[J]. 生物技术通报，2014(11)：14-23.
GAO Li. CLE peptide signaling molecules in plant[J]. Biotechnology Bulletin, 2014(11): 14-23(in Chinese).

（收稿日期：2021-02-27；修回日期：2021-04-04）