

## 细胞穿膜肽的穿膜活性与序列特征的关系\*

曹赞霞<sup>1, 2)\*\*</sup> 董川<sup>3)\*\*</sup> 赵立岭<sup>1, 2)</sup> 王吉华<sup>1, 2)\*\*\*</sup>

<sup>1)</sup> 德州学院山东省生物物理重点实验室, 德州 253023; <sup>2)</sup> 德州学院物理与电子信息学院, 德州 253023;

<sup>3)</sup> 电子科技大学生命科学与技术学院生物信息学中心, 成都 610054

**摘要** 细胞穿膜肽(cell penetrating peptides, CPPs)是一种小分子多肽, 能够容易地穿过细胞膜. 这类分子, 尤其是具有靶向功能的 CPPs 为高效率投送药物到靶细胞带来希望. 因此, 对其展开研究对于生物医学有着一定的意义. 本工作主要从序列水平对具有不同穿膜活性的 CPPs 进行研究, 试图找出影响 CPPs 穿膜活性的因素, 以及不同活性 CPPs 与非穿膜肽(NonCPPs)序列上的差异, 并引入一种分析生物序列的方法. 我们基于 CPPsite 数据库和不同的文献获取 CPPs 和 NonCPPs 序列, 并进一步从 CPPs 序列中提取具有高、中、低穿膜活性的穿膜肽(HCPPs、MCPPs、LCPPs)用于构建数据集. 基于这些数据集, 开展了以下研究: 首先, 利用方差分析的方法, 对不同活性的 CPPs 以及 NonCPPs 的氨基酸及二级结构组成进行分析, 发现氨基酸的静电与疏水相互作用对 CPPs 的穿膜活性起到了重要影响, 同时螺旋结构和无规卷曲也会影响 CPPs 的穿膜活性; 其次, 使用理化性质与长度将不同活性的 CPPs 展示在二维平面上, 发现在某些特殊的性质下不同活性的 CPPs 与 NonCPPs 可以产生聚簇现象, HCPPs、MCPPs 以及 LCPPs 和 NonCPPs 被分成了三簇, 这种现象显示了它们之间的差异; 最后, 本文引入了生物序列理化质心的概念, 将组成序列的残基看作质点, 进而把序列抽象成质点系进行研究, 并将此方法应用到 CPPs 的分析中, 通过 PCA 方法将不同活性的 CPPs 投射到三维平面上, 结果发现绝大部分 CPPs 聚在一起, 部分 LCPPs 与 NonCPPs 聚在一起. 此工作对于 CPPs 的设计, 以及理解不同活性 CPPs 序列上的差异具有一定的意义. 另外, 本文引入的生物序列理化质心的分析方法也可以用于其他生物问题的分析, 同时它们可以作为某些生物分类问题的输入参数, 在模式识别中起到一定的作用.

**关键词** 细胞穿膜肽(CPPs), 穿膜活性, 序列差异, 理化质心

**学科分类号** Q615, Q7

**DOI:** 10.16476/j.pibb.2015.0271

在药物的开发过程中, 药物分子在靶细胞中的递送效率是限制药物疗效的一个主要因素, 而细胞穿膜肽(cell-penetrating peptides, CPPs)或称之为蛋白转导域(protein transduction domains, PTDs)<sup>[1]</sup>的发现, 为进一步提高药物分子的递送效率, 克服这一障碍带来了一线光明. CPPs 是一类能够穿膜的小分子多肽, 氨基酸残基数目多集中于 5~30 个残基, 富含精氨酸与赖氨酸, 多数带有正电荷, pH 近似中性<sup>[2]</sup>. CPPs 具有强的货物运输能力, 如可以作为蛋白质、siRNA<sup>[3]</sup>、核酸分子<sup>[4-5]</sup>和脂质体(liposomes)等的递送载体, 亦可以作为药物分子的递送载体. 作为理想的递送分子, CPPs 与其他非天然的分子相比具有较低的细胞毒性, 可以携带多

种不同大小和性质的生物活性物质进入细胞<sup>[6]</sup>. 然而, 随着研究的深入, 现有 CPPs 作为药物递送分子存在一定的不足. 例如, 非靶向功能的 CPPs 不能专一性地递送药物分子到病变的靶细胞<sup>[7]</sup>. 此外, CPPs 如果带有过多的正电荷, 这或许会对靶

\* 国家自然科学基金(31000324, 61271378, 11447004), 山东省自然科学基金(ZR2012CL09, ZR2014AL014, ZR2014JL006)及山东省泰山学者资助项目.

\*\* 共同第一作者.

\*\*\* 通讯联系人.

Tel: 0534-8985933, E-mail: jhw25336@126.com

收稿日期: 2015-09-01, 接受日期: 2015-12-01

细胞有一定的毒性<sup>[8]</sup>. 因此, 对 CPPs 进行深入研究, 进而规避 CPPs 作为药物递送分子的弊端具有重要的意义. 目前, 国内外学者针对 CPPs 分子已经进行了一系列的研究: 一方面, 从实验上寻找或设计各类疾病特异性靶向穿膜肽是药物研发的热点; 另一方面, 通过计算的方法, 对如何设计新型穿膜肽提供指导. 计算方法主要包含两种: 一种是分子动力学模拟. 2014 年, 我们就 CPPs 跨膜分子机制及其靶向改造的分子动力学模拟研究进行了评述<sup>[9]</sup>. 另一种是生物信息学的手段, 很多研究小组通过发展算法及开发预测平台对细胞穿膜肽以及 CPPs 的穿膜活性进行预测<sup>[10-13]</sup>. 例如, Ankur Gautam 等曾使用蛋白质序列谱、二肽的频率作为 CPPs 的特征向量, 以此来预测 CPPs, Hansen 等则基于氨基酸的理化性质, 使用 Z-Score 来预测 CPPs. 此外, 还有很多课题组为了提高 CPPs 的穿膜活性, 通过实验手段对它们进行人工修饰, 进而提高 CPPs 的药物递送效率. 为了系统地研究 CPPs, Gautam 等<sup>[14]</sup>建了 CPPsite 数据库(<http://crdd.osdd.net/raghava/cppsite1/>). CPPsite 数据库 1.0 版本收录的 CPPs 已达到 843 条, 既包含天然的, 也包含人工设计的 CPPs, 并根据穿膜活性将其分为高穿膜活性的 CPPs (CPPs with high penetrating activities, HCPPs)、中穿膜活性的 CPPs (CPPs with medium penetrating activities, MCPPs) 和低穿膜活性的 CPPs (CPPs with low penetrating activities, LCPPs).

本工作中, 一方面, 我们分析比较了具有不同活性的细胞穿膜肽在序列上的差异, 另一方面, 我们提出理化质心的概念用于 CPPs 的分析, 该方法还可用于其他生物学问题的研究, 为生物学分类问题的模式识别提供新的输入特征.

## 1 材料与方法

### 1.1 数据集的构建

CPPsite 数据库中收录了自然存在和人工修饰后的 CPPs, 从此数据库的 1.0 版本中获取 CPPs 序列, 并进一步排除含有 D 型氨基酸的 CPPs. 最终, 得到了仅由 L 型氨基酸构成的 776 条 CPPs. 考虑到位点特异性残基可能会影响到 CPPs 的穿膜活性, 研究中没有对数据集进行去冗余处理, 在其他研究中也见到过类似的做法<sup>[1]</sup>. 不同穿膜活性的 CPPs 其序列上必然存在差异, 从 CPPs 数据集中提

取了穿膜活性具有明确表述的穿膜肽, 例如 “High”、“Medium”、“Low”, 排除那些穿膜活性不确定的 CPPs 后, 最终得到 56 条 HCPPs, 53 条 MCPPs 和 81 条 LCPPs. 另外, 从文献中选取 34 条非穿膜肽 (Non-cell penetrating peptides, NonCPPs)<sup>[10-11]</sup>组成非穿膜肽数据集. 各组的长度分布如表 1 所示.

**Table 1 Length distribution for CPPs with different penetrating activities and NonCPPs**

Type of CPPs	Range of length
HCPPs <sup>1)</sup>	5~28
MCPPs <sup>2)</sup>	5~22
LCPPs <sup>3)</sup>	3~34
NonCPPs <sup>4)</sup>	8~36

<sup>1)</sup> CPPs with high penetrating activities; <sup>2)</sup> CPPs with medium penetrating activities; <sup>3)</sup> CPPs with low penetrating activities;

<sup>4)</sup> Non-cell penetrating peptides.

### 1.2 方差分析 (ANOVA)

方差分析 (analysis of variance, ANOVA) 可以用于多个样本之间均数显著性差异的分析<sup>[15]</sup>, 其基本思想是通过分析每种因素对总变异的贡献关系, 来确定各因素对总变异产生的影响, 进而确定影响总变异的主要因素. 本工作构建了 4 个数据集, 分别由高、中、低及非穿膜活性的多肽组成, 对每种数据集中的 20 种氨基酸的含量进行了计算, 并对每个组别的相同氨基酸的含量进行了方差分析. 在统计检验过程中通常将  $P$  的阈值设置为 0.05 或者 0.01, 以此阈值作为判断显著性差异的标准. 为了使分析结果更加可靠, 我们采用了 0.01 这一阈值, 当  $P \leq 0.01$  时, 则认为它们的含量具有显著性差别, 并认为它们对 CPPs 的穿膜活性产生了影响.

### 1.3 序列的理化性质与长度分布

所构建的数据集中, 还有一些意义不明确的氨基酸 X、B、Z、U. 其中, X 表示 20 种常见氨基酸的任一种, U 代表硒代半胱氨酸, Z 表示谷氨酸或谷氨酰胺, B 表示天冬氨酸或天冬酰胺. 在接下来的讨论中, 如果不同活性的 CPPs 或者 NonCPPs 中的多肽含有这些氨基酸残基, 将会对这些残基的

理化性质做一个近似的处理: X 用 20 种常见氨基酸的平均理化性质来代替; U 用半胱氨酸的理化性质来代替; B 用天冬氨酸与天冬酰胺的平均理化性质来代替; Z 的理化性质则用谷氨酸与谷氨酰胺理化性质的平均值来代替. 从文献中选取了氨基酸的疏水性指数(H1)、亲水性指数(H2)、氨基酸的分子质量(M)、氨基酸氮端与碳端的解离常数值(pK<sub>1</sub>和pK<sub>2</sub>), 以及氨基酸的等电点(pI)等性质用于分析理化性质对 CPPs 穿膜活性的影响. 所有的这些性质可以从 <http://www.csbio.sjtu.edu.cn/bioinf/PseAA/ParaValue.htm> 获取. 对组成肽链的所有氨基酸, 将所关注的性质按照公式(1)进行计算, 即得到该性质的  $\langle p \rangle$  值:

$$\langle p \rangle = \frac{\sum_{i=1}^L p^0(AA_i)}{L} \quad (1)$$

上式(1)中  $p^0(AA_i)$  表示氨基酸  $AA_i$  的理化性质  $p^0$ ,  $L$  表示每类数据集中最长穿膜肽的序列长度. 于是每一条肽可以用点  $(\langle p \rangle, l)$  表示 ( $l$  表示对应 CPP 的真实长度), 由于选取了 6 种理化性质, 所以对于每条序列可以计算 6 组  $(\langle p \rangle, l)$ , 我们将以此来展示不同活性的 CPPs 与 NonCPPs 序列上的差异, 以及理化性质对其穿膜活性的影响.

#### 1.4 生物序列的理化质心

为了进一步从全局角度对不同活性的 CPPs 与 NonCPPs 进行研究. 我们从物理学中质点系的角度引入了生物序列理化质心的概念, 其基本思想是将生物序列等效成一系列的点来研究. 为了引入生物序列理化质心这一概念, 首先规定序列中每个氨基酸残基的坐标: 如果一条序列从 N 端到 C 端从 1 开始对序列进行编号, 考察长度为  $L$  的生物学序列的第  $i$  个残基, 那么  $(x_i, y_i)$  表示的是第  $i$  个氨基酸的一级结构序列坐标.  $x_i$  和  $y_i$  可以由下面的公式计算得到:

$$\begin{aligned} x_i &= i-1 \\ y_i &= l-i \end{aligned} \quad (2)$$

如果给定  $(x_i, y_i)$ , 那么就可以用上述的公式(2)计算出  $i$  与  $l$ . 因此, 坐标  $(x_i, y_i)$  包含残基的位置与生物序列的长度这两个方面的信息. 如果把生物序列看成一个质点系, 而组成它的残基看成质点系中的质点. 于是, 进一步模仿物理学中质心的概念, 将其引入生物序列的分析之中. 对于给定的任何一条生物学序列, 都可以计算这条序列的理化质心, 计算公式如下:

$$\begin{aligned} X_c &= \frac{\sum_{i=1}^L p(AA_i)x_i}{\sum_{i=1}^L p(AA_i)} \\ Y_c &= \frac{\sum_{i=1}^L p(AA_i)y_i}{\sum_{i=1}^L p(AA_i)} \end{aligned} \quad (3)$$

公式(3)中  $p(AA_i)$  表示肽链的第  $i$  个氨基酸残基的性质  $p$ ,  $p$  是选取的 6 种理化性质之一的标准化值(公式(4)).  $(X_c, Y_c)$  表示性质  $p$  下的理化质心, 它可以从整体的角度反应所考察序列的性质. 由于上述公式包含了之前定义的坐标信息, 所以点  $(X_c, Y_c)$  可以从一定程度上反映出序列的长度信息与氨基酸的位置信息. 此外, 通过计算 6 组性质下的  $(X_c, Y_c)$ , 我们就可以使用这些点来研究生物学序列. 一方面, 它能反映整个序列某个性质分布最集中的点; 另一方面, 通过序列的理化质心还可以将一条抽象的序列转化成具体的点进行考察. 某些序列上的差异可以通过这些点的分布进一步在几何图像上显示, 更形象地显示出不同活性 CPPs, 以及 NonCPPs 序列上的一些差异, 这是该参数的优势之一. 其次, 它也可以作为某些生物模式识别的输入参数, 根据所选取的理化参数的数目,  $(X_c, Y_c)$  数目也会相应增长, 提取到更多有价值的信息. 基于此, 我们认为它的引入对于 CPPs 的分析是有一定帮助的. 由于不同氨基酸的理化性质之间的量纲不同, 这就导致了不同类型的性质之间存在较大的差异. 因此, 在计算生物序列的理化质心之前, 先用下面的公式对各种性质进行标准化处理.

$$p = \frac{p^0 - \langle p^0 \rangle}{\text{std}(p^0)} \quad (4)$$

式(4)中  $p^0$  表示氨基酸标准化以前的性质,  $\langle p^0 \rangle$  为  $p^0$  的平均值,  $\text{std}(p^0)$  是  $p^0$  的标准差,  $p$  表示标准化以后的性质. 所有的性质可以在 <http://www.csbio.sjtu.edu.cn/bioinf/PseAA/ParaValue.htm> 获取.

## 2 结果与讨论

### 2.1 不同穿膜活性 CPPs 中具有显著差异的氨基酸

我们分别计算了高、中、低活性的 CPPs 以及 NonCPPs 数据集中每条短肽的每种氨基酸的含量, 并用单因素方差分析方法对相同氨基酸在不同数据集中的含量进行了分析. 如果  $P \leq 0.01$ , 则认为该氨基酸在不同的数据集中含量具有显著的差异, 这

些氨基酸在决定 CPPs 的穿膜活性方面起了关键的作用. 图 1 中横坐标表示氨基酸, 纵坐标表示对应

氨基酸含量在不同类别数据集中的  $P$  值,  $P \leq 0.01$  的氨基酸在图 1 中用箭头指示.

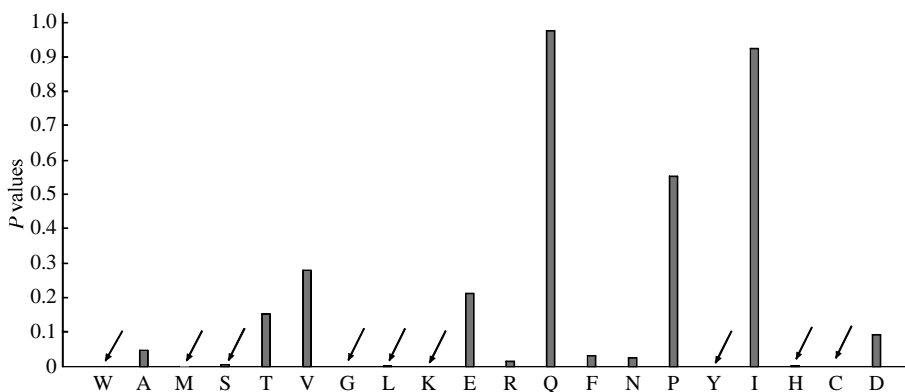


Fig. 1 P values for 20 kinds of amino acids

The amino acids with  $P \leq 0.01$  are indicated by arrows. Obviously, W, M, S, G, L, K, Y, H, C have significant difference among HCPPs, MCPPs, LCPPs and NonCPPs, respectively.

图 1 显示, 氨基酸 W、M、S、G、L、K、Y、H 和 C 的  $P \leq 0.01$ , 说明这些氨基酸在不同活性的 CPPs 中含量具有显著性差异, 它们对于 CPPs 的穿膜活性起到了重要的作用. 进一步分析可以发现, 这些氨基酸按照其理化性质分成两类: 极性和非极性的氨基酸 (S、G、Y、W、M、L、C) 以及带电荷的氨基酸 (K、H). 前者的含量可以影响 CPPs 与 NonCPPs 的整体疏水性, 而疏水性的强弱会影响多肽与脂膜的相互作用; 后者则影响了 CPPs 与 NonCPPs 整体的带电水平, 脂膜外层的磷酸基团则带有一定的负电荷, 所以小肽的带电水平可以影响其与脂质膜的静电相互作用. 这提示, 在人工设计 CPPs 时, 这些氨基酸是重点考虑的对象之一. 最近有研究发现 H16 (HHHHHHHHHHHHHHHH-NH<sub>2</sub>) 具有很高的穿膜活性<sup>[16]</sup>, 该条肽中的 H 含量与数据集中的 MCPPs、LCPPs 中含有的 H 差别明显, 图 1 也表明 H 的含量在不同活性的 CPPs 与 NonCPPs 中也具有较小的  $P$  值, 这与我们的研究是一致的. 此外, 有的研究也指出, 带正电荷的氨基酸对于 CPPs 的穿膜活性起了重要的作用<sup>[16]</sup>, 而图 1 显示具有显著性差异的氨基酸 K、H 也带有正电荷, 二者是一致的. 值得注意的是如果将显著性差异的阈值  $P$  由 0.01 调整为 0.05, 非极性氨基酸 F, 极性氨基酸 N, 以及另外一个正电荷氨基酸 R 也分别被包含进极性、非极性和带电荷的氨基酸分

类中.

## 2.2 理化性质与长度分布显示了不同活性穿膜肽之间的差异

使用 1.3 的方法计算了不同穿膜活性的 CPPs 以及 NonCPPs 中每条肽的  $\langle p \rangle$  值. 每一条 CPPs 与 NonCPPs 都可以用  $\langle p \rangle, l$  这样一个点来表示. 分别选取氨基酸的疏水性指数 ( $H1$ )、亲水性指数 ( $H2$ )、分子量 ( $M$ )、氨基酸氮端与碳端的解离常数 ( $pK_1$  和  $pK_2$ ) 以及氨基酸的等电点 ( $pI$ ) 计算  $\langle p \rangle$ , 将不同活性的 CPPs 以及 NonCPPs 展现在二维图上, 如图 2 所示. 图 2 中蓝色菱形表示 MCPPs, 红色实心圆圈表示 HCPPs, 绿色五角星表示 LCPPs, 黑色正方形表示 NonCPPs. 在图 2a~c, f 中, 发现代表不同穿膜活性 CPPs 的点混在一起, 表明这些氨基酸相应性质的参数 (疏水性指数  $H1$ 、亲水性指数  $H2$ 、分子量  $M$  以及氨基酸的等电点  $pI$ ) 不能很好地区分 CPPs 的穿膜活性. 而在图 2d, e 中, 对应低穿膜活性 LCPPs 的点 (绿色) 和对应非穿膜肽 NonCPPs 的点 (黑色) 混在一起, 分别与对应高穿膜活性 HCPPs 点 (蓝色)、中穿膜活性 MCPPs 点 (红色) 区别明显. 这说明 CPPs 的序列长度与每条肽的  $pK_1$ 、 $pK_2$  计算出的  $\langle p \rangle$  可以用来分析 CPPs 的穿膜能力. 进一步分析可知,  $pK_1$ 、 $pK_2$  与 CPPs 侧链的结合解离有关, 这一点直接影响到 CPPs 在内环境中的电荷水平, 进而影响 CPPs 与脂膜的相

互作用. 图 2d, e 表明 NonCPPs 与 LCPPs 在上述 6 种性质下的不可分现象, 说明它们之间在某些性质上比较接近, 这提示某些基于性质的分类器可能会对两者产生错分, 从而影响预测效果. 此外, 在长度限定下, 由  $pK$  计算出的  $\langle p \rangle$  对 CPPs 的分簇现象可能是由于长度会影响到 CPPs 的二级结构所致. CPPsite 数据库的 1.0 版本将 CPPs 的结构分成了 4 类 (<http://crdd.osdd.net/raghava/cppsite1/dssp-com.php>): a. 螺旋类, 包含 H( $\alpha$ -螺旋)、G(310-螺旋)、I( $\pi$ -螺旋); b.  $\beta$  折叠类, 包含 B(单个的桥, 即  $\beta$  桥)、E(延伸的  $\beta$  桥, 即  $\beta$  股); c. 转角类, 包含 T(转角)、S(卷曲); d. 卷曲(coil, C). 我们从 CPPsite 1.0 版本获得了数据集中每个肽的二级

结构, 并计算了每个肽中这 4 类结构的含量, 使用方差分析分别对每一类结构在 3 类不同穿膜活性 CPPs 中的含量做分析, 螺旋( $P=0.0036$ )结构与无规则卷曲( $P=0.0063$ )在 3 类 CPPs 中的含量具有显著性的差异, 说明这两种结构可能会影响到 CPPs 的穿膜活性. HCPPs、MCPPs、LCPPs 中螺旋结构的平均含量为 0.25、0.09 和 0.18, HCPPs 含有更多的螺旋结构, MCPPs 和 LCPPs 则含有相对较低的螺旋结构. 而 HCPPs 相比于 MCPPs 和 LCPPs 则具有较少的无规则卷曲结构. 由此可以得出结论, CPPs 的二级结构、CPPs 与脂膜的电荷相互作用及疏水相互作用会共同影响其穿膜活性.

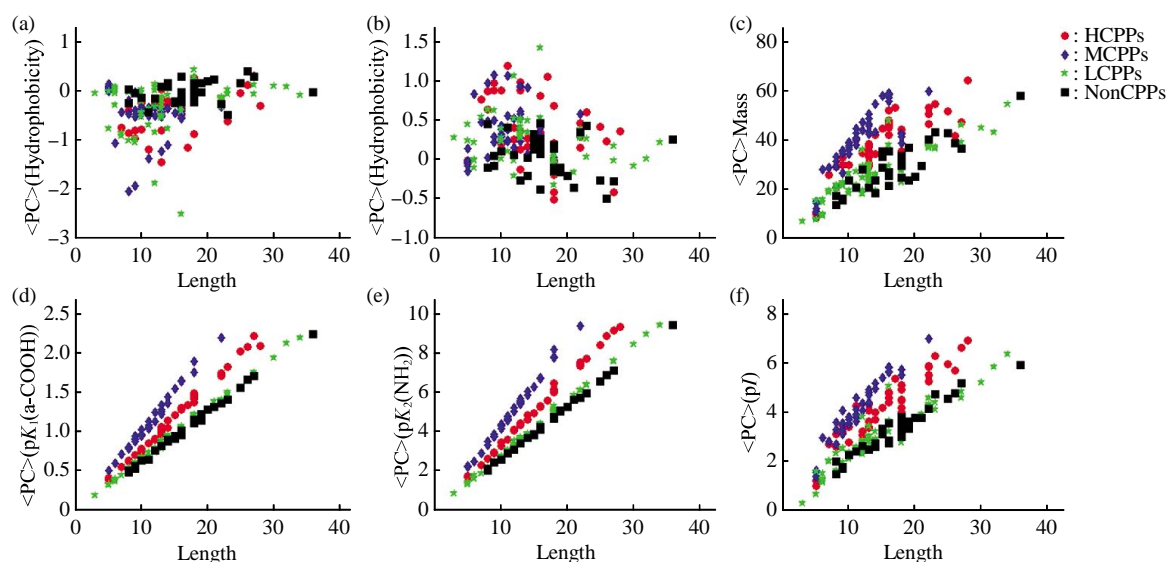


Fig. 2 Impact of CPPs' physico-chemical properties on CPPs penetrating activities

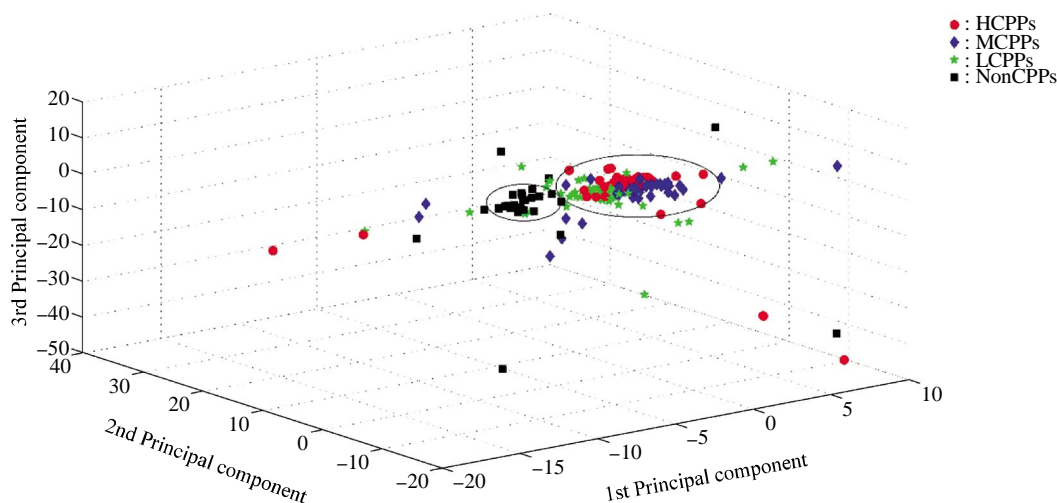
HCPPs, MCPPs, LCPPs and NonCPPs are denoted by red circles, blue diamond five-pointed stars, black squares, respectively. From (a) to (c) and (f),  $y$  represents the  $\langle p \rangle$  values of CPPs calculated by hydrophobicity ( $H1$ ), hydrophilicity ( $H2$ ), mass ( $M$ ), and  $pI$ . For (d) and (e) denoted the  $\langle p \rangle$  values calculated by  $pK$  (using  $pK_1$  to represent N-terminal,  $pK_2$  for C-terminal) and  $x$  represents the length of CPPs sequence.

### 2.3 不同活性的 CPPs 在理化质心下的分布

对数据集中的每个 CPPs 或 NonCPPs, 根据前述残基的 6 种理化性质, 分别计算了序列的理化质心, 并使用每种理化性质对应的  $Xc/Yc$  的比值来代替 CPPs 或 NonCPPs. 由于选取了 6 种性质, 所以每条序列可以用 6 个数值来描述. 对不同活性的 CPPs 以及 NonCPPs 分别进行主成分分析, 结果显示, HCPPs 第一、二、三主成分的累积贡献率达到了 90.54%, MCPPs 的第一、二、三主成分的累积贡献率达到了 90.49%, LCPPs 的第一、二、三主成分的累积贡献率达到了 81.14%, NonCPPs 的第一、二、三主成分的累积贡献率达到了

97.84%. 根据 PCA 分析的前 3 个主成分将研究体系投射到三维平面中, 如图 3 所示.

图 3 显示出 HCPPs、MCPPs 以及 LCPPs 的第一、二、三主成分的 score 之间没有明显的界线, 表明 HCPPs、MCPPs、LCPPs 以及 NonCPPs 在整体理化性质方面是逐渐变化的, 而不是跳跃式变化, 当这种变化达到一定的程度, CPPs 就会由一种活性转变为另一种活性. 此外, 这种分布显示不同活性 CPPs 序列上的差异. 同时, 图 3 中 NonCPPs 与 LCPPs 产生了与图 2d, e 类似的现象, 即它们之间不易分开, 表明 LCPPs 与 NonCPPs 序列的理化性质之间确实存在某种相似性.



**Fig. 3 Graph of top three principal components**

The dataset were separated into two groups by the top three principal components. In Figure 3 we used different colors and symbols to represent the CPPs with different penetrating activities.

### 3 总 结

高活性的 CPPs 不但可以高效地投送药物到细胞, 而且某些具有靶向功能的 CPPs 还可以起到靶向递送的作用. 因此, 研究 CPPs 对于生物医学有重要意义. 在本研究中, 我们开展了以下 3 个方面的工作: a. 基于方差分析, 分析了不同活性 CPPs 以及 NonCPPs 在氨基酸组成及二级结构组成的差异性, 具有显著组成差异的氨基酸包含一些极性与非极性氨基酸 (W、M、S、G、L、K、Y、H、C) 和带正电荷的氨基酸 (K、H), 表明 CPPs 与细胞膜的疏水与静电相互作用对其穿膜活性有着重要的影响. 同时无规卷曲和螺旋结构在三类不同穿膜活性的 CPPs 中的含量具有显著性差异. b. 通过  $\langle p \rangle$ ,  $l$  来表示不同活性的 CPPs 与 NonCPPs, 发现  $pK$  值与长度可以对不同活性的 CPPs 进行区分, 结果表明 CPPs 的带电水平与结构会影响它们的穿膜活性 (螺旋与卷曲). c. 受到质点力学中质心概念的启发, 我们将其引入到了生物序列的分析之中, 并将其应用到不同活性 CPPs 与 NonCPPs 的序列分析中.

本研究分析了影响 CPPs 穿膜活性的因素, 以及 CPPs 序列上的差异, 这对于 CPPs 的设计与预测具有一定的意义. 在下一步的工作中我们将结合

CPPs 与 NonCPPs, 以及不同活性的 CPPs 的差异对其进行预测分析, 并搭建在线的预测平台.

### 参 考 文 献

- [1] Gautam A, Chaudhary K, Kumar R, *et al.* In silico approaches for designing highly effective cell penetrating peptides. *Journal of Translational Medicine*, 2013, **11** (13): 1-12 (DOI:10.1186/1479-5876-11-74)
- [2] Jiang T, Zhang Z, Zhang Y, *et al.* Dual-functional liposomes based on pH-responsive cell-penetrating peptide and hyaluronic acid for tumor-targeted anticancer drug delivery. *Biomaterials*, 2012, **33**(36): 9246-9258
- [3] Gooding M, Browne L P, Quinteiro F M, *et al.* siRNA delivery: from lipids to cell-penetrating peptides and their mimics. *Chemical Biology & Drug Design*, 2012, **80**(6): 787-809
- [4] Bai H, You Y, Yan H, *et al.* Antisense inhibition of gene expression and growth in gram-negative bacteria by cell-penetrating peptide conjugates of peptide nucleic acids targeted to rpoD gene. *Biomaterials*, 2012, **33**(2): 659-667
- [5] Huang Y W, Lee H J, Tolliver L M, *et al.* Delivery of nucleic acids and nanomaterials by cell-penetrating peptides: opportunities and challenges. *BioMed Research International*, 2015, 834079: doi: 10.1155/2015/834079
- [6] Wang F, Wang Y, Zhang X, *et al.* Recent progress of cell-penetrating peptides as new carriers for intracellular cargo delivery. *Journal of Controlled Release: Official Journal of the Controlled Release Society*, 2014, **174**(1): 126-136
- [7] Vives E, Schmidt J, Pelegrin A. Cell-penetrating and cell-targeting

- peptides in drug delivery. *Biochimica et Biophysica Acta*, 2008, **1786**(2): 126-138
- [8] El-Aneed A. An overview of current delivery systems in cancer gene therapy. *Journal of Controlled Release: Official Journal of the Controlled Release Society*, 2004, **94**(1): 1-14
- [9] 曹赞霞, 刘磊, 王吉华. 分子动力学模拟研究穿透肽的跨膜机制及引导新肽设计. *科学通报 (中文版)*, 2014, **59**(22): 2160-2168  
Cao Z X, Liu L, Wang J H. *Chinese Science Bulletin*, 2014, **59**(22): 2160-2168
- [10] Hansen M, Kilk K, Langel U. Predicting cell-penetrating peptides. *Advanced Drug Delivery Reviews*, 2008, **60**(4-5): 572-579
- [11] Hallbrink M, Karelson M. *Cell-Penetrating Peptides*. New York: Springer New York, 2015: 39-58
- [12] Dobchev D A, Mager I, Tulp I, *et al.* Prediction of cell-penetrating peptides using artificial neural networks. *Current Computer-Aided Drug Design*, 2010, **6**(2): 79-89
- [13] Sanders W S, Johnston C I, Bridges S M, *et al.* Prediction of cell penetrating peptides by support vector machines. *PLoS Computational Biology*, 2011, **7**(7): e1002101
- [14] Gautam A, Singh H, Tyagi A, *et al.* CPPsite: a curated database of cell penetrating peptides. *Database: the journal of biological databases and curation*, 2012, **2012**: bas015
- [15] Freund R J, Littell R C. *SAS for linear models: a guide to the ANOVA and GLM procedures*. Sas Institute, 1981
- [16] Iwasaki T, Tokuda Y, Kotake A, *et al.* Cellular uptake and *in vivo* distribution of polyhistidine peptides. *Journal of Controlled Release: Official Journal of the Controlled Release Society*, 2015, **210**: 115-124

## Relationship Between Cell Penetrating Peptides (CPPs) With Different Activities and Their Sequence Characteristics\*

CAO Zan-Xia<sup>1,2)\*\*</sup>, DONG Chuan<sup>3)\*\*</sup>, ZHAO Li-Ling<sup>1,2)</sup>, WANG Ji-Hua<sup>1,2)\*\*\*</sup>

<sup>1)</sup> Key Laboratory of Biophysics of Shandong, Dezhou University, Dezhou 253023, China;

<sup>2)</sup> The College of Physics and Electronic Information of Dezhou University, Dezhou 253023, China;

<sup>3)</sup> Center of Bioinformatics School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu 610054, China)

**Abstract** Cell penetrating peptides (CPPs) are kinds of small molecular peptides, which can penetrate cell membrane easily. Their discovery sheds light on delivering drugs into target efficiently. Therefore, the investigation of CPPs has crucial significance in biomedicine field. In this work we intended to find factors which have influence on CPPs penetrating activity and introduce a new method called physico-chemical mass center of bio-sequence. This method was enlightened by particle mechanics. We used it to analyze bio-sequence. In addition, the differences between all kinds of CPPs with different penetrating activities were also investigated. The CPPs and Non-cell penetrating peptides (NonCPPs) were obtained based on CPPsite database version 1.0 (<http://crdd.osdd.net/raghava/cppsite1/>) and references, respectively. After that we extracted CPPs with high, medium, and low penetrating activities (HCPPs, MCPPs, LCPPs), respectively. Firstly, the amino acids composition was calculated for every peptide in HCPPs, MCPPs, LCPPs datasets, individually. We used ANOVA to analyze the amino acids composition in HCPPs, MCPPs, LCPPs datasets. From the results we noticed that there has a significant difference for amino acids with electric charge and hydrophobic properties. These results revealed that electrostatic and hydrophobic interactions may play a key role in CPPs' penetrating activities. ANOVA analysis also revealed that helix and coil structure may also have influence on CPPs' activities. Secondly, CPPs and NonCPPs can yield a clustering phenomenon under some physicochemical features such as pK values, and their corresponding lengths. HCPPs, MCPPs, LCPPs, and NonCPPs can be divided into three groups. The differences between CPPs and NonCPPs can be reflected by the above phenomena. Lastly, as a case study we used the new concept named physico-chemical mass center of bio-sequence method encouraged by particle mechanics to study CPPs. The sequences can be represented by a series of points in the above method. Combining with principal component analysis (PCA), we found that the distribution of CPPs under their first, second and third scores can generate a cluster for CPPs. The differences between HCPPs, MCPPs, LCPPs and NonCPPs were also demonstrated by the above phenomenon. From this study, we can conclude that this work can help us to further understand the differences between them in terms of their primary structures and the factors affecting on their activities. Most important, the physico-chemical mass center can be used to analyze other biological issues as well as input features for biological pattern recognitions.

**Key words** cell penetrating peptides (CPPs), penetrating activities, sequence differences, physico-chemical mass center

**DOI:** 10.16476/j.pibb.2015.0271

\*This work was supported by grants from The National Natural Science Foundation of China (31000324, 61271378, 11447004), Natural Science Foundation of Shandong Province, China (ZR2012CL09, ZR2014AL014, ZR2014JL006) and the Taishan Scholar Program of Shandong.

\*\*These authors contributed equally to this work.

\*\*\*Corresponding author.

Tel: 86-534-8985933, E-mail: jhw25336@126.com

Received: September 1, 2015 Accepted: December 1, 2015