

线性短模体：介导蛋白质相互作用的新模块*

刘璐 张志超**

(大连理工大学化学学院, 精细化工国家重点实验室, 大连 116030)

摘要 线性短模体是天然无序蛋白实现生物学功能的重要组件。线性短模体具有柔性结构和短小的序列, 可以介导瞬时、可逆的蛋白质相互作用, 并在发生相互作用时表现出杂泛性。随着实验技术的更新和预测手段的发展, 越来越多的线性短模体被发现和重新定义, 例如 BH3 线性短模体。本文重点总结了线性短模体在结构、生物学功能以及进化等方面的特点。对线性短模体功能的研究将为解析细胞信号转导网络、疾病靶标确认、新药发现等领域带来新的思路。

关键词 线性短模体, 天然无序蛋白, 蛋白质相互作用, 进化

学科分类号 Q7, Q51

DOI: 10.16476/j.pibb.2016.0245

蛋白质相互作用 (protein-protein interactions, PPIs) 决定着生物体生长、发育、分化、凋亡等生命活动, 是生物调控机制的重要组成部分。蛋白质结构域的研究领域曾经普遍认为 PPIs 的实现必须通过良好折叠的、具有特定三维空间结构的结构域 (globular domains) 才能完成^[1]。但是, 随着结构生物学和功能蛋白质组学的进展, 近 10 年来的研究发现, 很多蛋白质中包含一些天然无序区, 虽然没有特定的三维结构, 但是在与伴侣蛋白结合时, 可以形成特定的三维结构, 并介导蛋白质相互作用^[2]。这些蛋白质被称为天然无序蛋白 (intrinsically disordered proteins, IDPs), 或者称为包含天然无序区 (intrinsically disordered regions, IDRs) 的蛋白。这类蛋白质广泛参与信号转导过程^[3]。天然无序蛋白主要通过线性短模体 (short linear motifs, SLiMs) 和无序结构域 (disordered domains) 发挥作用。线性短模体^[4]是天然无序蛋白中较为普遍的功能模块, 是由连续分布的氨基酸残基形成的短而紧凑的蛋白质作用区域 (位点), 又称为真核生物线性模体 (eukaryotic linear motifs, ELMs)^[5]、迷你模体 (minimotifs)^[6]等。由于线性短模体的序列短小、结构柔性强, 可以与多个伴侣蛋白发生相互作用, 尤其是发生瞬时的、低亲和力的、动态可逆的相互作用^[7]。

研究线性短模体的结构特征、生物学功能和进化特点, 有助于我们进一步理解细胞功能调控的分子机制, 发现和确认新的线性短模体, 丰富对于蛋白质相互作用的认知。

1 线性短模体增加 PPIs 的复杂性

1.1 线性短模体的数量

研究者们预测线性短模体的数量高达 1 000 000 个^[8]。并且, 由于结构柔性, 一个线性短模体能够与多个伴侣蛋白发生相互作用^[9], 进一步增加了蛋白质相互作用网络的复杂性。另外, 可变剪切可以使蛋白增加或丢失线性短模体, 产生具有不同功能和相互作用特征的蛋白质亚型, 使蛋白质的数目比编码蛋白质的基因数目多几个数量级^[10]。除此以外, 某些线性短模体被修饰后可以获得新的伴侣蛋白, 进一步丰富蛋白质相互作用的数量^[11]。Stumpf 等^[12]根据节点取样的方法预测人类细胞内存

* 国家自然科学基金 (81570129, 81430083, 21372036, 21402022, 21502015) 和中央高校基本科研业务费专项资金资助项目。

** 通讯联系人。

Tel: 13942696903, E-mail: zczhang@dlut.edu.cn

收稿日期: 2016-10-31, 接受日期: 2017-01-18

在大约 650 000 个蛋白质相互作用对. 虽然这个结果已经远远高出传统方法所预测的相互作用对的数量, 但是随着越来越多的线性短模体被发现和确认, 由线性短模体介导的 PPIs 的数量也会不断增加.

1.2 线性短模体的结构特点

1.2.1 密集分布与结构柔性

相比结构域和无序结构域, 线性短模体在蛋白质中的分布密度更大, 相互作用的效率也更高. 结构域的序列长度大多在 100 个氨基酸残基左右^[13], 无序结构域的长度在 20~50 个氨基酸左右^[14], 而绝大多数的线性短模体的序列长度仅有 3~10 个氨基酸残基, 平均只有 3~4 个氨基酸残基参与 PPIs^[14]. 同一个蛋白质中甚至可以包含多个序列重叠或位置邻近的线性短模体. 对现有的线性短模体进行统计和数据分析发现, 至少有 5.4% 的线性短模体与另一个线性短模体在序列上部分重叠, 14.7% 与另一个线性短模体在序列上的距离小于 10 个氨基酸残基^[14]. 例如, 在 MYC 蛋白的₅₅LLPTPPLSPSP₆₅ 10 个氨基酸残基组成的序列中, 就包含了 2 个序列重叠但是功能不同的线性短模体. 其中线性短模体₅₉PPLSPSP₆₅ 可以被 MAPK 激酶识别, 导致 MYC 蛋白的磷酸化^[15], 线性短模体₅₅LLPTPPLS₆₂ 可以被泛素连接酶 FBW7 识别, 促使 MYC 蛋白降解^[16](图 1).



Fig. 1 Overlapping SLiMs in MYC and their functions

图 1 MYC 蛋白中序列重叠的 SLiMs 及其功能

线性短模体除了分布密度高之外, 还具有结构柔韧性和适应性. 线性短模体主要存在于 IDPs 和 IDRs^[17], 在分子识别过程中可以发挥分子变阻器的功能, 即无序构象在结合时会转变为与伴侣蛋白结构更为互补的新构象^[18], 既可能获得良好折叠的构象, 又可能仍然是无规则构象. 例如, 位于 p53 蛋白 C 端的线性短模体在与钙结合蛋白同型二聚体 S100ββ 结合时, 折叠成 α 螺旋^[19], 与 sirtuin 蛋白结合时则形成 β 折叠^[20], 与细胞周期蛋白(cyclin A)、CBP 蛋白结合时, 则分别形成不同的无规则结构^[21-22] (图 2). 这种结构柔性是线性短模体丰富蛋白质相互作用形式和种类的结构基础^[23], 使得同一个线性短模体结合不同的伴侣蛋白, 介导多种 PPIs.

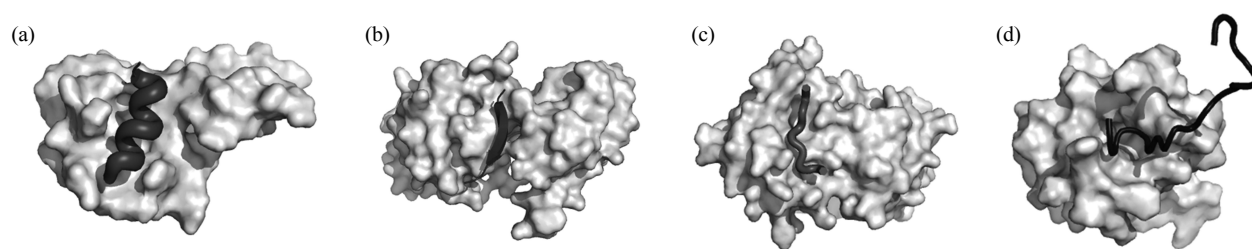


Fig. 2 Different binding modes of p53 C-terminus with different partners

图 2 p53 蛋白 C 端与不同蛋白的结合模式

p53 与 S100ββ(a), sirtuin(b), cyclin A(c)和 CBP 蛋白(d)复合物的晶体结构.

1.2.2 杂泛性

结构域介导的 PPIs 具有高度的特异性, 即只有一个或者少数几个伴侣蛋白. 线性短模体在与伴侣蛋白结合时则表现出程度不同的特异性, 甚至杂泛性 (promiscuity), 又称为混杂性, 也就是具有多个同源和非同源的伴侣蛋白. 例如: BH3 线性短模

体 LXXXXD 介导了 Bcl-2 家族蛋白网络的信号转导^[24](图 3). BH3-only 蛋白, 例如 Bim、Puma 蛋白通过 BH3 线性短模体与多种 Bcl-2 抗凋亡蛋白的结合^[25]. 非 Bcl-2 家族蛋白如载脂蛋白 Apol 6^[26]和自噬相关蛋白 ATG 12^[27]也因为包含 BH3 线性短模体, 与 Bcl-2 家族蛋白发生相互作用. 另外, 一些

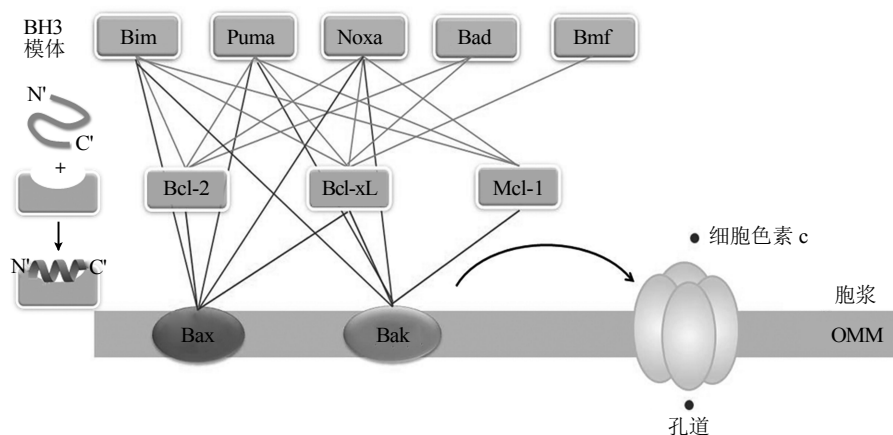


Fig. 3 BH3 motif mediates signaling network of interactions among the proteins of Bcl-2 family

图3 BH3 模体介导 Bcl-2 家族蛋白网络信号转导

病毒蛋白与 Bcl-2 家族蛋白虽然在一级序列上几乎没有相似性,但是具有与 Bcl-2 家族蛋白相似的三维结构^[28],例如牛痘病毒蛋白 N1^[29].它们也能利用 BH3 线性短模体的杂泛性与宿主细胞的 BH3-only 蛋白结合,抑制宿主细胞的凋亡.

线性短模体的杂泛性主要由内源因素和外源因素决定.内源因素指的是氨基酸的组成和序列.利用目前容量最大的线性短模体数据库——真核生物线性模体数据库(ELM database, <http://elm.eu.org>)进行数据挖掘研究,以蛋白质突变实验和直系同源蛋白序列比对为基础,结合 X 射线衍射晶体结构数据,Davey 等^[14]将线性短模体序列中的氨基酸残基分为如下几类:a.固定位置(fixed positions),处在这种位置的残基对结合能贡献较大,因此突变成任何残基都将导致亲和力的显著下降甚至完全丧失.这种位置的氨基酸残基通常是保守的.b.退化位置(degenerate positions),只有有限的几种结构或者化学性质相似的氨基酸残基之间的突变不会影响亲和力,例如丝氨酸与苏氨酸、赖氨酸与精氨酸、天冬氨酸与谷氨酸的相互替换.c.通用位置(wildcard positions),突变成任何氨基酸残基都不会对亲和力造成影响.d.禁止位置(prohibited positions),只有突变成特定氨基酸残基才会导致亲和力的下降,比如在某些 PPIs 中,侧链较大的氨基酸会产生空间位阻妨碍结合,或某些位置替换为脯氨酸会破坏对结合起关键作用的 α 螺旋的形成.以 BH3 线性短模体 LXXXXD 为例,亮氨酸残

基(L)和天冬氨酸残基(D)处在该线性短模体的固定位置,它们的突变会影响 Bim 与多个 Bcl-2 同源蛋白的结合^[30-32],这种氨基酸残基又被称为“hot spot”,是小分子 BH3 功能模拟物的分子设计中首要考虑的模拟对象.除此以外,BH3 线性短模体两侧的某些氨基酸残基,比如 G66E 的突变使 Bim 与所有 Bcl-2 同源蛋白的亲和力都大幅下降,所以 G66 处于禁止位置,其他位置氨基酸残基的突变并未对结合表现出较大的影响,这些位置就是通用位置^[30-32].将线性短模体序列中的氨基酸残基归纳成以上 4 种不同类型的位置,有助于我们初步判断一个线性短模体是否具备与目标蛋白结合的结构基础,或者发现可能与目标蛋白发生相互作用的新蛋白.

线性短模体与无序结构域的功能差异主要在于,线性短模体的固定位置直接决定其与伴侣蛋白的结合,无序结构域的功能则不仅仅由固定位置决定,与固定位置在一级序列上不连续甚至相距甚远的氨基酸残基也参与调控.例如,p27 蛋白的无序结构域必须通过保守序列 LFG 和与之相距较远的⁸⁸Y 的共同作用才能与 CDK 激酶结合并抑制 CDK 激酶的功能,⁸⁸Y 的磷酸化会直接导致这种功能的丧失^[33].

值得注意的是,与线性短模体发生相互作用的结构域并非是完全刚性的,其结构和构象也可能发生一些变化.例如,Mcl-1 单体蛋白的三维溶液结构显示,它的 α 4 螺旋显著地向 α 3 方向弯曲,导

致相互作用界面变小, 而 Mcl-1 与 BH3-only 蛋白 Bid 的 BH3 肽段结合的三维溶液结构显示, Mcl-1 的 $\alpha 4$ 沿螺旋长轴伸展, 蛋白质相互作用界面相应地变大, 容纳了 BH3-only 蛋白的疏水残基(图 4)。这种伴侣蛋白的结构柔性可能会干扰对线性短模体的固定位置氨基酸残基的确认。例如, Bcl-xL 蛋白与 Bim BH3 肽段的共结晶结构显示, Bim 的 F69 与 Bcl-xL 的疏水口袋发生相互作用, F69A 突变会导致 Bim 失去对 Bcl-xL 的亲和力, F69 被认为是 BH3 模体结合 Bcl-xL 的固定位置^[30]。然而, 通过 Split-Luciferase Reassembly 方法突变 Bcl-xL 蛋白上相应的疏水口袋的氨基酸残基, 却不影响 Bcl-xL 与 Bim 的结合亲和力^[34], 这又提示 F69 对于 Bcl-xL 并不是一个固定位置。这种相悖的结果可能源于 Bcl-xL 的结构域在与 Bim 结合时表现出一定的柔性。

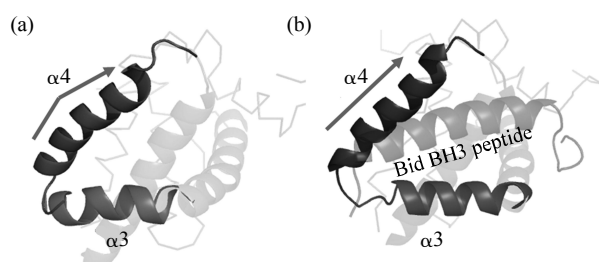


Fig. 4 NMR structure of Mcl-1(PDB ID: 2MHS)(a) and Mcl-1 complexed with Bid BH3 peptide(PDB ID: 2KBW)(b)
图 4 Mcl-1 单体蛋白 (PDB ID: 2MHS)(a)及 Mcl-1 蛋白与 Bid BH3 肽段结合(PDB ID: 2KBW)(b)的三维核磁结构

内源因素决定了线性短模体的杂泛性。外源因素, 例如细胞内环境, 则在一定程度上限制了这种杂泛性的发生。在蛋白质的亚细胞定位、细胞周期、组织特异性等影响下, 在体外具有杂泛性的线性短模体, 在细胞内往往由于和其他结构域或者线性短模体存在竞争关系、空间位阻、复合物组分之间的协同作用等原因, 仅能够和少数伴侣蛋白发生相互作用, 从而在一定程度上避免“脱靶”效应对细胞功能的干扰^[35]。

但是, 保证线性短模体在准确识别多个伴侣蛋白的同时降低“脱靶”效应的结构基础和外源因素, 仍然值得研究。

2 线性短模体的分类与功能

2.1 配体模体

蛋白质利用配体模体发生相互作用时不会发生化学反应, 在该相互作用被破坏后, 配体模体和伴侣蛋白也不会发生化学性质或结构上的变化。就其功能的不同, 可以将配体模体分为结合模体、降解模体、定位模体等。

结合模体在蛋白质相互作用中起到增强亲和力的作用。例如 p53 蛋白的线性短模体 KKLMF 能够与 cyclin A 结合, cyclin A 能够与 CDK 激酶形成复合物, 因此促使 CDK 激酶磷酸化 p53 蛋白。缺失这个模体的 p53 蛋白则不能结合 cyclin A, 被 CDK 激酶磷酸化的程度也随之大大降低^[36]。另外, 包含相同模体的蛋白可以竞争结合同一个伴侣蛋白, 例如 p21 和 CDC25A 都包含可结合 cyclin 的模体, 因此它们可以竞争结合 CDK 激酶, p21 发挥抑制细胞周期的功能, 而 CDC25A 则促进 $G_1 \rightarrow S$ 细胞周期的进展。这种竞争结合的平衡起到调控细胞周期的作用^[37]。

降解模体是指能够被 E3 连接酶识别的线性短模体。在 E3 连接酶的介导下, 蛋白质发生泛素化被蛋白酶体识别并降解^[38-39]。

定位模体在被细胞中的运输蛋白识别后, 可以将靶蛋白转运至某一特定细胞器。例如 p53 的核定位模体 $_{305}KRALPNNTSSSPQPKKKPL_{323}$ 和核输出模体 $_{339}EMFRELNEALELKD_{352}$, 可以分别被核输入蛋白 importin-alpha 和核输出蛋白 CRM1 识别, 赋予 p53 在细胞核与胞质之间穿梭往返的功能^[40-41]。

2.2 翻译后修饰模体

翻译后修饰模体中包含酶的修饰位点。蛋白质在该位点上被酶修饰后, 发生化学性质或结构上的变化, 蛋白质功能也发生相应的变化。在众多翻译后修饰过程中, 激酶的磷酸化作用非常重要, 常见于细胞周期调控和细胞信号传导等过程^[42]。例如在细胞周期调控因子 CDC25B 的翻译后修饰模体 $_{320}RSPSMP_{325}$ 中, S321 被 CDK1 磷酸化后, 会阻碍临近位置上 S323 的磷酸化, 从而抑制 14-3-3 与 CDC25B 结合, 维持 CDC25B 的活性^[43]。

裂解模体是一种特殊的翻译后修饰模体, 在被细胞中的蛋白酶识别后, 蛋白在该模体中的特定位点上被切割。Caspase 3 和 Caspase 7 是细胞凋亡下游通路的重要效应因子, 在被活化后可以识别多种蛋白的裂解模体, 并在一个固定位置天冬氨酸残基

上切割底物, 推进细胞凋亡的过程^[44-46].

3 线性短模体的协作机制与系统调节性机制

在各种生命过程中, 大量蛋白质通过多组线性短模体发生复杂的相互作用, 整合相关信号, 构建多种独特的信号转导网络. 各种生物学功能的实现既可以通过同一个蛋白质所包含的不同线性短模体来完成, 称之为协作机制(cooperative strategy), 也可以由不同蛋白质的多个线性短模体来完成, 称为系统性调节机制(systemic strategy).

3.1 线性短模体的协作机制

由于线性短模体具有密集分布的特点, 一个蛋白质可能包含多个功能完全不同的模体. 这些线性短模体之间的不同组合和相互作用方式赋予了蛋白质不同的活性和功能. 以 p21^{Cip1} 蛋白为例, 这个天然无序蛋白的 C 端包含 2 个线性短模体, 一个结合细胞周期蛋白, 抑制其结合多种细胞周期调控因子例如 CDC25A, 另一个结合 CDK 激酶, 使其无法激活细胞周期调控因子, 这种协作机制可以引起细胞周期的阻滞^[47].

3.2 线性短模体的系统性调节机制

系统性调节机制导致线性短模体调控的信号网络具有动态性、条件依赖性和可逆性. 这些特点在 T 细胞信号通路、缺氧信号通路的调控中都有明确的实验结论. 以 LCK 激酶的信号转导机制为例, 在非刺激条件下, CSK 激酶将 LCK 激酶 C 端的线性短模体₅₀₄QYQP₅₀₇ 磷酸化, 使其与 LCK 自身的 SH2 结构域结合, 将 LCK 激酶功能区封锁住, 呈现自抑制的状态, 无法激活 T 淋巴细胞受体^[48]. 另外, LCK 激酶和 PAG1 蛋白都具有包含半胱氨酸的线性短模体, 半胱氨酸可以与细胞膜上的十六烷酰结合, 使两个蛋白共定位在细胞膜上. 与此同时, 磷酸化的 PAG1 蛋白与 CSK 激酶结合, 导致 PAG1 向 LCK 激酶富集 CSK 激酶, 进一步强化 CSK 激酶对 LCK 激酶的磷酸化功能, 抑制 T 淋巴细胞受体的激活^[49]. 然而, 在细胞接受刺激条件下, 磷酸酶 CD45 会将 LCK 激酶的线性短模体₅₀₄QpYQP₅₀₇ 去磷酸化, 使其无法与 SH2 结构域结合, LCK 处于激活状态^[50], 促进 T 淋巴细胞受体的激活.

通过 ELM 数据库, 我们还可以获得线性短模体介导的相互作用网络信息. 例如, ELM 数据库可以被整合到 KEGG(Kyoto Encyclopedia of Genes and Genomes)数据库中^[51]. 将由线性短模体介导的

PPIs 加以标记, 描绘相关的信号通路会帮助我们在系统水平上研究线性短模体介导的不同的信号转导通路, 以及通路之间的交互作用(cross-talk), 扩展我们对线性短模体的理解.

4 线性短模体的进化

4.1 线性短模体具有从头进化的特征

不同于结构域和无序结构域, 线性短模体具有从头进化的特征, 即一个或几个氨基酸残基的突变就使一段本身没有功能的氨基酸序列进化为一个有特定功能的线性短模体. 类似的现象在基因进化研究中早有报道, 非编码 DNA 序列可以通过突变从头进化出编码基因^[52]. 线性短模体序列短, 介导 PPIs 的关键氨基酸残基数量少, 容易通过突变赋予蛋白新的功能. 这或许就解释了非同源蛋白进化出具有相同功能的线性短模体, 即趋同进化的现象. 据统计, 在 ELM 数据库中有 50% 以上的蛋白质含有趋同进化的线性短模体, 超过 10% 的蛋白质含有 10 个以上趋同进化的线性短模体^[14]. 这恰恰证明了线性短模体在进化上的这种特点. 另外, 通过对不同物种的氨基酸序列对比, 可以确定从头进化的线性短模体, 例如结合 SH3 域的模体 PXXP 只存在于哺乳动物的接头蛋白 CRK 中, 在蛙、鱼、蠕虫的同源蛋白中均未发现这个线性短模体, 说明这个线性短模体是通过从头进化获得的^[4].

4.2 线性短模体具有保守性低的进化特征

一个位点的突变就可以产生一个具有新功能的线性短模体, 一个位点的突变同样可以破坏原有线性短模体功能. 因此, 线性短模体的进化保守性低. 在 ELM 数据库中, 人与酵母之间的线性短模体的保守率只有 3.7%^[14]. 线性短模体在进化上的不保守性会导致蛋白质获得新的结合伴侣或者失去原有的结合伴侣, 乃至引起蛋白质相互作用网络的变化, 为进化提供了更多的选择. 不同类型的线性短模体的保守性也不尽相同. 结合模体的进化保守性相对最强, 序列和功能都具有一定的保守性; 定位模体虽然序列的保守性差, 但其功能仍可维持; 翻译后修饰模体的序列和功能的保守性都较差. 这反映了结合模体和定位模体的功能对不同的物种都具有较高的必要性^[14].

5 新的线性短模体: BH3 线性短模体

5.1 BH3 模体介导 Bcl-2 家族蛋白质的相互作用

Bcl-2 家族蛋白质介导细胞的线粒体凋亡通

路. 该家族蛋白质可以分为3类: 抗凋亡蛋白, 包括 Bcl-2、Bcl-xL、Mcl-1 等, 具有4个 Bcl-2 同源序列(BH1~BH4); 促凋亡蛋白 Bax、Bak 等, 具有3个同源序列(BH1~BH3); BH3-only 蛋白如 Bim、Bad、Bid、Noxa 等, 只包含 BH3 同源序列. 它们之间的相互作用通过共有的 BH3 motif 完成^[53-54].

在 Bcl-2 家族网络中, BH3-only 蛋白发挥着重要作用, 它们像哨兵一样将凋亡信号传递给效应因子 Bax 和 Bak, 既可以通过与抗凋亡蛋白结合, 释放 Bax 和 Bak, 又能直接激活 Bax 和 Bak, 促进 Bax 和 Bak 在线粒体膜上多聚并打孔, 释放细胞色素 C, 导致细胞凋亡^[53-54](图 3).

5.2 BH3 模体是线性短模体

随着研究的推进, BH3 模体逐渐呈现出很多线性短模体的特点, 在 2016 年, 它被正式定义为 BH3 线性短模体^[24]. 首先, 在目前已经发现的 8 个 BH3-only 蛋白中, 有 7 个是天然无序蛋白^[55], 它们的 BH3 motif 都是无序的, 只有在与抗凋亡蛋白结合时, 才折叠成 α 螺旋. 第二, BH3 模体包含 7~15 个氨基酸残基, 只有 4~6 个残基对结合起到重要作用, 包括对结合贡献较大的 2 个固定位置, 被称为“hot spot”的亮氨酸残基(L)与天冬氨酸残基(D)^[30-32]. 因此 BH3 线性短模体被简化为 LXXXXD. 第三, BH3 模体可能在 1 个蛋白质中多次出现, 例如, 鼠源 Noxa 就包含 2 个 BH3 线性短模体^[56], 表现出较高的分布密度. 第四, 部分 BH3-only 蛋白以“hit and run”的机制与 Bax/Bak 瞬时结合, 激活 Bax 和 Bak^[57], 这种瞬时作用和动态模式与线性短模体非常相似. 第五, 从进化的角度看, 从原核生物到酵母, 再到哺乳动物几乎全部生命形式都包含 BH3 线性短模体, BH3 线性短模体不但存在于 Bcl-2 家族蛋白中, 还在多种 Bcl-2 家族以外的非同源蛋白质中出现, 例如最新发现的 BH3-only 蛋白 E3 连接酶 Mule 和包含 BH3 模体的载脂蛋白 Apol 6, 说明这个线性短模体很有可能是通过趋同进化得来的. 第六, BH3 模体的功能受可变剪切的影响, 例如 Bax 的 BH3 模体在一些剪切变体中会缺失^[58]. 最后, BH3 同源序列中包含较多的丝氨酸残基, 所以磷酸化状态也会影响其功能. 例如 Bad 中 S112 或 S136 的磷酸化, 就可以促使 Bad 与 14-3-3 蛋白结合, 抑制其与 Bcl-2 或 Bcl-xL 结合并介导细胞凋亡^[59]. 综合这些特点, BH3 模体被定义为线性短模体.

和 SH3 结合模体 PXXP^[60]、内质网定位模体 KDEL^[61]等相似, BH3 模体 LXXXXD 的序列也不具有预测意义. 由于 BH3 模体在 8 个经典的 BH3-only 蛋白中具有保守性, 而在新发现的 Bcl-2 家族以外的非经典 BH3-only 蛋白中保守性却很低, 这是 BH3 模体不能为众多线性短模体预测软件准确识别的原因之一.

抑癌蛋白 p53 也可以像 BH3-only 蛋白一样与 Bcl-xL 结合, 并且其结合模式与 BH3-only 蛋白非常相似, 一些研究怀疑 p53 的这一段与 Bcl-xL 结合的序列 DLWKLL 是一个 BH3 线性短模体, 因为这一序列的顺序与已知的 BH3 线性短模体 LXXXXD 的顺序正好相反^[62]. 本课题组 2016 年发表的小分子骨架, 实现了对 p53 和 BH3-only 蛋白的 BH3 模体功能的双模拟^[63], 从功能上将 p53 的模体 DLWKLL 归入了 BH3 线性短模体. 这种情况给 BH3 线性短模体的预测和识别增加了难度.

5.3 BH3 线性短模体的特殊性

除了上述符合线性短模体的诸多特点, BH3 模体还表现出一些特殊性, 这也是 BH3 模体至今才被定义为线性短模体的原因. 首先, 不同于大多数线性短模体以无序状态存在于 IDP 中, BH3 模体还能够以 α 螺旋的有序构象存在于抗凋亡蛋白、促凋亡蛋白和 Bid 蛋白中. 其次, 线性短模体与伴侣蛋白的亲合力一般较低 ($>1 \mu\text{mol/L}$)^[4], 而大多数 BH3 模体与伴侣结合的亲合力都可以达到 nmol/L 级别^[64]. 这种高亲合力可能来源于 BH3 线性短模体之外的氨基酸残基对亲和力的贡献. 第三, 大部分已知的线性短模体均暴露于蛋白表面, 但是部分 BH3 motif 隐藏在结构域中, 只能有条件地暴露在蛋白质相互作用界面上. 例如在胞质中 Bax、Bak 的 BH3 motif 疏水结合面是掩埋在疏水沟槽中的, 只有通过激活机制, 致使 Bax、Bak 发生构象变化, 才能使 BH3 motif 暴露出来并发挥功能^[65-66].

6 前景与挑战

线性短模体的磷酸化修饰赋予了蛋白质功能多样化的特点, 可变剪切进一步增加了功能不同的蛋白质亚型的数量, 从头进化则扩展了蛋白质新的功能模块. 线性短模体在介导 PPIs 时表现的结构柔性和杂泛性大大丰富了蛋白质相互作用组的研究内容. 对于线性短模体的研究在以下几个方面具有重要意义, 并面临巨大挑战:

第一, 发现和鉴定新的线性短模体, 将对蛋白

质功能研究的相关领域做出贡献。由于线性短模体具有趋同进化的特点, 非同源蛋白可以适应性进化出相同的线性短模体, 并行使相同或相似的功能。这显然就为发现已知蛋白的新功能提供了便利。尤其是发现和鉴定那些进化保守性差, 但是在不同家族甚至不同物种中具有功能相关性的线性短模体, 将大大提高我们发现未知蛋白质和已知蛋白质新功能的效率。但是, 由于很多线性短模体结合亲和力较低, 而且为瞬时结合, 它们介导的 PPIs 无法被酵母双杂交、噬菌体展示等方法所发现^[67], 我们需要构建新的、高效的检测方法来发现和鉴定线性短模体。另外, 为了发现更多的线性短模体, 许多预测线性短模体的算法和软件应运而生, 例如 QSLiMFinder^[68]、SLiMFinder^[69]、SLiMPrints^[70]、Minimotif Miner^[6]等, 它们主要分为基于序列对比和非基于序列对比, 前者适用于预测属于同一家族蛋白中的线性短模体和已知功能的线性短模体, 后者更适用于预测非相关蛋白通过趋同进化获得的线性短模体。目前它们面临的共同困难是无法降低预测的假阳性。

第二, 发现和鉴定新的线性短模体, 将对疾病靶标确认领域做出贡献。以本课题组长期研究的 BH3 motif 为例, 随着其被定义为线性短模体, 它们在肿瘤信号转导通路中的作用和治疗价值也需要被重新评价。Bcl-2 家族以外的包含 BH3 模体的蛋白, 例如载脂蛋白 Apol 6^[26]、自噬相关蛋白 ATG 12^[27]和鞘氨醇激酶 SPHK 2^[71], 可能因为结构柔性、杂泛性的特点, 具有未知的伴侣蛋白, 介导未知的信号通路, 与已知的 Bcl-2 信号转导通路发生交互作用(cross-talk), 共同构建新的肿瘤相关的蛋白质相互作用网络。现有的小分子 BH3 功能模拟物(BH3 mimetics), 作为化学遗传学的工具, 将为这方面研究贡献更直接的结果。另外, BH3 功能模拟物也将作为分子探针, 在相对短的时间内, 在比较大的范围内钓取新的包含 BH3 模体的蛋白, 并鉴定其在肿瘤信号转导通路中的意义, 发现新的肿瘤靶标。

第三, 发现新的线性短模体的功能特征, 将为药物分子设计提供新的思路。以小分子 BH3 功能模拟物的设计为例: BH3 模体的性质研究, 与 BH3 功能模拟物的设计和合成是相辅相成的。虽然我们课题组早在 BH3 模体被定义为线性短模体之前, 就通过构效关系等研究证明, 优先和侧重对于这一线性模体中固定位置氨基酸残基的理化性质

的模拟, 有利于设计出高效结合 Bcl-2-like 蛋白的小分子^[72-73]。这种设计思路与在 1.2.2 中提到的 BH3 motif 氨基酸残基突变实验的结果是一致的。但是, 目前我们课题组和其他课题组所设计和合成的模拟 BH3 模体固定位置的小分子, 大部分并不能像含有 BH3 模体的如 Bad、Noxa 蛋白一样, 对不同的 Bcl-2 家族蛋白表现出选择性和特异性, 或者对它们的构效关系研究并未发现对于固定位置的功能模拟与抗癌活性的直接关系。这提示我们在分子设计中, 模拟氨基酸残基功能的范围可能需要扩展到线性短模体两侧的某些序列, 或者更多考虑其空间构象等因素。这些问题的解决, 还有赖于针对 BH3 模体本身性质的进一步研究结果。遗憾的是, 对于处在蛋白质天然无序区的线性短模体, 其结构、功能和计算方法等研究都受限于晶体学和生物信息学的方法要求。动力学和热力学的研究方法或可有助于我们进一步了解其调控生理过程的分子机制。后续研究还需要通过一系列的高通量实验, 在生理环境下或者在合适的模型体系中, 确定线性短模体实现其功能的分子机制。

参 考 文 献

- [1] 黄永棋, 刘志荣. 天然无序蛋白质: 序列 - 结构 - 功能的新关系. 物理化学学报, 2010, **26**(8): 2061-2072
Huang Y Q, Liu Z R. Acta Physico-Chimica Sinica, 2010, **26**(8): 2061-2072
- [2] 马 冲, 杨 冬, 姜 颖, 等. 无序蛋白质的判定及其结构、功能和进化特征. 生物化学与生物物理进展, 2015, **42**(1): 16-24
Ma C, Yang D, Jiang Y, *et al.* Progress in Biochemistry and Biophysics, 2015, **42**(1): 16-24
- [3] Wright P E, Dyson H J. Intrinsically disordered proteins in cellular signalling and regulation. Nature Reviews Molecular Cell Biology, 2015, **16**(1): 18-29
- [4] Van Roey K, Uyar B, Weatheritt R J, *et al.* Short linear motifs: ubiquitous and functionally diverse protein interaction modules directing cell regulation. Chemical Reviews, 2014, **114**(13): 6733-6778
- [5] Dinkel H, Van Roey K, Michael S, *et al.* ELM 2016—data update and new functionality of the eukaryotic linear motif resource. Nucleic Acids Research, 2015: gkv1291
- [6] Balla S, Thapar V, Verma S, *et al.* Minimotif Miner: a tool for investigating protein function. Nature Methods, 2006, **3**(3): 175-177
- [7] Guharoy M, Pauwels K, Tompa P. SnapShot: Intrinsic structural disorder. Cell, 2015, **161**(5): 1230-1230. e1
- [8] Tompa P, Davey N E, Gibson T J, *et al.* A million peptide motifs for the molecular biologist. Molecular Cell, 2014, **55**(2): 161-169
- [9] Babu M M, van der Lee R, de Groot N S, *et al.* Intrinsically

- disordered proteins: regulation and disease. *Current Opinion in Structural Biology*, 2011, **21**(3): 432–440
- [10] Weatheritt R J, Gibson T J. Linear motifs: lost in (pre) translation. *Trends in Biochemical Sciences*, 2012, **37**(8): 333–341
- [11] Deribe Y L, Pawson T, Dikic I. Post-translational modifications in signal integration. *Nature Structural & Molecular Biology*, 2010, **17**(6): 666–672
- [12] Stumpf M P H, Thorne T, de Silva E, *et al.* Estimating the size of the human interactome. *Proc Natl Acad Sci USA*, 2008, **105**(19): 6959–6964
- [13] Wheelan S J, Marchler-Bauer A, Bryant S H. Domain size distributions can predict domain boundaries. *Bioinformatics*, 2000, **16**(7): 613–618
- [14] Davey N E, Van Roey K, Weatheritt R J, *et al.* Attributes of short linear motifs. *Molecular BioSystems*, 2012, **8**(1): 268–281
- [15] Lutterbach B, Hann S R. Hierarchical phosphorylation at N-terminal transformation-sensitive sites in c-Myc protein is regulated by mitogens and in mitosis. *Molecular and Cellular Biology*, 1994, **14**(8): 5510–5522
- [16] Yada M, Hatakeyama S, Kamura T, *et al.* Phosphorylation-dependent degradation of c-Myc is mediated by the F-box protein Fbw7. *The EMBO Journal*, 2004, **23**(10): 2116–2125
- [17] Henderson B, Martin A C R. Protein moonlighting: a new factor in biology and medicine. *Biochemical Society Transactions*, 2014, **42**(6): 1671–1678
- [18] Boehr D D, Nussinov R, Wright P E. The role of dynamic conformational ensembles in biomolecular recognition. *Nature Chemical Biology*, 2009, **5**(11): 789–796
- [19] Rust R R, Baldisseri D M, Weber D J. Structure of the negative regulatory domain of p53 bound to S100B ($\beta\beta$). *Nature Structural & Molecular Biology*, 2000, **7**(7): 570–574
- [20] Avalos J L, Celic I, Muhammad S, *et al.* Structure of a Sir2 enzyme bound to an acetylated p53 peptide. *Molecular Cell*, 2002, **10**(3): 523–535
- [21] Lowe E D, Tews I, Cheng K Y, *et al.* Specificity determinants of recruitment peptides bound to phospho-CDK2/cyclin A. *Biochemistry*, 2002, **41**(52): 15625–15634
- [22] Resnick-Silverman L, Yan S, Mutjaba S, *et al.* Target structure-based discovery of small molecules that block human p53 and CREB binding protein association. *Chemistry & Biology*, 2006, **13**(1): 81–90
- [23] Schreiber G, Keating A E. Protein binding specificity versus promiscuity. *Current Opinion in Structural Biology*, 2011, **21**(1): 50–61
- [24] Aouacheria A, Combet C, Tompa P, *et al.* Redefining the BH3 death domain as a ‘Short Linear Motif’. *Trends in Biochemical Sciences*, 2015, **40**(12): 736–748
- [25] Cory S, Roberts A W, Colman P M, *et al.* Targeting BCL-2-like proteins to kill cancer cells. *Trends in Cancer*, 2016, **2**(8): 443–460
- [26] Liu Z, Lu H, Jiang Z, *et al.* Apolipoprotein L6, a novel proapoptotic Bcl-2 homology 3-only protein, induces mitochondria-mediated apoptosis in cancer cells. *Molecular Cancer Research*, 2005, **3**(1): 21–31
- [27] Rubinstein A D, Eisenstein M, Ber Y, *et al.* The autophagy protein Atg12 associates with antiapoptotic Bcl-2 family members to promote mitochondrial apoptosis. *Molecular Cell*, 2011, **44**(5): 698–709
- [28] Kvansakul M, Hinds M G. Structural biology of the Bcl-2 family and its mimicry by viral proteins. *Cell death & disease*, 2013, **4**(11): e909
- [29] Cooray S, Bahar M W, Abrescia N G A, *et al.* Functional and structural studies of the vaccinia virus virulence factor N1 reveal a Bcl-2-like anti-apoptotic protein. *Journal of General Virology*, 2007, **88**(6): 1656–1666
- [30] Lee E F, Czabotar P E, Van Delft M F, *et al.* A novel BH3 ligand that selectively targets Mcl-1 reveals that apoptosis can proceed without Mcl-1 degradation. *The Journal of Cell Biology*, 2008, **180**(2): 341–355
- [31] Day C L, Smits C, Fan F C, *et al.* Structure of the BH3 domains from the p53-inducible BH3-only proteins Noxa and Puma in complex with Mcl-1. *Journal of Molecular Biology*, 2008, **380**(5): 958–971
- [32] Dutta S, Gullá S, Chen T S, *et al.* Determinants of BH3 binding specificity for Mcl-1 versus Bcl-x L. *Journal of Molecular Biology*, 2010, **398**(5): 747–762
- [33] Tompa P, Fuxreiter M, Oldfield C J, *et al.* Close encounters of the third kind: disordered domains and the interactions of proteins. *Bioessays*, 2009, **31**(3): 328–335
- [34] Campbell S T, Carlson K J, Buchholz C J, *et al.* Mapping the BH3 binding interface of Bcl-xL, Bcl-2, and Mcl-1 using split-luciferase reassembly. *Biochemistry*, 2015, **54**(16): 2632–2643
- [35] Dinkel H, Van Roey K, Michael S, *et al.* The eukaryotic linear motif resource ELM: 10 years and counting. *Nucleic Acids Research*, 2013: gkt1047
- [36] Mitrea D M, Yoon M K, Ou L, *et al.* Disorder-function relationships for the cell cycle regulatory proteins p21 and p27. *Biological Chemistry*, 2012, **393**(4): 259–274
- [37] Saha P, Eichbaum Q, Silberman E D, *et al.* p21CIP1 and Cdc25A: competition between an inhibitor and an activator of cyclin-dependent kinases. *Molecular and Cellular Biology*, 1997, **17**(8): 4338–4345.
- [38] Tian W, Li B, Warrington R, *et al.* Structural analysis of human Cdc20 supports multisite degron recognition by APC/C. *Proc Natl Acad Sci USA*, 2012, **109**(45): 18419–18424
- [39] Wei W, Ayad N G, Wan Y, *et al.* Degradation of the SCF component Skp2 in cell-cycle phase G1 by the anaphase-promoting complex. *Nature*, 2004, **428**(6979): 194–198
- [40] Liang S H, Clarke M F. Regulation of p53 localization. *European Journal of Biochemistry*, 2001, **268**(10): 2779–2783
- [41] O’Keefe K, Li H, Zhang Y. Nucleocytoplasmic shuttling of p53 is essential for MDM2-mediated cytoplasmic degradation but not ubiquitination. *Molecular and Cellular Biology*, 2003, **23**(18): 6396–6405
- [42] Manning G, Whyte D B, Martinez R, *et al.* The protein kinase

- complement of the human genome. *Science*, 2002, **298** (5600): 1912–1934
- [43] Astuti P, Boutros R, Ducommun B, *et al.* Mitotic phosphorylation of Cdc25B Ser321 disrupts 14-3-3 binding to the high affinity Ser323 site. *Journal of Biological Chemistry*, 2010, **285** (45): 34364–34370
- [44] Fattman C L, Delach S M, Dou Q P, *et al.* Sequential two-step cleavage of the retinoblastoma protein by caspase-3/-7 during etoposide-induced apoptosis. *Oncogene*, 2001, **20**(23)
- [45] Ellerby L M, Andrusiak R L, Wellington C L, *et al.* Cleavage of atrophin-1 at caspase site aspartic acid 109 modulates cytotoxicity. *Journal of Biological Chemistry*, 1999, **274**(13): 8730–8736
- [46] Datta R, Kojima H, Yoshida K, *et al.* Caspase-3-mediated cleavage of protein kinase C θ in induction of apoptosis. *Journal of Biological Chemistry*, 1997, **272**(33): 20317–20320
- [47] Mitrea D M, Yoon M K, Ou L, *et al.* Disorder-function relationships for the cell cycle regulatory proteins p21 and p27. *Biological Chemistry*, 2012, **393**(4): 259–274
- [48] Yamaguchi H, Hendrickson W A. Structural basis for activation of human lymphocyte kinase Lck upon tyrosine phosphorylation. *Nature*, 1996, **384**(6608): 484–489
- [49] Brdicka T, Pavlišťová D, Leo A, *et al.* Phosphoprotein associated with glycosphingolipid-enriched microdomains (PAG), a novel ubiquitously expressed transmembrane adaptor protein, binds the protein tyrosine kinase csk and is involved in regulation of T cell activation. *The Journal of Experimental Medicine*, 2000, **191** (9): 1591–1604
- [50] Sheno H, Seavitt J, Zheleznyak A, *et al.* Regulation of integrin-mediated T cell adhesion by the transmembrane protein tyrosine phosphatase CD45. *The Journal of Immunology*, 1999, **162** (12): 7120–7127
- [51] Gibson T J, Dinkel H, Roey K, *et al.* Experimental detection of short regulatory motifs in eukaryotic proteins: tips for good practice as well as for bad. *Cell Communication and Signaling*, 2015, **13**(1): 42
- [52] McLysaght A, Guerzoni D. New genes from non-coding sequence: the role of de novo protein-coding genes in eukaryotic evolutionary innovation. *Phil Trans R Soc B*, 2015, **370**(1678): 20140332
- [53] Czabotar P E, Lessene G, Strasser A, *et al.* Control of apoptosis by the BCL-2 protein family: implications for physiology and therapy. *Nature Reviews Molecular Cell Biology*, 2014, **15**(1): 49–63
- [54] Moldoveanu T, Follis A V, Kriwacki R W, *et al.* Many players in BCL-2 family affairs. *Trends in Biochemical Sciences*, 2014, **39**(3): 101–111
- [55] Delbridge A R D, Strasser A. The BCL-2 protein family, BH3-mimetics and cancer therapy. *Cell Death & Differentiation*, 2015, **22**(7): 1071–1080
- [56] Weber A, Ausländer D, Häcker G. Mouse Noxa uses only the C-terminal BH3-domain to inactivate Mcl-1. *Apoptosis*, 2013, **18**(9): 1093–1105
- [57] Wei M C, Lindsten T, Mootha V K, *et al.* tBID, a membrane-targeted death ligand, oligomerizes BAK to release cytochrome c. *Genes & Development*, 2000, **14**(16): 2060–2071
- [58] Akgul C, Moulding D A, Edwards S W. Alternative splicing of Bcl-2-related genes: functional consequences and potential therapeutic applications. *Cellular and Molecular Life Sciences CMLS*, 2004, **61**(17): 2189–2199
- [59] Zha J, Harada H, Yang E, *et al.* Serine phosphorylation of death agonist BAD in response to survival factor results in binding to 14-3-3 not BCL-X L. *Cell*, 1996, **87**(4): 619–628
- [60] Chandra B R, Gowthaman R, Akhouri R R, *et al.* Distribution of proline-rich (PxxP) motifs in distinct proteomes: functional and therapeutic implications for malaria and tuberculosis. *Protein Engineering Design and Selection*, 2004, **17**(2): 175–182
- [61] Raykhel I, Alanen H, Salo K, *et al.* A molecular specificity code for the three mammalian KDEL receptors. *The Journal of Cell Biology*, 2007, **179**(6): 1193–1204
- [62] Xu H, Ye H, Osman N E, *et al.* The MDM2-binding region in the transactivation domain of p53 also acts as a Bcl-XL-binding motif. *Biochemistry*, 2009, **48**(51): 12159–12168
- [63] Wang Z, Song T, Feng Y, *et al.* Bcl-2/MDM2 dual inhibitors based on universal pyramid-like α -helical mimetics. *Journal of Medicinal Chemistry*, 2016, **59**(7): 3152–3162
- [64] Chen L, Willis S N, Wei A, *et al.* Differential targeting of prosurvival Bcl-2 proteins by their BH3-only ligands allows complementary apoptotic function. *Molecular Cell*, 2005, **17** (3): 393–403
- [65] Dewson G, Kratina T, Sim H W, *et al.* To trigger apoptosis, Bak exposes its BH3 domain and homodimerizes *via* BH3: groove interactions. *Molecular Cell*, 2008, **30**(3): 369–380
- [66] Gavathiotis E, Reyna D E, Davis M L, *et al.* BH3-triggered structural reorganization drives the activation of proapoptotic BAX. *Molecular Cell*, 2010, **40**(3): 481–492
- [67] Blikstad C, Ivarsson Y. High-throughput methods for identification of protein-protein interactions involving short linear motifs. *Cell Communication and Signaling*, 2015, **13**(1): 38
- [68] Palopoli N, Lythgow K T, Edwards R J. QSLiMfinder: improved short linear motif prediction using specific query protein data. *Bioinformatics*, 2015, **31**(14): 2284–2293
- [69] Davey N E, Haslam N J, Shields D C, *et al.* SLiMfinder: a web server to find novel, significantly over-represented, short protein motifs. *Nucleic Acids Research*, 2010: gkq440
- [70] Davey N E, Cowan J L, Shields D C, *et al.* SLiMprints: conservation-based discovery of functional motif fingerprints in intrinsically disordered protein regions. *Nucleic Acids Research*, 2012, **40**(21): 10628–10641
- [71] Igarashi N, Okada T, Hayashi S, *et al.* Sphingosine kinase 2 is a nuclear protein and inhibits DNA synthesis. *Journal of Biological Chemistry*, 2003, **278**(47): 46832–46839
- [72] Zhang Z, Wu G, Xie F, *et al.* 3-Thiomorpholin-8-oxo-8 H-acenaphtho [1, 2-b] pyrrole-9-carbonitrile (S1) based molecules as potent, dual inhibitors of B-cell lymphoma 2 (Bcl-2) and

myeloid cell leukemia sequence 1 (Mcl-1): structure-based design and structure-activity relationship studies. *Journal of Medicinal Chemistry*, 2011, **54**(4): 1101–1105

[73] Zhang Z, Li X, Song T, *et al.* An anthraquinone scaffold for putative, two-face Bim BH3 α -helix mimic. *Journal of Medicinal Chemistry*, 2012, **55**(23): 10735–10741

Short Linear Motifs (SLiMs): New Functionally Diverse Modules Regulating Protein-protein Interactions*

LIU Lu, ZHANG Zhi-Chao**

(State Key Laboratory of Fine Chemicals, School of Chemistry, Dalian University of Technology, Dalian 116030, China)

Abstract Short linear motifs (SLiMs) are important interaction modules of intrinsically disordered proteins. Characterized with structural flexibility and short sequence, SLiMs are promiscuous and mediate transient, reversible protein-protein interactions. With the advancement of experimental measures and motif-search tools, an increasing number of short linear motifs are being discovered. The BH3 domain of the Bcl-2 family, for example, was recently redefined as a short linear motif. Here, we review the characteristics of SLiMs in structure, function and evolution. Subsequent studies about their functions will shed new light on cell signaling networks researches, therapeutic targets identification and drug discovery.

Key words short linear motif, protein-protein interaction, intrinsically disordered protein, evolution

DOI: 10.16476/j.pibb.2016.0245

* This work was supported by grants from The National Natural Science Foundation of China (81570129, 81430083, 21372036, 21402022, 21502015) and the Fundamental Research Funds for the Central Universities.

**Corresponding author.

Tel: 86-13942696903, E-mail: zczhang@dlut.edu.cn

Received: October 31, 2016 Accepted: January 18, 2017