聚酮聚肽类天然产物生物合成过程中“多余”结构单元的引入、去除及其生物学意义

陈明, 陈华, 刘文*

中国科学院上海有机化学研究所, 生命有机化学国家重点实验室, 上海 200032

* 联系人, E-mail: wliu@mail.sioc.ac.cn

2017-02-14 收稿, 2017-05-03 修回, 2017-05-04 接受, 2017-06-28 网络版发表

国家自然科学基金(21472231, 31430005, 21520102004, 81302674)、上海市科学技术委员会重点项目(14JC1407700, 15JC1400400)、中国科学院战略性先导科技专项(SQYZDJ-SSW-SLH1037, XDB20020200)和 K. C. Wong 教育基金资助

摘要 来源于动植物或微生物的天然产物, 特别是聚酮类(polyketides, PKs)、非核糖体聚肽类(non-ribosomal peptides, NRPs)及其杂合化合物, 由于其广泛而良好的生物活性而在现代医药产业中占有重要地位, 引起人们极大的研究兴趣. 近年来, 在PKs, NRPs及其杂合类化合物的生物合成研究中, 一些结构单元先被引入, 再在后续反应中被去除的现象在某些分子中陆续被发现. 这些结构单元包括脂肪酰基、氨基酸残基、非天然氨基酸残基和脂肪酰基-氨基酸等, 其引入往往发生在聚酮合成酶或聚肽合成酶的装配线上, 而去除发生在线下, 并且去除机制差异巨大. 更特别的是, 这些结构单元看似多余, 但实际上起到了非常重要的作用, 例如作为生化反应的保护基团、控制后续生化反应指导分子组装、特殊的自身抗性保护机制等等. 本文总结了2000年以来报道的这一类分子中额外结构单元的引入以及去除的生化过程, 探讨这一过程在生命活动中的重要作用, 从这些有限的例子中发现一些规律, 为分析其他复杂生化反应和生物体生命活动提供线索, 为其他化合物的生物合成途径研究提供参考, 并且可以作为基因信息与化合物结构关联的重要桥梁为现代生物信息学分析提供依据.

关键词 聚酮聚肽, “多余”结构单元, 引入, 去除, 生物学功能

从动植物或微生物来源的天然产物是现代医药的重要来源之一, 特别是许多产生于微生物的天然产物具有良好的抗菌、抗肿瘤、抗病毒或者免疫抑制等活性, 在医药产业中占有重要地位^[1-3]. 目前已报道的微生物天然产物数量庞大, 其中聚酮类(polyketides, PKs)、非核糖体聚肽类(non-ribosomal peptides, NRPs)及其杂合化合物不仅占了相当大的比例, 而且表现出突出的生物活性, 对维护人类健康和各种疾病的治疗做出了重大贡献. 例如, 柔红霉素被用作临床一线抗癌药物, 万古霉素和达托霉素被用于治疗所有抗生素均无效的严重感染, 藤霉素被广泛用于减少

或阻断外科移植、移植物抗宿主病以及自身免疫疾病治疗中经常出现的免疫排斥反应等等. 因此, PKs, NRPs及其杂合类化合物的生物合成过程引起了人们广泛的研究兴趣.

近年来, 在针对PKs, NRPs及其杂合类化合物的生物合成研究中发现, 在某些分子的骨架生物合成过程中, 会先引入一些结构单元, 再在后续反应中去除. 由于这些结构单元不会参与最终成熟分子的骨架构建, 导致它们显得多余. 但实际上, 深入研究发现, 这类结构单元在这些分子的生物合成过程中不可或缺, 起到了超出人们意料之外的重要作用. 本文

引用格式: 陈明, 陈华, 刘文. 聚酮聚肽类天然产物生物合成过程中“多余”结构单元的引入、去除及其生物学意义. 科学通报, 2017, 62: 3548–3560
Chen M, Chen H, Liu W. The incorporation and removal of “auxiliary” building blocks and their functions in the biosynthesis of polyketides and non-ribosomal polypeptides (in Chinese). Chin Sci Bull, 2017, 62: 3548–3560, doi: 10.1360/N972017-00170

总结了2000年以来报道的这一类分子中额外结构单元的引入以及去除的生化过程,并探讨这一过程在生命活动中的重要作用。

1 聚酮聚肽类天然产物骨架中结构单元的引入

PKs和NRPs的骨架分别由聚酮合酶(polyketide synthases, PKSs)和非核糖体聚肽合成酶(non-ribosomal peptide synthetases, NRPSs)催化合成. PKSs催化小分子羧酸前体之间形成C-C键,而NRPSs催化氨基酸前体之间形成C-N键,它们都遵循着一种模板化的生物合成逻辑^[4]. 这两类酶都是含有1个或多个催化模块的大型多功能酶,每一模块负责一次C-C键或C-N

键的形成,其催化功能的实现都需要至少3种不同功能域的参与,并且共享一个来源于辅酶A(CoA)的泛酰巯基乙胺臂所活化、负责挂载底物的硫酯化功能域(thiolation domain, T)^[5](图1(a)). 对于PKSs,酰基转移酶(acyltransferase, AT)识别和激活特定的底物,转移至T功能域上,在酮基合成酶(β -ketoacylsynthase, KS)的催化下通过克莱森缩合促使C-C键的形成^[6](图1(b)). 而NRPSs中的腺苷化功能域(adenylation, A)识别并活化特定的氨基酸,将其转移至T功能域上,在缩合功能域(condensation, C)催化下,通过亲核反应形成C-N键^[7](图1(b)). 由于PKSs和NRPSs催化反应与蛋白组织结构的相似性^[8],二者的模块之间可以进行置换和融合,最终产生了更为复杂的PK-NRPs

图1 PKS, NRPs及其杂合类天然产物骨架中结构单元的引入过程. (a) Apo形式的硫酯功能域或蛋白(T)被来源于辅酶A(CoA)的泛酰巯基乙胺活化成为 Holo形式. (b) PKSs, NRPSs及其杂合蛋白的模块之间催化形成C-C键或者C-N键

Figure 1 Assembly-line chemistry of PKSs, NRPSs and their hybrids. (a) Conversion of the thiolation (T) domain/protein from an inactive Apo-form into an active holo-form via phosphopantetheinyl transfer. (b) Modular PKSs and NRPSs for the formation of C-C and C-N bonds

或者NRP-PKs杂合化合物(图1(b)). 此外, 在每一步延伸过程中, 可能还会有一些额外的功能域或者其他游离蛋白对挂载于T功能域上的结构单元或中间体进行修饰. 例如由酮基还原酶(ketoreductase, KR)、脱水酶(dehydratase, DH)、烯基还原酶(enoylreductase, ER)等功能域对 β -酮基的进一步修饰^[9]; 由甲基化酶(methyltransferase, MT)、异构酶(epimerase, E)等功能域催化氨基酸发生甲基化、异构化等反应^[10]. 最后, 再由一个C末端的硫酯酶(thioesterase, TE)功能域解离完成延伸的中间体, 得到线性或者环状的分子骨架^[11]. 大多数PKs, NRPs及其杂合类天然产物的化学结构都体现了蛋白在组织结构和催化功能上一一对应的线性逻辑关系.

在本文中, 所提及分子最终成熟的骨架结构与按照蛋白组织结构推测的理论结构相比, 缺失了一部分结构单元, 导致负责引入这些结构单元的功能域在一开始被误认为是不起作用的. 直到后续的深入研究, 才揭示这些功能域不仅能够正常行驶其功能, 并且引入的结构单元具有非常重要的作用, 但它们会在后续反应中被切除. 因此, 本文将重点探讨这些结构单元的去除以及这一看似多余的“引入-去除”机制背后的生物学意义. 下文将按照所缺失结构单元的类型及其去除机制进行分类探讨.

2 缺失氨基酸

目前, 已经明确报道的引入多余氨基酸结构单元的分子有3类, 分别为 β -内酰胺类化合物诺卡菌素(Nocardicins), 2,2'-联吡啶类化合物青蓝霉素(Caerulomycins, CAEs)与克里斯霉素(Collismycins, COLs)以及大环内酰胺类化合物文森他汀(Vicenistatin), 它们的氨基酸去除机制各不相同.

在 β -内酰胺类化合物Nocardicins的生物合成基因簇中, NRPS蛋白NocA, NocB共有5个模块, 理论上应该得到五肽化合物, 然而实际上的成熟分子却是一系列被不同程度修饰的三肽(图2(a)). Townsend课题组^[12,13]通过体内、体外实验证实, 5个模块中的A功能域分别识别L-对-羟基苯甘氨酸(L-pHPG)、精氨酸(L-Arg)、L-pHPG、丝氨酸(L-Ser)和L-pHPG, 对应的产物为五肽L-pHPG-L-Arg-D-pHPG-L-Ser-L-pHPG(模块3中有1个异构化功能域E, 负责使L-pHPG发生异构化反应得到D-pHPG), 并且在体外验证胰蛋白酶可以水解五肽为L-pHPG-L-Arg和D-pHPG-L-Ser-

L-pHPG两个片段. 作者对Nocardicins产生菌的基因组进行分析, 在其生物合成基因簇外找到了3个胰蛋白酶的类似蛋白, 推测其中之一催化了该水解反应. 随后, Townsend课题组^[14]对Nocardicins结构中的 β -内酰胺环的生物合成机制进行了深入研究, 同时也揭示了N端“多余”二肽L-pHPG-L-Arg的生物学功能(图2(b)): 挂载于NocB-T4功能域上的四肽中间体中的Ser残基首先发生脱水反应, 紧接着与挂载于NocB-T5上的L-pHPG反应, 但是L-pHPG的氨基不再直接进攻羰基, 而是先进攻脱水丝氨酸的双键发生1,4-亲核加成, 然后再发生分子内加成消除反应, 完成四肽到五肽延伸过程的同时形成 β -内酰胺环. 此外, 体外实验证实, NocB-TE功能域专一性的识别已经形成 β -内酰胺环的底物, 对五肽底物的反应活性要高于三肽底物, 并且该功能域还能异构化组成 β -内酰胺环的D-pHPG为L-pHPG^[15]. 因此, 可以推测NocA, NocB反应初始引入的两个“多余”氨基酸是下游蛋白进行底物识别的标签, 有利于NocB上第4个模块到第5个模块之间的 β -内酰胺环形成以及TE结构域负责的异构化和水解过程.

2,2'-联吡啶类化合物CAEs和COLs结构高度相似, 区别仅在于A环C5位取代的不同. 近几年的研究表明, 2,2'-联吡啶环这一核心结构来源于NRPS-PKS-NRPS杂合装配线, 线上各个模块识别的前体分别为吡啶甲酸、丙二酸、L-半胱氨酸(L-Cys)和L-亮氨酸(L-Leu)^[16-18](图3). 根据同位素标记实验, 联吡啶骨架的形成仅需要吡啶甲酸、丙二酸和L-Cys的参与^[17], 但是体内研究发现L-Leu的延伸必不可少, 敲除负责延伸该单元的基因完全中断了相关化合物的产生^[16-18]. 并且, L-Leu的去除会在中间体从装配线上解离下来以后立即发生, 催化该反应的蛋白CaeD/ColD(另一命名分别为CrmL, ClmAH)同属于金属离子依赖的酰胺水解酶, 当敲除其编码基因以后, 分离得到带有L-Leu单元的中间体^[16,18](图3). 虽然引入“多余”L-Leu结构单元的生物学意义尚不明确, 但对这类化合物生物合成机制的深入研究给出了一些启示. 现有证据表明, 2,2'-联吡啶环核心结构是在一个黄素依赖的氧化还原酶CaeB1/ColB1参与下, 挂载于装配线上完成^[4,17](图3): 装配线的前3个模块分别以各自的前体进行链延伸反应, 但在L-Cys的延伸过程中, 可能是由巯基代替 α -氨基作为亲核基团进攻上游羰基; 然后, 该直链中间体要么直接发生分子内环

图2 诺卡菌素(Nocardins)的生物合成途径。(a) Nocardin G和Nocardin A的生物合成过程。(b) 在NRPS装配线上形成 β -内酰胺的过程。红色标记的L-pHPG-L-Arg为“多余”结构单元

Figure 2 Proposed biosynthetic pathway of Nocardins. (a) Nonribosomal assembly and removal of the “unnecessary” residues. (b) The formation of β -lactam on the assembly line. The “auxiliary” L-pHPG-L-Arg residues are highlighted in red

化反应, 要么继续延伸L-Leu后再发生环化, 形成一个七元杂环, 接着发生分子内重排得到六元环; 最后, 六元环经过不同的芳构化方式得到CAEs(C5位脱硫)或者COLs(C5位脱氢)中间体。虽然第2个吡啶环的具体构筑机制还不明晰, 但可以确定L-Leu在此过程中起到非常重要的作用, 如果先环化再延伸, 可能延伸L-Leu的过程对C5位不同取代的中间体具有一定选择性; 如果先延伸再环化, 则该结构单元可能对之后的芳构化方式具有重要的影响。在装配线上协助构筑C5位不同取代的联吡啶以后, L-Leu单元的作用已经完成, 但它的存在封闭了C7位羧基, 阻碍了发生在该位置上的一系列后续反应, 因此从装配

线上解离下来的中间体发生的第一步反应即是去除L-Leu, 引发其他后修饰反应最终得到C6位被不同基团取代的CAEs和COLs。

Vicenistatin是来源于PKS途径的大环内酰胺类化合物, 其生物合成过程中包含一个非常独特的起始单元3-氨基-2-甲基丙酸(3-amino-2-methylpropionate, Amp)。2011年, Eguchi课题组^[19]报道了该特殊单元的合成以及上载过程, 从中发现了一个“多余”L-丙氨酸(L-Ala)单元的引入-去除过程, 但与以上两类分子中引入和去除的时机不同, L-Ala是在与Amp一起进入PKS装配线之前被引入, 并且挂载于装配线上被去除(图4)。首先, L-谷氨酸(L-Glu)经过多步转化之后

图3 青蓝霉素A(CAE-A)和克里斯霉素A(COL-A)生物合成过程. 红色标记的L-Leu为“多余”结构单元

Figure 3 Proposed biosynthetic pathway of Caerulomycin A (CAEs) and Collismycin A (COL A). The “auxiliary” L-Leu is highlighted in red

形成挂载于游离载体蛋白VinL上的前体, 得到Amp-VinL; 然后, ATP依赖的连接酶VinM催化L-Ala连接到Amp单元的β-氨基上, 形成Ala-Amp-VinL; 之后, 酰基转移酶VinK将Ala-Amp转移至PKS中的T0功能域上, 在PKS的作用下完成碳链的延伸. 然而, 延伸完成的碳链需要先经过肽酶VinJ的水解作用脱除β-氨基上的Ala, 再在TE功能域的催化下形成具有大环内酰胺骨架中间体Vicenilactam, 最后发生糖基化反应形成Vicenistatin. 在此过程中, L-Ala起到了保护Amp单元上β-氨基的作用, 降低了其亲核性, 避免Amp起始单元在早期的PKS模块(特别是第1个)上完

成二碳单元延伸之后, 氨基亲核进攻自身的硫酯, 提前终止碳链延伸(图4). 而这一作用主要受到VinK和VinJ这两个蛋白的控制: 酰基转移酶VinK特异性的识别Ala-Amp这一结构并将其转移至PKS上, 如果没有L-Ala, 该酰基转移反应不能正常发生; 此外, 肽酶VinJ(属于α/β水解酶家族)只能识别完成多轮延伸得到一定碳链长度的的中间体(底物通道大约长16 Å), 杜绝了L-Ala在早期链延伸过程中被水解的可能, 确保得到正确碳链长度的Vicenistatin前体. 值得注意的是, 这一保护作用并非孤例, 在萨利内酰胺(Salinilactam), BE-14106, ML-449等其他以β-氨基酸

图4 文森他汀(Vicenistatin)的生物合成途径. 红色标记的L-Ala为“多余”结构单元, 来源于L-Glu的结构单元Amp标记为蓝色, 灰色途径为VinK和VinJ不识别的旁支反应

Figure 4 Proposed biosynthetic pathway of Vicenistatin. The “auxiliary” L-Ala is highlighted in red, the building block Amp is highlighted in blue, and the side reactions that VinK and VinJ do not recognize are in gray

作为起始单元的大环内酰胺类化合物的生物合成基因簇中, 同样发现了VinJ, VinK, VinL, VinM, VinN的同源蛋白^[20], 因此在这些分子的生物合成过程中极可能也要经过一个相似的氨基酸保护和脱保护过程.

3 缺失脂肪酰-氨基酸

缺失单元为脂肪酰-氨基酸的分子可以分为两类, 其中一类结构上并不相似, 但缺失的单元及其去除机制非常相似, 因此放在一起讨论; 另外一类是结构特异的四氢异喹啉类化合物.

对前一类分子, 最先开展研究并且结果比较清晰的是来源于嗜线虫致病杆菌(*Xenorhabdus nematophila*, 革兰氏阴性菌)的帕克素(Xenocoumacin)和来源于枯草芽孢杆菌(*Bacillus* spp., 革兰氏阳性菌)的双效菌素(Zwittermicin)^[21~23], 这两个分子生物合成中的“多余”结构单元均是位于N端的长链脂肪酸-

D-天门冬酰胺. 在这两个化合物的装配线中, 负责引入该单元的NRPS模块组织结构均为C-A-T-E(图5(a)), 催化的生化反应包括脂肪酸与L-天门冬酰胺(L-Asn)的缩合以及L-Asn到D-Asn的异构化过程. 之后, 脂肪酰-D-Asn作为起始单元引发后续的链延伸、线下后修饰等反应, 分别得到前体Prexencoumacin和Prezwittermicin, 两个前体最后分别由肽酶XcnG和ZmaM负责去除多余单元得到成熟分子(图5). 值得注意的是, XcnG和ZmaM的蛋白结构并不相同. XcnG是一个由细胞周质肽酶功能域和3个跨膜螺旋组成的双功能酶, 而ZmaM是由肽酶结构域与具有9个跨膜螺旋的ABC-转运蛋白融合而成(图5(a)). 综合考虑脂肪酰-D-Asn单元被引入、去除的时机和机制, 再结合对前体以及成熟分子的生物活性测试, 该过程被认为是一种新颖的抗性机制: 带有多余单元的前体是没有生物活性的前药, 不会对自身造成损害, 但在前

图5 帕克素(Xenocoumacin)和双效菌素(Zwittermicin)的生物合成途径及其化学结构。(a)以Xenocoumacin(来源于革兰氏阴性菌*X. nematophila*)和Zwittermicin(来源于革兰氏阳性菌*Bacillus* spp.)两个分子生物合成途径为代表的前药保护机制, XcnG和ZmaM为带有不同结构跨膜结构域的肽酶, 负责切除脂肪酰-D-Asn单元。黄色五角星和蓝色三角代表成熟药物分子, 绿色星形代表脂肪酰-D-Asn单元。(b) 前药Prexencoumacin B, Prezwittermicins和Predidemnins的结构。红色标记部分为“多余”结构单元

Figure 5 The proposed biosynthetic pathways and structures of Xenocoumacin and Zwittermicin. (a) The prodrug activation mechanisms in the biosynthesis of Xenocoumacin (produced by *X. nematophila*) and Zwittermicin (produced by *Bacillus* spp.); XcnG and ZmaM are peptidases with different transmembrane domains that catalyze the removal of acyl-D-Asn. The yellow stars and blue triangles are drugs, while the green stars are the protective groups. (b) The structures of Prexencoumacin B, Prezwittermicins and Predidemnins. The “auxiliary” residues are highlighted in red

体被转运出细胞的同时, 肽酶切除多余单元, 得到具有良好活性的成熟分子, 帮助产生菌适应周围环境(图5(a))。此外, 这一前药机制还在Amicoumacin, 大肠杆菌毒素(Colibactin), Paenilamicin和膜海鞘素(Didemnins)等分子中被发现^[24-27]。特别是, Predidemnins中的多余单元为脂肪酰-谷氨酰胺, 并且切除该单元时断裂的是酯键(图5(b)), 推测该反应由具有跨膜结构域的酯酶催化完成^[27]。

另外一类缺失脂肪酰-氨基酸结构单元的现象是在萘啶霉素(naphthyridinomycin, NDM)和喹诺卡星(quinocarcin, QCN)生物合成中被发现^[28,29]。然而, 在番红霉素 A(Saframycin A, SFM A)中却只缺失脂肪酰单元^[30]。虽然缺失单元有所不同, 但NDM, QCN和

SFM A同属于结构特异的四氢异喹啉类化合物, 其生物合成途径也高度相似, 因此放在一起进行讨论。目前为止, 对这一类化合物中“多余”结构单元的去除过程研究主要来自于NDM和QCN(图6(a))。唐功利课题组^[28]在NDM的生物合成中发现, 脂肪酰-L-丙氨酸单元的去除是由具有9个跨膜结构的肽酶NapG负责, 敲除*napG*基因以后, NDM的产量大大降低, 证实了这一过程; 而在QCN生物合成中, 肽酶Qcn1负责该过程, 并且其功能同样通过基因敲除得到证实。对这一“多余”结构单元的生物学意义的探索主要来自于对SFM A生物合成机制的研究^[30]。2010年, Oikawa课题组^[30]报道了SFM-A的生物合成过程(图6(b))。PKS-NRPS蛋白SfmA和SfmB以14个碳的直链脂肪酸

图6 四氢异喹啉类化合物的生物合成途径。(a) 萘啶毒素(NDM)和喹诺卡星(QCN)生物合成中“多余”脂肪酰-氨基酸单元的去除。(b) 番红霉素A(SFM-A)中四氢异喹啉类骨架的生物合成过程。红色标记的脂肪酰-氨基酸或者脂肪酰基为“多余”结构单元

Figure 6 Proposed biosynthetic pathways of tetrahydroisoquinoline alkaloids. (a) Removal of the “auxiliary” residues in the biosynthesis of naphthyridinomycin (NDM) and quinocarcin (QCN). (b) Proposed overall scheme for the pentacyclic core skeleton assembly of Saframycin A (SFM-A). The “auxiliary” residues are highlighted in red

作为起始单元, 延伸一个L-甘氨酸(L-Gly)后发生还原解离得到末端醛基化合物中间体; 该中间体与挂载在另一NRPS蛋白SfmC上的底物发生关键的Pictet-Spengler反应(P-S反应), 得到的中间体再次还原解离后重复发生P-S反应, 形成四氢异喹啉的特殊结构;

最后还原解离的中间体再发生分子内环化、切除脂肪链等等后修饰过程, 得到成熟的SFM-A分子。体外实验显示, SfmC中的还原功能域(R或者RE)比较特异的识别长链脂肪酰基(12, 14, 16个碳原子), 不识别短链的乙酰基。同时, 脂肪链的存在对于反应中的第一

次还原至关重要,这也直接决定SfmC催化的第一步P-S反应能否发生,继而影响整个分子的合成.除了这3个分子,对比其他四氢异喹啉类化合物海鞘素743(Ecteinascidin 743)、番红霉素 Mx1(Saframycin Mx1)和番红菌素 A(Safracin A)的NRPS装配线,可以发现其装配线起始蛋白都是一样的AL-T0-C1-A-PCP结构(Safracin A为C1-A-PCP结构)^[28-32],因此可以推测该引入-去除机制广泛存在于这一类化合物生物合成中,对RE功能域还原解离过程的底物识别起到关键作用,决定下游的P-S反应能否发生,最终控制四氢异喹啉结构的合成.

4 缺失脂肪链

除了部分四氢异喹啉类化合物,缺失脂肪酸单元的现象也在铜绿假单胞菌铁载体(Pyoverdine, PVD)和远霉素(Telomycin, TEM)的生物合成过程中被发现,虽然这两个分子的产生菌和分子结构不同,但它们去除脂肪链的机制相似.

PVD的生物合成研究开展的比较早,在2006年,Lamont课题组^[33]已经发表综述总结了当时的研究进展(图7(a)). NRPS蛋白PvdL的起始模块结构为AL-T0-C-A-T,负责引入脂肪酸作为起始单元;之后,在其他NRPS蛋白PvdI, PvdJ, PvdD的共同作用下发生肽链延伸并最终从装配线上解离下来;紧接着,带脂肪酰基的中间体在N端亲和水解酶(N-terminal nucleophile hydrolase, Ntn hydrolase)PvdQ作用下除去脂肪酸单元,再经过其他后修饰反应得到成熟的PVD分子.2011年,Gulick课题组^[34]进一步报道了PvdL中AL功能域和PvdQ的功能并解析了PvdQ蛋白的晶体结构.体外实验表明,AL功能域对链长为10, 12, 14个碳的脂肪酸具有较好的识别和上载活性,其中C14脂肪酸活性最好.同时,该课题组从敲除基因*pvdQ*的突变株发酵产物中分离得到带C14脂肪链的PVD前体,支持AL功能域的体外实验结果.另一方面,以该前体为底物对蛋白PvdQ进行体外测活,顺利实现脂肪链的剪切.对蛋白PvdQ的晶体结构解析发现,这是一个典型的Ntn水解酶,一开始表达得到的蛋白没有活性,需要在关键的217位丝氨酸(S217)残基处自催化断裂成为两个亚基,再形成异源二聚体,才成为一个活性蛋白(图7(b)). S217即催化活性位点,其羟基作为亲核试剂进攻酰胺键中的羰基,脂肪链以共价结合的方式挂载在蛋白上,实现酰胺键

的水解^[34,35](图7(b)).该课题组以不同链长的脂肪酰-PVD与PvdQ蛋白进行共结晶,检测到了S217上的羟基与C14脂肪酸共价结合形成酯的结构,验证了该催化机制.

对TEM的生物合成研究要更晚一些,其中脂肪酰基的上载过程和去除时机与PVD相比略有不同^[36](图7(c)). (1) 分别负责识别活化和挂载脂肪酰基的蛋白Tem18(脂肪酸连接酶,对应PvdL-AL), Tem19(酰基载体蛋白,对应PvdL-T0)是两个游离蛋白; (2) 带脂肪酰基的前体需要先经过一系列的线下后修饰反应,再由Ntn水解酶家族蛋白Tem25切除脂肪链得到成熟TEM分子.当基因*tem25*被敲除之后,带有6-甲基庚酰基的TEM前体被分离得到,并且以此为底物在体外测得Tem25的水解活性.此外, Tem25还在体外表现出对其他中等碳链长度脂肪酰底物(*n*-C₆H₁₃CO-TEM和*n*-C₇H₁₅CO-TEM)的水解活性,但不能水解长碳链(*n*-C₁₂H₂₅CO-TEM)化合物.

关于脂肪链在PVD和TEM产生菌细胞中的生化意义,并没有明确的结论.由于对PVD的生物合成过程认识有限,对几个细胞周质蛋白PvdO, PvdP以及PvdN的功能不甚了解^[37],而PvdQ催化的酰胺水解反应并非该化合物生物合成的最后一步,因此推测脂肪链可能起到协助PVD跨越内膜进入周质空间的作用,然后脂肪链-PVD再被几个周质蛋白(PvdO/P/N/Q)催化发生后修饰反应.而对成熟TEM分子和带脂肪链前体的生物活性测试结果显示,对大部分测试菌株而言,脂肪酰-TEM与TEM具有相当甚至更好的生物活性^[36].这说明,针对作者选用的测试菌株而言,TEM引入的脂肪酰基并没有起到前药保护的作用.考虑到PVD和TEM脂肪酰基上载和去除途径的相似性,推测其生化意义可能具有相似之处.除PVD和TEM外,在负责化合物博来霉素(Bleomycin)、他利霉素(Tallysomycin)、拉来霉素(Zorbamycin)和表面活性肽(Surfactin)生物合成的NRPS蛋白中也发现了同样的AL-T0-C-A-T起始模块结构^[38,39],但一直没有证据证实其功能,因此脂肪链的存在与否及其生物学意义都还是未解之谜.

5 总结与展望

聚肽类天然产物中,除了上文提到的NRPs,还有一类与之生源途径完全不同、依赖于核糖体合成肽链骨架的化合物,称之为核糖体翻译后修饰肽(ribosomally

图7 铜绿假单胞菌铁载体(PVD)和远霉素(TEM)的生物合成途径。(a) PVD中脂肪酰单元的上载和去除过程。(b) 没有催化活性的Pro-PvdQ蛋白自催化断裂并得到具有活性的PvdQ异源二聚体的过程以及PvdQ蛋白催化酰胺键水解的机制。(c) TEM中脂肪酰单元的上载和去除过程。红色标记的脂肪酰基为“多余”结构单元

Figure 7 Proposed biosynthetic pathways of Pyoverdine (PVD) and Telomycin (TEM). (a) The incorporation and removal of acyl in the biosynthesis of PVD. (b) Autoproteolysis and catalytic mechanism of PvdQ. (c) The incorporation and removal of the acyl group in the biosynthesis of TEM. The “auxiliary” residues are highlighted in red

synthesized and post-translationally modified peptides, RiPPs). 在RiPPs生物合成中, 由核糖体翻译得到的前体肽链可根据氨基酸序列在肽链中的位置(从N端到C端)和生化功能依次分为信号肽、先导肽、核心肽以及后识别序列, 其中先导肽和核心肽是所有该类化合物生物合成中共有的序列, 信号肽和后识别序列则只在一部分化合物中被发现。然而, 在最终的RiPPs化合物中, 只呈现出核心肽的序列, 其余部分都被切除^[40]。目前的研究结果认为, “多余的”信号

肽、先导肽和后识别序列在前体肽的翻译后修饰过程中通过蛋白-蛋白相互作用起到了底物识别、控制后修饰反应等重要作用^[41~43]。在本文所列分子中, 除了2,2'-联吡啶类化合物Caerulomycins和Collismycins之外, 其他分子中的“多余”结构单元均是作为起始单元或起始单元的一部分被引入PKs, NRPs及其杂合化合物骨架, 因此这可以看做是特殊的“前体肽”, 只是其类型并不仅仅局限于天然氨基酸, 引入和去除的机制更多样, 所起到的生物学功能更复杂。

对于 β -内酰胺化合物Nocardicins, 2,2'-联吡啶类化合物Caerulomycins和Collismycins, 四氢异喹啉类化合物Naphthyrudinomycin, Quinocarcin以及Saframycin A而言, 虽然引入的“多余”结构单元及其去除机制互不相同, 但这一单元都对形成各自特殊分子结构的生化反应具有一定的选择和控制作用, 这可以为我们分析其他带有这些特殊结构的分子生物合成途径提供参考依据. 在大环内酰胺类化合物Vicenistatin生物合成中发现的“多余”氨基酸单元的引入-去除过程, 揭示了一种新颖的生化保护机制, 这一机制很可能同时存在于其他以 β -氨基酸作为特殊起始单元的同类化合物生物合成过程中. 与以上具有一定结构特点分子不同, 以Xenocoumacin和Zwittermicin为代表的一类分子结构上互不相同, 但引入和去除“多余”脂肪酰-D-Asn(Didemnin中为脂肪酰-L-Gln)单元的机制高度相似, 这一过程的发现揭

示了一种通过引入多余单元形成前药的全新抗性机制. 目前报道的含有这类前药保护机制的分子来源于多个菌属, 既有革兰氏阳性菌, 也有阴性菌, 预示着该机制很可能广泛存在于微生物中. 最后, 虽然对Pyoverdine和Telomycin中的脂肪酰基引入和去除机制已经有了深入认识, 但该单元所起到的作用目前还不是很清楚, 有待进一步的研究.

综上所述, 虽然目前报道的生物合成过程中含有引入-去除“多余”结构单元这一前体肽机制的PKs, NRPs及其杂合类分子数量并不多, 但每一个发现都使得人们对生化反应或者生物体复杂的调控、适应系统有了新的认识. 同时, 我们还可以从这些有限的例子中发现一些规律, 将其作为分析复杂生化反应和生物体生命活动的线索, 为其他化合物的生物合成途径研究提供参考, 并且可以作为基因信息与化合物结构关联的重要桥梁为现代生物信息学分析提供依据.

参考文献

- Mishra B B, Tiwari V K. Natural products: An evolving role in future drug discovery. *Eur J Med Chem*, 2011, 46: 4769–4807
- Shen B, Thorson J S. Expanding nature's chemical repertoire through metabolic engineering and biocatalysis. *Curr Opin Chem Biol*, 2012, 16: 99–100
- Newman D J, Cragg G M. Natural products as sources of new drugs over the 30 years from 1981 to 2010. *J Nat Prod*, 2012, 75: 311–335
- Chen M, Liu J, Duan P, et al. Biosynthesis and molecular engineering of templated natural products. *Natl Sci Rev*, 2016, doi: 10.1093/nsr/nww045
- Walsh C T. Insights into the chemical logic and enzymatic machinery of NRPS assembly lines. *Nat Prod Rep*, 2016, 33: 127–135
- Hertweck C. The biosynthetic logic of polyketide diversity. *Angew Chem Int Ed*, 2009, 48: 4688–4716
- Sieber S A, Marahiel M A. Molecular mechanisms underlying nonribosomal peptide synthesis: Approaches to new antibiotics. *Chem Rev*, 2005, 105: 715–738
- Walsh C T. The chemical versatility of natural-product assembly lines. *Acc Chem Res*, 2008, 41: 4–10
- Keatinge-Clay A T. Stereocontrol within polyketide assembly lines. *Nat Prod Rep*, 2016, 33: 141–149
- Marahiel M A. Working outside the protein—Synthesis rules: Insights into non-ribosomal peptide synthesis. *J Pept Sci*, 2009, 15: 799–807
- Du L, Lou L. PKS and NRPS release mechanisms. *Nat Prod Rep*, 2010, 27: 255–278
- Davidson J M, Townsend C A. *In vivo* characterization of nonribosomal peptide synthetases NocA and NocB in the biosynthesis of nocardicin A. *Chem Biol*, 2012, 19: 297–306
- Davidson J M, Bartley D M, Townsend C A. Non-ribosomal propeptide precursor in nocardicin A biosynthesis predicted from adenylation domain specificity dependent on the MbtH family protein NocI. *J Am Chem Soc*, 2013, 135: 1749–1759
- Gaudelli N M, Long D H, Townsend C A. β -lactam formation by a non-ribosomal peptide synthetase during antibiotic biosynthesis. *Nature*, 2015, 520: 383–387
- Gaudelli N M, Townsend C A. Epimerization and substrate gating by a TE domain in beta-lactam antibiotic biosynthesis. *Nat Chem Biol*, 2014, 10: 251–258
- Garcia I, Vior N M, Brana A F, et al. Elucidating the biosynthetic pathway for the polyketide-nonribosomal peptide collismycin A: Mechanism for formation of the 2,2'-bipyridyl ring. *Chem Biol*, 2012, 19: 399–413
- Qu X, Pang B, Zhang Z, et al. Caerulomycins and collismycins share a common paradigm for 2,2'-bipyridine biosynthesis via an unusual hybrid polyketide-peptide assembly Logic. *J Am Chem Soc*, 2012, 134: 9038–9041
- Zhu Y, Fu P, Lin Q, et al. Identification of caerulomycin A gene cluster implicates a tailoring amidohydrolase. *Org Lett*, 2012, 14: 2666–2669

- 19 Shinohara Y, Kudo F, Eguchi T. A natural protecting group strategy to carry an amino acid starter unit in the biosynthesis of macrolactam polyketide antibiotics. *J Am Chem Soc*, 2011, 133: 18134–18137
- 20 Miyanaga A, Kudo F, Eguchi T. Mechanisms of β -amino acid incorporation in polyketide macrolactam biosynthesis. *Curr Opin Chem Biol*, 2016, 35: 58–64
- 21 Luo Y, Ruan L F, Zhao C M, et al. Validation of the intact zwittermicin A biosynthetic gene cluster and discovery of a complementary resistance mechanism in *Bacillus thuringiensis*. *Antimicrob Agents Chemother*, 2011, 55: 4161–4169
- 22 Reimer D, Pos K M, Thines M, et al. A natural prodrug activation mechanism in nonribosomal peptide synthesis. *Nat Chem Biol*, 2011, 7: 888–890
- 23 Reimer D, Bode H B. A natural prodrug activation mechanism in the biosynthesis of nonribosomal peptides. *Nat Prod Rep*, 2014, 31: 154–159
- 24 Li Y, Xu Y, Liu L, et al. Five new amicoumacins isolated from a marine-derived bacterium *Bacillus subtilis*. *Mar Drugs*, 2012, 10: 319–328
- 25 Brotherton C A, Balskus E P. A prodrug resistance mechanism is involved in colibactin biosynthesis and cytotoxicity. *J Am Chem Soc*, 2013, 135: 3359–3362
- 26 Muller S, Garcia-Gonzalez E, Mainz A, et al. Paenilamicin: Structure and biosynthesis of a hybrid nonribosomal peptide/polyketide antibiotic from the bee pathogen *Paenibacillus larvae*. *Angew Chem Int Ed*, 2014, 53: 10821–10825
- 27 Xu Y, Kersten R D, Nam S J, et al. Bacterial biosynthesis and maturation of the didemnin anti-cancer agents. *J Am Chem Soc*, 2012, 134: 8625–8632
- 28 Pu J Y, Peng C, Tang M C, et al. Naphthyridinomycin biosynthesis revealing the use of leader peptide to guide nonribosomal peptide assembly. *Org Lett*, 2013, 15: 3674–3677
- 29 Hiratsuka T, Koketsu K, Minami A, et al. Core assembly mechanism of quinocarcin/SF-1739: Bimodular complex nonribosomal peptide synthetases for sequential mannich-type reactions. *Chem Biol*, 2013, 20: 1523–1535
- 30 Koketsu K, Watanabe K, Suda H, et al. Reconstruction of the saframycin core scaffold defines dual Pictet-Spengler mechanisms. *Nat Chem Biol*, 2010, 6: 408–410
- 31 Koketsu K, Minami A, Watanabe K, et al. Pictet-spenglerase involved in tetrahydroisoquinoline antibiotic biosynthesis. *Curr Opin Chem Biol*, 2012, 16: 142–149
- 32 Le V H, Inai M, Williams R M, et al. Ecteinascidins: A review of the chemistry, biology and clinical utility of potent tetrahydroisoquinoline antitumor antibiotics. *Nat Prod Rep*, 2015, 32: 328–347
- 33 Visca P, Imperi F, Lamont I L. Pyoverdine siderophores: From biogenesis to biosignificance. *Trends Microbiol*, 2007, 15: 22–30
- 34 Drake E J, Gulick A M. Structural characterization and high-throughput screening of inhibitors of PvdQ, an NTN hydrolase involved in pyoverdine synthesis. *ACS Chem Biol*, 2011, 6: 1277–1286
- 35 Bokhove M, Nadal Jimenez P, Quax W J, et al. The quorum-quenching N-acyl homoserine lactone acylase PvdQ is an Ntn-hydrolase with an unusual substrate-binding pocket. *Proc Natl Acad Sci USA*, 2010, 107: 686–691
- 36 Fu C, Keller L, Bauer A, et al. Biosynthetic studies of telomycin reveal new lipopeptides with enhanced activity. *J Am Chem Soc*, 2015, 137: 7692–7705
- 37 Nelson K E, Weinel C, Paulsen I T, et al. Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. *Environ Microbiol*, 2002, 4: 799–808
- 38 Cosmina P, Rodriguez F, de Ferra F, et al. Sequence and analysis of the genetic locus responsible for surfactin synthesis in *Bacillus subtilis*. *Mol Microbiol*, 1993, 8: 821–831
- 39 Galm U, Wendt-Pienkowski E, Wang L, et al. Comparative analysis of the biosynthetic gene clusters and pathways for three structurally related antitumor antibiotics: Bleomycin, tallysomyin, and zorbamycin. *J Nat Prod*, 2011, 74: 526–536
- 40 Arnison P G, Bibb M J, Bierbaum G, et al. Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat Prod Rep*, 2013, 30: 108–160
- 41 Oman T J, van der Donk W A. Follow the leader: The use of leader peptides to guide natural product biosynthesis. *Nat Chem Biol*, 2010, 6: 9–18
- 42 Burkhart B J, Hudson G A, Dunbar K L, et al. A prevalent peptide-binding domain guides ribosomal natural product biosynthesis. *Nat Chem Biol*, 2015, 11: 564–570
- 43 Fuchs S W, Lackner G, Morinaka B I, et al. A Lanthipeptide-like N-Terminal leader region guides peptide epimerization by radical SAM epimerases: Implications for RiPP evolution. *Angew Chem Int Ed*, 2016, 128: 12518–12521

Summary for “聚酮聚肽类天然产物生物合成过程中‘多余’结构单元的引入、去除及其生物学意义”

The incorporation and removal of “auxiliary” building blocks and their functions in the biosynthesis of polyketides and non-ribosomal polypeptides

CHEN Ming, CHEN Hua & LIU Wen*

State Key Laboratory of Bioorganic and Natural Products Chemistry, Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, Shanghai 200032, China

* Corresponding author, E-mail: wliu@mail.sioc.ac.cn

Natural products that produced by animals, plants or microorganisms, especially polyketides (PKs), non-ribosomal peptides (NRPs) and their hybrids, exhibit an extremely wide range of biological activities, include antibacterial, antifungal, antitumor, antiviral and immunosuppressive activities, which underlie the critical roles of natural products in medical industry as various drugs. The biosynthetic pathways of PKs, NRPs and their hybrids share a templated multifunctional enzymology for skeleton assembly. Polyketide synthases (PKSs) usually catalyze C–C bond formation using short carboxylic acid as monomers, whereas non-ribosomal peptide synthetases (NRPSs) incorporate amino acids through C–N bond formation. Consistent with their catalytic cycles, both PKSs and NRPSs are often giant enzymes organized into modules, which each contains a common thiolation (T) domain for loading building blocks. A prototype of a PKS module consists of an acyltransferase (AT), a T domain and a β -ketoacyl synthase (KS) domain. A minimum NRPS module contains an adenylation (A) domain, a T domain and a condensation (C) domain. AT or A domain is responsible for recognition and activation of building blocks, while KS or C domain catalyzes the elongation of the growing skeleton. In general, modular PKSs or NRPSs appear to function as assembly lines, which program monomer polymerization/modification and chain tailoring/termination following a non-iteratively “one domain, one function” like co-linearity rule. Thus, the domain composition of each module and the number of modules can be used to predict the structure of skeleton. However, there are exceptions reported in recent years, as modules are now known to be both skipped and iterated during the normal biosynthetic processes. In addition, another origin for a lack of correlation between domain composition of the multienzyme complex and the final product structure has been identified recently. In this case, the missing building blocks are normally incorporated on the assembly lines, but they will be removed by the post-assembly modifications. These kind of building blocks which would not appear in the final products seem “unnecessary”, but they are essential in the biosynthetic pathways exactly. In this review, we summarize these “auxiliary” building blocks, focus on the chemical mechanisms of the incorporation and removal of them and discuss their biological functions that have been reported since 2000. The related building blocks here include fatty acyls, amino acids, unnatural amino acids and fatty acyl-amino acids. Some of these “auxiliary” build blocks are involved in the biosynthesis of natural products with specific structures, such as 2,2'-bipyridine cores, tetrahydroisoquinoline alkaloids, β -lactam and macrocyclic lactam antibiotics. In this case, building blocks are incorporated by the PKSs, NRPSs or their hybrids assembly lines, but the chemical mechanisms are unusual. The biological functions of these kind of building blocks are considered to be strategies for protection of the reactive groups and mediation of molecular assembly and modifications. On the other hand, some of the “auxiliary” building blocks are present in biosynthesis of natural products which are structurally different from each other, however, the “unnecessary” building blocks are almost same and their chemical mechanisms of incorporation and removal are similar. These kind of building blocks are proposed to be very important for protection the chemical producers by prodrug generation. In summary, the molecules mentioned in this review could increase the knowledge regarding the “auxiliary” building blocks, add another layer of complexity to natural product biosynthesis, help to hypothesize biosynthetic pathways of other natural products with specific structures and assist the association of molecules and their gene clusters.

Polyketides and peptides, “auxiliary” building blocks, incorporation, removal, biological functions

doi: 10.1360/N972017-00170