

蛋白甲基化识别结构域抑制剂的研究进展

周理炜¹, 樊甜甜², 刘同超², 吴小余^{1*}, 熊兵^{2*}

(1. 上海大学理学院化学系, 超分子化学与催化中心, 上海 200444; 2. 中国科学院大学, 中国科学院上海药物研究所, 上海 201203)

摘要: 随着质谱和蛋白质组学技术的发展, 人们对蛋白质翻译后甲基化修饰的研究也取得了长足进步, 发现和鉴定了一大批甲基化识别相关蛋白。蛋白翻译后修饰中的甲基化主要发生于赖氨酸和精氨酸, 其中赖氨酸可形成单甲基化、双甲基化和三甲基化, 而精氨酸则可形成单甲基化、对称或不对称双甲基化。甲基化识别结构域可精确识别不同甲基化程度的赖氨酸或精氨酸, 传递甲基化信号, 调控下游多种细胞过程, 包括基因表达调控、RNA 剪接翻译、细胞周期调控等。随着研究的深入, 人们发现这些识别蛋白的异常与肿瘤发生发展也密切相关, 因而这些甲基化识别蛋白已成为小分子干预的潜在药物靶点。鉴于该领域发展迅猛, 本文对多种蛋白甲基化识别结构域进行综述, 简介各结构域中重要蛋白及其抑制剂的研究进展, 希望能为后续的药物研发提供靶点选择和抑制剂设计的参考。

关键词: 甲基化; 甲基化识别结构域; 抑制剂; 表观遗传学

中图分类号: R914 文献标识码: A 文章编号: 0513-4870(2021)11-3014-16

Research progress of protein methylation reader domain inhibitors

ZHOU Li-wei¹, FAN Tian-tian², LIU Tong-chao², WU Xiao-yu^{1*}, XIONG Bing^{2*}

(1. Center for Supramolecular Chemistry and Catalysis and Department of Chemistry, College of Sciences, Shanghai University, Shanghai 200444, China; 2. University of Chinese Academy of Sciences, Chinese Academy of Sciences Shanghai Institute of Materia Medica, Shanghai 201203, China)

Abstract: The development of mass spectrometry and proteomics significantly advanced our understanding of post-translational methylation of proteins. In recent years, a large number of proteins containing the methylation recognition domain have been identified. Protein methylation mainly occurs on lysine and arginine residues. Both lysine and arginine have three different methylation states. Lysine can be mono-methylated, di-methylated, and tri-methylated, while the arginine residue can be modified as mono-methylation, symmetrical di-methylation, and asymmetrical di-methylation. Methylation recognition domains can accurately identify lysine or arginine with different state of methylation, transfer methylation signals, and perform functions in a variety of cellular processes, including gene expression regulation, RNA splicing and translation, cell cycle regulation, *etc.* In recent years, researchers have found that the abnormalities of these recognition proteins are also closely related to the genesis and development of tumors. Therefore, these methylation recognition proteins were considered as potential drug targets for small molecule intervention. In this review, we summarized the researches on the recognition domains of protein methylation as well as their inhibitors, hoping to provide the basis for further drug development in this field.

Key words: methylation; methylation recognition domain; inhibitor; epigenetics

收稿日期: 2021-04-06; 修回日期: 2021-07-29.

基金项目: 国家“新药创制”科技重大专项基金项目 (2018ZX09711002-004); 自然科学基金面上资助项目 (81773572).

*通讯作者 Tel: 13661970096,

E-mail: bxiong@simm.ac.cn; wuxy@shu.edu.cn

DOI: 10.16438/j.0513-4870.2021-0486

甲基化是生物体非常普遍的修饰模式, 可发生在 DNA、RNA 和蛋白质上。利用这一简单的修饰模式, 生物体可以极大地拓展核苷酸、氨基酸的种类, 从而增加基因、蛋白质的多样性, 实现经济高效的精确调控。蛋白甲基化是研究较为充分的蛋白翻译后修饰的方

式,一般发生于赖氨酸和精氨酸的侧链。在2012年的一篇综述中,Arrowsmith等^[1]这样描述:蛋白的甲基化是一种动态的、可逆的翻译后修饰形式,其甲基化调控的整个过程主要涉及三大类蛋白,包括可以“写入”修饰的蛋白甲基化转移酶、“擦除”修饰的蛋白去甲基化酶和“阅读”修饰的甲基化识别蛋白。甲基转移酶将甲基从S-腺苷蛋氨酸(SAM)转移到相应赖氨酸或精氨酸残基上,而去甲基化酶则可通过FAD辅酶或者通过铁离子催化的方式氧化去除甲基,二者活性的平衡控制着蛋白的甲基化状态^[2,3]。而蛋白甲基化识别蛋白一般包含多个结构域,它们通过甲基化识别结构域来解读甲基化的信号,再通过其他结构域传递信号给下游的蛋白或执行相应的催化功能。正是通过这三类蛋白的协同调控,细胞可实现甲基化信号的传递,从而对细胞的多个过程进行精确的调控。

蛋白甲基化修饰可呈现不同的修饰状态,如图1所示,赖氨酸甲基化可具有单甲基化(Kme1)、双甲基化(Kme2)和三甲甲基化(Kme3)三种,而精氨酸甲基化具有单甲基化(Rme1)、对称双甲基化(Rme2s)以及不对称双甲基化(Rme2a)^[4]。

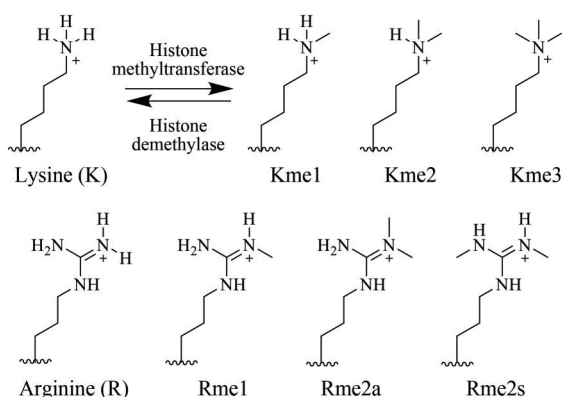


Figure 1 Methylation states of lysine and arginine in histones

不同于组蛋白乙酰化修饰,甲基化不改变氨基酸侧链的形式电荷,仅仅改变体积、电荷分布以及亲脂性。甲基化识别蛋白可识别不同状态的甲基化修饰,从而发挥精确调控功能^[4]。随着研究的深入,人们发

现甲基化识别蛋白与多种癌症的发生发展密切相关^[5,6],例如:HP1 α (heterochromatin protein) 是一种通过染色质域介导特异性识别甲基化H3K9的甲基化识别蛋白^[7],其有 α 、 β 、 γ 三种亚型)的下调与乳腺癌细胞高侵袭潜能有关^[8-10];HP1 α 下调也在甲状腺乳头状癌^[11]、成神经管细胞瘤^[12]中被发现。这些研究表明甲基化识别蛋白是一类潜在的药物靶标。

随着人类基因组和蛋白组学研究的进展,目前对于甲基化识别蛋白已有较为深入的了解。其中,甲基化赖氨酸识别结构域包括MBT (malignant brain tumor domains) 结构域、染色质 (chromodomains) 域、Tudor 结构域、PWWP (proline-tryptophan-tryptophan-proline domains) 结构域、PHD (plant homeodomain fingers) 结构域、WD40 (WD40 repeat domains) 结构域、ANK (ankyrin repeats) 结构域、BAH (bromo adjacent homology domains) 结构域^[13,14]。这些结构域可识别不同甲基化程度的赖氨酸,在某些蛋白中的Tudor结构域^[13,15]和WD40结构域^[16,17]也可识别甲基化的精氨酸残基。从结合模式上分析,这些甲基化识别结构域一般是通过2~4个芳香残基形成的“芳香笼”来精确识别甲基化的赖氨酸残基,甲基化赖氨酸结合位点附近的氨基酸残基也通过控制构象变化等方式进一步提高识别的特异性^[4](图2)。

本文将根据识别蛋白的抑制剂研究进展并按照识别结构域类别选择代表性蛋白进行介绍。结合蛋白-抑制剂的复合物晶体结构,简要分析抑制剂的结合模式,并特别关注了相关抑制剂的发现历程,希望为后续抑制剂设计提供参考。

1 染色质域及其抑制剂

人源染色质域蛋白HP1于2001年首次被发现可与组蛋白上的甲基化赖氨酸结合^[18,19]。目前在人类基因组中已确认有29个含染色质域蛋白。该结构域的抑制剂研发集中于两个蛋白家族:HP1和多梳蛋白家族(chromobox, CBX蛋白)。其中HP1结合H3K9me3^[19],而CBX蛋白特异性识别H3K27me3^[20],两者均主要参与基因的表达抑制调控过程。通过已解析的晶体结构

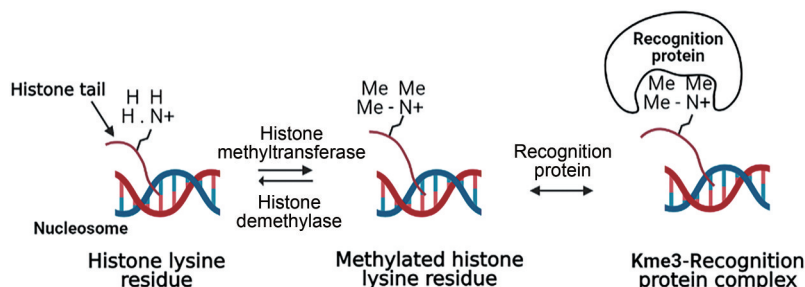


Figure 2 The pattern of action of the recognition domain of methylation proteins

可知染色质域的芳香笼口袋稍浅,这使其成为一个极具挑战性的小分子药物靶标。与多肽的相互作用分析显示,三甲基化的赖氨酸残基占据芳香笼并与周围残基的芳香环形成多个阳离子- π 相互作用,其主链可以和蛋白质结合槽之间形成多个特异性的氢键^[21,22],从而实现底物选择性。

CBX7是研究比较充分的多梳蛋白,它通过染色质域特异性识别H3K27me3。CBX7与多种癌症息息相关,例如:CBX7的高表达可引发并加速淋巴瘤的发生,并与原癌基因Myc协作产生侵袭性淋巴瘤^[23];CBX7的过表达与卵巢透明细胞腺癌(clear cell adenocarcinoma of the ovary, OCCA)的不良预后有关,且CBX7的敲除能够诱导OCCA细胞的凋亡^[24]。

第一个CBX7抑制剂是通过改造与CBX7有较强结合的多肽获得的。2014年Simhadri等^[25]根据CBX7的相互作用蛋白SETDB1结合界面的三甲基化肽段(RQVAVKme3STRGFALKme3STHGIAIKme3STN, $K_d = 1.2 \mu\text{mol}\cdot\text{L}^{-1}$, PDB: 4X3S)进行简化,合成只有5个氨基酸的短肽(Ac-FALKme3S-NH₂),该短肽对CBX7有一定的活性($\text{IC}_{50} = 12 \mu\text{mol}\cdot\text{L}^{-1}$)。如图3所示,其与蛋

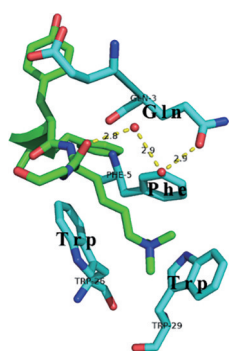


Figure 3 The binding interactions of Ac-FALKme3S-NH₂ in CBX7 complex

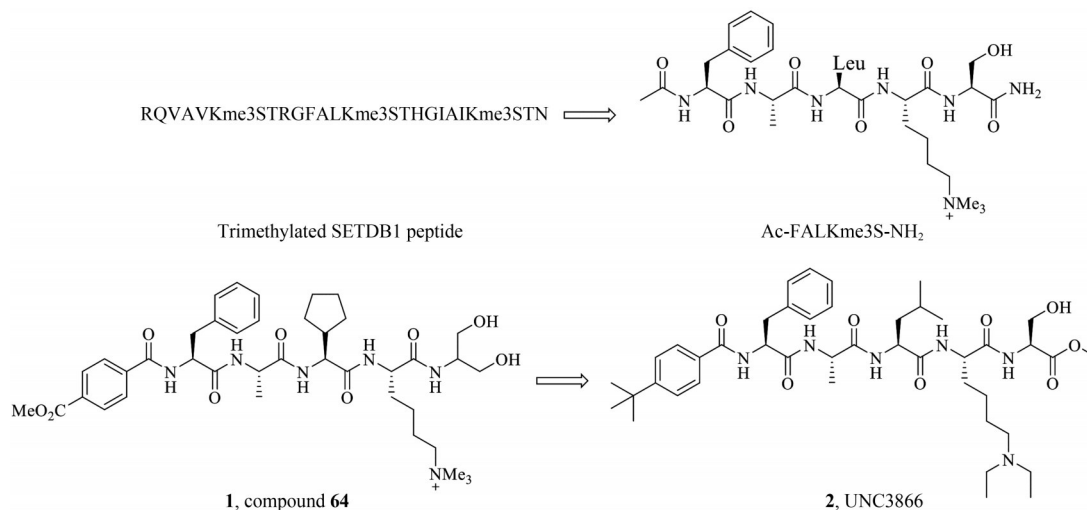


Figure 4 The design strategy of CBX7 peptidomimetic inhibitors

白的晶体结构(PDB: 4MN3)显示肽段与蛋白的结合主要是通过三甲基化赖氨酸和组成芳香笼的芳香残基之间的阳离子- π 相互作用来介导;三甲基化赖氨酸残基邻近的丝氨酸残基与CBX7蛋白的Gln3之间形成了氢键作用,通过水分子介导与Gln3和Glu2之间酰胺键的NH形成氢键相互作用,增强了与蛋白的结合;体积较小的丙氨酸残基插入CBX7蛋白一个小的疏水区域,使这个短肽衍生物具有较强的亲和性及选择性。通过进一步的分子优化,他们发现在短肽的N端添加疏水性芳香基团,可将活性提升40倍左右。最后他们得到具有百纳摩尔活性的拟肽类化合物**1**(图4, $K_d \approx 200 \text{ nmol}\cdot\text{L}^{-1}$)。

为了克服化合物**1**三甲基化赖氨酸侧链季铵盐膜渗透性差的问题,Stuckey等^[26]选择叔胺和仲胺作为电子等排体,开展优化,期望获得具有较好细胞渗透性的CBX7抑制剂,化合物**2**(UNC3866)的改造策略如图4。他们发现采用二乙基赖氨酸残基的拟肽对CBX7的结合能力与化合物**1**相当。随后,他们对N端的疏水基团进行优化,得到化合物**2**($K_d = 97 \text{ nmol}\cdot\text{L}^{-1}$, PDB: 5EQ0)。**2**在后续的生物评价中也表现出较好的抗细胞增殖活性。该研究不仅得到具有较好细胞活性的化合物,还证明了三甲基赖氨酸季铵盐并不是三甲基化赖氨酸识别结构域抑制剂的必要片段,为后续相关抑制剂的开发提供了宝贵的经验。

2014年Ren等^[27]通过高通量筛选得到非肽类选择性的CBX7抑制剂**3**(MS452, $K_i = 33.1 \mu\text{mol}\cdot\text{L}^{-1}$)。根据CBX7-MS452复合物的晶体结构(PDB: 4X3T, 图5左)可以看出:双甲氧基苯环将Trp35外推以占据由Phe11、Trp32、和Trp35构成的芳香笼;哌嗪环夹在Phe11和Trp32之间,形成“夹心”堆积作用;连接二甲氧基苯与哌嗪环的羰基氧可以通过水介导与His5形成

氢键相互作用。将形成芳香笼的 Phe11、Trp35 或 Tyr39 突变为 Ala 后发现, 突变体 CBX7 不能与 **3** 结合, 证明芳香笼对于 **3** 的结合是必不可少的。他们通过序列比较, 分析了 **3** 具有较高选择性的原因: 在 HP1 蛋白中, Trp35 变成了共轭体系更小的 Phe, 而 Phe 的苯环与双甲氧基苯形成 π - π 共轭作用相对较弱; 在 CBX7 的其他同族蛋白中, Tyr39、Val13 和 His47 由其他氨基酸代替, 导致 **3** 与相应蛋白的结合能力显著下降。

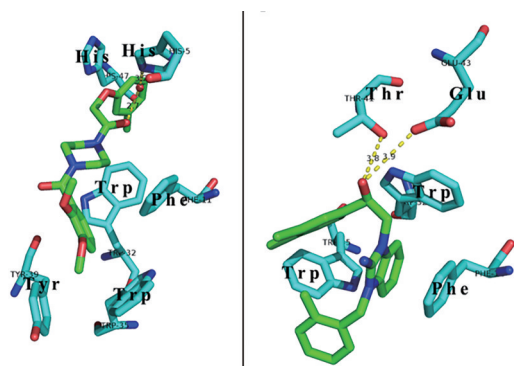


Figure 5 The binding interactions of MS452 (**3**, left), MS351 (**5**, right) and CBX7

在后续工作中, 他们通过结构生物学研究发现甲苯与二甲氧基苯采取不同构象时化合物与蛋白结合模式有显著变化。如图 6 所示, 当甲苯与二甲氧基苯采取反式构象的时候, 两个羰基的氧原子通过水分子介导分别与 His5 和 Tyr39 形成氢键相互作用 (图 6 左); 当采取顺式构象的时候, 甲苯向内“插入”蛋白中, 破坏了 β -折叠, 不利于与蛋白的结合 (图 6 右)。由此他们猜测引入与 CBX7 额外的氢键或疏水作用来稳定反式构象, 可能会提高配体的亲和力。他们对甲苯进行衍生化, 其中化合物 **4** (图 7, MS508)^[28] 的活性与 **3** 相比有较

大的提升 ($K_i = 4.8 \mu\text{mol} \cdot \text{L}^{-1}$)。但所合成的化合物的细胞活性均很差。因此, 他们进一步考虑替换化合物的母核。他们通过筛选得到包含亚氨基苯并咪唑结构的化合物 **5** (MS351)^[28], 其与 CBX7 的晶体结构 (PDB: 5EJW, 图 5 右) 显示: 亚氨基苯并咪唑核心占据着芳香笼, 与 Phe11 形成 π - π 堆积作用; 亚氨基苯并咪唑垂直于 Trp32 的吡啶环, 增加了 **5** 与芳香笼的相互作用强度; 连接部分的羟基与 Thr41、Glu43 形成氢键相互作用。实验证明, **5** 可以有效诱导小鼠 ES 细胞和人前列腺癌 PC3 细胞中包括 p16INK4a 在内的 CBX7 靶基因的转录去抑制。

另外 CBX6 也有拟肽类抑制剂^[29]报道, 由于篇幅有限, 本文不做赘述。

2 MBT 域及其抑制剂

MBT 结构域存在于 L3MBT 蛋白和多疏蛋白中, 主要识别单甲基化和双甲基化的赖氨酸残基。该结构域有以下特点: ① MBT 结构域由高度保守的芳香残基 (Phe、Tyr、Trp) 和亲水基团 (Asp) 形成芳香笼, 亲水残基还可以与甲基化赖氨酸残基的 NH 形成氢键相互作用; ② 配体“插入”深而窄的亲水腔, 与蛋白结合, 所以 Kme1/Kme2 与蛋白的结合能力要强于 Kme0/Kme3。

在恶性脑肿瘤的遗传学研究中, SCML2L3 (一种多梳蛋白)、L3MBTL2 和 L3MBTL3 的 MBT 域在髓母细胞瘤中纯合缺失^[30]。L3MBTL1、L3MBTL3 和血液恶性肿瘤有关^[31-33], L3MBTL4 在乳腺癌患者中被证明是失调的^[34]。尽管相关蛋白与癌症的联系、相关抑制剂及其结合的模式都有报道, 但人们对其在体内的复杂作用机制知之甚少。

2011 年, Herold 等^[35]报道了第一个靶向 L3MBTL1 蛋白的 MBT 结构域的小分子抑制剂, 该抑制剂也是针

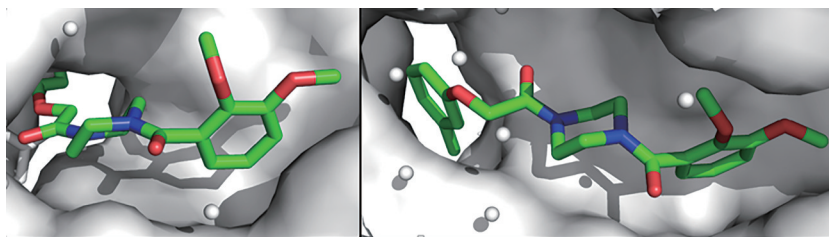


Figure 6 The binding interactions of **3** and CBX7 complex with toluene and dimethoxybenzene in different conformations (left, *cis*; right, *trans*)

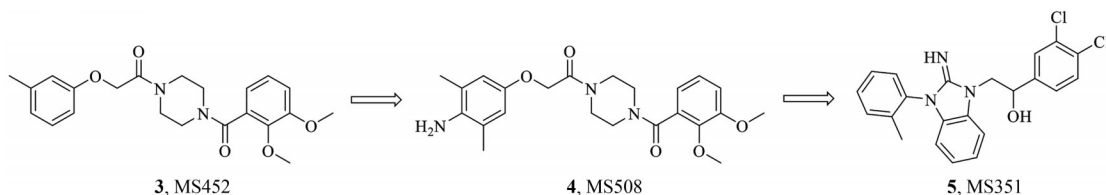


Figure 7 The optimization process from **3** to **5**

对赖氨酸甲基化识别结构域所报道的第一个小分子抑制剂。他们基于 H4K20Me2 与 L3MBTL1 复合物的共晶结构, 设计、合成了拟肽类衍生物。构效关系研究发现采用四氢吡咯环修饰的赖氨酸侧链来模拟二甲甲基化的赖氨酸残基能提高化合物的结合活性。他们得到活性较优的拟肽化合物 **6** (图 8, $K_d = 37 \mu\text{mol}\cdot\text{L}^{-1}$); 进一步简化拟肽结构得到化合物 **7** ($K_d = 26 \mu\text{mol}\cdot\text{L}^{-1}$) 和 **8** ($K_d = 25 \mu\text{mol}\cdot\text{L}^{-1}$); 将连接部分改造成哌啶环, 得到化合物 **9** (UNC669), 活性提高了约 5 倍 ($K_d = 5 \mu\text{mol}\cdot\text{L}^{-1}$)^[36]。 **9** 与 L3MBTL1 复合物的晶体结构 (PDB: 3P8H, 图 9 左) 显示吡咯烷的 N 质子化后与芳香笼形成阳离子- π 共轭作用, 分子中的 NH 还可以与 Asp355 形成氢键。

尽管到目前为止还没有研究表明 MBT 结构域可以与组蛋白多价结合, 但在 9 种人源蛋白中存在重复的 MBT 结构域^[37], James 等^[38]猜测 MBT 结构域可能通过多价结合的方式与底物结合。为此, 他们设计了“双碱”类化合物, 得到化合物 **10** (UNC1215, $K_d = 0.12 \mu\text{mol}\cdot\text{L}^{-1}$, 图 9 右和图 10) 和 **11** (UNC2533, $K_d = 0.062 \mu\text{mol}\cdot\text{L}^{-1}$)^[39]。 UNC1215-L3MBTL3 复合物的晶体结构 (PDB: 4FL6)

显示: 苯胺间位的吡咯烷深埋进 L3MBTL3 的 MBT II 结构域; 苯胺邻位的吡咯烷与第二分子 L3MBTL3 的 MBT I 结构域在其表面结合; 另外一分子的 **10** 采用上述结合方式, 交叉形成 2-2 的二聚物。他们通过选择性评价发现, 虽然 L3MBTL1 和 L3MBTL3 同源性很高, 参与结合芳香笼的结构以及氨基酸序列几乎一样, 但 **10** 对 L3MBTL3 的活性要比对 L3MBTL1 高 100 倍以上, 这是因为 **10** 和 L3MBTL1 采取单价的结合模式。在细胞水平, **10** 也有较好的增殖抑制能力, 后续的生物评价表明 **10** 是一个较理想的 L3MBTL3 探针分子。他们的工作提出多价结合的模型, 并证明了靶向同一识别蛋白的多个结合口袋策略的有效性。后续他们在 **10** 的基础上, 又引入异二氢吡啶环得到化合物 **12** (UNC1679), 经过构效关系探索, 得到了选择性更高的化合物 **13** ($K_d = 0.35 \mu\text{mol}\cdot\text{L}^{-1}$)^[40]。

3 Tudor 域及其抑制剂

Tudor 结构域蛋白是人类基因组中一个相对较大的蛋白家族, 由 41 个成员组成。Tudor 结构域具有 50~60 个氨基酸基序的共性序列, 可作为衔接蛋白和招募

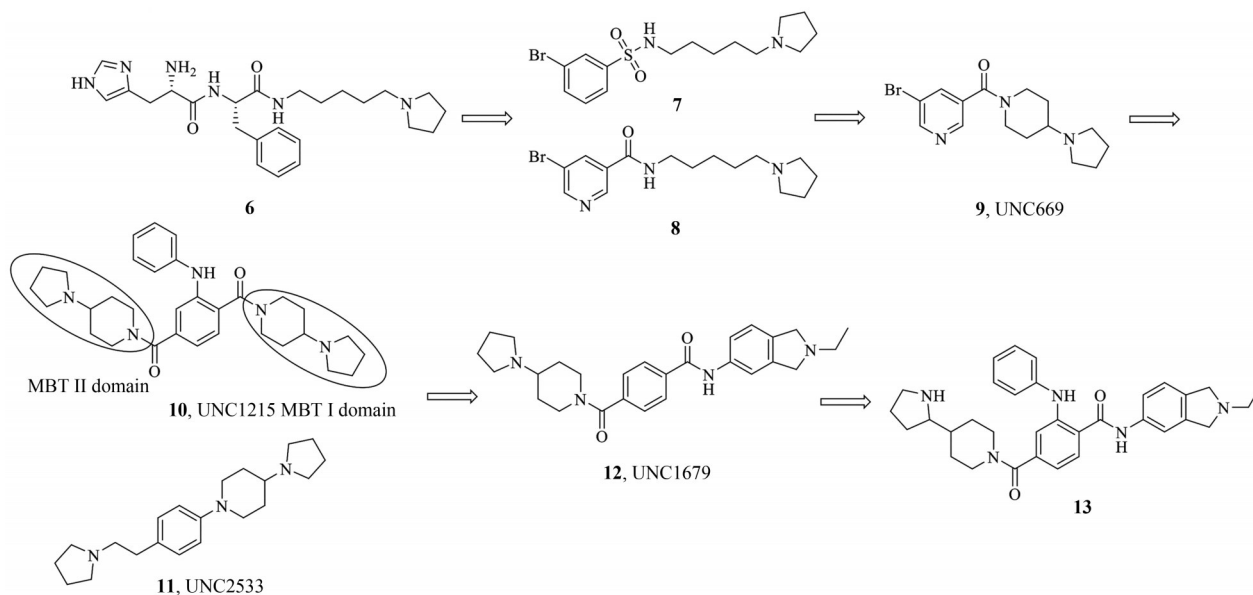


Figure 8 The optimization for modifying peptidomimetic structure **6** to compound **13**

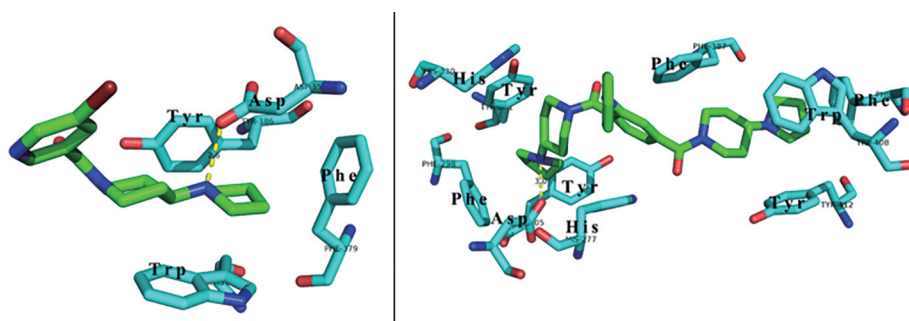


Figure 9 The binding interactions of UNC669 (**9**, left), UNC1215 (**10**, right) with L3MBTL3

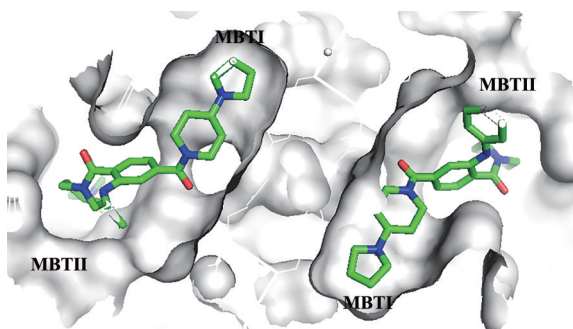


Figure 10 The pattern of 2-2 dimer formed with **10** and L3MBTL3

蛋白形成蛋白-蛋白相互作用, 促进多蛋白复合物的形成^[41]。它们在RNA代谢、组蛋白修饰、DNA损伤修复过程中都发挥着重要作用^[42]。Tudor结构域主要识别甲基化赖氨酸残基, 但在TDRD3、SMN、SPF30和Spindlin1蛋白中可识别甲基化精氨酸残基。

53BP1蛋白中含有串联的Tudor结构域, 它与H4K20me2结合, 从而影响DNA的损伤修复^[43]; 同时其也可以与肿瘤抑制蛋白P53的K382me2^[44]结合。敲除实验表明53BP1在发育、免疫应答和肿瘤发生中具有重要作用^[45]。2014年, Perfetti等^[46]通过对包含10种甲基化识别结构域的蛋白进行筛选得到具有较好选择性的53BP1 Tudor抑制剂**14**(图11, UNC2170, $IC_{50} = 30 \mu\text{mol}\cdot\text{L}^{-1}$)。该化合物对53BP1的活性比对其他甲基化赖氨酸识别蛋白要高至少17倍。其与53BP1 Tudor蛋白的晶体结构(PDB: 4RG2, 图12)显示: 酰胺的NH与Asp521可以形成氢键; 叔丁基胺锚定在赖氨酸结合口袋, 与内源性的赖氨酸甲基化的肽段竞争结合。与内源肽仅结合其中一个Tudor结构域不同, 该抑制剂同时与串联的两个Tudor结构域发生作用。通过结构优化研究, 他们发现连接部分的脂肪链替换为刚性基团会降低活性, 这可能是刚性连接器使叔丁胺偏离结合位点引起的。

Spindlin1 (SPIN1) 含有3个Tudor结构域, 既可以识别甲基化赖氨酸残基, 又可以识别甲基化精氨酸残

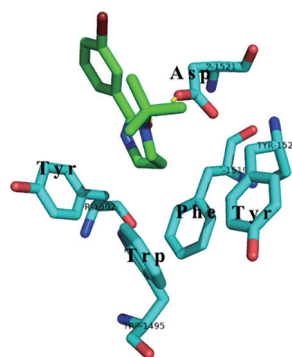


Figure 12 The binding interactions of **14** and 53BP1 Tudor domain

基。H3K4me3与SPIN1 Tudor II结合, H3K4me3R8me2a则结合在SPIN1 Tudor I口袋^[47]。其中, 后者的结合还可以加强前者的结合能力。

2017年, Sweis等^[48,49]发现化合物**15**(A366, G9a抑制剂)对SPIN1有一定程度的抑制($IC_{50} = 190 \text{ nmol}\cdot\text{L}^{-1}$)。他们通过构效关系研究发现A366细微的结构变动都会导致对SPIN1活性的显著减少。他们随后通过蛋白微阵列筛选了50多个**10**(L3MBTL3 MBT结构域选择性抑制剂)类似物, 得到对酶谱扩大的化合物**16**(EML405), 该化合物对SPIN1也有抑制作用。**16**与SPIN1蛋白复合物的晶体结构(PDB: 5JSJ, 图13左)显示: **16**的对称结构两端的哌啶环分别与两个Tudor结构域结合; 酰胺的羰基氧通过水分子介导与蛋白形成氢键; 在蛋白中, 苯胺所处位置还有较大的空腔未被占据。随后, 他们合成化合物**17**(EML631)^[50], 引入吡咯烷去占据带负电的口袋。等温滴定量热实验表明, **17**的结合常数比**16**有所提高。从**17**与SPIN1的晶体结构(PDB: 5JSJ, 图13右)可以看出: 对称的四氢吡咯环分别占据了两个Tudor结构域的芳香笼; 苯胺的NH与蛋白通过水分子介导形成氢键; 酚醚的氧原子可以与Tyr177形成氢键; 苯胺侧链的四氢吡咯环占据了两个Tudor结构域中间的结合空腔, 且该环上的N原子与Asp173之间有静电相互作用。他们通过此改造大大提高了化合

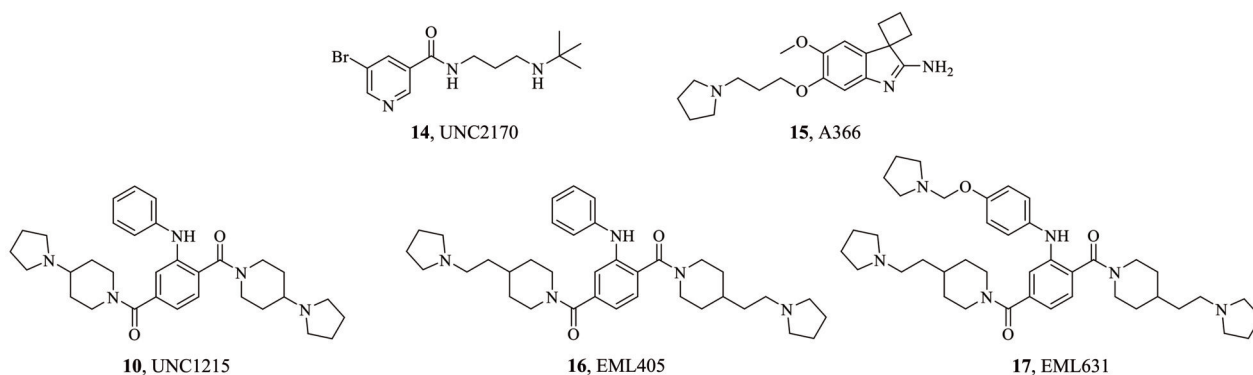


Figure 11 The chemical structures of **14** and **15**, as well as the optimization strategy from **10** to **17**

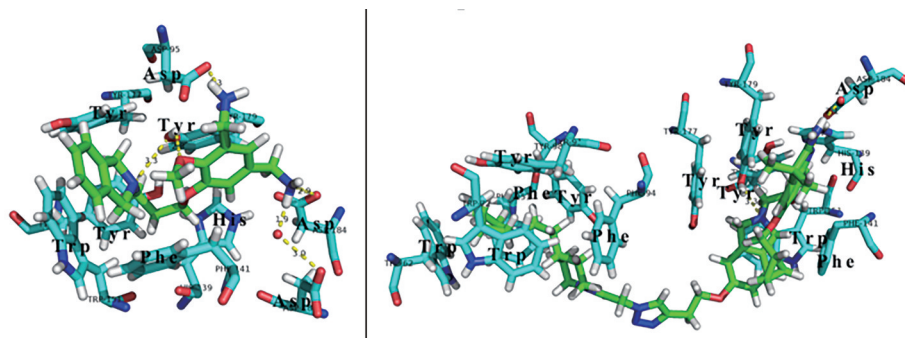


Figure 15 The binding interactions of MS31 (20, left), VinSpinIn (22, right) with SPIN1

应来引入三氮唑结构来控制连接子的长度, 经过新一轮优化后, 得到化合物 **22** (VinSpinIn)^[52]。化合物 VinSpinIn 与 SPIN1 复合物的晶体结构 (PDB: 6I8B, 图 15 右) 显示在与 Tudor II 结构域的结合模式保持不变的情况下, 新引入的四氢吡咯环占据 Tudor I 结构域的芳香笼。虽然化合物 VinSpinIn 的分子质量很大, 但由于其有着较好的水溶性、稳定性以及膜透性, 该化合物仍是 SPIN1 蛋白的重要探针分子。

另外, Upadhyay 等^[53]基于 2D-NMR 技术的片段筛选方法获得了与 KDM4A 串联 Tudor 结构域特异性结合的化合物 **23** (图 16)。Mader 等^[54]筛选了包含 252 个分子的化合物库得到对 SETDB1 串联 Tudor 结构域只有较弱活性的片段, 经过一系列的结构优化工作, 最后得到具有中等抑制活性的化合物 **24**。Senisterra 等^[55]也通过筛选得到 UHRF1 串联 Tudor 结构域抑制剂 **25** (NV03)。Liu 等^[56]通过片段筛选得到了靶向 TDRD3 Tudor 结构域精氨酸结合口袋的抑制剂 **26**。

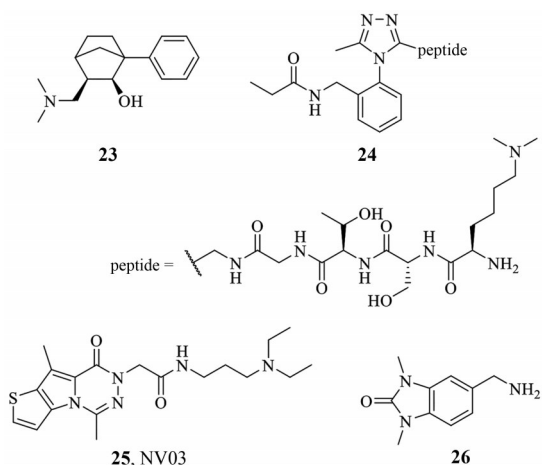


Figure 16 The chemical structures of 23–26

4 PWWP 结构域及其抑制剂

PWWP 结构域最早在 WHSC1 蛋白中被发现, 并因具有 Pro-Trp-Trp-Pro 基序而被命名 PWWP 结构域^[57]。

包含 PWWP 结构域的蛋白通常参与染色质相关的生物学过程, 如转录调控和 DNA 修复。PWWP 结构域是一个包含 100~150 个氨基酸的基序, 其突变与多种人类疾病有关。结构研究表明, PWWP 结构域具有保守的芳香笼用来识别组蛋白甲基化赖氨酸, 结合组蛋白或 DNA, 这有助于其与核小体的结合和染色质定位。此外, PWWP 结构域经常与其他组蛋白和 DNA 识别或修饰结构域协作, 影响表观遗传相关的调控过程^[58]。

2019 年, Böttcher 课题组^[59]利用饱和转移差-核磁共振波谱、差示扫描荧光测定法以及二维 ¹H/¹⁵N 横向弛豫优化核磁共振谱作为粗筛手段, 筛选了包含 1 899 个分子的片段库。他们确定了几个与 NSD3 PWWP 蛋白有结合的片段并获得了化合物 **27** (NMR, $K_d = 330 \mu\text{mol}\cdot\text{L}^{-1}$, PDB: 6G25, 图 17 左)、**28** (NMR, $K_d = 650 \mu\text{mol}\cdot\text{L}^{-1}$, PDB: 6G27, 图 17 右) 与 NSD3 PWWP 复合物的晶体结构。之后, 他们通过虚拟筛选得到 108 个小分子片段, 其中与蛋白结合 K_d 值最高的化合物 **29** (SPR, $K_d = 160 \mu\text{mol}\cdot\text{L}^{-1}$) 与蛋白的结合模式与化合物 **27**、**28** 相应的结合模式接近。以化合物 **29** 作为起始片段, 他们开展了结构优化工作, 最终得到具有较好选择性、细胞增殖抑制活性的第一个 NSD3 PWWP 结构域的探针分子 **30** (BI-9321; SPR, $K_d = 0.166 \mu\text{mol}\cdot\text{L}^{-1}$)。 **30** 与蛋白的晶体结构 (PDB: 6G2O, 图 18 左) 表明: 苯胺与 Glu318 之间可以形成氢键相互作用; 甲基咪唑环占据芳香笼; 咪唑的 N 与 Ser314 形成氢键; 咪唑环上的甲基与芳香残基 Trp284 之间形成 C-H- π 共轭作用。

Ferreira 等^[60]采用 ZMYND11 PWWP 结构域 (具有结构明确的甲基化赖氨酸结合袋, 被认为可用于代替 PWWP 结构域进行虚拟筛选) 虚拟筛选了大约 2 百万个可购买的小分子片段, 最后购买了其中 39 个分子来进行活性测试。他们通过表面等离子共振技术测试片段与 NSD2 PWWP 结构域的结合强度, 发现化合物 **31** ($K_d = 41 \mu\text{mol}\cdot\text{L}^{-1}$, SPR, 图 19) 可以稳定 N-端 NSD2 PWWP 蛋白。他们采用骨架跃迁的策略来优化化合

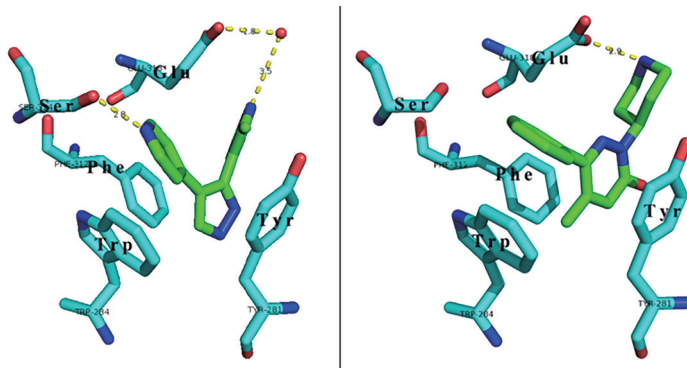


Figure 17 The binding interactions of compound **27** (left), compound **28** (right) and NSD3 PWWP domain

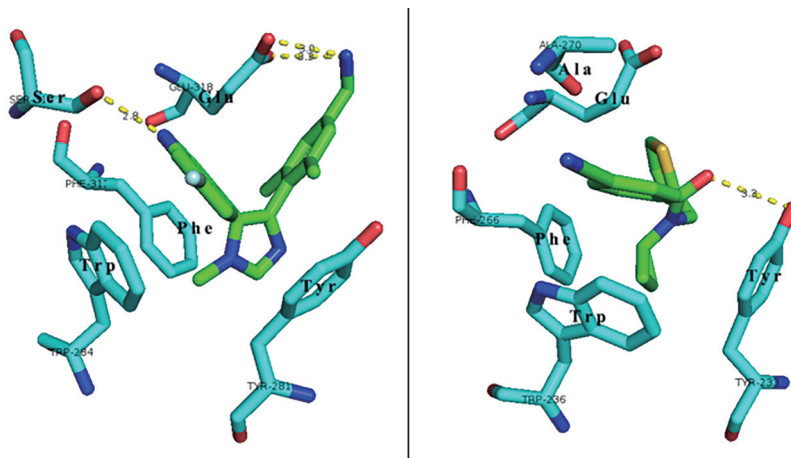


Figure 18 The binding interactions of BI-9321 (**30**) and NSD3 PWWP domain (left), MR837 (**34**) and NSD2 PWWP domain (right)

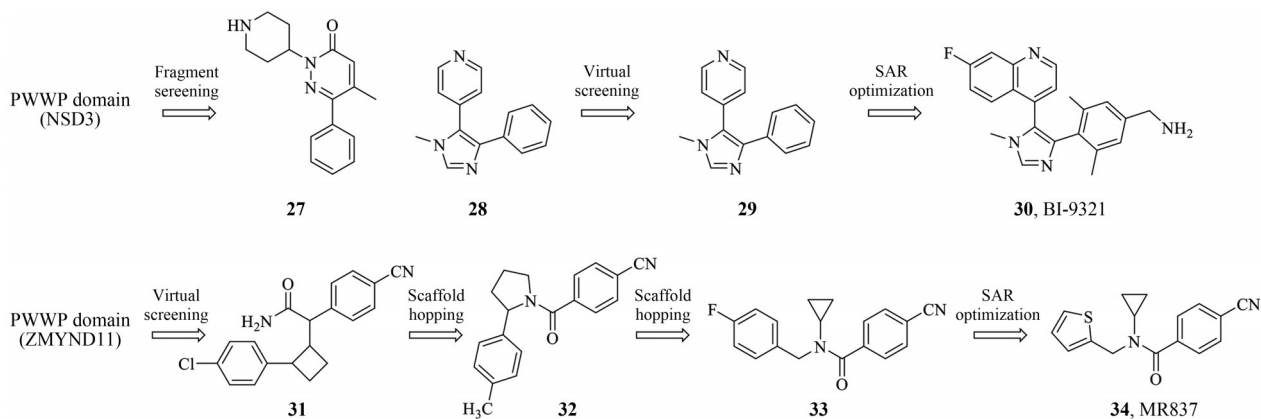


Figure 19 The optimization strategies of BI-9321 and MR837

物**31**的活性、选择性以及溶解度，其中只有衍生化合物**32** ($K_d = 175 \mu\text{mol}\cdot\text{L}^{-1}$, SPR)对蛋白表现出中等程度的结合。他们接着购买了一些化合物**32**的类似物，探索构效关系，最后发现苯甲腈的结构对于化合物与蛋白的活性至关重要。他们固定苯甲腈的结构，又购买一些类似化合物，并测得化合物**33** ($K_d = 7 \mu\text{mol}\cdot\text{L}^{-1}$, SPR)的活性相比化合物**32**有了很大提升。化合物**33**不仅消除了手性中心，也更容易衍生化。经过深入优

化，他们获得了NSD2 PWWP结构域第一个抑制剂**34** (MR837, $K_d = 3.4 \mu\text{mol}\cdot\text{L}^{-1}$, SPR)。化合物**34**与蛋白的共晶结构 (PDB: 6UE6, 图18右)显示：苯甲腈CN上的N原子与主链上Ala270 (Ala和Asp)构成酰胺键的NH形成氢键相互作用；N-环丙基酰胺则是占据着芳香口袋；噻吩环伸向蛋白内部，与蛋白形成疏水相互作用。

5 PHD结构域及其抑制剂

PHD结构域蛋白家族由超过40种组蛋白识别蛋

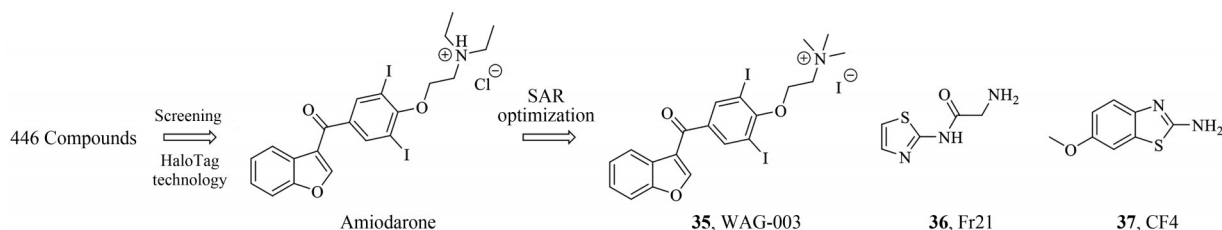


Figure 20 The optimization strategy of WAG-003 and the chemical structures of compounds Fr21 and CF4

白组成, 主要识别组蛋白上甲基化的赖氨酸侧链, 以及未修饰的组蛋白^[61]。生物学研究已经确定了依赖 PHD 结构域的组蛋白识别蛋白在染色质重塑复合物和转录编码中的功能。此外, PHD 结构域家族成员 Pygo、JARID1A 和 BAZ2A 的表达水平升高与各种人类癌症的致癌基因有关, 包括结直肠癌^[62]、急性髓细胞性白血病^[63]和前列腺癌^[64]。PHD 结构域在不同类型癌症中的作用表明, 靶向 PHD 结构域的小分子具有作为新型治疗手段的潜力。

2012年, Wagner等^[65]利用 HaloTag 技术测试了 446 个临床研究化合物对 JARID1A PHD3 结构域的活性, 发现双硫仑、胺碘酮 (胺碘酮, 属 III 类抗心律失常药, 是具有轻度非竞争性的 α 、 β 肾上腺素受体阻滞剂)、替加色罗这 3 个化合物与蛋白有较好的结合。他们对胺碘酮进行结构优化, 得到活性、选择性较好的化合物 35 (WAG-003, 图 20)。在结构优化过程中, 他们发现胺碘酮的季铵盐结构对于和蛋白的结合至关重要。

同年, Amato等^[66]通过片段筛选得到了靶向 BAZ2A PHD 结构域的化合物 36 (Fr21)。2014年, Miller等^[67]通过片段筛选得到苯并噻唑类衍生物, 可以特异性地结合到 Pygo PHD 结构域 H3K4 的结合位点。经过结构优化后, 他们得到化合物 37 (CF4)。该化合物与蛋白的晶体结构 (PDB: 4UP5, 图 21) 验证了他们的猜想: 氨基和 Thr240、噻唑环的 N 和 Asp380 可以形成氢键相互作用; 苯并噻唑的苯环与 Phe354、Ala332, 噻唑环的 S 与 Thr359 通过形成疏水相互作用, 提高对蛋白的结合能力。

6 WD40 结构域

WD40 结构域一般由 4~16 个包含色氨酸-天冬氨酸 (WD) 基序组成的单元, 整体结构呈现典型 β -折叠桶^[68]。WD40 结构域作为骨架蛋白, 参与许多生物过程, 包括染色质组装、细胞信号转导、细胞周期控制、RNA 转录和免疫应答^[69]。

PRC2 (polycomb repressive complex 2) 复合物包括 3 个核心的亚基: 组蛋白甲基转移酶 EZH2 (enhancer of zeste homologue 2)、EED (embryonic ectoderm development) 蛋白和 SUZ12 (suppressor of zeste 12) 蛋白, 其可以催化 H3K27 的甲基化^[70]。其中, 具有 WD40 结构域

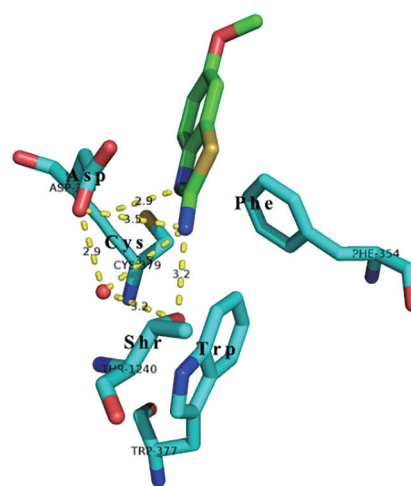


Figure 21 The binding interactions of compound CF4 (37) and Pygo PHD domain

的 EED 蛋白可以识别 H3K27me3 和招募其他蛋白形成 PRC2 复合物。针对 EED 蛋白已开展了较多的抑制剂研发工作, 其中进展最快的是诺华公司靶向 EED 蛋白的抑制剂 MAK-683 (图 22), 现已进入临床 II 期试验阶段。然而至今未见文献报道该化合物的研发历程, 本文对该公司已经披露的前期研究进行简要介绍。

2017年, 诺华公司从 EED-H3K27me3 出发通过高通量筛选得到具有较高活性的化合物 38^[71]。化合物 38 与蛋白复合物的晶体结构 (PDB: 5H19, 图 23 左) 显示: 5-位的氨基 NH 与 Asn194、2-位的 N 与 Lys211 和 Tyr365、3-位的 N 与 Tyr148 可以形成氢键相互作用, 且 CN 的 N 可以由水分子介导与 Arg367 形成氢键; 双环共轭体系与 Tyr365、Tyr148 之间通过形成 π 共轭作用增加化合物与蛋白的亲合力; 咪唑环与 Tyr365 可以形成 edge-to-face 共轭作用。通过对比化合物 38、H3K27Me3 与蛋白复合物的晶体结构, 他们发现化合物与 Arg367、Trp364、Tyr365 的结合会引起这几个残基构象的变化, 产生更大的结合口袋。他们发现化合物 38 的片段化合物 39 ($IC_{50} = 1.00 \mu\text{mol}\cdot\text{L}^{-1}$) 活性与化合物 38 ($IC_{50} = 4.03 \mu\text{mol}\cdot\text{L}^{-1}$) 相当, 且拥有更高的配体效率和亲脂效率。随后, 他们从化合物 39 出发开始 SAR 优化, 最后得到化合物 40 (EED226, $IC_{50} = 0.022 \mu\text{mol}\cdot\text{L}^{-1}$)^[72], 该化

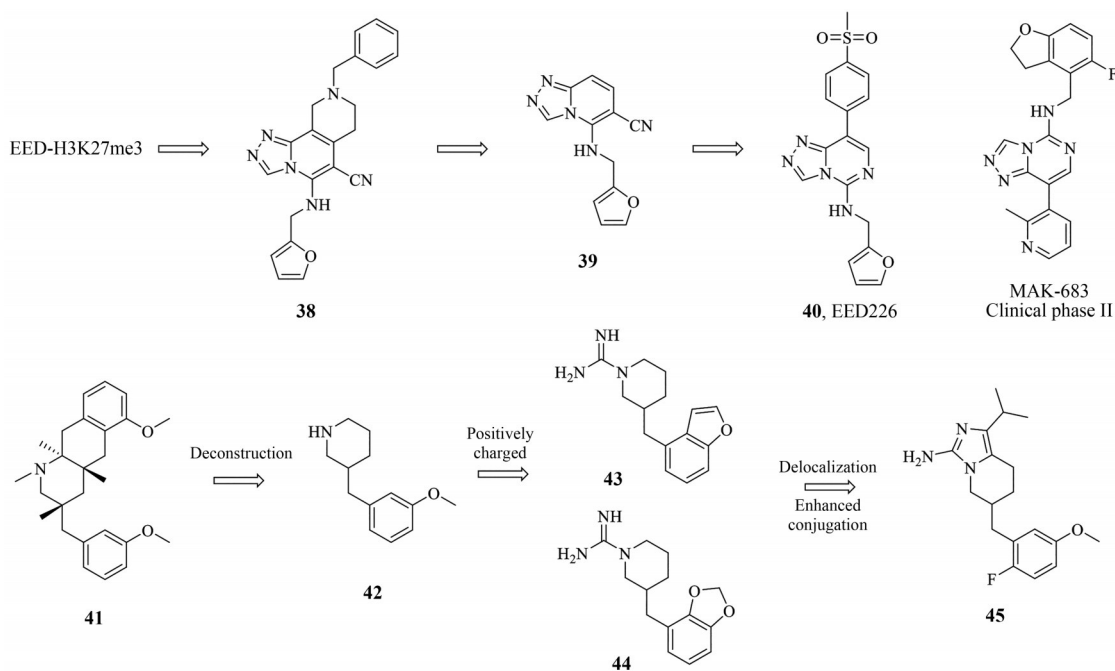


Figure 22 The optimization strategies of **40** and **45**, and the chemical structure of MAK-683

化合物的活性得到明显提升。晶体结构 (PDB: 5GSA, 图 23 右) 表明: 化合物 **40**、**38** 与蛋白结合的模式类似; 对甲磺酰基苯可以提高 π - π 相互作用。

在细胞水平, 化合物 **40** 可以特异性地抑制 H3K27 的甲基化。其还可以抑制 PRC2 复合物的催化活性并调控基因转录, 这一点与 SAM 竞争性的 EZH2 抑制剂的效果类似。小鼠口服给药 ($40 \text{ mg}\cdot\text{kg}^{-1}$, 32 天) 可以有效诱导肿瘤的完全消退。在动物体内耐受性良好, 其对体重的影响有限, 没有明显的不良反应。在 EZH2 耐药的细胞模型中, 其仍可以有效抑制细胞增殖以及 H3K27 的三甲基化。耐药的 EZH2 突变一般通过广泛使用不同的 SAM 竞争性的 EZH2 抑制剂治疗, **40** 联合 SAM 竞争性 EZH2 的抑制剂可能提供一种更有效的癌症治疗策略^[73]。

中等活性的化合物 **41** ($\text{IC}_{50} = 2.5 \text{ }\mu\text{mol}\cdot\text{L}^{-1}$, PDB: 5H17, 图 24 左), 考虑到合成的难易程度以及可优化的空间, 他们将化合物 **41** 解构, 来确定最小的药效团, 得到

化合物 **42** ($\text{IC}_{50} = 95 \text{ }\mu\text{mol}\cdot\text{L}^{-1}$, PDB: 5U5K)。他们随后在哌嗪环处引入胍基, 希望增强化合物与芳香笼形成的阳离子- π 共轭作用, 得到化合物 **43** ($\text{IC}_{50} = 2.1 \text{ }\mu\text{mol}\cdot\text{L}^{-1}$) 和化合物 **44** ($\text{IC}_{50} = 3.9 \text{ }\mu\text{mol}\cdot\text{L}^{-1}$, PDB: 5U5T)。他们引入 2-氨基咪唑环, 希望在降低化合物 pKa 值的同时, 保持其正电性。最后, 经过 SAR 优化后, 他们得到化合物 **45** ($\text{IC}_{50} = 1.3 \text{ }\mu\text{mol}\cdot\text{L}^{-1}$, PDB: 5U5H, 图 24 右)^[74]。

另外艾伯维的 Curtin 等^[75,76] 针对 EED 蛋白通过高通量筛选得到化合物 **46** ($K_i = 600 \text{ nmol}\cdot\text{L}^{-1}$, PDB: 5U69, 图 25), 经过合理的 SAR 优化得到化合物 **47** (A-395, $K_i = 0.31 \text{ nmol}\cdot\text{L}^{-1}$, PDB: 5K0M)。其通过占据 EED 的 H3K27me3 口袋, 抑制 PRC2 复合物的活性。Western blot 结果显示, GSK126 (SAM 竞争性 EZH2 抑制剂) 和 **47** 对 H3K27me3 的调节活性相似^[77]。

研究表明, K116 三甲基化的 Jarid2 (jumonji, AT-rich interactive domain 2) 蛋白是 EED 的结合底物, 同

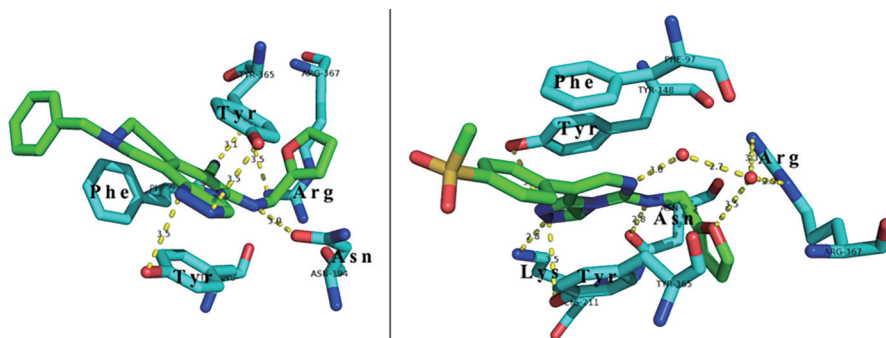


Figure 23 The binding interactions of compound **38** (left), **40** (right) and EED WD40 domain

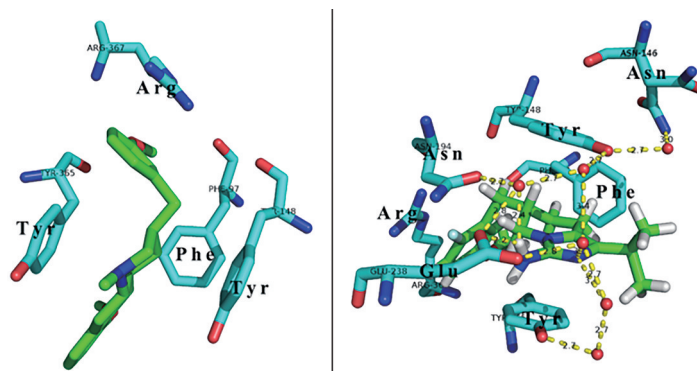


Figure 24 The binding interactions of compound 41 (left), compound 45 (right) and EED WD40 domain

时也是PRC2复合物的变构调节剂^[78]。2017年Barnash等^[79]利用组合化学和基于结构的优化策略从Jarid2-K116me3 (PDB: 4X3E) 出发得到拟肽**48** (UNC5114, $K_d = 0.68 \pm 0.05 \mu\text{mol}\cdot\text{L}^{-1}$) 和**49** (UNC5115, $K_d = 1.14 \pm 0.14 \mu\text{mol}\cdot\text{L}^{-1}$)。

2019年Read等^[80]从专利报道的EED抑制剂(**50**)出发, 获得其晶体结构, 了解其与蛋白的结合模式, 并合成衍生物。利用荧光偏振探针技术消除之前高通量筛选的假阳性化合物, 最终得到化合物**51** ($\text{IC}_{50} = 0.2 \mu\text{mol}\cdot\text{L}^{-1}$)。

2020年Rej等^[81]报道了他们的研究工作。他们基于化合物**40** ($\text{IC}_{50} = 17.6 \text{ nmol}\cdot\text{L}^{-1}$) 进行基于结构的设计与分子优化, 得到化合物**52** (EEDi-5285, $\text{IC}_{50} = 0.2 \text{ nmol}\cdot\text{L}^{-1}$, PDB: 6W7F)。在用KARPAS422细胞株进行的细胞增殖实验中, 化合物**52** ($\text{IC}_{50} = 0.5 \text{ nmol}\cdot\text{L}^{-1}$)

活性比化合物**40** ($\text{IC}_{50} = 180 \text{ nmol}\cdot\text{L}^{-1}$) 高出300多倍。

WDR5 (WD repeat domain 5) 作为WD40结构域另一个重要成员, 也开展了较多的抑制剂研发工作。WDR5通过调控MLL1甲基转移酶复合物的形成, 调节H3K4的甲基化状态, 从而对染色质重塑和MLL1基因的表达产生影响^[82], 因而被认为是重要的抗肿瘤药物靶点。针对WDR5蛋白, 目前已报道WDR5-MLL1的拟肽类^[83-85]、非肽类小分子^[86-89]抑制剂, 另外还有靶向WDR5 WBM位点的小分子抑制剂^[90,91]被报道。

7 总结

本文对甲基化识别结构域及其抑制剂进行了简要介绍, 并对相关抑制剂的优化策略以及抑制剂与蛋白的结合模式进行了分析。综合上述多个识别结构域的研究结果, 总结如下:

甲基化赖氨酸识别结构域具有鲜明的结构特点:

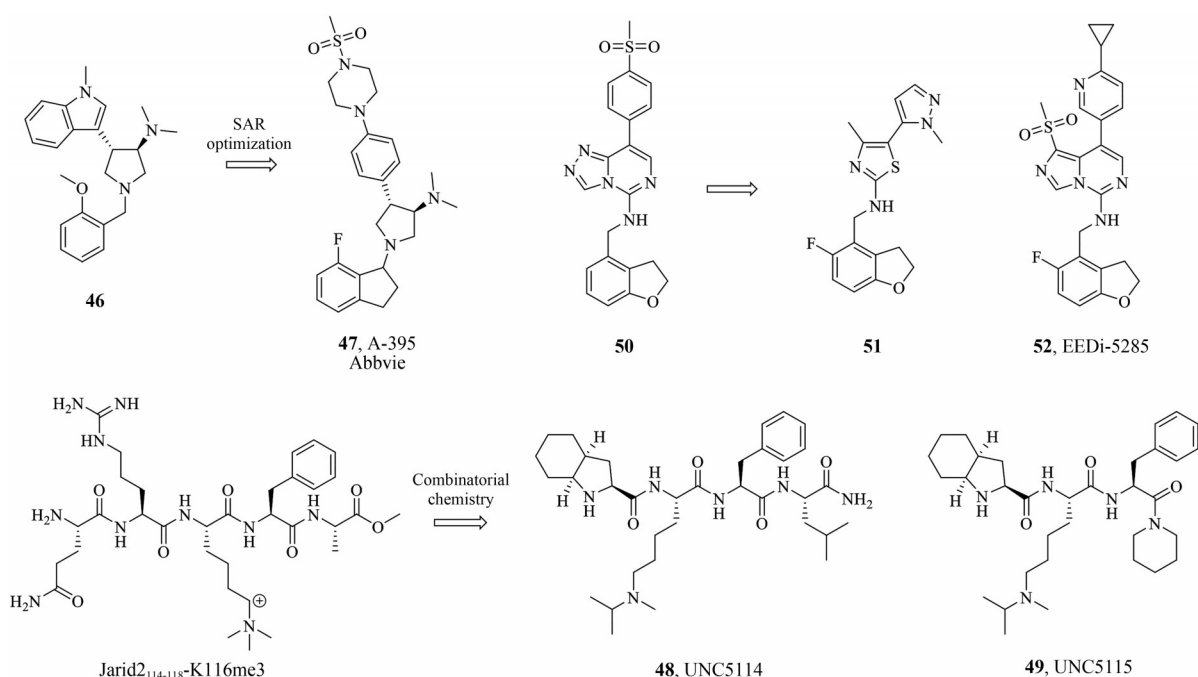


Figure 25 The optimization strategies of compounds 47, 48, 49 and 51 and the chemical structure of EEDi-5285

① 都具有由芳香环残基组成的芳香笼, 来结合并识别相应的甲基化赖氨酸残基; ② 识别 Kme0、Kme1 和 Kme2 结构域的结合口袋一般很深, 此时, 疏水残基与内源肽形成的疏水相互作用将会有利于蛋白与组蛋白的结合; ③ 识别 Kme3 结构域的结合口袋一般较浅, 这时蛋白表面的亲水残基与内源肽形成的氢键相互作用将会对识别相应修饰不可或缺。

甲基化赖氨酸识别结构域抑制剂的优化策略主要包括: ① 从与相应蛋白有较强活性的内源肽出发, 确定与蛋白结合的关键结构, 进行结构改造, 得到活性以及选择性有很大提高的拟肽类结构; ② 通过筛选获得与蛋白有结合的小分子片段, 然后进行结构优化, 特别是根据小分子与蛋白复合物的晶体结构来进行结构改造, 快速获得活性较好的小分子化合物; ③ 通过叠合或者对接的方式来分析其他蛋白的抑制剂可能与蛋白的结合方式, 再通过分析蛋白间的序列不同之处, 来合理改造化合物并实现化合物活性的“靶点跨越”。

纵览抑制剂的结构特征, 可以发现一般需要包括 4 个结构模块与蛋白进行结合, 如图 26 所示: 模拟甲基化赖氨酸残基的片段占据芳香笼; 极性基团与蛋白的亲水性残基形成氢键; 疏水性基团与蛋白的疏水残基形成疏水相互作用; 连接基团调整各个结合模块在蛋白中的位置。对于拟肽类抑制剂而言, 其主链就扮演着连接基团的角色; 有些抑制剂其他的结合模块则可能同时发挥着连接基团的作用。

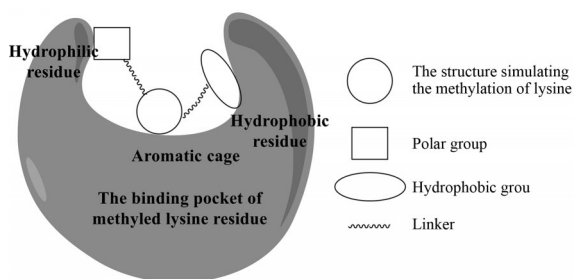


Figure 26 The schematic diagram of binding patterns of methylated lysine recognition domain inhibitors to proteins

总之, 甲基化赖氨酸识别结构域作为一种新兴的表现遗传相关的潜在靶点, 有望在多种癌症的治疗中发挥作用。然而, 目前该领域的研究与蛋白甲基化转移酶等相比还远不成熟。希望本文的综述可以为后续靶向这类蛋白的药物开发提供基础, 促进该领域的研究。

作者贡献: 周理炜负责综述的文献资料收集和主要内容撰写, 绘制图表。樊甜甜、刘同超对综述的选题和内容框架提出了重要建议并撰写、修改了部分内容。吴小余教授、熊

兵研究员对本文提出了重要的指导和修改意见。

利益冲突: 本文不涉及个人与组织利益冲突。

References

- [1] Arrowsmith CH, Bountra C, Fish PV, et al. Epigenetic protein families: a new frontier for drug discovery [J]. *Nat Rev Drug Discov*, 2012, 11: 384-400.
- [2] Allis CD, Berger SL, Cote J, et al. New nomenclature for chromatin-modifying enzymes [J]. *Cell*, 2007, 131: 633-636.
- [3] Smith BC, Denu JM. Chemical mechanisms of histone lysine and arginine modifications [J]. *Biochim Biophys Acta*, 2009, 1789: 45-57.
- [4] Taverna SD, Li H, Ruthenburg AJ, et al. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers [J]. *Nat Struct Mol Biol*, 2007, 14: 1025-1040.
- [5] Wang GG, Allis CD, Chi P. Chromatin remodeling and cancer, part I: covalent histone modifications [J]. *Trends Mol Med*, 2007, 13: 363-372.
- [6] Chi P, Allis CD, Wang GG. Covalent histone modifications--miswritten, misinterpreted and mis-erased in human cancers [J]. *Nat Rev Cancer*, 2010, 10: 457-469.
- [7] Grewal SI, Jia S. Heterochromatin revisited [J]. *Nat Rev Genet*, 2007, 8: 35-46.
- [8] Norwood LE, Moss TJ, Margaryan NV, et al. A requirement for dimerization of HP1H α in suppression of breast cancer invasion [J]. *J Biol Chem*, 2006, 281: 18668-18676.
- [9] De Koning L, Savignoni A, Boumendil C, et al. Heterochromatin protein 1 α : a hallmark of cell proliferation relevant to clinical oncology [J]. *EMBO Mol Med*, 2009, 1: 178-191.
- [10] Varier RA, Timmers HT. Histone lysine methylation and demethylation pathways in cancer [J]. *Biochim Biophys Acta*, 2011, 1815: 75-89.
- [11] Pomeroy SL, Tamayo P, Gaasenbeek M, et al. Prediction of central nervous system embryonal tumour outcome based on gene expression [J]. *Nature*, 2002, 415: 436-442.
- [12] Wasenius VM, Hemmer S, Kettunen E, et al. Hepatocyte growth factor receptor, matrix metalloproteinase-11, tissue inhibitor of metalloproteinase-1, and fibronectin are up-regulated in papillary thyroid carcinoma: a cDNA and tissue microarray study [J]. *Clin Cancer Res*, 2003, 9: 68-75.
- [13] Botuyan MV. Chapter 8-Tudor domains as methyl-lysine and methyl-arginine readers [M]//*Chromatin Signaling & Diseases* (Ed. by Binda O, Fernandez-Zapico ME). Amsterdam: Elsevier Inc., 2016: 149-165. <https://doi.org/10.1016/B978-0-12-802389-1.00008-3>.
- [14] Milosevich N, Warmerdam Z, Hof F. Structural aspects of small-molecule inhibition of methyllysine reader proteins [J]. *Future Med Chem*, 2016, 8: 1681-1702.
- [15] Friesen WJ, Massenet S, Paushkin S, et al. SMN, the product of

- the spinal muscular atrophy gene, binds preferentially to dimethylarginine-containing protein targets [J]. *Mol Cell*, 2001, 7: 1111-1117.
- [16] Guccione E, Bassi C, Casadio F, et al. Methylation of histone H3R2 by PRMT6 and H3K4 by an MLL complex are mutually exclusive [J]. *Nature*, 2007, 449: 933-937.
- [17] Migliori V, Muller J, Phalke S, et al. Symmetric dimethylation of H3R2 is a newly identified histone mark that supports euchromatin maintenance [J]. *Nat Struct Mol Biol*, 2012, 19: 136-144.
- [18] Bannister AJ, Zegerman P, Partridge JF, et al. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain [J]. *Nature*, 2001, 410: 120-124.
- [19] Lachner M, O'Carroll D, Rea S, et al. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins [J]. *Nature*, 2001, 410: 116-120.
- [20] Eissenberg JC. Structural biology of the chromodomain: form and function [J]. *Gene*, 2012, 496: 69-78.
- [21] Min J, Zhang Y, Xu RM. Structural basis for specific binding of polycomb chromodomain to histone H3 methylated at Lys 27 [J]. *Genes Dev*, 2003, 17: 1823-1828.
- [22] Kaustov L, Ouyang H, Amaya M, et al. Recognition and specificity determinants of the human cbx chromodomains [J]. *J Biol Chem*, 2011, 286: 521-529.
- [23] Scott CL, Gil J, Hernando E, et al. Role of the chromobox protein CBX7 in lymphomagenesis [J]. *Proc Natl Acad Sci U S A*, 2007, 104: 5389-5394.
- [24] Shinjo K, Yamashita Y, Yamamoto E, et al. Expression of chromobox homolog 7 (CBX7) is associated with poor prognosis in ovarian clear cell adenocarcinoma *via* TRAIL-induced apoptotic pathway regulation [J]. *Int J Cancer*, 2014, 135: 308-318.
- [25] Simhadri C, Daze KD, Douglas SF, et al. Chromodomain antagonists that target the polycomb-group methyllysine reader protein chromobox homolog 7 (CBX7) [J]. *J Med Chem*, 2014, 57: 2874-2883.
- [26] Stuckey JI, Dickson BM, Cheng N, et al. A cellular chemical probe targeting the chromodomains of polycomb repressive complex 1 [J]. *Nat Chem Biol*, 2016, 12: 180-187.
- [27] Ren C, Morohashi K, Plotnikov AN, et al. Small-molecule modulators of methyl-lysine binding for the CBX7 chromodomain [J]. *Chem Biol*, 2015, 22: 161-168.
- [28] Ren C, Smith SG, Yap K, et al. Structure-guided discovery of selective antagonists for the chromodomain of polycomb repressive protein CBX7 [J]. *ACS Med Chem Lett*, 2016, 7: 601-605.
- [29] Milosevich N, Gignac MC, McFarlane J, et al. Selective inhibition of CBX6: a methyllysine reader protein in the polycomb family [J]. *ACS Med Chem Lett*, 2016, 7: 139-144.
- [30] Northcott PA, Nakahara Y, Wu X, et al. Multiple recurrent genetic events converge on control of histone lysine methylation in medulloblastoma [J]. *Nat Genet*, 2009, 41: 465-472.
- [31] Gurvich N, Perna F, Farina A, et al. L3MBTL1 polycomb protein, a candidate tumor suppressor in del(20q12) myeloid disorders, is essential for genome stability [J]. *Proc Natl Acad Sci U S A*, 2010, 107: 22552-22557.
- [32] Perna F, Gurvich N, Hoya-Arias R, et al. Depletion of L3MBTL1 promotes the erythroid differentiation of human hematopoietic progenitor cells: possible role in 20q-polycythemia vera [J]. *Blood*, 2010, 116: 2812-2821.
- [33] Honda H, Takubo K, Oda H, et al. HEMP, an mbt domain-containing protein, plays essential roles in hematopoietic stem cell function and skeletal formation [J]. *Proc Natl Acad Sci U S A*, 2011, 108: 2468-2473.
- [34] Addou-Klouché L, Adelaide J, Finetti P, et al. Loss, mutation and deregulation of L3MBTL4 in breast cancers [J]. *Mol Cancer*, 2010, 9: 213.
- [35] Herold JM, Wigle TJ, Norris JL, et al. Small-molecule ligands of methyl-lysine binding proteins [J]. *J Med Chem*, 2011, 54: 2504-2511.
- [36] Herold JM, James LI, Korboukh VK, et al. Structure-activity relationships of methyl-lysine reader antagonists [J]. *MedChemComm*, 2012, 3: 45-51.
- [37] Bonasio R, Lecona E, Reinberg D. MBT domain proteins in development and disease [J]. *Semin Cell Dev Biol*, 2010, 21: 221-230.
- [38] James LI, Barsyte-Lovejoy D, Zhong N, et al. Discovery of a chemical probe for the L3MBTL3 methyllysine reader domain [J]. *Nat Chem Biol*, 2013, 9: 184-191.
- [39] Camerino MA, Zhong N, Dong A, et al. The structure-activity relationships of L3MBTL3 inhibitors: flexibility of the dimer interface [J]. *MedChemComm*, 2013, 4: 1501-1507.
- [40] James LI, Korboukh VK, Krichevsky L, et al. Small-molecule ligands of methyl-lysine binding proteins: optimization of selectivity for L3MBTL3 [J]. *J Med Chem*, 2013, 56: 7358-7371.
- [41] Pek JW, Anand A, Kai T. Tudor domain proteins in development [J]. *Development*, 2012, 139: 2255-2266.
- [42] Chen C, Nott TJ, Jin J, et al. Deciphering arginine methylation: Tudor tells the tale [J]. *Nat Rev Mol Cell Biol*, 2011, 12: 629-642.
- [43] Iwabuchi K, Basu BP, Kysela B, et al. Potential role for 53BP1 in DNA end-joining repair through direct interaction with DNA [J]. *J Biol Chem*, 2003, 278: 36487-36495.
- [44] Kachirskaja I, Shi X, Yamaguchi H, et al. Role for 53BP1 Tudor domain recognition of p53 dimethylated at lysine 382 in DNA damage signaling [J]. *J Biol Chem*, 2008, 283: 34660-34666.
- [45] Ward IM, Minn K, van Deursen J, et al. p53 binding protein 53BP1 is required for DNA damage responses and tumor suppression in mice [J]. *Mol Cell Biol*, 2003, 23: 2556-2563.
- [46] Perfetti MT, Baughman BM, Dickson BM, et al. Identification of a fragment-like small molecule ligand for the methyl-lysine binding protein, 53BP1 [J]. *ACS Chem Biol*, 2015, 10: 1072-1081.
- [47] Su X, Zhu G, Ding X, et al. Molecular basis underlying histone

- H3 lysine-arginine methylation pattern readout by Spin/Ssty repeats of Spindlin1 [J]. *Genes Dev*, 2014, 28: 622-636.
- [48] Sweis RF, Pliushchev M, Brown PJ, et al. Discovery and development of potent and selective inhibitors of histone methyltransferase g9a [J]. *ACS Med Chem Lett*, 2014, 5: 205-209.
- [49] Wagner T, Greschik H, Burgahn T, et al. Identification of a small-molecule ligand of the epigenetic reader protein spindlin1 *via* a versatile screening platform [J]. *Nucleic Acids Res*, 2016, 44: e88.
- [50] Bae N, Viviano M, Su X, et al. Developing spindlin1 small-molecule inhibitors by using protein microarrays [J]. *Nat Chem Biol*, 2017, 13: 750-756.
- [51] Xiong Y, Greschik H, Johansson C, et al. Discovery of a potent and selective fragment-like inhibitor of methyllysine reader protein spindlin 1 (SPIN1) [J]. *J Med Chem*, 2019, 62: 8996-9007.
- [52] Fagan V, Johansson C, Gileadi C, et al. A chemical probe for tudor domain protein spindlin1 to investigate chromatin function [J]. *J Med Chem*, 2019, 62: 9008-9025.
- [53] Upadhyay AK, Judge RA, Li L, et al. Targeting lysine specific demethylase 4A (KDM4A) tandem TUDOR domain - a fragment based approach [J]. *Bioorg Med Chem Lett*, 2018, 28: 1708-1713.
- [54] Mader P, Mendoza-Sanchez R, Iqbal A, et al. Identification and characterization of the first fragment hits for SETDB1 Tudor domain [J]. *Bioorg Med Chem*, 2019, 27: 3866-3878.
- [55] Senisterra G, Zhu HY, Luo X, et al. Discovery of small-molecule antagonists of the H3K9me3 binding to UHRF1 tandem Tudor domain [J]. *SLAS Discov*, 2018, 23: 930-940.
- [56] Liu J, Zhang S, Liu M, et al. Structural plasticity of the TDRD3 Tudor domain probed by a fragment screening hit [J]. *FEBS J*, 2018, 285: 2091-2103.
- [57] Stec I, Wright TJ, van Ommen GJ, et al. WHSC1, a 90 kb SET domain-containing gene, expressed in early development and homologous to a Drosophila dysmorphia gene maps in the Wolf-Hirschhorn syndrome critical region and is fused to IgH in t(4;14) multiple myeloma [J]. *Hum Mol Genet*, 1998, 7: 1071-1082.
- [58] Qin S, Min J. Structure and function of the nucleosome-binding PWWP domain [J]. *Trends Biochem Sci*, 2014, 39: 536-547.
- [59] Böttcher J, Dilworth D, Reiser U, et al. Fragment-based discovery of a chemical probe for the PWWP1 domain of NSD3 [J]. *Nat Chem Biol*, 2019, 15: 822-829.
- [60] Ferreira de Freitas R, Liu Y, Szcweczyk MM, et al. Discovery of small-molecule antagonists of the PWWP domain of NSD2 [J]. *J Med Chem*, 2021, 64: 1584-1592.
- [61] Sanchez R, Zhou MM. The PHD finger: a versatile epigenome reader [J]. *Trends Biochem Sci*, 2011, 36: 364-372.
- [62] Thompson B, Townsley F, Rosin-Arbesfeld R, et al. A new nuclear component of the Wnt signalling pathway [J]. *Nat Cell Biol*, 2002, 4: 367-373.
- [63] Wang GG, Song J, Wang Z, et al. Haematopoietic malignancies caused by dysregulation of a chromatin-binding PHD finger [J]. *Nature*, 2009, 459: 847-851.
- [64] Gu L, Frommel SC, Oakes CC, et al. BAZ2A (TIP5) is involved in epigenetic alterations in prostate cancer and its overexpression predicts disease recurrence [J]. *Nat Genet*, 2015, 47: 22-30.
- [65] Wagner EK, Nath N, Flemming R, et al. Identification and characterization of small molecule inhibitors of a plant homeodomain finger [J]. *Biochemistry*, 2012, 51: 8293-8306.
- [66] Amato A, Lucas X, Bortoluzzi A, et al. Targeting ligandable pockets on plant homeodomain (PHD) zinc finger domains by a fragment-based approach [J]. *ACS Chem Biol*, 2018, 13: 915-921.
- [67] Miller TC, Rutherford TJ, Birchall K, et al. Competitive binding of a benzimidazole to the histone-binding pocket of the Pygo PHD finger [J]. *ACS Chem Biol*, 2014, 9: 2864-2874.
- [68] Xu C, Min J. Structure and function of WD40 domain proteins [J]. *Protein Cell*, 2011, 2: 202-214.
- [69] Jain BP, Pandey S. WD40 repeat proteins: signalling scaffold with diverse functions [J]. *Protein J*, 2018, 37: 391-406.
- [70] Beltran M, Yates CM, Skalska L, et al. The interaction of PRC2 with RNA or chromatin is mutually antagonistic [J]. *Genome Res*, 2016, 26: 896-907.
- [71] Li L, Zhang H, Zhang M, et al. Discovery and molecular basis of a diverse set of polycomb repressive complex 2 inhibitors recognition by EED [J]. *PLoS One*, 2017, 12: e0169855.
- [72] Huang Y, Zhang J, Yu Z, et al. Discovery of first-in-class, potent, and orally bioavailable embryonic ectoderm development (EED) inhibitor with robust anticancer efficacy [J]. *J Med Chem*, 2017, 60: 2215-2226.
- [73] Qi W, Zhao K, Gu J, et al. An allosteric PRC2 inhibitor targeting the H3K27me3 binding pocket of EED [J]. *Nat Chem Biol*, 2017, 13: 381-388.
- [74] Lingel A, Sendzik M, Huang Y, et al. Structure-guided design of EED binders allosterically inhibiting the epigenetic polycomb repressive complex 2 (PRC2) methyltransferase [J]. *J Med Chem*, 2017, 60: 415-427.
- [75] Curtin ML, Pliushchev MA, Li HQ, et al. SAR of amino pyrrolidines as potent and novel protein-protein interaction inhibitors of the PRC2 complex through EED binding [J]. *Bioorg Med Chem Lett*, 2017, 27: 1576-1583.
- [76] Wang Y, Edalji RP, Panchal SC, et al. Are we there yet? Applying thermodynamic and kinetic profiling on embryonic ectoderm development (EED) hit-to-lead program [J]. *J Med Chem*, 2017, 60: 8321-8335.
- [77] He Y, Selvaraju S, Curtin ML, et al. The EED protein-protein interaction inhibitor A-395 inactivates the PRC2 complex [J]. *Nat Chem Biol*, 2017, 13: 389-395.
- [78] Sanulli S, Justin N, Teissandier A, et al. Jarid2 methylation *via* the PRC2 complex regulates H3K27me3 deposition during cell differentiation [J]. *Mol Cell*, 2015, 57: 769-783.

- [79] Barnash KD, The J, Norris-Drouin JL, et al. Discovery of peptidomimetic ligands of EED as allosteric inhibitors of PRC2 [J]. *ACS Comb Sci*, 2017, 19: 161-172.
- [80] Read JA, Tart J, Rawlins PB, et al. Rapid identification of novel allosteric PRC2 inhibitors [J]. *ACS Chem Biol*, 2019, 14: 2134-2140.
- [81] Rej RK, Wang C, Lu J, et al. EEDi-5285: an exceptionally potent, efficacious, and orally active small-molecule inhibitor of embryonic ectoderm development [J]. *J Med Chem*, 2020, 63: 7252-7267.
- [82] Dou Y, Milne TA, Ruthenburg AJ, et al. Regulation of MLL1 H3K4 methyltransferase activity by its core components [J]. *Nat Struct Mol Biol*, 2006, 13: 713-719.
- [83] Karatas H, Townsend EC, Bernard D, et al. Analysis of the binding of mixed lineage leukemia 1 (MLL1) and histone 3 peptides to WD repeat domain 5 (WDR5) for the design of inhibitors of the MLL1-WDR5 interaction [J]. *J Med Chem*, 2010, 53: 5179-5185.
- [84] Karatas H, Townsend EC, Cao F, et al. High-affinity, small-molecule peptidomimetic inhibitors of MLL1/WDR5 protein-protein interaction [J]. *J Am Chem Soc*, 2013, 135: 669-682.
- [85] Karatas H, Li Y, Liu L, et al. Discovery of a highly potent, cell-permeable macrocyclic peptidomimetic (MM-589) targeting the wd repeat domain 5 protein (WDR5)-mixed lineage leukemia (MLL) protein-protein interaction [J]. *J Med Chem*, 2017, 60: 4818-4839.
- [86] Orlicky S, Tang X, Neduva V, et al. An allosteric inhibitor of substrate recognition by the SCF(Cdc4) ubiquitin ligase [J]. *Nat Biotechnol*, 2010, 28: 733-737.
- [87] Bolshan Y, Getlik M, Kuznetsova E, et al. Synthesis, optimization, and evaluation of novel small molecules as antagonists of WDR5-MLL interaction [J]. *ACS Med Chem Lett*, 2013, 4: 353-357.
- [88] Senisterra G, Wu H, Allali-Hassani A, et al. Small-molecule inhibition of MLL activity by disruption of its interaction with WDR5 [J]. *Biochem J*, 2013, 449: 151-159.
- [89] Getlik M, Smil D, Zepeda-Velazquez C, et al. Structure-based optimization of a small molecule antagonist of the interaction between WD repeat-containing protein 5 (WDR5) and mixed-lineage leukemia 1 (MLL1) [J]. *J Med Chem*, 2016, 59: 2478-2496.
- [90] Aho ER, Wang J, Gogliotti RD, et al. Displacement of WDR5 from chromatin by a WIN site inhibitor with picomolar affinity [J]. *Cell Rep*, 2019, 26: 2916-2928e13.
- [91] Tian J, Teuscher KB, Aho ER, et al. Discovery and structure-based optimization of potent and selective WD repeat domain 5 (WDR5) inhibitors containing a dihydroisoquinolinone bicyclic core [J]. *J Med Chem*, 2020, 63: 656-675.