



G四链体互作蛋白及其研究方法进展

黄海燕, 周翔*

武汉大学化学与分子科学学院, 武汉 430072

* 联系人, E-mail: xzhou@whu.edu.cn

2023-02-26 收稿, 2023-03-29 修回, 2023-04-17 接受, 2023-04-24 网络版发表

国家自然科学基金(22107086, 22037004, 21721005, 92153303)和中国博士后科学基金(2021M692468)资助

摘要 G四链体是一种非经典的核酸二级结构, 并且已经被证实大量存在于多种生物基因组及转录组中的关键调节位点. 由于其特殊的四链结构, G四链体能够与多种功能的蛋白发生相互作用, 从而参与生物体多种生理及病理过程的调控. 因此, G四链体互作蛋白的结构及功能研究将有利于进一步揭示G四链体的生命活动调节机制, 促进开发G四链体及其互作蛋白作为疾病治疗靶点的新型治疗方法. 近年来, 得益于核酸-蛋白互作研究方法的进步及G四链体互作蛋白研究新技术的开发, 科学家已经发现了多种功能各异的G四链体互作蛋白. 本文简要介绍了已知G四链体互作蛋白的种类和功能, 综合评述近年来G四链体互作蛋白研究方法的发展, 并对未来的发展方向进行展望.

关键词 G四链体, G四链体互作蛋白, 解旋酶, 复制, 转录, 端粒

除了经典的DNA双螺旋结构, 生物体内还存在着多种不同的非经典二级结构, 共同调节着生物体的性状及遗传. 其中, G四链体(G-quadruplex, G4)由于其特殊的四链结构及重要的生理功能, 从发现之初就受到科学家的广泛关注^[1,2]. G四链体是一种具有多态性的核酸二级结构, 通常在细胞中的单链DNA和RNA区域更容易形成这种结构. 在生物体内, G四链体的形成、稳定、解开都被严格且精细地控制着. 研究发现, 生物体内存在大量与G四链体相互作用的蛋白, 这些蛋白通过对G四链体的识别和作用控制众多生理或者病理通路的关键步骤. 对G四链体互作蛋白的氨基酸组成、结构特性以及作用方式的研究, 有助于G四链体识别机制的揭示, 也为将G四链体及其互作蛋白作为药物靶点提供结构基础. 对G四链体互作蛋白时空表达的广泛研究将为药物设计提供新的靶点, 为疾病新型治疗方法的开发奠定基础.

然而, G四链体互作蛋白的鉴定存在着一些难点.

首先, 由于G四链体形成的动态性和细胞环境的复杂性, 体外G四链体与蛋白质之间的相互作用可能无法模拟体内真实发生的事件. 其次, G四链体和蛋白复合物三维结构的精细解析难度较大, 而这正是计算机模拟建模和G四链体精确靶向的关键. G四链体互作蛋白研究方法及技术的开发将帮助解决这些问题. 本文首先介绍G四链体的结构和功能、G四链体互作蛋白的主要分类及功能, 然后系统介绍G四链体互作蛋白的检测方法及技术, 并分析了各种方法的优势及存在的缺点. 最后, 我们讨论了如何在现有研究方法的基础上, 改进和开发更为有力的G四链体互作蛋白鉴定技术, 从而为G四链体及其互作蛋白的深入研究及科学应用提供思路.

1 G四链体

G四分体(G-quartet)是由4个鸟嘌呤碱基之间两两形成的4个Hoogsteen氢键连接而成的平面结构^[1], 是组

引用格式: 黄海燕, 周翔. G四链体互作蛋白及其研究方法进展. 科学通报, 2023, 68: 3887-3898

Huang H Y, Zhou X. Progress in the study of G-quadruplex interacting proteins (in Chinese). Chin Sci Bull, 2023, 68: 3887-3898, doi: 10.1360/TB-2023-0180

成G四链体的基本单元,多个G四分体之间通过 π - π 堆积作用形成特殊核酸二级结构称为G四链体^[3].每两个G四分体平面的空腔由于G平面上鸟嘌呤碱基上的羰基形成强烈的电负性,一些一价阳离子(K^+ 、 Na^+ 、 Li^+ 、 NH^+ 等)及二价阳离子(Ca^{2+} 、 Sr^{2+} 、 Ba^{2+} 等)能够进入空腔,并稳定G四链体结构^[4],如图1所示.由于loop环的长度和方向、组成G四链体的G四分体数量以及中心离子的不同,G四链体能够形成不同的拓扑结构^[5].例如,DNA G四链体可以形成平行、反平行或混合结构;而RNA G四链体由于糖苷键的反平行构象,通常形成平行结构,而且具有更高的热稳定性.G四链体可以被一些化学小分子高亲和力结合,我们将这类小分子称为G四链体配体.这些配体通常通过与G四分体平面、沟槽或者loop环结合等方式结合G四链体^[6,7].

研究人员通过对 $G_XN_{1-7}G_XN_{1-7}G_XN_{1-7}G_XN_{1-7}$ ($X \geq 3$, N可以是任意碱基)的算法分析,获得潜在G四链体位点(putative G-quadruplex forming sequences, PQS)^[8].通过算法预测的G四链体位点还需要通过实验验证,以进一步确定该位点是否真实形成G四链体结构.这些方法通常包括圆二色谱法^[9]、核磁法^[10]、晶体法^[11]等.此外,

得益于基因测序技术的高速发展,科学家已经开发了多种G四链体测序技术,如基于DNA聚合酶停滞的G4-seq法^[12]、基于逆转录停滞的rG4-seq法^[13]、基于染色质免疫共沉淀测序法(chromatin immunoprecipitation followed by sequencing, ChIP-seq)的G4 ChIP-seq法^[14]以及基于重组蛋白技术的G4P人工蛋白富集测序法等^[15].而且越来越多的技术将适用于高精度、高灵敏度以及原位的G四链体检测,如最近开发的CUT&Tag技术能够实现在单细胞水平中对G四链体进行原位检测^[16].上述方法对G四链体在基因组级转录组中分布位置的鉴定也为后续G四链体功能的研究提供了基础.研究表明,G四链体在体内的分布不是随机的,而且大部分G四链体分布在基因组的重要功能区域,如端粒区域、启动子区域、转录起始位点等^[17].由此可见,G四链体结构与基因组的功能存在密切关联.

2 G四链体互作蛋白及其功能

G四链体的稳定性受到多种因素影响,包括序列本身的特性、离子环境、基因组开放度以及周围蛋白的作用等^[18].虽然G4序列在细胞中广泛存在,但并不是所

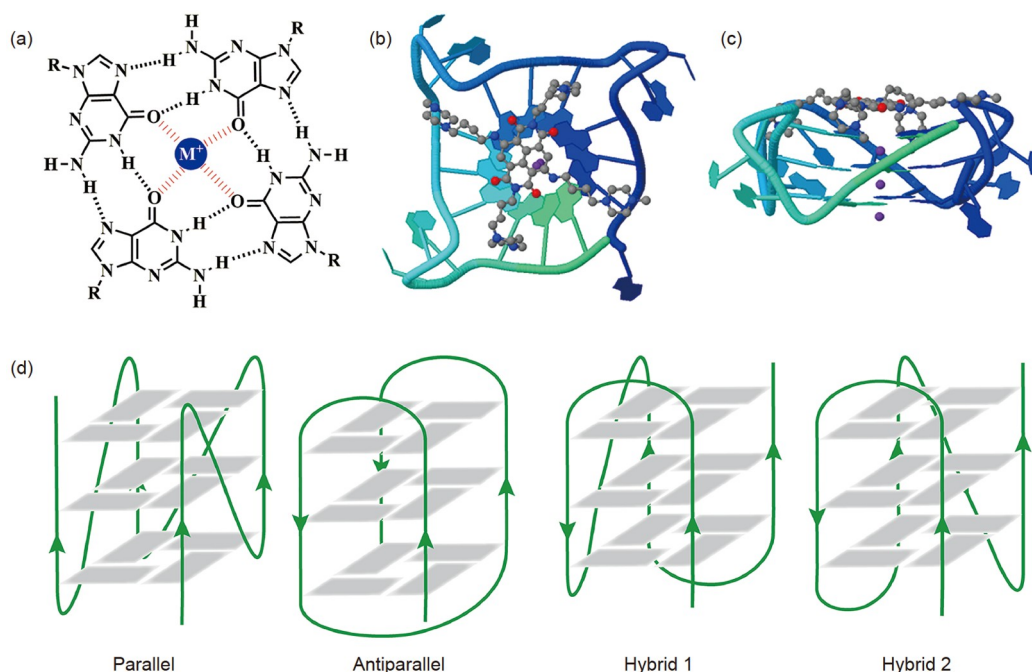


图1 G四分体及G四链体结构. (a) G四分体的Hoogsteen氢键连接方式,中间的 M^+ 金属离子稳定G四分体结构;人端粒DNA G四链体的晶体结构(Protein Data Bank: 3T5E)的俯视图(b)和正视图(c). (d) 一些经典构象的G四链体结构示意图

Figure 1 The structure of G-quartet and G-quadruplex. (a) Hoogsteen base-pairing of G-quartet and the central cation (M^+) stabilized G-quartet. Crystal structure of human telomeric DNA G4s (Protein Data Bank: 3T5E): (b) Top view; (c) side view. (d) Schematic illustration of some classical G4 topologies

有的G4序列都能在体内形成G四链体结构,而且不同细胞类型以及不同细胞状态下的G四链体结构具有很大的异质性^[17,19]。因此,了解G四链体结构在细胞环境中如何被诱导和稳定,是一个至关重要的问题。根据G四链体互作蛋白的功能,可以将细胞中的G四链体互作蛋白分为3种:G四链体结合蛋白(G4BP)、G四链体解旋蛋白以及被G四链体招募的功能性蛋白(图2)。已鉴定的G四链体互作蛋白已有详细的介绍^[20~22]。本文主要通过列举G四链体参与的主要生理过程及具有代表性的G四链体互作蛋白,以说明G四链体互作蛋白的不同作用方式及其功能。

2.1 DNA G四链体互作蛋白

由于常见的DNA聚合酶无法通过G四链体结构形成的位点,G四链体的异常堆积会造成基因组表达异常,从而对机体造成严重的损伤。因此,为了保持正常的生理活动,细胞中表达着多种多样的G四链体解旋酶。解旋酶占据了很大一部分已鉴定的G四链体互作蛋白,许多典型的解旋酶家族成员被证明是G四链体解旋蛋白^[23,24],如RecQ家族、DEAD-box或者DEAH-box解旋酶家族。RecQ家族成员BLM解旋酶(bloom helicase)是第一个被验证的G四链体解旋酶,可以解开分子内和分子间G四链体结构^[25~27]。G四链体解旋酶的表达异常可能导致细胞异常增殖,引起癌症等疾病的发生^[28,29]。与正常细胞相比,BLM缺陷的细胞中基因转录起始位点(transcription start site, TSS)的G四链体结构大量堆积,并且主要堆积在mRNA的第一个内含子中,该缺陷引发布鲁姆综合征^[30]。

复制叉前进中遇到的G4结构有可能阻止DNA聚合酶的通过,从而导致复制叉的阻滞,这可能造成基因组的缺失和突变。G四链体结构及其相关蛋白在复制机制调节过程中发挥重要作用(图2)。其中,G四链体结合蛋白Rif1(replication timing regulatory factor 1)通过在核膜附近生成高阶染色质结构和招募蛋白磷酸酶来决定基因组的哪些片段在细胞核中何时何地复制^[31]。FANCD1(fanconi anemia group J protein)参与复制叉形成过程中的G四链体解旋,以保证DNA聚合酶的顺利前进,FANCD1缺陷将使得复制叉上的G四链体堆积,造成复制的停滞,并最终造成DNA损伤^[32,33]。

基因组转录过程中的G四链体结构具有双重作用(图3)。最近的研究也表明,启动子中的G四链体结构是多种转录因子的结合位点,这些与G4相互作用的转录

因子造成了细胞转录组的特异性。多种转录因子已经被证明是G四链体互作蛋白。研究发现,人类基因组的启动子附近分布有大量的潜在G四链体位点,这表明G4可能在基因的表达过程中起着至关重要的作用。G四链体基序在原癌基因启动子上出现的频率要高于管家基因和抑癌基因^[34]。近年来,原癌基因的启动子G四链体基序得到了最为深入的研究,结果发现大量原癌基因启动子能够在生理环境下形成G4结构,并且调控癌症的发生和发展。致癌基因KRAS的异常表达将导致胰腺癌的发生,KRAS启动子位置可以折叠形成G四链体结构。恶性胰腺癌患者的体内通常APE1蛋白(DNA-(apurinic or apyrimidinic site) endonuclease)高表达与KRAS启动子G四链体结构结合并稳定其结构,导致KRAS的异常表达。APE1的下调降低MAZ(Myc-associated zinc finger protein)转录因子装载到KRAS启动子上,从而降低PDAC细胞中KRAS的表达^[35]。具有RNA识别结构域(RNA recognition motif, RRM)的蛋白LARK是一种近来发现的G4折叠蛋白^[36]。研究表明,家蚕的LARK蛋白能够与转录因子POUM2(POU domain protein 2)基因启动子上的G四链体结构结合,在结合后促进G四链体结构的形成和稳定,而且促进了靶基因的转录^[37]。

除了解旋酶,有趣的是,研究者在实验过程中还发

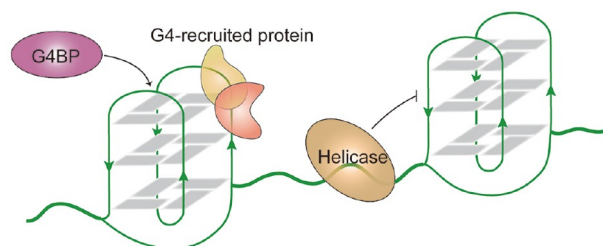


图2 G四链体互作蛋白的分类
Figure 2 The classification of G4-interacting proteins

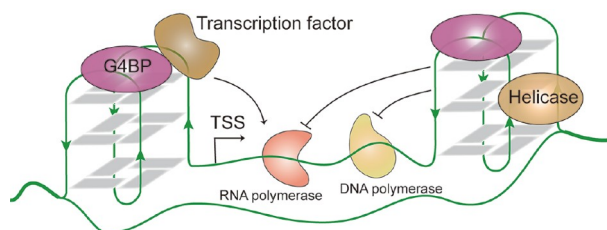


图3 G四链体互作蛋白在DNA复制和转录中的调控作用
Figure 3 Regulatory roles of G4-interacting proteins in DNA replication and transcription

现了一些对G四链体结构的稳定和解旋表现出双重作用的蛋白,如拓扑异构酶.耐辐射奇球菌的拓扑异构酶1B在体外解旋预先形成的分子内G四链体,而且促进其形成分子间G四链体,但是研究者仅通过跑胶快慢来表征形成的是哪种G四链体,具有一定的局限,其具体的作用方式有待进一步研究^[38].这种对分子间G四链体和分子内G四链体具有选择性结合的蛋白还有科凯恩综合征B蛋白(cockayne syndrome B protein, CSB),CSB的缺失将导致早衰.该蛋白能够特异性结合,并解开核糖体DNA上的分子间G四链体结构,然而对分子内G四链体却没有该功能^[39].此外,Vimentin蛋白被报道具有对G四链体重复序列的高度亲和性,而不与单个G四链体结合^[40].

人类端粒DNA序列是最早被发现的G4结构之一,其末端保守的TTAGGG回文重复序列极易形成G4结构,使用G4抗体及配体成像也证实了端粒区域在细胞环境中形成G4结构^[19,41].端粒蛋白复合体不仅能够防止核酶对染色体末端序列的降解以及DNA修复机制将它们识别为断裂片段,而且能够调节并维持端粒长度,防止由于端粒异常引起的癌症发生.哺乳动物端粒DNA被端粒蛋白复合体(shelterin)结合,端粒蛋白复合体中的端粒保护蛋白POT1(Protection of telomeres protein 1)已经被证实了能够与端粒G4相互作用,并在维持染色体结构和稳定性方面起着至关重要的作用^[42,43].在端粒酶存在的情况下,端粒通过常规复制机制和端粒酶之间的密切配合进行复制.未解开的端粒G4结构也可以抑制端粒酶活性,导致端粒缩短.因此,细胞中存在许多G4解旋酶能够与端粒上的G四链体结合并打开这些结构,包括DNA解旋酶以及单链DNA结合蛋白(图4).如人类单链结合蛋白CST(CTC1-STN1-TEN1)以及RPA(replication protein A)可以与端粒G四链体结合并打开G四链体结构,帮助解决端粒的复制压力^[44-46].

一些DNA解旋酶,如Pif、BLM、WRN、RTEL1和DNA2等也都具有解开端粒G四链体的能力^[47].

2.2 RNA G四链体互作蛋白

相比于DNA,由于其单链存在的特征, RNA在生理条件下更易形成G4结构.通过高通量测序^[13]及荧光成像技术^[48]也已经证实了细胞中大量存在RNA G四链体(rG4)结构,而且rG4在主要富集分布在5'-UTR(5'-untranslated region)和3'-UTR(3'-untranslated region)区域.研究表明RNA G四链体在选择性剪接、翻译调控、转录终止等过程中发挥着重要作用^[49,50].因此,参与这些生理过程的多种蛋白也已经被证明通过与G四链体的相互作用调控相关的生理过程,例如核糖体蛋白和剪接因子等.

G4基序在mRNA翻译过程中的作用取决于G4基序的位置和相关的结合蛋白.在5'-UTR和开放阅读框(open reading frame, ORF)区域的G4四链体结构被认为是翻译抑制剂,阻碍核糖体的扫描和移位.因此,一些解旋酶和特定的结合蛋白在转录过程中帮助打开这些G四链体结构(图5).RNA解旋酶DHX36已被表明具有这一功能,并被证明通过解开G4结构来协助*Nkx2-5* mRNA的翻译^[51].脆性X染色体综合征,以及自闭症谱系障碍的一些表现,是由于缺乏脆性X智力迟钝蛋白(fragile X mental retardation protein, FMRP)导致的RNA调控不当所致,FMRP蛋白可以与自身mRNA编码区的G四链体结构结合,从而负反馈调节自身的翻译^[52].同样,在ORF区域的G四链体结构也被认为是聚合酶延伸的障碍.CNBP(cellular nucleic acid binding protein)蛋白可以阻止其mRNA结合位点ORF区域的G四链体结构形成^[53].当CNBP缺失时,核糖体密度在富G结合位点附近显著增加,而下游约200 nt的核糖体密度急剧下降,这表明CNBP的缺失使核糖体在富含G的

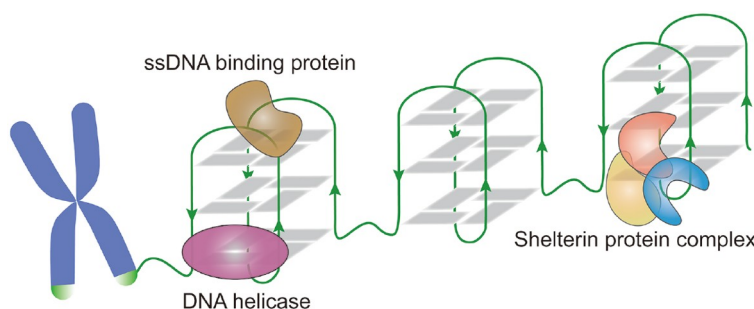


图4 端粒相关的G四链体互作蛋白及其分类
Figure 4 Telomere associated G4-interacting proteins and their classification

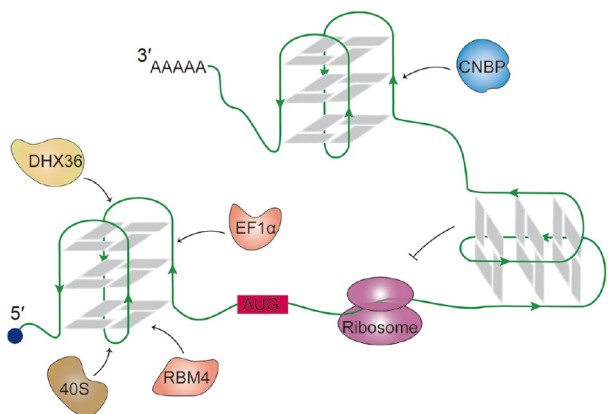


图 5 G四链体交互蛋白在mRNA翻译中的调控作用
Figure 5 Regulatory roles of G4-interacting proteins in mRNA translation

位点停滞。

上述G四链体对翻译的抑制作用主要是基于结构效应，它们作为屏障干扰翻译起始、延伸、移位等一系列相关过程。在这些过程中，翻译机制和蛋白质因子需要克服G四链体障碍，因此，参与该过程的通常是具有解旋酶活性的G四链体交互蛋白。

WTAP(Wilms' tumor 1-associating protein)复合体通过RNA G四链体结合蛋白与位于盒式外显子侧边内含子上的RNA G四链体序列相互作用被招募到剪接位点(或通过WTAP复合体的其他RNA结合蛋白被招募到其他区域)，并通过RNA G四链体序列干扰剪接位点的识别，从而导致外显子跳跃^[54]。

此外，RNA G四链体结构还可以通过协助一些翻译相关的蛋白因子与mRNA的结合或定位来影响mRNA的翻译。G4基序对帽独立翻译(Cap-independent translation)启动的调节作用就是基于这种效应。帽独立翻译在很大程度上取决于mRNA 5'-UTR中的内部核糖

体入口位点(the internal ribosome entry site, IRESs)。据报道，人血管内皮生长因子(human vascular endothelial growth factor, hVEGF) mRNA有两个IRESs, IRES-A和IRES-B, 而IRES-A具有RNA G4基序。当G4基序通过突变被破坏时，IRES-A的翻译启动活性被完全抑制。随后的一项研究提出，这一过程需要G4基序招募40S核糖体亚基。RBM4(RNA-binding protein 4)也随后被报道参与G4介导的IRES表达^[55]。G4相关的VEGF IRES活性对G4配体敏感，表明G4基序对VEGF表达的影响以及医学研究的潜力。

在应激条件下，科学家观察到G4辅助的延伸因子EF1a(elongation factor 1-alpha)与mRNA的结合。在H₂O₂处理引起的氧化应激下，EF1a通过结合核因子NRF2(nuclear factor (erythroid-derived 2)-like 2) mRNA的5'-UTR中的G四链体结构来促进翻译。由于NRF2对于防止组织损伤至关重要，NRF2 mRNA中的G四链体结构可能通过与EF1a相互作用的翻译调节细胞防御机制。上述例子表明RNA G四链体结构在mRNA翻译中发挥着复杂的调节作用，由于这些结构的位阻效应以及与相关蛋白质的相互作用对翻译的双重效应，导致翻译结果的多样化。

如图6所示，一些G4交互蛋白在核酸表观遗传修饰的调节过程中发挥重要作用。例如，DNA甲基转移酶(DNA methyltransferases, DNMTs)在哺乳动物细胞中催化CpG岛中5-甲基胞嘧啶(5-methylcytosine, 5mC)的形成，而且大多数DNA甲基转移酶的结合位点为CpG岛中G4潜在位点，这预示着G4可能通过DNA甲基转移酶的招募调节DNA甲基化水平。研究表明，DNMTs可能被G四链体捕获，从而降低附近CpG岛的甲基化水平^[56]。同样地，RNA甲基转移酶METTL14(methyltransferase-like 14)也被证明能够特异性与G四链体结构结

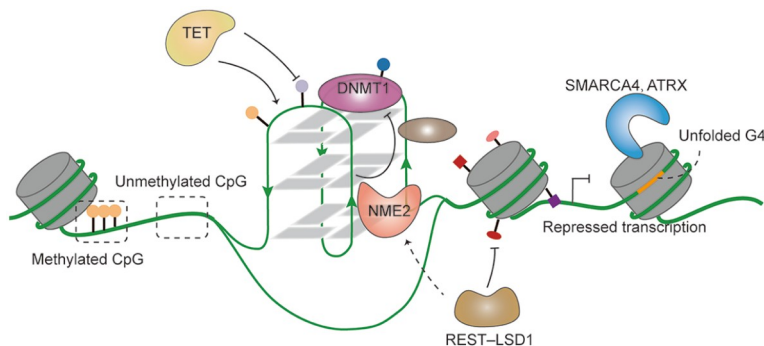


图 6 G四链体交互蛋白在表观遗传修饰中的调控作用
Figure 6 Regulatory roles of G4-interacting proteins in epigenetic modification

合^[57]。相反, DNA甲基化也可以影响G4基序的拓扑结构, 从而调节其他G4互作蛋白的结合^[58]。例如, 在TET (ten eleven translocation)去甲基化酶缺失的人弥漫性大B细胞淋巴瘤中, 科学家观测到G4结构堆积引起的稳态干扰, 而这种干扰在对DNMT1进行敲低后得到了缓解, 这说明甲基化可能对G4结构起去稳定作用^[59]。

一些G4互作蛋白在组蛋白甲基化及染色质重塑过程中发挥重要作用。例如, 染色质重塑蛋白ATRX的结合位点集中在富含GC的CpG岛上, 该位点富含潜在的G4基序, 而且在ATRX低表达的恶性胶质瘤细胞中更易形成由G4堆积引起的复制压力及DNA损伤^[60]。这表明G4的选择性稳定可能是一种有前景的用于ARTX突变肿瘤的治疗策略。最近, Shankar课题组^[61]在细胞中验证了SMARCA4蛋白能够与基因组启动子区域的G四链体结构相互作用, 该蛋白是染色质重塑复合物BAF的核心马达亚基, 可能参与基因组转录的调控。启动子G四链体及其相关蛋白如NME2, 可以募集REST-LSD1(RE1沉默转录因子与赖氨酸特异性组蛋白去甲基化酶的复合物)以去除组蛋白(H3K4)的基因激活甲基化, 并抑制基因表达^[62]。

3 G4互作蛋白的研究技术和方法

不同的细胞类型和染色质状态下, G四链体结构是高度变化的。同时, G4结构在整个基因组和转录组中的形成与解开都受到G4互作蛋白的直接或间接调控, 从而影响多种生物学过程。因此, 对G4互作蛋白的鉴定和深入研究有助于全面解释G4和蛋白的相互作用机制及其在体内的生物学作用。

3.1 G四链体与蛋白相互作用的研究技术

蛋白与G四链体的作用方式及不同的调控过程可以通过化学生物学实验进行验证。G四链体与相关蛋白相互作用的研究技术主要涉及电泳迁移实验(electrophoretic mobility shift assay, EMSA)、酶联免疫吸附实验(enzyme linked immunosorbent assay, ELISA)、蛋白免疫印迹法(Western blot)、蛋白质谱鉴定法、染色质免疫共沉淀测序法(ChIP-seq)、交联免疫共沉淀测序法(cross-linking and immunoprecipitation methods coupled to high-throughput sequencing, CLIP-seq)等。

电泳迁移实验和酶联免疫吸附实验常被用于研究特定蛋白与特定G4结构相互作用^[15,61]。如图7(a)所示,

通常使用P³²标记的G4链, 并使其在KCl溶液中退火, 形成稳定的G4结构, 再与待研究的特定蛋白孵育后, 进行凝胶电泳。蛋白的结合使核酸链的电泳迁移速率减小, 放射性自显影时在凝胶上呈现相对于自由核酸链较为滞后的条带。通常, 通过该实验方法对G4与蛋白的相互作用进行验证时, 还需要设置对照组, 对照组的设置包括G4位点突变的单链、发卡结构(hairpin)双链等。当某个蛋白只对G4链呈现高亲和性的结合, 可认为该蛋白为G4结合蛋白。电泳迁移实验需要使用纯化后的蛋白以及放射性元素标记的G四链体核酸链, 通量较低。

蛋白免疫印迹法和蛋白质谱鉴定法主要用于对亲和富集后的大量混合蛋白进行分析^[61,63]。蛋白免疫印迹法可以对蛋白混合物中的不同蛋白进行检测, 省略了电泳迁移实验中复杂的蛋白表达及纯化的过程, 通量较高。如图7(b)所示, 蛋白免疫印迹法的主要流程为用十二烷基磺酸钠(SDS)将待检测的蛋白变性并带上负电荷, 通过凝胶电泳将不同分子量的蛋白分开呈列在凝胶的不同位置上, 再通过转膜将凝胶上的蛋白转移到印迹膜上, 最后用待检测蛋白的抗体及辣根过氧化物酶(HRP)标记的二抗进行孵育和显影。该方法常用于对G4互作蛋白特异性探针富集的G4相关蛋白混合物中的一种或者几种蛋白进行检测。同样地, 蛋白质谱法亦可用于对G4互作蛋白特异性探针富集的混合蛋白样品进行表征, 该方法通过将蛋白样品酶解成多肽混合物后进行质谱鉴定, 再与已知的多肽库进行比对, 可以对复杂体系中的所有蛋白进行精确的定量和鉴定, 具有高通量和精准度高的特点。

ChIP-seq和CLIP-seq常用于研究活细胞中G四链体与蛋白之间的相互作用^[61,63]。由于下一代测序技术(next-generation sequencing)的巨大进步, ChIP-seq和CLIP-seq技术能够提供不同蛋白在基因组和转录组上更为精确及更为广泛的结合位点。如图7(c)所示, 该方法通过紫外照射使蛋白与结合的核酸形成共价交联, 再使用特异性抗体对与蛋白结合的核酸进行富集, 最后对富集后的核酸进行二代测序。该方法常用于对体外验证过的G4互作蛋白与基因组及转录组的结合位点进行检测, 再通过结合位点与潜在G4位点分布的比对, 以确定该蛋白在真实生理条件下的结合位点是否富含G4形成序列。相较于上述几种方法, 该方法能够真实反映蛋白在生理状态下的G4结合位点, 为G4生理功能的揭示提供有力的支持。

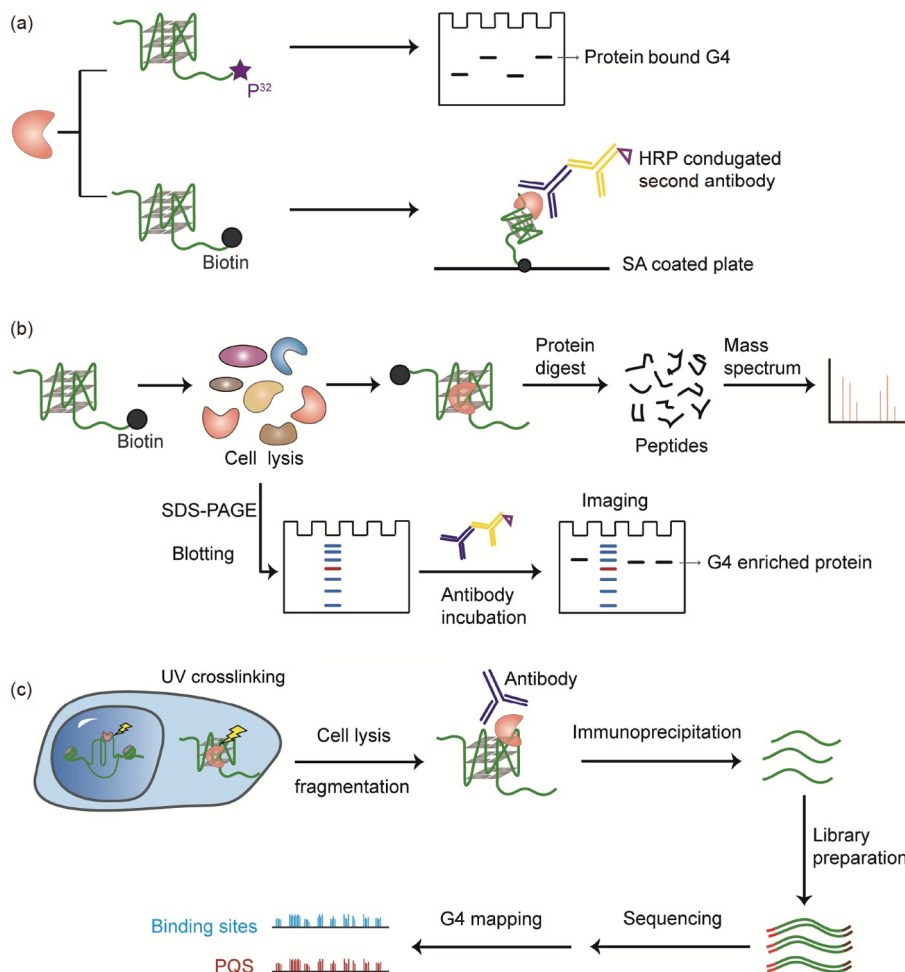


图7 G四链体与蛋白相互作用的研究技术。(a) 电泳迁移实验和酶联免疫吸附实验; (b) 蛋白免疫印迹法和蛋白质谱鉴定法; (c) 染色质免疫共沉淀法(ChIP-seq)和交联免疫共沉淀法(CLIP-seq)

Figure 7 Research techniques for the study of interaction between G4 and G4-interacting proteins. (a) Electrophoretic mobility shift assay (EMSA) and enzyme linked immunosorbent assay (ELISA); (b) Western blot and protein mass spectrometry; (c) chromatin immunoprecipitation followed by sequencing (ChIP-seq) and cross-linking and immunoprecipitation methods coupled to high-throughput sequencing (CLIP-seq)

3.2 G4互作蛋白的鉴定方法

目前, G4互作蛋白的鉴定主要依赖于亲和蛋白质组学实验, 即使用RNA或DNA G四链体寡核苷酸链作为诱饵, 从细胞提取物中分离与其结合的蛋白。如图8(a)所示, Shankar课题组^[64]选取了*NRAS* 5'-UTR上一段G四链体序列, 并以相应的突变单链和发卡结构链作为负对照, 成功地从HeLa细胞中筛选出G四链体结合蛋白DDX3X。值得注意的是, 这种方法只能使用特定的序列, 而且只能从细胞裂解液中对相关蛋白进行钓取, 而细胞裂解液无法模拟真实细胞环境中G四链体与蛋白的相互作用。

另外, 计算机分析技术也可以作为一种发现G4互作蛋白的辅助方法。计算分析方法包括基于特定生理过程的核酸结合蛋白结合位点, 并结合潜在的G4位点加以评估, 或利用meta分析寻找具有已知G4结合蛋白的共享结构特征, 来预测新的G4结合蛋白。G四链体基序在启动子附近大量富集, 这说明转录调控因子可能通过与G四链体的相互作用调控转录过程。TFIIH(transcription factor II H)是一种通用转录因子, 在转录起始和核苷酸切除修复中起作用。TFIIH包含11个亚基, 其中两个是高度保守的解旋酶XPB和XPD^[65]。XPB和XPD在人类细胞和大多数真核生物中具有十分重要的调节功能。例如, XPB和XPD的特定突变等位基因可能

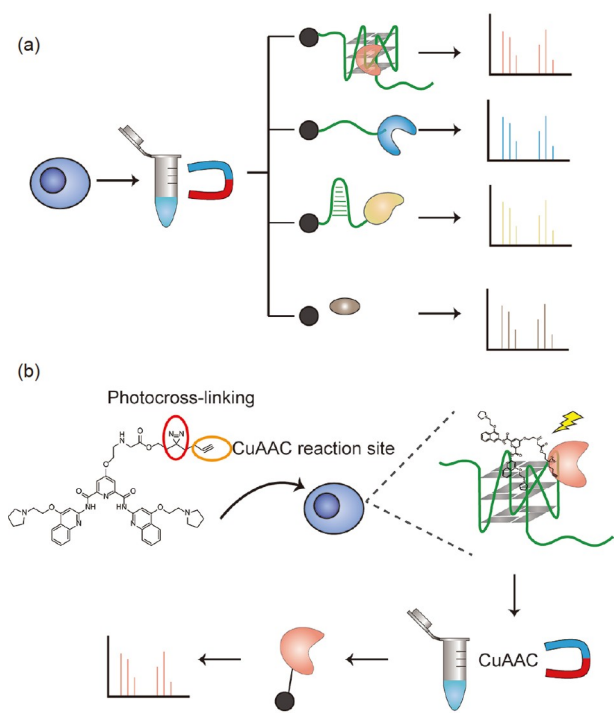


图 8 G四链体互作蛋白的检测方法。(a) 利用RNA或DNA G四链体寡核苷酸链及阴性对照链作为诱饵, 从细胞提取物中分离G四链体互作蛋白; (b) 光交联G4靶向配体分子探针用于G四链体互作蛋白的筛选

Figure 8 Methods for the detection of G4-interacting proteins. (a) Using RNA or DNA G4 oligonucleotide and negative control chains as decoys to isolate G4-interacting proteins from cell extracts; (b) optical cross-linked G4-targeting ligand probes for the detection of G4-interacting proteins

引起色素性干皮等疾病^[66]。通过对人类细胞的ChIP-seq分析, Gray等人^[67]发现, 40%转录相关解旋酶(XPB和XPD)的结合位点与G4基序高度重叠, 表明这些蛋白质可能被招募到基因组DNA形成的G四链体结构上。他们进一步通过实验证明, XPD是一种强大的DNA G四链体解旋酶, 并且在约20%基因的转录起始位点附近富集, 在高转录基因附近, 这种现象更加显著。这种鉴定方法的不足是只能鉴定已知功能的蛋白, 而蛋白功能的执行依靠多个蛋白之间组成的复杂蛋白网络, 这种间接的鉴定方法无法确定未知功能蛋白与G四链体之间的相互作用。

总结已报道的G4相关蛋白的氨基酸序列规律, 根据该规律预测新的G4相关蛋白, 通过对大量G4结合蛋白的分析, 研究人员发现多个参与到G4识别中的关键结构域。如DHX36与G4的晶体结构表明, DHX36特异性基序(DSM)和C端类OB折叠域对DHX36识别G4结构

起关键作用^[68]。此外, 富含精氨酸与甘氨酸的RGG结构域也在G4结合蛋白中经常出现, 其中部分RGG结构域直接参与蛋白对G四链体结构的识别^[69,70]。中山大学黄志纾团队^[69]从1000多个含有RGG结构域蛋白中, 筛选出具有不同序列特征的12条富含RGG结构域的多肽片段, 发现并证实了含有7个RGG重复单元的多肽12能选择性识别G四链体结构。基于上述结果, 他们对蛋白进行数据库搜索, 寻找含有该RGG多肽序列特征的蛋白。结果发现了CIRBP蛋白, 并通过一系列实验证实, 该蛋白能在细胞内外有效识别G四链体结构。而且通过基因突变改变或去掉该RGG结构域, 直接导致蛋白对G4核酸识别能力的下降, 说明在该蛋白与G4核酸的识别过程中, RGG结构域起到关键作用。该研究首次通过对RGG多肽序列与G4相互作用的系统研究, 发现了全新的G四链体结合蛋白, 为将来发现更多的G4结合蛋白提供了一种新途径。基于计算方法推测的G四链体互作蛋白还需要进一步通过实验方法加以验证。

亲和蛋白质组学方法及基于计算分析的方法都不考虑天然的染色质状态, 因此本课题组^[61]以及Shankar课题组^[63]先后独立发表了化学小分子化合物共价结合介导的蛋白质分析方法, 用于活细胞中G4互作蛋白的检测(图8(b))。两个课题组不约而同地采用了相似的策略来设计小分子, 可见该小分子应用于在G4互作蛋白的检测中, 具有显而易见的优势。简单来说, 该小分子为三官能探针, 其核心部分是吡啶双喹啉衍生物pyridostatin(PDS), 用于活细胞中G4的靶向; 第二个部分是具有光催化交联功能的双吡丙啶基团, 用于与蛋白质进行交联; 第三部分是叠氮基团, 用于进行铜催化的生物正交反应。相较于传统的亲和蛋白质组学, 该方法可以克服序列偏好性, 并且能够在活细胞环境中对G4互作蛋白进行捕获。利用该方法获得的G4互作蛋白结果能够更加精确地表现不同细胞状态下的G4互作蛋白组。相较于计算机算法评估法, 该方法具有更高的鉴定通量, 能够更好地反映G4蛋白的相互作用网络。

4 总结和展望

G四链体互作蛋白在调节细胞中G四链体结构的形成、稳定和解旋方面至关重要, 并且这种调节机制具有空间和时间的特异性。由于基因组及转录组上的G四链体主要分布在功能调控区域, 因此G四链体互作蛋白常常通过对G四链体结构的调控, 以调节相关的生理过程。而且G四链体由于其特殊的结构特征, 对特定的

蛋白表现出高亲和性, 这种特征使其能够通过自身的折叠和解旋招募或者释放特定功能蛋白到特定靶标位置, 行使相应的功能. G四链体互作蛋白在与疾病相关的不同细胞通路中起着至关重要的作用, G四链体的失调将引起基因组表达的失调, 从而造成严重的肿瘤等疾病. 具体而言, DNA G四链体互作蛋白主要涉及端粒的维持、DNA复制、转录、表观遗传修饰等调控, 而RNA G四链体互作蛋白则主要参与mRNA的成熟和翻译调控等过程.

通过现有的分析方法, 已经鉴定出大量的G四链体互作蛋白. 研究人员还对这些蛋白质的共有基序及结构域进行了大量研究, 提出多个具有一定保守性的基序及结构域. 这为G四链体互作蛋白及其与G四链体结构之间相互作用的分子机制揭示提供了重要的基础. 然而, 由于内源性的G四链体结构和G四链体互作蛋白涉及多种生物过程的调控, 并且是动态变化的, 我们应该关注现有鉴定方法的改善和开发新型鉴定方法. 基于G四链体邻近标记小分子或G四链体靶向的蛋白生物素连接酶能够在生理条件下对G四链体相关蛋白进

行捕获, 这将为揭示更加真实的G四链体相关蛋白作用网络, 提供有力的工具. 但是解决小分子及生物酶对原有作用网络的干扰也是需要解决的难点之一.

基于结构模拟的筛选需要较为精确的结构预测及大通量的筛选. 由于G四链体结构的多态性, 外源表达的G四链体结构与内源存在的G四链体结构存在较大差异, G四链体的体外鉴定方法限制了一些精细结构的表征, 因此需要进一步开发生理条件下G四链体结构的鉴定方法, 以克服这一难点. 在此基础上, 在计算机水平上, 对G四链体及G四链体互作蛋白的相互作用进行模拟, 将有望进一步在分子水平上揭示其内在的作用机制. 过去10年间, 计算模型的巨大进步提高了结构建模的精确度. 如我们可以通过使用基于深度学习的AlphaFold技术等, 在理论水平上通过氨基酸序列精确预测蛋白的结构, 辅助解析G四链体相关蛋白的结构特征及二者之间更为精确的作用方式. 由此, 进一步寻找和发现G四链体与其相关蛋白作用方式的保守特点及特殊方式, 为新的G四链体相关蛋白的发现提供结构基础.

参考文献

- Gellert M, Lipsett M N, Davies D R. Helix formation by guanylic acid. *Proc Natl Acad Sci USA*, 1962, 48: 2013–2018
- Sen D, Gilbert W. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature*, 1988, 334: 364–366
- Zimmerman S B, Cohen G H, Davies D R. X-ray fiber diffraction and model-building study of polyguanylic acid and polyinosinic acid. *J Mol Biol*, 1975, 92: 181–192
- Sen D, Gilbert W. A sodium-potassium switch in the formation of four-stranded G4-DNA. *Nature*, 1990, 344: 410–414
- Burge S, Parkinson G N, Hazel P, et al. Quadruplex DNA: Sequence, topology and structure. *Nucleic Acids Res*, 2006, 34: 5402–5415
- Andreev D V, Tikhomirov A S, Shchekotikhin A E. Ligands of G-quadruplex nucleic acids. *Russ Chem Rev*, 2021, 90: 1–38
- Tao Y, Zheng Y, Zhai Q, et al. Recent advances in the development of small molecules targeting RNA G-quadruplexes for drug discovery. *Bioorg Chem*, 2021, 110: 104804
- Todd A K. Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res*, 2005, 33: 2901–2907
- Gray D M, Wen J D, Gray C W, et al. Measured and calculated CD spectra of G-quartets stacked with the same or opposite polarities. *Chirality*, 2008, 20: 431–440
- Amrane S, Adrian M, Heddi B, et al. Formation of pearl-necklace monomorphous G-quadruplexes in the human CEB25 minisatellite. *J Am Chem Soc*, 2012, 134: 5807–5816
- Campbell N H, Parkinson G N. Crystallographic studies of quadruplex nucleic acids. *Methods*, 2007, 43: 252–263
- Chambers V S, Marsico G, Boutell J M, et al. High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat Biotechnol*, 2015, 33: 877–881
- Kwok C K, Marsico G, Sahakyan A B, et al. rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. *Nat Methods*, 2016, 13: 841–844
- Hänsel-Hertsch R, Spiegel J, Marsico G, et al. Genome-wide mapping of endogenous G-quadruplex DNA structures by chromatin immunoprecipitation and high-throughput sequencing. *Nat Protoc*, 2018, 13: 551–564
- Zheng K, Zhang J, He Y, et al. Detection of genomic G-quadruplexes in living cells using a small artificial protein. *Nucleic Acids Res*, 2020, 48: 11706–11720

- 16 Hui W W I, Simeone A, Zyner K G, et al. Single-cell mapping of DNA G-quadruplex structures in human cancer cells. *Sci Rep*, 2022, 12: 23641
- 17 Huppert J L. Prevalence of quadruplexes in the human genome. *Nucleic Acids Res*, 2005, 33: 2908–2916
- 18 Spiegel J, Adhikari S, Balasubramanian S. The structure and function of DNA G-quadruplexes. *Trends Chem*, 2020, 2: 123–136
- 19 Henderson A, Wu Y, Huang Y C, et al. Detection of G-quadruplex DNA in mammalian cells. *Nucleic Acids Res*, 2014, 42: 860–869
- 20 Brázda V, Hároníková L, Liao J, et al. DNA and RNA quadruplex-binding proteins. *Int J Mol Sci*, 2014, 15: 17493–17517
- 21 Sun Z Y, Wang X N, Cheng S Q, et al. Developing novel G-quadruplex ligands: From interaction with nucleic acids to interfering with nucleic acid-protein interaction. *Molecules*, 2019, 24: 396
- 22 Shu H, Zhang R, Xiao K, et al. G-quadruplex-binding proteins: Promising targets for drug design. *Biomolecules*, 2022, 12: 648
- 23 de Almeida C R, Dhir S, Dhir A, et al. RNA helicase DDX1 converts RNA G-quadruplex structures into R-loops to promote IgH class switch recombination. *Mol Cell*, 2018, 70: 650–662.e8
- 24 Sauer M, Paeschke K. G-quadruplex unwinding helicases and their function *in vivo*. *Biochem Soc Trans*, 2017, 45: 1173–1182
- 25 Huber M D. G4 DNA unwinding by BLM and Sgs1p: Substrate specificity and substrate-specific inhibition. *Nucleic Acids Res*, 2002, 30: 3954–3961
- 26 Budhathoki J B, Ray S, Urban V, et al. RecQ-core of BLM unfolds telomeric G-quadruplex in the absence of ATP. *Nucleic Acids Res*, 2014, 42: 11528–11545
- 27 Sharma T, Kundu N, Kaur S, et al. Recognition and unfolding of human telomeric G-quadruplex by short peptide binding identified from the HRDC domain of BLM helicase. *RSC Adv*, 2022, 12: 21760–21769
- 28 Hänsel-Hertsch R, Simeone A, Shea A, et al. Landscape of G-quadruplex DNA structural regions in breast cancer. *Nat Genet*, 2020, 52: 878–883
- 29 Biffi G, Tannahill D, Miller J, et al. Elevated levels of G-quadruplex formation in human stomach and liver cancer tissues. *PLoS One*, 2014, 9: e102711
- 30 Nguyen G H, Tang W, Robles A I, et al. Regulation of gene expression by the BLM helicase correlates with the presence of G-quadruplex DNA motifs. *Proc Natl Acad Sci USA*, 2014, 111: 9905–9910
- 31 Alavi S, Ghadiri H, Dabirmanesh B, et al. G-quadruplex binding protein Rif1, a key regulator of replication timing. *J Biochem*, 2021, 169: 1–14
- 32 Wu C G, Spies M. G-quadruplex recognition and remodeling by the FANCI helicase. *Nucleic Acids Res*, 2016, 44: 8742–8753
- 33 Bosch P C, Segura-Bayona S, Koole W, et al. FANCI promotes DNA synthesis through G-quadruplex structures. *EMBO J*, 2014, 33: 2521–2533
- 34 Westdorp K N, Terhune S S. Impact of RNA polymerase I inhibitor CX-5461 on viral kinase-dependent and -independent cytomegalovirus replication. *Antiviral Res*, 2018, 153: 33–38
- 35 Pramanik S, Chen Y, Song H, et al. The human AP-endonuclease 1 (APE1) is a DNA G-quadruplex structure binding protein and regulates *KRAS* expression in pancreatic ductal adenocarcinoma cells. *Nucleic Acids Res*, 2022, 50: 3394–3412
- 36 Niu K, Xiang L, Jin Y, et al. Identification of LARK as a novel and conserved G-quadruplex binding protein in invertebrates and vertebrates. *Nucleic Acids Res*, 2019, 47: 7306–7320
- 37 Chen Y, Niu K, Song Q, et al. Effect of G-quadruplex loop mutations on the G-quadruplex formation, protein binding and transcription of BmPOUM2 in *Bombyx mori*. *Arch Insect Biochem Phys*, 2022, 110: e21876
- 38 Kota S, Misra H S. Topoisomerase IB of *Deinococcus radiodurans* resolves guanine quadruplex DNA structures *in vitro*. *J Biosci*, 2015, 40: 833–843
- 39 Liano D, Chowdhury S, Di Antonio M. Cockayne syndrome B protein selectively resolves and interact with intermolecular DNA G-quadruplex structures. *J Am Chem Soc*, 2021, 143: 20988–21002
- 40 Ceschi S, Berselli M, Cozzaglio M, et al. Vimentin binds to G-quadruplex repeats found at telomeres and gene promoters. *Nucleic Acids Res*, 2022, 50: 1370–1381
- 41 Biffi G, Tannahill D, McCafferty J, et al. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat Chem*, 2013, 5: 182–186
- 42 Ray S, Bandaria J N, Qureshi M H, et al. G-quadruplex formation in telomeres enhances POT1/TPP1 protection against RPA binding. *Proc Natl Acad Sci USA*, 2014, 111: 2990–2995
- 43 Xu M, Axhemi A, Malgowska M, et al. Active and passive destabilization of G-quadruplex DNA by the telomere POT1-TPP1 complex. *J Mol Biol*, 2021, 433: 166846
- 44 Bhattacharjee A, Wang Y, Diao J, et al. Dynamic DNA binding, junction recognition and G4 melting activity underlie the telomeric and genome-wide roles of human CST. *Nucleic Acids Res*, 2017, 45: 12311–12324
- 45 Safa L, Delagoutte E, Petruseva I, et al. Binding polarity of RPA to telomeric sequences and influence of G-quadruplex stability. *Biochimie*, 2014, 103: 80–88
- 46 Wang Y R, Guo T T, Zheng Y T, et al. Replication protein A plays multifaceted roles complementary to specialized helicases in processing G-quadruplex DNA. *Iscience*, 2021, 24: 102493
- 47 Mendoza O, Bourdoncle A, Boulé J B, et al. G-quadruplexes and helicases. *Nucleic Acids Res*, 2016, 44: 1989–2006

- 48 Biffi G, Di Antonio M, Tannahill D, et al. Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. *Nat Chem*, 2014, 6: 75–80
- 49 Sauer M, Juranek S A, Marks J, et al. DHX36 prevents the accumulation of translationally inactive mRNAs with G4-structures in untranslated regions. *Nat Commun*, 2019, 10: 2421
- 50 Zhang J, Harvey S E, Cheng C. A high-throughput screen identifies small molecule modulators of alternative splicing by targeting RNA G-quadruplexes. *Nucleic Acids Res*, 2019, 47: 3667–3679
- 51 Nie J, Jiang M, Zhang X, et al. Post-transcriptional regulation of Nkx2-5 by RHAU in heart development. *Cell Rep*, 2015, 13: 723–732
- 52 Edwards M, Joseph S. The fragile X proteins differentially regulate translation of reporter mRNAs with G-quadruplex structures. *J Mol Biol*, 2022, 434: 167396
- 53 Benhalevy D, Gupta S K, Danan C H, et al. The human CCHC-type zinc finger nucleic acid-binding protein binds G-rich elements in target mRNA coding sequences and promotes translation. *Cell Rep*, 2017, 18: 2979–2990
- 54 Horiuchi K, Kawamura T, Hamakubo T. Wilms' tumor 1-associating protein complex regulates alternative splicing and polyadenylation at potential G-quadruplex-forming splice site sequences. *J Biol Chem*, 2021, 297: 101248
- 55 Niu K, Zhang X, Song Q, et al. G-quadruplex regulation of VEGFA mRNA translation by RBM4. *Int J Mol Sci*, 2022, 23: 743
- 56 Mao S Q, Ghanbarian A T, Spiegel J, et al. DNA G-quadruplex structures mold the DNA methylome. *Nat Struct Mol Biol*, 2018, 25: 951–957
- 57 Yoshida A, Oyoshi T, Suda A, et al. Recognition of G-quadruplex RNA by a crucial RNA methyltransferase component, METTL14. *Nucleic Acids Res*, 2022, 50: 449–457
- 58 Li P T, Wang Z F, Chu I T, et al. Expression of the human telomerase reverse transcriptase gene is modulated by quadruplex formation in its first exon due to DNA methylation. *J Biol Chem*, 2017, 292: 20859–20870
- 59 Shukla V, Samaniego-Castruita D, Dong Z, et al. TET deficiency perturbs mature B cell homeostasis and promotes oncogenesis associated with accumulation of G-quadruplex and R-loop structures. *Nat Immunol*, 2022, 23: 99–108
- 60 Wang Y, Yang J, Wild A T, et al. G-quadruplex DNA drives genomic instability and represents a targetable molecular abnormality in ATRX-deficient malignant glioma. *Nat Commun*, 2019, 10: 943
- 61 Zhang X, Spiegel J, Martínez Cuesta S, et al. Chemical profiling of DNA G-quadruplex-interacting proteins in live cells. *Nat Chem*, 2021, 13: 626–633
- 62 Saha D, Singh A, Hussain T, et al. Epigenetic suppression of human telomerase (hTERT) is mediated by the metastasis suppressor NME2 in a G-quadruplex-dependent fashion. *J Biol Chem*, 2017, 292: 15205–15215
- 63 Su H, Xu J, Chen Y, et al. Photoactive G-quadruplex ligand identifies multiple G-quadruplex-related proteins with extensive sequence tolerance in the cellular environment. *J Am Chem Soc*, 2021, 143: 1917–1923
- 64 Herdy B, Mayer C, Varshney D, et al. Analysis of NRAS RNA G-quadruplex binding proteins reveals DDX3X as a novel interactor of cellular G-quadruplex containing transcripts. *Nucleic Acids Res*, 2018, 46: 11592–11604
- 65 Compe E, Egly J M. TFIIH: When transcription met DNA repair. *Nat Rev Mol Cell Biol*, 2012, 13: 343–354
- 66 Kamileri I, Karakasilioti I, Garinis G A. Nucleotide excision repair: New tricks with old bricks. *Trends Genet*, 2012, 28: 566–573
- 67 Gray L T, Vallur A C, Eddy J, et al. G quadruplexes are genomewide targets of transcriptional helicases XPB and XPD. *Nat Chem Biol*, 2014, 10: 313–318
- 68 Chen M C, Tippana R, Demeshkina N A, et al. Structural basis of G-quadruplex unfolding by the DEAH/RHA helicase DHX36. *Nature*, 2018, 558: 465–469
- 69 Huang Z L, Dai J, Luo W H, et al. Identification of G-quadruplex-binding protein from the exploration of RGG motif/G-quadruplex interactions. *J Am Chem Soc*, 2018, 140: 17945–17955
- 70 Goering R, Hudish L I, Guzman B B, et al. FMRP promotes RNA localization to neuronal projections through interactions between its RGG domain and G-quadruplex RNA sequences. *eLife*, 2020, 9: e52621

Summary for “G四链体互作蛋白及其研究方法进展”

Progress in the study of G-quadruplex interacting proteins

Haiyan Huang & Xiang Zhou*

College of Chemistry and Molecular Sciences, Wuhan University, Wuhan 430072, China

* Corresponding author, E-mail: xzhou@whu.edu.cn

G-quadruplex (G4) is a noncanonical nucleic acid secondary structure formed by regular guanine repeat. G4 structure has been proven to participate in various physiological and pathological processes, such as telomere maintenance, transcription, translation, and carcinogenesis. Although Putative G-quadruplex forming sequences (PQSs) have been proven to be widely present in genome and transcriptome, the structure is highly heterogeneous among different cell types and cell cycles. G4 structures are regulated by corresponding proteins (G4-interacting proteins), for instance, nucleolin was reported to stabilize G4 structures, and DHX36 on the contrary unfolds G4s. Therefore, the study of the G4 structure and their interacting proteins will be beneficial to further reveal the regulation mechanism of biological activity and promote the development of new disease therapeutic methods.

There are multiple G4 helicases in cells, in order to maintain normal physiological activity, such as the RecQ family, DEAD-box family, and DEAH-box family. The G4 structure encountered in the progression of the replication fork may prevent the passage of DNA polymerase, resulting in the arrest of the replication fork, which may result in deletion and mutation of the genome. The G quadruplex structure in the process of genome transcription has a double function. On the one hand, the G4 structure blocks the passage of RNA polymerase, resulting in the decrease of transcription level. On the other hand, studies have shown that the special structure of G4 can be used as the binding site for transcription factors to regulate transcription. Interestingly, researchers have also found some proteins that play a dual role in the folding and unfolding of G4 structures, such as topoisomerase. Human telomere DNA sequence is one of the first G4 structures to be discovered, and POT1 in the telomere protein complex is able to interact with telomere G4 and play a crucial role in maintaining chromosome stability. In addition, some G4 helicases such as CST and RPA can be bind to the G4 and open the G4 structure, helping to solve the replication pressure of the telomere. Compared with DNA, RNA is more likely to form a G4 structure under physiological conditions due to its single-stranded characteristics. High throughput sequencing and fluorescence imaging techniques have also confirmed that a large number of RNA G-quadruplex (rG4) exist in cells, and rG4 is mainly concentrated in the 5' and 3' UTR regions. Studies have shown that RNA G-quadruplex plays an important role in alternative splicing, translation regulation, and transcription termination. It is reported that G4 interacting proteins play an important role in the regulation of nucleic acids epigenetic modification, such as DNA methyltransferase DNMTs and RNA methyltransferase METTL14, which have been shown to bind specifically to G4 and regulate the nucleic acid methylation process. Moreover, G4 interacting proteins play important roles in histone methylation and chromatin remodeling.

Therefore, the identification and in-depth study of G4-interacting proteins will help to reveal the mechanism of G4-protein interaction and its biological role *in vivo*. In recent years, benefiting from the progress in the study of nucleic acid-protein interaction and the development of new technology for the study of G-quadruplex interacting proteins, scientists have discovered a variety of G-quadruplex interacting proteins involved in a variety of biological processes. In this paper, we introduced the types and functions of known G4 interacting proteins, reviewed the recent development of research methods of G-quadruplex interacting proteins, and blueprint the forward development.

G-quadruplex, G4-interacting proteins, helicase, replication, transcription, telomere

doi: [10.1360/TB-2023-0180](https://doi.org/10.1360/TB-2023-0180)