PIBB 生物化学与生物物理进展
Progress in Biochemistry and Biophysics
2012, 39(9): 843~852
www.pibb.ac.cn

核小体定位的转录调控功能研究进展 *

刘 辉[1)] 壮子恒[1)] 关佶红[2)] 周水庚[3)**]
([1)]常州大学制药与生命科学学院，常州 213164；[2)]同济大学计算机系，上海 201804；
[3)]上海市智能信息处理重点实验室(复旦大学)，上海 200433)

摘要 核小体是真核生物染色质的基本组成单位，组蛋白八聚体在 DNA 双螺旋上精确位置称为核小体定位．核小体定位已被证实在基因转录调控、DNA 复制与修复、调控进化等过程中扮演着重要的角色．随着染色质免疫共沉淀 - 芯片(ChIP-chip)与染色质免疫共沉淀 - 测序(ChIP-seq)等高通量技术的出现，已测定了多种模式生物全基因组核小体定位图谱，掀起了一股核小体定位及其功能的研究热潮，并取得了一定的成果．本文介绍了核小体定位的概念，总结了核小体在启动子与编码区域内定位的基本模式．在此基础上，综述了核小体定位在转录起始、转录延伸、基因表达模式多样化以及可变剪接等方面的功能研究进展．

关键词 核小体定位，转录调控，可变剪接，基因表达，染色质
学科分类号 Q34，Q71
DOI: 10.3724/SP.J.1206.2011.00252

核小体(nucleosome)是构成真核生物染色质的基本结构单位，由 DNA 缠绕在组蛋白八聚体上构成．围绕在组蛋白上的 DNA 称为核心 DNA(core DNA)或者核小体 DNA(nucleosomal DNA)，其长度为 146 bp，围绕组蛋白约 1.65 圈．连接相邻核小体之间的 DNA 称为连接 DNA(linker DNA)，长度在 20～54 bp 不等[1-2]．图 1 显示了核小体的基本结构，核心组蛋白由 4 种组蛋白 H2A、H2B、H3 和 H4 各 2 个分子形成一个组蛋白八聚体，连接组蛋白 H1 在八聚体与 DNA 之间起加固作用，核小体之间通过 linker DNA 连接形成念珠状结构．核小体定位(nucleosome positioning)指组蛋白八聚体在 DNA 双螺旋上精确位置，具体可分为描述核小体 DNA 的中心点相对于染色体座位的平移定位(translational positioning)和描述 DNA 双螺旋与组蛋白八聚体结合表面的旋转定位(rotational positioning)[3]．真核细胞内 75%～90%的 DNA 缠绕在组蛋白上，组蛋白的静电位阻作用会排斥其他蛋白质分子与核小体 DNA 结合，从而阻断包括那些负责基本生命过程的蛋白质(如：RNA 聚合酶、转录因子和 DNA 复制与修复酶等)与 DNA 的接触机会．而核小体缺乏区域与核小体之间的连接 DNA 允许这些蛋白质因子进入，并与顺式调控元件结

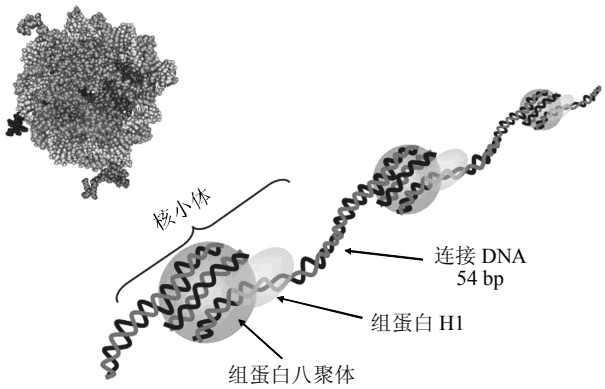

Fig. 1 Illustration of nucleosomal structure
图 1 核小体结构示意图
核小体由 DNA 缠绕在组蛋白(histone)构成，4 种组蛋白 H2A、H2B、H3 和 H4 各 2 个分子形成一个组蛋白八聚体，约 146 bp 的 DNA 双螺旋分子盘绕在组蛋白八聚体 1.65 圈，连接组蛋白 H1 在八聚体与 DNA 之间起加固作用，核小体之间通过 linker DNA 连接形成念珠状结构．

* 国家重点基础研究发展计划(973)(2010CB126600)，国家自然科学基金(31100954, 61173118, 61272380)，上海市智能信息处理重点实验室开放研究计划(IIPL-2010-011)和常州大学博士培养基金资助项目(CCZUPY1001).
** 通讯联系人.
Tel: 021-65654504, E-mail: sgzhou@fudan.edu.cn
收稿日期：2011-08-05，接受日期：2011-09-08

合，启动基因转录，使 DNA 上的遗传信息得以表达[4-6]. 因此，核小体在 DNA 序列上的位置对该段 DNA 序列参与的所有功能都会产生重要的影响[4-5]. 实际上，最近的研究表明核小体定位在基因转录调控[7]、DNA 复制与修复、可变剪接[8-9]等基本生命过程中扮演着重要的角色，甚至对 DNA 序列与基因表达调控的进化也有着重要影响[10-12]. 此外，核小体的重要组成部分——组蛋白的修饰异常与肿瘤等重大疾病的发生、发展也有紧密联系. 因此，核小体定位及其功能的研究有助于揭示染色质结构对基因表达的影响，为重大疾病特别是癌症的治疗提供更多的机会. 然而，核小体在 DNA 上的定位非常复杂，DNA 既要缠绕在组蛋白上进行高度的压缩，又要对基因的表达进行精确的调控，那么核小体在 DNA 上是如何定位，以及在基因转录过程中有哪些调控功能呢？目前的研究结果仍无法准确地回答这个问题.

因此，核小体定位及其功能研究是一项复杂而意义重大的任务，已经在国内外掀起了一股研究热潮. 许多研究者使用包括生化实验、生物信息学等多种方法来揭示体内核小体在基因组上的定位规律及其功能. 特别是最近 5 年来，染色质免疫共沉淀 - 芯片(ChIP-chip)与染色质免疫共沉淀 - 测序(ChIP-seq)等高通量技术的发展，多种模式生物的全基因组核小体定位图谱被测定并公开，促进了核小体定位的研究，并为探索核小体定位及其在基因表达调控中的功能奠定了基础. 本文先简要描述目前人们对核小体定位的基本认识，然后综述了核小体定位在转录起始、延伸、基因表达模式多样化以及可变剪接过程中标记外显子(exon)的功能.

1　核小体定位研究进展

目前，人们对核小体的分布规律有了基本的认识：a. 在基因的启动子、终止子和转录因子结合位点(TFBS)区域内核小体比较缺乏[13-15]，而编码区内核小体相对较密集；b. 在基因转录起始位点与转录终止位点附近分别存在一个核小体缺乏区域(nucleosome free region，NFR) 5′NFR 与 3′NFR，且 5′NFR 的上下游分布着+1 和−1 两个位置高度确定的核小体，3′NFR 上游也有一个定位较强的核小体[16-17]；c. 编码区其他位置上核小体定位的确定性随着与 5′NFR 与 3′NFR 的距离增大而降低[18-19]，如图 2 所示；d. linker DNA 的长度是量子化的，其长度服从(10n+5) bp 规则[20]. 对于上述核小体定位的具体机制目前主要有两种主流的观点：DNA 序列定位和统计定位.

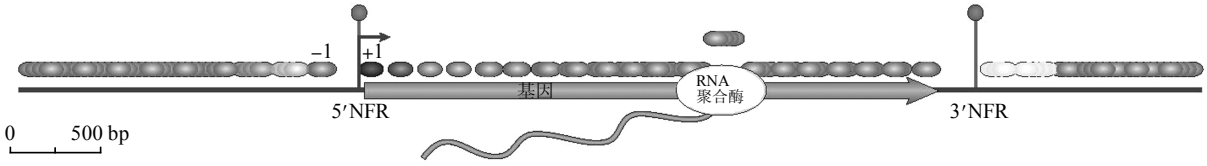

Fig. 2　Illustration of nucleosome positioning[5]

图 2　核小体定位示意图[5]

基因转录起始位点与转录终止位点附近分别存在一个核小体缺乏区域(nucleosome free region，NFR)，5′NFR 的上下游分布着+1 和−1 两个位置高度确定的核小体，3′NFR 上游也有一个定位较强的核小体；编码区其他位置上核小体定位的确定性随着与 5′NFR 与 3′NFR 的距离增大而降低.

DNA 序列定位是指不同 DNA 序列片段与组蛋白的亲和力不同，构成核小体的能力存在内源性的差别，从而决定了核小体的定位. Segal 领导的研究小组率先提出了 DNA 序列定位的观点[17-19]，他们先后采用 ChIP-seq 技术测定了酵母体内(*in vivo*)和体外(*in vitro*)纯化的 DNA 核小体定位图谱，发现体内与体外核小体定位具有较高的重合度，且核小体的核心区 AA/TT/TA 二联体与 CC/CG/GC/GG 二联体交替出现，间隔为 5 bp. 包含这两种二联体的序列具有较强可塑性，易形成核小体. 而启动子区域、DNA 复制起始区域与基因的 3′端 Poly (dA:dT)序列含量较多，该序列刚性较强，不易弯曲，致这些区域内核小体缺乏，这恰好有利于其他蛋白质分子与其上的靶位点结合. Segal 等将 DNA 序列的这种特性称为核小体序列偏好性(nucleosome sequence preference)，并由此推断体内核小体定位

主要是由 DNA 序列决定的.

统计定位则基于屏障模型[21]，认为 NFR、+1 与–1 核小体或 DNA 结合蛋白构成了若干屏障，处于屏障位置上的核小体定位是高度确定的，而其他区域内的核小体由于静电位阻作用而服从统计定位[22-24]，即离屏障的距离越远核小体位置的确定性越低. Fu 等[25]与 Schones 等[26]分析了酵母与人类核小体定位的差异，支持了统计定位的观点. Zhang 等[27]则更进一步指出，DNA 序列的偏好主要决定核小体的旋转定位，而对核小体平移定位的影响是有限的. Stein 等[28]的研究结果进一步支持了 Zhang 等的结论.

高分辨率的核小体定位图谱为生物信息方法预测活体内(*in vivo*)核小体定位提供了数据样本，目前已经有多种计算预测算法被开发出来[29-32]. Zhao 等[33]使用多样性增量二次判别方法对酵母、人类与果蝇 3 个物种的核小体偏好序列与排斥序列进行分类，准确率分别达到 94.94%、77.60%和 86.81%，支持了 DNA 序列定位的观点. Liu 等[34]提出了一个曲率相位(curvature profile)模型用于预测核小体定位，发现在 TSS 与 TFBS 序列排斥核小体，SNP 位点附近也有一个 NFR，而且 miRNA 靶位点则倾向于形成核小体.

概括地说，目前人们对核小体定位达成了初步的共识：NFR 与 ployA 序列是低等真核生物(酵母与蠕虫)中核小体定位的主要因素，但在高等生物中，DNA 序列只决定了一部分的核小体定位，更多的核小体服从统计定位. 此外，CpG 岛的甲基化[14, 35]、组蛋白变异与翻译后修饰[36]、以及染色质改构复合物[37]也不同程度地影响核小体定位.

2 核小体定位的基因表达调控功能研究进展

随着多种模式生物，包括酵母(*Saccharomyces cerevisiae*)[16-17, 22]、线虫(*Caenorhabditis elegans*)[38]、果蝇(*Drosophila melanogaster*)[24, 39]、小鼠(*mouse*)[40]，人(human)[41-42]等，全基因组核小体定位图谱公开，许多研究者对核小体定位在基因表达调控中的功能开始进行探索. 研究结果表明，核小体定位在基因转录起始阶段、转录延伸阶段、基因表达模式多样化 (expression pattern divergence) 以及可变剪接 (alternative splicing，也称为选择性剪接)等过程中发挥着重要的调控作用，分述如下：

2.1 核小体参与转录起始调控

当核小体定位于启动子区域内并且占据 TFBS 时，将阻碍转录因子与 TFBS 的结合，从而抑制基因转录. Yuan 等[22]采用嵌合芯片(tiling array)测定了啤酒酵母第Ⅲ号染色体以及其他 223 个调控区域共 2 278 个核小体的定位，发现大多数功能性 TFBS 位于核小体缺乏区域，–1 核小体上的结合位点通常位于背向组蛋白八聚体的双螺旋大沟上，这恰好从反面说明占据了 TFBS 的核小体通常会阻碍转录起始. Bernstein 等[15]分析了酵母全部启动子区域核小体的占位水平，发现启动子区域核小体的数量与启动子下游基因转录速率呈负相关. Lee 等[16]测定了啤酒酵母全基因组的核小体定位图谱，发现核小体的分布特征与转录速率以及转录本的丰度密切相关，并且发现功能相关的基因正好可以根据它们启动子区域内核小体的分布模式聚为一类. 而且有些基因即使处于沉默状态时，启动子区域也存在稳定的 NFR，这表明 NFR 是启动子的一种开放稳定的状态. 对于缺乏核心启动子元件的基因，转录复合物无法通过常规的方法识别转录起始位点(TSS)，但 TSS 与+1 核小体在同一物种中总是成对出现且间距相等[6]，这表明转录机制极有可能是依靠 NFR 来识别转录的起始位点[4]. Leimgruber 等[43]则对核小体缺乏程度与基因表达的关系进行了研究，发现人类 MHC Ⅱ 类基因 HLA-DRA 的启动子区域核小体严重缺乏，如果替换 NFR 但不改变核心启动子序列，将导致 TSS 发生偏移，这更进一步暗示，HLA-DRA 基因群的转录机制是通过 NFR，而不是启动子序列来识别 TSS 的. 此外，Wan 等[44]统计分析了多个酵母的核小体定位数据集，发现启动子区域内核小体呈现周期性分布，且启动子区域内核小体分布的周期性越强，基因转录水平越低，反之表达水平则越高，这说明启动子内的核小体定位模式影响基因转录水平.

根据启动子跨越 DNA 序列的长度以及转录起始模式，可将后生动物(metazoan)的启动子分为“聚集型”与“分散型”两类. Rach 等[45]分析了果蝇启动子区域的核小体定位，发现这两类不同的启动子恰好具有不同的染色质结构：“分散型”启动子 TSS 下游的核小体定位较强，NFR 区域更确定；“聚集型”启动子的核小体排列较弱，但 RNA P Ⅱ 出现的概率要高，H2A.Z 变体与 H3K4 甲基化也有差别. 而且这些差别在哺乳动物与果蝇之间是保守的，依据这两种类型的核小体对启动子进行分类，比用 CpG 岛的出现与缺乏或者 RNA 聚合酶延伸减速来划分启动子更清晰，更能反映核小体在转录调

控中的作用．Poirier 等[46]通过使用荧光共振能转换(FRET)的方法观察大肠杆菌染色质的动态变化，发现核小体分布具有“紧凑”与“松散”两种可以相互转换的构象状态，“紧凑”的核小体结构是高度动态可变的，在 ms～s 的时间粒度内便产生构象上的改变，瞬时暴露结合位点，从而使得 TF 能结合到被核小体遮盖的靶位点，与靶位点结合的 TF 促使核小体与 DNA 解离，使“紧凑”构象向“松散”构象转换．

上述研究结果表明，核小体定位与转录相关的蛋白质因子(包括 RNA 聚合酶、TF、染色质改构复合物等)之间是相互联系的，它们以一种协调的、系统的方式共同调控基因的转录水平．Raveh-Sadka 等[47]将核小体定位、TF、RNA PⅡ、TFBS 等整合到一个定量预测转录水平的热力学模型中，比不考虑核小体定位能更准确地预测基因转录水平．

2.2 核小体参与转录延伸调控

最近研究表明，核小体不仅在转录起始阶段扮演着重要的角色，在转录延伸阶段也起着重要作用．Belch 等[48]发现，酵母基因外显子区域的核小体定位较弱，核小体能快速地解离与重构，便于 RNA 聚合酶顺利地通过．编码区 DNA 序列偏好性对核小体定位起着主导作用，但个别核小体定位受转录相关的各种蛋白质因子的影响．在基因间区域，核小体定位受序列偏好程度、屏障核小体以及其他蛋白质因子相互作用的影响，共同控制着 TFBS 的可结合性，从而调控基因的转录水平．Fan 等[49]认为，常规染色质免疫共沉淀(ChIP)实验中使用的微球菌核酸酶(MNase)趋向于在启动子与终止子区域进行切割，在一定程度上造成观察到的核小体定位模式 5′NFR 与 3′NFR 比真实的显著，于是他们使用一种改进的方法测定了酵母的核小体定位图谱，证实了 5′NFR 的存在，但 3′NFR 并不如预料中的显著，却发现 3′转录终止区的核小体与相邻基因的方向、距离以及酵母的生长条件具有较强的相关性．而且 RNA PⅡ迅速从转录终止区脱离，便于核小体迅速重构，表明 3′NFR 的形成不仅与 RNA PⅡ的延伸有关，而且与 3′端的构象相关．Hodges 等[7]通过光镊实验观察单个 RNA PⅡ在转录延伸阶段的行为，发现编码区的核小体会导致 RNA PⅡ多次停顿，RNA PⅡ恢复延伸的速度较缓，导致转录速率下降.与 Belch 等[48]所观察的结果不同的是，RNA PⅡ在延伸阶段并不会导致 DNA 从组蛋白上解离，而是像棘轮一样沿着念珠状的 DNA 向前推进，当碰到核小体时，DNA 临时弯曲成环状使组蛋白位置临时平移，帮助棘齿正好咬合核小体．由此可见，RNA 聚合酶在延伸阶段是如何读取缠绕在核小体上 DNA 的机制尚不清楚，亟待更多的研究．

2.3 核小体定位的动态性与基因表达模式变化密切相关

真核细胞内核小体定位是动态变化的，而且这种动态性与基因表达模式的变化密切相关[10-11, 50-52]．多个研究小组[50-52]发现包含 TATA 盒的基因(简称为 TATA 基因)启动子区域内核小体较多，不包含 TATA 盒的基因(简称为非 TATA 基因)启动子区域核小体较缺乏．有趣的是，TATA 基因通常是应激反应基因，转录水平的可塑性(plasticity)较强，表达调控进化速度较快．而非 TATA 基因通常是“看家”基因，转录水平连续稳定且进化较慢．这极有可能是核小体与 TF 竞争结合到 TATA 位点，使得 TF 与 TATA 的结合不稳定，导致转录水平出现快速的起伏．Tirosh 等[50-52]分析了酵母启动子区域的核小体定位模式与表达模式可塑性之间的关系，得到了与上述一致的结果，并将这两种核小体定位模式总结为启动子调控基因表达的两种策略：第一种启动子的 NFR 较宽且靠近 TSS，上游的核小体定位高度确定，TFBS 集中分布在 NFR 中，这类基因的转录水平连续稳定；第二种启动子区域内 NFR 缺乏，核小体均匀动态定位且与 TSS 的距离较近，TFBS 分散在启动子区域内，TFBS 侧翼序列的变异使其容易形成核小体，这类基因转录水平起伏变化较大，可塑性较强．这两种启动子特征在人类中是保守的，核小体在这两种启动子上的分布特征与两类基因的表达模式是一致的，说明核小体的动态定位增强了基因表达模式的可塑性，有利于生物适应环境的改变，是基因表达调控进化的一种方式，这正是当前表观遗传学研究的热点[51-52]．Vaillant 等[53]发现，酵母编码区内核小体定位有两种较稳定的模式，两种模式都是由 5′NFR 与 3′NFR 之间的距离决定的．第一种是 5′NFR 与 3′NFR 之间的正好容纳 n 个规则排列的核小体，构成“类晶体”结构；第二种模式则形成 n 个微压缩的核小体或者 n+1 个高度压缩的核小体结构，即在 n 与 n+1 个核小体两种结构状态之间转换，构成“双稳态”结构．“双稳态”结构与染色质的动态性质相关，相对于“类晶体”结构，“双稳态”型

基因具有更高的转录可塑性，对染色质改构复合物更敏感，并且具有较高的 H3 转换率和较低的 H2A.Z 丰度．这些结果表明，编码区核小体定位动态性、染色质结构的稳定性与压缩级别也是一种转录调控手段．He 等[54]发现，前列腺癌细胞中雄激素受体主要结合在增强子(enhancer)区域，而雄激素治疗会导致雄激素受体结合位点的一个中心核小体解离．Schones 等[26]较系统地研究了核小体定位的动态调控特性，他们使用 Solexa 高通量测序仪测定了静息与活化的人类 CD4+T 细胞的核小体分布，发现活化细胞中+1 核小体定位的尖峰在+40 bp 处，而静息细胞+1 核小体定位的尖峰在+10 bp 处，表明随着细胞状态的改变核小体定位相应地动态调整至另一种均衡状态．更重要的是，他们发现 TCR 信号诱导 CD4+T细胞表达后导致了核小体重新定位，诱导表达基因的+1 与+2 核小体定位增强，但–1 核小体定位无显著变化，而通过 TCR 信号抑制细胞表达，则发现+1 与–1 核小体定位都显著增强．还发现静息细胞与活化细胞功能性增强子元件 AI1 与 CNS1 区域的核小体定位也表现出显著的差异，启动子附近的组蛋白修饰也会导致核小体定位的改变．Nie 等[55]也发现，CD4+T 静息细胞的 NF-κB 结合位点区域核小体较密集，当 CD4+T细胞活化之后，NF-κB 结合位点区域的核小体显著减少．

上述研究表明，核小体定位的动态改变与基因表达模式的变化之间是密切相关的，虽然核小体定位的改变不是驱使基因表达模式改变的源始动力，但至少反映了基因表达模式的改变是通过暴露或隐藏启动子区域内某些 TF 结合位点或增强子元件、或者调整局部染色质结构，使基因的转录水平改变而最终达到的．

2.4 核小体定位在可变剪接中标记外显子

研究还表明核小体在基因编码区内倾向定位于外显子上，具有标记外显子的功能．陈伟等[56]根据核小体偏好序列和排斥序列的碱基分布特征，使用多样性增量二次判别方法(IDQD)对核小体在人类基因组剪接位点(GT/AG)邻近序列中的分布方式进行了研究，发现外显子序列通常倾向参与核小体的形成，对应的 RNA 片段具有较强的刚性，而剪接位点及其邻近的内含子(intron)序列则排斥核小体的形成，对应的 RNA 片段具有较强的柔性，而且 DNA 序列的核小体偏好 / 缺失和 RNA 的刚性 / 柔性具有统计相关性．Tilgner 等[8]比较分析了秀丽隐杆线虫与人类外显子区域的核小体定位，也发现外显子上的核小体较多，而内含子区域的核小体较缺乏，而且核小体趋向于定位在外显子的中心区域而非剪接点两侧．沉默基因上的核小体定位也有类似规律，这些结果显示，核小体定位可能包含 RNA 剪接的信号．Schwartz 等[9]分析了人类、果蝇与线虫全基因组范围内的核小体定位，也得到了同样的结果，并且发现外显子区域内核小体占位水平与 GC 的含量相关，而在内含子区则没有观察到这种相关性，说明核小体偏向于定位在外显子区域无法用序列偏好性来解释，而是与组蛋白修饰和剪接有关．更有趣的是，他们还发现，核小体的占据率与被剪接机制识别为外显子的可能性成正比，即内含子上的核小体数量最少，选择性拼接外显子包含较少的核小体，而组成性外显子上的核小体数量最多．Tilgner 等[8]也发现假外显子(pseudoexons)上核小体相对缺乏．由此可见，外显子与内含子上的核小体定位差异有助于剪接机制识别外显子．Keren 等[57]提出了一个减速带模型，如图 3 所示，RNA 聚合酶在转录过程中通过核小体缺乏区域时速度较快，到达核小体密集区域时其转录速度下降，如图 3a 所示．当 RNA 聚合酶通过核小体时，与 RNA 聚合酶耦联的剪接因子(splicing factor，SF)结合到 mRNA 前体上外显子 3′与 5′端的剪接位点(splice site，SS)，便于剪接机制识别出外显子．由此，核小体定位实现了标记外显子的功能，如图 3b 所示．Hodges 等[7]的光镊实验观察到密集核小体区域导致 RNA PⅡ多次停顿且恢复转录的速度变缓，为核小体减速带模型提供了有力的证据．

外显子上的核小体还可通过组蛋白的修饰作用，帮助剪接机制以更高的效率识别外显子[58-61]．Kolasinska-Zwierz[59]等发现线虫外显子上的组蛋白容易发生 H3K36me3 修饰． Schwartz 等[9]使用 ChIP-seq 技术测定了人类与小鼠核小体定位，也发现核小体趋向定位于 3′端外显子上，H3K36me3 修饰与 RNA PⅡ也都倾向于发生在外显子上的核小体, 人类外显子还易于发生 H3K79me1、H4K20me1 与 H2BK5me1 三种组蛋白修饰，而且这种修饰作用是跨物种保守的，这些结果都表明外显子上组蛋白的修饰与剪接有很大的关系．Chodavarapu 等[60]通过微球菌核酸酶将拟南芥染色质降解为单个核小体的形式，然后用 Illumina GAⅡ测序仪对每个核小体 DNA 进行测序，并且和单个核苷酸位点分辨率的 DNA 甲基化模式进行比较，发现核小体 DNA

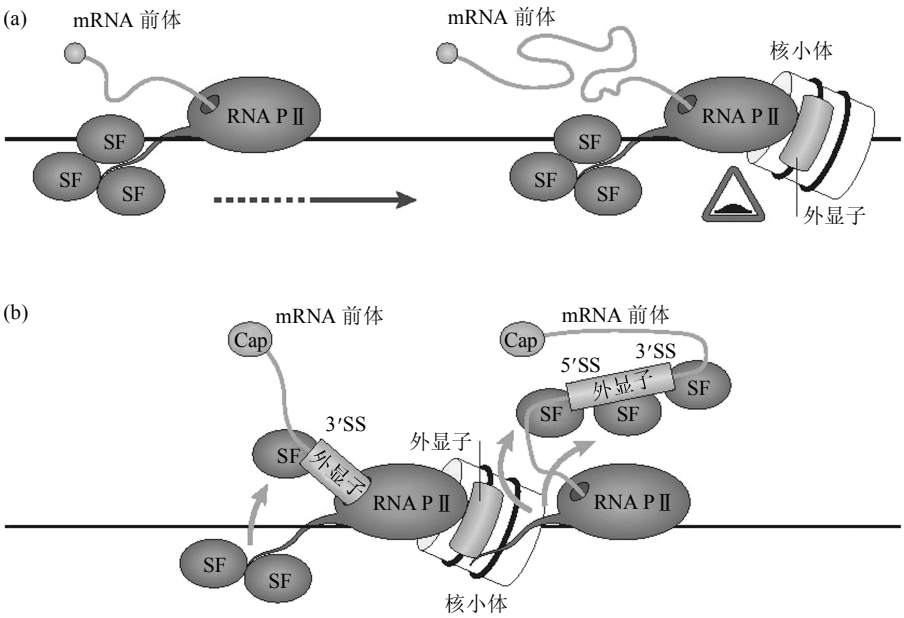

Fig. 3　Speed bump model of nucleosome occupancy[57]

图 3　核小体减速带模型[57]

(a)核小体倾向于定位在外显子(Exon)使得 RNA 聚合酶(RNA P Ⅱ)遇到外显子时延伸速度下降；(b)与 RNA 聚合酶耦联的剪接因子(splicing factor，SF)结合到 mRNA 前体(pre-mRNA)上外显子 3′与 5′端的剪接位点(splice site，SS)，表明核小体在可变剪接中具有标记外显子的作用.

的甲基化水平比连接 DNA 的甲基化水平要高．这表明全基因组范围内的核小体定位影响 DNA 甲基化模式，即 DNA 甲基化酶倾向于对缠绕在组蛋白上的 DNA 进行修饰，而且这种甲基化规律在人类的染色质上是保守的．Chodavarapu 等也观察到核小体偏向定位于外显子区域，甲基化和 RNA 聚合酶也偏向于发生在外显子上．不同的是，核小体倾向于定位在外显子与内含子交界处而不是外显子中心．尽管如此，上述研究的结论基本上是一致的，表明核小体定位与 DNA 甲基化具有标记外显子的作用，而且影响 RNA 聚合酶的活性．Luco 等[61]的研究结果则提供了更直接的证据，他们通过改变正常前列腺上皮细胞(PNT2)与成人骨髓间质干细胞(hMSC)中 FGFR2 基因的 H3K36me3 修饰水平后，然后观察组蛋白修饰与可变剪接(alternative splicing，AS)产物中是否包含外显子序列．他们发现，组蛋白修饰与剪接产物之间具有对应关系，而且调整组蛋白修饰信号会相应地导致剪接位点的改变，这种改变是通过组蛋白修饰介导多聚嘧啶序列结合蛋白(PTB)影响剪接调控子的募集而达到的．然而，Dhami 等[62]认为情况并不是如此简单，他们分析了人类三种细胞类型(K562 细胞、U937 细胞、CD14+单核细胞)的核小体定位，发现在组蛋白修饰与 RNA P Ⅱ只是标记了部分外显子，在 K562 与 U937 中甚至趋向于发生在内含子中，表明核小体定位标记外显子的机制仍不清楚，有待更深入的研究．

核小体定位于外显子的机制及其功能可能不仅限于上述所讨论的．例如，外显子缠绕在核小体在一定程度上能保护核小体 DNA 免受紫外线辐射和其他的损伤，从而降低了外显子序列的变异速率，增强了其保守性[63].

3　总结与展望

现有的研究结果充分说明核小体不仅是压缩 DNA 的物质基础，而且在基因转录调控中起着重要的作用．然而，核小体定位与转录调控之间不是一种简单、线性的关系，而是与转录相关蛋白质之间相互作用，使染色质压缩程度与 DNA 序列的可结合性处于一种动态平衡的状态[64]，从而准确地调控基因的转录水平．因此，我们可以从以下两个方面来总结核小体定位对转录调控的影响．

一方面，核小体抑制转录并且影响 DNA 序列与调控的进化．组蛋白占据 TFBS，阻碍转录因子、RNA 聚合酶等蛋白质与 DNA 结合，抑制基因转录，起着负调控子的功能．核小体定位对 DNA 序列的进化也有一定影响，进化分析表明，离核小体中心越近的序列多态性变异越大，两端序列的多态性变异程度随着与核小体中心的距离增大而降低，至连接序列达到最低[65-67]．一种合理的解释就是靠近核小体中心的序列，接触 DNA 修复蛋白的机会少，变异的概率增大．另一种解释是序列变异速率是相同的，但纯化选择淘汰了连接序列中的不利变异，而核小体序列中的中性变异固定下来．孙啸研究小组最近的研究还表明，核小体定位不仅对编码基因的转录具有调控作用，对 microRNA 的转录同样具有调控作用[68]．

另一方面，越来越多的证据表明，核小体定位在基因表达调控中的作用是有限的，并且是由转录机制支配的．当核小体占据了 TFBS 时，并不是通过精确地占据与释放 TFBS 来调控转录，而是形成一个稀疏的核小体区域，以便 TF 与靶位点结合[69]．在细胞周期中，基因表达水平起伏变化时，核小体定位整体上也没有出现大幅度的变化[70]，这说明转录机制会在需要的时候调整染色质结构，使组蛋白从 DNA 解离或者沿 DNA 平移，暴露出功能性 TFBS，使转录机制能顺利地完全装配与转录．要注意的是，组蛋白修饰(甲基化、乙酰化等)对于核小体定位、转录调控也是非常重要的，三者之间密切相关．Zhang 等[71]分析了人类 CD4+T 细胞甲基化组蛋白 ChIP-seq 数据，发现甲基化修饰的核小体主要分布在基因启动子、DNase Ⅰ敏感位点以及 TFBS 区域．Wang 等[72]也发现 p16 的表达受组蛋白去乙酰化的调控．

此外，核小体定位的动态性与表观遗传调控(epigenetic regulation)之间的关系已经成为了另一个研究热点．核小体定位的动态性主要体现在不同细胞类型、细胞的不同状态、不同发育阶段以及特定刺激等条件下，核小体定位从一种均衡状态转换至另一种均衡状态．这种动态均衡特性是受核小体定位序列、ATP 依赖的染色质改构复合物、组蛋白修饰、DNA 结合蛋白等因子的共同作用下形成的[73]，而且核小体定位的动态均衡特性与表现遗传的稳定性是密切相关的[74]．不仅如此，核小体定位还表现出一种微观状态下的动态平衡性，Poirier 等的研究结果提供了直接的证据(见 **2.1** 节)[41]，Poirier 等发现“紧凑”的核小体阵列能在 ms～s 时间粒度内产生构象上的微小变动，正是这种变动使 TFBS 瞬间暴露，TF 才能结合到位于核小体中心的靶位点，进而促使核小体定位从“紧凑”构象向“松散” 构象转变，我们推测核小体定位的这种微观扰动 可能是核小体定相(nucleosome phasing)转换的生化基础．

尽管当前人们对核小体定位及其功能方面开展了许多研究并取得初步的认识，但还有许多问题亟待深入研究．我们相信以下三个方面将是核小体定位机制及其功能研究的热点与重点：a．核小体定位与转录机制之间的相互作用关系及其对基因表达的影响；b．核小体定位在序列进化与调控进化中的功能研究；c．核小体定位动力学与表观遗传基因调控之间的关系．

核小体定位及其功能的研究是一项艰巨的任务，幸运的是，随着下一代测序技术的发展及应用，越来越多的核小体定位图谱将会公布出来．通过对这些海量数据的分析与现有研究结果的整合，帮助我们理解核小体定位及其在基因表达调控方面的功能并建立新的理论模型，最终揭示核小体定位在生命过程及重大疾病发生与发展中的作用．

参 考 文 献

[1] Luger K, Mader A W, Richmond R K, *et al*. Crystal structure of the nucleosome core particle at 2.8Å resolution. Nature, 1997, **389**(6648): 251−260

[2] Richmond T J, Davey C A. The structure of DNA in the nucleosome core. Nature, 2003, **423**(6936): 145−150

[3] Albert I, Mavrich T N, Tomsho L P, *et al*. Translational and rotational settings of H2A.Z nucleosomes across the *Saccharomyces cerevisiae* genome. Nature, 2007, **446**(7135): 572−576

[4] Kornberg R D, Lorch Y. Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. Cell, 1999, **98**(3): 285−294

[5] Jiang C, Pugh B F. Nucleosome positioning and gene regulation: advances through genomics. Nat Rev Gene, 2009, **10**(3): 161−172

[6] Segal E, Widom J. What controls nucleosome positions?. Trends in Genetics, 2009, **25**(8): 335−343

[7] Hodges C, Bintu L, Lubkowska L, *et al*. Nucleosomal fluctuations govern the transcription dynamics of RNA polymerase Ⅱ. Science, 2009, **325**(5940): 626−628

[8] Tilgner H, Nikolaou C, Althammer S, *et al*. Nucleosome positioning as a determinant of exon recognition. Nat Struc Mol Biol, 2009, **16**(9): 996−1002

[9] Schwartz S, Meshorer E, Ast G. Chromatin organization marks exon-intron structure. Nat Struc Mol Biol, 2009, **16**(9): 990−995

[10] Choi J K, Kim Y J. Epigenetic regulation and the variability of gene

expression. Nat Gene, 2008, **40**(2): 141-147

[11] Choi J K, Kim Y J. Intrinsic variability of gene expression encoded in nucleosome positioning sequences. Nat Gene, 2009, **41**(4): 498-503

[12] Stunkel W, Kober I, Seifart K H. A nucleosome positioned in the distal promoter region activates transcription of the human U6 gene. Mol Cell Biol, 1997, **17**(8): 4397-4405

[13] Lee C K, Shibata Y, Rao B, *et al*. Evidence for nucleosome depletion at active regulatory regions genome- wide. Nature Genetics, 2004, **36**(8): 900-905

[14] Sekinger E A, Moqtaderi Z, Struhl K. Intrinsic histone-DNA interactions and low nucleosome density are important for preferential accessibility of promoter regions in yeast. Mol Cell, 2005, **18**(6): 735-748

[15] Bernstein B E, Liu C L, Humphrey E L, *et al*. Global nucleosome occupancy in yeast. Gen Biol, 2004, **5**(9): R62

[16] Lee W, Tillo D, Bray N, *et al*. A high-resolution atlas of nucleosome occupancy in yeast. Nat Gene, 2007, **39**(10): 1235-1244

[17] Segal E, Fondufe-Mittendorf Y, Chen L, *et al*. A genomic code for nucleosome positioning. Nature, 2006, **442**(7104): 772-778

[18] Field Y, Kaplan N, Fondufe-Mittendorf Y, *et al*. Distinct modes of regulation by chromatin encoded through nucleosome positioning signals. PLoS Compu Biol, 2008, **4**(11): e1000216

[19] Kaplan N, Moore I K, Fondufe-Mittendorf Y, *et al*. The DNA-encoded nucleosome organization of a eukaryotic genome. Nature, 2009, **458**(7236): 362-366

[20] Wang J P, Fondufe-Mittendorf Y, Xi L, *et al*. Preferentially quantized linker DNA lengths in *Saccharomyces cerevisiae*. PLoS Comput Biol, 2008, **4**(9): e1000175

[21] Kornberg R D, Stryer L. Statistical distributions of nucleosomes: Nonrandom locations by a stochastic mechanism. Nucl Acid Res, 1988, **16**(14A): 6677-6690

[22] Yuan G C, Liu Y J, Dion M F, *et al*. Genome-scale identification of nucleosome positions in *S. cerevisiae*. Science, 2005, **309**(5734): 626-630

[23] Mavrich T N, Ioshikhes I P, Venters B J, *et al*. A barrier nucleosome model forstatistical positioning of nucleosomes throughout the yeast genome. Genome Res, 2008, **18**(7): 1073-1083

[24] Mavrich T N, Jiang C, Ioshikhes I P, *et al*. Nucleosome organization in the *Drosophila* genome. Nature, 2008, **453**(7193): 358-362

[25] Fu Y, Sinha M, Peterson C L, *et al*. The insulator binding protein CTCF positions 20 nucleosomes around its binding sites across the human genome. PLoS Genet, 2008, **4**(7): e1000138

[26] Schones D E, Cui K, Cuddapah S, *et al*. Dynamic regulation of nucleosome positioning in the human genome. Cell, 2008, **132**(5): 887-898

[27] Zhang Y, Moqtaderi Z, Rattner B P, *et al*. Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions *in vivo*. Nat Struct Mol Biol, 2009, **16**(8): 847-852

[28] Stein A, Takasuka T E, Collings C K. Are nucleosome positions *in vivo* primarily determined by histone-DNA sequence preferences?. Nucl Acid Res, 2009, **38**(3): 709-719

[29] Morris R T, O′Connor T R, Wyrick J J. Ceres: software for the integrated analysis of transcription factor binding sites and nucleosome positions in *Saccharomyces cerevisiae*. Bioinformatics, 2010, **26**(2): 168-174

[30] Yi X, Cai Y D, He Z, *et al*. Prediction of nucleosome positioning based on transcription factor binding sites. PloS One, 2010, **5**(9): e12495

[31] Tolstorukov M Y, Choudhary V, Olson W K, *et al*. NuScore: A web- interface for nucleosome positioning predictions. Bioinformatics, 2008, **24**(12): 1456-1458

[32] Cai L, Zhao X, Pei Z, *et al*. Predicting nucleosome positioning based on DNA sequence. Proceeding of 4th Ch on Bioinformatics & Systems Biology, 2010: 158-160

[33] Zhao X, Pei Z, Liu J, *et al*. Prediction of nucleosome DNA formation potential and nucleosome positioning using increment of diversity combined with quadratic discriminant analysis. Chrom Res, 2010, **18**(7): 777-785

[34] Liu H, Duan X, Yu S, *et al*. Analysis of nucleosome positioning determined by DNA helix curvature in the human genome. BMC Genomics, 2011, **12**: 72

[35] Nathan D, Crothers D M. Bending and flexibility of methylated and unmethylated EcoRI DNA. J Mol Biol, 2002, **316**(1): 7-17

[36] Yang Z, Zheng C, Hayes J J. The core histone tail domains contribute to sequence-dependent nucleosome positioning. J Biol Chem, 2007, **282**(11): 7930-7938

[37] Yuan G C, Liu J S. Genomic sequence is highly predictive of local nucleosome depletion. PLoS Compu Biol, 2008, **4**(1): e13

[38] Valouev A, Ichikawa J, Tonthat T, *et al*. A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning. Genome Res, 2008, **18**(7): 1051-1063

[39] Guertin M J, Lis J T. Chromatin landscape dictates HSF binding to target DNA elements. PLoS Gen, 2010, **6**(9), pii: e1001114

[40] Li Z, Schug J, Tuteja G, *et al*. The nucleosome map of the mammalian liver. Nat Struct Mol Biol, 2011, **18**(6): 742-746

[41] Lidor Nili E, Field Y, Lubling Y, *et al*. p53 binds preferentially to genomic regions with high DNA-encoded nucleosome occupancy. Genome Res, 2010, **20**(10): 1361-1368

[42] Valouev A, Johnson S M, Boyd S D, *et al*. Determinants of nucleosome organization in primary human cells. Nature, 2011, **474**(7352): 516-520

[43] Leimgruber E, Seguin-Estevez Q, Dunand-Sauthier I, *et al*. Nucleosome eviction from MHC class Ⅱ promoters controls positioning of the transcription start site. Nucl Acid Res, 2009, **37**(8): 2514-2528

[44] Wan J, Lin J, Zack D J, *et al*. Relating periodicity of nucleosome organization and gene regulation. Bioinformatics, 2009, **25**(14): 1782-1788

[45] Rach E A, Winter D R, Ashlee M. Transcription initiation patterns

indicate divergent strategies for gene regulation at the chromatin level. PLoS Genetics, 2011, **7**(1): e1001274

[46] Poirier M G, Oh E, Tims H S, *et al*. Dynamics and function of compact nucleosome arrays. Nat Struc Mol Biol, 2009, **16** (9): 938−944

[47] Raveh-Sadka T, Levo M, Segal E. Incorporating nucleosomes into thermodynamic models of transcription regulation. Genome Res, 2009, **19**(8): 1480−1496

[48] Belch Y, Yang J, Liu Y, *et al*. Weakly positioned nucleosomes enhance the transcriptional competency of chromatin. PLoS ONE, 2010, **5**(9): e12984

[49] Fan X, Moqtaderi Z, Jin Y, *et al*. Nucleosome depletion at yeast terminators is not intrinsic and can occur by a transcriptional mechanism linked to 3′-end formation. Proc Natl Acad Sci USA, 2010, **107**(42): 17945−17950

[50] Tirosh I, Barkai, N. Two strategies for gene regulation by promoter nucleosomes. Genome Res, 2008, **18**(7): 1084−1091

[51] Tirosh I, Weinberger A, Bezalel D, *et al*. On the relation between promoter divergence and gene expression evolution. Mol Sys Biol, 2008, **4**: 159

[52] Tirosh I, Reikhav S, Levy A A, *et al*. A yeast hybrid provides insight into the evolution of gene expression regulation. Science, 2009, **324**(5927): 659−662

[53] Vaillant C, Palmeira L, Chevereau G, *et al*. A novel strategy of transcription regulation by intragenic nucleosome ordering. Genome Res, 2010, **20**(1): 59−67

[54] He H H, Meyer C A, Shin H, *et al*. Nucleosome dynamics define transcriptional enhancers. Nature Genetics, 2010, **42**(4): 343−348

[55] Nie Y, Sun X. Dynamic Nucleosome Positioning around Functional Transcription Factor Binding Sites in the Promoters of Inducible NF-kappaB Target Genes. Bioinformatics and Biomedical Engineering(iCBBE), 2011: 1−3

[56] 陈 伟, 罗辽复, 张利绒, 等. 核小体定位与 RNA 剪接. 生物化学与生物物理进展, 2009, **36**(8): 1035−1040
Chen W, Luo L F, Zhang L R, *et al*. Prog Biochem Biophys, 2009, **36**(8): 1035−1040

[57] Keren H, Lev-Maor G, Ast G. Alternative splicing and evolution: diversification, exon definition and function. Nat Rev Genetics, 2010, **11**(5): 345−55

[58] Andersson R, Enroth S, Rada-Iglesias A, *et al*. Nucleosomes are well positioned in exons and carry characteristic histone modifications. Genome Res, 2009, **19**(10): 1732−1741

[59] Kolasinska-Zwierz P, Down T, Latorre I, *et al*. Differential chromatin marking of introns and expressed exons by H3K36me3. Nat Gen, 2009, **41**(3): 376−381

[60] Chodavarapu R K, Feng S, Bernatavichute Y V, *et al*. Relationship between nucleosome positioning and DNA methylation. Nature, 2010, **466**(7304): 388−392

[61] Luco R F, Pan Q, Tominaga K, *et al*. Regulation of alternative splicing by histone modifications. Science, 2010, **327**(5968): 996−1000

[62] Dhami P, Saffrey P, Bruce A W, *et al*. Complex exon-intron marking by histone modifications is not determined solely by nucleosome distribution. PLoS ONE, 2010, **5**(8): e12339

[63] Kaplan C D. Revealing the hidden relationship between nucleosomes and splicing. Cell Cycle, 2009, **8**(22): 3633−3634

[64] Cairns B R. The logic of chromatin architecture and remodelling at promoters. Nature, 2009, **461**(7261): 193−819

[65] Warnecke T, Batada N N, Hurst L D. The impact of the nucleosome code on protein-coding sequence evolution in yeast. PLoS Genetics, 2008, **4**(11): e1000250

[66] Washietl S, Machne R, Goldman N. Evolutionary footprints of nucleosome positions in yeast. Trends Genetics, 2008, **24** (12): 583−587

[67] Choi J K, Kim Y J. Implications of the nucleosome code in regulatory variation, adaptation and evolution. Epigenetics, 2009, **4**(5): 291−295

[68] Liu H D, Zhang D J, Xie J M, *et al*. Analysis of nucleosome positioning in promoters of miRNA genes and protein-coding genes. Chin Sci Bull, 2010, **55**(22): 2347−2352

[69] Goh W S, Orlov Y, Li J, *et al*. Blurring of high-resolution data shows that the effect of intrinsic nucleosome occupancy on transcription factor binding is mostly regional, not local. PLoS Compu Biol, 2010, **6**(1): e1000649

[70] Hogan G J, Lee C K, Lieb J D. Cell cycle-specified fluctuation of nucleosome occupancy at gene promoters. PLoS Genetics, 2006, **2**(9): e158

[71] Zhang Y, Shin H, Song J S, *et al*. Identifying positioned nucleosomes with epigenetic marks in human from ChIP-Seq. BMC Genomics, 2008, **9**: 537

[72] 王秀莉, 冯云鹏, 赵 静, 等. 可逆的组蛋白乙酰化调控 p16 的表达. 生物化学与生物物理进展, 2010, **36**(6): 600−606
Wang X L, Feng Y P, Zhao J, *et al*. Prog Biochem Biophys, 2010, **36**(6): 600−606

[73] Henikoff S. Nucleosome destabilization in the epigenetic regulation of gene expression. Nat Rev Gen, 2008, **9**(1): 15−26

[74] Korber P, Becker P B. Nucleosome dynamics and epigenetic stability. Essays Biochem, 2010, **48**(1): 63−74

Transcriptional Regulation Functions of Nucleosome Positioning: a Survey*

LIU Hui[1], Zhuang Zi-Heng[1], GUAN Ji-Hong[2], ZHOU Shui-Geng[3]**

([1]) *School of Pharmaceutical Engineering & Life Science, Changzhou University, Changzhou* 213164, China;

[2] *Department of Computer Science & Technology, Tongji University, Shanghai* 201804, China;

[3] *Shanghai Key Laboratory of Intelligent Information Processing, Fudan University, Shanghai* 200433, China)

Abstract Nucleosome is the building unit of eukaryotic chromatin, the location of histone octamer on the DNA sequence is called nucleosome positioning. With the advent of high throughout technologies such as ChIP-chip and ChIP-seq. large-scale nucleosome positioning atlas of multiple model organisms have been measured, which attract many researchers to investigate nucleosome positioning and its functions on transcriptional regulation. In this paper, we first introduce the concept of nucleosome positioning and summarize the regular patterns of nucleosome positioning on genetic region, then review the major advances of functions of nucleosome positioning on transcriptional initiation, elongation, divergence of expression patterns and alternative splicing.

Key words nucleosome positioning, transcriptional regulation, alternative splicing, gene expression, chromatin

DOI: 10.3724/SP.J.1206.2011.00252

* This work was supported by grants from National Basic Research Program of China (973)(2010CB227100), The National Natural Science Foundation of China(31100954, 61173118, 61272380), Open Research Fund Program of Shanghai Key Laboratory of Intelligent Information Processing, Fudan University (IIPL-2010-011), and PhD Program Scholarship Fund of Changzhou University (CCZUPY1001).

**Corresponding author.

Tel: 86-21-65654504, E-mail: sgzhou@fudan.edu.cn

Received: August 5, 2011 Accepted: September 8, 2011