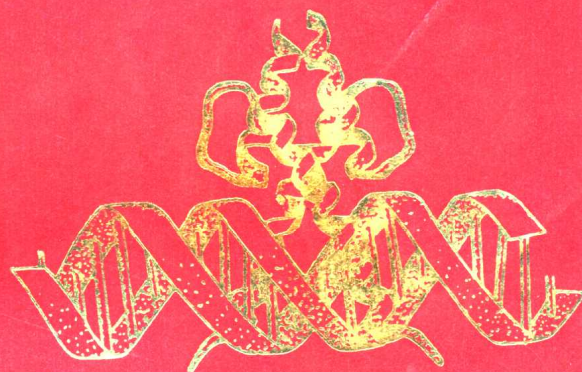


真核基因表达调控

THE CONTROL OF
EUKARYOTIC GENE EXPRESSION

沈翊珩 方福德 主编



高等教育出版社

真核基因表达调控

吴阶平题



主 编
沈翊珩 方福德

高等教育出版社

(京)112号

内 容 提 要

真核基因表达调控是当代生命科学研究的前沿、热点。本书共18章,从基因外调控、基因表达各层次等方面,多角度地介绍了真核基因表达调控的基础知识和基本研究方法。内容详实,图文并茂,既注重基本原理和研究进展的阐述,也注重相关的研究方法和研究技术的介绍,书末附有简要的名词解释和索引。

本书可供高等学校师生和生命科学研究人员参考,特别适用于综合性大学生物系和医学院校研究生教学。

责任编辑:谭丽霞、吴雪梅

图书在版编目(CIP)数据

真核基因表达调控. —北京:高等教育出版社,1996

ISBN 7-04-005541-4

I. 真… II. Ⅲ. 调节基因:基因表达—遗传控制:遗传
调节 IV. Q756

中国版本图书馆 CIP 数据核字(95)第 12984 号

*

高等教育出版社出版

北京沙滩后街 55 号

邮政编码:100009 传真:4014048 电话:4054588

新华书店总店北京发行所发行

国防工业出版社印刷厂印刷

*

开本 787×1092 1/16 印张 22 字数 540 000

1996年2月第1版 1996年2月第1次印刷

印数 0001—1 921

定价 31.40 元

凡购买高等教育出版社的图书,如有缺页、倒页、脱页等
质量问题者,请与当地图书销售部门联系调换。

版权所有,不得翻印

序

对基因的深入认识是本世纪生物科学最重要的研究成就,它奠定了分子生物学快速发展的基础。基因研究的重点是活细胞中基因的结构和功能。基因功能的体现取决于其结构和表达调控状况。如果把基因结构比作“硬件”,那么基因表达的调控就是把这些“硬件”变换成多种多样基因功能的“软件”和操作者。现在我们已经初步看到,基因表达的调控是一个十分复杂而协调有序的运作过程,这一过程不仅与基因本身的功能,也与细胞及机体的功能表现密切相关。毋庸置疑,对基因表达调控的研究将进一步阐明重要的生命现象,解释细胞行为和疾病的发生机理,从而在分子水平上为人类疾病的诊断、治疗和预防提供科学依据和实用技术。

由于上述原因,基因表达的调控已成为生命科学中重要的前沿研究领域。为适应该研究领域的发展,中国医学科学院、中国协和医科大学的一些专家近十余年来一直坚持在国内开展与基因表达调控有关的研究,并为研究生开设了《真核基因表达的调控》课程,系统地讲授基因表达调控的基本知识和最新研究进展。经过不断地充实和完善,最近他们对讲稿进行了更新和整理,并聘请少数院外专家参加撰写,形成我国第一部有关《真核基因表达调控》的研究生教学参考书。我们相信,这本书的出版必将有助于我国相关领域研究生教学水平的提高,为我国生命科学的发展作出贡献。

由于基因表达调控的研究进展极其迅速,每个专业工作者都面临不断学习和更新知识的任务,因此我希望这本教学参考书也能经常修订、充实和提高,以便更好地、及时地反映出国际上科学研究前沿的进展。

吴昱

中国科学院院士
1994年9月23日

前 言

真核基因表达的调控是当前分子生物学这一前沿学科中的前沿研究领域,现有的研究表明,许多重要生命现象的深层次问题都集结于此,形成了众多的热点探索课题。在一定程度上可以说基因表达的调控是基因生物学乃至分子医学的真谛所在。

为了适应国际上在基因表达的调控研究中日新月异的发展,培养我国跨世纪的科学人才,及时将有关的新知识、新概念、新研究方法和新技术系统地介绍给在读研究生和有关的科学工作者,当属十分必要和迫切。有鉴于此,我们所在的中国协和医科大学和中国医学科学院研究生院早在 80 年代初期就开始在研究生课程中设置了《真核基因表达的调控》专题,每学年的教学已达 60 个学时。担负授课任务的教授主要来自中国医学科学院的医学分子生物学国家重点实验室和协和医大,他们都是活跃在相关领域科研、教学第一线的有经验的专家。研究生经过这段学习后,专业知识水平得到了显著的提高,科学底蕴得到了充实,为他们日后在生物医学有关领域进行创造性研究工作打下了坚实的基础。因此,本课程的设置受到了各方面的欢迎。

经过十余年来经验的积累,每位教授对自己所讲授的内容进行反复揣摩,不断充实,使其基本上已能反映当前国际发展的现状和趋势。从总体上说,这些教学内容和教材已成为我国生命科学领域研究生培养事业的宝贵知识财富。科教界一些同仁提出,若能将这些内容加以整理、出版和传播,必将为提高国内相关学科研究生教学质量作出更大的贡献。在这种情况下,我们约请了中国医学科学院基础医学研究所、中国协和医科大学基础医学院参加教学的有关教授,重新整理和补充讲稿,并特邀中国科学院生物化学研究所和现在美国 Duke 大学工作的有关领域专家根据他(她)们的专长撰写了有关内容,编辑成本书,奉献给教育界和科技界的广大读者。

本书共包括 18 章,其内容涵盖真核基因表达的调控的若干主要方面。在组织编写的过程中,我们始终力求在保证严谨的科学性和高水平的基础上,兼顾教学和科学研究的特点,突出以下特色:①在内容安排上,力求全面、系统,既立足于阐述基础知识和基本概念,也注重介绍最新研究成果和进展,以便做到既适合于研究生学习,又可作为研究人员和高校教师从事研究和教学的参考。②在结构安排上,突出强化意识。本书前五章比较系统地介绍真核基因多级调控的重点内容及有关最新进展;第 6~10 章着重以真核细胞为中心介绍多种生命现象中基因表达调控的作用;第 11、12 章有关酵母和病毒基因的调控是作为高等真核生物的简单模型和重要的可操作体系,以便加深对真核基因调控的认识;第 13~15 章对 3 个性质的基因家族进行个别的系统论述。我们希望这种安排会有助于读者从纵横、垂直、交叉等多方位、多角度理解基因调控机制及其复杂性,并从中反复领会基本概念和基本知识,启迪思维。③在阐述基本理论的同时,还介绍主要的研究方法和技术,值得提出的是,我们特别介绍了当今国际研究热点中以完整基因组为对象的人类基因组研究以及以转基因或基因“敲除”的整体动物为对象的有关研究技术,力求理论联系实际,使读者能较快捷地开展有关的研究工作,因而具有一定的实用价值。④专门安排了辅助阅读资料(第 18 章),广泛介绍了真核基因表达调控的研究

方法,并附有专业基础术语及其简释。对于读者特别是初学者进一步巩固和扩大相关知识可能会有所助益。这一章由于涉及内容较多,因而有较多同志参加编写,他们是沈翊珩、方福德、崔莲仙、陈兰英、吴宁华、沈岩、杜恺、莽锐、程小款、施一江、张敏磊、徐晓石、邢光新。

参加本书编写的各位作者,无论所负任务轻重,都全身心投入,从而使本书基本上达到了预期的要求。我们向这些同仁表示衷心的感谢和谢忱!

但是,正如上述,基因表达的调控是一个十分活跃的前沿研究领域,进展迅速,新的研究成果层出不穷,有些现在看起来合理和被接受的观点和结论,过一段时间后有可能被新的证据所补充、修正或推翻。因此,如若读者发现本书所介绍的某些资料和论据与最新文献资料和论据相抵触,是毫不奇怪的。科学的发展就是一个不断修正、不断前进的过程。基于此点认识,我们深感,对本书进行经常性的修订和再版是非常必要的。我们在此恳切期望有更多的国内外同仁今后能参与本书的修订和再版工作,使之不断完善。

此外,尽管我们在组织书稿时曾力求各章之间在行文风格、篇幅大小和掌握内容等方面大致保持协调一致,但由于没有相应的国内外图书作为编写的依据以及其它实际因素的限制,做得并不十分理想。此外,每一位编写人员,除承担有关的教学任务以外,都工作在繁忙的第一线科研岗位上,许多作者还承担了大量的、繁琐的业务行政管理工作,不仅写作的时间很难得到保证,同时每位作者所从事的研究工作领域也很不相同,对本书所涉及的有关领域的理论或研究进展的理解也不尽相同。因此我们遵循“百家争鸣”的基本原则,尊重每位作者的学术观点,不过分强求对稿件的组织和写作习惯上的统一。希望得到读者的谅解和指正,以便再版时加以改进。

在本书即将出版之际,我们怀着感激的心情,鸣谢高等教育出版社和北京协和医学音像出版社所给予的真诚无私的帮助!正是由于他们的帮助,本书才得以在较短时间内顺利付梓出版。本书在编写过程中还得到了老一辈科学家的热情关怀,中国科学院院士吴阶平、吴旻两位教授分别为本书题写书名和序,中国医学科学院李士澍、吴冠芸等教授审阅了部分章节,并提出了宝贵意见,在此一并致以诚挚的感谢。

沈翊珩 方福德

1994年9月20日于北京

谨以此书献给

中国协和医科大学生物化学系创建七十周年
(1924~1994)

目 录

第一章 真核细胞染色质结构与基因活化	沈玥珩 (1)
一、真核生物染色质的基本结构	(1)
二、染色质水平上的基因活化调节	(5)
三、核基质与基因活化	(10)
四、DNA 甲基化与真核基因表达	(11)
第二章 真核基因转录水平的调控	沈玥珩 (14)
一、真核基因的基础转录	(14)
二、基因转录水平的顺式调节	(19)
三、基因转录的反式作用因子	(23)
四、真核细胞基因转录调控的特征	(29)
第三章 转录后遗传信息的“扩展”	沈玥珩 (34)
一、mRNA 前体加工的分子基础	(34)
二、内含子的选择性剪接	(41)
三、“反式”剪接	(45)
四、RNA 编辑	(46)
五、结语	(50)
第四章 信使 RNA 与翻译水平的调控	强伯勤 (51)
一、蛋白质合成的起始作用	(51)
二、翻译因子的可逆磷酸化与蛋白质合成的控制	(53)
三、酵母 GCN4mRNA 翻译的调控	(57)
四、5' - 非翻译区结构与翻译起始作用的调节	(59)
五、3' - UTR 结构对翻译的调控作用	(62)
六、转铁蛋白 mRNA 的翻译调控	(63)
七、mRNA 的稳定性	(64)
八、小结	(67)
第五章 蛋白质磷酸化与真核基因调控	沈玥珩 (69)
一、真核细胞中的蛋白质磷酸激酶	(70)
二、蛋白质磷酸化的调节机制	(73)
三、转录水平的磷酸化调节	(76)
四、转录后的磷酸化调控	(82)
第六章 真核细胞的信息传递与基因表达	裴 钢 (84)
一、细胞膜表面受体的分类和简介	(84)
二、与 G 蛋白耦联受体媒介的信息传递	(85)

三、单一跨膜片段受体媒介的信息传递	(89)
四、信息传递研究中的新进展	(94)
第七章 细胞分化的基因调控	刘德培(99)
一、胚胎细胞分化	(99)
二、机体成型过程中的基因表达	(102)
三、成体组织中基因的特异性表达	(118)
四、细胞分化的基因调控机制	(118)
第八章 细胞程序化死亡的基因调控	郑德先、刘士廉(122)
一、程序化细胞死亡	(122)
二、程序化细胞死亡与个体发育	(123)
三、免疫系统中程序化细胞死亡	(125)
四、程序化细胞死亡与肿瘤	(128)
五、程序化细胞死亡相关基因及其表达调控	(130)
六、程序化细胞死亡与信息传递	(136)
七、展望	(138)
第九章 细胞周期的基因调控	吴宁华(140)
一、细胞周期及其中心控制体系	(140)
二、中心控制体系在酵母细胞增殖中的作用	(143)
三、哺乳动物细胞周期的基因调控	(147)
第十章 原癌基因与抑癌基因的表达与调控	方福德(159)
一、原癌基因和抑癌基因	(159)
二、原癌基因的表达	(163)
三、原癌基因表达的调控	(167)
四、抑癌基因表达的调控	(180)
第十一章 酵母基因表达的调控	敖世洲(182)
一、酵母的启动子元件	(182)
二、酵母的转录调控因子	(183)
三、酵母的 RNA 聚合酶 I	(186)
四、转录的控制机制	(187)
五、进化	(188)
第十二章 病毒基因表达的调控	琦祖和、袁建刚(191)
一、乙肝病毒基因表达的调控	(191)
二、SV40 病毒基因表达的调控	(201)
三、HIV 的基因表达调控	(206)
第十三章 类固醇激素受体超家族基因的调控	张永莲(222)
一、类固醇激素与真核基因表达调控的关系	(222)
二、作为反式作用因子的甾体激素受体	(223)
三、甾体激素受体基因表达的特性	(232)

四、甾体激素受体与疾病	(233)
五、结束语	(235)
第十四章 珠蛋白基因表达的调控	刘德培、张俊武 (236)
一、珠蛋白基因结构与功能	(236)
二、珠蛋白基因表达的调控	(238)
三、珠蛋白基因表达异常	(245)
四、珠蛋白基因表达调控研究与医学	(253)
第十五章 热休克基因表达的调控	刘巨洪、沈翊珩 (255)
一、真核细胞的热休克应答	(255)
二、热休克元件的调控作用	(257)
三、热休克因子的调控作用	(261)
四、热休克基因的诱导表达	(266)
五、热休克基因转录后调控	(268)
六、热休克基因表达的自身调控	(270)
七、结束语	(271)
第十六章 人类基因组研究	沈 岩 (273)
一、人类基因组研究	(273)
二、人类基因组研究的策略	(275)
三、人类基因组研究中常用的方法	(281)
第十七章 转基因技术在基因调控研究中的应用	陈兰英 (287)
一、在顺式调控元件研究中的应用	(287)
二、转基因动物作为研究外源基因调控的动物模型	(289)
三、用基因敲除技术研究基因表达	(290)
四、应用转基因技术研究发育中的基因表达与调控	(291)
五、基因多级调节系统	(292)
第十八章 辅助阅读资料	(295)
一、真核基因调控的主要研究方法简介	(295)
二、专业基础术语及其简释	(322)
索引	(333)
作者名单	(338)
编后记	(339)

第一章 真核细胞染色质结构与基因活化

生物体的基因组就存在于每个个体的各个有核细胞之中。比较原核生物与真核生物细胞时,会发现它们之间最主要的差别就是细胞核的存在。细胞核的大小一般占真核细胞的 10%,其中包含细胞中主要的 DNA,并以核被膜(nuclear envelope)结构与细胞质隔开。面向细胞质的核膜外层有核糖体存在并与内质网膜连接,同时也与细胞质中的微丝、微管、中间纤维等骨架元件相联,协助固定细胞核的位置。紧密附着于核质(nucleoplasm)表面的核膜内层是一薄层纤维样物质称为核纤层(nuclear lamina),它由基模蛋白组成,支撑着核膜结构并参与核孔的骨架结构。核膜上 10%~25%的面积为核孔所占据,有核孔复合体(nuclear pore complex)结构,其中央通道直径仅为 10nm,只能被动通过分子量在 50×10^3 以下的球形蛋白质,较大的活性物质则不能通过核孔自由地扩散。真核生物中储存遗传信息的基因被限定在细胞核内,核膜的存在会对 RNA 和蛋白质的出入有很大的选择性,使得基因的表达受到细胞核内、外诸多层次的调节,即真核基因表达的多级调控。

基因组 DNA 都是以很长的线性分子而存在。真核生物的另一个特征在于真核基因组 DNA 在细胞核中存在着以核小体(nucleosome)为基本单位的染色质(chromatin)结构,典型的间期核染色质可分为高度密集状态的异染色质(heterochromatin)和较为松散的常染色质(euchromatin)。在常染色质中大约 10%处于更为开放的疏松状态即为活性染色质。这些结构特征产生了真核细胞基因转录前在染色质水平上独特的调控机制。

一、 真核生物染色质的基本结构

1. 真核染色质的结构特点

(1)核小体是染色质的基本单位 染色质是真核细胞的特征性结构。1974 年在电镜下研究间期细胞核时发现“串珠状”(beads on a string)结构,这些串珠就是作为染色质基本单位的核小体。在核小体中,DNA 绕在 4 对组蛋白 H2A、H2B、H3、H4 所组成的“八聚体”核心外两周,并含有连接区的一个分子的 H1 组蛋白。一般核小体 DNA 的长度是 180~210bp。经微球菌核酸酶短期处理后,在电泳上可以看到由于降解后核小体聚合体数目不同而形成的“梯”带。进一步酶解产生含有 H1 组蛋白在内,由 166bpDNA 环绕两周的“染色质体(chromatosome)”;继续酶解则分离出的核心颗粒由 146bp 的 DNA 环绕核心 1.8 周。包绕核小体核心外的 DNA 成弧形向核心弯曲,每周 DNA 双螺旋向内弯折 50° (DNA bending)。

H3 和 H4 两对组蛋白可构成“亚核小体颗粒”,它保持着染色质基本单位的物理性质,在核小体的组装上起决定性作用。核小体的组成为三步,即两对 H3/H4 组蛋白组成的四聚体首先与组成核小体的 DNA 中段 120bp 相结合,随后,两个 H2A/H2B 的组蛋白二聚体才分别与其上下游的 DNA 结合,形成完整的核小体核心。两个核小体之间为 DNA 连接区,其长度与核小体相位、也与种属特异性有关,并受到进化压力的影响变异较大,可由 0~80bpDNA 构成。

(2)核小体相位(phased positioning) 核小体相位主要包括两方面的含义:

• 1 •

A. 核小体的“旋转”定位(rotational positioning) 指核小体核心与 DNA 双螺旋在空间结构中的相互关系: 主要涉及 DNA 双螺旋的小沟是面向抑或背向核心结构, 并与连接区 H1 组蛋白的结合和进一步组装有关。其特征可由 DNA 酶 I (或羟基基团) 的水解产物显示: 当小沟面向外时, DNA 酶 I 可以接触到每一周的 DNA 双螺旋结构上, 因此酶解产物为重复的 10bp 片段。核小体的热力学性质决定富含 AT 的 DNA 小沟面向组蛋白八聚体核心, 而小沟中富含 GC 时面向外方, 因为 DNA 双链弯折导致双螺旋小沟结构压缩的可能性在富含 AT 时, 会较 GC 为主的序列更容易实现。

B. 核小体的“转译”定位(translational positioning) 指核小体与特定基因 DNA 序列的结合位置和方式: 着重在转录活性相关的 DNA 元件(启动子、增强子)等序列与核小体相应位置的关系上, 例如一个核小体可以定位在某基因转录起始位点上游 -300/-144bp 上。核小体对特定 DNA 序列有一定的选择性, 其选择性较其它序列高约 1 000 倍。DNA 双链与序列特异性的蛋白质紧密结合着时, 会阻止核小体结构的形成, 以致数百个碱基对的 DNA 双链上都没有核小体结构, 并导致对 DNA 酶 I 的高敏感性。

(3) 染色质的高级结构 人类细胞核平均直径为 $5\mu\text{m}$ 。细胞中 3×10^9 碱基对的 DNA 存在于 22 对常染色体和 X、Y 染色体中。如按每条染色体含 1.5×10^8 个碱基对的 DNA 线性分子, 每个核小体 DNA 平均为 200bp 计算, 应有 7.5×10^5 个核小体存在, 其伸展长度约为 7mm, 远非细胞核所能容纳。X 射线衍射确定核小体直径约为 11nm, 在有关核蛋白的协同下组成以 6 个核小体为一周、直径 30nm 的紧密结构, 即光镜下可见的染色质纤维。此时人类基因组已压缩到 1mm 的长度。在两栖类未成熟的卵母细胞间期核中, 可在光镜下见到的一种特殊的灯刷状染色体(lampbrush chromosome)结构, 是在转录高度活跃的 RNA 与蛋白质形成的核蛋白复合体基础上构成的, 灯刷状染色体进一步形成大环状结构(looped domain), 通过结合在大环基部的蛋白质相互连接或基部直接连接在染色体的中轴上, 每个环的长度约为 20~100kb, 高度约为 300nm。含有这种环状结构的染色质 DNA 再进而盘绕成直径为 700nm 的染色单体。经过这一系列的组装, 人类 DNA 可通过约 2 000 个大环最终浓缩 30 万倍而组装在染色体中(图 1-1)。每条染色体的基本结构包括中心粒、两个端粒和复制起始点。

(4) 染色质结构的区间性

A. 基因座(位)控制区(locus control region, LCR) LCR 是一种顺式作用元件, 它使得相关基因或基因簇在基因组任何位置上都能表达; 换言之, LCR 具有稳定染色质疏松结构的功能。在人 β 珠蛋白基因有 4 个 LCR, 其中 3 个必须在稳定的整合状态表达时才具有增强子样活性, 提示这种活性与染色质结构有关。在 LCR 结构域中包含多种反式作用因子的 DNA 结合序列, 可能参与协同蛋白的结合作用使得启动子处于无组蛋白状态; 其中多种反式作用因子的结合序列则可保证 DNA 复制时与启动子结合的因子仍然保持在原位上。不同的 LCR 可由特异结合因子导向而结合在相关基因的启动子上, 因此也具有增强子样附带的功能。LCR 的这种性质必然决定了转基因在动物中, 带有 LCR 结构域基因都能高效表达。

B. 隔离子(insulator) 真核细胞染色体普遍具有“位置效应间隔(position-effect variegation)”现象。即染色体可以自然地分隔为互不影响的独立功能单位。能防止处于阻遏状态与活化状态下的染色体结构域之间的结构特点向两侧扩展的序列称为“隔离子”。

酵母细胞中, 整合到沉寂的交配型基因座位或端粒附近的基因会被抑制, 这种沉寂效应至

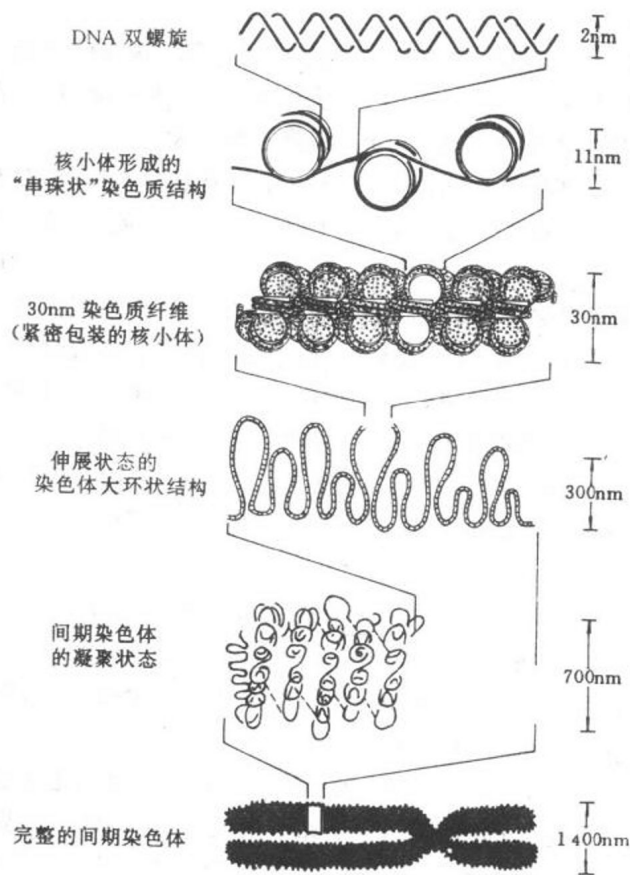


图 1-1 真核细胞中基因组 DNA 的组装模式

少可扩展 5~10kb 而不会达到 20~30kb。在果蝇中发现, 将一个正常的活性基因导入与异染色质邻近的一段染色体上, 该基因转录即被阻扼。首先发现果蝇热休克基因座位两端的核酸酶高敏位点的外界是一段 250~300bp 的 DNA, 这些“特化染色体结构 (specialized chromosome structure, scs)”位于正在转录的 hsp70 基因座位所在的疏松染色质和相邻的紧密结构之间。scs 元件的存在一方面可防止异染色质的转录抑制作用扩散; 若插入启动子与增强子之间又有阻断来自增强子的转录活化作用; 但在其限定的结构域内, scs 自身没有转录活化或抑制作用。scs 元件所形成的核蛋白复合体至今尚不清楚, 但一种由 gypsy 转座元件与毛翅阻扼蛋白 [su (HW)] 形成的核蛋白复合体具有 scs 类似的隔离子性质。在进行转基因动物实验时也发现“报告基因 (reporter gene)”随机整合到哺乳动物基因组不同位置上, 其转录水平会有很大的差异。而将 LCR 连接在报告基因上则可消除上述位置效应。最近证实鸡 β 珠蛋白基因染色体结构域的边界确有隔离子存在, 这些隔离子元件在果蝇中同样有效, 表明一切生物体都有类似的隔离

子。

2. 真核细胞核内的蛋白质

高等真核生物细胞核内的染色质总体上蛋白质与 DNA 的比例为 2:1, 细胞核蛋白质中包括组蛋白与非组蛋白(non-histone chromosomal proteins)两大类, 其中组蛋白的含量约占一半, 即组蛋白与 DNA 之比为 1:1。

(1) 组蛋白 组蛋白是一类较小而带有丰富正电荷(赖氨酸和精氨酸)的核蛋白, 与 DNA 有很高的亲和力, 它们在细胞中大量存在, 平均每个细胞有约 6×10^7 个分子。5 种组蛋白可分为核小体核心组蛋白(H2A、H2B、H3、H4)和 H1 组蛋白两大类。

核心组蛋白在进化上较保守, 分别由 102~135 个氨基酸残基组成, 对 DNA 盘绕成核小体结构形成以及核小体位相起关键作用; H3 和 H4 组蛋白富含精氨酸, 它们的 N 端含有多个可以受乙酰化和去乙酰化调节的赖氨酸残基。H3 是哺乳动物细胞中唯一含有半胱氨酸残基的组蛋白, 其 110 位半胱氨酸上的巯基可在疏松结构的染色质中与有机汞亲和柱结合。

H1 组蛋白较大, 约有 220 个氨基酸残基, 与核小体的稳定和高级结构的组装有关。H1 组蛋白具有进化上保守的中段球状核心区, 其伸展的 N 端和“长臂”状的 C 端较少保守性。H1 的球状核心区结合于 DNA 进入和脱离核小体核心的部位, 当核小体上有 H1 组蛋白结合时, 用微球菌核酸酶分离出的 DNA 片段为 166bp, 表明其“长臂”结构可以保护部分的 DNA 连接区, 甚至结合于相邻核小体的核心组蛋白上, 促使相邻核小体更加靠近。禽类的 H5 组蛋白与 H1 的功能相似。

(2) 非组蛋白 非组蛋白指细胞核内除组蛋白以外的各种蛋白质。它们的分子量一般比组蛋白大。已知的非组蛋白中主要有多种参与核酸代谢和修饰的酶类, 其它序列特异性的 DNA 结合蛋白的种类繁多, 而每种蛋白在细胞内至多只有 10 000 个分子, 其中许多是具有重要调控作用的反式作用因子(trans acting factors)。

(3) 高迁移率(high mobility group, HMG)蛋白质 高迁移率蛋白是细胞核内非组蛋白中的一组较丰富而不均一、富含电荷的蛋白质, 分子量一般 $\leq 30 \times 10^3$ 因其在聚丙烯酰胺电泳中迁移率很高而被命名为“高迁移率组 HMG”蛋白。在高等真核生物中, 这些不均一的蛋白质可分为 3 大类: HMG-1/2、HMG-I(Y)以及 HMG-14/17 等。

HMG-1/2 有 3 个结构域, 其中 A 和 B 结构域彼此很相似, 第三个含有酸性的羧端尾部能与组蛋白 H1 在体外结合。最近发现 RNA 聚合酶 I 体系的转录因子 UBF 的 DNA 结合单元是类似 HMG-1/2 的 A、B 结构域的 85 个氨基酸残基重复区, 因此将这种特征性结构命名为“HMG 结构域”; 在 DNA 中, 这一结构域的相应序列称为“HMG 盒”。具有一个或多个 HMG 结构域的蛋白质称为“HMG 结构域蛋白”。

HMG 结构域蛋白的组织特异性: HMG-1/2 家族的 HMG 结构域蛋白可按照 HMG 结构域拷贝数、序列识别特异性和进化关系分为两个亚族, 分别具备下列特点之一:

A. 较大量、广泛存在于各种细胞中, 如 HMG-1、HMG-2、UBF 等。一般有多个 HMG 结构域, 没有明显的 DNA 识别特异性。这类蛋白质各个 HMG 结构域的功能不同: 如 UBF 含有 4~6 个 HMG 结构域, 其中只有第一个结构域能以二聚体形式结合在 rRNA 的启动子上, 第三和第四个结构域能促进 UBF 与其它蛋白质和 DNA 的结合, 而第二个结构域对特异启动子起导向作用。

B. 一般只存在于特定的细胞中,能识别特异的 DNA 序列,包括仅有一个 HMG 结构域的 DNA 结合蛋白,这类蛋白质中有淋巴细胞增强子结合因子 LEF-1 和 TCF-1、性别决定因子 SRY 等。HMG 结构域蛋白的最大特点是结合在 DNA 双螺旋的小沟中,并仅以 40 倍的优势选择富含嘧啶的核苷酸元件。

二、染色质水平上的基因活化调节

转录前染色质结构发生的一系列重要变化是基因转录的前提,因此染色质的活化又称为转录前的基因调整(modulation)或基因的“开关”。

1. 活性染色质(active chromatin)呈“开放”的疏松型结构

非活化的染色质 DNA 不能被转录,其特点是以 30nm 的间期核染色质纤维为基础,并已浓缩 40~50 倍包装在压缩状态。30nm 的纤维“开启”为可转录的疏松染色质结构时仅压缩 6 倍,成为 11nm 直径的单核小体伸展状态,即电镜下的“串珠状”结构。在细胞中染色质的活性与非活性状态是可以逆转的,并且存在着二者间的过渡状态。为便于讨论,本章将以 $\leq 10\text{nm}$ 直径的单核小体结构视为“疏松”或“活化”的染色质结构;而以压缩到 $\geq 30\text{nm}$ 的染色质纤维作为“紧密”或“非活化”的状态。

(1) 活化(或有活化潜能)的基因处于染色质的疏松结构中 果蝇幼虫细胞可以在没有细胞分裂的情况下 DNA 多次复制,成为可含多达 1 000 个以上拷贝染色体 DNA 的巨型细胞,在唾液腺等多种分泌细胞中,复制后的同源染色体平行排列,形成一条巨大的多线染色体(polytene chromosome)。在光镜下多线染色体中 85% 的区域呈现出因结构紧密而染色较深的“带(band)”,以及占 15% 的较浅的“间带(interband)”区,每个带约有 3 000 至 300 000bp。用 ^3H 尿嘧啶核苷标记时发现最活跃的染色体转录区结构疏松,形成“蓬松区(chromosome puff)”结构。电镜下发现多线染色体蓬松区 DNA 处于 30nm 以下的疏松状态染色质中,同时也存在着与灯刷状染色体相似的大环结构。

(2) 染色质疏松或紧密结构的可逆转换 在非活化的染色质紧密结构中,核小体核心 H3 组蛋白 110 位的巯基处于封闭状态。转录时核小体转变为疏松结构导致 H3 组蛋白的巯基活性可被检测到。如在 NIH3T3 细胞中研究原癌基因 c-fos 和 c-myc 的诱导表达时,发现在基因诱导转录的数十分钟或数小时期间内,正在转录的基因绝大部分都分布在疏松结构(即活性)染色质中,与细胞中相应 mRNA 前体的转录水平相平行。一旦转录终止,这些基因又重新分布到非活化的紧密结构中,表明转录时核小体疏松或紧密结构的转换是可逆的。

2. “开放”(或疏松)染色质结构的形成

核小体结构的松散甚至缺失可能由于起阻扼作用的组蛋白与 DNA 的亲合力降低;或特异调控蛋白、HMG 蛋白等与 DNA 以高亲合力结合,将核小体组蛋白部分置换改变了核小体的空间构象。它又可因为组蛋白的修饰或突变,以及 DNA 构象改变所导致二者间结合力降低。现将以上诸因素分别讨论如下:

(1) DNA 结构与活性染色质 前面介绍了核小体相位对染色质活化状态的影响,除此而外,富含 GC 达到 15 次重复的 Z-DNA 存在时可导致核心组蛋白八聚体与 DNA 的亲合力降低;同时 DNA 拓扑异构酶可以持续地调整 DNA 的超螺旋状态,如拓扑异构酶 2 抑制剂

novobiotin 阻断果蝇热休克基因的转录,拓扑异构酶 1 与多线染色体中具有转录活性的基因座位分布相等等都显示出 DNA 结构与基因活化的关系。

(2) 组蛋白的修饰

A. H1 组蛋白磷酸化 H1 组蛋白对核小体起装配作用,确定核小体方向性;并将基本的“串珠”进一步组装到 30nm 的纤维中。活化染色质对 H1 的亲合力极低,H1 的磷酸化也直接影响染色质的活性。在有丝分裂时的染色体中,H1 组蛋白高度磷酸化,可达到每个分子中有 5~6 个丝氨酸磷酸化,分裂后则下降到峰值的 20%;到下个周期的染色体凝聚开始前的间期再次达到高峰。有丝分裂中 H1 磷酸化导致其对 DNA 的亲合力下降,可能与具有转录活性的染色质疏松状态相似。

B. 核心组蛋白的修饰

①核心组蛋白的结构域 组蛋白 H2A/H2B 二聚体与两对 H3/H4 组蛋白的四聚体之间的联系最为脆弱,因此在核小体松解时首当其冲的是两对 H2A/H2B 二聚体从两端脱离核心结构。

组蛋白 H3 和 H4 的 N 端不直接参与维持 DNA 在核小体核心上的结合,但这些富含正电荷的区域与 DNA 的磷酸二酯键骨架有关。在酵母和人之间,H3、H4 都十分保守。H4 组蛋白的 N 端尾部的强碱性区及其相邻的基本非极性区对酵母交配基因座位是阻扼作用;但对 Gal-1 启动子和其它基因是活化作用。然而 H4 组蛋白氨基端激活的基因,却受到 H3 组蛋白的氨基端尾部的阻扼作用,表明 H3 和 H4 组蛋白尾部可能与调控因子起特异的调节作用。

H3 组蛋白羧基端球形结构域中第 105 和 118 位氨基酸突变时,基因可在不需要激活复合物的状态下表达。这是因为突变点正好处于组蛋白 H3 结合核小体二重对称轴上的 DNA,以及与另一个组蛋白 H3 相连接的位置,影响了蛋白与蛋白、蛋白与 DNA 之间的接触,造成核小体结构的不稳定性;并有可能导致组蛋白 H2A/H2B 以及组蛋白 H1 自核小体上解离。因此认为(H3/H4)₂与 DNA 的结合点是转录活化复合体起作用的关键位置。

H4 的氨基端的氨基酸置换或删除会改变特定基因上的核小体相位。H4 的氨基端尚有特殊的氨基酸残基可与基因特异性的激活因子相互作用。

酵母基因的诱导性启动子受核小体结构影响。当核心组蛋白中的 H2B 或 H4 合成受阻时,某些诱导表达的基因,如 PHO5 等可被激活,出现部分基因表达,但通常低于正常的诱导表达活性,表明有关基因的表达需要两种因素的共同作用:首先是核小体阻扼结构的破坏(染色质活化);其次是特异性的基因转录因子与基础转录运转器件的相互作用(基因转录)。

②核心组蛋白的乙酰化 核小体核心组蛋白最主要的修饰方式是其赖氨酸残基上的 ϵ -氨基乙酰化。它促使相邻核小体的聚合受阻,并影响泛素与组蛋白 H2A 分子的结合,导致蛋白质的选择性降解。

H3、H4 是蛋白修饰酶的主要底物,其氨基端结构域的乙酰化和磷酸化通常是信息传递途径影响核小体稳定性的靶位。H3、H4 组蛋白乙酰化可能有类似促旋酶(gyrase)的活性,它使得核小体间的 DNA 因产生较多的负超螺旋而容易从核小体上脱离,对核酸酶的敏感性增高,并有利于转录调控因子的结合。一般活性染色质是高度乙酰化的,乙酰化是可逆的,同时乙酰化的位点不是随机的。H4 氨基端第 5 或第 8 位 Lys 乙酰化在果蝇多线染色体中是活化的;而第 12 位乙酰化则不在活性状态出现。在鸡红细胞中, α D 珠蛋白探针在高乙酰化的活性染色质中

富集 15~30 倍,而不表达的卵清蛋白则不富集。

核小体中乙酰化和去乙酰化酶活跃,每个乙酰基分子在组蛋白上平均只保留 10 分钟左右。乙酰化部位常在组蛋白的碱性氨基端,即通过转乙酰基酶自乙酰辅酶 A 将乙酰基转移到 Lys 残基上,并中和一个正电荷。从而减弱核小体中碱性蛋白与 DNA 链之间的静电吸引力,并降低相邻核小体间的聚集。用去乙酰化酶抑制剂丁酸钠处理 HeLa 等多种细胞均可导致 H3 和 H4 分子的多乙酰基化,这一过程随丁酸盐的去除而迅速逆转(在处理 20 小时后 H4 的乙酰基化分子可达 80%,移除 15 分钟后 H4 完全恢复到非乙酰化状态),其它如磷酸化、甲基化、以及可能与 DNA 修复或复制时的多聚(ADP)糖基化等也是重要的修饰形式。

(3)其它 DNA 结合蛋白的作用

A. 上游调控因子 蛋白质因子、组蛋白和 DNA 三维复合体的形成可改变核小体结构,甚至特定区域内的无核小体存在。然而仅有极少数的蛋白质可以直接结合核小体,最早发现的染色质结合蛋白是糖皮质激素受体(GR)。该诱导因子结合于小鼠乳腺肿瘤病毒(MMTV)启动子区的核小体上,使核小体表面的 DNA 的双螺旋结构发生特殊相位的旋转,主沟从面对组蛋白八聚体而转向外,同时保持 GR 应答元件(GRE)处在较易接近的核小体位置上。核因子 1(NF1)也是 MMTV 启动子的结合蛋白,但只在 GR 结合以后才能结合。此外管家基因(housekeeping gene)所通常含有的一些结合位点也有持续存在的蛋白质因子(如 SP1)与之结合。有证据表明 SP1 与基因上的多个结合位点相结合可以阻止核小体 H1 组蛋白的阻扼作用。

核小体抑制 RNA 聚合酶 II 启动子上转录起始复合体的形成,而 TF II D(或 TATA 盒结合蛋白 TBP)的结合可以防止核小体阻扼启动子的功能。由此推论 TBP 的有效结合可能是在核小体模板上形成转录起始复合体的速度限制环节。

在果蝇热休克 hsp70 基因的研究中也发现类似的特点:组成性表达的 GAGA 转录因子的结合导致核小体“动态解聚”,机理包括核小体核心八聚体解离为组蛋白 H2A/H2B 二聚体和 (H3/H4)₂四聚体;由于组蛋白-DNA 间相互作用减弱整个八聚体核心可以自启动子序列滑动开,或在启动子上移除了八聚体;GAGA 因子与其它蛋白质间的相互作用可使其间的 DNA 承受到某种扭转力以致核小体稳定性降低等。GAGA 因子有赖于 ATP 的功能导致核小体解聚。

B. HMG 结构域蛋白 最近发现 HMG 结构域蛋白可识别某些异型的 DNA 结构。LEF-1 等特异蛋白可使 DNA 链产生 90~130°的弯折。一般情况下,短于 150bp 的 DNA 双链很难自发形成环状结构,然而在非特异的 HMG-1 蛋白存在下,可以改变 DNA 的双螺旋结构以致形成 66bp 的 DNA 环。因此认为 HMG 结构域蛋白的功能与 DNA 弯折和 DNA-蛋白质复合体高级结构的形成有关。HMG 结构域蛋白的作用可以是唯一的因素,也可作为其它反式作用因子参与的支撑结构导致 DNA 高级结构的改变、反式作用因子的结合和基因活化等有关。

具有转录活性的核小体常缺乏 H1,却有分子量在 30×10^3 以下、没有的组织特异性的非组蛋白 HMG14、17 存在。在全部染色质中大约每 10 个核小体中可有一个核小体存在着这两种 HMG 蛋白。HMG14 和 HMG17 在每个核小体上有两个高亲和力的结合位点,同时也是目前已知唯一直接结合在核小体核心颗粒上的核蛋白。最近发现在爪蟾卵提取物体系中 HMG17 参与到 DNA 复制后新组装的染色质中会促进 RNA 聚合酶 III 的基因转录;而 HMG14 则可直接参与 RNA 聚合酶 II 对染色质中基因的转录等,都说明这些 HMG 蛋白与核小体的结合是产