

表观遗传学前沿

● 主编 蔡 禄

*Biaoguan
Yichuanxue
Qianyan*

清华大学出版社



表观遗传学前沿

● 主编 蔡 禄

清华大学出版社
北京

内 容 简 介

本书共分 11 章,包括染色质结构与功能、核小体定位、染色质重塑、组蛋白修饰、DNA 甲基化、RNA 可变剪接的表观遗传学机制、非编码 RNA 研究进展、假基因研究进展、表观遗传学研究实验技术简介、表观遗传学相关数据库简介及表观遗传学的功能。

在主要阐述了表观遗传学的基本知识的基础上,结合本领域国内外最新科研成果,本书系统总结介绍了当前表观遗传学领域的前沿进展。

本书可供从事生命科学研究的人员参考,也可作为高等院校遗传学等相关专业研究生的教材。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

表观遗传学前沿/蔡禄主编. —北京:清华大学出版社,2012.12

ISBN 978-7-302-30817-1

I. ①表… II. ①蔡… III. ①发育遗传学—研究 IV. ①Q344

中国版本图书馆 CIP 数据核字(2012)第 287014 号

责任编辑:罗 健 王 华

封面设计:戴国印

责任校对:王淑云

责任印制:沈 露

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦 A 座 邮 编:100084

社总机:010-62770175 邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质 量 反 馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

印 装 者:北京国马印刷厂

经 销:全国新华书店

开 本:185mm×260mm 印 张:18.5 插 页:4 字 数:493千字

版 次:2012年12月第1版

印 次:2012年12月第1次印刷

印 数:1~2000

定 价:49.80元

产品编号:048177-01

编者名单

主 编

蔡 禄

副 主 编

赵秀娟 刘国庆 崔向军

编 者

(按姓名拼音排序)

蔡 禄 崔向军 刘国庆 马利兵

邢永强 赵宏宇 赵秀娟

人类基因组计划的完成大大促进了人们对遗传信息组织、传递及表达规律的认识，同时也使我们意识到细胞内遗传信息表达机制的异常复杂性。为了彻底破译生命这本天书，各种组学(Omics)计划包括蛋白质组学、转录组学、代谢组学、RNA组学相继实施，特别是近两年才出现的表观遗传组学使多年来分子生物学领域公认的中心法则遇到了前所未有的挑战。表观遗传学研究在DNA序列不发生改变的情况下，由于DNA甲基化、染色质结构变化等因素的改变，使基因功能发生可遗传的变化并最终导致表型变异的遗传学机制。基因组携带有两类遗传信息：一类提供生命必须的蛋白质的模板，称为遗传编码信息；另一类提供基因选择性表达(何时、何地、何种方式)的指令，称为表观遗传信息。基因表达调控机制的研究一直是遗传学研究的中心问题，表观遗传信息将大大丰富遗传学研究内容。只有将遗传编码信息和表观遗传信息的组织、传递和表达机制研究清楚，才有可能真正解读细胞内的生命过程。表观遗传信息对于细胞组织特异性分化、发育、疾病发生发挥重要作用。

近几年国内外出现了一些优秀的表观遗传学专著或教材，但总体上讲还太少。由于表观遗传学相关内容发展迅速，一些最新的研究成果往往在书籍中难以找到系统介绍。作者在研究工作中深深感到研究人员，特别是年轻的研究生朋友，急需系统介绍表观遗传学前沿知识的书籍。事实上，这也正是本书作者们虽然深感自己学识浅薄，仍然坚持撰写本书的主要原因。参与编写本书的人员均工作在表观遗传学实验和理论研究前沿，他们通过认真学习国内外相关优秀书籍，在大量阅读最新文献的基础上，经系统总结完成本书。全书共分11章，包括染色质结构与功能、核小体定位、染色质重塑、组蛋白修饰、DNA甲基化、RNA可变剪接的表观遗传学机制、非编码RNA研究进展、假基因研究进展、表观遗传学研究实验技术简介、表观遗传学相关数据库简介及表观遗传学的功能。赵秀娟副教授编写第1、2、5章和第9章第1~4节，第3章由邢永强博士编写，崔向军副教授编写第4、10章第1~4节及第11章第3~5节，刘国庆副教授编写了第7、8章，赵宏宇博士编写第9章第5~7节和第10章第5节，马利兵教授编写第11章第1~2节，蔡禄教授编写第6章，并负责全书的审核。

我们热切希望这本书能为本领域研究人员、教师和学生提供有价值的新知识，但由于这一领域的快速发展，加之尚有许多问题学术界仍在探讨未有定论，特别是编者学术水平有限，书中肯定有诸多不足和错误之处，敬请读者批评指正。

蔡 禄

2012年10月

| | | | |
|--|----|--|--|
| 第 1 章 染色质的结构与功能 | 1 | | |
| 1.1 细胞核的结构与功能 | 1 | | |
| 1.1.1 细胞核的结构 | 1 | | |
| 1.1.2 细胞核的功能 | 3 | | |
| 1.2 染色质结构 | 5 | | |
| 1.2.1 染色质纤维 | 5 | | |
| 1.2.2 染色质分为常染色质和 异染色质 | 6 | | |
| 1.3 核小体 | 8 | | |
| 1.3.1 发现历史 | 8 | | |
| 1.3.2 核小体的结构 | 9 | | |
| 1.3.3 相位 | 9 | | |
| 1.4 染色体 | 10 | | |
| 1.4.1 染色体的化学组成 | 10 | | |
| 1.4.2 染色体的 DNA | 10 | | |
| 1.4.3 染色体的特征性结构 | 10 | | |
| 1.5 核仁结构与功能 | 11 | | |
| 1.5.1 核仁的化学组成 | 11 | | |
| 1.5.2 核仁的结构 | 12 | | |
| 1.5.3 核仁的形成 | 12 | | |
| 1.5.4 核仁的功能 | 12 | | |
| 参考文献 | 12 | | |
| 第 2 章 核小体定位 | 15 | | |
| 2.1 基因组上核小体定位的实验 图谱 | 16 | | |
| 2.1.1 基因组尺度核小体定位 作图的基本策略 | 16 | | |
| 2.1.2 基因组尺度核小体定位 图谱 | 17 | | |
| 2.2 基因组上核小体分布的一般 模式 | 17 | | |
| 2.2.1 DNA 转录相关位点邻近 核小体分布 | 17 | | |
| 2.2.2 DNA 复制起始位点邻近区 核小体分布 | 18 | | |
| 2.2.3 核小体定位的“统计 定位”模型 | 19 | | |
| 2.3 核小体定位的理论研究 | 20 | | |
| 2.3.1 从 DNA 序列到核小体 定位 | 20 | | |
| 2.3.2 基于 DNA 结构的物理性质 预测核小体定位 | 23 | | |
| 2.3.3 核小体定位的反式因子 | 24 | | |
| 2.4 邻近核小体和染色质高级结构 影响核小体定位 | 26 | | |
| 2.4.1 核小体串珠 | 26 | | |
| 2.4.2 高级染色质结构的影响 | 26 | | |
| 2.5 核小体定位与基因表达调控 | 26 | | |
| 2.6 总结与展望 | 27 | | |
| 参考文献 | 28 | | |
| 第 3 章 染色质重塑 | 34 | | |
| 3.1 染色质重塑概述 | 34 | | |
| 3.2 SWI/SNF 复合物 | 36 | | |
| 3.2.1 SWI/SNF 复合物的组成 | 36 | | |
| 3.2.2 SWI/SNF 复合物的结构域 和功能 | 37 | | |
| 3.2.3 SWI/SNF 复合物与癌症 | 38 | | |
| 3.3 ISWI 复合物 | 39 | | |
| 3.3.1 ISWI 复合物的组成 | 39 | | |
| 3.3.2 ISWI 复合物的功能 | 41 | | |
| 3.4 CHD 复合物 | 41 | | |
| 3.4.1 CHD 复合物的组成 | 41 | | |
| 3.4.2 Chd 蛋白的位置和组织表达 模式 | 43 | | |

| | | | |
|--|-----|--|-----|
| 3.4.3 Chd 蛋白在染色质组装和 重塑过程的作用 | 43 | 稳定性 | 103 |
| 3.4.4 CHD 家族成员的多亚基 复合物 | 44 | 5.4.1 DNA 甲基化与 DNA 复制 起始 | 104 |
| 3.4.5 Chd 蛋白和转录延伸 | 44 | 5.4.2 DNA 甲基化与错配 修复 | 104 |
| 3.4.6 发育和分化阶段的 Chd 蛋白 | 45 | 5.4.3 DNA 甲基化与转座子 失活 | 104 |
| 3.4.7 Chd 蛋白与人类疾病 | 45 | 5.5 DNA 甲基化与基因表达调控 .. | 104 |
| 3.4.8 总结与展望 | 46 | 5.5.1 DNA 甲基化直接影响一些 转录因子的结合活性 .. | 104 |
| 3.5 INO80 复合物 | 46 | 5.5.2 DNA 甲基化结合蛋白与转 录抑制 | 105 |
| 3.5.1 INO80 复合物的组成 | 47 | 5.6 DNA 甲基化与其他表观遗传修饰 的关系 | 105 |
| 3.5.2 INO80 复合物的功能 | 48 | 5.6.1 DNA 甲基化与核小体 定位 | 105 |
| 3.5.3 SWR1 复合物的组成 | 50 | 5.6.2 DNA 甲基化与组蛋白 修饰 | 106 |
| 3.5.4 SWR1 复合物的功能 | 52 | 5.6.3 DNA 甲基化与非编码 RNA | 108 |
| 3.5.5 总结与展望 | 53 | 参考文献 | 108 |
| 3.6 染色质重塑模式和机制 | 53 | 第 6 章 RNA 可变剪接的表观遗传学 机制 | 111 |
| 3.6.1 染色质重塑模式 | 53 | 6.1 引言 | 111 |
| 3.6.2 染色质重塑机制 | 54 | 6.2 可变剪接的基本机制 | 111 |
| 3.7 总结与展望 | 56 | 6.2.1 可变剪接的基本概念 .. | 111 |
| 参考文献 | 57 | 6.2.2 可变剪接的基本类型 .. | 112 |
| 第 4 章 组蛋白修饰 | 62 | 6.2.3 可变剪接的基本机制 .. | 112 |
| 4.1 组蛋白修饰概述 | 62 | 6.3 转录与剪接同时进行的机制 .. | 113 |
| 4.1.1 组蛋白的分类和性质 | 62 | 6.4 可变剪接受 RNA 聚合酶 II 延伸 速率的控制 | 113 |
| 4.1.2 组蛋白修饰的种类 | 63 | 6.5 可变剪接的表观遗传学机制 .. | 114 |
| 4.2 组蛋白修饰酶及其相关复合物 .. | 70 | 6.5.1 可变剪接与核小体定位 .. | 114 |
| 4.2.1 乙酰化酶及其复合物 | 72 | 6.5.2 组蛋白修饰诱导可变 剪接 | 115 |
| 4.2.2 组蛋白甲基化酶 | 83 | 6.5.3 可变剪接与 DNA 甲基化 | 115 |
| 4.3 组蛋白变体及其修饰 | 85 | 6.5.4 可变剪接与组蛋白 变体 | 116 |
| 4.3.1 组蛋白变体概述 | 85 | 6.5.5 RNA 与可变剪接 | 116 |
| 4.3.2 组蛋白变体修饰 | 89 | 6.5.6 可变剪接的新模型 | 116 |
| 4.4 组蛋白密码 | 89 | | |
| 参考文献 | 94 | | |
| 第 5 章 DNA 甲基化 | 98 | | |
| 5.1 DNA 甲基化概况 | 98 | | |
| 5.2 真核生物 DNA 甲基化分布 模式 | 99 | | |
| 5.3 真核生物 DNA 甲基化修饰 系统 | 100 | | |
| 5.3.1 DNA 甲基转移酶 | 100 | | |
| 5.3.2 甲基化结合蛋白 | 101 | | |
| 5.4 DNA 甲基化与遗传物质的 | | | |

| | | | |
|--|-----|--|-----|
| 6.5.7 可变剪接相关生物大分子 (蛋白质、DNA 和 RNA) 相互作用网络 | 117 | 8.5 假基因的分布 | 168 |
| 6.6 总结与展望 | 118 | 8.6 假基因的功能 | 169 |
| 参考文献 | 118 | 8.6.1 一氧化氮合酶 (NOS) 假基因的功能 | 170 |
| 第 7 章 非编码 RNA 研究进展 | 123 | 8.6.2 <i>Makorin1-p1</i> 假基因的 功能及作用机制 | 170 |
| 7.1 RNA 干扰 | 124 | 8.6.3 假基因干预细胞的基因沉默 机制 | 173 |
| 7.2 siRNA 的研究进展 | 125 | 8.6.4 假基因产生 siRNAs | 173 |
| 7.2.1 siRNA 的简介 | 125 | 8.6.5 假基因保护 snoRNA 的 作用 | 175 |
| 7.2.2 siRNA 的作用机制 | 126 | 8.6.6 假基因的转录与 保守性 | 176 |
| 7.2.3 siRNA 的特点 | 129 | 8.6.7 假基因产生基因 多样性 | 177 |
| 7.2.4 RNAi 的应用 | 131 | 8.6.8 假基因的命运 | 177 |
| 7.3 miRNA 的研究进展 | 133 | 8.6.9 假基因的弊端 | 178 |
| 7.3.1 miRNA 的简介 | 133 | 8.7 总结与展望 | 179 |
| 7.3.2 miRNA 的产生机制 | 133 | 参考文献 | 179 |
| 7.3.3 miRNA 的作用机制 | 135 | 第 9 章 表观遗传学研究实验技术 简介 | 184 |
| 7.3.4 miRNA 的功能 | 140 | 9.1 染色质免疫共沉淀技术 | 184 |
| 7.3.5 总结与展望 | 142 | 9.1.1 ChIP 技术的原理 | 184 |
| 7.4 piRNA 的研究进展 | 143 | 9.1.2 ChIP on chip | 184 |
| 7.4.1 piRNA 的发现 | 143 | 9.1.3 ChIP-Seq | 185 |
| 7.4.2 piRNA 的结构特征 | 144 | 9.2 全基因组定位技术 | 185 |
| 7.4.3 piRNA 的分布 | 144 | 9.2.1 基因表达系列分析的原理和 实验路线 | 185 |
| 7.4.4 piRNA 的产生机制 | 145 | 9.2.2 全基因组定位技术原理与 步骤 | 186 |
| 7.4.5 piRNA 的功能 | 146 | 9.2.3 应用 | 187 |
| 7.4.6 piRNA 的总结与展望 | 148 | 9.3 体外组装核小体技术 | 188 |
| 7.4.7 小 RNA 的总结与展望 | 148 | 9.3.1 盐透析法体外组装 核小体 | 188 |
| 7.5 长链非编码 RNA 的研究进展 | 149 | 9.3.2 依赖于 ATP 的体外组装 核小体方法 | 189 |
| 7.5.1 lncRNA 的产生机制 | 149 | 9.3.3 体外组装单个核小体的检测 方法 | 190 |
| 7.5.2 lncRNA 的功能 | 150 | 9.3.4 长片段 DNA 序列体外组装 染色质的检测方法 | 193 |
| 7.5.3 lncRNA 与疾病 | 154 | 9.4 核小体相位分析 | 197 |
| 7.5.4 总结与展望 | 154 | | |
| 参考文献 | 154 | | |
| 第 8 章 假基因研究进展 | 162 | | |
| 8.1 假基因 | 162 | | |
| 8.1.1 假基因的发现 | 162 | | |
| 8.1.2 假基因的结构特征 | 163 | | |
| 8.2 探讨假基因的生物学意义 | 163 | | |
| 8.2.1 假基因是 Junk DNA? | 163 | | |
| 8.2.2 假基因的生物学意义 | 164 | | |
| 8.3 假基因的识别 | 165 | | |
| 8.4 假基因的进化 | 167 | | |

| | | | | | |
|---------------------------------|----------------------------------|-----|-------------------------|---|-----|
| 9.4.1 | 基本原理 | 197 | 10.5.1 | siRNA 数据库介绍 | 229 |
| 9.4.2 | 主要步骤 | 198 | 10.5.2 | miRNA 数据库介绍 | 230 |
| 9.4.3 | 注意事项 | 199 | 10.5.3 | piRNA 数据库介绍 | 233 |
| 9.4.4 | 应用 | 199 | 10.5.4 | lncRNA 相关数据库 介绍 | 233 |
| 9.5 | DNA 甲基化分析技术 | 199 | 10.5.5 | 非编码 RNA 组数据库 NONCODE | 235 |
| 9.5.1 | 基因组整体水平甲基化 分析 | 201 | 参考文献 | | 235 |
| 9.5.2 | 特异性位点的 DNA 甲基化的检测 | 202 | 第 11 章 表观遗传学的功能 | | 239 |
| 9.5.3 | 甲基化新位点的寻找 | 210 | 11.1 干细胞的分化 | | 239 |
| 9.6 | 染色质 DNA 酶 I 高敏感位点的 检测 | 210 | 11.1.1 | 胚胎干细胞的分化 | 239 |
| 9.6.1 | 基本原理 | 211 | 11.1.2 | 成体干细胞的分化 | 243 |
| 9.6.2 | 基本步骤 | 212 | 11.2 X 染色体失活 | | 254 |
| 9.6.3 | 注意事项 | 213 | 11.2.1 | X 染色体随机失活的 起始 | 255 |
| 9.6.4 | 应用 | 213 | 11.2.2 | <i>Xist</i> RNA 介导的 X 染色体 沉默以及失活 X 染色体的 异染色质化 | 256 |
| 9.7 | 体内 DNA 足迹法 | 213 | 11.3 基因组印记 | | 264 |
| 9.7.1 | 基本原理 | 213 | 11.3.1 | 哺乳动物的印记基因 | 264 |
| 9.7.2 | 操作步骤 | 215 | 11.3.2 | 基因组印记的周期及 机制 | 265 |
| 9.7.3 | 注意事项 | 215 | 11.3.3 | 基因组印记的进化 | 267 |
| 9.7.4 | 应用 | 216 | 11.4 衰老的表观遗传学 | | 268 |
| 参考文献 | | 216 | 11.4.1 | 衰老过程中的表观 遗传学 | 268 |
| 第 10 章 表观遗传学相关数据库 简介 | | 219 | 11.4.2 | 衰老相关疾病的表观 遗传学 | 270 |
| 10.1 核小体定位相关数据库简介 | | 219 | 11.5 记忆过程中的表观遗传学 | | 271 |
| 10.1.1 | 传统实验技术条件下的 核小体定位数据 | 219 | 11.5.1 | 组蛋白的乙酰化修饰与依赖 于脑海马的记忆行为 | 271 |
| 10.1.2 | 高通量实验技术条件下的 核小体数据 | 220 | 11.5.2 | 组蛋白的乙酰化修饰与不依 赖脑海马的记忆行为 | 273 |
| 10.2 可变剪接数据库 | | 221 | 11.5.3 | 其他形式的组蛋白修饰与 记忆行为 | 274 |
| 10.2.1 | 可变剪接数据库概述 | 221 | 11.5.4 | DNA 甲基化与记忆的 形成及储存 | 275 |
| 10.2.2 | ASD 数据库简介 | 222 | 参考文献 | | 276 |
| 10.2.3 | ASAP 数据库简介 | 224 | | | |
| 10.3 组蛋白修饰数据库 | | 225 | | | |
| 10.3.1 | 人组蛋白修饰数据库 | 225 | | | |
| 10.3.2 | 酵母组蛋白修饰 数据库 | 227 | | | |
| 10.4 DNA 甲基化的相关数据库 | | 227 | | | |
| 10.5 非编码 RNA 相关数据库简介 | | 229 | | | |

1.1 细胞核的结构与功能

真核细胞的重要特征是有细胞核，电镜观察真核细胞时，细胞核是最大的可见部分，核内有一些相当难以辨别的结构（图 1-1）。细胞核包含了真核生物几乎所有的遗传物质（在线粒体和叶绿体中还含有少量的 DNA）。细胞核与细胞质之间有双层的核膜相隔，核膜中间有一个内腔与内质网相连，核孔复合体（NPC）横跨核被膜，是细胞核与细胞质之间输送生物大分子的通道，能将蛋白质及 RNA 运入、运出细胞核。内质网中的空间与两层核膜间的空间是连续的，内外层核膜的脂双层与每一个核孔是连续的。两个中间丝网络提供了核膜的机械支持；细胞核中的中间丝形成一个特殊的结构，称为核纤层。细胞核的亚单位核仁是核糖体 RNA 合成的场所，没有膜结构。

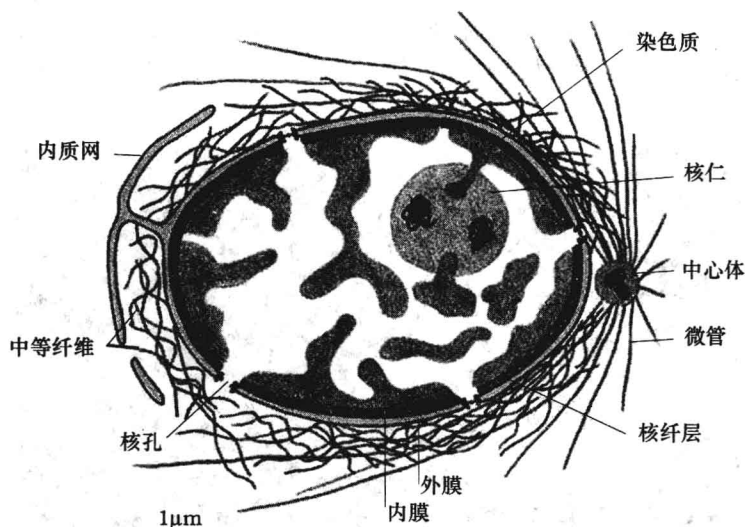


图 1-1 间期细胞核的形态结构（引自 Alberts et al, 1998）

细胞核的大小范围从直径 $1\mu\text{m}$ 到大于 $10\mu\text{m}$ 不等，细胞核的大小与其所包含的 DNA 的多少有关，酵母细胞的细胞核只有 $1\mu\text{m}$ 左右，而非洲爪蟾的卵母细胞细胞核直径约为 $400\mu\text{m}$ 。大多数细胞有一个细胞核，但有的细胞也有多个细胞核，少数细胞没有细胞核。球形和椭圆形是细胞核最常见的形状，不同物种细胞核所占细胞总体积的比例是不一样的，细胞核的形态和表型是区分不同细胞类型的两个要素。

1.1.1 细胞核的结构

1. 核膜与核纤层 细胞核以核被膜为界限，核被膜由外核膜和内核膜组成（图 1-2）

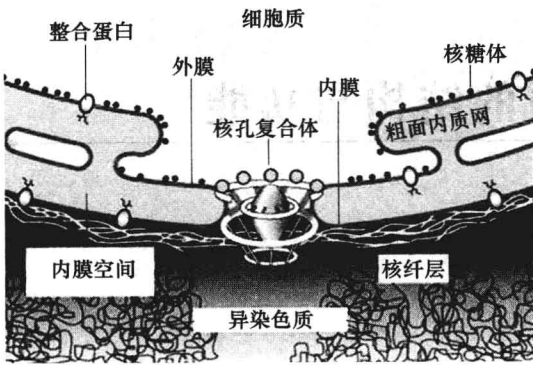


图 1-2 核膜 (引自 Andersson et al, 1999)。显示了双膜层、核孔复合物、核纤层以及外膜层同粗面内质网的连续性

(Andersson et al, 1999; Fahrenkrog et al, 2001; Rout et al, 2001)。核膜由磷脂双分子层和特定的蛋白质组成。外核膜与内质网连通, 表面附有核糖体, 负责蛋白质的合成。外核膜与内核膜之间的膜间腔与内质网的膜间腔连通。

核纤层位于核内膜之下, 核纤层是多细胞动物细胞核的一个普遍特征, 它是纤维状的网络结构 (Broers et al, 2004; Goldman et al, 2002; Gruenbaum et al, 2005; Hutchison et al, 2001; Taddei et al, 2004)。核纤层蛋白被称为中间纤维蛋白, 其成丝的大小 (直径 10~20nm) 在肌动蛋白微丝 (7nm) 和微管 (直径 25nm) 之间, 核孔复合体锚定在核纤层。核纤层相关蛋白 (LAP)

中的一些蛋白质介导核纤层和内核膜之间的相互作用。核纤层支持细胞核结构, 同时也与染色质相互作用。

2. 核孔复合体 NPC 核孔复合体位于内核膜和外核膜的融合部位, 是一个双向结构 (Fahrenkrog et al, 2001; Rout et al, 2001; Schwartz et al, 2005), 由 30 多种不同的多肽组成, 它们是核孔蛋白, 而且许多核孔蛋白含有短序列重复, 例如 Gly-Leu-Phe-Gly、X-X-Phe-Gly, 这些短序列重复在转运过程中与转运因子相互作用。核孔复合体的结构类似于桶, 镶嵌在双层膜上, 形成一种环状的结构 (图 1-3)。

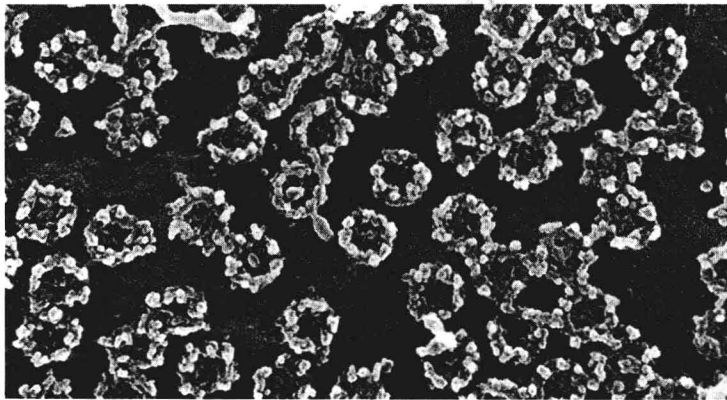


图 1-3 核孔复合体的高分辨率扫描电镜图 (引自 Lodish et al, 1999)。两栖类卵细胞分离的 NPS 的细胞质面, 细胞质颗粒覆盖在 NPS 的胞质环

3. 染色体占据 细胞核内的每一条染色体都有其特定的位置, 占领一定的区域 (Gilber et al, 2005; Parada et al, 2002; Sanger et al, 1997), 染色体末端的端粒被锚定在核被膜上, 这对防止染色体纠缠有作用。细胞核包含染色体区和染色体间区, 有染色体定位的区域即染色体区, 附近没有染色体的区域为染色体间区, 染色体间区存在着带有 poly (A) 尾巴的 RNA。

4. 核仁 细胞核内最明显的亚单位是核仁。多数细胞核只有一个核仁, 但有的细胞核内也有多核仁, 核仁的大小与细胞中核糖体数量有关。核糖体 RNA 合成和加工, 核糖体亚单位装配都是在核仁中进行的。核仁中含有编码 rRNA 的 DNA, 当 rRNA 转录被人为中断时, 核仁消失, 当

转录被允许时，核仁重新出现。

1.1.2 细胞核的功能

水及相对分子质量小于 100 的不带电荷的分子，可以通过磷脂双分子层自由扩散，其他的分子和生物大分子必须通过核孔复合体进出细胞核。

1. 通过核孔复合体选择性运输蛋白质

(1) 蛋白质的入核：核孔是一个选择性通道，它只允许携带有正确氨基酸信号的蛋白质进入。一般情况下，被转运蛋白质包含定位于细胞核的信息，没有定位信号的蛋白质也可以通过与有定位信号的蛋白结合而被运输，而且蛋白质进入细胞核之后，定位信号依然存在不会被去除。这种定位信号被称为细胞核定位信号（NLS），是一段短链氨基酸序列，普遍包含基本氨基酸——赖氨酸和精氨酸。入核蛋白的运输是受体介导的过程，实验表明蛋白质入核是可饱和的，入核受体常称为核转运蛋白家族（Goldfarb et al, 2004）。

核蛋白入核可以分为两个步骤（Harel et al, 2004）：一是核孔处依赖 NLS 的结合（停泊 docking）；二是迁移进入核质。促进结合活性的 NLS 受体，这种受体伴随着运输物质通过 NPC，在细胞核内释放运输物质，之后再回到细胞质继续下一轮的转运，通过 PNC 增强迁移活性的物质是 Ran GTPase。目前已鉴定了两种类型的 NLS 受体：可以直接结合到核孔上的 NLS 受体，这种受体是由 α 输入蛋白和 β 输入蛋白组成的；核孔上需要其他的特殊受体介导结合的 NLS 受体，以确保有效的输入，有的运输蛋白在入核过程中并不止利用一个核转运受体，另外，并非所有的蛋白质都通过核转运蛋白输入到细胞核中，例如信号分子 β -catenin，直接与核孔蛋白相互作用进入细胞核。

(2) 蛋白质的出核：从细胞核中输出蛋白质也是受体介导的，出核信号被称为核输出信号（NES）。在细胞质中合成的某些蛋白质，被输入细胞核，然后再输出细胞核，这种蛋白质大多是转录因子，通过它们的定位来控制其活性。一些蛋白质在细胞核和细胞质之间连续循环的过程被称为细胞核-细胞质穿梭（核-质穿梭），这些蛋白不仅有 NLS 使之定位到细胞核中，而且有 NES。

核转运蛋白和核孔蛋白的相互作用对于通过核孔的迁移是至关重要的，核运输的方向性部分决定于核转运蛋白与某些核孔蛋白的特异作用。

Ran GTPase 控制细胞核运输的方向（Kalab et al, 2002；Dasso, 2002）。Ran 是一个小的 GTPase，它普遍存在于所有的真核细胞中，在细胞核和细胞质中都有分布。Ran-GTPase 通过 Ran 促进 GTPase 水解，而 Ran-GEF 促进了 Ran 上的 GDP 交换为 GTP。Ran-GAP 存在于细胞质中，而 Ran-GEF 定位在细胞核中。Ran 通过结合核孔蛋白控制细胞核转运，并影响它们结合物质的能力。

核转运蛋白与大多数核孔蛋白通过苯基丙氨酸-氨基乙酸（FG）重复相互作用，这些 FG 重复排列在核孔中央通道内；研究表明一些核转运蛋白与特定的核孔蛋白有高度亲和性；Ran 不仅介导运输物质与核转运蛋白的相互作用，而且介导核转运蛋白和核孔蛋白的相互作用；虽然核转运蛋白可以双向穿过核孔，但大多数只是单方向携带运输物。

(3) 核转运模型：一个简单的细胞核转运基质模型显示通过易化扩散穿过通道的运动（Rout et al, 2001；Becksei et al, 2005）。这个模型中，转运受体可以与核孔蛋白有短暂的低亲和性相互作用，没有方向性结合反应，运动的方向由核孔任一侧 Ran 依赖的终止步骤决定。这个模型的一个更复杂的解释是，转运的方向是由于随着运输物/受体复合物穿过 NPC 通道，转运受体和核孔蛋白之间亲和力梯度的增加。

另一个模型是观察排列于核孔通道内的许多包含 FG 的疏水核孔蛋白（Ben-Efraim et al,

2001)。这些核孔蛋白的 FG 结构域被认为是没有组织的,在核孔通道中创造疏水的环境。在选择阶段模型中,核孔蛋白富含 FG 区之间的相互吸引力构成了核孔蛋白质的中央屏障,这个屏障使大多数蛋白质不能通过。这个模型进一步提出在 NPC 通道这个特殊的环境中,转运蛋白具有选择性的亲和力。这将核孔排除许多蛋白质通过核孔的能力与核转运蛋白复合体的易化扩散相联系。这个模型的证据依然处于初级阶段,但是入核的快速动力学支持这个模型,而且观察到核转运蛋白与核孔蛋白的 FG 重复相互作用。然而,这个模型没有解释大的复合物,如 mRNA 复合物如何转运。

荧光显微镜技术开始用于探测单个分子,并且这些方法正在被应用到细胞核转运的研究中(Ribbeck et al, 2001; Yang et al, 2004)。研究表明,蛋白复合物通过 NPC 通道的运动非常快(平均 10ms),大多数核孔蛋白运输物复合体的运动在 NPC 通道是随机的,并且速率限制步骤可能脱离中央通道进入细胞核。这些研究还说明 1 个 NPC 可以同时运输至少 10 个底物分子(和它们的受体),并且一个 NPC 每秒可以转运约 1000 个分子。

在真核生物中核被膜的存在在一定水平上使基因表达和细胞周期得以调控,这在原核生物中是不可能的。蛋白质输入和输出都是可调控的,对于核运输调控的重要性有许多例子。细胞通过对转录因子在细胞核和细胞质之间运动的调控来控制转录,从而实现对应激和生长控制信号的应答。例如,通过调控周期性和非周期性(永恒)转录因子的核转运来控制生理节律(Zhu et al, 2003)。此外,在细胞核和细胞质之间蛋白激酶及其调节子的运动对于推动细胞周期的进行和外界刺激的应答是非常重要的。

蛋白质的人核和出核可以在多个水平上受到调控:第一,运输物质与转运受体之间的相互作用的能力可通过对运输物质的直接修饰而调控,如磷酸化修饰等;第二,运输物质可以锚定在细胞的一个区域,因此锚定解除前它不能与转运子一起移动;第三, NPC 本身转运蛋白质的能力是被调控的对象。

2. 多种 RNA 的输出 在真核细胞中,几乎所有的 RNA 都是在细胞核中产生, mRNA/tRNA 和包含 rRNA 的核糖体亚基必须从细胞核中输出。通过 NPC 转运到细胞质中,在细胞质中行使翻译功能。大多数 RNA 不包含输出信号,但是为了输出细胞核必须结合在包含 NES 的蛋白质上。事实上,细胞中的 RNA 总是与蛋白质形成复合物,以保护 RNA 分子与其他细胞组分的相互作用。作用于蛋白质转运的 NPC 也同样用于 RNA 输出。RNA 输出是受体介导和能量依赖的;每种 RNA 的转运需要不同的可溶性转录因子。每一类型的 RNA 的输出至少需要一类特殊类型的限速因子,所有输出的 RNA 都利用相同的 NPC,所以 NPC 的数量是不限制的。这些研究没有说明每类 RNA 转运中涉及多少因子,输出单一种类的 RNA 有多少因子是共享的和有多少因子是唯一需要的。

(1) tRNA 在细胞核中转录,它们在细胞核中加工,然后被运到细胞质,在细胞质中它们被氨基化并参与翻译。只有完全加工好的 tRNA 才能被输出。细胞核 tRNA 的输出是由专一的转运受体介导的,并且需要 Ran, 输出蛋白 t 是 tRNA 的转运受体,属于核转运蛋白家族中的一员(Shaheen et al, 2005; Takano et al, 2005)。输出蛋白 t 是唯一能直接结合 tRNA 的核转运蛋白,它优先与完全加工的 tRNA 结合。

(2) mRNA 的转录和加工发生在细胞核内部,需要 hnRNP 从转录位点转移到细胞核周边的 NPC 上(Aguilera et al, 2005; Saguez et al, 2005)。转录位点也是大多数 mRNA 的加工位点,加工完成以后, mRNA 从染色体区域释放到染色质间区。mRNA 通过染色质间区扩散到细胞核周边。mRNA 以 RNA-蛋白质复合物的形式输出细胞核。转录过程中与 mRNA 确定加工的位点,同时包装 mRNA 进行细胞核输出。在细胞核中与 mRNA 结合的蛋白质在核输出后,与 mRNA 解离并返回细胞核。一小部分蛋白质在核输入之前即迅速地与 mRNA 分离。mRNA 输出信号可能存

在于与 mRNA 结合的蛋白质中, mRNA 的输出是可调控的, 但调控机制尚不清楚。

(3) 小核糖体蛋白颗粒 (snRNP) 是在 mRNA 前体剪接和其他核 mRNA 加工中起重要作用的 RNA-蛋白复合物。snRNP 中存在的 U snRNA 在细胞核内产生。然而, 在哺乳动物细胞中功能性 snRNP 的形成需要 U snRNA 的核输出、在细胞质中修饰、与胞质中的蛋白质结合, 以及 U snRNA 复合体的核输入, 进入核内的 U snRNA 复合物最终装配成为 snRNP。

核糖体包括两个亚基, 它是由大约 80 个蛋白质和 4 个核糖体 RNA (rRNA) 组成的大型复合物。亚基分别在核仁中组装并被输出到细胞质进行最后的组装 (Johnson et al, 2002)。核糖体亚基是可以通过 NPC 运输的最大的复合物成员, 核糖体亚基是核孔通道可以通过的上限。由于这些亚基的尺寸很大, 当一个核糖体亚基通过 NPC 通道时其他的大分子就不能被运输。

核糖体亚基的组装和输出是非常复杂的过程, 并且高度依赖细胞核运输。核糖体蛋白在细胞质中合成, 之后被输入到细胞核, 进入核仁。在核仁中, 核糖体蛋白与 rRNA 前体相互作用, 在核仁中被转录并组装成因子。随着正确的组装, 每一个亚基通过 NP 被输出。在细胞质中, 亚基与转录起始因子、mRNA 组装形成成熟的翻译核糖体。

迄今为止已经被研究的所有的核糖体蛋白和细胞核输入都使用核转运蛋白家族成员和 Ran GTPase。但是, 至少还有一些核糖体蛋白利用不止一个专一的输入因子。例如在酵母中, 核转运家族的两个不同成员介导相同核糖体蛋白的输入 (Cullen et al, 2003; Libri et al, 2003)。因为正确的核糖体产生对于细胞的存活是至关重要的, 所以要确保有充分的组分用于组装。

核糖体亚基的输出是可饱和的, 表明它是受体介导的。核糖体亚基不与其他的 RNA 蛋白复合物相互竞争输出, 表明核糖体亚基使用不同转运受体。甚至细菌核糖体也可被输出, 表明真核生物的输出组件可以识别这些核糖体 (Oeffinger et al, 2004)。

(4) microRNA 在细胞核内通过转录产生, 局部加工形成有发夹结构的前体, 通过输出蛋白 v 输出细胞核, 最终在胞质中加工成熟 (Ambros et al, 2004)。microRNA 在基因表达的调控方面具有重要作用, 可调控许多生理途径, 包括发育、分化、细胞程序性死亡 (细胞凋亡)、器官形成和细胞增殖 (Kim et al, 2005)。microRNA 与胞质中的靶 mRNA 结合, 从而阻断它们的翻译甚至促进这些生命过程的逆转 (Zamore et al, 2005)。

1.2 染色质结构

细胞分裂时, 核内染色质凝缩而成的线状结构, 对碱性染料染色很深, 故名染色质 (chromatin)。染色质是由 DNA 与蛋白质组成的复合物, 也是构成染色体的结构, 存在于真核生物的细胞核内。原核生物也有染色质, 存在于拟核内 (Postow et al, 2004)。

1.2.1 染色质纤维

人细胞平均含有 60 亿碱基对的 DNA, 分布在 46 条染色体上, 每个碱基对的长度是 0.34nm, 60 亿碱基对的 DNA 加起来有 2m 长, 每个碱基对结合 6 分子水, 直径只有 10 μ m 的细胞核如何能容下 2m 长的水溶性 DNA 染色质呢? 研究发现遗传物质以核小体作为基本结构逐步进行包装压缩, 经 30nm 染色质纤维、超螺线管、最后压缩包装成染色体, 总共经过四级包装。由 DNA 和组蛋白组成的染色质纤维细丝是许多核小体连成的念珠状结构。

1. 从 DNA 到核小体 核小体的装配是染色体装配的第一步, 一个核小体的直径是 10nm, 由

167 个碱基对的 DNA 组成缠绕在组蛋白八聚体周围，每个碱基对长度为 0.34nm，一个核小体伸展开来的长度是 70nm，由此推算，DNA 包装成核小体，大约压缩了 7 倍。

2. 核小体到螺线管 (solenoid) 核小体通过自身形成螺旋的方式形成致密的、外径为 30nm 的管状结构，称为螺线管，又称 30nm 染色质纤维。从完整的细胞中常常可分离到 10nm 的核小体纤维和 30nm 纤维，只要改变提取液的盐浓度可将二者分开。

10nm 纤维是由核小体串联成的染色质细丝，主要在低离子强度及无 H1 组蛋白情况下产生，当离子强度较高级 H1 存在时，以 30nm 纤维为主。

30nm 纤维染色质丝，是由 10nm 纤维压缩形成的粗丝。有人认为 30nm 纤维染色质丝是由细丝螺旋盘绕而成的螺线管，这种螺线管是分裂间期染色质和分裂中期染色体的基本组分。螺线管进一步压缩形成超螺旋；有人认为粗丝是 Z 型锯齿结构模型，人们认为 30nm 纤维从中心轴开始形成圆环，最后环状的链至少经过了超过一种的包装而形成有丝分裂染色体。

3. 从螺线管到超螺线管 30nm 的染色质纤维进一步螺旋化，形成一系列的螺旋域 (coiled domain) 或环 (loop)，这些环附着在支架 (scaffold) 蛋白上。螺旋域的直径是 300nm。然后，螺旋环进一步形成超螺旋环 (supercoiled loop)，或超螺旋域，此时的直径为 700nm，每个环估计含有 50~100kbDNA，推测染色质环仍然是基因协同结合在一起，其中包括 II 型拓扑异构酶，推测该酶调节 DNA 的超螺旋程度。DNA 从螺线管的超螺线管估计又压缩了 40 倍。

4. 从超螺线管到染色体 超螺线管进一步螺旋化，形成直径为 1~2 μ m，长度为 2~10 μ m 的中期染色体，从超螺线管到染色体大约压缩了 5 倍。由此推算，DNA 经核小体到染色体，总共压缩了 8400 倍。将 DNA 经由核小体包装成染色体的基本过程见图 1-4。

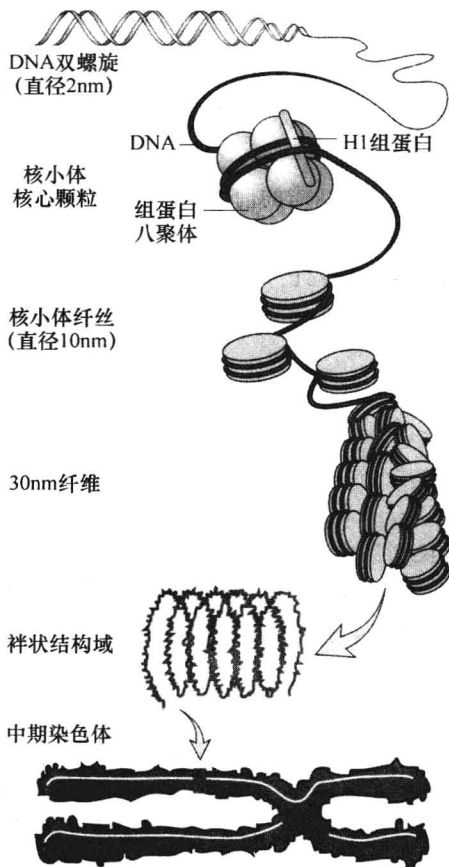


图 1-4 DNA 经由核小体包装成染色体的基本过程

1.2.2 染色质分为常染色质和异染色质

在细胞中，常染色质 DNA 只占全部 DNA 的一小部分，其他大多数区域是异染色质。

1. 常染色质 (enchromatin) 是基因密度较高的染色质，多在细胞周期的 S 期进行复制，且通常具有转录活性，能够生产蛋白质。常染色质染色较浅且均匀且松散，常染色质在真核生物与原核生物的细胞中皆存在。与原核生物不同，真核生物基因组 DNA 的遗传信息首先在细胞核内由基因转录为 mRNA 前体，经剪接加工后，mRNA 在胞质的核糖体中翻译成为多肽，经过折叠、修饰等成为具有生物功能的蛋白质，也就是说，细胞内特定蛋白质的选择性表达首先取决于基因组中特定蛋白编码基因的活化。

2. 异染色质 (heterochromatin) 是间期细胞核中，染色质丝折叠程度高，处于凝缩状态，碱性染料着色深的那一部分染色质。它具有早凝集晚复制的特点，通常无法转录成为 mRNA。异染色质以浓集状态存在。在细

胞周期的S期中，异染色质比常染色质更晚进行复制，且只在真核生物中存在，着丝粒及端粒皆属于异染色质，雌性体内去活化的X染色体（也就是巴尔氏体）也是异染色质。

(1) 组成型异染色质和兼性异染色质。异染色质可分为组成型异染色质和兼性异染色质。

组成型异染色质较稳定，通常位于着丝粒、端粒及核仁组织区，DNA常不转录，是永久性的异染色质结构，不含或含很少的结构基因。

兼性异染色质，含正常编码的DNA，可自由伸展，使附近基因失活，表现为斑点位置效应(PEV)，可和常染色质相互转化，其基因呈抑制状态。人们认为不同组织的细胞，或不同发育期的细胞，正是由于染色质的不同部位形成不同程度的高级有序的异染色质，导致基因开启-关闭，从而细胞表现出不同的生理功能、代谢类型和结构状态。

异染色质和常染色质是可以相互转化的，有资料表明，组蛋白尾富含的碱性氨基酸（如Lys和Arg）使其带的正电荷对异染色质的形成和解聚是至关重要的，而各种修饰可以改变组蛋白尾的静电荷，从而使染色质处于不同水平的凝缩状态。如H3 Lys9甲基化促进异染色质的形成，它是异染色质蛋白(HP1、SWi6)的特异结合位点，而H3 Lys4甲基化抑制异染色质的形成。异染色质的形成也直接受基因的控制，有两类拮抗基因E(var)和Su(var)，前者抑制异染色质的形成，而促进常染色质的形成；后者促进异染色质形成，抑制常染色质形成。

异染色质最突出的功能是对基因的调控，包括两个相反的过程，一个是通过一种与“异染色质化”有关的过程，使许多碱基的染色质结构关闭；另一方面是通过稳定更多的已开放的染色质结构来避免这种关闭的结构状态的存在。此外，异染色质的形成对印记(imprinting)、剂量补偿、重组和染色质浓缩是必要的。

(2) 异染色质的形成。在染色质上并无特征性的因素决定哪里是异染色质形成的位点。异染色质蛋白复合物优先向重复DNA部分聚集，如常见的高等真核生物的臂间异染色质和基因间区域。斑点位置效应表明异染色质能沿染色质扩散，从而引起附近DNA序列遗传的不活泼。组蛋白尾的修饰是异染色质形成的重要原因。

3. 染色质结构的顺式调控和反式调控 尽管核小体的基本结构相同，但基因一旦处在30nm直径以上的高级结构甚至异染色质中，该基因就不可能转录。除此以外，真核细胞的基因及其转录活化需要的顺式调控元件(cis acting element)在染色质中的状态对转录效率也至关重要。如果一个基因和它的重要转录元件(如启动子、增强子等)被特定的染色质结构间隔开，也不能进行转录。

体内核小体结构的动态调整过程表现为核小体有序的周期性结构因特异转录因子的加入或去除而改变。染色质结构调整导致基因起始转录过程中的4个阶段：①“串珠状”核小体甚至更高级结构中的基因处于非活化的基态；②当蛋白质因子通过其DNA结合结构域结合到染色质上，局部的染色质转变为去阻遏状态，这一过程不需要ATP和转录因子的反式活化结构域的参与；③蛋白质因子结合染色质后，依赖于ATP的存在介导了染色质结构的调整，使染色质成为活化状态；④结合于染色质的蛋白反式活化结构域参与募集启动子结合蛋白和转录起始前复合体，基因开始转录。由此可见，染色质结合蛋白中的DNA结合结构域是染色质结构动态调整的关键，而其中的反式活化结构域主导基因的转录。

4. X染色体和Y染色体 X染色体即巴氏小体或X小体，为紧贴细胞核膜内面的团块状结构，直径约 $1\mu\text{m}$ ，染色程度较其他染色质深。其形态不一，常呈三角、半圆、平凸或球形。利用放射自显影技术的研究发现，女性的两条X染色体中有一条DNA复制延迟，称迟复制X。迟复制的X染色体在间期时表现为X染色质。当细胞内有一条以上X染色体时，在间期时除一条X染色体外，其余的X染色体均表现为X染色质，因此间期细胞核中的X染色质数目等于X染色体数

减去1。当X染色体结构异常时，X染色质的形态也会有相应的改变。如X等臂染色体时，出现大的X染色质，双着丝粒X染色体时，出现双叶或大的X染色质。

Y染色质又称Y小体或荧光小体。Y染色体用荧光染料染色后，呈亮暗不一的荧光带，在Y染色体长臂的远侧段呈明亮的荧光区。在间期时Y染色体长臂远侧段的强荧光特性仍然存在，经荧光染色后，呈强荧光亮点，直径为 $0.25\sim 0.3\mu\text{m}$ ，位于细胞核内的任何部位。

1.3 核小体

真核生物基因组DNA储存在细胞核内的染色体中。核小体(nucleosome)是构成真核生物染色质的基本结构单位，各核小体串联而成染色质纤维。核小体DNA长度约为165个碱基对，其中缠绕在组蛋白八聚体周围的核心DNA(core DNA)约1.65圈，约合147个碱基对；而相邻的核小体之间的自由区域(linker DNA)为20~50个碱基的长度，也就是说基因组的75%~90%被核小体所占据。组蛋白八聚体由进化上高度保守的H2A、H2B、H3和H4各两个拷贝组成。核小体核心颗粒在组蛋白H1的作用下形成稳定结构，核小体的进一步组装会干扰DNA复制、基因表达和细胞周期进展的过程。

1.3.1 发现历史

早在1956年，为双螺旋模型提供X衍射证据的Wilkins和另一位科学家Vittorio Luzzati对染色质进行了X衍射研究，发现染色质中具有间隔为10nm的重复性结构。蛋白质和DNA本身的结构从来不会表现出这种重复性。推测可能是组蛋白和DNA的结合方式迫使DNA折叠或缠绕成具有10nm周期的重复结构。

1971年，Clark和Felsenfeld首先用葡萄球菌核酸酶(staphylococcal nuclease)来作用于染色质，发现有一些区域对核酸酶敏感，有一些则不敏感，不敏感的区域比较均一，这暗示染色体中存在着某些亚单位。

Hewish和Burgoyun(1973)用内源核酸酶消化细胞核，再从核中分离出DNA，结果发现一系列DNA片段，它们相当于长约200bp的一种基本单位的多聚体。表明组蛋白结合在DNA，以一种有规律的方式分布，以致产生对核酸酶敏感的只是某些限定区域。

Noll(1974)用外源核酸酶处理染色质，然后进行电泳，证实了以上结果，他测得前三个片段的长度分别为205bp、405bp、605bp长，每个片段相差200bp，即染色质可能以200bp为一个单位。这正好和以下电镜观察的结果相印证。

与此同时Olins夫妇(1974)和Pierre Chambon等(1975)在电镜下观察到大鼠胸腺和鸡肝染色质的“绳珠”状结构，小球的直径为10nm；Olins并把这种小球称为n小体(n-body即nu body)。

X衍射图表明组蛋白的多聚体都是紧密相联，并无可容纳像DNA分子那样大小的孔洞，所以不可能由DNA之“绳”穿过组蛋白之“珠”，而只可能是DNA缠绕在“珠”的表面。电泳的结果和电镜观察到“绳珠”结构之间是什么样的关系呢？

1974年，哈佛大学Kornberg和Thomas用实验回答了这一问题。他们先用小球菌核酸酶稍稍消化一下染色质，切断一部分200核苷酸对单位之间的DNA，使其中含有单体、二聚体、三聚体和四聚体等。通过凝胶电泳证明每一组分子大小及纯度，并用电镜来观察。结果发现单体均为10nm小体，二聚体则是两个相联的小体，三聚体和四聚体分别由3个小体和4个小体组成，表明200核