

- 9 Yamamoto F, Clausen H, White T *et al.* Nature, 1990; **345**: 229
- 10 Yamamoto H, Hakomori S. J Biol Chem, 1990; **265**: 19257
- 11 Yamamoto K, Konami Y, Osawa T *et al.* J Biochem, 1992; **111**: 87
- 12 Galili U, Thall A, Macher B A. Trends Glycosci Glycotechnol, 1990; **2**: 304
- 13 Galili U, Swanson K. Proc Natl Acad Sci USA, 1991; **88**: 7401
- 14 Furukawa K. Trends Glycosci Glycotechnol, 1990; **2**: 243
- 15 Yamashita K, Tachibana Y, Ohkura T *et al.* J Biol Chem, 1985; **260**: 3963

密码结构域及其功能表现

劳为德

(中国科学院生物物理研究所, 北京 100101)

摘要 能组织成超级结构的各种序列组块,因其具有特定一级序列、或者具有某种卷曲的螺旋构象和某种非 β 螺旋结构,可以作为有特异功能的结构域、被特异的结合蛋白质识别和结合,因而可称为密码结构域。密码结构域作为特异的分子相互作用和过程的遗传指令,参与细胞周期的各种事件乃至发育和分化过程中基因有区别的表达。

关键词 密码结构域, 染色质超级结构, 串列重复序列, 着丝粒, 端粒, 核基质, 同源异型基因, 发育与分化

从事分子生物学研究的人大都在习惯上把编码特性与 DNA 序列联系起来,看看该序列是否有某种读出框架可转译出某种氨基酸序列。DNA 这种编码能力约在 30 年前(1966 年)已得到认识。基本的遗传密码广泛存在于迄今研究过的古今生物种类而且高度保守。但现在越来越明显,DNA 具有比此要大得多的编码潜力来完成其它重要的任务。除了编码特异的 RNA(如 rRNA, snRNA 和 tRNA)分子之外,DNA 特异结构型式和序列型式本身也呈现出调节密码和表达密码的遗传活性。比如,染色质密码规定着核小体中 8 体的组蛋白复合物的相位调整;转译框架密码决定蛋白质合成时三联子的正确辨认;剪接过程中套环密码则组织 RNA 新生链的单链与蛋白质的相互作用;剪接密码使相继的 5' 和 3' 拼接点的正确剪接。这些 DNA 密码,大多数毫无例外地是基于 DNA 的一级序列本身;但看来也包涵于相应的超级结构的特异性特征,因为各种 DNA 密码

都是在间期和细胞分裂时在细胞核内起重要作用的特定分子相互作用(或过程)的遗传指令。

1 串列重复 DNA 结构的编码能力

大多数真核生物基因组都存在具有不同总长度的串列重复序列结构,其特征是两个以上的序列单位以头对尾的形式重复排列,其总长度在两个与上千个重复之间。这种串列的组块式结构或是位于特异的染色体位点,如着丝粒和端粒的染色体区域;或是弥散于基因组各处。各种组块结构几乎被组织成某种重复的座位特异的超级结构^[1]。

1.1 串列重复 DNA 序列之座位特异的超级结构

虽然串列重复的 DNA 序列结构都可在每个染色体中找到,但它们的总体序列组成(超级结构)都是座位特异的。这种特异性或是基于各种较小的重复序列组块在特异的座位上的扩增,或是基于基因组其它部分的非同源序列

的嵌入. 染色体特异的超级结构突出的例子是果蝇染色体的着丝粒异染色质中高度重复的卫星 DNA 序列^[2], 和人染色体着丝粒异染色质高度重复的 α 类 DNA 序列^[3].

特异性还基于某种共同的序列单位 (共识序列, consensus sequence) 座位特异的趋异, 以及基于在特异的染色体位点上这些序列的拷贝数高度的变化不定. 具有这种结构的大多数序列家族, 如 $(T)_n$, $(G)_n$, $(TG)_n$, $(TC)_n$ 在任何真核基因组中都是高度保守的^[4]. 突出的例子是酵母、果蝇、小鼠和人都具有的 $(GATA)_n$ 序列; 和小卫星序列的超级家族, 这些序列见于鸟类、小鼠和各种家畜.

除了这种保守性外, 同样的序列家族在基因组中也是遗传变异的主要来源. 许多科学家把这些序列视为刍狗 (junk) DNA 是能理解的 (“刍狗”一词源自《老子》: “天地不仁, 以万物为刍狗; 圣人不仁, 以百姓为刍狗.” 取郭象注: “废弃之物, 于时无用”). 但最近的一些研究结果说明不可以世俗的眼光笼而统之视之为刍狗^[5].

1.2 串列重复 DNA 序列结构中的刍狗 DNA

许多研究结果表明, 串列重复 DNA 结构中的不同序列区可与特异的蛋白质相结合. 为这种具有与蛋白质相结合的功能编码的密码形式, 可以是特异的一级序列框 (sequence box), 也可以是某种卷曲的螺旋 DNA 构象, 或是某种非 β 螺旋的 DNA 结构. 这些有功能的序列区一般包埋于更大的重复序列结构中, 由此很清楚, 特定的串列重复序列家族中不是每种序列拷贝在功能上都同等重要的; 重复的刍狗 DNA 和有功能的重复 DNA 序列有着共同的起源.

人们推断, 不同染色体区域中特异染色质折叠结构的出现和稳定下来是由某种串列重复的超级结构的座位特异性引起的^[1]. 这种假说看来可从最近有关着丝粒、端粒和核基质附着区的串列重复序列的座位特异性超级结构的实验得到支持.

2 染色体的着丝粒区域

在人染色体中, 着丝粒区域的染色质折叠结构的特异性可以通过热处理后作 Giemsa 特异性染色 (C 分带) 反应出来. 着丝粒区域有着各种高度重复的序列家族, 如 α 类序列家族, 和卫星 DNA 家族 I, II 和 III. 它们都是具有某种结构形态的座位特异性序列. 最近发现, 人染色体中有一种特异的着丝粒蛋白质 CENP-B, 直接与 α 类序列家族的特异性序列 (CENP-B 框: 5' CTTCGTTGGAAACGGGA 3') 相结合^[6]. CENP-B 框序列只是 α 类序列单位中的一部分. 免疫电镜研究结果表明 CENP-B 蛋白质分布于整个着丝粒异染色质中, 这与 α 类 DNA 丛的分布颇相似. CENP-B 与 DNA 之间形成的主要复合物中含有两种 DNA 分子, 这说明 CENP-B 框之间的相互作用在着丝粒区染色质的高级形态的组织中起着重要的作用.

最近从人染色体着丝粒中分离到了主要存在于卫星 DNA 家族 II 和 III 的高度保守的序列单位 $(GGAAT)_n$ ^[7]. 这与酵母着丝粒的关键序列相似. $(GNAAT)_n$ 的解链曲线表明, 这些寡核苷酸有一种折回或是多链结构, 这是由于单纯的嘌呤链在形成双链时有相同的热稳定性. 这种特异的折叠结构是由 G·A 碱基配对引起的. 稳定的 G·A 配对是 tRNA 三级结构形成时一般的结构形式. DNA 迁移率漂移实验表明, $(GNAAT)_n$ 的序列单位与 HeLa 细胞粗提物中的特异性核蛋白质具有很高的结合亲和力. 酵母的 CBF₃ 蛋白在与裂殖酵母中同源的着丝粒序列单位 CDE III 也有类似的亲和力. 这一单位有 110kb 长, 也是由不同的重复序列组块组成. 由此组织成一种染色体特异的超级结构.

电子显微镜观察也表明家鼠 (*Mus domesticus*) 大卫星家族以及邻近小卫星序列单位也有某种染色体特异的超级结构.

按照假定的着丝粒区中串列重复序列的编码能力^[1], 现在有可能详细说明这种结构性 DNA 密码的表现. 也许, 在着丝粒区域中, 某

些序列结构域可通过其特异的蛋白质结合能力,或特异的有力卷曲能力,或采取非 β 螺旋DNA构象的能力而被选择.它们可以定义为“密码结构域”.这种密码结构域即如作为单元的CENP-B框和具有其特异折叠结构的高度保守的重复序列组块(GNAAT)_n等.在鼠着丝粒卫星序列中有一种保守的内在卷曲的序列元件也是类似的密码结构域.通过周围的重复序列结构域(“间隔结构域”)有选择的放大和重排,使得总的重复序列超级结构并因此高级的折叠结构也变成座位特异的.

由于不同序列单位持续不断的放大、嵌入和缺失,从而使大多数串列重复序列成就了座位特异性^[1].不难想象,重复的密码结构域只有处于串列重复序列中才能成就其恰当的空间组织形态并得以稳定下来.顺理思之,人的卫星序列与酵母着丝粒的CDE III序列的同源性是意味深长的^[9].酵母着丝粒这一区域中AT含量很丰富,看来它对有功能的CDE I和CDE III序列单位起到间隔的作用.人的卫星I序列单位在人的着丝粒中也可能有类似的间隔作用.

目前尚不清楚着丝粒区的染色体结构域是如何折叠,以及是哪一种DNA序列组块参与了动粒的形成.最近的实验表明,电镜中所看到的动粒的板状结构是通过一种线状DNA纤维折叠而成的,这种DNA纤维则是由一些接头将一些串列重复的亚单位连接而成的.这一模型可能有助于说明着丝粒的重复DNA序列的特异性折叠结构在动粒形成中所起的作用.

3 染色体的端粒区

人染色体的端粒区有一种高度保守的小卫星重复序列家族(TTAGGG)_n^[10].在每一人染色体端粒上有250—1000个这种6体单位组织成特异的超级结构.在锥体虫的端粒上也发现同样的小卫星序列家族.它们与特异的核蛋白质相结合,并形成一种多链的折叠结构.这是由G四联(G-quartet)结构造成的;在G四联中,每一残基既为供体又为受体与邻近的G-G Hoogsteen碱基对相作用^[11].不同的端粒序列

单位都可形成G四联结构,因为只要序列中有成串的G即可.显然,保守的(TTAGGG)_n折叠结构是由保守的G四联结构衍生出来的.最近用寡核苷酸所作的实验表明,成串C的双链,通过互相间置的方式也可以形成C四联结构^[12].重复序列中其它核苷酸则为这种染色体末端结构提供适当间隔.因此,染色体端粒区成串的G可以解释为这种串列重复序列结构的密码结构域.

有证据表明,染色体的长度是多态性的^[13].这可能是由于端粒序列本身重复度迅速改变的缘故,也可能是由于在近端粒区的重复序列组块的座位特异超级结构发生不同重排所致.对细胞核中的染色体来说,维持长度的独特柔性可能是重要的:柔性可能参与间期核染色体三维折叠结构的最佳格局的形成.

4 染色体的核基质附着区

在细胞核内,DNA以30nm粗的染色质丝进行盘绕;在间期核内,它形成长度不等的绊环附着于核骨架上^[14].可以想象,弥散在基因组中的串列重复序列之密码结构域会被用来组织成每一染色体上的座位特异的绊环结构^[1].

细胞核内染色质结构域的拓扑状态看来是由核基质结构所维系的.这种结构主要由一套复杂核蛋白质和RNA组成.在核基质处不仅有转录和复制发生,而且核内初级转录物的加工,如剪接和聚A加成看来也由核基质所介导.

DNA附着于核基质结构的位点有人称为MAR,有人称为SAR位点.这些位点与核基质结合时,使DNA螺旋局部解旋,从而舒缓近邻的超盘绕的染色质绊环之间的超螺旋张力.MAR/SAR位点可以组成性地被利用,也可以被短暂地使用,主要取决于毗邻的染色质结构域遗传活性出现时间的长短^[15].

MAR/SAR位点一般含有70%—75%的AT.除了拓扑异构酶I结合位的共识序列之外,还有两种序列单位是不同物种的MAR/SAR区均为保守的.它们均是10bp的序列.一

种是所谓 A 元件, 即 AATAAA^T/_AAAA; 另一种是 T 元件, TT^A/_TTT^T/_ATT^T/_ATT. A 元件常常以几个拷贝形成串列的结构. 显然, 简单序列家族 (T)_n 和 (A)_n 较短的串列重复这种序列结构是 DNA 与核基质相结合的主要部分. 这些区域的 DNA 是卷曲的^[16], 并有一种特异的染色质折叠结构. 这可以在拓扑异构酶 II 结合位和 DNA 双螺旋局部解链部位看到.

虽然 MAR/SAR 位点的短重复 DNA 组块的密码结构域尚待界定, 但短重复简单序列使 MAR/SAR 位点出现特异的染色质折叠结构却是颇具启迪性的. 最近的一些研究结果表明, 短重复简单序列组块对特异的核蛋白质的结合亲和力不是由特异的一级结构型式, 而是由其特异的折叠结构决定的. 如上所述, 着丝粒区和端粒区外部确实存在着染色质的特异折叠结构. 由于这些折叠结构只是在基本的染色质结构域处于转录或复制时才出现或才具功能, 因此要在活体内进行实验跟踪显然是有困难的.

在转基因动物中, 转基因的表达一般是随其在染色体上的整合位点的变化而变化, 因而带有不可预测性. 这种所谓的位置效应意味着染色质组织成拓扑限制性的绊环对有功能的遗传结构域具有限定的作用. 真核生物体内有功能的转录单位已通过培养细胞的转染实验被证明是以 MAR/SAR 为边界来缓冲毗邻染色质的影响. 转基因动物的实验也证明, MAR 序列即所谓 A 元件可以使转基因的调节与其整合位点无关^[17]. MAR 显然是在染色质水平上发生作用的. 它们所具有的染色质边界和隔绝性质在将有功能的遗传组织成特定的形态上起着重要作用^[18].

5 同源异型框

最先在果蝇调节发育的蛋白质中发现有一种保守的 61 个氨基酸片段, 即同源异型结构域. 编码同源异型蛋白质的同源异型基因大都在其上游的远端、近端或下游的调节区中具有同源异型蛋白质的结合位. 这就有力地说明在

同源异型蛋白质和同源异型基因之间存在着一个自身调节和交叉调节的复杂网络. 相当大量的其它基因, 尤其是在发育期间表达型式发生改变的基因, 在它们的调节区中都含有同源异型蛋白质结合位. 同源异型蛋白质与其它转录蛋白质因子的相互作用、它们在核内的拷贝数、以及它们和特定调节位点的亲和力等会决定着哪些蛋白质去占据某个基因的调节区, 而不同的细胞类型和不同的发育阶段在这些方面都会有所差别^[19].

用纯化的同源异型蛋白质与同源异型基因的调节序列进行的足迹法实验已鉴别出它们在 DNA 上一些同族的结合位点序列. 大部分同源异型蛋白质的核心识别序列是 ATTA, ATT-TA 和 ATTTTA, 以及它们的互补形式 TAAT, TAAAT 和 TAAAAT. 哺乳动物的同源异型蛋白质也识别上述同样的序列. 另一种同源异型蛋白质的识别序列是如果蝇 Ubx 蛋白质所识别的 TAA 三核苷酸或类似于它的 TAATCG 的串列重复. X 射线晶体学分析表明, 61 个氨基酸的同源异型结构域识别 DNA 的 9bp, 看来, 像 ATTA 和 ATTTA 这些序列之间的距离会在 10—30 核苷酸左右, 使与之结合的蛋白质得以相互作用. 这与 MAR/SAR 位点的间隔序列长度相似.

不难看出, 在富于 AT 的区域中的 MAR/SAR 位点既与染色质的特异结构域有关, 而且也具有同源异型蛋白质的识别位, 以及像 AT-TA 和 ATTTA 作为复制起点的功能序列. 实际上, 人的载脂蛋白 B 基因全部 555bp 的 MAR/SAR 序列就是 TAAT, TAAAT, AT-TA, ATTTTA, TAAAAT 和 ATTTA 的嵌合形式^[20]. 可以想象, 这类 DNA 片段经与未知的核基质蛋白质相互作用而将基因锚固到核基质上. 从这一意义上说, 核基质附着区附近特异的高级结构所寓意的密码结构域或许是相当复杂的.

6 密码结构域可能的遗传功能

座位特异的超级结构中串列的重复序列组

块的编码容量可能从着丝粒、端粒和核基质附着区最近的分子生物学研究的结果得到揭示。特异的序列结构域即所谓密码结构域只有在串列重复序列组块中才易于以特异的折叠结构进行排列。有功能的和与狗 DNA 的重复序列同属一个序列家族, 因为一个特定的串列重复序列家族中不必各个成员功能上同样重要。由串列重复序列组成的座位特异超高级结构的出现和得以稳定看来不是由 DNA 序列本身所决定, 而是因有关染色质部位上特异高级结构的需要。

这种染色质折叠结构也应该是座位特异的, 因为建筑在 DNA 序列结构基础上的折叠结构看来可由特异的核蛋白质来使之稳定。因此串列重复序列组块潜在的遗传功能不是通过 DNA 的一级序列, 而是通过其能够呈现和维持特异的染色质结构来描述^[1]。

串列重复序列结构中的密码结构域是很保守的。例如, 用 CENP-B 框序列进行的实验表明, 这种密码结构域不仅在人染色体着丝粒的 α 类超高级结构中而且在小鼠染色体的小卫星序列家族中也存在。

MAR/SAR 位点与复制起点以及同源异型蛋白质识别序列交错镶嵌的结构模式可能为认识密码结构域的功能提供某种线索: 核基质结合位点附近的结构域不仅在 DNA 复制上有着重要的作用, 而且控制着发育期间基因有次序的表达。

核基质参与着发育期间复制起点有次序的激活。在特定的发育阶段, 每一细胞被激活的复制起点只不过是其潜在的全部复制起点的一小部分。发育过程中活跃的复制起点一般都趋减少^[21]。一个生物体内一种确立了的细胞类型的复制起点有次序的激活可能对基因有次序的表达起着某种作用^[22]。

此外, 在 S 期中, 基因的活动可能是活跃的和 not 活跃的基因有区别地复制的结果。由于转录和复制看来都发生在核基质上, 核基质也可能参与 S 期染色质不同部分复制的起始。S 期基因复制的顺序可能在发育期间是很重要的: 先复制的活跃基因可能抢先消耗细胞核内

的转录因子, 而使后转录基因的调节区无可利用。因此, 核基质可能确实是 S 期基因复制、发育期间复制起点有次序的激活、以及在同源异型蛋白质参与的基因有次序的表达, 乃至发育期间动物体型的形成等等多种作用荟萃和交融的场所。由此看来, 分子生物学在发育研究上的主要任务在某种意义上说就是破译密码结构域的含义, 进而阐明它们级联式的生物学功能。

参 考 文 献

- 1 Vogt P. Human Genet, 1990; **84**: 301
- 2 Peacock W J, Lohe A R, Gerlach N L *et al.* Cold Spring Harb Symp Quant Biol, 1978; **42**: 1121
- 3 Willard H F. Am J Hum Genet, 1985; **37**: 524
- 4 Tautz D, Renz M. Nucleic Acids Res, 1984; **12**: 4127
- 5 Zuckerkandl E. J Mol Evol, 1992; **34**: 259
- 6 Masumoto H, Masukata H, Muro Y *et al.* J Cell Biol, 1989; **109**: 1963
- 7 Grady D L, Ratliff R L, Robinson D L *et al.* Proc Natl Acad Sci USA, 1992; **89**: 1695
- 8 Murakami S, Masumoto T, Niwa O *et al.* Chromosoma, 1991; **101**: 214
- 9 Zinkowski R P, Meyne J, Brinkley B R. J Cell Biol, 1991; **113**: 1091
- 10 Raghuraman M K, Dunn C J, Hicke B J *et al.* Nucleic Acids Res, 1989; **17**: 4235
- 11 Williamson J R, Raghuraman M K, Cech T R. Cell, 1989; **59**: 871
- 12 Gehring K, Leroy J-L, Gueron M. Nature, 1993; **363**: 561
- 13 Foote S J, Kemp D J. Trends Genet, 1989; **5**: 337
- 14 Jackson D A. Trends Biochem Sci, 1986; **19**: 249
- 15 Manuelidis L. Science, 1990; **250**: 1533
- 16 Homberger H P. Chromosoma, 1989; **98**: 99
- 17 Mcknight R A, Shamay A, Sankaran L *et al.* Proc Natl Acad Sci USA, 1992; **89**: 6943
- 18 Phi-Van L, von Kries J P, Ostertag W *et al.* Mol Cell Biol, 1990; **10**: 2302
- 19 McGinnis W, Hart C P, Gehring W J *et al.* Cell, 1984; **38**: 675
- 20 McGinnis W, Krumlauf R. Cell, 1992; **68**: 283
- 21 Beachy P A, Krasnow M A, Gavis E R *et al.* Cell, 1988; **55**: 1069
- 22 Levy-Wilson B, Fortier C. J Biol Chem, 1989; **264**: 21196

Code Domains and Their Potential Functions.

Lao Weide. (*Institute of Biophysics, Academia Sinica, Beijing 100101*). *Prog. Biochem. Biophys.* (China). 1994; 21 (1): 13

Various sequence blocks, due to their specific primary sequence box, and/or a curved helical DNA conformation, and/or a non- β -helical DNA structure, can be organized in the corresponding higher order structures, and behave as functional sequence domains recognized and bound by specific proteins. They may be defined as "code domains". Code domains are genetically instructive for specific molecular interactions or processes, important not only in nucleus during interphase and during cell division, but also in differential gene expression during development and differentiation.

Key words code domain, chromatin superstructure, centromere, telomere, nuclear matrix, homeobox gene, differentiation and development

Structure and Function of Respiratory Chain Enzyme System. Xu Jianxing. (*Institute of Biophysics, Academia Sinica, Beijing 100101*). *Prog. Biochem. Biophys.* (China). 1994; 21 (1): 18

The studies in this lab on the enzyme system of respiratory chain were reviewed briefly. Two inhibitors which affect simultaneously on three Q-related enzymes of complex I, II and III were synthesized. They could be an useful tools in the studies of ubiquinone reactions in respiratory chain. The direct interaction between TTFA inhibitive site of $\text{QH}^+ \rightarrow \text{QH}_2$ of complex I and cytochrome b_{562} of complex III was found. This kind of interactions may play some roles in modulating the electron transfer from complex I to complex III. The maxi-

mum reconstitutive activity of lipid-depleted succinate-cytochrome c reductase with mixed phospholipids PC : CL : PE = 2 : 2 : 1 were obtained, this means not only the lipid but also thier composition were important for regulating the enzyme activiy of respiratory chain.

Key words ubiquinone, inhibitor, interaction

Progress in Topography of Ribosomal RNA and RNA N-Glycosidase Research (I). Zhang Jinsong, Liu Wangyi. (*Shanghai Institute of Biochemistry, Academia Sinica, Shanghai 200031*). *Prog. Biochem. Biophys.* (China). 1994; 21 (1): 23

Studies on ribosomal RNA topography play an important role toward elucidation of the functional roles of rRNAs in protein synthesis. RNA N-glycosidases are a group of ribosome-inactivating proteins which inactivate ribosomes by hydrolyzing the N—C glycosidic bond of a special adenylic acid in rRNA and releasing the adenine. Ricin A-chain is the first determined RNA N-glycosidase which has been studied in the greatest detail. Up to the present, 25 glycosidases have been characterized. The action site of RNA N-glycosidases lies in the α -sarcin domain of 28S rRNA. Treatment of RNA N-glycosidases induces conformational change of ribosomes and leads to the inactivation.

Key words ribosome topography, ribosome-inactivating protein (RIP), RNA N-glycosidase, trichosanthin

Latent Forms of Transforming Growth Factor- β : Structure and Function. Long Jianyin, Wang Huixin. (*Institute of Basic Medical Sciences, Academy of Military Medical Sciences, Beijing 100850*). *Prog. Biochem. Biophys.*