

同源异形域结构及对 DNA 的识别*

杨歧生

(浙江大学生物科学与技术系, 杭州 310027)

摘要 同源异形域蛋白是真核生物中一类重要的转录因子。根据同源盒基因及其同源异形域产物的肽链结构可以分为多种家族。文章综述了同源异形域与 DNA 结合的一般特点, 并叙述了 Antp、POU 等重要类型的转录因子如何识别 DNA 位点、HTH 及其他蛋白质在识别中如何起作用。

关键词 同源异形域结构, 蛋白质-DNA 相互作用, 转录因子, HTH 结构

同源异形域蛋白 (homeodomain protein) 是真核基因转录的一大类调控蛋白。最初认识的 Antp 是果蝇控制发育的基因, 在染色体上长达 103 kb。用该 Antp cDNA 作探针进行杂交试验, 发现在果蝇染色体上有 10 多个位点可以杂交, 是与 Antp 有同源性的一类基因。DNA 序列分析表明, 这些基因均含有约 180 bp 的同源盒 (homeobox) 结构, 不同基因的同源盒之间有较大的保守性, 并保持相同的阅读框架, 编码约 60 个氨基酸残基的肽段。这一肽段称为“同源异形域” (homeodomain, HD)。HD 与螺旋-转角-螺旋 (helix-turn-helix, HTH)、锌指结构、亮氨酸拉链等一样, 是真核生物 DNA 结合蛋白中一类特殊的结构式样^[1]。

1 同源异形域的类型

根据同源盒基因在染色体上的分布, 可将基因分为复合型和分散型。复合型基因在染色体上成簇串联分布^[2]。果蝇的 Hom 同源盒基因簇在染色体上至少有 8 种基因 (图 1), 其分布顺序与果蝇发育中躯体前后轴向表达顺序相同, 即所谓线性排列规律。这种规律也适用于高等脊椎动物的同源盒基因。Hom 还可进一步分为 3 种主要类型: Antp、Abd-B 和 lab。其中 Antp 可以分成若干家族 (family): Dfd, Scr, Antp, pb, XloX 和 Gsh 等, 它们之间的 HD 有 98% 同源性。哺乳动物有 4 簇同源盒基因, 分

别称为 Hox a、b、c 和 d, 共约 13 种基因, 来源于同一原始基因, 通过基因重复和交换逐渐增多。Hox 成员的同源盒结构有共同序列 (consensus sequence)。Hom 和 Hox 簇各成员之间有一定的对应性, 因此复合型也称为 Hom/Hox 簇。位于 Hom 中央的 Antp 与 Hox 6、7 的 HD 较类似, 并与共同序列偏离较小。基因簇两侧的基因在进化上分开最早, 它们的 HD 差异较大^[3]。

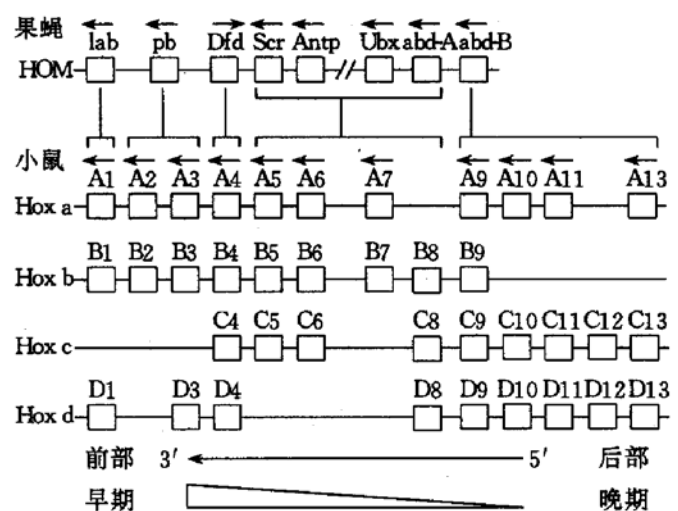


图 1 果蝇 Hom 和脊椎动物 Hox a~d 同源盒基因的成簇排列

箭头表示各基因的转录方向。

*国家自然科学基金资助项目。

收稿日期: 1995-02-20, 修回日期: 1995-08-02

分散型的同源盒基因数量很多, 分布在各染色体上. 主要根据同源异形盒蛋白的结构, 可以分为约 16 种类型, 如 eve、ems、Dll、cad、Hlx、msh、TLC/NEC、NK-1、NK-2、en、prd、prd-like、cut、LIM、ZF 和 POU 等. 它们之间

的差异包括 HD 的同源性程度不同、分子内其他肽段结构不同, 显示出蛋白质水平上的多样性. 复合型的 Antp 和 lab 类型、分散型的 cad、ems 和 TLC 等类型除了有各自的 HD 结构之外, 均有保守的六肽 IYPWMK 结构, 因此统称为 HEX 超类 (superclass). en 类型有 4 个高度保守的肽段 EH, 分别位于 HD 的两侧. prd 类型具有高度保守的 130 个氨基酸残基的肽段, 称为“配合域” (paired domain), 它是 prd 第二个结合 DNA 的结构域. POU 蛋白的名称来自最初分离到的转录因子 Pit-1、Oct-1、2 和 Unc-86, 它们都含有 2 个结合 DNA 的结构域, 即 POU-HD 和约 80 个氨基酸的 POU 特异性结构域 (POU-specific domain). 其他主要类型的 HD 蛋白基本结构见图 2.

2 同源异形域的结构

2.1 HD 的共同序列

HD 是由同源盒编码的一段高度保守的肽段, 一般由 60 个氨基酸残基构成. 图 3 表示根据 346 种 HD 得到的共同序列^[4].

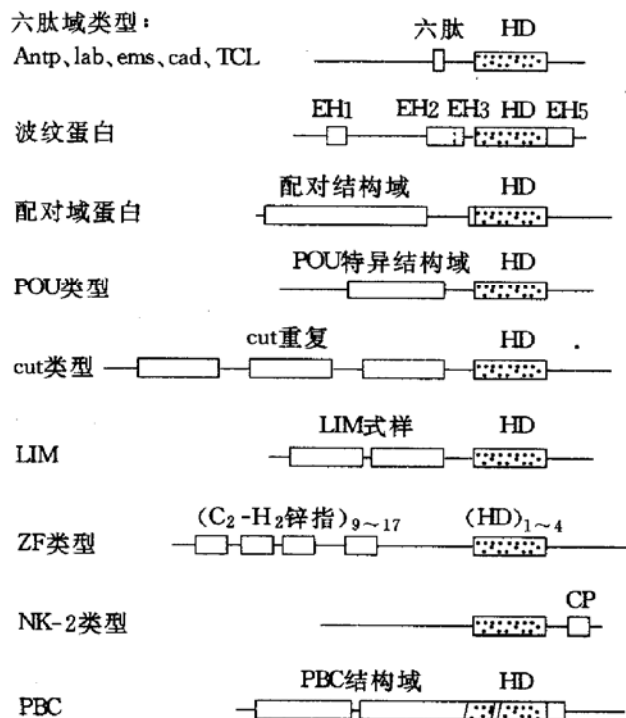


图 2 各种同源异形域蛋白的结构式样

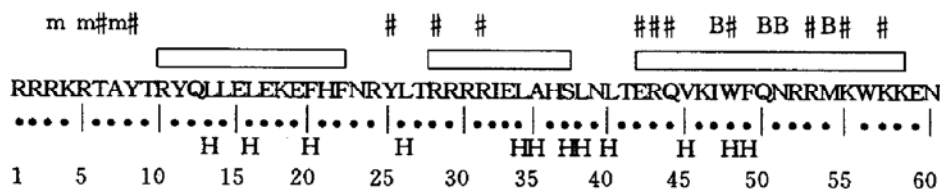


图 3 HD 的共同序列

HD 形成 3 个 α 螺旋. H 的残基形成疏水核心, B 的残基同 DNA 大沟接触, m 的残基同小沟接触, # 的残基同磷酸基团接触.

2.2 HD 的立体结构

NMR 和 X 衍射分析表明 HD 形成 3 个 α 螺旋. 螺旋 I 和 II 平行, 螺旋 III 与前两个螺旋基本垂直^[5,6]. Antp 的 HD^[7] 在第 52/53 残基形成一个结, 使螺旋 III 分成 2 个螺旋 (III/IV). II、III 和它们之间的转折形成保守的 HTH 结构. HD 的立体构架是由 11 个残基折叠成核心结构, 其中 10 个残基呈疏水性, 6 个残基 (Leu46、Leu38、Leu40、Ile/Val45、Trp48 和 Phe49) 高

度保守. 由于核心内这些保守残基的存在, 使不同类型的 HD 均可折叠成十分相似的立体结构. HD 的 N 端容易弯曲, 不影响 HD 的基本结构框架, 但与 DNA 结合时显示重要的特异性.

3 同源异形域与 DNA 的结合

用 NMR 和 X 衍射分析果蝇 Antp HD 蛋白与人工合成的 14 bp 双链 DNA 的复合物, 获得了它们相互作用、识别的大致轮廓^[8,9]. HD

蛋白与 DNA 结合主要通过 HTH 样识别螺旋起作用. 与原核 HTH 结构不同, 真核的 HD 接触 DNA 涉及更多位点, 往往还需要多种特定蛋白质参与, 保证识别的特异性, 并且 HD 蛋白多以单体形式结合 DNA. 以 Antp 蛋白为例可以了解 HD 对 DNA 结合的一般特性.

3.1 识别螺旋对 DNA 大沟的接触

与很多真核转录因子类似, HD 的识别螺旋大致可以分为两部分, C 端的 Arg/Lys 残基倾向于接触 DNA 的磷酸基团, N 端残基负责同碱基相互作用, 因此具有定位和识别的双重功能. Antp HD 的螺旋 III/IV 与 DNA 位点结合, 作用包括对 DNA 两条主链 (α 和 β 链) 的非特异性接触、对特定碱基识别两个方面. Arg43、Arg52 和 Arg55 等的侧链与 β 主链的磷酸基团形成盐键, Arg53 接触 α 主链. 这样, HD 被定位到大沟两侧的主链骨架上. 同碱基的结合包括 Ile47 对碱基 T8 和 A9, Gln50 对 C7 和 T8, Met54 对 C7 等. 另外, Ile47 同 β 主链、Met54 同 α 主链分别有疏水作用, 对接触加以补充.

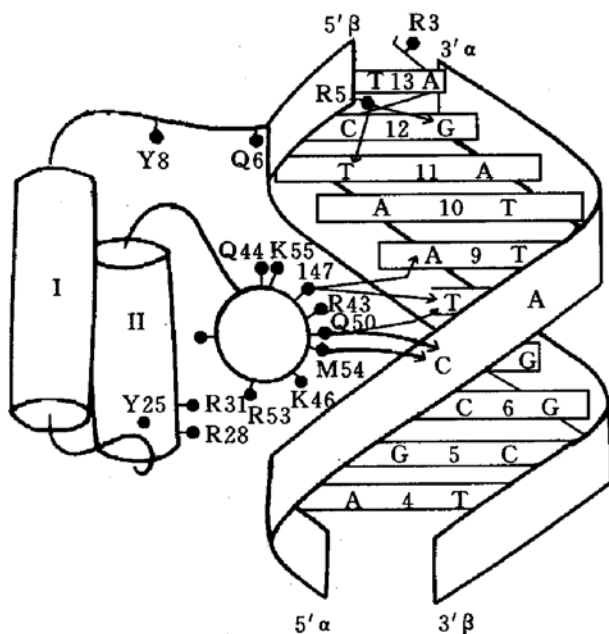


图 4 Antp HD-DNA 复合物内各种相互作用的概况

3.2 邻近残基和环区对 DNA 的接触

HD 的 N 端容易弯曲, 进入 DNA 小沟内

参与接触. Arg3 同磷酸基团形成盐键, Arg5 先同糖基形成疏水接触, 然后对碱基 T11 和 G12 形成氢键. 螺旋 I 和 II 之间的环区内残基也参与接触, 如 Tyr25 接触糖基, 环区邻近的 Arg28、Arg31 分别与磷酸基团形成盐键.

3.3 某些残基有高度亲和性的识别能力

HD 蛋白优先作用的 DNA 位点大多含有 ATTA (互补链 TAAT) 核心序列. 在 en 或者 Antp HD-DNA 复合物中, 保守残基 Ile47 和 Asn51 在大沟内同 ATTA 相互作用 (在 en 复合物中为 TAAT, Antp 复合物中为 TAAT). Arg3 和 Arg5 在小沟内同 ATTA 相互作用 (en 复合物中为 TAAT, Antp 复合物中为 TAAT). 这 4 个氨基酸残基非常保守, 对 DNA 核心序列表现高度的亲和性^[10,11]. 这些保守残基的改变, 识别的核心序列也相应改变. 例如酵母 MAT α 2 蛋白中 HD 的第 47 位是 Asn (不是 en 和 Antp 的 Ile47), 结果能高度亲和识别的核心序列是 GTAA.

HD-DNA 复合物的另一特点是当识别螺旋第 50 位为 Lys 时, 它能识别的核心序列 ATTA 之前通常有 GG 序列, 即 GGATTA. 如 bcd、otd、prd (K50)、ftz (K50) 等优先结合于有 GG 的 ATTA 核心序列. 有些 HD 第 50 位是 Gln, 就倾向优先结合于 ATTA 之前有 CC 的核心序列. 因此, 第 50 位点是特异性作用的重要残基之一. 如 Gln 代替 Lys50, bcd 会降低对有 GG 的核心序列的亲和性. 若相反, Ser50 或 Gln50 换成 Lys50, 就会增强对有 GG 的核心序列的结合, 同时降低对前面无 G 的核心序列的亲和性. prd (K50) 和 ftz (K50) 的情况说明如此^[12]. 近来在果蝇胚胎发育的研究中证实了 HD 第 50 位残基是对 DNA 结合特异性的主要决定因素^[13].

3.4 其他蛋白参与 HD 蛋白的功能特异性

在离体实验中, 差异很大的 HD 蛋白往往识别相似的 DNA 序列元件, 特别是许多蛋白有相似的 HD, 这与调控蛋白在细胞中起着各不相同的、微小差异的生物学功能十分矛盾. HD 蛋白如何实现各自结合 DNA 的特异性? 主

要通过两种方式共同起作用: a. 除 HD 外, 其邻近残基也参与对 DNA 序列特异性的接触; b. 通过与其他辅助蛋白、转录因子的相互作用, HD 蛋白显示出各种识别特异性。

酵母 MAT α 2 蛋白能同两种不同的操纵基因相互接触, 控制两种基因的表达, 这两种操纵基因的序列元件均是 CATGTAA. MAT α 2 通过同两种不同 DNA 结合蛋白 (MCM1 和 MAT α 1) 结合, 选择性地结合到某一种操纵基因上. MAT α 2 分子内紧挨 HD 的 N 端一小段肽链负责与 MCM 1 蛋白作用, 而 MAT α 2 的 C 端 20 个残基呈无立体构象的伸展状态, 介导同 MAT α 1 相互作用. 不同蛋白质参与, 造成 MAT α 2 结合相同序列元件时有不同的空间构架和方向, 从而分辨两种不同的操纵基因. 因此, 尽管调控因子有相似的 HD, 但有不同的 N 端和 C 端, 可与不同的辅助蛋白作用, 对 DNA 序列有很不同的亲和性^[14,15].

4 POU 同源异形域蛋白

POU 蛋白是一类重要的调控因子, 参与哺乳动物脑基因表达和激素对基因的调控. 现对约 30 余种 POU 蛋白结构作了分析, 其中对 Oct-1 等研究得较详尽.

4.1 POU-DNA 复合物的整体结构

POU 蛋白的特点是具有 POU 特异性结构域和 POU-HD. 两者偶联存在, 是独立的 HTH 结构单位, 都直接结合到 DNA 序列上, 对碱基和磷酸基团作广泛的接触, 两者是高亲和性地识别 DNA 序列所必需的^[16]. 以 Oct-1 为例说明复合物的一般特点. 它同八核苷酸 ATGCAAAT 结合. 其中 POU 特异性结构域同 5' 端半位点 ATGC 接触, POU-HD 同 3' 端半位点 AAAT 接触. 复合物的结构分析表明, 两个结构域分别结合于双链 DNA 相反两面的大沟内. 两个结构域之间没有蛋白质-蛋白质的直接相互作用, 各以 HTH 单位内的识别螺旋定位于相背的大沟内. 两个结构域之间 24 个残基的肽段在晶体中无定形, 使蛋白质分子内两个结合 DNA 的部位结合得更精确, 识别更具

特异性.

4.2 POU-特异性结构域与 DNA 相互作用

POU 特异性结构域由 4 个 α 螺旋组成, 有高度保守的紧密的疏水核心, 许多亲水性残基位于蛋白质一侧表面. 螺旋 II、III 及其转折区形成 HTH 单位. 螺旋 II 和转折区要比原核的 HTH 长数个残基.

Oct-1 中螺旋 III 是识别螺旋, Thr45 同 ATGC 序列中的 T2 和 C3 形成氢键, Arg49 同 G3 和 G4 形成氢键. 如果把 Arg49 换成 Ala, 就会破坏 Oct-1 对 DNA 的接触. Gln44 同 A1 形成氢键. 结构域内其他保守性残基也参与接触, 如 Arg20 和 Gln27 与碱基及磷酸基团的氧原子形成氢键, 使整个氢键网络更加稳定^[17].

4.3 POU-HD 与 DNA 相互作用

POU-HD 在 DNA 大沟内与 3' 端半位点 AAAT 结合. 关键性的结合由识别螺旋内 Val47、Asn51 和 Cys50 的侧链进行. Asn51 与 A7 形成氢键, Val47 与 T8 的甲基发生疏水性作用. Cys50 是 POU-HD 特有的, 其 —SH 基与八核苷酸之外的位点接触, 其原因尚不清楚. 另外一些保守残基如 Arg5、Thr6 和 Arg13 同八核苷酸一条主链上磷酸基团形成氢键, Lys25、Ser48、Arg46 和 Arg53 与另一主链的磷酸基团接触. 尽管不同 POU 蛋白的 HD 有较大差异, 但与 DNA 结合的方式却十分相似, 特别是 Asn51 非常保守, 总是同 AAAT 的第 3 个 A (即 A7) 相互作用, Arg53 和第 25 位残基 (Oct-1 中是 Lys25) 与磷酸基团形成盐键.

4.4 POU 两个结构域与 DNA 作用的协同性

Oct-1 分离的 POU-特异性结构域和 POU-HD 分别与靶序列 ATGCAAAT 进行结合, 测定解离常数. 发现单独的 POU-特异性结构域的结合很弱, POU-HD 的结合较好, 而含有两者的肽段结构对这一靶序列的亲和性很强. 如果在该靶序列的两个半位点之间增加 1~3 bp (如 ATGCAGCAAAT 等), 就会大大降低亲和性. POU 两个结构域之间的肽链使得两侧的 POU 结构域更有利于定位到八核苷酸序列上. 两个 POU 结构域对 DNA 半位点的结

合能力具有协同作用^[10,18].

POU 蛋白不仅仅依靠两个保守的结构域来分辨 DNA 序列, 也通过辅助因子来调节自身的识别功能. 例如 Oct-1 与辅助因子 VP16 形成复合物, 结合于 TAATGARAT 序列, 获得转录激活活性的能力, 显示转录调控因子的特性. Oct-1 与 VP16 相互作用主要来自 POU-HD 内第 22 位 Glu 残基. 如果用 Ala 替换 Glu22, 就会明显降低 Oct-1 同 VP-16 的偶联能力. 而相反, Oct-2 的 HD 内 Ala22 换成 Gln, 就会得到能同 VP-16 偶联的 Oct-2 蛋白. 一般情况下 Oct-2 不与 VP-16 结合.

同源盒基因发现 10 多年来, 已对其产物同源异形域蛋白的结构和功能作了大量研究, 认识到在真核生物个体发育中它们具有调控因子的作用, 对很多转录因子的编码基因、生长因子编码基因进行作用, 参与脑基因的调控^[19]. 可以预期同源异形域蛋白具有更广泛的生物学功能. 它们的结构上有一定相似性, 在功能上又有差异. 与 DNA 结合时, 除了特定氨基酸残基显示特异性之外, 其他蛋白因子也参与识别和介导对基因的调控.

参 考 文 献

- 1 Mc Ginnis W, Levine M S, Hafen E *et al.* Cell, 1984; 37: 403
- 2 Kaufman T C, Seeger M A, Olsen G. Adv Genet, 1990; 27: 307
- 3 Akam M. Cell, 1989; 57: 347
- 4 Bürglin T R. In: Duboule D ed. Guidebook to the homeobox genes. New York: Oxford U Press. 1994: 27
- 5 Billeter M, Qian Y Q, Otting G *et al.* J Mol Biol, 1990; 214: 183
- 6 Giintert P, Qian Y Q, Otting G *et al.* J Mol Biol, 1991; 217: 531
- 7 Gehring W J, Müller M, Affolter M *et al.* Trends Genet, 1990; 6: 323

- 8 Otting G, Qian Y Q, Billeter M *et al.* EMBO J, 1990; 9: 3085
- 9 Billeter M, Qian Y Q, Otting G *et al.* J Mol Biol, 1993; 234: 1084
- 10 Kissinger C R, Lir B, Martin-Blanco *et al.* Cell, 1990; 63: 579
- 11 Qian Y Q, Resendey-Perey D, Gehring W J *et al.* Proc Natl Acad Sci USA, 1994; 89: 10738
- 12 Percival-Smith A, Müller M, Affolter M *et al.* EMBO J, 1990; 9: 3967
- 13 Schier A F, Gehring W J. Nature, 1992; 356: 804
- 14 Mak A, Johnson A D. Genes Dev, 1993; 7: 1862
- 15 Vershon A K, Johnson A D. Cell, 1993; 72: 1
- 16 Li P, He X, Gerrero M R. Genes Dev, 1993; 7: 2483
- 17 Dekker N, Cox M, Boelens R *et al.* Nature, 1993; 362: 852
- 18 Wright P E. Curr Opin Stru Biol, 1994; 4: 22
- 19 Fujii H, Hamada H. Neuron, 1993; 11: 1197

Homeodomain Structure and Its Recognition to DNA. Yang Qisheng (*Department of Biological Science and Technology, Zhejiang University, Hangzhou 310027, China*).

Abstract Homeodomain proteins are very important regulation factors in eukaryotic transcription and development. They are classified into many families according to structure of homeobox genes and homeodomain peptide chains. The interaction of homeodomain and DNA was characterized. How the Antp and POU homeodomains recognized the binding sites of DNA sequences was explained. The function of HTH structure and other proteins in recognition were also described.

Key words homeodomain structure, protein-DNA interaction, transcription factor, HTH structure