

ADP in a manner competent to form bound ATP. The second is that during net ATP formation multiple catalytic sites on the synthase participate in strongly cooperative sequence. Rotation of the γ subunit in F_1 is thought to deform the catalytic sites to give binding change. When

the crystal structure of the F_1 -ATPase was eventually solved, direct evidences for rotation of subunits during catalysis of F_1 -ATPase were provided.

Key words ATP synthase, binding change mechanism, rotational catalysis

转录因子与 DNA 识别的立体化学规律

杨岐生

(浙江大学生物科学与技术系, 杭州 310027)

摘要 转录因子-DNA 的识别包含普遍性的化学规律和特异性的立体化学规律. 转录因子的识别螺旋以特异性方式结合于 DNA, 位于一侧的大沟内. 识别螺旋的“残基连线”和碱基的“碱基连线”之间吻合使两者完全匹配. 它通常涉及最多 3 圈 α 螺旋和 3~5 bp. 决定氨基酸残基和碱基相互关系的结合几何图形表示为立体化学图, 它表明了识别的特异性. 在该基础上总结成转录因子识别 DNA 的立体化学规律.

关键词 转录因子, 识别螺旋-DNA 的相互关系, 立体化学图, 立体化学识别规律

学科分类号 Q523

认识转录因子与 DNA 的识别对研究基因表达和调控具有重要意义, 也是结构生物学的重要内容. 两者识别有什么规律? 研究这种规律使我们能够: a. 预期蛋白质-DNA 的相互作用; b. 改变已有转录因子或 DNA 序列的结合特异性; c. 设计结合于特定 DNA 序列的新的蛋白质分子, 因此具有重要的理论和实践意义.

蛋白质-DNA 识别主要由两类规律组成: 化学规律和立体化学规律, 化学规律是普遍性的, 立体化学规律则对各类 DNA 结合蛋白家族有特异性. 本文讨论 α 螺旋作为识别螺旋的情况. β 折叠的结构式样将另作讨论.

1 DNA 结合蛋白的结构域

识别螺旋可以分为 HTH、PH (probe 螺旋, 包括同源异形域和碱性拉链蛋白等家族)、锌指结构、C4 锌指蛋白 (包括甾体激素受体家族和 GATA 蛋白), 以及新的结构式样 Myb 等. 20 种氨基酸残基按其大小分为 4 类: 大、

中、小和芳香族残基. 残基的大小涉及到侧链与 DNA 接触和配合的能力.

原先以为蛋白质与 DNA 结合时, 蛋白质的柔性能使识别螺旋足够地变形, 以适应 DNA 的结构. 现在认为不是蛋白质, 而主要是 DNA 结构发生变化, DNA 柔性加强了蛋白质-DNA 的相互作用^[1]. 识别螺旋沿着 DNA 弯曲的大沟, 在大沟一侧接触到碱基, 但接触的碱基数目有限, 一般不超过 5 bp.

转录因子的识别螺旋长度只有几圈. 它的残基大致可分为 3 类: a. 同 DNA 碱基接触的残基; b. 同主链磷酸基团接触的残基; c. 同蛋白质其余部分接触的残基. 各种结构式样中这些类型的残基都有比较保守的位置. a 类大多是极性的, 也有酸性或碱性残基; b 类大多是碱性氨基酸; c 类往往以疏水性残基背向 DNA, 与蛋白质分子其余部分结合, 限制了

螺旋的转动，对结合的几何图形是重要的。这 3 类残基排列成单个识别螺旋的方式，对每类转录因子家族都是特异性的，使它在 DNA 上结合时有特定的结合几何图形。

2 识别的化学规律

化学规律是指一定的氨基酸残基与碱基之间非共价的弱作用，两者间形成氢键、疏水作用以及部分离子效应的内在能力^[2]。结合的特异性主要来自氨基酸侧链与碱基之间的化学接触。蛋白质与 DNA 核糖-磷酸主链骨架之间

的接触不是特异性识别的主要来源。

化学规律主要包括：a. 残基与碱基之间氢键的形成；b. 疏水作用，T 在大沟内有强烈的疏水作用；c. 有的残基可以与同一 DNA 链或不同链（W 链和 C 链）的两个碱基之间相互接触，构成“桥联”；d. 离子效应的作用。

能够接触某一碱基的氨基酸残基很容易被大小相似的同一类残基所取代。氨基酸残基同碱基识别的这种相互关系是一种普遍的化学规律，可以综合成“化学识别表”（表 1）。

表 1 氨基酸残基与碱基相互作用的化学规律

小残基 (s)	中等残基 (m)	大残基 (l)	芳香族残基
A (Cys Ser Thr)	Asn Asp (His)	Gln Glu (Arg Lys Met)	(Tyr Trp)
T Ala (Cys Ser Thr)	Val Ile (Asn His)	Leu Met (Gln Arg Lys)	Tyr Phe Trp
G (Cys Ser Thr)	His (Asn)	Arg Lys (Gln)	(Tyr)
C Val (Cys Ser Thr)	Asp (Asn His Ile)	Glu (Gln Leu Met)	Tyr Phe Trp

注：黑体字残基表示强的结合，括号内残基有弱结合。

3 相互识别的主要特性

立体化学规律是指结构式样与 DNA 相互匹配时三维结构上的一般规律性。同一类转录因子家族有共同的结构式样，所以有特异性的 DNA 结合几何图形^[3]。

a. 结合的结构要素。识别螺旋的结构式样是识别特异性、亲和性的关键，包括螺旋的二级结构特性、长度、3 类残基的布局，DNA 沟的宽度、深度和序列，识别螺旋在大沟内倾斜的角度，接触的碱基对数目（一般为 4~5 bp）及其排列位置等。

b. DNA 位点只有几个碱基对。DNA 大沟要有足够的宽度，恰好可以接受 α 螺旋。 α 螺旋基本上是挺直的，进入弯曲的大沟中处于 DNA 的一侧，只能接触到 5 个连续的碱基对，甚至更少，选择自己在化学规律和立体化学上可以匹配的序列，有多个接触点。PH 和 HTH 的识别螺旋同 DNA 大沟呈切线结合。这里受到两种因素的影响。一是蛋白质的结合诱导 DNA 构象的弯曲，二是转录因子结合位点往

往是两个不连续的序列，受间隔序列的影响。

c. 识别螺旋只有几圈。在一般的识别螺旋与 DNA 的复合物中，5 个碱基对中有 8 个碱基可参与相互作用，即 C1~ C4 和 W2~ W5，很难接触到 W1 或 C5 等碱基。为了识别这 8 个位点的碱基，识别螺旋的 3 圈内最多有 9 个氨基酸残基可被利用。 α 螺旋的螺距是 0.54 nm，面向 DNA 的 3 圈螺旋的跨度是 1.08 nm，而 3 个碱基对绕 DNA 螺旋轴相应距离是 1.0 nm，两者基本可以配合^[4]。如果识别螺旋有第 4 圈就会与 DNA 主链上磷酸基团接触了。

d. DNA 位点上大沟的宽度有变化。一般情况下，B-DNA 双螺旋的大沟和小沟平均宽度稍有波动，如 HTH 或 PH 蛋白质结合在识别位点上，大沟的中央变得最窄^[5]。蛋白质结合时，大沟和小沟呈相反的变化。

e. DNA 的弯曲有利于蛋白质的结合。某些蛋白质结合可使 DNA 有弯曲的构象。DNA 碱基台阶 (step) 的卷动引起 DNA 双螺旋轴弯曲^[6]，大沟内 2 个碱基对之间的距离减小，

小沟距离增加，使 DNA 螺旋向大沟一侧弯曲，小沟变宽，大沟变窄。在各种卷动的碱基台阶中，以嘧啶-嘌呤序列（如 TG/CA, CG, TA）为主，容易造成 DNA 轴向大沟一侧弯曲。DNA 分子的柔性改变其结构，立体结构有很大变化，能更好地与转录因子结构式样表面吻合、匹配。DNA-蛋白质复合物中，蛋白质在结构没有大的差别^[7,8]，而 DNA 的结构却与标准 B-DNA 有很大的不同。

f. 识别螺旋有一套残基决定与 DNA 的立体化学关系。识别螺旋在 DNA 大沟内的倾斜是立体化学识别中很重要的内容，这种倾斜是由识别螺旋上特异性结构要素固定下来的。例如 PH 类型的识别螺旋具有保守的 Arg/Lys，它们结合到 DNA 磷酸基团上，使螺旋在大沟内有一定的倾斜，因而大致固定了结合的几何图形。每个结构式样都有一定的 α 螺旋，使用一套特定的残基去识别碱基。这种关系能总结成一张立体化学图，表示出优先使用、识别的残基的大小和特异性。经数据统计，建立模型，作出立体化学关系的一般性认识^[9]。

4 识别螺旋与 DNA 相互关系的参数

识别螺旋有 3 种情况^[10]：a. 只用 1 圈螺旋作为识别螺旋，如 TR、C6（包括 Gal4 等）。与大沟结合时，识别螺旋大致同 DNA 大沟垂直，因此称为“垂直配合”，识别螺旋的一端（N 或 C 端）插入大沟；b. 用 3 圈螺旋作识别，如 PH、C4，识别螺旋同大沟平行，因此称为“平行配合”。c. 用 2 圈螺旋进行识别，如各种 HTH、ZnF 蛋白，识别螺旋以范围较广的角度结合在大沟内。这些结合的情况下，参数 α 和 β 是识别螺旋在大沟内的角度，参数 d 是识别螺旋的中心点与 DNA 双螺旋轴的最短距离， α 表示识别螺旋轴同 d 之间的夹角， β 为识别螺旋轴同双链 DNA 轴之间的夹角。从大量已知蛋白质-DNA 复合物晶体结构的分析数据表明， α - β 图是二条直线， $\beta = 0.67\alpha + 25$ 或 $\beta = -0.67\alpha + 25$ ， β 依赖 α ，或者相互依赖。所以识别螺旋在大沟内的几何

图形基本上为单的自由度，可由角度 α 来表示特征。在决定识别螺旋的倾斜程度时， α 角起着重要作用。

5 表面匹配和结合的几何图形

特异性结合是通过 α 类残基与特定碱基的作用来达到的。接触表面只占蛋白质或 DNA 分子的很小部分，约 28 nm^2 ，局部表面的氨基酸残基-碱基之间必须紧密匹配，涉及数十个化学接触，包括非共价的特异性接触、残基大小匹配等。一个长的侧链能触到大沟内较远位点的碱基。当某个小残基很容易匹配的结构，用大残基却不能匹配。立体化学规律要表示出这类相互关系。

根据各种已知复合物的立体结构进行分析，同一转录因子家族往往具有相同的接触残基和碱基位点。识别螺旋内面向大沟的残基连成“氨基酸残基连线”，DNA 位点内的碱基也连成“碱基连线”。相互结合时，两连线重叠，氨基酸残基之间的距离必须接近于它们所接触的碱基之间的距离，并在化学规律上完全匹配，识别螺旋与大沟才形成稳定的相互作用。蛋白质-蛋白质识别也有类似情况^[11]。

不同的转录因子家族取不同的结合几何图形，因而有不同的立体化学图。立体化学图

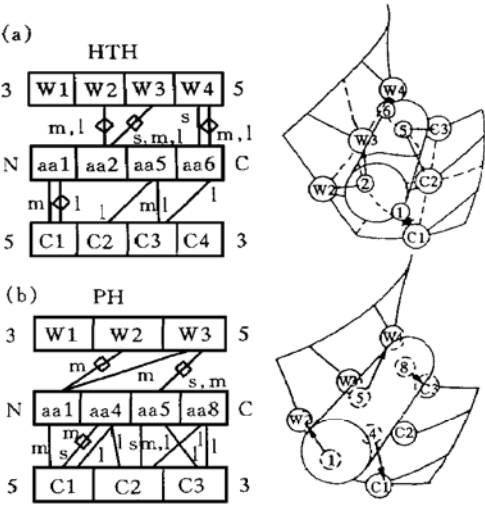


图 1 HTH 和 PH 识别螺旋与 DNA 识别时的立体化学规律

(图 1) 表示出识别螺旋结合到 DNA 大沟上的大致轮廓. 一个转录因子家族中, 接触 DNA 碱基的氨基酸残基不是非常保守的, 通过改变这些残基, 家族的成员就可以在 DNA 序列之间作出辨别. 立体图形仍保持类似的特异性, 但识别位点发生了改变^[12].

6 识别密码表

把化学规律和立体化学图结合起来, 立体化学上可以成对的氨基酸和碱基构建识别密码表^[4]. 表 2 表示 C4 和 A 型锌指结构 (ZnF)

识别螺旋的识别密码表. 例如, C4 识别 DNA 位点时, aa1 残基与 W2 碱基结合, 往往要求 aa1 为中等 (m) 或大 (l) 残基. 当 W2 为 A 时, aa1 以疏水或亲水的 Asn、Asp、Gln 或 Glu 与之结合. 大残基 aa9 与 W4 位点结合, 当 W4 为 C 碱基时, 亲水性的 Glu 与 C 有较强的结合, 疏水性的大残基 Leu 或 Met 结合能力较弱.

这类识别密码表可以初步估测相互结合的规律, 为改造转录因子的特性, 构建新的转录因子打下了基础.

表 2 DNA- 识别螺旋识别的立体化学规律和 C4、ZnF 的识别密码表

C4	W4 aa9 (l)	W3 aa5 (m, l)	W2 aa1 (m, l)
A	Gln Glu	Asn Asp Gln Glu	Asn Asp Gln Glu
T	Leu Met	Val Ile Leu Met	Val Ile Leu Met
G	Arg Lys	His Arg Lys	His Arg Lys
C	Glu (Leu Met)	Asp Glu (Leu Met Ile)	Asp Glu (Leu Met Ile)
ZnF	W4 aa7 (l)	W3 aa4 (m, l)	W2 aa1 (l)
A	Gln Glu	Asn Asp (Gln Glu)	Gln Glu
T	Leu Met	Val Ile (Leu Met)	Leu Met
G	Arg Lys	His (Arg Lys)	Arg Lys
C	Glu (Leu Met)	Asp (Glu Leu Met Ile)	Glu (Leu Met)

注: 氨基酸残基以黑体字表示强的识别, 括号内表示弱结合.

7 间隔序列的规律有待研究

原核的 HTH 家族成员与 DNA 作用的模式相似, 结合序列都有相似性, 但通过结合位点之间间隔序列的差异, 不同转录因子能分辨出不同的 DNA 元件. 真核生物的 PH、C4、锌指结构等家族的转录因子, 也是通过间隔序列的变化, 使同一家族的不同转录因子能分辨出不同基因的调控元件. 因此, 间隔序列在识别和调控上具有重要作用. 总之, 转录因子对 DNA 识别的一般性规律尚未很好认识, 需在更高的复杂水平上作研究和分析.

参 考 文 献

1 Wener M H, Gronenborn A M, Clore G M. Intercalation, DNA kinking, and the control of transcription. *Science*, 1996, **271**: 776~ 784

2 杨岐生. 真核转录因子对 DNA 序列的识别. *生命的化学*, 1996, **16** (2): 7~ 9

3 Suzuki M, Brenner S E, Gerstein M *et al.* DNA recognition code of transcription factor. *Protein Engineering*, 1995, **8**: 319~ 328

4 Heer W, Cleary M A. An in the groove view of DNA structure in complex with protein. *Genes & Development*, 1996, **255**: 677~ 687

5 Suzuki M, Gerstein M. Binding geometry of α helices that recognize DNA. *Protein Struct Funct Genet*, 1995, **23**: 525 ~ 535

6 Chasman D I, Flaherty K M, Sharp P A *et al.* Stereochemical basis of DNA bending by transcription factor. *Nucleic Acids Res*, 1995, **23**: 2083~ 2091

7 Kim J L, Nikolov D B, Burley S K. Crystal structure of a yeast TBP/TATA-box complex. *Nature*, 1993, **365** (6446): 512~ 519

8 Kim Y, Geiger J H, Hohn S *et al.* Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature*, 1993, **365** (6446): 520~ 527

9 Suzuki M, Yagi N. DNA recognition code of transcription

- factors in the helix-turn-helix probe helix, hormone receptor, and zinc finger families. Proc Natl Acad Sci USA, 1994, **91**: 12357~ 12361
- 10 Suzuki M. Common features in DNA recognition helix of eukaryotic transcription factors. EMBO J, 1993, **12**: 3221~ 3226
- 11 Chothia C. Principles that determine the structure of proteins. Ann Rev Biochem, 1984, **53**: 537~ 572
- 12 Suzuki M, Yagi N, Gerstein M. DNA recognition and superstructure formation by helix-turn-helix proteins. Protein Engineering, 1995, **8**: 329~ 538

Stereochemical Rules of DNA Recognition by Transcription Factors. YANG Qi-sheng (Department of Biological Science and Technology, Zhejiang University, Hangzhou 310027, China).

Abstract DNA recognition of the transcription factors involves both general chemical rules and

specific stereochemical rules. Recognition helix of transcription factor families binds to DNA, positing at a major groove. Pairing between the “residues line” of recognition helix and “base line” of base positions makes full fitting, usually involving 3 turns of α -helix and 3~ 5 base pairs. The binding geometry determined by interaction of the residues and bases in recognition area is indicated in the stereochemical chart, which shows recognition specificity. Stereochemical rules of DNA recognition by transcription factor family are summarized at bases of the chart.

Key words transcription factor, recognition helix-DNA interaction, stereochemical chart, stereochemical rules of recognition

肽核酸的分子生物学效应及应用

李晓旭 张亮仁 张礼和

(北京医科大学天然及仿生药物国家重点实验室, 北京 100083)

陈耀祖

(兰州大学化学系, 兰州 730000)

摘要 肽核酸 (PNA) 是一类以酰胺键连接骨架替代核酸中核糖磷酸二酯键骨架构成的核酸类似物, 其中 N-乙基甘氨酸骨架 PNA 与核酸链以 Watson-Crick 碱基配对形式稳定互补结合, 具有广泛生物学效应, 包括调节 DNA 识别蛋白质的功能以及调节转录和翻译. 在分子生物学研究中 PNA 作为新的工具在多方面得到应用. 除它的 DNA (RNA) 结合特性外 PNA 在生物稳定性、细胞摄取、结构修饰多方面的研究进展显示出作为基因调节药物具有良好前景.

关键词 肽核酸, 生物学效应, 转录, 翻译, 基因调节药物

学科分类号 Q524

肽核酸 (peptide nucleic acid, PNA) 是以肽键为骨架的核苷酸类似物, 其中 Nielsen 和 Egholm 等提出的以 N-氨基乙基甘氨酸为单元, 通过甘氨酸 α -N 酰甲基与碱基相连的 PNA (图 1), 其基本理化特性. 分子生物学效应及应用研究十分活跃, 显示了良好的应用前景.

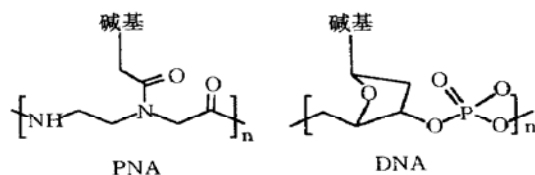


图 1 PNA 和 DNA 结构

收稿日期: 1996-10-30, 修回日期: 1997-01-24