

锌指类基因调控蛋白^{*}

——生物无机化学和分子生物学发展的新领域

黄仲贤 顾伟强

(复旦大学化学系, 上海 200433)

胡红雨

(中国科学院上海生物化学研究所, 上海 200031)

摘要 生命必需元素锌, 除了以锌酶的形式, 参与生物体的各类代谢过程外, 近年来发现, 锌还以各种锌蛋白的结构方式, 包括锌指、锌纽、锌带和锌簇等, 参与生物体的基因转录、复制及蛋白质的合成等各种基因调节和控制过程。联想到金属硫蛋白在锌的体内平衡中所扮演的角色, 很有可能锌是通过金属硫蛋白和锌指类蛋白的逐级调控, 成为生物体生长发育的调控中心。癌基因和人免疫缺陷病毒(HIV) 的许多调节蛋白也具有锌指类结构, 这给癌症和爱滋病的治疗提供了新思路。

关键词 锌, 锌指蛋白, 基因调控

早在 1869 年, 锌就被发现是真核生物所必需的微量元素。1940 年, Kelin 和 Mann 首次发现了含锌酶——碳酸酐酶, 人们开始从本质上认识到锌在生命体中的作用。至今, 发现的含锌酶已达 300 多种^[1]。近年来, 发现了许多直接参与 DNA 转录和复制的核蛋白是含锌的蛋白质。锌指蛋白(zinc finger proteins) 是一类最重要最典型的 DNA 结合蛋白。因而, 锌在生命体中不仅存在于各种酶中, 参与机体新陈代谢, 而且大量存在于锌指类蛋白中, 控制着细胞的生长、分裂和机体的生长发育。在高等动物体内, 锌还以 Zn_n-MT 这一特殊的形式存在, 维系着生物体内的内稳态平衡(homeostasis), 以至著名的生物无机化学家 R. J. P. Williams^[2]指出, 锌(与钙一起)很有可能在真核生物中起着中心调控的作用。

1983 年, Hanas 等在研究爪蟾卵母细胞时, 发现了第一个控制基因转录的锌指蛋白——转录因子 TFIIIA^[3,4]。此后的 10 年中, 已发现在病毒、酵母、哺乳动物等各种控制复制和转录的 500 多种蛋白中具有这种锌指类结构顺序^[1]。而且还有许多不同于 TFIIIA 结构的锌指类蛋白; 除了能与 DNA 结合外, 还能与 RNA 结合。目前有四类结构不同的 DNA 结合

蛋白控制着 DNA 的转录: 螺旋-转折-螺旋(helix-turn-helix), 亮氨酸拉链(leucine zipper), 锌指蛋白(zinc finger) 和 β 带(β-ribbon)^[5,6], 而锌指蛋白是真核细胞中最普遍存在的一类 DNA 结合蛋白。突出的是只有锌指结构才具有选择结合的能力, 而其它金属如 Fe、Cu、Mn、Ni 等均不能代替锌的作用。

锌指类蛋白结构域除了上述 TFIIIA 中 Cys 和 His 通过 Zn-S 和 Zn-N 与中心金属锌配位外, 也有以激素受体为代表的 Cys₂Cys₂ 配位。随着对锌指类基因调控蛋白的深入研究以及 X 射线衍射和 2D-NMR 结构分析技术的发展, 已发现有 4 种典型的锌指类结构域的结构模式, 根据锌的配位和结构域的性质可分为锌指、锌纽(zinc twist)、锌带(zinc ribbon) 和锌簇(zinc cluster)(见图 1)。锌纽结构主要在糖皮质激素和雌激素受体蛋白中^[7]。这些受体蛋白通常有 3 个结构域: 第一个结构域结合特定的激素如糖皮质激素或雌激素, 第二个结构域结合 DNA, 第三个结构域则与 RNA 聚合酶结合。糖皮质激素的 440—525 位氨基酸残基中富含 Cys, 但没有 His, 它必须结合两个锌离

* 国家自然科学基金资助课题。

收稿日期: 1994-05-11, 修回日期: 1994-06-02

子后形成扭状键合 DNA 的结构。

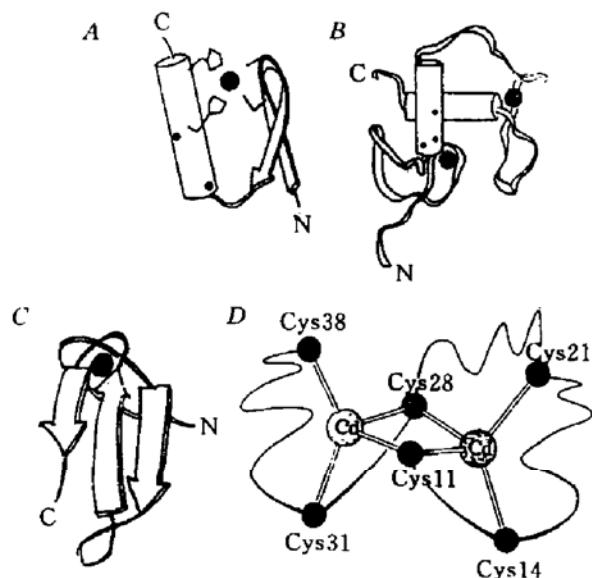


图 1 四种常见锌指类蛋白的结构

A: 锌指 B: 锌紐 C: 锌带 D: 锌簇。

含有 881 个残基的 GAL4 是酵母的转录因子蛋白^[8]。它通过 N 端的 62 个氨基酸残基中的 6 个 Cys 与两个锌离子结合, 形成 Zn₂Cys₆ 锌硫簇结构, 与 DNA 上游激活片段键合, 启动乳糖代谢基因。它的双核锌硫簇结构与 MT 中的 Zn₃Cys₉、Zn₄Cys₁₁ 金属硫簇很相似, 但它们却分属于两类功能的蛋白: 前者结合 DNA, 为基因调控蛋白; 而后者主要用于调节锌离子的内稳态平衡。人大脑内含有类似

MT 的蛋白, 称为生长抑制因子 (growth inhibitor factor, GIF), 其氨基酸序列几乎与 MT 同源, 并可结合 3 个锌和 4 个铜, 然而其生物功能与 MT 相差很大, 可能与神经细胞的生长有关。一个饶有兴趣的问题是, 生物体中的这些锌簇结构相似, 但为何功能各异。很有可能这是生物进化中构象块或结构域基因的选择和拼接造成的结果。

来自噬菌体 T4 的基因 32 蛋白 (g32p)^[9], 含一个锌, 以 Cys₃His 的形式配位, 它除了能与单链 DNA 结合, 影响 DNA 复制、重组、修复外, 还能与 mRNA 结合, 抑制 g32p 本身 mRNA 的转译。

因而, 在锌指类蛋白的配位和结构特征来看, 从起始的 Cys₂His₂, Cys₂Cys₂, 到 Cys₃His 配位, 从指状结构到簇结构, 锌指蛋白已成为一类含锌的核酸结合蛋白的总称。4 种结构域的特征氨基酸序列、中心金属配位性质、结构域构象以及与 DNA 作用的方式有所不同 (见表 1), 但它们的共同特征是由螺旋 (或折叠)-环-螺旋 (或折叠) 所构成的超二级结构, 并络合中心金属离子形成的空间构象。锌指类结构域一般较其它蛋白结构域小, 约 30—60 个氨基酸, 正是锌的配位才可能使维持这样小的空间构象得以稳定。作者认为这种小结构域很可

表 1 锌指类蛋白的类型及结构特点

类型	锌和配体	结构特点	Zn—Zn 距离 / Å	识别 DNA 序列	例子	序列
锌指	ZnCys ₂ His ₂ 1Zn/位点	反平行折叠, 转折, 螺旋指状突起	27	GCG · · · · GCG 5' -GCGTGGGCG-3'	TFⅢA Zif268	CX ₄ C-X ₁₂ -H-X ₄ -H
锌紐	(ZnCys ₄) ₂ 2Zn/位点	螺旋-环-螺旋	12—13		GR	CXXC-X ₇ -H-X ₅ -CXXC- X ₁₁ -H-X ₃ C-X ₅ -C-X ₉ - CXXC-X ₄ C
锌带	ZnCys ₄ 1Zn/位点	三条反平行 β 折叠			TFⅢS	CX ₂ C-X ₂₄ -CX ₂ C-
锌簇	Zn ₂ Cys ₆ 2Zn/位点	螺旋-伸展链-环-螺旋 -伸展链	3	5' -CGGN ₁₁ CCG-3'	GAL4	CXXC-X ₆ -C-X ₆ -CXXC- X ₆ -C
其它	ZnCys ₃ His	Zn-S (N) 配合物 稳定蛋白质二聚体		结合单链 DNA	g32p	CXXXH-X ₅ -CXXC-

能有利于其与 DNA 的凹沟的嵌合，调节 DNA.

锌指类蛋白既然能够选择性地结合于核酸，那么，它们与核酸结合的三维结构必定是人们最感兴趣的^[10]. 到目前为止，已测定了 3 个锌指蛋白结合区肽段与核酸特定序列复合物的晶体结构，它们分别是 Zif268-DNA, GR-DNA 和 GAL4-DNA 复合物，正好分属于 Cys₂His₂、Cys₂Cys₂ 和 簇状的 Zn₂Cys₂ 3 种不同结构类型的锌指蛋白^[11].

Zif268 是小鼠的 DNA 转录因子，有 3 个类似 TFIIIA 中 Cys₂His₂ 结构的锌指。1991 年，Pavletich 和 Pabo^[12] 获得了分辨率为 2.1 Å 的 Zif268-DNA 复合物的单晶结构。在 Zif268 与其靶 DNA 5'-GCGTGGGCG-3' 作用时，3 个锌指依次地沿着 DNA 的大沟，以排列成类似 C 字形的夹子环绕于 DNA 连续的 9 个碱基对上。每个锌指结构域中的 α 螺旋与 3 个碱基对相互作用。α 螺旋的轴线大致与 DNA 的大沟平行。β 折叠与碱基无相互作用，而其 C 端与 DNA 一条链的糖-磷酸主骨架链相接触。3 个锌指与 DNA 间形成了 11 个氢键，而所有这些作用都发生在上述富含 G 碱基片段上；在每个锌指中，位于 α 螺旋前的一个残基与 α 螺旋中 2、3、6 位的 4 个残基在锌指与 DNA 的作用中起着至关重要的作用（在锌指 1、3 上，它们为 Arg、Asp、Glu、Arg，在锌指 2 上为 Arg、Asp、His、Thr）。每个锌指与 DNA 有两个主要的作用点，3 个锌指中 α 螺旋前的一个残基，即 Arg18、Arg46、Arg74 与相应的 GCG、GGT、GCG 三组碱基对中的第一个鸟嘌呤上的 N7 和 O6 原子生成氢键，而在每个 α 螺旋中，位置同样是保守的天冬氨酸（即 Asp20、Asp48、Asp76）用它们的羧酸基与精氨酸胍基中的 N_ε 和 N_η 生成氢键和盐键，以辅佐稳定 Arg-G 之间的作用。此外，除了 β 折叠的 C 端以外，各个锌指中与锌配位的组氨酸（即 His25、His53、His81），除了通过咪唑环上的 N3 与锌配位以外，它们的 N1 分别与 5'-DNA 上的磷酸基结合。图 2 显示了这些相互作用的细节。从这一

单晶结构可以看出，在锌指上的 α 螺旋和 β 折叠都与 DNA 的骨架主链上的磷酸基团氧原子相结合，以帮助稳定锌指与 DNA 的键合。

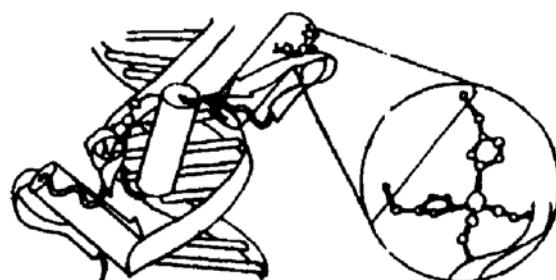


图 2 Zif268 三个锌指肽段与 11-bp DNA 片段相互作用的结构图

具有 Cys₂Cys₂ 型锌指结构的糖皮质激素受体 (glucocorticoid receptor, GR) 是一个含 777 氨基酸的 DNA 结合蛋白，其中 421—486 位氨基酸对应的肽段是 DNA 的结合部位。1991 年，Luisi 等^[13] 获得了 86 个残基的 GR 片段（对应于 GR 本身的 DNA 结合区）与 DNA 的复合物单晶结构（图 3）。

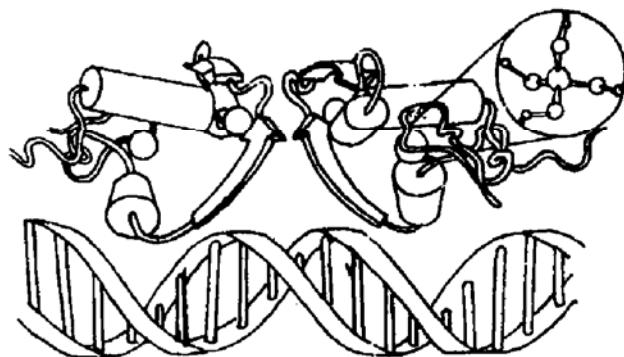


图 3 糖皮质激素 (GR) 二聚体与 19bp DNA 片段相互作用的结构图

在该结构图中，GR 以二聚体的方式二重对称地作用于 DNA. GR 分为 3 个部分：α 螺旋，连接区 (linker) 和二聚体区域。GR 与 DNA 的结合主要来自 α 螺旋表面的 Arg466、Val462 和 Lys463.

GAL4 是一个酵母 DNA 转录因子，含有 881 个氨基酸，其中 1—62 位为含有 Zn₂Cys₂ 簇的肽段，是 DNA 的结合部位。在 1992 年发

表的 GAL4-DNA 复合物单晶结构中^[14], 每个 GAL4 均含有 3 个特征的区域: 8—40 位为含金属的结构域, 41—49 为连接区, 50—64 为 α 螺旋, 用于形成二聚体。在 GAL4 与 DNA 作用时, 含金属的结构域对称地位于目标 DNA 两端的大沟中, 与 5'-CCG-3' 作用, 连接区与 DNA 仅发生有限的作用, 而二聚区则可以看作与 DNA 的小沟发生了作用(图 4)。

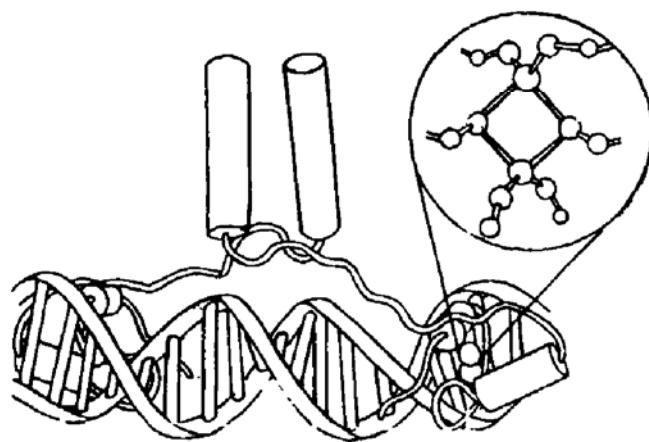


图 4 酵母 DNA 转录因子 GAL4 二聚体与 17bp DNA 片段相互作用

最近, Reece 等人^[15]研究了 GAL4、PUT3、PPR1 三个类似的酵母转录因子, 发现它们在与 DNA 相互作用中, 作为 DNA 结合区的金属硫簇可以相互替换而不丧失对目标 DNA 的结合能力, 而连接区(linker)的长度则决定了二聚体中两个金属结构域间的距离, 从而在与 CGGN_xCCG 类型的目标 DNA 结合时, 长度 x 也各不相同。

锌指类蛋白的发现有重要意义, 因为锌在人体中是含量仅次于铁的微量必需生命元素, 在人体中主要以生物大分子的方式存在。锌一直被认为作为酶的重要辅助因子, 参与生物体小分子的利用和储运、肽链的解合、糖的代谢等一系列生物化学过程, 而发挥其重要作用。然而, 流行病学和病理学的临床研究表明, 锌与生物体的生长发育有着十分紧密的联系, 以前总把它归之于一些含锌酶如 DNA 聚合酶、RNA 聚合酶、氨基 RNA 合成酶等形式

参与了细胞发育, 包括 DNA 复制、转录, 调控蛋白的合成等一系列过程。然而, 锌指蛋白的发现, 使人们醒悟到锌蛋白直接参与了基因表达的调控, 直接影响着生物的生长发育。锌指类蛋白是基因调控蛋白中最重要的一类含锌金属蛋白, 在真核细胞的基因调控中起着关键的作用。由 cDNA 序列推测, 很多“gap”类基因产物, 如果蝇的 Kr 蛋白 2, 哺乳动物细胞 SP1 和人的 ZFY、ZFX 性别决定因子, 含有锌指类同源的氨基酸序列 Cys₂His₂ 盒。这些 Cys、His 可能是锌的结合位点, 其发挥功能, 结合 DNA 或 RNA, 与其形成锌指类结构有关。随着基因序列分析的发展, 现已发现愈来愈多的基因编码蛋白含有(Cys、His)盒, 这些蛋白很可能存在着锌指类结构域。可以认为, 锌在生物体中是最重要的调节离子, 锌指类结构域是基因调控最广泛的模式^[16]。目前探明的锌指类蛋白仅是基因调节蛋白的“端倪”。因此, 可以预言, 愈来愈多的事实会确定锌在生长发育和基因调控中的中心地位。含锌基因调控蛋白的结构和功能也将是分子生物学和发育生物学的研究热点。

一个耐人寻味的问题是, 锌缺乏会引起动物生长迟缓、发育受阻, 但过量的锌离子引起动物中毒的后果却从未见报道, 也即动物体内似乎存在着一个相当有效的机制, 消除过量的锌离子。遗憾的是目前人们对生物体内锌离子的体内平衡机制(homeostasis)仍然一无所知, 特别是生物体是否存在一个锌库(zinc pool)或相应的锌储存蛋白尚无肯定的答案。有些作者(包括本实验室)指出, 金属硫蛋白至少部分地承担着生物体内锌离子调节的作用。它可以作为许多锌酶的锌供体(MT 即使在部分镉中毒的情况下, 仍然具备这种供锌能力^[17]; apo-MT 可以夺取锌指蛋白中的锌, 使锌指结构不能形成, 从而失去选择结合核酸的能力^[18](但 Zn₋MT 提供锌离子, 促使锌指蛋白形成锌指结构的证据尚未得到)。很有可能, 机体通过 Zn-MT/apo-MT 来调节体内锌离子, 从而调制各种锌酶和锌蛋白的活性。

从结构与功能上来看，锌指蛋白通过多种方式来创造它们各不相同的特异的核酸选择性的位点结合特性：用不同的配位方式 Cys₂His₂、Cys₃His、Cys₂Cys₂……；不同的簇结构(finger、twist、ribbon、cluster)；配位位点之间的不同的氨基酸的数目(即不同的 spacer)；与核酸结合的不同指数；不同的锌指聚合度(单个锌指，双聚锌指等)；锌指上 α 螺旋和 β 折叠的不同组合；以及锌指与核酸主骨架上糖-磷酸酯键的结合方式等，来生成结合于不同用途、不同核酸的特异选择性。那么，当从分子水平上认清了它们的相互作用及特异选择性的本质后，人们完全可以从分子识别与分子设计出发，来调制基因，来合成各种特异的基因工程药物分子，或者制造各种承担特殊功能的特异的人工蛋白分子。例如，现在已经知道，许多癌症基因，原癌基因和肿瘤抑制基因编码的 DNA 结合蛋白，也具有典型的锌指结构序列^[19]。推测存在着锌指结构域。例如真核细胞转录延长因子(transcriptional elongation factor, TFIIS)，可以与 RNA 聚合酶以及单链或双链 DNA 结合，组成三元复合物，它们很可能控制着原癌基因和病毒，包括 HIV 病毒的基因表达^[20]。研究表明，一种称为 Wilm's 肿瘤的肾脏肿瘤，起因于一种蛋白的遗传突变，从而干扰了它的锌指结合区与 DNA 的正确键合^[3]。值得注意的是，在这些癌基因指状结构中的锌离子，可以为 Cd²⁺、Hg²⁺ 和 Co²⁺ 所替代，因此有可能以这些金属为基础，发展出一类防癌和治癌的新金属络合物药剂。对于 Aids 病毒(human immunodeficiency virus, HIV)，也有不少研究证实，它们与锌指类蛋白有关。例如，逆转录病毒 HIV-1 的核壳体(nucleocapsid, NC)蛋白能同病毒核中的 RNA 基因组结合。体外试验表明，它们可以促使引物的 tRNA 退火为病毒的 RNA。研究指出，NC 蛋白的 N 端的 18 个氨基酸结合 1 个锌，形成锌指类结构后，可以选择性地键合于核酸^[21]。体外模拟研究 Zn (HIV-F1) 与 d (ACGCC) 寡聚核苷酸的结合，其结合常数为 $K_{ass} = 2 \times 10^5 \text{ (mol/L)}^{-1}$ 。而且

在它的 N 端的锌指结构-C-X₂-C-X₄-H-X₄-C 中(X 为各种氨基酸)，X 必须有疏水氨基酸和碱性氨基酸的存在，它们直接参与对病毒基因组的识别。如果用特定的抗病毒试剂把锌驱走，就将使病毒无法结合于核酸。这对寻找新的抗 HIV 病毒药物提供了新思路^[22]。此外，HIV 病毒的 tat 蛋白，以及在病毒 DNA 复制中所需要的单纯疱疹-1 病毒蛋白 HSV-1 也有可能含有锌指结构域。锌指类基因调控蛋白的发现，以及结构和作用机理的阐明，将进一步了解癌基因的活动和病毒的生理活动的本质，正象以上所述，这很可能为寻找有效的药物治疗提供一条途径。

锌在哺乳动物脑中的含量相当丰富，从神经分子生物学的水平上看，锌指类结构域也可能存在于神经受体中，并对神经细胞的基因表达进行调控。前已提及，在人脑中存在着一种类金属硫蛋白，即生长抑制因子^[23]。它可以抑制神经细胞营养因子的活性，从而避免神经细胞过早死亡。它比肝肾的金属硫蛋白多 7 个氨基酸，其中第 6 位插入了一个 Ser，在第 55 位插入了以 Glu、Asp 为主的 6 个氨基酸。它同样有 7 个金属离子，只不过是 Cu₄Zn₃，也形成金属硫簇。有意思的是 GIF 在老年痴呆症(Alzheimer disease, AD)患者的大脑中含量特别低，因此，是否是由于锌不足或者其它金属离子(例如 Al³⁺，也可以是四面体配位)的竞争，使老年痴呆症患者的大脑中不能生成足够的 GIF，或者是金属置换的 GIF，失去了抑制神经营养因子活性的能力。可以一提的是，在老年痴呆症病人的血浆中胸腺素(thymulin)，含量并不降低，但结合于其它蛋白载体的活性的胸腺素特别低。用锌可以激活胸腺素，这似乎与糖皮质激素和雌激素的结合情况有类似之处^[24]。

锌指类蛋白及其结构的揭示，似乎已经把以锌为主的基因调控蛋白的面纱在渐渐地拉开，推上了舞台的中心。其它金属的调节蛋白如汞调节蛋白(MerR)和铁调节因子(IRF)以及 Cup2 (ACE1) 等很有可能都是基因调节水

平上的演化。即这些金属离子通过篡改锌指类结构，在基因水平上调控自身的结合蛋白^[25]。显然，与基因表达和基因调控密切有关的锌指类蛋白的确定、结构分析以及与DNA作用（基因调控）的基础研究已成为分子生物学和发育生物学研究的新兴领域。同时，在基因调控水平上，将人工设计的锌指结构域类似物，应用与基因治疗，具有广阔前景。

参 考 文 献

- 1 Vallee B L, Falchuk K H. Physiol Rev, 1993; **73**: 79
- 2 Williams R J P. Biochem Society Transection, 1990; **18**: 689
- 3 Rhodes D, Klug A. Sci Amer, 1993; Feb: 32
- 4 胡红雨, 鲁子贤. 生物化学与生物物理进展, 1992; **19**: 250
- 5 Qian X-Q, Jeon C J, Yoon H S et al. Nature, 1993; **365**: 277
- 6 Harrison S C. Nature, 1993; **353**: 715
- 7 Freedman L P, Luisi B F, Korszum Z R et al. Nature, 1988; **334**: 543
- 8 Pan T, Coleman J E. Proc Natl Acad Sci USA, 1989; **86**: 3145
- 9 Gibson T J, Postma J P, Brown R S et al. Protein Eng, 1988; **2**: 209
- 10 Desjalais J R, Berg J M. Proc Natl Acad Sci USA, 1992; **89**: 7345
- 11 Berg J M. Curr Opin Struct Biol, 1993; **3**: 11
- 12 Pavletich N P, Pabo C O. Science, 1991; **252**: 809
- 13 Luisi B F, Xu W X, Otwinowski Z O et al. Nature, 1991; **352**: 497
- 14 Marmorstein R, Carey M, Harrison S C. Nature, 1992; **356**: 408
- 15 Reece R J, Patashne M. Science, 1993; **261**: 909
- 16 Berg J M. Annu Rev Biophys Biophys Chem, 1990; **19**: 405
- 17 Huang Z X, Hu H Y, Gu W Q et al. Chinese J Chem, 1993; **11**: 246
- 18 Zeng J, Vallee B L, Kagi J H R. Proc Natl Acad Sci USA, 1991; **88**: 9984
- 19 Sunderman F W Jr, Barber A M. Annu Clin Lab Sci, 1988; **18**: 267
- 20 Bentley D L, Groudine M. Nature, 1986; **321**: 702
- 21 Delahunty M D, South T L, Summers M F et al. Biochemistry, 1992; **31**: 5151

- 22 Summers M F, South T L, Lee B et al. J Inorg Biochem, 1993; **51**: 60
- 23 Uchida Y, Takio K, Titani K et al. Neuron, 1991; **7**: 337
- 24 Davis L J. Adv Biosci, 1993; **87**: 279
- 25 O'Halloran T V. Science, 1993; **261**: 715

Gene Regulation Proteins Containing Zinc Finger Structure: New Developments on Bioinorganic Chemistry and Molecular Biology. Huang Zhongxian, Gu Weiqiang (*Department of Chemistry, Fudan University, Shanghai 200433, China*); Hu Hongyu (*Institute of Biochemistry, Academia Sinica, Shanghai 200031, China*).

Abstract It is well known that zinc is an essential element for life. It takes part in a variety of important metabolisms in biological systems in the form of zinc containing metalloenzymes. In the past decade, it found that zinc is also involved in the gene transcription, duplication and protein synthesis processes as a regulator and mediator. Till now, a number of zinc proteins have been characterized such as zinc finger, zinc twist, zinc ribbon and zinc cluster. Among them the zinc-metallothionein, Zn_n-MT is one of the well known zinc-cluster protein which is suggested playing a crucial role in the zinc homeostasis. It is very possible that zinc occupies the central part in controlling the growth and development of organisms by means of successive regulation of MT and zinc finger proteins. In addition, it was found that some of regulation protein binding to oncogene and human immunodeficiency virus also contains zinc finger amino acid sequence, that will open a new field to fight against cancer and aids diseases.

Key words zinc, zinc protein, zinc finger protein, gene regulation