

鉴定水稻中一个新的专一结合 GCC元件的 AP2/EREBP 族转录因子 *

陈 峰¹⁾ 余军平²⁾ 方晓红^{2)**} 张贵友^{1)**}

(¹⁾清华大学生物科学与技术系, 北京 100084; ²⁾中国科学院化学研究所, 北京 100080)

摘要 TSH1, 是通过搜寻 GenBank 的 EST 库而获得的一个来源于水稻的含 AP2/EREBP 保守结构域的蛋白质. 为了详细分析 TSH1 蛋白与其 DNA 顺式元件的结合特性, 首先采用传统的凝胶阻滞实验和酵母单杂交技术, 证实 TSH1 在体内和体外均专一性地结合于 GCC 元件, 然后利用原子力显微镜技术精确测量了 TSH1 蛋白与 GCC 元件在单分子水平的相互作用力. 结果表明, GST-TSH1 与 DRE 元件没有特异性的结合, 而 GST-TSH1 与 GCC 元件结合力的大小为 (83.9 ± 2.2) pN, 这种特异性的结合可以被加入的游离 TSH1 蛋白明显降低. GST 蛋白和突变 GCC 元件作为负对照显示出与 GCC 元件无特异性作用力. 以上结果充分证明, TSH1 是专一性地与 GCC 元件相作用的转录因子, 而且原子力显微镜对于检测转录因子与 DNA 相互作用时单碱基的突变十分灵敏. 通过比较几种评估转录因子与 DNA 顺式元件结合特异性的方法, 阐述了原子力显微镜技术的特点及优越性.

关键词 TSH1, AP2/EREBP, 转录因子, 原子力显微镜

学科分类号 Q71

伴随基因组研究的蓬勃发展, 在 GenBank 中登录的基因数目急剧增加, 因此重视这一发展趋势, 充分利用网上资源, 将使筛选新基因及鉴定基因功能的工作更方便快捷^[1,2]. 目前这方面的研究很多集中在对转录因子基因的克隆与功能鉴定上. 真核生物中各种转录因子在转录水平的调控是基因表达调控的重要方式之一. 转录因子一般都含有 DNA 结合区与转录调控区, 通过与特异性的 DNA 元件的相互作用来调控生物体内相关靶基因在特定条件(如环境胁迫)、特定组织中定时定量地表达. 要阐明细胞内各种信号转导与调控基因表达的机制, 鉴定各种调控基因表达的 DNA 元件和转录因子至关重要.

AP2/EREBP 转录因子家族的第一个成员 APETALA2 (简称 AP2) 由 Jofuku 等^[3]从拟南芥 (*Arabidopsis thaliana*) 分离出来, 它由 432 个氨基酸残基组成, 分子质量约为 48 ku, 调控与花和种子发育相关的基因表达. 序列分析表明, 它的 DNA 结合区中含有 2 个由 68 个氨基酸残基组成的 AP2 结构域 (AP2-R1 和 AP2-R2), 二者之间的氨基酸同源性为 69%, 均包含有可形成双亲性 α 螺旋的区

段, 这个 α 螺旋结构可能介导蛋白质间的相互作用^[4]. 随后 Ohme 等^[5]从烟草 (*Nicotina tabacum*) 中克隆了 4 个与顺式作用元件 GCC (核心序列为 GCCGCC) 特异结合的转录因子, 它们都含有与 APETALA2 转录因子的 AP2 结构域同源性高的 DNA 结合区. 由于 GCC 元件普遍存在于受乙烯诱导基因的启动子区域, 这些调控乙烯诱导基因表达的转录因子分别被命名为 EREBP1~4 (ethylene responsive element binding proteins). 不久, Liu 等^[6]利用 *rd29A* 基因启动子的 DRE (dehydration responsive element) 顺式作用元件通过酵母单杂交方法 (yeast one-hybrid method), 分别从低温和干旱处理的拟南芥 cDNA 文库中, 克隆了 3 个与低温胁迫耐性相关的转录因子 DREB1A~C 和 2 个与干旱胁迫反应相关的转录因子 DREB2A~B. 拟南芥

*国家自然科学基金资助项目(39970166, 20225516 和 10334100).

** 通讯联系人.

张贵友. Tel: 010-62772254, Fax: 010-62772678

E-mail: zhang-gy@mail.tsinghua.edu.cn

方晓红. Tel: 010-62650024, E-mail: xfang@iccas.ac.cn

收稿日期: 2006-04-27, 接受日期: 2006-05-19

DREB转录因子与烟草 EREBP 转录因子一样, 都含一个由 58 个左右氨基酸残基组成、与 AP2 结构域同源性高的 DNA 结合区域。

随着对转录因子重要功能研究的日益重视, 人们相继从拟南芥、烟草、番茄 (*Lycopersicon esculentum*)、水稻 (*Oryza sativa*)、玉米 (*Zea mays*) 等多种植物中分离鉴定了许多含有 AP2/EREBP 结构域的转录因子。这些转录因子具有几个共同特征: a. DNA 结合域均为保守的 AP2/EREBP 结构域, 该结构域以外的氨基酸序列较少有同源性; b. 转录因子的 N 端常含有由短的碱性氨基酸残基组成的核定位区, 如 AP2 转录因子含有由 10 个氨基酸残基组成的核定位区; c. 含有转录调控区, 如 AP2 转录因子 N 端富含丝氨酸的酸性区, 以及 EREBP1~4 和 DREB 转录因子 C 端的酸性氨基酸区域。

根据所含 AP2/EREBP 结构域的数目, 可将 AP2/EREBP 转录因子分为 AP2 和 EREBP 2 个亚族^[4]。AP2 亚族转录因子含有 2 个 AP2/EREBP 结构域 (AP2-R1 和 AP2-R2), 2 个结构域之间的连接区也相对保守, 其中包括 AP2、ANT、RAP2.7、G115、IDS1、ZM、PHAP2A、PHAP2B 转录因子等^[7], 它们主要调控植物的生长发育。这类转录因子中目前仅有 ANT 已用实验证实其与特定的 DNA 序列结合^[8], 结合序列为 gCAC(A/G)N(A/T)TcCC(a/g)ANG(c/t), ANT 中任何一个单独的 AP2/EREBP 结构域均不能独立与该 DNA 序列结合。EREBP 亚族转录因子仅含一个 AP2/EREBP 结构域, 包括 TINY、DREB1、DREB2、ABI4、RAV1/2、EREBP、Pti4/5/6、EBF1、AtEBP 和 RAPs 等^[7], 这类转录因子调节植物对激素(乙烯)、病原、低温、干旱及高盐等的分子应答。根据识别及结合的 DNA 顺式作用元件不同, EREBP 亚族转录因子的研究又主要集中在 A、B 两类上。A 类转录因子如 DREB1A~C 和 DREB2A~B 识别 DRE 元件(核心序列为 A/GCCGAC)^[6]; B 类转录因子如 EREBP1~4 等的靶序列为 GCC 元件(GCCGCC)^[4]。

利用计算机分析, 从 GenBank 数据库发现大量含 AP2/EREBP 结构域的蛋白质, 它们广泛分布于多种植物。仅拟南芥中就已发现许多 AP2/EREBP 家族转录因子, 编码这些转录因子的基因分布于拟南芥全部 5 条染色体上。AP2 位于染色体 VI 的 93.2 cM 处, 在分子标记 m214 与 g2486 之间。DREB1A、1B、1C 3 个基因依 1B-1A-1C 间隔顺次

排列于第 IV 染色体的 72.8 cM 位置, 与分子标记 m600 和 PG11 相邻, 而受其调控的 *cor78*、*cor15a*、*cor6.6*、*cor47* 等基因则位于其他染色体上。DREB2A 与 DREB2B 分别定位于第 V 和第 III 条染色体上。此外, RAP2.8、RAP2.3 和 TINY 分别定位于染色体 I、III、V 上, ABI4 定位于染色体 II 上, 靠近分子标记 nga168 处^[9~11]。目前拟南芥的基因组全序列已测完, 通过序列同源性比较, 预测其所含的 AP2/EREBP 基因总数近 150 个^[12]。

在重要经济作物中克隆和鉴定 AP2/EREBP 家族转录因子具有十分重要的意义。特别是 EREBP 亚族转录因子, 由于它们调节植物对激素(乙烯)、病原、低温、干旱及高盐等的分子应答, 因此对利用分子生物学技术来改良和提高植物的抗逆性具有重要的理论研究价值和广泛的应用前景。这是目前植物分子生物学研究领域引人注目的重要前沿课题之一。本研究通过同源搜索的办法, 从 GenBank 中快速分析了水稻中含 AP2/EREBP 保守结构域的 TSH1 蛋白, 并应用多种试验方法和手段, 详细鉴定了 TSH1 蛋白与其 DNA 顺式元件之间的相互作用。

1 材料与方法

1.1 材料来源

tsh1 基因 (即 AU056391, 插入 pBluescript SK⁺, 插入位点为 5' *Sal* I - 3' *Not* I) 由日本水稻基因组研究中心友好提供。

1.2 凝胶阻滞试验

DNA 元件的制备及标记: 用 *Hind* III 完全消化 4×GCC/SK⁺ 质粒, 在 4% 的 NuSieve GTG 琼脂糖 (Sigma 公司) 凝胶上回收约 200 bp 的 GCC 片段 (纯化回收试剂盒购于 QiaGen 公司)。以 50 ng 的 GCC 片段为模板, 65℃ 下变性 5 min, 置于冰上。加入 0.5 mmol/L dATP、0.5 mmol/L dTTP、0.5 mmol/L dGTP 各 1 μl, 10× Klenow Buffer 2 μl, α-³²PdCTP 2 μl, Klenow fragment 酶 1 μl, 双蒸水补足体积至 12 μl。37℃ 下反应 30~60 min。然后经 G-250 凝胶 (Sigma 公司) 过柱纯化, 纯化后加入 75 μl TE 缓冲液, 在 Quick-count 仪器 (Bioscan 公司) 上测量放射强度。

TSH1 蛋白的提取: 通过构建 *tsh1* 基因编码框全长到 pGEX-4T-1 载体上, 转化到大肠杆菌 BL21 菌株里, 诱导表达后利用亲和层析的方法获得纯化蛋白质。

元件与蛋白质的结合反应体系为 3 μl 10×

结合缓冲液, 1.5 μl 80%甘油, 20 g/L BSA, 3 μl 2 g/L poly[dl-dC], 10 μl 蛋白质溶液, 9 μl dH₂O, 2 μl 标记后的元件(2000 DPM/ μl), 25℃下静置 30 min(10×结合缓冲液: 0.25 mol/L HEPES-KOH pH=7.9, 0.5 mol/L KCl, 5 mmol/L EDTA). 聚丙烯酰胺凝胶电泳在 150 V 恒压下, 1.5 h. 然后在 75℃下干燥凝胶 1 h, 压磷屏过夜.

1.3 酵母单杂交与 β -半乳糖苷酶活性分析

用 SD 基础酵母液体培养基活化待测目的酵母菌 1~2 次, 然后接种入相同的 3 ml SD 基础酵母液体培养基中, 培养细胞至对数生长的中后期. 取 1 ml 菌悬液于 600 nm 处测菌浓度 (以对应培养基作空白对照). 另外取 1 ml 菌悬液于 12 000 r/min 离心 30 s 收集菌体, 用 500 μl 的 Z 缓冲液重悬菌体, 在每个样品反应管中加入 15 μl SDS (1%) 和 30 μl 氯仿, 旋涡振荡 10~15 s, 在 28℃水浴中平衡 15 min. 加入 0.1 ml 4 g/L 的 ONPG (ONPG 被溶于 pH 7.0, 0.1 mol/L 磷酸钾缓冲液中), 在旋涡混合器上振荡 5~10 s, 以便充分混匀, 置 28℃水浴中, 并开始计时. 反应 1 h 后 (依溶液变黄的程度可适当改变反应时间), 加入 0.25 ml 1 mol/L Na₂CO₃ 终止反应. 12 000 r/min 离心细胞 2~3 min, 转移上清至新的 Eppendorf 管中, 测定上清的 A₄₂₀ 和 A₅₅₀, 同时以无菌体的样品作空白对照.

按下列公式计算 β -半乳糖苷酶的活性(U)

$$U = \frac{500 \times [(A_{420}) - (1.75 \times A_{550})]}{t \times v \times A_{600}}$$

公式中, t 为反应时间(min), v 为用作检测活性的菌悬液体积(ml), A_{600} 为检测开始时细胞的密度, A_{420} 为邻硝基苯酚的光吸收值和细胞碎片的光散射之和, A_{550} 为细胞碎片的光散射值.

1.4 原子力显微镜测量蛋白质与 DNA 相互作用的单分子力

1.5 cm×1.5 cm 的单晶硅片经过清洗后, 被用作固定 DNA 序列的基底. 经过预处理的基底立即放入 1.0% APTES (3-氨基-三乙氧基硅烷) 的甲苯溶液中 2 h 后接上氨基. 用溶剂冲洗基底完全去除未结合在基底上的硅烷. 将经过硅烷化的硅片放入 10%戊二醛 (PBS 缓冲液为溶剂) 中反应 1 h. 以戊二醛作为交联剂, 一端可与硅烷化形成的氨基反应, 一端与 DNA 分子中修饰后带上的氨基反应, 以固定 DNA 分子. 再用 PBS 缓冲液清洗.

双链 DNA 由 SBS Genetech 公司合成. 野生型 GCC 元件的序列: 5' NH₂-CGCAGACATA-

GCCGCCATTT 3'; 突变型 GCC 元件的序列: 5' NH₂-GATATACTACCGAGATGAGTTC 3'. 野生型 DRE 元件的序列: 5' NH₂-GATATACT-GCCGACATGAGTTC 3' (2 个 DNA 元件的序列都是以双链的正义链形式给出, 下划线部分是元件与蛋白质相互作用的核心序列). 合成时按照实验的要求在双链中的一条链 5'端加上—NH₂. 然后将引物配制成 5.0×10⁻⁷ mol/L Oligo 1, 5.0×10⁻⁷ mol/L Oligo 2, 5.0×10⁻² mol/L NaCl, 1.0×10⁻⁴ mol/L EDTA 溶液. 将经硅烷化处理过的基底浸入上述双链 DNA 溶液中. 4℃下 10 h, 将 DNA 固定在基底上. 利用双链 DNA 与基底反应是为确保 DNA 链中仅 5'端的一NH₂ (其中被修饰的那一条链) 与基底上的戊二醛发生交联反应而被固定在基底上. 最后用 PBS 缓冲液冲洗, 去除未被固定的 DNA. 在 4℃ PBS 缓冲液中保存待用.

蛋白质在针尖上的固定, 是将经过清洗和硅烷化处理的针尖浸入 5 mg/L 的待测蛋白质溶液中 (溶于 PBS 缓冲液), 室温下固定 0.5 h. 反应完成后用 PBS 缓冲液冲洗针尖. 为了检测蛋白质在针尖上的固定效果及固定的数量, 在相同条件下同样处理一块单晶硅片, 然后用 Nano Scope IIIa AFM (DI) 扫描检测硅片上蛋白质固定的情况. 修饰好的针尖在 4℃下 PBS 缓冲液中保存待用.

利用 A Nano Scope IV AFM (DI, Santa Barbara, CA) 来测定蛋白质与 DNA 序列间的相互作用力, 测量中以 PBS 溶液作为缓冲体系, 参见 Jiang 等方法^[13]. 针尖微悬臂的弹性系数按照热波动方法进行校正. 作用力测定时的基底加载速率为 4.8×10⁴ pN/s^[14].

2 结 果

2.1 生物信息学方法预测 TSH1 蛋白

通过同源序列分析查询 GenBank 的 EST (expressed sequence tag) 库, 经序列比较, 我们发现 33 个登录的水稻 cDNA 为同一个基因. 这些基因包括 (1) AU030886, (2) AU033077, (3) AU056391, (4) AU056392, (5) AU057089, (6) AU057090, (7) AU057298, (8) AU063844, (9) AU064640, (10) C19752, (11) C25883, (12) C72081, (13) C72300, (14) C97645, (15) C98138, (16) D15531, (17) D24976, (18) D39181, (19) D39201, (20) D39250, (21) D39720, (22) D39851, (23) D40526, (24)

D40562, (25) D40546, (26) D40553, (27) D41183, (28) D41268, (29) D41525, (30) D41572, (31) D47184, (32) D47296, (33) D47722. 通过对序列的比较和衔接, 我们推断其基因的完整序列如图 1 所示, 其推断的氨基酸序列中含

(a)

```

1  gaaagaaaagcatcacaaatctcacgcactgtctctcgttcgcgcaagcaecgtgctt
61  tctccgctttgcgagcaccatagcctagcccaccatgtgcggcgccaccctctccga
      M C G G A I L S D
121 cctcatcccgccgcccggcggtaccgcggcgacctctggctggagaagaccaagaa
      L I P P P R R V T A G D L W L E K T K K
181 gcagcagcagcagaagaagaagaacaaggcgcgaggaggtgccactgcgcaagagga
      Q Q Q Q K K K N K G A R R L P L R Q E E
241 ggaggtgatttcgagcgacgttcgaggatgcagggtgattccggcgagtgaggat
      E D D F E A D F E E F E V D S G E W E V
301 ggagtcgacgcccagcagggcgaagccgctcgcgcgccccggagcggtctcgtaaag
      E S D A D E A K P L A A P R S G F A K G
361 tggattgaaaacactactgttctgggtcgtgatggcgctcagcaaggtctcgtaaaag
      G L K N T T V A G A D G P A A R S A K R
421 gaagagaagaaccaattcagggtatccgacgagcgccatggggcaaatgggctcgga
      K R K N Q F R G I R Q R P W G K W A A E
481 aatcagagatcctcgaaggtgtccgcgtctggttggcaccctcaactctcctgagga
      I R D P R K G V R V W L G T F N S P E E
541 agctgccagagcttatgatctgaagcagcaggatcagggcaagaaggccaaggtcaa
      A A R A Y D A E A R R I R G K K A K V N
601 ttcccgatggggctccagtggtcttcagaggagtcagctgagccctcctcatgaa
      F P D G A P V A S Q R S H A E P S S M N
661 catgcctctttcagcagcgaagagaagcgccgctcatgtcagcaggaacaaaacat
      M P A F S I E E K P A V M S A G N K T M
721 gtacaacacaaatgcttatgctaccctgctgttgagtacaccttaccagagccattgt
      Y N T N A Y A Y P A V E Y T L Q E P F V
781 gcagattcagaatgtctcattgttctcgaatgaacgcgattgaggatacttctgtaa
      Q I Q N V S F V P A M N A I E D T F V N
841 cctgtctctgatcaaggagcgaactccttgggtgctcggactttagcaggagaatga
      L S S D Q G S N S F G C S D F S Q E N D
901 tatcaagaccctgacatacttccatgcttcgacgacacatgacaggtgttgatgactc
      I K T P D I T S M L A P T M T G V D D S
961 cgcattctccagaaacatgcagtgatgcaatgggtacctctgtgtggggaagtctag
      A F L Q N N A S D A M V P P V M G N A S
1021 cattgatcttgcctgacctggagccgtacatgaaattctgatcgatggtggttcggatga
      I D L A D L E P Y M K F L I D G M S D E
1081 gtgcattgacacccttctgagctctgatgattcagagatgtggccagtagcatggacct
      S I D T L L S S D G S Q D V A S S M D L
1141 ttggagcttcgatgacatgcccggtgcggcgagtttactgaggggtttgggtgtagc
      W S F D D M P V S A E F Y *
1201 aactgggtcgttatataaaggacaaatggaataaacattctggacatccaagaagcgg
1261 catgtgtctgtcggcgcttcttagttgcgtatataagctatgttagtattgtatgtg
1321 ctgtgtctagcttagatgctgaagtctcaagtactatttggcagtgaaactatctatctg
1381 taactgtatatagagctggaacaagtacttagcttctaccttatctgacttgctata
1441 gtggctgtgaaccttgtgagctgaactctgaagccaatgtttactatataatgtggttg
1501 gttttataaactctagtgttatttggaccctgtcaatggctcatgctatggctggggatta
1561 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaggcgccgccaccgcggtggag
1621 ctcacagct
  
```

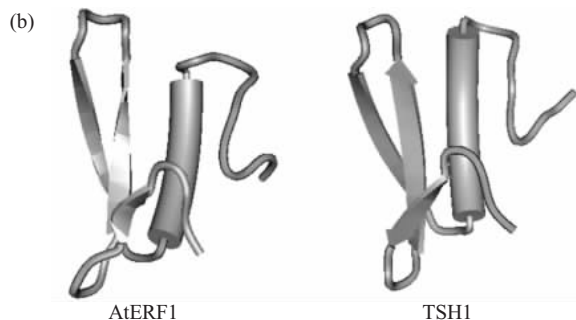


Fig. 1 Analysis of TSH1 sequence and structure

(a) cDNA sequence of *tsh1* and its coding protein. Underline shows the conserved AP2/EREBP domain, The hatching shows basic amino acid-rich region, both underline and hatching shows the acidic amino acid-rich region. (b) Comparing of conserved DNA-binding domain structure models of TSH1 and AtERF1.

AP2/EREBP 结构域. 从中我们发现基因 AU056391 的 5' 端完整, 为此我们从日本水稻基因组研究中心索取了该克隆, 经测序发现, 该基因是一个有完整开放阅读框的全长基因, 序列与图 1 显示的完全相同. 我们命名该基因为 *tsh1* (其编码的蛋白质为 TSH1), GenBank 登录号为 Af190770. 除了具有保守的与 DNA 元件结合的 AP2/EREBP 结构域以外, TSH1 还含有另外 2 个典型的转录因子特征: 其 N 端第 26 和第 38 位氨基酸残基之间的碱性氨基酸富集区, 可推断为核定位区; 其 N 端第 48 和第 75 位氨基酸残基之间的酸性氨基酸富集区, 为转录激活域.

关于 EREBP 亚族中的 B 类转录因子与 GCC 元件相互作用的研究较为深入, 利用核磁共振技术对 AtERF-1 转录因子的 AP2/EREBP 结构域进行分析, 发现其空间结构中含有一个由 3 条反向平行的 β 链组成的折叠片层结构和一个 α 螺旋, AP2/EREBP 结构域通过 β 折叠片层与 GCC 元件的大沟相结合, 其中有 7 个残基直接参与与顺式元件的结合. 计算机模拟预测, 其他一些 EREBP 亚族转录因子成员如 EREBP1~4、CBF1、TINY 等的 AP2/EREBP 结构域也可形成类似的空间结构. 在 B 类转录因子中与顺式元件直接结合的这 7 个残基绝对保守, 因此它们与 AtERF-1 可能具有相同的与 GCC 元件结合的模式. 通过序列同源性比较, 我们也发现 TSH1 在对应位置的这 7 个氨基酸残基也完全与 AtERF-1 相同. 而且结构预测显示 TSH1 的 AP2/EREBP 结构域也可形成类似的空间结构 (图 1b), 因此我们认为 TSH1 与 AtERF-1 可能具有相似的 DNA 结合模式.

2.2 竞争性凝胶阻滞试验

具有单个 AP2/EREBP 保守结构域的转录因子通过与不同 DNA 顺式元件 (GCC 或 DRE) 结合, 可以行使不同的生理功能. 为了鉴定 TSH1 究竟与哪种 DNA 元件特异性地结合, 我们构建了 GST-TSH1 融合蛋白表达质粒 (pGEX-4T-1 系统), 将其转化入大肠杆菌 BL21 菌株内, 诱导表达并体外纯化出了含有 AP2 结构域的 GST-TSH1 融合蛋白.

将 GST-TSH1 融合蛋白与同位素 ^{32}P 标记的 DNA 元件探针混合后进行聚丙烯酰胺凝胶电泳. 通过放射性自显影的结果, 来判断 DNA 与蛋白质是否因为特异性结合而产生滞后的条带. 如图 2 所示, GST-TSH1 融合蛋白与野生型 GCC 元件有结合,

而与野生型 DRE 元件不结合，而作为对照的 GST 蛋白与野生型元件以及 GST-TSH1 融合蛋白与突变型元件都没有结合. 由此可以看出 TSH1 蛋白与 DNA 元件结合的专一性.

图 2 显示的竞争性试验结果表明，当在体系中分别加入 10 倍和 100 倍非标记的野生型 GCC 元件时，DNA 与蛋白质结合产生的放射性自显影滞后条带随之减弱，暗示着非标记的探针对 ³²P 标记的探针与 TSH1 蛋白结合产生了竞争性的抑制，进一步说明了 GST-TSH1 蛋白与 GCC 元件结合的特异性.

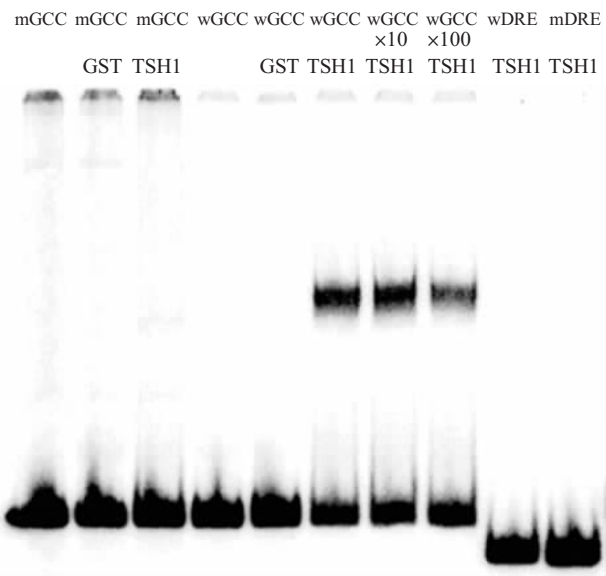


Fig. 2 Gel shift assay for TSH1 and specific DNA elements
On top of the electrophoresis lanes showed the specific DNA elements and proteins. ×10 and ×100 showed the folds of the competed DNA elements unlabeled ³²P in the samples. The bands in the middle and bottom indicate the DNA-protein complex and free DNA.

2.3 酵母单杂交激活试验

将水稻 *tsh1* 基因构建到 YepGAP 酵母表达载体中，分别转化到含有 DRE 元件和 GCC 元件的酵母单杂交双报告系统中 (含 pHISi-1 和 pLacZi 双报道子的酵母 YM4271 菌株). 然后在 SD/-His-Ura-Trp 液体培养基中于 30℃ 培养 22 h 后测菌体内的 β 半乳糖苷酶活力. 从图 3 可知，相对于空的 YepGAP 质粒，*tsh1* 基因转化入酵母后，使含 GCC 元件的酵母 β 半乳糖苷酶活力提高了约 24.5 倍，而对含 mGCC 或 DRE 元件的酵母 β 半乳糖苷酶活力无明显影响. YepGAP 酵母表达载体上不含有 GAL4 转录激活域，因此，若转化后的酵母细胞能显现出明

显的半乳糖苷酶活性，则表明该转录因子能与酵母报告系统中的元件结合并具有转录激活能力.

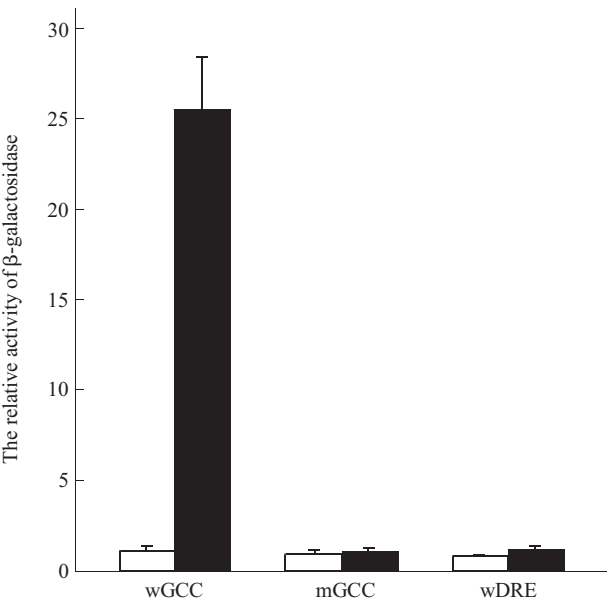


Fig. 3 Activation of the *lacZ* gene in yeast reporter cells by TSH1

β-galactosidase activity of the yeast reporter cells containing the wGCC or wDRE elements upstream of the *lacZ* gene was detected. Clones carrying *tsh1* and YEP vector are shown as black and white columns. The activities of the yeast reporter cells containing the YEP vector served as negative control. Yeast cells were grown on a selective liquid medium of SD/His-Ura-Trp- for 22 h. (n=6). □: YEP vector; ■: YEP vector+*tsh1*.

2.4 原子力显微镜测量 DNA 与蛋白质结合的单分子力

1986 年，Quate 和 Gerber 在扫描隧道显微镜 (STM) 的基础上发明了原子力显微镜 (atomic force measurement, AFM). AFM 的早期应用主要是对周期性的二维表面成像. 对 Si 表面的 (7 × 7) 重构，Au 和芳香化合物吸附物等的原子级成像，对组织化的有机薄膜分子间距的测量，和获得 DNA 的形貌图. 这种新型显微技术不仅可以获得高时空分辨率的形貌图，还可在纳米尺度上测量微观组分间的相互作用力. 而脱离真空及要求材料导电的限制，能在大气、液态和生理条件下直接测量，更使 AFM 在许多学科特别是生物学研究中崭露头角.

AFM 检测的基本原理如下：装在微悬臂上的很小的针尖在基底上做垂直移动，通过记录微悬臂的形变量得出对应的力值随针尖 - 表面距离变化的曲线，即力 - 距离曲线，来研究分子间 / 分子内特异的相互作用力及其性质.

图 4 显示了利用原子力显微镜对 GST-TSH1 蛋白与 GCC 元件 (核心序列 GCCGCC) 相互作用力的分析. 由于原子力显微镜测定单分子作用力时存在外界扰动, 如溶液扰动, 温度扰动等, 使得测得的值不是一个确定的值, 而是符合正态分布的一个范围, 因而 AFM 测定单分子间的作用力均用高斯分布图来得到单分子力. 图 4a 纵坐标表示蛋白质与 DNA 元件结合成键的概率, 横坐标表示每次结合的力的大小. 蛋白质与每种元件的测量次数为 300 次, 测量的统计学结果符合高斯曲线分布规律. GST 蛋白和突变 GCC 元件作为负对照显示无特异性作用力. 经计算以后得到 GST-TSH1 与 GCC 元件结合力的大小为 (83.9 ± 2.2) pN.

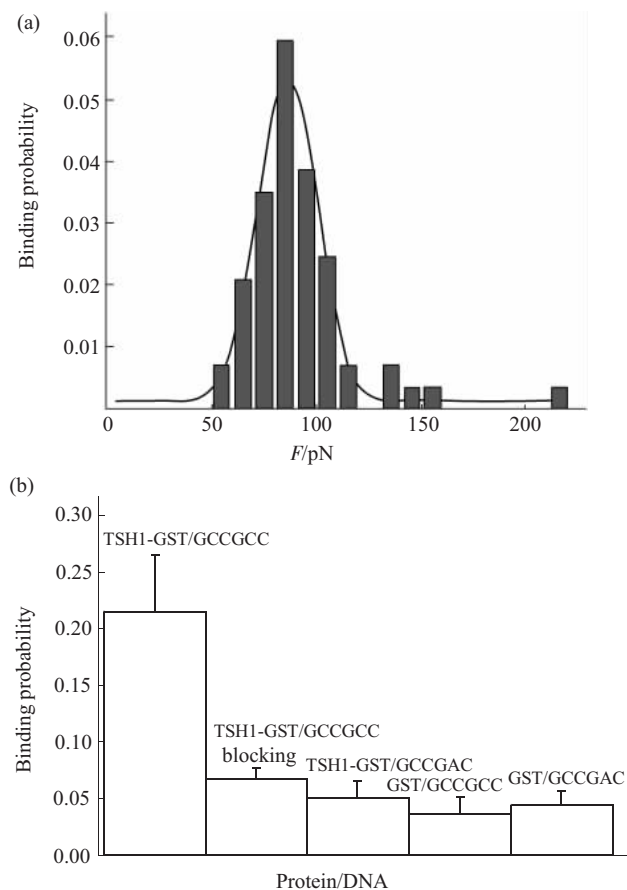


Fig. 4 Analysis of the interaction of TSH1 with the specific DNA elements

(a) The histogram of the adhesion forces of TSH1/GCC, which was obtained from about 300 force-distance cycles. The bars represent the experimental data while the solid line represents a theoretical Gaussian curve. (b) Binding probabilities obtained with various protein-modified tips and different DNA-modified substrates.

为了进一步验证原子力显微镜检测 TSH1 蛋白与 DNA 元件单分子作用力的特异性. 我们用游离

于探针针尖的蛋白质作为抑制剂, 观察在相同体系下加入游离蛋白对于蛋白质与 DNA 结合力检测的影响. 图 4b 结果显示, 加入游离的 TSH1 蛋白质明显降低了探针上 TSH1 蛋白与基底特异性 DNA 元件(GCCGCC)结合成键的概率, 与 GCC 元件的结合几率从 $(21.5 \pm 5.0)\%$ 下降到了 $(6.7 \pm 1.0)\%$. 而 TSH1 蛋白修饰的针尖与 DRE 元件 (GCCGAC) 修饰的基底之间的成键几率只有 $(5.1 \pm 1.5)\%$. 这个结果同时也表明了原子力显微镜对于检测蛋白质与 DNA 相互作用力时单碱基的突变也是十分灵敏的.

3 讨 论

通过以上结果可以证明, 水稻 TSH1 蛋白具有转录因子的典型特征, 其不能与 DRE 元件相互作用, 而专一性结合于 GCC 元件. 因此我们可以确定 TSH1 为 EREBP 亚族的 B 类转录因子.

至今, 在包括植物在内的真核生物中已经发现了许多转录因子, 对其结构和体内、体外功能的研究虽然取得了一些进展, 但在分子水平上尚未完全清楚地阐明转录因子参与调控的机制. 如特异的 DNA 结合蛋白影响转录起始速度的机制, 不同的调控蛋白是如何激活基因的转录起始从而控制不同的生命过程, 如细胞分化、器官发育、形态建成、光合作用与生物合成等, 转录因子是如何接受外源信号以及如何控制靶基因的表达等, 都仍是必须深入研究的课题. 随着实验方法的不断改进, 越来越多物种中不同家族的转录因子基因也被克隆出来, 对于这些转录因子功能的分析也越来越得到人们的重视. 对于水稻 TSH1 转录因子的研究结果不仅有助于人们更好地了解基因表达的调控机理, 同时利用这些结果为基因工程的研究提供更多的资源, 生产出人类所期望的基因工程产品.

目前对于转录因子功能鉴定的传统方法包括酵母单杂交和凝胶阻滞试验. 酵母单杂交方法根据转录因子与 DNA 顺式作用元件结合, 从而能调控报道基因表达的原理来克隆或者验证编码目的转录因子的基因. 该方法将已知的特定顺式作用元件构建到最基本启动子上游, 基本启动子下游连接报道基因. 进行 cDNA 融合表达文库筛选或特定基因功能验证时, 编码目的转录因子的 cDNA 融合表达载体被转化进入酵母细胞后, 其编码产物 (转录因子) 与顺式作用元件结合, 就可以激活基本启动子, 并促使报道基因表达. 根据报道基因的表达, 筛选或确认与已知顺式元件结合的转录因子. 酵母单杂交

方法是在细胞内分析鉴定转录因子与顺式作用元件结合的有效方法.

除此之外, 原子力显微镜 (AFM) 作为一种直接检测分子间相互作用力的工具, 可以用于揭示 DNA 与蛋白质间的相互作用. 与传统的凝胶阻滞实验相比, AFM 具有定量、准确、灵敏、微量、无需同位素操作等优点. 基于这些优势, 我们认为 AFM 在大规模鉴定和发现 DNA 顺式作用元件的研究上具有较好的应用前景. 通常, 要确定一个关键的 DNA 元件, 首先要分析得到一系列表达模式相似的基因, 然后根据它们的启动子序列特征进行启动子的删减实验, 最终得到由 4~8 个碱基构成的起关键性作用的核心序列, 即 DNA 顺式作用元件. 然后, 通过酵母单杂交的方法分离克隆与这样的 DNA 序列结合的转录因子, 从而揭示这一系列基因表达调控的一个可能模式. 随着基因组学的发展, 在模式生物中越来越多的基因已经被分离克隆, 现在已经知道在拟南芥中大约有 1 533 个转录因子, 果蝇、线虫和酵母分别为 635、669 和 209 个. 我们可以通过已知的转录因子的结构特征, 推测功能未知的一些转录调控蛋白的结构, 但是同一类转录因子即使具有高度相似的 DNA 结合域, 它们所识别的 DNA 序列却可能不同, 所参与的调控途径和生物学功能可能也完全不同. 因此, 要研究一个转录因子的功能, 其关键是要确定其参与的调控途径或者发现其目标 DNA 元件. 传统的凝胶滞留实验作为体外验证某个 DNA 序列是否为某个转录因子靶序列的一种实验手段, 已经难以胜任大规模筛选和发现某个未知功能的转录因子的要求了. 然而, AFM 技术由于其要求的样品量极少, 并且可以将不同 DNA 序列的阵列排列到某个单位面积的基底上, 这样, 用未知功能的转录因子修饰的针尖与这个基底上的每个不同的点接触, 测定它们之间的相互作用力, 我们可以同时得到一个蛋白质分子与上千种 DNA 序列的结合力, 从而发现该转录因子的目的序列. 这一结果可以为进一步鉴定该转录因子的目的基因提供有价值的信息.

参 考 文 献

- Pandey A, Lewutter F. Nucleotide sequence database: a gold mine for biologist. *Trends Biochem Sci*, 1999, **24** (7): 276~280
- 王华春, 陈清轩. 充分利用 EST 数据库资源. *生物化学与生物物理进展*, 2000, **27** (4): 442~444
Wang H C, Chen Q X. *Prog Biochem Biophys*, 2000, **27** (4): 442~444
- Jofuku K D, den Boer B G W, Van Montagu M, *et al.* Control of *Arabidopsis* flower and seed development by the homeotic gene APETALA2. *Plant Cell*, 1994, **6** (9): 1211~1225
- Okamuro J K, Caster B, Villarreal R, *et al.* The AP2 domain of APETALA2 defines a large new family of DNA binding proteins in *Arabidopsis*. *Proc Natl Acad Sci USA*, 1997, **94** (13): 7076~7081
- Ohme T M, Shinshi H. Ethylene-inducible DNA binding proteins that interact with an ethylene-responsive element. *Plant Cell*, 1995, **7** (2): 173~182
- Liu Q, Kasuga M, Sakuma Y, *et al.* Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA-binding domain separate two cellular signal transduction pathways in drought- and low-temperature-responsive gene expression in *Arabidopsis*. *Plant Cell*, 1998, **10** (8): 1391~1406
- Riechmann J L, Meyerowitz E M. The AP2/EREBP family of plant transcription factors. *Biol Chem*, 1998, **379** (6): 633~646
- Nole-Wilson S, Krizek B A. DNA binding properties of the *Arabidopsis* floral development protein AINTEGUMENTA. *Nucleic Acids Res*, 2000, **28** (21): 4076~4082
- Gilmour S J, Zarka D G, Stockinger E J, *et al.* Low temperature regulation of the *Arabidopsis* CBF family of AP2 transcriptional activators as an early step in cold-induced COR gene expression. *Plant J*, 1998, **16** (4): 433~442
- Medina J, Bargues M, Terol J, *et al.* The *Arabidopsis* CBF gene family is composed of three genes encoding AP2 domain-containing proteins whose expression is regulated by low temperature but not by abscisic acid or dehydration. *Plant Physiol*, 1999, **119** (2): 463~470
- Nakashima K, Shinwari Z K, Sakuma Y, *et al.* Organization and expression of two *Arabidopsis* DREB2 genes encoding DRE-binding proteins involved in dehydration- and high-salinity-responsive gene expression. *Plant Mol Biol*, 2000, **42** (4): 657~665
- Riechmann J L, Heard J, Martin G, *et al.* *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. *Science*, 2000, **290** (5499): 2105~2110
- Jiang Y X, Qin F, Li Y Q, *et al.* Measuring specific interaction of transcription factor ZmDREB1A with its DNA responsive element at the molecular level. *Nucleic Acids Res*, 2004, **32** (12): 101~108
- Qin F, Jiang Y X, Ma X Y, *et al.* A study of specific interaction of the transcription factor and the DNA element by atomic force microscopy. *Chin Sci Bull*, 2004, **49** (13): 1376~1380

Characterization of a Novel AP2/EREBP Transcription Factor TSH1 Specifically Binding to GCC Elements*

CHEN Feng¹⁾, YU Jun-Ping²⁾, FANG Xiao-Hong^{2)**}, ZHANG Gui-You^{1)**}

(¹⁾Department of Biological Sciences and Biotechnology, Tsinghua University, Beijing 100084, China;

(²⁾Institute of Chemistry, The Chinese Academy of Sciences, Beijing 100080, China)

Abstract TSH1, a novel transcription factor in rice containing conserved AP2/EREBP domain, was screened from the EST database. The specific interaction of TSH1 and DNA elements was detected by two kinds of traditional methods, electronic mobility assay and yeast one-hybrid assay, which suggested that TSH1 specifically binds to GCC element *in vitro* and *in vivo*. Furthermore, It was confirmed by atom force microscopy (AFM) that TSH1 binds to GCC element (core sequence GCCGCC) not DRE element (core sequence GCCGAC), and the single molecular force between TSH1 and GCC element was quantitatively measured. The molecular force of TSH1 with GCC element is (83.9 ± 2.2) pN. This interaction can be observably blocked by the dissociate TSH1 protein in the solution. These above results proved that TSH1 recognizes and binds to GCC with specifically interaction, which also demonstrate the high sensitivity of the AFM measurements for the single base substitution of the GCC core sequence greatly reduced the binding affinity. Comparing these methods for identification the binding of transcription factor with its DNA element, it was considered that AFM is expected to be a sensitive and reliable method that offers valuable information for the characterization of transcription factors.

Key words TSH1, AP2/EREBP, transcription factor, AFM

*This work was supported by grants from The National Natural Science Foundation of China (39970166, 20225516 and 10334100).

**Corresponding author.

Zhang Gui-You. Tel: 86-10-62772254, Fax: 86-10-62772678, E-mail: zhang-gy@mail.tsinghua.edu.cn

Fang Xiao-Hong. Tel: 86-10-62650024, E-mail: xfang@iccas.ac.cn

Received: April 27, 2006 Accepted: May 19, 2006