

大豆種子貯蔵たん白質の遺伝子発現制御機構とその遺伝的変異に関する研究

吉野道子¹・堤 賢一²・阿部 純¹・島本義也¹・金澤 章^{*1}

¹北海道大学大学院農学研究科 ²岩手大学農学部附属寒冷バイオシステム研究センター

Analysis of the Regulation of Gene Expression and Genetic Variations in Seed Storage Proteins in Soybean

Michiko YOSHINO¹, Ken-ichi TSUTSUMI², Jun ABE¹,
Yoshiya SHIMAMOTO¹ and Akira KANAZAWA¹

¹Graduate School of Agriculture, Hokkaido University, Sapporo 060-8589

²Cryobiosystem Research Center, Faculty of Agriculture, Iwate University, Iwate 020-8550

ABSTRACT

The 7S globulin (β -conglycinin), a major component of seed storage proteins in soybean, comprises three subunits, α , α' and β . The nucleotide sequence of a 7.6-kb *Eco*RI-digested fragment of genomic DNA, which contains a gene, CG-3, of the β -conglycinin gene family, was analyzed. By comparing the sequence with cDNAs of β -conglycinin, we found that CG-3 encodes the α -subunit. The gene contained 6 exons and 5 introns, of which positions were conserved among genes for α -, α' - and β -subunits. A major transcription start site of the α -subunit gene was located at 56 bp upstream of the initiation codon by a primer extension analysis. Binding of proteins in the DNA region upstream of the α -subunit gene was examined by DNase I footprint analysis. Binding of proteins was found at 4 sites in the region -163 to +7. We focused our analysis on the DNA-protein interactions on a 39-bp sequence designated as Box I, which is conserved among upstream regions of genes for 7S-type seed storage proteins. Gel mobility shift assay showed that at least three different protein factors bind to three distinct elements located within the Box I sequence. *Soy Protein Research, Japan* 4, 11-18, 2001.

Key words : soybean, β -conglycinin, α -subunit, cis-acting element, DNA-protein interaction

大豆種子貯蔵たん白質の主要な成分である 7S グロ

ブリン (β -コングリシニン) は, α , α' および β という 3 種のサブユニットによって構成される. β -コングリシニン各サブユニットの遺伝子に関してはその

*〒 060-8589 札幌市北区北 9 条西 9 丁目

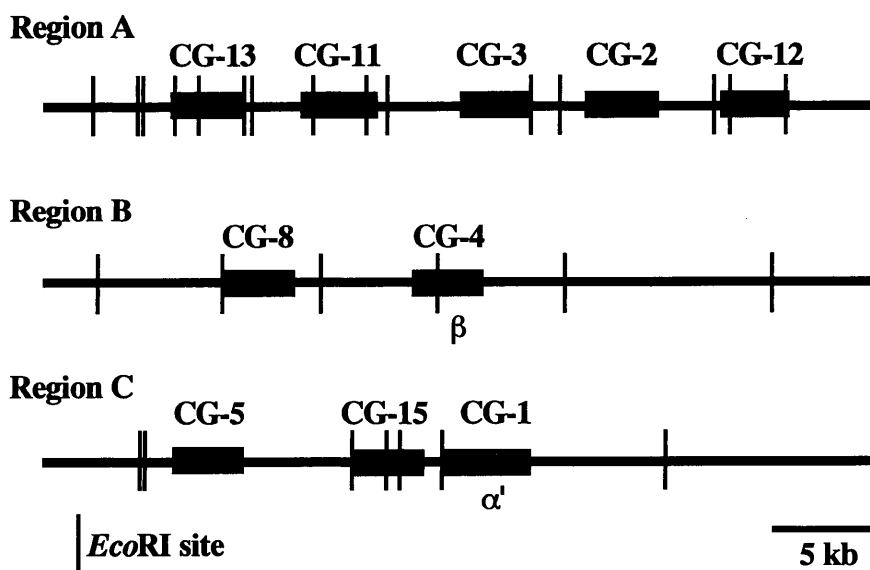


Fig. 1. Location of β -conglycinin genes in the soybean genome. The gene family of β -conglycinin comprises at least 15 genes designated as CG-1 to CG-15. These are located in several regions of soybean genome including the three regions shown here where genes are linked each other (see ref. 4).

cDNA 配列が明らかにされている¹⁻⁶⁾。それらの cDNA を用いてこれらのサブユニットの遺伝子を含むゲノム DNA 領域のクローン化が行われ、 α' および β サブユニット遺伝子についてはゲノム DNA の構造が明らかにされている^{2,4)}。 β -コングリシニン遺伝子は少なくとも 15 個の非常によく似た遺伝子 (CG-1 ~ CG-15 と名付けられている) からなる遺伝子族を構成しており、これらの遺伝子のいくつかは染色体の一部の領域に並列して存在していることが判明している⁴⁾ (Fig. 1)。ノーザン解析における転写産物のサイズから、Region C とよばれる DNA 領域内の CG-1、および Region A 内の CG-2 および CG-3 の 3 つの遺伝子が、 α もしくは α' サブユニットの遺伝子であると考えられた⁴⁾。このうち CG-1 は、 α' サブユニットを発現していない大豆品種「毛振」において欠失していることから、 α' サブユニットをコードすることが明らかになっている⁷⁾。残りの CG-2 および CG-3 は α サブユニット遺伝子の候補と考えられてきた。

3 種のサブユニットの遺伝子の発現は、種子の登熟過程において制御を受けている。転写制御に関わる因子に関しては、トランスジェニック植物を用いた解析によって、 α' および β サブユニット遺伝子に関して詳細に解析されている⁸⁻¹¹⁾。しかしながら、 α サブユニット遺伝子のゲノム DNA が単離されていなかったため、 α サブユニットの発現制御機構は未解明であった。種子登熟過程における遺伝子発現制御機構を解明する

ことは、大豆の育種を推し進める上で必須の課題であると考え、本研究では α サブユニット遺伝子のゲノム DNA を同定し、さらに α サブユニット遺伝子上流域におけるたん白質因子の結合に関する解析を行った。

方 法

材料

α サブユニット遺伝子の転写開始点の解析には、大豆品種「毛振」を用いた、DNase I フットプリント法およびゲル移動度シフト法による遺伝子上流域へのたん白質の結合の解析には「Forrest」を用いた。 α サブユニット遺伝子を同定する目的で、「Forrest」の核ゲノムの Region A に存在し、 α サブユニットをコードする遺伝子の候補である CG-3⁴⁾を含む 7.6 kb の *Eco*RI 断片をもつプラスミド DNA (p α 7.6#7, 千葉大学中村郁郎博士より分譲) を解析に用いた。

塩基配列の解析

CG-3 を含むプラスミド DNA に関してエキソヌクレアーゼ III によるデリーションを行い、塩基配列を解析した。塩基配列の解読にはオートシーケンサー (ALF express TM DNA sequencer, Pharmacia Biotech AB) を用いた。

プライマー伸長法による転写開始点の決定

グアニジン・イソチオシアネート法によって登熟中の種子より抽出した全 RNA を用いて、Terui らの方

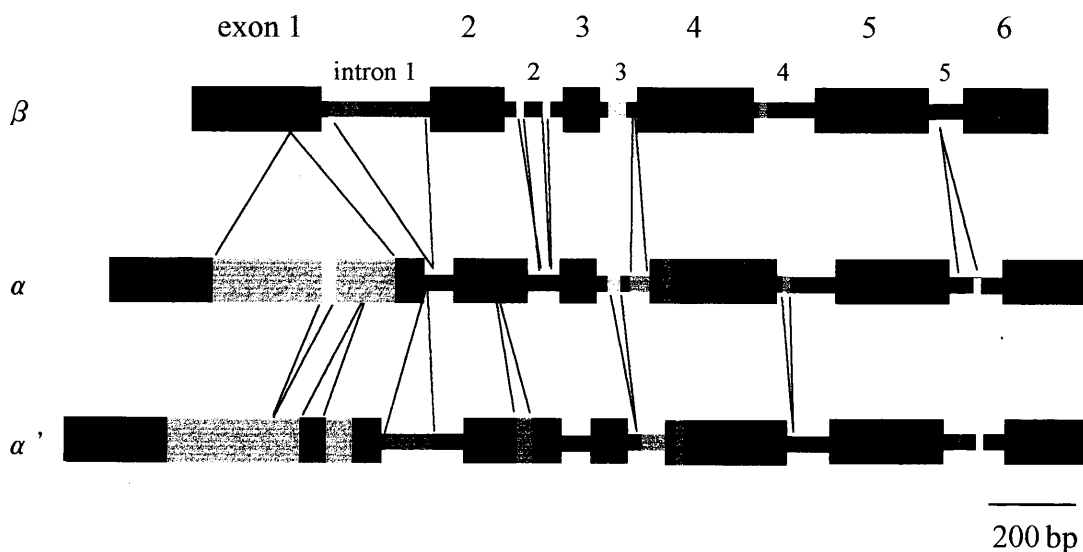


Fig. 3. Comparison of the genomic DNA structures of genes encoding α -, α' - and β -subunits of β -conglycinin.

α サブユニット遺伝子 5' 側上流領域の塩基配列の特徴

α サブユニット遺伝子の 5' 側上流領域の塩基配列中に、これまでに報告されている転写制御に関与するシスエレメントが存在するかを解析した. soybean embryo factors (SEF) 1, および SEF4 と呼ばれる転写制御因子が結合し, 子葉における組織特異的発現に関与すると考えられている配列¹¹⁾は, α' および β サブユニット遺伝子では翻訳開始点の上流 1,000 bp 以内の領域に多数検出されているのに対し, α サブユニット遺伝子では同じ領域には 2 箇所しか検出されず, 1,000 bp よりも上流の領域において多数検出された (Fig. 4). 同様に α' サブユニット遺伝子上流領域に存在する SEF3 と呼ばれる転写制御因子 SEF3 が結合する配列は, α サブユニット遺伝子の 5' 側上流領域では検出されな

かった. さまざまな種子貯蔵たん白質遺伝子上流領域にひろく存在する配列 5'-CAACACATG-3' やレグミンボックスのコア配列である RY repeat^{2, 4, 14, 15)}は, α サブユニット遺伝子においても存在した. これらに加えて, α , α' およびファゼオリン β タイプ遺伝子間の 5' 側上流領域に共通して存在する 4 つの領域 (Box I, II, III, および IV) が見つかった (Fig. 5). Box I は 7S グロブリン型のたん白質遺伝子の多くに共通して存在し, 組織特異的発現に関与すると考えられているビシリンボックス¹⁶⁻¹⁸⁾を含んでいた.

α サブユニット遺伝子の 5' 上流領域における核たん白質の結合

α サブユニット遺伝子上流領域における核たん白質の結合を DNase I フットプリント法で解析した

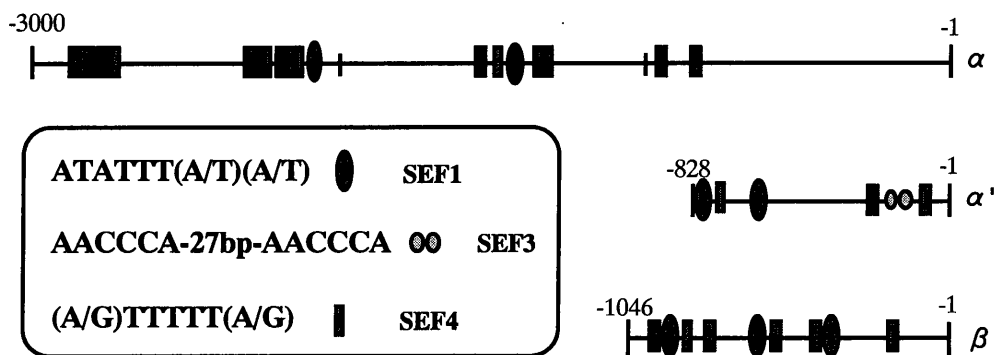


Fig. 4. Location of SEF1, SEF3 and SEF4 elements in the upstream regions of β -conglycinin genes.

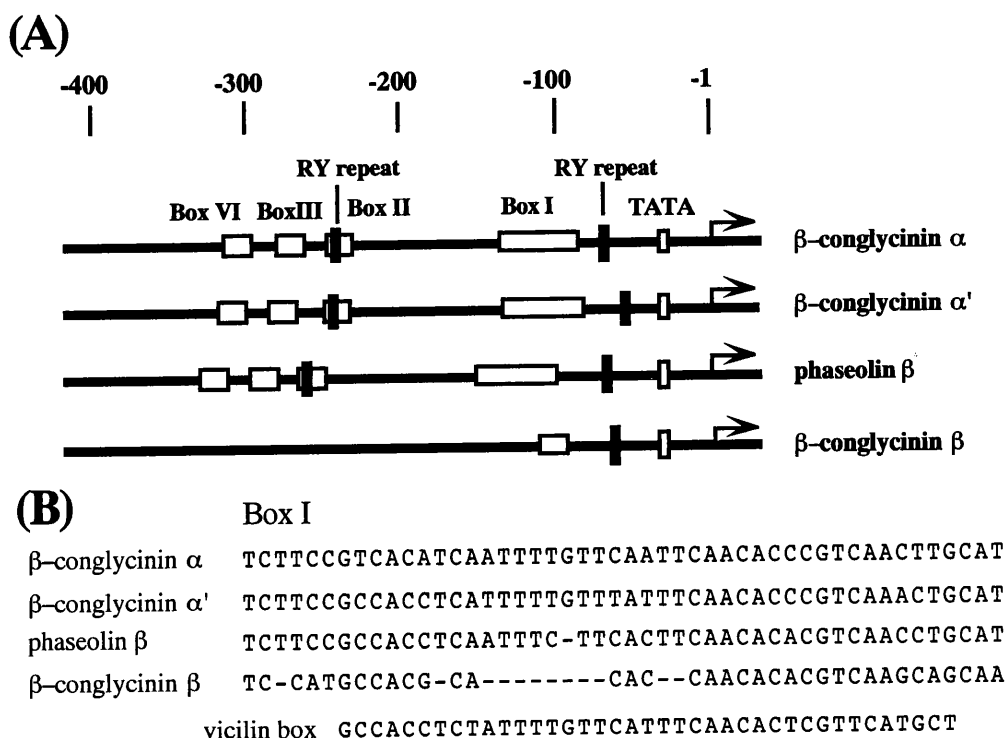


Fig. 5. Conserved elements in the upstream regions of 7S-type seed storage proteins. (A) Locations of Box I-BoxIV and RY repeat. (B) Nucleotide sequence of Box I. Note that the Box I contains a sequence similar to vicilin box.

(Fig. 6). 核抽出物は「Forrest」の開花後 42 日目の登熟中の未熟種子から得た。-230 から +76 bp までの領域を PCR で増幅した DNA 断片のプロープを核抽出物と複合体を形成させた後、DNase I で消化し、電気泳動で分離したところ、4 つの領域、-163 から -150 (フットプリント 1)、-145 から -86 (フットプリント 2)、-77 から -64 (フットプリント 3)、および -60 から +7 (フットプリント 4) でバンドが消失し、これらの領域に核たん白質が結合することが明らかになった。フットプリント 2 の配列はビシリンボックスを含み、フットプリント 3 の配列はレグミンボックスを含んでいた。フットプリント 4 の配列は TATA ボックスに結合する因子群や RNA ポリメラーゼ II が結合する領域を含んでいた。

Box I 配列に対する核たん白質の結合

ビシリンボックスを含む Box I への核たん白質の結合を、ゲル移動度シフト法によって解析した。Box I をプロープに用いてゲル移動度シフト法による解析を行ったところ、4 種類の DNA-たん白質複合体が検出された。非標識の Box I の 5' 側領域 (TT)、Box I の中心部 (TCA) をそれぞれ競合 DNA として、複合体

形成の競合実験を行ったところ、TT あるいは CA の一方でのみ競合するバンドと、両者で競合するバンドが見られ、Box I がいくつかの異なるたん白質の結合領域となっていることが示唆された。

そこで、TT、TCA および CA (Box I の 3' 側) をプロープとして同様のゲル移動度シフト法による解析を行った (Fig. 7)。TT、TCA、CA のそれぞれによる競合の有無から、Box I に結合する核たん白質は少なくとも 3 種類あることが示唆された (Fig. 8)。一つは TT、TCA および CA のすべてに結合する核たん白質因子 (これを SNF-A、seed nuclear factor A と名付けた) である。二つめは TCA および CA に結合する核たん白質因子 (SNF-B) であり、三つ目は CA にのみ結合する核たん白質因子 (SNF-C) である。TT、TCA および CA の全てに結合する SNF-A は、これらのプロープに共通な TCAATT を認識する因子であると考えられた。同様に SNF-B の認識に重要な配列は TCA と CA で重複する配列 (TCAATTCAACAC) と考えられた。また、SNF-C の認識配列はプロープ CA に特有で、TT および TCA にはない配列と考えられた。

α' 遺伝子⁸⁾やファゼオリン遺伝子¹⁹⁾のプロモーター

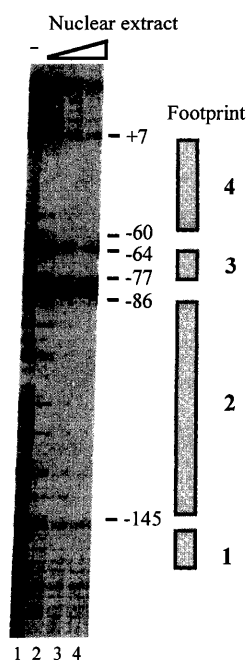


Fig. 6. DNase I footprint analysis. Binding reaction was performed in the presence of no nuclear extract (lane 1), 5 μ g (lane 2), 10 μ g (lane 3), and 20 μ g (lane 4) of nuclear extracts. The result indicates that DNA-protein interactions occur 4 sites within the region -163 to +7.

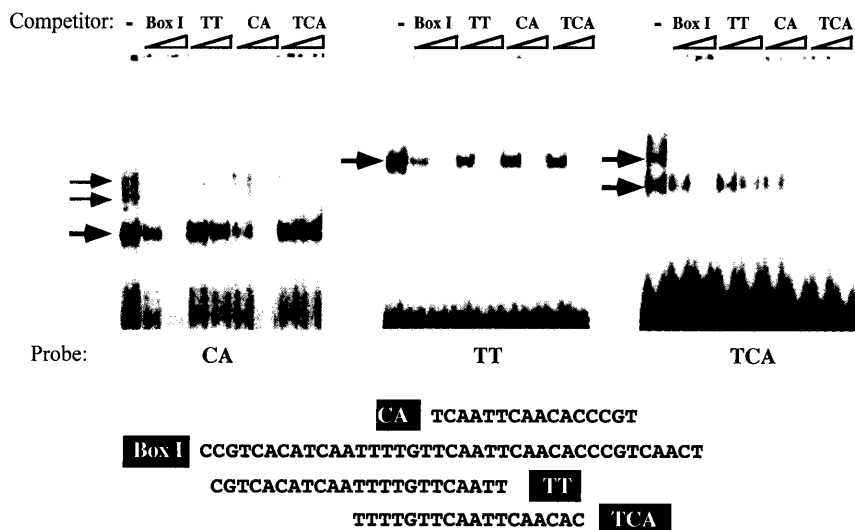


Fig. 7. Gel shift analysis of nuclear extract using three probes of Box I, CA, TT and TCA. 10 $\times$ or 50 $\times$ amounts of competitors were added in each reaction mixture. When CA was used as a probe, three DNA-protein complexes were detected: the formation of the complex of higher mobility competed with Box I and CA, but not with TT and TCA; the formation of the other two complexes was competed with all the four sequences. When TT was used as a probe, a single complex was detected: the formation of the complex was competed with all the four sequences. When TCA was used as a probe, two complexes were detected: the formation of the complex of higher mobility competed with Box I, TCA and CA, but not with TT; the formation of the complex of the lower mobility competed with all the four sequences. As a consequence, binding of at least three proteins to the Box I sequence was suggested.

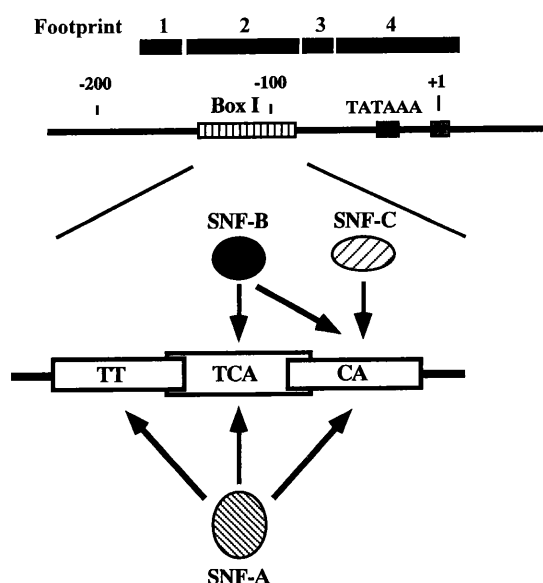


Fig. 8. Summary of DNA-protein interactions on the upstream region of the α -subunit gene. Box I sequence located within the footprint 2 of DNase I footprint analysis. Within the Box I sequence, three different factors, designated as SNF-A, SNF-B and SNF-C bind to three distinct elements. SNF-A binds to TT, TCA and CA. SNF-B binds to TCA and CA. SNF-C binds to CA.

において、Box I を含む領域が欠失すると活性が大きく低下することを考えると、ここで解析した Box I の核たん白質結合領域が転写制御に関わっていることが十分に考えられる。 α' サブユニット遺伝子では、典型的な SEF4 結合サイトが Box I 内にあり、実際に SEF4 が結合することが報告されている¹⁰⁾。SEF4 結合サイト (A/G) TTTT (A/G) に似た配列 AATTTTG が、 α サブユニット遺伝子の Box I の TT 配列中にあるが、この部分を特異的に認識する核たん白質は検出されなかった。 α' サブユニット遺伝子のプロモーターと異なり、 α サブユニット遺伝子の 5' 側上流 400 bp 内に SEF3 や SEF4 の結合サイトは見あたらない。本研究の結果は、 α サブユニット遺伝子と α' サブユニット遺伝子の Box I では質的に異なる核たん白質の結合が起こっていることを示しており、 β -コングリシニンのサブユニット間での協調的発現に、これまで考えられてきた SEF の結合による制御に加えて、より複雑な制御機構が介在していることを示唆している。

要 約

大豆種子貯蔵たん白質の主要な成分である 7S グロブリン (β -コングリシニン) の α サブユニット遺伝子のゲノム DNA を含むと考えられた 7.6 kb の DNA 断片の塩基配列を解析し、この断片上に α サブユニット遺伝子が実際に存在すること、およびその遺伝子構造、発現制御を行うと考えられた領域の塩基配列を明らかにした。この遺伝子は 6 個のエクソンからなり、イントロンの位置は α' および β サブユニット遺伝子におけるものと類似していた。プライマー伸長法によって転写開始点を解析したところ、主要な転写開始点は、翻訳開始コドンの 56 塩基上流に同定された。また、DNase I フットプリント法により、遺伝子上流域におけるたん白質の結合を解析したところ、転写開始点に対して -163 bp から +7 の領域内で 4 つの領域にたん白質が結合することが明らかになった。これらの中で転写制御に深く関わりと示唆された領域において、他の種子貯蔵たん白質遺伝子との間で高度に保存された 39 bp の配列を見だし、Box I と名付けた。この Box I 配列に対するたん白質の結合をゲルシフト法によって解析した。さまざまな DNA 配列を用いた解析を行うことで、たん白質の結合に関して、Box I 配列はさらに 3 つの領域に分けられること、また、この配列には少なくとも 3 種類のたん白質が結合することを明らかにした。

文 献

- 1) Schuler MA, Ladin BF, Pollaco JC, Freyer G and Beachy RN (1982): Structural sequences are conserved in the genes coding for the α , α' and β -subunits of the soybean 7S seed storage protein. *Nucl Acids Res*, **10**, 8245-8261.
- 2) Doyle JJ, Schuler MA, Godette WD, Zenger V and Beachy RN (1986): The glycosylated seed storage proteins of *Glycine max* and *Phaseolus vulgaris*: structural homologies of genes and proteins. *J Biol Chem*, **261**, 9228-9238.
- 3) Tierney ML, Bray EA, Allen RD, Ma Y, Drong RF, Slightom J and Beachy RN (1987): Isolation and characterization of a genomic clone encoding the β -subunit of β -conglycinin. *Planta*, **172**, 356-363.
- 4) Harada JJ, Barker SJ and Goldberg RB (1989): Soybean β -conglycinin genes are clustered in several DNA regions and are regulated by transcriptional and posttranscriptional processes. *Plant Cell*, **1**, 415-425.
- 5) Sebastiani FL, Farrell LB, Schuler MA and Beachy RN (1990): Complete sequence of a cDNA of α subunit of β -conglycinin. *Plant Mol Biol*, **15**, 197-201.
- 6) Lelievre JM, Dickinson CD, Dickinson LA and Nielsen NC (1992): Synthesis and assembly of soybean β -conglycinin *in vivo*. *Plant Mol Biol*, **18**, 259-274.
- 7) Ladin BF, Doyle JJ and Beachy RN (1984): Molecular characterization of a deletion mutation affecting the α' subunit of β -conglycinin of soybean. *J Mol Appl Genet*, **2**, 372-380.
- 8) Chen Z-L, Schuler MA and Beachy RN (1986): Functional analysis of regulatory elements in a plant embryo specific gene. *Proc Natl Acad Sci USA*, **83**, 8560-8564.
- 9) Chen Z-L, Pan N-S and Beachy RN (1988): A DNA sequence element that confers seed-specific enhancement to a constitutive promoter. *EMBO J*, **7**, 297-302.
- 10) Allen RD, Bernier F, Lessard PA and Beachy RN (1989): Nuclear factors interact with a soybean β -conglycinin enhancer. *Plant Cell*, **1**, 623-631.
- 11) Lessard PA, Allen RD, Bernier F, Crispino JD, Fujiwara T and Beachy RN (1991): Multiple nuclear factors interact with upstream sequence of differentially regulated β -conglycinin genes. *Plant Mol Biol*, **16**, 397-413.
- 12) Terui Y, Tsutsumi K, Kidou S, Sawazaki T, Kuroiwa Y, Yamaki M and Ejiri S (1998): A novel variant of translation elongation factor-1 β : isolation and characterization of the rice gene encoding EF-1 β 2. *Biochim Biophys Acta*, **1442**, 369-372.
- 13) Lee W, Mitchell P and Tjian R (1987): Purified transcription factor AP-1 interacts with TPA-inducible enhancer elements. *Cell*, **49**, 741-752.
- 14) Nielsen NC, Dickinson CD, Cho T-J, Thanh VH, Scallon BJ, Fischer RI, Sims TL, Drews GN and Goldberg RB (1989): Characterization of the glycinin gene family in soybean. *Plant Cell*, **1**, 313-328.
- 15) Fujiwara T and Beachy RN (1994): Tissue-specific and temporal regulation of a β -conglycinin gene: roles of the RY repeat and other cis-acting elements. *Plant Mol Biol*, **24**, 261-272.
- 16) Brown D, Ellis THN and Gatehouse JA (1988): The sequence of a gene encoding convicilin from pea (*Pisum sativum* L.) shows that convicilin differs from vicilin by an insertion near the N-terminus. *Biochem J*, **251**, 717-726.
- 17) Higgins TJV, Newbiggin EJ, Spencer D, Llewellyn DJ and Craig S (1988): The sequence of a pea vicilin gene and its expression in transgenic tobacco plants. *Plant Mol Biol*, **11**, 683-695.
- 18) Newbiggin EJ, de Lumen BO, Chandler PM, Gould A, Blagrove RJ, March JF, Kortt AA and Higgins JV (1990): Pea convicilin: structure and primary sequence of the protein and expression of a gene in the seeds of transgenic tobacco. *Planta*, **180**, 461-470.
- 19) Bustos MM, Begum D, Kalkan FA, Battraw MJ and Hall TC (1991): Positive and negative cis-acting DNA domains are required for spatial and temporal regulation of gene expression by a seed storage protein promoter. *EMBO J*, **10**, 1469-1479.