

高等植物におけるpol-II依存性プロモーター

Promoter architecture of higher plants

山本 義治

(岐阜大学 応用生物科学部)

はじめに

動植物を問わずタンパク質をコードする遺伝子はRNA polymerase IIにより転写される。高等植物のpol-II依存性プロモーターの研究は環境変化や植物ホルモンなどによる転写調節領域の同定を目的とする研究が長く主流であった。こういった研究は分子生物学的な手法の発達や植物への遺伝子導入法の確立のおかげで技術的に十分手の届く研究分野であり、またその成果はアグロバイテクなどへのアウトプットが見込めるために需要があったので研究者人口も多い。一方、転写制御ではなく転写そのものを可能にするコアプロモーターの研究に関しては、研究活動としては非常にまばらな状況が長く続いているのは寂しい限りである。本レビューでは著者らの研究を中心に高等植物のプロモーター構造についての知見を紹介していきたい。

TATAボックスとTATA-lessプロモーター

植物遺伝子の構造決定が行われると動物同様植物プロモーターにもTATAボックスがあることが知られるようになり、1987年には当時構造決定されていた79の植物遺伝子（RuBisCO小サブユニットやCAB、Lectin、Leghaemoglobinなどが多かった）より得られたTATAボックスのコンセンサス配列がまとめられた(1)。この報告では84%もの遺伝子について明瞭なTATA配列が確認されており、TATA配列が無い遺伝子は例外的であった。このコンセンサス配列はその後四半世紀にわたって用いられることになった。

この少し前に植物の形質転換技術が開発され、プロモーター上の機能配列の検定が行えるようになった(2)。この技術を用いてTATAボックスが植物の転写に必要であることが1987年にde Paterらにより確かめられている(3)。その後、システムチックな点変異の導入により、TATA配列と転写活性の関係が*in vivo*転写系(4)や*in vivo*での実験系(5)により詳細に解析されている。余談になるが、TATAボックスに結合するタンパク質であるTATA-Binding Protein (TBP)の立体構造はシロイヌナズナの遺伝子産物を用いて初めて解明された(6)。

先に述べた1987年のJoshiらの論文ではTATAボックス以外の植物コアプロモーター因子として転写開始点近傍のコンセンサス配列であるInitiator (Inr)が同定されている。

筆者は1989年に大学院修士課程に入学し、博士課程を修了するまで北海道大学小保方潤一研究室のメンバーとしてタバコ光合成遺伝子の構造解析などを行っていたが、当時構造決定したタバコpsaDbプロモーターはTATA配列を持たないものであった(7)。配列からは認識出来ないが機能的にはTATA様の因子を持つ、という可能性を検討するために、コアプロモーター領域に塩基置換を導入し機能解析を行った結果、確かにTATAボックスは存在しないこと、また、psaDbに存在していたInr配列が転写に必要であることが研究室の中邨らにより確認された(8)。

哺乳類のコアプロモーターに関しては植物のものと比較するとより多くの知見が得られており、TATAボックスを持つプロモーターは全体の3割程度であること、TATA-lessのプロモーターにはCpGアイランドが見られること、TATA型は組織特異的発現を示すプロモーターに多く、CpG型は恒常的な発現をするものに多いことからコアプロモーターのタイプに役割の分化が見られること、等が知られている(9, 10, 11)。CpG型の転写機構に関してはよくわかっていないが、哺乳類に関してはTATA型とCpG型という二つのコアタイプでゲノムに含まれるおおよそのプロモーターをカバーすることが出来ている。従って、コア構造が認識出来ない、という動物プロモーターは少数派である。

さて、植物プロモーターにはTATAボックスとInrという2つのコア因子が存在することが明らかになったものの、植物プロモーターの全体像は不明なままである。植物のTATA-less型プロモーターの存在は明らかになったが、CpG型が植物に存在するのかどうか、もし存在しないとすれば何か代替りの因子があるのかどうかなど、いくつかの重要な疑問が残された。

高等植物のプロモーター構成配列の抽出

筆者は2001年頃から光ストレス応答を示す遺伝子群をマイクロアレイ解析により同定し、新規シス配列の抽出を試みていた。ストレス誘導されるプロモーター群に特徴的な短い配列を探そうというものである。しかし、頻度解析だけでは信頼度の高い予測配列は難しく、何か別の指標による「転写制御配列らしさ」の計測の必要を感じていた。

そんな折、ヒトプロモーターの統計解析の結果から、ある種の転写制御配列はプロモーター1kb程度の領域に均一に存在するのではなくて下流側により多く現

れる、という現象が報告された (12)。筆者はこの報告にヒントを得て、「プロモーター上の特定の位置に頻出する配列を抽出すれば機能配列をまとめて取得出来るのではないか」と考えた。そこで、とりうる全ての6塩基配列 ($4^6 = 4,096$) について局在性を示す配列を全て抽出したところ、TATAボックスやG Box、ABRE、Site IIなどよく知られている転写に関わる配列が多数含まれていた。そこで、さらに抽出の条件検討を行い、解析する塩基の長さは長い方が強いシグナルが得られること (配列特異性が高まるためである)、しかし長くなると配列あたりの出現回数が減少するために統計解析に耐える出現数が得られない解析不能配列が多くなってしまふこと (未発表データ)、TATAボックスの詳細な活性-配列相関解析の結果 (5) と照らし合わせると配列の局在性と転写活性には非常に強い相関があること (13)、などのことが分かった。著者らはこの抽出手法をLDSS (Local Distribution of Short Sequences)法と名付け、抽出された配列はLDSS陽性配列と呼ぶことにした (14)。

設定した条件に従い8塩基配列の総当たり ($4^8 = 65,536$) でLDSS解析を行い、陽性配列をゲノム当たり1,000程度抽出しそれらを分布パターンに従ってクラスタリングした結果、既に機能配列として知られているグループとして、REG (Regulatory Element Group)、TATA、Inr、Kozakが、さらに新規のプロモーター構成因子としてY Patch、CA、GAを同定することが出来た (図1) (13, 14, 15)。後者のグループは新規のコア因子であると考えられる。

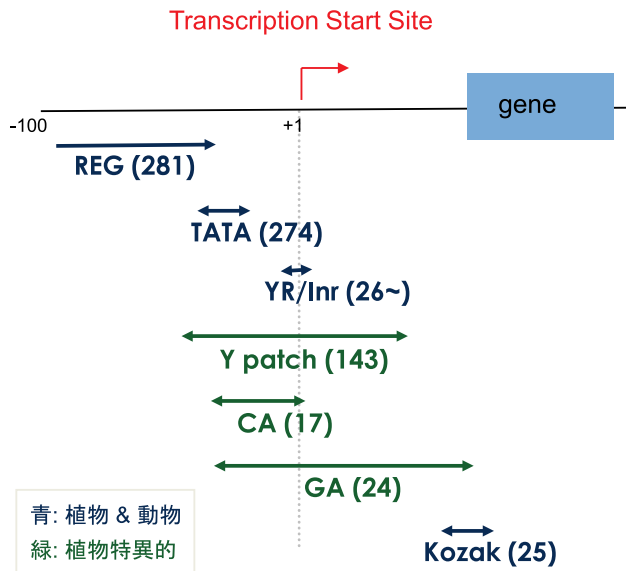


図1. 植物から抽出されたプロモーター構成配列

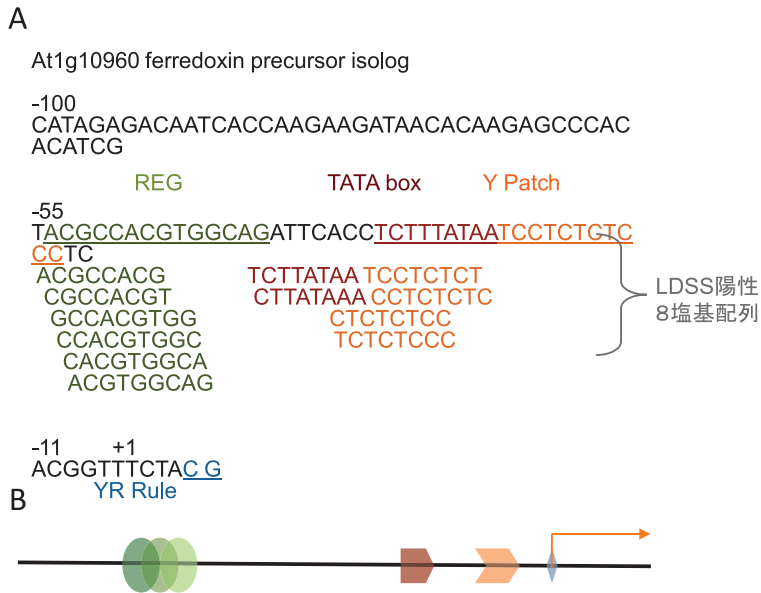


図2. LDSS陽性配列を用いたプロモーター構造認識

A. ゲノム配列からLDSS陽性の8塩基配列を検出する。B. 模式図

当初の予想通りマイクロレイデータをもとにした転写制御配列候補の中からさらにLDSS陽性なもの（REG）を選べば非常に確度の高い予測配列を得ることが出来るようにはなったが、転写制御配列の中にはLDSS陰性のものも含まれているので（14, 16）、予測配列がLDSS陰性だからといって転写制御配列ではないとは言えない。従って、当初期待したような確実な配列フィルタリング手法を確立できた、というほどの成果が得られた訳ではないが、それでも機能未知の確度の高い制御配列候補を100以上得られたのは大きな進展であった。

高等植物のコアプロモーター構造の異種性

抽出されたLDSS陽性配列を用いるとプロモーター構造を認識することが可能になる（図2）。著者らによるシロイヌナズナのゲノムワイドなプロモーター構造認識の結果から、TATA型プロモーターは動物同様3割程度であり、残り7割がTATA-lessであることが分かった。1987年のJoshiによる解析（1）では8割の遺伝子がTATA型であり筆者らの結果とは大分違うが、結局この頃クローニングされていた遺伝子は高発現なものばかりであり、ゲノム全体のプロモーター集団が

ら見ると相当偏っていたのであった。筆者らの結果に戻ると、TATA-less型の植物プロモーターのうちY Patch、CA、GAのいずれかが検出されるものが4割程度あり、プロモーター構造が検出できない（もしくは構造を持たない）corelessプロモーターは3割程度にまで下げることが可能になった。結局、著者らの手法により7割の植物プロモーターについて構造検出が可能になった。

転写開始点情報はプロモーター解析において非常に有用な情報を与えてくれる(11)。そこで、筆者らはCT-MPSSという新規手法によりシロイヌナズナの高品質転写開始点タグを16万同定し、約2万4千のプロモーターを特定することに成功した(15)。得られたデータから動物同様植物においてもプロモーターあたりの転写開始点は複数用いられるものが多いことや転写ノイズと思われるプロモーター活性が遺伝子間領域に多数存在すること("Orphan Promoter")などを観察した。さらに、遺伝子性のプロモーターについて定量的な解析を行ったところ以下のようなことがわかってきた。

- 1) TATAボックスはY Patch、Inrと共局在すること、高発現・単一ピーク型の転写開始点の形状をとるという特徴が共通していること、TATAボックスとInrの間に共進化的な関係が見られること、などからこれら3つの因子は協調的に機能していることが示唆された。
- 2) 上記のTATAグループとは独立した因子としてGAグループが存在する。これはTATA-lessのプロモーターに多く、TATAグループとは異なりブロード型の転写開始点を持つ。
- 3) マイクロアレイデータからTATA型プロモーターは環境応答性の発現パターンを示し、GA型やCore-less型は恒常的な発現を示す。

以上のことからシロイヌナズナにはTATAグループとGAグループという2つの主要なプロモーターグループが存在し、役割分化していることが明らかになった(図3)。動物においてはTATA型プロモーターは組織特異的発現をする遺伝子に多いことが報告されている(17)が、植物では環境応答を示す遺伝子に多いという点は、on/offの制御を受ける遺伝子という面では同じであるが、制御を受ける状況に動植物間の違いが現れており興味深い。

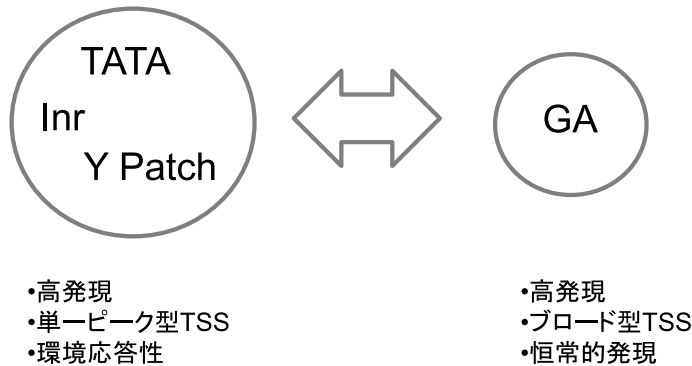


図3. シロイヌナズナのコアプロモーター因子間の関係
文献(15)のデータによる。

Plant Promoter Database (ppdb)の開発

筆者らはLDSS法により抽出された陽性配列を用いて検出されるプロモーター構造をデータベース化し、一般公開している (ppdb, <http://ppdb.gene.nagoya-u.ac.jp>)。対応しているゲノムはシロイヌナズナ、イネ、及びヒメツリガネゴケであり、特徴としては、独自に抽出したLDSS陽性配列をもとに構造認識していること、ゲノムごとに異なる機能配列リスト (LDSS陽性配列) を用いていること、転写開始点情報が充実していること、などが挙げられる。ppdbによりゲノム配列からプロモーター構造がわかり機能配列が特定出来るが、その生物学的な役割についての情報が乏しいのが現在の課題である。今後転写制御配列の発現パターンを実験的に明らかにしていく予定であるので、そう遠くない将来にはプロモーター配列から発現パターンの予測が行えるようになるかも知れない。

動植物間のコアプロモーターの分化

著者らはLDSS解析をイネ、ヒト、マウスのプロモーターに対しても行い、動植物間のプロモーター構造の比較を行った (13, 14)。その結果REG、TATA、Inrについては動植物間に共通して存在していたが、CpGアイランドは植物プロモーターには見られないこと、シロイヌナズナから発見されたY Patch、GA Element、CA Elementについては動物には存在せず植物特異的なコアプロモーター因子であること、などを明らかにした (表1)。これらのことからコアプロモーター構造が動植物間で分化していることが分かった。

	REG	TATA	Inr	Y Patch	GA	CpG Island
シロイヌ ナズナ	あり	あり	あり	あり	あり	なし
イネ	あり	あり	あり	あり	あり	なし
ヒト	あり	あり	あり	なし	なし	あり
マウス	あり	あり	あり	なし	なし	あり

表1. 主要なプロモーター因子の動植物での保存性LDSS解析による結果 (13) を示す。

哺乳類ではCpG型のプロモーターはブロード型の転写開始点を持ちかつ恒常的発現をするハウスキーピング遺伝子に多い (17, 18) 。前述のように植物にはCpGアイランドはプロモーター領域には存在しないが、GA Elementが同様の特徴を持っておりこれが動物CpGの代わりを果たしているものと考えられる。まず気になるのはCpGとGAの間には配列上の関連はないという点である。CpGアイランドのCはメチル化の制御を受けるが、メチル化されたCは脱アミノ化を經由してTへ変異する速度がはやいために、CpGはTGへと変異する傾向があるが、この点を考慮してもGA Elementとの配列上の関連はやはり見あたらないので、CpGからGA、もしくはGAからCpGへと遷移するような進化上の経路も現在のところ考えられない。従って、CpGとGAは独立に発生したものと想像されるが、動植物共通してこれらが発生する利点が無ければこのような状況は考えにくい。では動物も植物もTATA型のプロモーターだけではなくもう一つ別のプロモータータイプを必要としているのであろうか？

Schugらのヒトプロモーターの解析によれば、恒常的発現や胚での発現をする遺伝子にはTATA型プロモーターを持つものが少ない (17) 。もしかすると胚発生の際などでTATA型プロモーターでは発現できない、もしくは具合が悪いような状況が存在しており、そうした状況下においてはTATA-less型のプロモーターによる発現に何らかのメリットがあるのかも知れない。

まとめ

情報解析と実験解析とを組み合わせることで、植物プロモーターの構成因子についてほぼ全体像を概観することができた。今後は合成プロモーターを用いた研究により各プロモーター構成因子の機能や、ひとつひとつの転写制御配列が実現する転写調節パターン、転写制御配列の組み合わせによる相互作用、などについて解明していきたい。また、プロモーターの進化に関する研究はまだ緒についたばかりであり、どのような発見が出てくるのか非常に興味深い。

参考文献

- 1 Joshi C. P. (1987) An inspection of the domain between putative TATA box and translation start site in 79 plant genes. *Nucleic Acids Res.* 15, 6643-6653.
- 2 Bevan M. (1984) Binary *Agrobacterium* vectors for plant transformation. *Nucleic Acids Res.* 12, 8711-21.
- 3 de Pater B. S., de Kam R. J., Hoge J. H. and Schilperoort R. A. (1987) Effects of mutations in the TATA box region of the *Agrobacterium* T-cyt gene on its transcription in plant tissues. *Nucleic Acids Res.* 15, 8283-92.
- 4 Yamaguchi Y., Itoh Y., Takeda Y. and Yamazaki K. (1998) TATA sequence requirements for the initiation of transcription for an RNA polymerase II in vitro transcription system from *Nicotiana tabacum*. *Plant Mol. Biol.* 38, 1247-52.
- 5 Kiran K., Ansari S. A., Srivastava R., Lodhi N., Chaturvedi C. P., Sawant S. V. and Tuli R. (2006) The TATA-box sequence in the basal promoter contributes to determining light-dependent gene expression in plants. *Plant Physiol.* 142, 364-76.
- 6 Nikolov D. B., Hu S. H., Lin J., Gasch A., Hoffmann A., Horikoshi M., Chua N. H., Roeder R. G. and Burley S. K. (1992) Crystal structure of TFIID TATA-box binding protein. *Nature* 360, 40-6.
- 7 Yamamoto Y., Tsuji H. and Obokata J. (1993) Structure and expression of a nuclear gene for the PSI-D subunit of photosystem I in *Nicotiana glauca*. *Plant Mol. Biol.* 22, 985-994.
- 8 Nakamura M., Tsunoda T. and Obokata J. (2002) Photosynthesis nuclear genes generally lack TATA-boxes: a tobacco photosystem I gene responds to light through an initiator. *Plant J.* 29, 1-10.
- 9 Butler J. E. and Kadonaga J. T. (2002) The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes Dev.* 16, 2583-92.
- 10 Smale S. T. and Kadonaga J. T. (2003) The RNA polymerase II core promoter. *Annu Rev Biochem* 72, 449-79.
- 11 Carninci P., Sandelin A., Lenhard B., Katayama S., Shimokawa K., Ponjavic J., Semple C. A., Taylor M. S., Engstrom P. G., Frith M. C., Forrest A. R., Alkema W. B., Tan S. L., Plessy C., Kodzius R., Ravasi T., Kasukawa T., Fukuda S., Kanamori-Katayama M., Kitazume Y., Kawaji H., Kai C., Nakamura M., Konno H., Nakano K., Mottagui-Tabar S., Arner P., Chesi A., Gustincich S., Persichetti F., Suzuki H., Grimmond S. M., Wells C. A., Orlando V., Wahlestedt C., Liu E.

- T., Harbers M., Kawai J., Bajic V. B., Hume D. A. and Hayashizaki Y. (2006) Genome-wide analysis of mammalian promoter architecture and evolution. *Nat. Genet.* 38, 626-35.
- 12 Elkon R., Linhart C., Sharan R., Shamir R. and Shiloh Y. (2003) Genome-wide in silico identification of transcriptional regulators controlling the cell cycle in human cells. *Genome Res.* 13, 773-80.
- 13 Yamamoto Y. Y., Ichida H., Abe T., Suzuki Y., Sugano S. and Obokata J. (2007) Differentiation of core promoter architecture between plants and mammals revealed by LDSS analysis. *Nucleic Acids Res.* 35, 6219-6226.
- 14 Yamamoto Y. Y., Ichida H., Matsui M., Obokata J., Sakurai T., Satou M., Seki M., Shinozaki K. and Abe T. (2007) Identification of plant promoter constituents by analysis of local distribution of short sequences. *BMC Genomics* 8, 67.
- 15 Yamamoto Y. Y., Yoshitsugu T., Sakurai T., Seki M., Shinozaki K. and Obokata J. (2009) Heterogeneity of Arabidopsis core promoters revealed by high density TSS analysis. *Plant J.* 60, 350-362.
- 16 FitzGerald P. C., Shlyakhtenko A., Mir A. A. and Vinson C. (2004) Clustering of DNA sequences in human promoters. *Genome Res.* 14, 1562-74.
- 17 Schug J., Schuller W. P., Kappen C., Salbaum J. M., Bucan M. and Stoeckert C. J., Jr. (2005) Promoter features related to tissue specificity as measured by Shannon entropy. *Genome Biol.* 6, R33.
- 18 Saxonov S., Berg P. and Brutlag D. L. (2006) A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc. Natl. Acad. Sci. U. S. A.* 103, 1412-7.