

ROŚLINNE PROMOTORY POLIMERAZY RNA II – STRUKTURA I IDENTYFIKACJA

PLANT PROMOTERS OF RNA POL II –STRUCTURE AND IDENTIFICATION

Tomasz SAKOWICZ, Seweryn FRASIŃSKI

Katedra Genetyki Ogólnej, Biologii Molekularnej i Biotechnologii Roślin,
Uniwersytet Łódzki

Streszczenie: Promotory to krótkie odcinki DNA niezbędne do precyzyjnego kierowania procesem inicjacji transkrypcji, pełniąc rolę sekwencji regulatorowej genów, same nie ulegają transkrypcji. Zazwyczaj zlokalizowane są powyżej sekwencji kodującej genów. Ich strukturę tworzy kombinacja elementów rozpoznawanych przez polimerazę RNA zależną od DNA oraz liczne, współdziałające z nią czynniki transkrypcyjne, które wiążąc się z DNA umożliwiają rozpoczęcie i kontynuację transkrypcji. Poznanie struktury promotorów wydaje się kluczowe dla opisu maszynarii transkrypcyjnej i wyjaśnienia mechanizmów regulacji ekspresji genów. Największe zainteresowanie badaczy wzbudzają promotory polimerazy RNA II odpowiedzialne za syntezę frakcji mRNA i microRNA. Wśród promotorów genów eukariotycznych znacznie lepiej scharakteryzowane są te pochodzące z genomów zwierzęcych. Aktualna wiedza o nich pozwala formułować pewne uogólnienia dotyczące budowy promotorów, ich modeli i zasad działania. Na tym tle, informacje o promotorach roślinnych, chociaż stale wzbogacane, są uboższe i w głównej mierze pochodzą z analiz genomowych sekwencji *Arabidopsis thaliana* i *Oryza sativa*, gatunków najlepiej charakteryzowanych na poziomie molekularnym. Analizy porównawcze, poza podobieństwami, ujawniają też szereg różnic świadczących o specyfice tych elementów wynikającej z odmiennych kierunków i tempa ewolucji genomów zwierząt i roślin. Odmienności strukturalne to głównie pochodna obecności lub braku pewnych motywów sekwencyjnych spotykanych w jednej, a nieobecnych w drugiej z grup. Do takich należą elementy: CpG, BRE, MTE, DPE, CCAAT-box czy Y-Patch. Pewne różnice dotyczące lokalizacji i sekwencji najwyższej zgodności (*consensus*) wskazano także dla elementów obecnych zarówno w promotorach genów zwierzęcych i roślinnych jak TATA-box czy Inr.

Słowa kluczowe: promotory roślinne, promotor rdzeniowy, miejsce inicjacji transkrypcji (TSS), czynniki transkrypcyjne, polimeraza RNA II

Summary: Promoters are short fragments of DNA that are necessary to create initiation complex and begin transcription. Typically, they are located upstream of the gene coding sequence. Their structure creates a combination of elements which are identified by RNA polymerase, and numerous transcription factors that bind to DNA, enabling start of transcription. Knowledge of the promoters facilitates understanding of the mechanisms of gene expression regulation, one of the most intriguing aspects of the genomes functioning. Main attention is focused on promoters recognized by RNA polymerase II, responsible for the synthesis of mRNA and microRNAs. Among eukaryotic promoters much better characterized ones are these associated with the genes of animals, the knowledge of the corresponding structures of the plant is significantly poorer and mainly comes from analysis of *Arabidopsis thaliana* and *Oryza sativa* genomes. Comparative analysis, besides similarities, demonstrates a number of differences that determine the specificity of these components which result from different directions and pace of the evolution of whole plant and animal genomes. Structural differences are an outcome of the presence or absence of certain sequence motifs (CpG, BRE, MTE, DPE, CCAAT-box or Y Patch).

Key words: plant promoters, core promoter, transcription start site (TSS), transcription factors, RNA polymerase II

WSTĘP

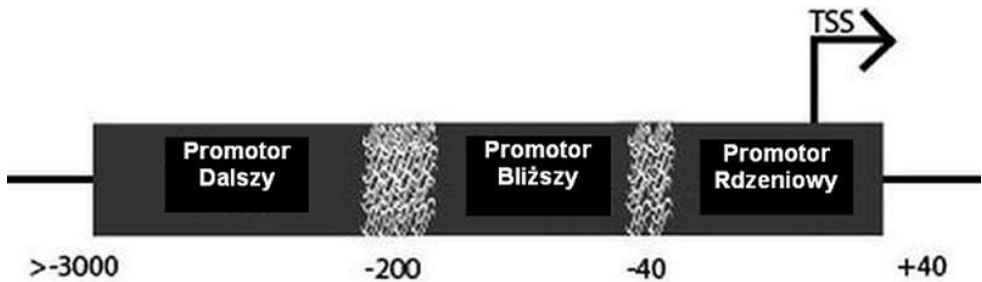
Punktem odniesienia w badaniach nad budową i funkcją promotorów roślinnych jest bogata wiedza o analogicznych strukturach w genomach zwierząt. Badania nad nimi rozpoczęto wcześniej stąd aktualna pula danych jest znacząco większa. Podczas identyfikacji charakterystycznych motywów w promotorach roślinnych poszukiwano sekwencji, które wcześniej opisano u *D. melanogaster*, *C. elegans*, *M. musculus* czy *H. sapiens* [8, 20, 42, 44]. To u nich po raz pierwszy wykazano obecność elementów: TATA-box, BRE, DPE, MTE, CCAAT, wyspy CpG czy Inr [9]. Wykorzystano też informacje o promotorach *S. cerevisiae* [27]. W efekcie, opisano szereg sekwencji specyficznych dla regionów promotorowych.

Charakteryzując promotory roślinne poszukiwano więc odpowiedzi na pytania czy różnią się one od promotorów zwierzęcych, jeśli tak to czym, czy zachowały elementy uznane zazachowawcze, czy pojawiły się w nich sekwencje specyficzne dla genomów roślinnych?

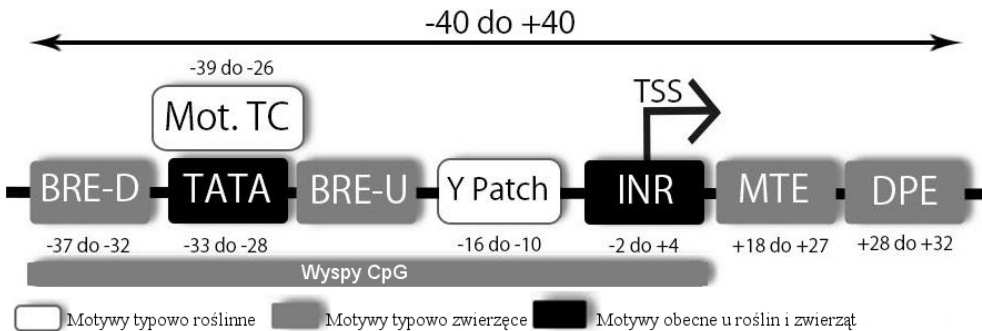
Korzystając z technik eksperymentalnych i metod obliczeniowych potwierdzono u roślin występowanie elementów TATA-box czy Inr. Brak natomiast dowodów na obecność u roślin innych jak: BRE, DPE, MTE czy CpG opisanych w promotorach genów zwierzęcych, co wydaje się najbardziej istotną różnicą w strukturze promotorów obu grup organizmów.

Promotory eukariotyczne wykazują złożoność i zróżnicowanie struktury. Osiągają długość ok. kilkuset par zasad. W tym obszarze wyróżniane są trzy części funkcjonalne: I – promotor rdzeniowy/minimalny (ang. *core promoter region*)

zlokalizowany ok. kilkadziesiąt (40-50) pz od miejsca startu transkrypcji (TSS), II – promotor bliższy (ang. *proximal promoter*) oraz III – promotor dalszy (ang. *distal promoter*), bogaty w motywy wiążące liczne czynniki regulatorowe (ryc. 1.). Kluczową rolę odgrywa rdzeniowa część promotorów skupiająca sekwencje krytyczne dla aktywności promotorów/genów. To swoista platforma dla tworzenia kompleksu preinicjacyjnego (PIC) z podstawowymi czynnikami transkrypcji (TF) i polimerazą RNA II (ryc. 2.). Strukturę promotorów tworzy kombinacja kilku podstawowych modułów reprezentowanych przez krótkie motywy sekwencyjne [39].



RYCINA 1. Podstawowe regiony promotorów eukariotycznych. TSS – miejsce startu transkrypcji
FIGURE 1. Basic regions of eukaryotic promoters. TSS – Transcription Start Site



RYCINA 2. Elementy rdzeniowego regionu promotorów. Kolorem białym zaznaczono elementy typowe dla promotorów roślin, szarym – elementy typowe dla zwierząt, czarnym – elementy obecne zarówno w promotorach roślin jak i zwierząt (pozostałe szczegóły w tekście)

FIGURE 2. Elements of core promoter regions White – plant-specific motifs, grey – animal-specific motifs, black – motifs present in both animals and plants

ELEMENTY PROMOTORÓW RDZENIOWYCH

○ **kaseta TATA** (ang. *TATA-box*) – rozpoznawana przez TBP (białka wiążące TATA, ang. *TATA-box Binding Protein*), tj. kilkanaście czynników TFIID. Kaseta często otoczona sekwencjami bogatymi w GC, obecna głównie w tzw. skupionym modelu promotora. Wcześniej uznawana za najbardziej zachowawczy element promotorów, obecnie wiadomo, że występują jedynie w ok. 10-20% promotorów, zarówno u roślin jak u zwierząt [32]. Błędy w szacunkach wynikały m.in. z faktu, że był to element łatwy w identyfikacji, a starsze opracowania dotyczyły mało licznych populacji genów [8, 53]. Obecność TATA najczęściej wiązana jest z promotorami o wyższej specyficzności tkankowej [35, 51]. Lokalizowane w pozycjach -35 do -25 wobec TSS. Analizy korelacji między obecnością TATA, a innymi cechami genów sugerują powiązania z długością elementu 5'UTR. W promotorach z TATA u *A. thaliana* średnia długość 5'UTR to 108 pz, ten sam parametr w promotorach bez TATA osiąga 138 pz [35]. Analiza regionów flankujących miejsce TSS nie wykazała istotnych różnic statystycznych w składzie nukleotydów promotorów z TATA i bez nich [32]. Geny z promotorami TATA wykazują pozytywną korelację z poziomem ekspresji [52]. Promotory pozbawione TATA powiązane są z genami metabolizmu podstawowego, podczas gdy te zawierające TATA podlegają regulacji m.in. przez bodźce stresu czy czynniki biotyczne [56].

○ **Inr** (ang. *Initiator*) – bogata w pirymidyny krótka sekwencja obejmująca miejsce inicjacji transkrypcji (TSS), rozpoznawana przez białka TFIID. Element najwyraźniej zaznaczony w modelu promotorów skupionych. Analizy Inr są szczególnie istotne ze względu na obecność w nim miejsca TSS. Pojawia się tu kilka motywów sekwencyjnych: YR, YYR, TCA, TCA_Y, YTCA_Y (podkreślono TSS, +1), a adenina jest najczęstszym miejscem startu transkrypcji. Tzw. reguła YR, zaproponowana na podstawie analiz sekwencji *A. thaliana* i *O. sativa*, dotyczy ok. 80% promotorów roślinnych [50], obserwowana jest też u ssaków [11,16].

○ **DPE** (ang. *Downstream Promoter Element*) – sekwencja rozpoznawana przez białka TAF kompleksu TFIID, zlokalizowana poniżej miejsca TSS w regionie +28 do +33. Identyfikowana w obszarze rdzeniowych promotorów *D. melanogaster*, drożdży, człowieka. Częstość jego występowania jest porównywalna do TATA box. Niekiedy traktowany jako substytut kasety TATA w promotorach, które utraciły ten element [22].

○ **Y-Patch** – krótki region bogaty w pirymidyny bezpośrednio poprzedzający Inr. Wydaje się być elementem charakterystycznym dla części promotorów roślinnych. Identyfikowany za pomocą różnego typu analiz (LDSS, MEME, AlignACE) [51]. Brak dostatecznej puli informacji na temat jego funkcjonalności. Obecny w gatunkach roślin jedno- i dwuliściennych. Obserwowano lekką pozytywną korelację między jego obecnością w promotorach, a poziomem transkrypcji. Obec-

ność Y-Patch jest niezależna od występowania elementu TATA, niemniej wspólnie z nimi wykazuje kooperatywne działanie i wpływa na wysoką aktywność promotorów [52]. Można mówić o pozytywnej kolokalizacji elementów TATA, Y-Patch i Inr, w przeciwieństwie do promotorów bogatych w motywy GA [43]. Promotory zawierające Y-Patch nie wykazują znaczących preferencji wobec genów określonej kategorii. Liczba wykrytych promotorów z Y Patch znacząco rośnie wraz z precyzyjnie określonym regionem Inr [52]. Y-Patch występuje w pozycji -13 wobec TSS (uśredniona lokalizacja w regionie -20 do -6). U *A.thaliana* Y Patch przyjmuje najczęściej postać TTCTTC i TTCTCT [50].

○ **MTE** (ang. *Motif Ten Element*) dziesięcionukleotydowy motyw między Inr a DPE (+18 do +27), wzmacnia powinowactwo czynnika TFIID wobec promotora. Sekwencja konsensusowa dla promotorów zwierzęcych to CSARCSSAAC [37]. Wykazuje działanie synergistyczne z TATA i DPE [28]. U roślin brak danych o obecności takiego motywu w promotorach.

○ **BRE** (ang. *TFIIB Recognition Element*) – sekwencja bogata w GC zlokalizowana powyżej (ang. *upstream*) miejsca TSS – rozpoznawana przez TFIIB. Wskazano dwie lokalizacje BRE, częstsza, przed sekwencją TATA (*BRE^u upstream*) i rzadszą (*BRE downstream*). Odpowiedzialna jest za tworzenie kompleksu TFIIB-TBP. *Consensus BRE upstream* -SSRCGCC, dla *BRE downstream* – RTDKKKK [21]. *BRE upstream* ma pozytywny efekt na działanie promotora [14]. U roślin nie wykazano analogicznych elementów.

○ **DCE** (*Downstream Core Element*) – położony za TSS, uczestniczy w wiązaniu TAF1. Działa w połączeniu z TATA oraz Inr jak i w promotorach pozbawionych motywu TATA [28].

○ **Wyspy CpG** – typowe dla promotorów genów zwierzęcych, częsty motyw rozproszonych modeli promotorów [13, 15]. Powszechne w *housekeeping genes* tj. genach niezbędnych w utrzymaniu podstawowych funkcji życiowych komórki. Blisko połowa promotorów ludzkich leży w obszarze wysp CpG [41]. Możliwe, że odpowiednikiem odcinków bogatych w CpG promotorów zwierzęcych są u roślin regiony bogate w motyw GA [53].

Wymienione wyżej elementy pojawiają się jedynie w części promotorów [8]. I tak, zarówno u zwierząt jak i roślin identyfikowano TATA-box i Inr. Niektóre z baz danych (np. AGRIS, Softberry), deponujące informacje na temat tysięcy promotorów roślinnych grupują je m.in. właśnie pod kątem obecności lub braku w nich tych elementów. W świetle aktualnych wyników wydaje się, że za najbardziej typowy motyw promotorów roślinnych można uznać Y-Patch obecny w regionie rdzeniowym czy motyw CCAAT w obszarze promotora bliższego. Jednocześnie, w strukturze promotorów roślinnych nie odnaleziono takich elementów jak: BRE, MTE, DPE, co należy uznać za kolejną cechę odróżniającą je od promotorów zwierzęcych (ryc.2).

TYPY PROMOTORÓW

Promotory genów eukariotycznych wyróżnia duża złożoność strukturalna. Jednym z kryteriów klasyfikacji promotorów jest liczba i charakter miejsc startu transkrypcji (TSS) w promotorze rdzeniowym [21]. Biorąc pod uwagę ten parametr wyróżnia się dwa modele:

- **Skupione, zwarte** (ang. *focused*) – region z jednym, wyraźnie wyodrębnionym miejscem TSS lub klastrami złożonymi z kilku nukleotydów (pozycje -3 do +3) w tym jednym, zdecydowanie dominującym (region Inr). Taki model spotykany jest w promotorach regulowanych [22], starszy ewolucyjnie od rozproszonych [21, 27], zapewnia inicjację transkrypcji w odpowiednim miejscu i czasie.

- **Rozproszone** (ang. *dispersed*) zawierają liczne TSS ze słabo zaznaczonym nukleotydem inicjującym, rozproszone na dystansie 50-100 pz (u kręgowców w regionach CpG-bogatych) [2]. Promotory tego typu (klasy) są trudniejsze do identyfikacji, słabiej zbadane niż wyżej opisane. Są kojarzone z genami ulegającymi konstytutywnej ekspresji konstytutywnymi [22]. Mniej wiadomo o mechanizmach regulacji transkrypcji w tego typu promotorach. Charakteryzuje je deficyt sekwencji ATG („*ATG desert*”). Region *ATG-desert* to unikatowa cecha genów kodujących mRNA, transkrybowanych przez polimerazę RNA II, podobnej prawidłowości nie obserwowano w otoczeniu genów rRNA, snoRNA czy tRNA. Większość genów eukariotycznych inicjuje transkrypcję z licznych miejsc TSS [23]. Ten model wiązany jest z brakiem elementu TATA.

Kryterium stosowanym do klasyfikacji promotorów jest też fakt obecności lub braku w nich sekwencji TATA-box (odpowiednio ang. *TATA-containing*, *TATA-less* lub *non-TATA promoters*) [16]. Mnogość promotorów pozbawionych TATA zweryfikowała pierwotną wiedzę na temat ich architektury, była też sygnałem iż regulacja ekspresji nie ma charakteru uniwersalnego dla wszystkich genów. U kręgowców promotory pozbawione TATA określa się jako model CpG [8]. Możliwe, że ich odpowiednikiem u roślin jest typ opisany jako GA u *A. thaliana*. Konstytutywny charakter ekspresji genów związanych z takimi promotorami nie jest jednak tak wyraźny jak w modelu CpG [53]. Wśród wielu promotorów roślinnych pozbawionych TATA w regionie w regionie odpowiadającym jego zwykłej lokalizacji pojawiają się trakty mikrosatelitarne (TC)_n czy (TTC)_n (u *A. thaliana* >18% promotorów). Analogiczna analiza w promotorach zwierzęcych nie ujawniła podwyższonej obecności TC. U roślin mogą one stanowić klasę regulatorowych elementów modulujących ekspresję genów [5]. Struktura promotora powiązana jest też z organizacją genu. Geny z promotorem TATA charakteryzuje bardziej „kompaktowa” budowa niż tych bez TATA (liczba eksonów, długość intronów, rozmiary mRNA) [53].

Odmienne mechanizmy regulacji transkrypcji umożliwiły grupowanie promotorów na podstawie podobieństw lub odmienności ich działania. Szczególne zainteresowanie w kontekście aplikacyjnym budzą specyficzne promotory aktywne

w wybranych tkankach, warunkach, stadiach rozwojowych także te o silnej, nie-swoistej ekspresji we wszystkich organach. Każdą wyróżniają charakterystyczne cechy jednak elementy decydujące o specyficzności organowej czy tkankowej zlokalizowane są raczej poza regionem promotorów rdzeniowych [17, 24, 57].

Promotory grupować można również wg innych niż wymienione cechy. I tak, na podstawie częstości występowania wybranych motywów wyodrębniono modele: *AT-based*, *CG-based*, *ATCG-compact*, *ATCG-balanced*, *ATCG-middle*, *ATCG-less*, *AT-less*, *CG-less*, *AT-spike*, *CG-spike*. Trudno uznać taki podział za powszechnie stosowany, niemniej ujawnia on pewne ciekawe różnice w budowie promotorów. Dotyczą one zarówno promotorów roślinnych (*A. thaliana* i *O. sativa*) jak zwierzęcych (*D. melanogaster* i *H. sapiens*), pozwalają też na wskazanie podobieństw i różnic między. U *A. thaliana* szczególnie częste okazały się modele: *AT-based*, *ATCG-compact* i *AT-spike* zaś niemal nieobecne *CG-based*, *AT-less* oraz *ATCG-middle*, *less*. Analogiczna analiza u *O. sativa* jako najczęstsze wskazuje *CG-spike* i *ATCG-balanced*, a nieobecne to *AT-less*, *ATCG-middle* *ATCG-less*. Co ciekawe, nieliczne modele promotorów były wspólne dla obu wymienionych gatunków roślin (których szlaki ewolucyjne rozdzieliły się ok. 150 mln lat temu) [11]. Gagniuc i Ionescu-Tirgoviste, autorzy proponowanych wyżej modeli tworzyli je na podstawie analiz sekwencji ponad 20 tysięcy promotorów pobranych z różnych baz danych. Poczynili obserwacje dotyczące obecności lub braku określonego motywu w każdym z modeli, zwrócili też uwagę na pewne podobieństwa i różnice związane z wybranym gatunkiem. I tak zauważają, że profil modelu wynikający z uwzględnienia zawartości par GC promotory *O. sativa* wykazywały więcej podobieństw strukturalnych wobec promotorów *H. sapiens* niż wobec analogicznych regionów *A. thaliana*. Te ostatnie, wykazały za to liczne podobieństwa wobec promotorów *D. melanogaster*. Niestety, autorzy nie podejmują w tym opracowaniu próby skomentowania swoich obserwacji w kontekście ewolucyjnym, zwracając raczej uwagę na kwestie doboru odpowiednich danych uwzględniających specyfikę baz, z których pozyskiwali sekwencje do analiz (m.in. TiGER -*Tissue-specific Gene Expression*, EPD – *Eukaryotic Promoter Database*, *PlantProm DB*) [18].

Kolejna (lub: „inna”) klasyfikacja wyróżnia trzy grupy promotorów. I – z klasyczną wersją kasety TATA, II – z licznymi wariantami motywu TATA, III – z elementami TC obecnymi w regionie -39 do -26, gdzie w innych promotorach umiejscowione są elementy TATA i BRE [21]. W ponad połowie analizowanych sekwencji *A. thaliana* i *O. sativa* nie wykazano znaczącej statystycznie przewagi któregokolwiek z trzech wymienionych modeli, zaś 18% promotorów *A. thaliana* zawierało elementy TC (inne niż Y Patch). Analogiczne badania populacji promotorów genów człowieka i myszy nie ujawniły podwyższonej częstości TC, co pozwala widzieć w nich element charakterystyczny dla grupy promotorów roślinnych. Strukturalne i funkcjonalne cechy genów zawierających w swoich promotorach opisywany element są odmienne od genów z promotorami typu TATA. Geny z promotorami TC

są zwykle dłuższe i cechuje je specyficzność ekspresji, zaś geny zawierające kasetę TATA przeważnie są niewielkie i wykazują wysoki poziom ekspresji. Elementy TC obecne w regionie -39 do -26 mogą stanowić klasę elementów regulatorowych modulujących ekspresję genów roślinnych [5]. Lokalizacja TC pozwala widzieć w nich odpowiednik elementu BRE częstego w strukturze promotorów zwierzęcych i rozpoznawanego przez TFIIB [14].

Identyfikacja i klasyfikacja promotorów na podstawie obecności takich elementów jak: TATA-box, GC-box, CCAAT-box, BRE, TC czy INR nie jest zadaniem łatwym. Stąd brane są pod uwagę również inne cechy omawianych regionów. I tak badana jest: podatność na zgięcia, tworzenie krzywizn, stabilność DNA. Wartość wymienionych, fizycznych parametrów DNA w obszarze promotorów i sąsiadujących z nimi sekwencji o innym charakterze, wykazuje znaczące różnice co leży u podstaw metod pozwalających na identyfikację i wyróżnienie regionów promotorowych na tle pozostałych elementów genomu [1, 25, 59, 33].

IDENTYFIKACJA REGIONÓW PROMOTOROWYCH I PERSPEKTYWY ICH ANALIZ

Spektrum metod umożliwiających identyfikację promotorów, obok technik molekularnych, obejmuje też zaawansowane techniki obliczeniowe. Początkowo, metody identyfikacji opierały się na odnajdywaniu motywów uznanych za charakterystyczne dla promotorów (TATA, CpG). Występują one jedynie w ograniczonej puli promotorów, co wymusiło poszukiwania bardziej uniwersalnych metod.

Obok badań sekwencji nukleotydowych niezwykle przydatne są też analizy czynników transkrypcyjnych (TF), które wskazują miejsca oddziaływań z wybranymi sekwencjami genomu. Liczne bazy deponują informacje o sekwencjach promotorów i związanych z nimi czynnikami transkrypcyjnymi i regulatorowymi (PLACE, PlantCARE, PlantProm, TRANSFAC, TRRD DB, PlantPAN, MAPPER, JASPAR, Grassius, LegumeTFDB, Plant TF DB) [10, 12, 26, 29, 36, 40, 58]. Niektóre bazy dotyczą konkretnych gatunków roślin (DATF i AGRIS dla *A. thaliana*, DRTF dla *O. sativa*, DPTF dla *P. trichocarpa*, SoyDB dla soi) [30, 48]. U roślin stosunkowo najlepiej opisano motywy sekwencyjne rozpoznawane przez TF *A. thaliana*, powstają też bazy zawierająca charakterystyki czynników (kilkanaście tys. rodzin) dla gatunków o dużym znaczeniu gospodarczym (winorośl, topola, maniok – TreeTF database) [31]. Wskazanie najliczniejszych rodzin czynników (np. bHLH, PHD, AP2_ERE BP) obecnych w szerszym spektrum gatunków wraz z rozpoznawanymi przez nie motywami sekwencyjnymi (np. E box – CANNTG i G-box – CACGTG dla TF bHLH) stanowi przesłankę w definiowaniu konserwatywnych motywów promotorów [31]. Do programów ukierunkowanych na identyfikację miejsc wią-

zania czynników transkrypcyjnych należą: CONREAL, PromH, Reista, R-Motif, TFBIND [45]. Wśród narzędzi o charakterze bioinformatycznym wymienić warto te, które przy pomocy algorytmów pozwalają na przyrównania multisekwencyjne (ClustalW, Lagan, AVID, Bays-Block Aligner, BlastZ, MEME, Consensus czy Gibbs Sampler z grupy Markova) [46]. Badając wybrane cechy sekwencji, analizować można lokalny rozkład krótkich motywów nukleotydowych (met. LDSS) między regionami bogatymi i pozbawionymi genów [50]. Zwykle określa się częstość motywów tetra- do heksamerowych [3, 4]. Pod uwagę brane są też specyficzne cechy drugorzędowych struktur DNA jak np. analiza krzywizn wynikająca z sekwencji danego fragmentu (program BEND) czy wielkość względnej entropii regionów promotorowych wyróżniających je spośród innych fragmentów genomu [35].

W sytuacji gdy duża część promotorów nie zawiera elementu TATA, kluczowe okazało się precyzyjne wskazanie regionu Inr z miejscem (lub miejscami TSS, co umożliwiałoby zdefiniowanie czynników transkrypcyjnych zaangażowanych w inicjację i regulację transkrypcji. Jedną z większych baz deponujących podobne dane jest DPTSS [54].

Równoległe z analizami *in silico* opracowywane metody eksperymentalne (w tym liczne modyfikacje metody PCR). Część technik nastawiona jest na wskazanie preferowanych lokalizacji motywów wiążących TF (np. DNA *footprinting* w żelach denaturujących czy *gel-shift* w warunkach niedenaturujących obrazująca różnice w prędkości migracji fragmentów DNA powiązanych z białkami i wolnych od nich). Niezwykle przydatne w poszukiwaniach i analizach promotorów są pule cDNA pełnej długości (*full-length cDNA*) uzyskiwane m.in. za pomocą szybkiej amplifikacji końców cDNA (RACE) [7]. Inne metody nastawione są na identyfikację miejsc TSS. Do ich mapowania wykorzystuje się np. nukleazę S1 (*S1 protection assay*) i różnice w migracji dupleksów traktowanych i nietraktowanych tą nukleazą. Również modyfikacje metody *primer extension* wykorzystują różnice w migracji hybrydowych dupleksów do wskazania obecności/braku w ich obszarze miejsc TSS [55].

W omawianych analizach pomocne są też techniki SAAB (ang. *Selected And Amplified Binding site*) czy SELEX (ang. *Systematic Evolution of Ligands by Exponential Enrichment*), rodzaje chromatografii powinowactwa połączonej z reakcją PCR [1]. Dany TF wiązany jest na kolumnie, a ta przemywana jest mieszaniną olii gonukleotydów losowo generowanych sekwencji. Fragmenty związane do złoża są eluowane. Procedura powtarzana jest do chwili uzyskania puli jednorodnych sekwencji wiązanych przez badany czynnik transkrypcyjny [34].

Różnorodności stosowanych metod sprawia, że skuteczność identyfikacji omawianych regionów genomów stale rośnie. Jednocześnie, głównie w metodach *in silico*, wzrasta prawdopodobieństwo przypadkowego odnajdywania pewnych motywów i niezbędna staje się weryfikacja za pomocą metod eksperymentalnych. Trudność stanowi m.in. precyzyjna lokalizacja regionu/ów Inr i miejsc TSS. Istotną jest nie tylko sama sekwencja DNA, ale też struktura i stan chromatyny w regionie

Inr/TSS (metylacja, acetylacja) [47]. Badania nad czynnikami TF i oddziałujących z nimi motywów mogą wzbogacić wiedzę na temat struktury promotorów, ale także wyjaśnić mechanizmy regulacji transkrypcji. Poznano jedynie część miejsc wiązania, a niektóre z nich mogą wiązać różne czynniki w różnych typach komórek czy stadiach rozwojowych organizmów [31, 49]. Problemem w jednoznacznej identyfikacji miejsc wiązania TF wynika z faktu, że zwykle są to bardzo krótkie sekwencje i pojawiają się w licznych miejscach genomów, nie tylko tych bezpośrednio sąsiadujących z genami. Sam fakt istnienia miejsca wiązania dla TF nie musi oznaczać też, że dochodzi tam do jego interakcji z DNA. Wydaje się, że lepsze perspektywy stoją przed globalnymi przyrównaniami multisekwencyjnymi niż częstszymi dzisiaj rodzajami przyrównań lokalnych [6].

Wymienione trudności wskazują, że już kwestia identyfikacji omawianych elementów genomu pozostaje poważnym wyzwaniem naukowym wymagającym wielokierunkowych badań, a stanowi ona jedynie wstęp do wyjaśnienia interakcji pomiędzy elementami regulatorowymi promotorów i wiążących się z nimi białkami. Bogatsza wiedza na ten temat może przyczynić się do rozwikłania niewiadomych związanych z mechanizmami regulacji ekspresji genów czy ustalenia warunków w jakich konkretne geny ulegają procesowi aktywacji.

PODZIĘKOWANIA

Autorzy dziękują mgr. Adamowi Janowskiemu za opracowanie i wykonanie rytciny zamieszczonej w niniejszej pracy. Praca finansowana z umowy nr **506/041157**

LITERATURA

- [1] ABEEL T, SAEYS Y, BONNET E, ROUZÉ P, VAN DE PEER Y. Generic eukaryotic core promoter prediction using structural features of DNA. *Genome Res* 2008; **18**: 310-23.
- [2] ANISH R, HOSSAIN MB, JACOBSON RH, TAKADA S. Characterization of transcription from TATA-less promoters: identification of a new core promoter element XCPE2 and analysis of factor requirements. *PLoS One* 2009; **4**: e5103.
- [3] ANWAR F, BAKER SM, JABID T, MEHEDI HASAN M, SHOYAIB M, KHAN H, WALSH R. Pol II promoter prediction using characteristic 4-mer motifs: a machine learning approach. *BMC Bioinformatics* 2008; **9**: 414.
- [4] AZAD AK, SHAHID S, NOMAN N, LEE H. Prediction of plant promoters based on hexamers and random triplet pair analysis. *Algorithms Mol Biol* 2011; **6**: 19.
- [5] BERNARD V, BRUNAUD V, LECHARNY A. TC-motifs at the TATA-box expected position in plant genes: a novel class of motifs involved in the transcription regulation. *BMC Genomics* 2010; **11**: 166.
- [6] BLANCO E, MESSEGUER X, SMITH TF, GUIGÓ R. Transcription factor map alignment of promoter regions. *PLoS Comput Biol* 2006; **2**: e49.

- [7] BOWER NI, JOHNSTON IA. Targeted rapid amplification of cDNA ends (T-RACE) – an improved RACE reaction through degradation of non-target sequences. *Nucl Acids Res* 2010; **38**: e194.
- [8] BUTLER JE, KADONAGA JT. The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes Dev* 2002; **16**: 2583-92.
- [9] CARNINCI P. Genome-wide analysis of mammalian promoter architecture and evolution. *Nat genetics* 2006, **38**, 6.
- [10] CHANG WC, LEE TY, HUANG HD, HUANG HY, PAN RL. PlantPAN: Plant promoter analysis navigator, for identifying combinatorial cis-regulatory elements with distance constraint in plant gene groups. *BMC Genomics* 2008; **9**: 561.
- [11] CHAW SM, CHANG C, CHEN HL, LI WH. Dating the monocot-dicot divergence and the origin of core eudicots using whole chloroplast genomes. *J Mol Evol* 2004; **58**: 424-41.
- [12] DAVULURI RV, SUN H, PALANISWAMY SK, MATTHEWS N, MOLINA C, KURTZ M, GROTEWOLD E. BMC Bioinformatics. 2003; 4: 25. AGRIS: Arabidopsis gene regulatory information server, an information resource of Arabidopsis cis-regulatory elements and transcription factors. *BMC Bioinformatics* 2003; **4**: 25.
- [13] DEATON A, BIRD A. CpG islands and regulation of transcription. *Genes Dev* 2011; **25**:1010-22.
- [14] DENG W, ROBERTS SG. Core promoter elements recognized by transcription factor IIB. *Biochem Soc Trans* 2006; **34**: 1051-53.
- [15] FENOUIL R, CAUCHY P, KOCH F, DESCOSTES N, SPICUGLIA S, GUT M, GUT I, ANDRAU Jc. CpG islands and GC content dictate nucleosome depletion in a transcription-independent manner at mammalian promoters. *Genome Res* 2012; **22**: 2399-408.
- [16] FITZGERALD PC, STURGILL D, SHYAKHTENKO A, OLIVER B, VINSON C. Comparative genomics of *Drosophila* and human core promoters. *Genome Biol* 2006; **7**: R53.
- [17] FAUTEUX F, STRÖMVIK MV. Seed storage protein gene promoters contain conserved DNA motifs in Brassicaceae, Fabaceae and Poaceae. *BMC Plant Biol* 2009; **9**: 126.
- [18] GAGNIUC P, IONESCU-TIRGOVISTE C. Eukaryotic genomes may exhibit up to 10 generic classes of gene promoters. *BMC Genomics* 2012; **13**: 512.
- [19] GAN Y, GUAN J, ZHOU S. A comparison study on feature selection of DNA structural properties for promoter prediction. *BMC Bioinformatics* 2012; **13**: 4.
- [20] JUVEN-GERSHON T, CHENG S, KADONAGA JT. Rational design of a super core promoter that enhances gene expression. *Nat Methods* 2006; **3**: 917-922.
- [21] JUVEN-GERSHON T, HSU J, THEISEN JM, KADONAGA JT. The RNA Polymerase II Core Promoter – the Gateway to Transcription. *Curr Opin Cell Biol* 2008 ; **20**: 253-259.
- [22] JUVEN-GERSHON T, KADONAGA JT. Regulation of gene expression via the core promoter and the basal transcriptional machinery. *Dev Biol* 2010; **339**: 225-9.
- [23] LEE MP, HOWCROFT K, KOTEKAR A, YANG HH, BUETOW KH, SINGER DS. ATG deserts define a novel core promoter subclass. *Genome Res* 2005; **15**: 1189-1197.
- [24] LIM CJ, LEE HY, KIM WB, LEE BS, KIM J, AHMAD R, KIM HA, KWON SY. Screening of tissue-specific genes and promoters in tomato by comparing genome wide expression profiles of Arabidopsis orthologues. *Mol Cells* 2012; **34**: 53-9.
- [25] KANHERE A, BANSAL M. Structural properties of promoters: similarities and differences between prokaryotes and eukaryotes. *Nucl Acids Res* 2005; **33**: 3165-75.
- [26] LESCOT M, DÉHAIS P, THIJS G, MARCHAL K, VAN DE PEER Y, ROUZÉ P, ROMBAUTS S. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for *in silico* analysis of promoter sequences. *Nucl Acids Res* 2002; **30**: 325-7.
- [27] LUBLINER S, KEREN L, SEGAL E. Sequence features of yeast and human core promoters that are predictive of maximal promoter activity. *Nucl Acids Res* 2013; **41**: 1-13.
- [28] MASTON GA, EVANS SK, GREEN MR. Transcriptional regulatory elements in the human genome. *Annu Rev Genomics Hum Genet* 2006; **7**: 29-59.

- [29] MATYS V, KEL-MARGOULIS OV, FRICKE E, LIEBICH I, LAND S, BARRE-DIRRIE A, REUTER I, CHEKMENEV D, KRULL M, LEWICKI-POTAPOV B, SAXEL H, KEL AE, WINGENDER E. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucl Acids Res* 2006; **34**: D108-10.
- [30] MOCHIDA K, YOSHIDA T, SAKURAI T, YAMAGUCHI K, SHINOZAKI K, TRAN LS. In silico analysis of transcription factor repertoire and prediction of stress responsive transcription factors In soybean. *DNA Res* 2009; **16**: 353-69.
- [31] MOCHIDA K, YOSHIDA T, SAKURAI T, YAMAGUCHI-SHINOZAKI K, TRAN LS. TreeTFDB: An Integrative Database of the Transcription Factors from Six Economically Important Tree Crops for Functional Predictions and Comparative and Functional Genomics. *DNA Res* 2013; **20**: 151-62.
- [32] MOLINA C, GROTEWOLD E. Genome wide analysis of Arabidopsis core promoters. *BMC Genomics* 2005; **6**: 25.
- [33] MOREY C, MOOKHERJEE S, RAJASEKARAN G, BANSAL M. DNA free energy-based promoter prediction and comparative analysis of Arabidopsis and rice genomes. *Plant Physiol* 2011; **156**: 1300-15.
- [34] MUIÑO JM, ANGENENT G, KAUFMANN K. Visualizing in vivo DNA-binding events and direct target genes of plant transcription factors. *Methods Mol Biol* 2011; **754**: 293-305.
- [35] PANDEY SP, KRISHNAMACHARI A. Computational analysis of plant RNA Pol-II promoters. *Biosystems* 2006; **83**: 38-50.
- [36] PORTALES-CASAMAR E, THONGJUEA S, KWON Z, ARENILLAS D, VALEN E, YUSUF D, WASSERMAN WW, SANDELIN A. JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. *Nucl Acids Res* 2010; **38**: D105-10.
- [37] RACH EA, YUAN HY, MAJOROS WH, TOMANCAK P, OHLER U. Motif composition, conservation and condition-specificity of single and alternative transcription start sites in the Drosophila genome. *Genome Biol* 2009; **10**: R73.
- [38] REISS DJ, HOWARD FM, MOBLEY HL. A novel approach for transcription factor analysis using SELEX with high-throughput sequencing (TFAST) *PLoS One* 2012; **7**: e42761.
- [39] RHEE HS, PUGH BF. Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature* 2012; **483**: 295-301.
- [40] RIVA A. The MAPPER2 Database: a multi-genome catalog of putative transcription factor binding sites. *Nucl Acids Res* 2012; **40**: D155-61.
- [41] ROZENBERG JM, SHLYAKHTENKO A, GLASS K, RISHI V, MYAKISHEV MV, FITZGERALD PC, VINSON C. All and only CpG containing sequences are enriched in promoters abundantly bound by RNA polymerase II in multiple tissues. *BMC Genomics* 2008; **9**: 67.
- [42] SANDELIN A., CARNINCI P, LENHARD B, PONJAVIC J, HUME DA. Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nat Reviews* 2007; **8**(6): 424-436.
- [43] SEIZL M, HARTMANN H, HOEG, KURTH F, SÖDING J, CRAMER P. A conserved GA element in TATA-less RNA polymerase II promoters. *PLoS One* 2011; **6**: e27595.
- [44] SMALE ST, KADONAGA JT. The RNA polymerase II core promoter. *Annu Rev Biochem* 2003; **72**: 449-79.
- [45] SUN H, YUAN Y, WU Y, LIU H, XIE H. Tmod: toolbox of motif discovery. *Bioinformatics* 2010; **26**: 405-407.
- [46] THUIS G, MARCHAL K, LESCOT M, ROMBAUTS S, DE MOOR B, ROUZÉ P, MOREAU Y. A Gibbs sampling method to detect overrepresented motifs in the upstream regions of coexpressed genes. *J Comput Biol* 2002; **9**: 447-64.
- [47] WANG J, ZHUANG J, IYER S, LIN X, WHITFIELD TW, GREVEN MC, PIERCE BG, DONG X, KUNDAJE A, CHENG Y, RANDO OJ, BIRNEY E, MYERS RM, NOBLE WS. Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Res* 2012; **22**: 1798-812.
- [48] WANG Z, LIBAULT M, JOSHI T, ET AL. SoyDB: a knowledge database of soybean transcription factors. *BMC Plant Biol* 2010; **10**: 14.
- [49] WHITFIELD TW, WANG J, COLLINS PJ, PARTRIDGE EC, ALDRED SF, TRINKLEIN ND, MYERS RM. Functional analysis of transcription factor binding sites in human promoters. *Genome Biol* 2012; **13**: R50.

- [50] YAMAMOTO YY, ICHIDA H, MATSUI M, OBOKATA J, SAKURAI T, SATOU M, SEKI M, SHINOZAKI K, ABE T. Identification of plant promoter constituents by analysis of local distribution of short sequences. *BMC Genomics* 2007; **8**: 67.
- [51] YAMAMOTO YY, ICHIDA H, ABE T, SUZUKI Y, SUGANO S, OBOKATA J. Differentiation of core promoter architecture between plants and mammals revealed by LDSS analysis. *Nucl Acids Res* 2007; **35**: 6219-26.
- [52] YAMAMOTO YY, YOSHITSUGU T, SAKURAI T, SEKI M, SHINOZAKI K, OBOKATA J. Heterogeneity of Arabidopsis core promoters revealed by high-density TSS analysis. *Plant J* 2009; **60**: 350-62.
- [53] YAMAMOTO YY, YOSHIOKA Y, HYAKUMACHI M, OBOKATA J. Characteristics of core promoter types with respect to gene structure and expression in *Arabidopsis thaliana*. *DNA Res* 2011; **18**: 333-42.
- [54] YAMASHITA R, SUGANO S, SUZUKI Y, NAKAI K. DBTSS: DataBase of Transcriptional Start Sites progress report in 2012. *Nucl Acids Res* 2012; **40**: D150-4.
- [55] YANAGAWA T, KOGA H. Modified multiple primer extension method. *Methods Mol Biol* 2009; **578**: 425-35.
- [56] YANG C, BOLOTIN E, JIANG T, SLADEK FM, MARTINEZ E. Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. *Gene* 2007; **389**: 52-65.
- [57] YE R, ZHOU F, LIN Y. Two novel positive cis-regulatory elements involved in green tissue-specific promoter activity in rice (*O.sativa* L). *Plant Cell Rep* 2012; **31**: 1159-72.
- [58] YILMAZ A, NISHIYAMA, MY JR, FUENTES B, ET AL. GRASSIUS: a platform for comparative regulatory genomics across the grasses. *Plant Physiol* 2009; **149**: 171-80.
- [59] ZUO YC, LI QZ. Identification of TATA and TATA-less promoters in plant genomes by integrating diversity measure, GC-Skew and DNA geometric flexibility. *Genomics* 2011; **97**: 112-20.

Redaktor prowadzący – Michał Nowicki

Otrzymano: 28.11.2013

Przyjęto: 03.02.2014

Tomasz Sakowicz

Katedra Genetyki Ogólnej, Biologii Molekularnej i Biotechnologii Roślin

Wydział Biologii i Ochrony Środowiska, Uniwersytet Łódzki

Banacha 12/16, 90-237 Łódź

tel.: 42 6354428

e-mail: tomesakowicz@wp.pl

