**Struktura genu eukariotycznego**

**Materiały do zajęć *Genetyka* na Wydziale Biologii UW (v. 2014)**

Schemat ekspresji informacji genetycznej w zwierzęcej komórce eukariotycznej przedstawiono na poniższym schemacie:



**W jądrze komórek eukariotycznych występują trzy polimerazy RNA:**

**- polimeraza RNA I** – transkrybuje większość genów rRNA (lokalizacja: jąderko)

- **polimeraza RNA II** – transkrybuje wszystkie geny kodujące białka i niektóre geny snRNA (lokalizacja: nukleoplazma)

- **polimeraza RNA III** – transkrybuje geny tRNA, 5S rRNA U6 snRNA oraz geny kilku innych klas małych RNA (lokalizacja: nukleoplazma)

Celem niniejszych ćwiczeń jest prezentacja i omówienie poszczególnych elementów struktury jądrowego **genu eukariotycznego transkrybowanego przez polimerazę RNA II**. Ponadto, omówione zostaną niektóre aspekty budowy promotora, inicjacji i regulacji transkrypcji oraz potranskrypcyjnych etapów ekspresji genu eukariotycznego wraz z elementami strukturalnymi, warunkującymi ich właściwy przebieg. Poniższy schemat w sposób uogólniony i uproszczony przedstawia strukturę takiego genu.



**Regulacja ekspresji genów eukariotycznych na poziomie transkrypcji**

Regulacja ekspresji genów eukariotycznych zachodzi przede wszystkim na poziomie transkrypcji. Głównym transkrypcyjnym mechanizmem regulacji jest inicjacja transkrypcji kontrolowana przez oddziaływanie kompleksu polimerazy RNA II z promotorem genu. Promotory genów transkrybowanych przez polimerazę RNA II stanowią mniej lub bardziej rozległe obszary położone powyżej (po stronie 5’) miejsc inicjacji transkrypcji i zawierają bardzo rozmaite **elementy *cis***. Niektóre z nich umożliwiają prawidłowe rozpoczęcie transkrypcji i odpowiadają za podstawowy poziom mRNA, inne mają znaczenie regulacyjne - aktywują bądź (rzadziej) hamują ekspresję danego genu w danej sytuacji.

W odróżnieniu od enzymów prokariotycznych, polimerazy RNA *Eukaryota* nie są w stanie same rozpoznać i związać się z DNA. Za rozpoznanie sekwencji promotorowych i uformowanie funkcjonalnego kompleksu transkrypcyjnego odpowiedzialny jest zestaw **podstawowych czynników transkrypcyjnych**. Schemat powstawania tego kompleksu jest następujący:



Pierwszym etapem składania kompleksu transkrypcyjnego polimerazy RNA II jest wiązanie się podstawowego czynnika transkrypcyjnego TFIID z konserwowaną sekwencją DNA powyżej miejsca inicjacji transkrypcji, tzw. **elementem lub blokiem TATA** (ang. *TATA-box*). TFIID zawiera białko wiążące się z blokiem TATA zwane TBP (ang. *TATA-binding protein*) oraz kilka białek typu TAF (ang. *TBP-associated proteins*). Element TATA jest zasadniczym elementem transkrypcji na poziomie podstawowym, który jednak nie występuje we wszystkich promotorach dla polimerazy RNA II (patrz niżej). Sekwencja najwyższej zgodności elementu TATA to ciąg ośmiu par AT: TATAAATA (minimalny element TATA: TATAAA), występujący w kontekście sekwencji bogatych w pary GC. U większości *Eukaryota* TATA lokuje się w obszarze mniej więcej -25 - -30 i współdecyduje o wyznaczeniu miejsca startu transkrypcji. U drożdży pozycja elementu TATA nie jest tak ściśle określona i leży w zakresie od -40 do -120 (za wyznaczenie miejsca startu transkrypcji w większym stopniu odpowiada kontekst pozycji +1). Następnym etapem po związaniu TFIID jest precyzyjnie kontrolowane wiązanie wielu innych podstawowych czynników transkrypcyjnych i tworzenia aktywnego **kompleksu inicjacyjnego** **polimerazy RNA II** (zwanego również holoenzymem).

Transkrypcja zwykle zaczyna się od **A**, która występuje w luźno zdefiniowanym kontekście sekwencji regionu inicjatorowego **Inr** (ang. *Initiator region*). Sekwencja najwyższej zgodności Inr dla drożdży to: TCGA lub PuPuPyPuPu; zaś dla innych Eukaryota: PyPyCAPyPyPyPy.

Wiele promotorów genów *Metazoa* nie zawiera elementu TATA. Są to w większości bogate w nukleotydy G i C promotory wyrażanych konstytutywnie genów podstawowego metabolizmu komórkowego (ang. *house-keeping*). W tego rodzaju genach inicjacja transkrypcji może zachodzić w pojedynczym miejscu, w kilku miejscach zgrupowanych w niewielkim obszarze lub bardzo wielu miejscach oddalonych od siebie nawet o setki nukleotydów.

Skład kompleksu inicjacyjnego jest w zasadzie identyczny dla wszystkich genów transkrybowanych przez polimerazy RNA II. Jednakże wydajność jego składania oraz procesy związane z inicjacją transkrypcji, od których zależy poziom ekspresji, są bardzo zróżnicowane. Za kontrolę tych procesów odpowiadają białka wiążące się do promotora powyżej (rzadziej poniżej) miejsca składania kompleksu polimerazy RNA nazywane **regulacyjnymi czynnikami transkrypcyjnymi** (elementami regulacyjnymi *trans*). Regulacyjne czynniki transkrypcyjne rozpoznają krótkie sekwencje promotorowe nazywane **elementami regulacyjnymi *cis***. Wiążąc się do nich regulują kompleks polimerazy RNA (pozytywnie lub negatywnie) na zasadzie oddziaływań białko-DNA lub białko-białko, poprzez zmiany stężenia regulacyjnych czynników transkrypcyjnych lub powinowactwa do DNA. Znaczna większość regulacyjnych czynników transkrypcyjnych to **aktywatory**. Dodatkowymi czynnikami aktywującymi lub hamującymi aktywność regulacyjnych czynników transkrypcyjnych mogą być hormony (np. steroidowe), związki drobnocząsteczkowe (np. cAMP), jony (np. Ca2+) oraz peptydy lub białka. Zestaw elementów regulacyjnych *cis* może być wspólny dla określonej grupy genów. Synergiczne oddziaływania wielu regulacyjnych czynników transkrypcyjnych z promotorami genów danej grupy prowadzi do specyficznego profilu ich ekspresji.

Zwykle region genu zawierający najwięcej miejsc wiązania regulacyjnych czynników transkrypcyjnych mieści się w zakresie 200 - 400 bp powyżej miejsca inicjacji transkrypcji (+1). Dodatkowo, geny eukariotyczne mogą zawierać **sekwencje wzmacniacza (ang. *enhancers*)**, które również są elementami regulacyjnymi *cis*, do których wiążą się czynniki transkrypcyjne. Sekwencje te mogą występować w dowolnej lokalizacji (powyżej lub poniżej) i odległości (nawet tysiące bp) względem miejsca inicjacji transkrypcji. Sekwencje regulacyjne **UAS** (ang. *Upstream Activator Sequences)* drożdży uważa się za odpowiedniki wzmacniaczy wyższych *Eukaryota*, choć z reguły lokują się one w pobliżu genu i działają jedynie z pozycji powyżej miejsca inicjacji transkrypcji.

Bardzo prostym i jednym z najlepiej poznanych przykładów struktury i działania genu eukariotycznego są geny odpowiedzialne za wykorzystanie galaktozy u *S. cerevisiae*. Cały system obejmuje 12 genów struktury i genów regulacyjnych, z których najważniejsze to:

|  |  |
| --- | --- |
| Geny struktury  | Geny regulacyjne |
| GAL1 - galaktokinaza,GAL2 - permeaza galaktozy,GAL7 - galaktotransferaza,GAL10 - epimeraza UDP-gal | GAL4 - aktywator transkrypcji,GAL80 - inhibitor aktywatora (represor?). |

Ekspresja genów struktury regulowana przez galaktozę i glukozę. Do wysokiego poziomu transkrypcji konieczne jest związanie się aktywatora transkrypcji – białka GAL4 do sekwencji UASG znajdujących się w regionach promotorowych genów GAL1, GAL10 i GAL7. W przypadku genu GAL1 są położone około 250bp od miejsc inicjacji transkrypcji. Zawierają po 4 palindromowe sekwencje o długości 17 bp, przy czym pojedyncza kopia palindromu jest wystarczająca do niemal pełnej aktywacji. Sekwencja najwyższej zgodności dla pojedynczego palindromu: CGGA(c/g)GAC(A/T)GTC(g/c)TCCG.

Białko GAL4 ma 3 domeny funkcjonalne, konieczne dla (i) oddziaływania z DNA, (ii) aktywacji transkrypcji, (iii) oddziaływania z białkiem GAL80, które wiąże się z białkiem GAL4 w nieobecności galaktozy, blokując domenę aktywacji transkrypcji.



**Brak galaktozy**

Powstawanie kompleksu transkrypcyjnego jest uniemożliwione, ponieważ domena aktywująca Gal4 jest blokowana przez Gal80.

**Obecność galaktozy – indukcja**

Gal80 odłącza się od domeny aktywującej Gal4, co umożliwia złożenie kompleksu transkrypcyjnego. Gal80 jest wiązane w cytoplazmie przez Gal3 - wiązanie zależne od obecności galaktozy.

Traven et al. ***Yeast Gal4: a transcriptional paradigm revisited****.* EMBO Rep. May 2006; 7(5): 496–499

**Warto przeczytać (Dla zainteresowanych):**

1. Ana Traven, Branka Jelicic, Mary Sopta ***Yeast Gal4: a transcriptional paradigm revisited****.* EMBO Rep. May 2006; 7(5): 496–499. doi: 10.1038/sj.embor.7400679
2. Houseley J, Rubbi L, Grunstein M, Tollervey D, Vogelauer M. ***A ncRNA modulates histone modification and mRNA induction in the yeast GAL gene cluster.*** Mol Cell. 2008 Dec 5;32(5):685-95. doi: 10.1016/j.molcel.2008.09.027

**Potranskrypcyjne etapy ekspresji genów eukariotycznych**

W odróżnieniu od *Prokaryota* większość genów eukariotycznych ma strukturę nieciągłą, co oznacza, że obszary ulegające transkrypcji składają się z naprzemienne występujących **eksonów** i **intronów**. Sekwencje intronowe nie wchodzą w skład funkcjonalnego mRNA co wymusza istnienie mechanizmu ich usuwania. Ponadto transkrypcja i translacja są *u Eukaryota* rozdzielone w czasie i przestrzeni. Translacja zachodzi w cytoplazmie, w związku z czym powstający w jądrze mRNA musi być tam przetransportowany. W strukturze genu eukariotycznego powinny się więc znajdować elementy warunkujące powstawanie dojrzałego mRNA, jego stabilność i transport do cytoplazmy a także prawidłową translację.

Bezpośrednim produktem transkrypcji genu jest zawierająca introny cząsteczka **prekursorowego mRNA (pre-mRNA).** Pre-mRNA jest niestabilny i niezdolny do eksportu z jądra.Pre-mRNA ulega procesom **obróbki potranskrypcyjnej** (*dojrzewania*), na które składają się:

* formowanie końca 5'-dodanie *czapeczki*
* usuwanie intronów - składanie mRNA *(*ang. *splicing)*
* formowanie 3' końca - cięcie i poliadenylacja.

**Czapeczkę** dodawaną do końców 5' pre-mRNA stanowi cząsteczka zmetylowanej guanozyny (7-mG) dołączona do RNA przez trzy reszty fosforanowe. Dodanie czapeczki jest koniecznym warunkiem do rozpoczęcia składania mRNA*,* czapeczka jest jednym z sygnałów eksportu mRNA z jądra, zabezpiecza mRNA przed działaniem egzonukleaz, jest też niezbędna do prawidłowej inicjacjitranslacji.

**Składanie mRNA** jest bardzo skomplikowanym procesem, w którym introny są wycinane z RNA, a sąsiadujące ze sobą eksony odpowiednio łączone. W proces ten zaangażowanych jest wiele **czynników *cis i trans****.* Czynniki *cis* to charakterystyczne elementy w sekwencjach intronów oraz eksonów (na granicach z intronami), czynniki *trans* to różnorodne kompleksy białko - RNA i białka, odpowiedzialne za rozpoznanie intronów, katalizę reakcji składania mRNAoraz jego regulację. Składanie mRNA jest procesem dwuetapowym:



W pierwszym etapie następuje przecięcie cząsteczki pre-mRNA na **granicy 5' ekson/intron** (ang. *5' splice site*)i wolny koniec 5' intronu zostaje jednocześnie połączony z jedną z zasad tzw. **regionu rozgałęzienia** w obrębie intronu *(*ang. *branch-point region).*W wiązaniu tym uczestniczy grupa 2'-OH zasady z regionu rozgałęzienia i tworzy się struktura tzw. ***lariatu***(lassa). Drugi etap składania mRNA to przecięcie cząsteczki na **granicy 3' intron/ekson** *(*ang. *3' splice site) i* jednoczesna ligacja eksonów.

Struktura intronów różni się w obrębie *Eukaryota:*



Na schemacie "R" oznacza purynę; "Y" - pirymidynę.

Nieliczne, stosunkowo krótkie i luźno zdefiniowane elementy strukturalne intronów są wielokrotnie rozpoznawane przez różne czynniki *trans w* czasie formowania się *spliceosomu,* wielkiego kompleksu enzymatycznego, odpowiedzialnego za przeprowadzenie reakcji składania mRNA*.* Wielokrotne oddziaływania czynników *cis i trans* dają możliwość precyzyjnego rozpoznania sekwencji intronowych.

Szczególną kategorię czynników *trans* stanowią kompleksy niektórych małych jądrowych RNA (snRNA) z białkami, zwane kompleksami snRNP. W składanie mRNAzaangażowane są: U1snRNP, U2snRNP, U4/U6snRNP i U5snRNP. Kompleksy snRNP stanowią zasadniczą część dynamicznej struktury *spliceosomu*; sąodpowiedzialne za rozpoznanie intronu, likwidację ewentualnych struktur drugorzędowych w RNA intronu, odpowiednie sfałdowanie pre-mRNA w toku składania mRNA*,* wreszcie - samą katalizę reakcji transestryfikacji. Większość tych funkcji spełniają snRNA poprzez oddziaływania RNA-RNA, część - skompleksowane z nimi białka. Oprócz kompleksów snRNP, bardzo ważną rolę *w* składaniu mRNAodgrywają liczne (ponad 100) dodatkowe białka.

**Formowanie końca 3' mRNA** obejmuje szereg reakcji, m. in. **cięcie** i **poliadenylację** (dodanie *ogona* zbudowanego z reszt A). Proces terminacji transkrypcji polimerazy RNA II u *Eukaryota* odbywa się w rozległych, obszarach daleko za końcami 3' otwartych ramek odczytu (Patrz pozycja 1 i 2 poniżej, dla zainteresowanych). Procesami dobrze poznanymi natomiast są reakcje cięcia i poliadenylacji mRNA. Poliadenylacja mRNA spełnia szereg funkcji - jest niezbędna do właściwego przebiegu translacji, chroni mRNA przed degradacją i daje możliwość regulacji ekspresji genu. Postuluje się też jej możliwe znaczenie w eksporcie mRNA z jądra do cytoplazmy.

W proces formowania końca 3' zaangażowanych jest także szereg elementów cis i *trans.* Podstawowym elementem *cis w* przeważającej większości genów jest sekwencja sygnałowa 5’-AAUAAA-3’ niezbędna do dołączenia ogona poli(A) w trakcie obróbki pre-mRNA. Element ten rozpoznawany jest przez białko zwane CPSF *(*ang. *Cleavage and Polyadenylation Specificity Factor),* które rozpoczyna formowanie funkcjonalnego kompleksu obróbki końca 3', prowadząc do cięcia pre-mRNA w odległości 10-30bp poniżej sekwencji sygnałowej 5’-AAUAAA-3’, bardzo często tuż za dinukleotydem 5’-CA-3’. W skład kompleksu formowania 3’-końca mRNA wchodzą również czynniki: czynnik stymulacji cięcia CstF *(*ang. *Cleavage stimulation factor),* wiążący się w rejonie bogatym w GU (10-20bp za miejscem ciecia), CF1 (ang. *Cleavage Factor),* polimeraza poli(A) oraz białko wiążące się z ogonami poliA - PABII *(*ang. *PolyA Binding factor)*.

Warto zanaczyć, że poliadenylacja jest istotna do prawidłowej terminacji transkrypcji, której szczegółowy mechanizm nie jest omawiany na ćwiczeniach.

**Warto przeczytać (Dla zainteresowanych):**

1. Kuehner JN, Pearson EL, Moore C. ***Unravelling the means to an end: RNA polymerase II transcription termination.*** Nat Rev Mol Cell Biol. 2011 May;12(5):283-94. doi: 10.1038/nrm3098.
2. Proudfoot NJ. ***Ending the message: poly(A) signals then and now***. Genes Dev. 2011 Sep 1;25(17):1770-82. doi: 10.1101/gad.17268411.