

## TRANSCRIPTIA GENETICA

– principii, etape, ARN polimeraze, promotori și terminatori, factori de transcripție (NOTE DE CURS BM3/BM4)

- procesarea ARNm la eucariote (NOTE DE CURS BM4)

Atât celulele procariote cât și cele eucariote posedă mecanisme, pe de o parte comune, pe de altă parte distincte, pentru reglarea expresiei genelor și a fluxului de informație de la ADN la proteine. **Transcripția ADN**, realizată de **ARN polimeraze** constituie prima etapă, comună la toate organismele. La celulele procariote, ribozomii se fixează pe molecula de ARNm, în curs de sinteză, și realizează traducerea imediată a informației în proteine, în citoplasmă. Din contră, la eucariote membrana celulară separă locul în care se realizează transcripția de cel în care are loc traducerea. Transcriptul primar, ARNm imatur, conține o succesiune de secvențe netraduse, numite introni, care vor fi eliminate prin procedeul de scindare a intronilor și sudare a exonilor („splicing”), în procesul de maturare a ARNm. În urma acestei etape, localizată în nucleu, ARNm trece în citoplasmă unde este tradus. Maturarea ARNm prin „splicing” conferă celulei eucariote posibilitatea diversificării informației prin combinarea alternativă a exonilor, o altă etapă foarte importantă în dezvoltarea organismelor.

Expresia informației genetice conținută într-un segment de ADN este transcrisă în structura diferitelor tipuri de ARN. Astfel, **sinteza macromoleculelor de ARN direcționată de ADN** și realizată cu ajutorul unor enzime specializate, ARN polimeraze, poartă denumirea de **transcripție**. Acest proces se desfășoară, atât la procariote, cât și la eucariote în trei etape:

- inițierea transcripției;
- alungirea catenei ARN;
- terminarea transcripției;

### Transcripția la procariote

**Inițierea transcripției la procariote** (BM3, SLIDE 7) prezintă următoarele caracteristici:

- ARN polimeraza inițiază transcripția majorității genelor la nivelul unei poziții unice (**promotor**) situată în amonte de secvența codantă;

- perechea de baze la nivelul căreia se inițiază transcripția este denumită “situs de inițiere a transcripției” sau Punct START;

-prin convenție, situs de inițiere a transcripției de pe secvența ADN este desemnată cu poziția +1, bazele situate în sensul transcripției (downstream) sunt desemnate cu numere pozitive, iar cele poziționate în sens opus (upstream) sunt desemnate cu numere negative;

- diferitele proteine (ARN polimeraze, activatori, represori) interacționează cu ADN la nivelul promotorului sau în vecinătatea acestuia pentru a regla inițierea transcripției.

Interacțiunile proteine – ADN sunt frecvente în timpul transcripției și au fost evidențiate experimental prin diferite tehnici (*footprinting, gel-shift assays, etc.*).

### **Interacțiunile proteine-ADN identificate prin tehnica *footprinting* (SLIDE 8)**

În figură sunt prezentate schematic etapele tehnicii *footprinting* cu DN-aza I, comună pentru identificarea situsurilor de legare a proteinelor la ADN.

Un fragment de ADN care este marcat la unul din capete cu izotopul  $P^{32}$  (reprezentat în figura de bilele roșii), marcarea similară celei prezente în metoda de secvențiere Maxam și Gilbert. Porțiuni din probă sunt apoi digerate cu DN-aza I, în prezența și în absența unei proteine care se leagă la o secvență specifică din fragment. DN-aza I hidrolizează aleator legăturile fosfodiester. La concentrații scăzute, DN-aza I este utilizată astfel încât fiecare moleculă de ADN este clivată numai o dată. În absența proteinelor care se leagă la ADN, proba este clivată la toate pozițiile posibile între capătul marcat și nemarcat al moleculei. Cele două probe de ADN sunt apoi separate de proteine, denaturate pentru desfacerea catenelor și supuse electroforezei. Gelul rezultat este supus electroforezei și supus autoradiografiei pentru a detecta numai catenele marcate și pentru a identifica fragmentele care se întind de la capătul marcat până la locul de acțiune al DN-azei I. În dreapta se poate observa o autoradiogramă unde sunt analizate două fragmente de ADN în absența și în prezența unor proteine de legare la ADN după digestia cu DN-aza I. Pentru proba digerată în prezența proteinelor de legare la ADN se observă lipsa a două benzi; acestea corespund regiunii protejate de digestie prin legarea proteinelor (spunem că am realizat un *footprinting* privind legarea acestei proteine). Această regiune de protecție poate fi foarte ușor aliniată cu secvența ADN dacă se realizează o reacție de secvențiere a produșilor inițiali și a celor care s-au obținut în urma protecției furnizate de legarea proteinelor. Reacțiile de secvențiere și fragmentele rezultate în urma acțiunii DN-azei sunt trecute pe același gel.

### **Identificarea interacțiilor proteine-ADN prin tehnica *gel-shift assays* (SLIDE 9)**

În figură sunt prezentate rezultatele obținute în urma unei determinări a mobilității electroforetice prin tehnica de “întârziere în gel” pentru proteinele care se leagă la ADN (**EMSA – Electrophoretic Mobility Shift Assay**). În acest exemplu, fracții ale unei coloane cromatografice de schimb ionic au fost analizate pentru proteinele care se leagă la **regiunea promotor** a unei gene pentru ARNt de la EK. Un eșantion al fracției proteice este încărcată pe o coloană (ON) iar fracțiile eluate de pe coloană la concentrații saline crescute (numerele fracțiilor) au fost incubate cu o sondă (fragment de restricție) radiomarcată care include regiunea promotoare; fiecare probă a fost apoi supusă electroforezei într-un gel de poliacrilamidă. Sondele libere care nu se leagă la proteine migrează în față. O proteină prezentă în fracțiile 7 și 8 eluate de pe coloană se leagă la sondă, formând un complex ADN-proteină care migrează mult mai lent decât sonda liberă.

### **Evidențierea poziționării ARN polimerazei la nivelul regiunii control a represorului *lac* prin tehnica *footprint* (SLIDE 10)**

În imagine este reprezentat un experiment de tip “footprint” a legării ARN polimerazei și a represorului *lac* la regiunea control al ADN *lac*. Deoarece DN-aza I clivează unele legături fosfodiester mai frecvent decât altele, densitatea benzilor în absența proteinelor (linia 3) nu este la fel de uniformă ca și din figura din slide-ul 8. Parantezele din dreapta indică regiunile ADN protejate de digestia cu DN-aza I prin legarea ARN polimerazei (linia 4) sau a represorului *lac* (linia 5). Liniile 1 și 2 arată produșii celor două reacții de secvențiere Maxam și Gilbert: de pe aceste linii, benzile din gel pot fi corelate cu secvența nucleotidică a regiunii de control *lac*. Pozițiile în secvență sunt notate pe partea stângă; capetele săgeților indică direcția transcripției.

### **La procariote există o singură ARN polimerază.**

**ARN polimeraza** de la *E.coli* este o enzimă solubilă, cu MM – 480 kD, constituită din următoarele **subunități:  $\alpha$ ,  $\beta$ ,  $\beta'$ ,  $\sigma$** . Pentru a iniția transcripția, ARN polimeraza se fixează specific la nivelul secvențelor promotorului (SLIDE 12). În imagine este reprezentarea schematică modului de legare la ADN a formei majore a ARN polimerazei de la *E. coli*.

Prin convenție, situsul de inițiere a transcripției este numerotat, în general, cu +1. Perechile de baze care se extind în direcția transcripției sunt numite “downstream”, în aval, față de situsul de inițiere a transcripției; cele care sunt extinse în regiunea inversă sunt considerate “upstream” sau în amonte. **Subunitatea  $\sigma 70$**  se leagă la secvențe specifice situate lângă **pozițiile –10 la –35 ale**

**promotorului.** Subunitățile  $\alpha$  se leagă strâns la ADN din amonte, iar  $\beta$  și  $\beta'$  sunt asociate punctului de START.

## **Secvențe promotor la procariote**

**Diferențele de la nivelul secvențelor promotorului *E. coli* afectează frecvența inițierii transcripției** (SLIDE 13).

Figura reprezintă *Promotorii recunoscuți de către ARN polimeraza* (conținând subunitatea  $\sigma 70$ ) de la *E. coli*.

La procariote există mai multe secvențe promotor, care nu sunt identice, dar prezintă anumite nucleotide conservate. Astfel au fost evidențiate:

a) Secvențele unor promotori “puternici” cu spații introduse pentru a maximiza omologia la nivelul regiunii  $-10$  la  $-35$ . Aceste secvențe corespund catenei sens a promotorului, în condițiile în care procesul de transcripție se realizează în direcți  $5' \rightarrow 3'$  (de la stînga la dreapta). Bazele care corespund secvenței consens din regiunea  $-35$  la  $-10$  sunt reprezentate în figura respectivă prin colorare în galben. Șase dintre secvențe corespund secvențelor control ale genelor care codifică pentru ARN. Secvențele din structura regiunilor promotoare ale Lambda, T7 și fd, care sunt prezente în genoame virale care sunt direct transcrise de către ARN polimeraza celulei gazdă.

b) Secvențele consens ale regiunilor  $-35$  la  $-10$ , care sunt separate de 15-17 pb. Sunt descrise mutațiile cunoscute ca descrescând semnificativ frecvența transcripției unui anumit număr de promotori. În secvențele consens, frecvența cu care bazele indicate apar la fiecare poziție la diferiți promotori pentru  $\sigma 70$  este indicată după cum urmează: litere în roșu:  $> 70\%$ ; litere boldite negre: 50-70%; litere negre: 40-50%.

c) Secvențele regiunilor  $-35$  la  $-10$  ale promotorului *lac* care sunt diferite de secvența consens în patru poziții (reprezentate în albastru). Mutațiile “down” determină scăderea expresiei operonului *lac*. Cele două mutații “up” care cresc gradul de asemănare a regiunii  $-10$  cu secvența consens, cresc expresia operonului *lac*.

**Transcripția de la nivelul unor promotori este inițiată de factorii sigma ( $\sigma$ ) alternativi** (SLIDE 20).

Transcripția majorității promotorilor de la *E. coli* este demarată prin interacția promotorilor cu ARN polimeraza care conține factorul  $\sigma 70$ . Transcripția anumitor grupe de gene este realizată de

către ARN polimeraze care conțin una sau mai multe variante alternative de factori  $\sigma$ :  $\sigma_{28}$ ,  $\sigma_{32}$ ,  $\sigma_{38}$  și  $\sigma_{54}$ , care recunosc secvențe promotor consens diferite de cele recunoscute de  $\sigma_{70}$ . Atât  $\sigma_{28}$  și  $\sigma_{32}$  prezintă regiuni de omologie cu  $\sigma_{70}$  și recunosc secvențe situate tot în intervalul  $-35$  la  $-10$  (Tabel 10-1, Slide 20).

Un exemplu în ceea ce privește factorii sigma alternativi este reprezentat de activarea subunității  $\sigma_{54}$  a ARN polimerazei la nivelul promotorului *glnA* de către NtrC (Nitrogen regulatory Protein) (SLIDE 21).

*glnA* este o genă care codifică pentru enzima glutamin sintetaza care sintetizează glutamina pornind de la acidul glutamic și amoniac. ARN polimeraza se leagă la promotorul *glnA*, formând un complex unit (strâns), înainte de activare. Ca răspuns la concentrațiile mici de azot organic, o protein kinază numită NtrB fosforilează proteina dimeră NtrC, care apoi se leagă la doi “enhanceri” poziționați la  $-108$  și  $-140$  față de situsul de START al transcripției. Dimerii NtrC fosforilați legați la enhanceri interacționează cu subunitatea  $\sigma_{54}$  legată determinând formarea unei bucle la nivelul ADN implicat. Activitatea ATPazică a dimerilor NtrC stimulează ARN polimerază să desfacă catenele la situsul de START, formând un complex deschis și în acest moment transcripția genei *glnA* poate să înceapă.

Vizualizarea buclei ADN și a interacțiilor ADN-NtrC și ADN-subunitate  $\sigma_{54}$  a ARN polimerazei a fost posibilă prin micrografie electronică (SLIDE 22).

a) Micrografie electronică a fragmentului de restricție ADN cu cei doi dimeri NtrC fosforilați legați la regiunea enhancer lângă unul din capetele subunității sigma54 a polimerazei legată la promotorul *glnA* lângă celălalt capăt.

b) Micrografie electronică a aceluiași fragment demonstrând legarea dimerilor NtrC și a subunității sigma54 a polimerazei unul cu altul cu formarea buclei de ADN între aceștia.

### **Elongarea ARN de către ARN polimeraza** (vezi CURS BM4, SLIDE 2)

În regiunea care va fi transcrisă, dublul helix este desfăcut cu aproximativ un tur pentru a permite catenei sens să formeze un scurt segment hibrid ADN/ARN dublu catenar cu capătul 3' al ARN. Cu cât ARN polimeraza avansează de-a lungul matriței ADN, acesta este despiralizat în fața furcii de transcripție în deplasare și respiralizat în spatele acesteia, eliberând astfel ARN nou sintetizat de catena matriță (antisens).

Au fost propuse două mecanisme de elongare:

- a) O cale poate să fie urmată este aceea în care ARN polimeraza urmează deplasarea catenei matriță astfel încât transcriptul va rămâne atașat la nivelul elicei duplexului;
- b) O a doua posibilitate, mult mai plauzibilă, este aceea că ARN se deplasează în linie dreaptă, în timp ce matrița ADN se rotește dedesubtul său. În acest caz ARN nu se va răsuci în jurul ADN, dar ADN va rămâne suprarăsucit în fața sa (luați în considerare plasarea degetului între catrenele de ADN răsucit în acest model și împingând către dreapta). Modelul presupune că atât capetele ADN, cât și ARN polimeraza sunt protejate de atașamente în interiorul celulei.

### **Reglarea transcripției la procariote**

Operonii reprezintă un grup de gene localizate una lângă cealaltă. Toate genele din structura unui operon sunt exprimate ca o singură unitate. Acest tip de aranjament este comun genoamelor procariote și poate fi ilustrat la *E. coli* cu operonul lactozei care a fost descoperit în 1961 de către Jacob și Monod. Acesta conține trei **gene structurale** (*lacZ*, *lacY* și *lacA*) implicate în transformarea dizaharidului lactoză în unitățile sale componente (glucoză și galactoză), o regiune **promotor**, o regiune **operator** și o regiune **reglator**. Genele operonului *lac* sunt implicate în utilizarea lactozei ca sursă de energie de către *E. coli*. Deoarece, lactoza nu este o componentă comună a mediului înconjurător pentru *E. coli*, majoritatea timpului operonul nu este exprimat și enzimele pentru utilizarea lactozei nu sunt sintetizate de bacterie. Astfel, în absența lactozei, proteina **repressor** se leagă la regiunea operator și împiedică legarea ARN polimerazei și transcrierea genelor structurale ale operonului. Când lactoza devine disponibilă se realizează activarea operonului și cele trei gene sunt exprimate împreună. Operonul *lac* reprezintă un exemplu clasic de reglare a expresiei genelor la bacterii.

Mutațiile la nivelul promotorului, la care se leagă ARN polimeraza, sau la nivelul operatorului acționează în cis; aceasta înseamnă că ele afectează numai expresia genelor de pe aceeași moleculă de ADN în care apare mutația. Mutațiile în secvența unui operator care scad legarea unui repressor au drept rezultat o transcripție constitutivă. Mutațiile în secvența promotorului, care afectează afinitatea de legare a ARN polimerazei, pot fie să scadă (mutație “down”), fie să crească (mutație “up”) transcripția.

Majoritatea **secvențelor operator** la bacterii sunt scurte repetiții inversate (SLIDE 14). Secvența operatorului *lac* reprezintă o secvență repetitivă inversată aproape perfectă centrată în jurul unei secvențe GC situată în poziția 11. Secvența de 17 pb de pe catena sens care începe la -7 este identică cu secvența de 17 pb situată pe catena antisens și începe la +28, cu excepția

nucleotidelor indicate în italic. Fiecare din aceste secvențe repetitive este denumită “jumătate de situs operator” (Figura SLIDE 14). Majoritatea **represorilor** bacterieni sunt structuri dimerice care conțin elice  $\alpha$  care se inserează în fosele majore adiacente ale ADN (SLIDE 15).

### **Controlul pozitiv/negativ al operonului *lac* de către cAMP-CAP (SLIDE 17, 18)**

- a) În absența lactozei, nu se formează ARNm *lac* deoarece represorul se leagă la operatorul *lac* și împiedică transcripția;
- b) În prezența glucozei și lactozei, represorul *lac* se leagă la lactoză și suferă modificări conformaționale și astfel nu se mai poate lega la operatorul *lac*. Totuși, cAMP este scăzut, deoarece glucoza este prezentă și atunci complexul cAMP-CAP nu se leagă sitului CAP de pe operator. Ca rezultat, ARN polimeraza nu se leagă eficient la promotorul *lac* și este sintetizat numai puțin ARNm *lac*;
- c) În prezența lactozei și în absența glucozei, apare transcripție maximă la nivelul operonului *lac*. În această situație, represorul *lac* nu se leagă la operatorul *lac*, concentrația cAMP crește și complexul cAMP-CAP format se leagă la situsul CAP, stimulând legarea și inițierea de către ARN polimerază.

Legarea cooperativă a cAMP-CAP și a ARN polimerazei la regiunea de control a *lac* activează transcripția genelor operonului *lac*. Prin propria structura, ARN polimeraza și cAMP-CAP posedă afinități relative reduse pentru situsurile lor de legare la promotorul *lac*. Totuși, interacția dintre regiunile CAP și subunitatea alfa a ARN polimerazei formează un complex proteină-proteină care se leagă mult mai stabil la promotor decât fiecare dintre cele două proteine singure.

### **Concluzii privind procesul de transcripție la procariote (SLIDE 24, 25)**

-ARN polimerazele sunt proteine mari compuse din subunități  $\beta$ ,  $\beta'$  și două subunități  $\alpha$  care formează “core” polimeraza și o subunitate  $\sigma$ , din câteva alternative, care funcționează ca factor de inițiere.

-Inițierea începe când subunitatea  $\sigma$  a moleculei de polimerază se leagă la secvența promotor, formând un complex unit. Polimeraza separă apoi catenele pe o distanță de 12-13 pb la nivelul situsului de START a transcripției, formând un complex deschis. După ce aproximativ 10 ribonucleotide au fost polimerizate, subunitatea  $\sigma$  este eliberată și “core” polimeraza continuă transcripția matriței.

-“Puterea” unui promotor se referă la cât de frecvent ARN polimeraza inițiază transcripția pornind de la acesta. Subunitatea  $\sigma 70$ , factorul major în inițiere la *E. coli*, interacționează cu secvențele promotor situate în regiunea -10 la -35 față de situsul de START;

- Represorii se leagă la secvențele de ADN numite operatori, care se suprapun parțial cu regiunea promotoare la nivelul căreia se leagă ARN polimeraza. Legarea represorului interferă cu legarea ARN polimerazei și cu inițierea transcripției.

-Activatorii subunității  $\sigma 70$  a ARN polimerazei se leagă, în general, la ADN pe partea opusă a helixului față de polimerază, în regiunea -20 la -50, sau chiar în amonte de ARN polimerază, în apropierea -60;

- Activatorul cAMP-CAP stimulează transcripția prin formarea unui complex cu ARN polimeraza care prezintă o mai mare afinitate pentru situsuri ADN specific decât proteinele individuale. În plus față de stimularea legării polimerazei, activatorii pot stimula și formarea complexului deschis și începerea transcripției;

- Mulți represori bacterieni sunt dimeri. Fiecare monomer conține un  $\alpha$ -helix care se inserează în fosa majoră a operatorului ADN, astfel încât dimerul se leagă la fose majore succesive. Afinitatea mare de legare este rezultatul formării multor legături de hidrogen, ionice și van der Waals dintre proteine și secvențe ADN specific.

-Secvențele de ADN care leagă proteinele reglatoare dimere sunt secvențe repetitive inverse, care reprezintă fiecare jumătate de situsuri care leagă un monomer.

- Operonii transcriși de către subunitatea sigma54 sunt reglate prin activatori care se leagă la secvențe “enhancer” de aproximativ 100 pb în amonte de situsul de inițiere. Secvențele “enhancer” care leagă activatori inetracționează tranzitoriu cu ARN polimeraza legată la nivelul promotorului, stimulând formarea unui complex deschis și inițierea transcripției.

## **Transcripția la eucariote**

### **Controlul genelor eucariote: scop și principii generale (slide 26, 27)**

- spre deosebire de celulele bacteriene și eucariotele unicelulare, celulele organismelor pluricelulare au relativ puține gene reglate reversibil de condițiile de mediu;
- controlul genetic al organismelor pluricelulare este important pentru dezvoltare și diferențiere și, în general, nu este reversibil;



În interiorul unei celule bacteriene singulare, genele sunt reversibil induse și represate prin control transcripțional pentru a ajusta mașinăria enzimatică celulară mediului înconjurător nutrițional și fizic. Eucariotele monocelulare, cum sunt drojdiile, posedă și ele multe gene care controlează răspunsul la variația mediului (cum ar fi statusul nutrițional, aportul de oxigen și temperatura). Chiar în organele organismelor superioare, de exemplu, ficatul mamiferelor unele gene pot răspunde reversibil la stimulii externi cum sunt noxele chimice. Totuși, în general celulele eucariotelor pluricelulare sunt protejate de influențele externe imediate; adică majoritatea celulelor își desfășoară activitatea într-un mediu constant. Probabil, din acest motiv genele care răspund schimbărilor de mediu contribuie o fracție mult mai mică din numărul celulelor unui organism multicelular decât în organismul unicelular. Cea mai importantă caracteristică a controlului genetic la organisme multicelulare este exactitatea execuției deciziei precise de dezvoltare astfel încât gena necesară să fie activată într-o anumită celulă la un anumit moment din dezvoltarea organismului care conține un număr mare de tipuri celulare. În majoritatea cazurilor o dată ce etapa de dezvoltare a fost depășită de o celulă, ea nu este reversibilă. Astfel, aceste decizii sunt fundamental diferite de inducția sau represia bacteriană. În executarea programului lor genetic, multe tipuri celulare diferite (ex. celulele din piele, celulele roșii din sânge, celule producătoare de anticorpi, etc) marchează o cale către moartea celulară finală. Profilele determinate ale controlului genetic care conduc la diferențiere servesc necesitățile întregului organism și nu supraviețuirii individuale a celulei.

La eucariotele superioare reglarea multor gene se realizează prin controlul transcripției. Măsurătorile directe ale ratelor de transcripție ale mai multor gene în diferite tipuri celulare au arătat că reglarea inițierii transcripției este forma universală a controlului genetic atât la eucariote, cât și la bacterii.

### **Elementele reglatoare la eucariote (SLIDE 30)**

- reprezintă adesea mii de kilobaze în amonte sau în aval de START; Principiile de bază care controlează transcripția la procariote, există și la eucariote: transcripția este inițiată la nivelul unei regiuni specifice și este controlată de legarea proteinelor “trans-activatoare” (factori de transcripție) la secvențele “cis-activatoare” din structura ADN;

- elementele “cis-activatoare” sunt adesea mult mai departe de promotorul pe care îl reglează, transcripția la nivelul unui singur promotor poate fi reglată prin legarea mai multor factori de transcripție la elementele de control alternative;
- secvențele care controlează transcripția pot fi identificate prin analize de deleție în serie la capătul 5’.

Genele procariote și eucariote sunt structurate după aceeași logică de bază, dar cu diferențe importante în detaliile moleculare. O definiție sintetică a unei gene eucariote poate fi utilă pentru a rezuma esența acestor diferențe. Este general admis că o singură definiție dată genei eucariote nu poate satisface pe toată lumea sau să corespundă la toate exemplele. Cea pe care am adoptat-o este rezultatul structurii moleculare a genelor care asigură o gamă largă de funcțiuni în diverse organism eucariote. Definiția ține cont de diferențele de localizare și de tipurile de secvențe care influențează expresia genelor. Aceasta presupune implicit că mutațiile fenotipice observabile sunt rezultatul modificărilor la nivelul semnalelor de reglare dar și în secvențele codante.

Definim gena ca fiind o combinație de segmente de ADN care, împreună, constituie o unitate de expresie, o unitate care determină formarea unui produs specific și funcțional al genei și care poate să fie o moleculă de ARN sau un polipeptid. Segmentele de ADN care definesc gena includ:

1. **Unitatea de transcripție** care este constituită dintr-un segment AND continuu care codifică pentru secvența transcriptului primar; acesta include:

- a) secvența codantă a ARN matur sau a proteinei produse,
- b) intronii și
- c) secvențele leader 5’ și coadă 3’ prezente la nivelul ARNm matur ca și secvențele de separare care sunt eliminate în timpul maturării transcripturilor primary care codifică pentru ARN.

2. **Secvențele minimale** necesare pentru inițierea corectă a transcripției (**promotorul**) și pentru a da naștere extremității 3’ particulare din structura ARN matur.

3. Elementele secvenței care determină frecvența de inițiere a transcripției; acestea cuprind secvențele responsabile de inducerea și represia transcripției și de specificitatea celulară, tisulară și temporală a transcripției. Aceste regiuni sunt foarte variate structural, ca poziție și ca funcție pentru a putea purta un singur nume. Dintre acestea enumerăm **elementele activatoare (enhancers)** și **elementele moderatoare (silencers)**, care sunt secvențe ce influențează inițierea transcripției la distanță, independent de localizarea lor în raport cu situsul de inițiere.

În această definiție a genei nu sunt incluse secvențele ADN care influențează configurația unei gene în cromatină și nici cele care codifică proteine sau ARN care modulează expresia unei anumite unități de transcripție.

## **ARN Polimeraze la eucariote**

**La EK sinteza moleculelor de ARN este catalizată de 3 polimeraze** (slide 33).

Experimental, cele trei ARN polimeraze de la eucariote au fost separate și identificate prin cromatografie pe coloană (vezi figura slide 33). Un extract proteic din nucleii celulelor de broască a fost trecut pe o coloană de DEAE Sephadex, pe care proteinele încărcate se absorb diferit. Proteinele absorbite sunt eluate (curba albastră) cu o soluție cu concentrație de NaCl crescândă. Frațiile care conțin proteinele eluate sunt testate pentru capacitatea lor de a realiza transcripția ADN (curba roșie) în prezența a patru ribonucleotid-trifosfați. A fost măsurată și sinteza ARN de către fiecare fracție în prezența a 1  $\mu\text{g/ml}$  de  $\alpha$ -amanitină (curba bleu). La această concentrație  $\alpha$ -amanitina inhibă activitatea ARN polimerazei II, dar nu afectează pe cea a ARN polimerazelor I și III (ARN polimeraza III este sensibilă la 10  $\mu\text{g/ml}$  de  $\alpha$ -amanitină, în timp ce ARN polimeraza I este neafectată chiar și la concentrații mai mari).

Transcripția genelor din clasa I este realizată de **ARN polimeraza I** și reprezintă jumătate din activitatea de transcripție pentru majoritatea celulelor eucariote. Singurul produs al acestei transcripții este un **precursor de ARN ribozomic (pre-ARNr)**, care este transformat prin clivare secvențială pentru a forma **speciile de ARNr 5,8S, 18S și 28S** (cel de al patrulea tip de ARNr 5S este codificat de către o genă din clasa III). Numărul de gene care codifică pentru ARNr variază de la o sută la mai multe mii, în funcție de specie. Acestea sunt localizate pe unul sau pe mai mulți cromozomi specifici, în regiuni morfologice distincte numite organizatori nucleolari. În timpul interfazei, aceste regiuni sunt încorporate în nucleol, structură în care moleculele de ARNr sunt transcrise activ, se transformă pre-ARNr și se assemblează ribozomii.

Contrar **ARN polimerazelor II și III localizate în nucleoplasmă**, **ARN polimeraza I este asociată nucleolului** și poate fi izolată de nucleosol. Numai ARN polimeraza I catalizează transcripția grupului de gene ARNr. Specificitatea ARN polimerazei I pentru genele ARNr este sugerată de insensibilitatea sintezei ARNr și de rezistența polimerazei I la  $\alpha$ -amanitină.

Transcripția genelor din clasa II este realizată de **ARN polimeraza II**. Genele din clasa II codifică pentru toate moleculele de **ARNm citoplasmatic, moleculele ARN U din snRNP** (small nuclear ribonucleoproteine), **cu excepția U6**. Moleculele transcript ale genelor din clasa II

posedă modificări caracteristice la extremitatea 5' (**coif**): 7-metil guanozină pentru ARNm și 2, 2, 7- trimetil guanozină pentru ARN U. În ambele cazuri, guanozina metilată a coifului este legată de primul nucleotid transcris printr-o legătură trifosfat cu nucleotidele din pozițiile lor 5'. Aproape toate moleculele de ARNm posedă și o **coadă poliadenilată** (poli A) caracteristică care debutează la distanțe diferite dincolo de sfârșitul regiunii codante; coada poli A nu este codificată de gena care a determinat obținerea transcriptului ci este adăugată în urma unei reacții post-transcripționale. La drojdii și alte eucariote inferioare lungimea cozii poli A este de ordinul a 75 la 185 resturi. Moleculele de ARNm nuclear de la majoritatea vertebratelor au o coadă poli A de 200 la 300 nucleotide, dar extensia poli A a moleculelor ARNm citoplasmatică este mai scurtă și caracteristică fiecărui tip de ARNm în parte. Din contră, multe molecule de ARNm care codifică histone sau ARN U sunt lipsite de această coadă. Anumite molecule ARNm conțin adenin N6-metilate și toate moleculele ARN U au caracteristic un număr mare de resturi uracil modificate. Aceste modificări sunt probabil realizate în timpul sau după transcripție, funcțiile lor fiziologice fiind necunoscute.

Transcripția genelor din clasa III este realizată de **ARN polimeraza III**. Producții genelor din clasa III sunt molecule mici de ARN implicate în sinteza proteinelor (ARNt și ARNr 5S), în transportul intercelular al proteinelor (ARN 7 SL) și în maturarea post-transcripțională (ARN U6). Genele din clasa III codifică și multe alte molecule de ARN al căror rol fiziologic este necunoscut (ARN 7SK, ARN Alu). Genoamele adenovirusurilor posedă două gene de clasă III, VA1 și VA2, ai căror produși de transcripție (de 157 și 140 nucleotide) deturneză mașinăria de transcripție a celulei infectate făcând-o mai eficace pentru sinteza proteinelor necesare structurii virale. Două molecule mici de ARN, EBER I și EBER II (166 și 172 nucleotide) sunt și ele produse de gene din clasa III ale genomului virusului Epstein-Barr.

**Cele trei ARN polimerazele de la drojdii prezintă patru subunități centrale (core)** care exprimă aceeași omologie cu subunitățile  $\beta$ ,  $\beta'$  și subunitățile  $\alpha$  de la *E. coli* (SLIDE 35).

Cea mai mare subunitate (L') de la ARN polimeraza II conține și domeniul C terminal (CTD). ARN polimeraza I și III conțin aceleași două subunități *like- $\alpha$*  neidentice, în timp ce ARN polimeraza II are două copii a unei subunități diferite *like- $\alpha$* . Toate trei polimerazele prezintă alte cinci subunități comune (două copii ale celei mai mari dintre acestea). În plus, fiecare polimerază din drojdie conține de la 4 la 7 subunități mici unice.

## ARN polimeraza II

**Genele din clasa II sunt transcrise în nucleu de către ARN polimeraza II.** Enzima a fost izolată de la numeroase organisme, printre care drojdiile din genul *Physarum*, *Aspergillus*, *Acanthamoeba*, plantele superioare, insecte, *Xenopus* și mamifere. Compoziția în subunități a enzimelor din diferite organisme este foarte asemănătoare. ARN polimeraza II este constituită din 9-10 subunități. Toate enzimele izolate de la diferite tipuri de organisme eucariote și analizate conțin două subunități mari: o subunitate de ~ 220 kD, care este cea mai mare polipeptidă pe care o regăsim în structura celor trei tipuri de ARN polimeraze de la EK și o altă subunitate de ~140 kD. Trei subunități mici ale ARN polimerazei II sunt comune cu cele ale ARN polimerazelor I și III, celelalte patru sau cinci subunități sunt specifice enzimei. Existența unei similitudini structurale și chimice între cele două subunități mari la cele trei ARN polimeraze eucariote a fost inițial pusă pe seama reacțiilor imunologice încrucișate pe care acestea le dădeau. În plus, cele două subunități mari ale fiecăreia dintre cele trei polimeraze eucariote posedă secvențe în aminoacizi cu un înalt grad de asemănare cu secvențele subunităților mari ale ARN polimerazei de la *E. coli*.

Dintre cele trei ARN polimeraze, ARN polimeraza II este cea mai sensibilă la  $\alpha$ -amanitină. Toxina permite inițierea sintezei lanțului de ARN, dar blochează elongarea acestuia. Situsul de acțiune al  $\alpha$ -amanitinei pare să fie situat la nivelul subunității de 220 kD a ARN polimerazei II. La *Drosophila* și la *S. cerevisiae* mutațiile care conferă rezistența la amanitină sunt localizate în gena care codifică pentru subunitatea de 220 kD. Această subunitate mare este și subiectul reacțiilor de fosforilare și defosforilare la nivelul mai multor situsuri. Forma înalt fosforilată a enzimei, găsită *in vivo* atunci când activitatea de transcripție este importantă, este considerabil mai activă decât forma defosforilată a acesteia.

**Domeniul carboxi-terminal** al subunității celei mai mari a ARN polimerazei II de la drojdie, de la *Drosophila* și de la mamifere conține multiple **repetiții în tandem ale hexapeptidului Tyr-Ser-Pro-Thr-Ser-Pro-Ser**. Numărul acestor copii este 26 la drojdie, respectiv la 52 la șoarece. Această secvență repetitivă este unică și esențială pentru activitatea enzimei. Modificarea acestei regiuni în gena de drojdie demonstrează că subunitățile care conțin mai puțin de 13 copii nu sunt funcționale *in vivo*. Serinele și treoninele, abundente în această regiune, pot reprezenta situsurile de fosforilare menționate mai sus, dar **rolul precis al acestui heptamer nu este deplin elucidat** (slide 36).

## Factorii de transcripție la eucariote

ARN polimeraza II, ca și celelalte ARN polimeraze necesită o serie de proteine adiționale pentru a forma un complex de transcripție funcțional. Multe sunt proteine care se fixează pe ADN și recunosc una sau mai multe secvențe care, par să constituie promotorul genei. Anumiți **factori de transcripție** reacționează prin intermediul **interacțiilor proteină-proteină cu alți factori** de transcripție și sunt capabili astfel să modifice astfel specificitatea și afinitatea acestora. Prin **interacții mutuale sau cu diferite secvențe de ADN**, diferiții **factori de transcripție** formează **ansamble proteice complexe** care **reglează capacitatea ARN polimerazei II de a iniția transcripția**. Cea mai mare parte a complexelor reacționează pozitiv, crescând frecvența inițierilor, dar sunt cunoscute și altele care acționează negative, fiind deci represori. Există chiar și factori care stimulează transcripția unei gene, dar inhibă transcripția altei gene. Mulți dintre acești **factori** sunt **specifici unui țesut sau unor celule** și nu reacționează decât în anumite stadii de dezvoltare; acest lucru permite genelor să fie transcrise diferit în funcție de țesut sau de timp. Într-un anumit sens, **ARN polimeraza reprezintă mașinăria de transcripție, iar interacțiile factorilor cu semnalele de reglare de pe ADN sau între ei, determină unde, când și cu ce viteză va funcționa această mașinărie**. Numărul de factori proteici care acționând *în trans* sunt implicați în sistemele de transcripție utilizate de ARN polimeraza II este mereu în creștere, cu toate că nu toți au fost foarte bine caracterizați. Complexitatea factorilor de transcripție este atât de mare, încât anumiți factori se leagă la mai mult de o secvență de ADN și anumite secvențe de ADN reprezintă situsuri de fixare pentru mai multe tipuri de factori. În plus, anumite motive organizate în tandem se suprapun, creând adesea situsuri de legătură suplimentare absente în motivele individuale unde favorizează fixarea unui factor sau a altuia.

### Concluzii (SLIDE 39)

- principalul scop al controlului genetic în organismele multicelulare este realizarea deciziilor precise de dezvoltare astfel încât gene particulare sunt exprimate în celule particulare în timpul dezvoltării și diferențierii celulare;
- controlul transcripțional este principalul sens al reglării expresiei genice atât la eucariote, cât și la procariote;
- în genoamele eucariote, elementele de control care acționează *în cis* și care reglează expresia de la nivelul promotorului sunt adesea localizate la multe kb depărtare de situsul

de START. În contrast, elementele de control bacterian sunt dispuse, în general, pe o distanță de 60 pb față de promotorul pe care aceștia îl reglează;

Eucariotele conțin trei tipuri de ARN polimeraze nucleare. Toate trei conțin două subunități mari și două subunități mici care constituie structura centrală omologă cu subunitățile  $\beta$ ,  $\beta'$  și  $\alpha$  ale ARN polimerazei de la *E. coli*, cât și numeroase subunități mai mici adiționale. Unele dintre aceste subunități mici sunt comune, iar altele sunt unice pentru fiecare polimerază.

ARN polimeraza I sintetizează numai pre-ARNr, ARN polimeraza II sintetizează ARNm și unele molecule mici de ARN nuclear care participă la splicingul ARNm, iar ARN polimeraza III sintetizează ARNt, ARNr 5S și multe alte molecule relativ scurte și stabile de ARN.

O secvență heptapeptidică, domeniul carboxi-terminal (CTD), din subunitatea cea mai mare a ARN polimerazei II este fosforilată în timpul inițierii și rămâne fosforilată cât timp enzima transcrie matricea. Similar cu ARN polimerazele bacteriene, ARN polimeraza II inițiază, de obicei, transcripția genelor la nivelul unor perechi de baze specifice sau baze vecine alternative din ADN matricea. Nucleotidul din 5' căruia i se adaugă capșonul la nivelul ARNm corespunde nucleotidului de pe catena matricea la care transcripția este inițiată.

**Majoritatea genelor eucariote sunt reglate prin mai multe mecanisme de control** (slide 44, 45)

a) Genele organismelor multicelulare conțin atât elementele proximale promotorului și "enhancerii" (activatorii) cât și TATA box sau alte elemente promotoare. Ultimile poziționează ARN polimeraza II pentru a iniția transcripția la situsul de START și influențează rata transcripției. Activatorii pot fi poziționați fie în amonte fie în aval și la o distanță de aproximativ 50 kb față de situsul de START al transcripției. În unele cazuri, elementele din proximitatea promotorului apar la fel de bine și în aval de situsul de START al transcripției.

b) Majoritatea genelor de drojdii conțin numai o regiune reglatoare, numită "secvența activatoare în amonte" (Upstream Activating Sequence-UAS) și o TATA box, care este de aproximativ 90 pb în amonte de situsul de START.

**Factorii implicați în transcripție sunt identificați prin tehnici biochimice** (SLIDE 46).

Ca și la *E. coli*, diferitele elemente transcripționale de control prezente la eucariote sunt situsuri de legare a proteinelor reglatoare care acționează în trans. Experimental a fost realizată identificarea, purificarea și determinarea structurii acestor factori de transcripție care sunt activatori și represori ai genelor proteice care codifică pentru proteine.

După identificarea elementelor reglatoare cu ajutorul analizelor mutaționale descrise anterior, pasul următor îl reprezintă identificarea proteinelor care recunosc aceste structuri și se leagă specific la ele.

**Figura** prezentată pe slide evidențiază modul în care un extract proteic nuclear este supus mai multor etape cromatografice și apoi unor analize de “footprinting” cu DN-aza I sau de “întârziere în gel” folosind fragmente de ADN care conțin elementele reglatoare în cis deja identificate. Există probabilitatea ca fracțiile care conțin proteine ce se leagă la elementele reglatoare conțină și un posibil factor de transcripție. O tehnică bună pentru etapa de purificare finală a factorilor de transcripție o constituie un tip particular de cromatografie de afinitate specifică unei secvențe de ADN. Ca un test final pentru a evalua capacitatea proteinelor izolate de a stimula transcripția unei matrici care conține un situs corespunzător de legare a proteinei, se realizează o reacție de transcripție *in vitro*.

O dată ce factorul de transcripție a fost izolat, secvența sa parțială în aminoacizi poate fi determinată și folosită pentru clonarea genei sau a ADNc codificat de aceasta. Gena izolată poate fi folosită apoi pentru a testa capacitatea proteinei codificate de a stimula transcripția într-o reacție de transcripție *in vitro*.

În figura de pe slide-ul 46 sunt prezentate schematic etapele unui proces de ***purificare a factorilor de transcripție***.

- a) Pentru purificarea unui factor de transcripție care se leagă specific la un element reglator din structura ADN sunt necesare mai multe etape cromatografice. Purificarea finală este obținută prin cromatografie de afinitate pe o coloană care este cuplată cu catene lungi de ADN care conțin mai multe copii al situsului de legare a factorului proteic. Un preparat proteic parțial purificat care conține factorul de transcripție este aplicat pe o coloană cu o tărâie ionică scăzută (100 mM KCl). Proteinele care nu se leagă deloc la situsurile specifice vor fi eluate cu tamponare cu salinitate scăzută. Proteinele cu afinitate scăzută pentru situsul de legare sunt spălate de pe coloană cu tamponare de salinitate intermediară (300 mM KCl). În final, factorii proteici înalt purificați sunt eluați cu soluții saline concentrate (1 M KCl);
- b) Proteinele separate prin cromatografie pe coloană sunt testate pentru capacitatea lor de legare la un element reglator identificat. În acest exemplu, testările ADN “footprinting”



pentru proteinele separate prin cromatografie de schimb ionic indică că fracțiile de interes sunt 9-12. Aceste proteine pot fi testate și prin tehnica de întârziere în gel.

Ulterior izolării și purificării factorilor de transcripție este realizată determinarea *in vitro* a activității factorilor de transcripție (SLIDE 47). Sistemul experimental testarea *in vitro* a activității factorilor de transcripție prezentat în imagine necesită prezența a două plasmide. O plasmidă conține gena care codifică pentru un posibil factor de transcripție- proteina X.

Cea de a doua plasmidă conține o genă raportor și unul sau mai multe situsuri de legare a proteinei X. Ambele plasmide sunt simultan introduse în celulele gazdă cărora le lipsesc gena care codifică pentru proteina X și gena raportor și este măsurată transcripția genei raportor; alternativ, activitatea proteinelor codificate poate fi determinată. Dacă transcripția genei raportor este mai mare în prezența plasmidei care codifică proteina X, înseamnă că proteina X este un activator; dacă transcripția este mai scăzută, atunci proteina X joacă rol de represor. Prin folosirea unor plasmide care codifică un factor de transcripție mutant sau care conține o rearanjare, pot fi identificate domeniile importante ale proteinei.

### **Concluzii (SLIDE 63)**

Expresia genelor eucariote care codifică pentru proteine este reglată prin intervenția unor elemente multiple care acționează sub forma unor regiuni de control în cis. Unele elemente de control sunt localizate în apropierea situsului de START (elemente din proximitatea promotorului), în timp ce altele sunt localizate la distanță (activatori sau “enhanceri”).

Promotorii determină situsul de inițiere a transcripției și legarea directă a ARN polimerazei II. Au fost identificate trei tipuri de secvențe promotor pentru ADN de la eucariote, dintre care, TATA box reprezintă tipul cel mai comun și este dominant pentru genele transcrise rapid. Promotorii de tip inițiator sunt reprezentați cu o frecvență scăzută în anumite gene, în timp ce insulele CpG sunt caracteristice genelor transcrise.

Elementele din proximitatea promotorului sunt localizate într-un interval de 200 pb față de situsul de START. Multe astfel de elemente, conținând mai puțin de 20 pb, pot fi utile în reglarea unei gene particulare.

Activatorii, care au de obicei o lungime de aproximativ 100-200 pb, conțin mai multe elemente de control de 8 la 20 pb. Acestea pot fi localizate de la 200 pb la 10 kb în amonte sau în aval de promotor, în interiorul unui intron, sau în aval de exonul final al genei. Elementele din

proximitatea promotorului și activatorii sunt adesea specifici unui tip celular, funcționând numai tipuri celulare diferențiate specific.

Factorii de transcripție, care stimulează sau reprimă transcripția, se leagă la elemente situate în proximitatea promotorului și la “enhancers” (activatori) din structura ADN la eucariote;

Activatorii sunt în general proteine modulare, care conțin un singur domeniu de legare și unul sau mai multe domenii activatoare; diferitele domenii sunt adesea legate prin regiuni polipeptidice flexibile. Acestea permit domeniilor diferiților activatori să interacționeze chiar și când domeniile de legare la ADN recunosc situsuri separate de zeci de perechi de baze;

**Activatorii conțin, în general, mai multe situsuri de legare pentru factorii de transcripție** grupate sub formă de clustere. Legarea cooperativă a mai multor activatori la situsuri învecinate din structura unui activator formează un complex multi-proteic numit “**complex activator**”. Asamblarea acestuia necesită adesea proteine mici care se leagă la fosa minoră a ADN și determină curbarea rapidă a secvenței, permitând proteinelor de pe fiecare parte a curburii să interacționeze mult mai bine;

**Majoritatea represorilor** de la eucariote sunt **proteine modulare**. Similar cu activatorii, aceștia conțin de obicei un singur domeniu de legare și unul sau mai multe domenii represoare, și pot controla transcripția când sunt legați la situsuri situate la sute sau mii de baze față de situsul de START.

**Domeniile de legare la ADN ale factorilor de transcripție de la eucariote exprimă o varietate de structuri**. Printre cele mai comune motive structurale sunt **homeodomeniile, domeniile bazice “fermoar de leucină” (leucine zipper), HLH** și diferite tipuri de “degete zinc” (zinc finger).

În general una sau mai multe  $\alpha$ -elice din domeniul de legare la ADN interacționează cu fosa majoră la nivelul unor situsuri de recunoaștere. Capacitatea anumitor factori de transcripție de a forma heterodimeri crește numărul de situsuri ADN pe care acești factori le pot controla și modalitățile prin care le pot controla. Deși, anumite domenii de activare și de represare sunt bogate în aminoacizi particulari, aceste domenii funcționale manifestă o varietate de secvențe de aminoacizi și structuri proteice caracteristice diferiților factori transcripție.

### **Complexul ARN polimerazic II de inițiere a transcripției (SLIDE 64)**

Inițierea de către Pol II necesită **factori generali de transcripție**, care **poziționează Pol II la nivelul situsului de inițiere** și sunt necesari pentru **transcrierea majorității genelor** transcrise

de către această polimerază. Factorii generali de transcripție sunt structuri multimerice și înalt conservate.

Proteinele cuprinse în complexul ARN polimerazei II de inițiere a transcripției se asamblează într-o ordine specific *in vitro*, dar majoritatea proteinelor se pot combina pentru a forma un complex holoenzimatic *in vivo*.

Așa cum a fost precizat anterior, ARN polimeraza de la *E. coli* lipsită de subunitatea  $\sigma 70$  nu poate iniția transcripția. Totuși, când *core* polimeraza este asociată cu subunitatea  $\sigma 70$ , holoenzima rezultată poate iniția transcripția *in vitro*, pornind de la promotori puternici. Astfel, subunitatea  $\sigma 70$  funcționează ca un factor de inițiere care este absolut necesar pentru începerea transcripției și care este eliberat de pe matricea o dată ce s-a produs polimerizarea primilor aproximativ 10 ribonucleotidtrifosfați.

Din contră, la EK, transcripția *in vitro* cu ARN polimeraza II purificată necesită adăugarea unui număr mare de **factori de inițiere** care sunt separați de complexul polimerazic în timpul purificării. Acești factori de inițiere, care poziționează ARN polimeraza II la nivelul situsurilor de inițiere a transcripției, sunt numiți **factori generali de transcripție**, deoarece se consideră că ei sunt necesari pentru transcripția majorității genelor de către acest tip de polimerază. Un complex de inițiere a transcripției conține ARN polimeraza și diferiți factori generali implicați în transcripție care se leagă la regiunea promotoare.

Mulți dintre factorii generali de transcripție necesari polimerazei II pentru inițierea transcripției de la nivelul majorității promotorilor care conțin o TATA box au fost izolați și caracterizați. Aceste proteine au fost desemnate cu terminologia **TFIIA**, **TFIIB**, etc, majoritatea fiind proteine multimerice. **TFIID** este unul dintre cei mai mari factori având o masă moleculară de aprox 750 kDa. Acesta conține o singură proteină de aprox 38 kDa numită **TATA box-binding protein (TBP)** și **11 factori asociați TBP numiți TAF** (TATA Associated factors) care nu au fost foarte bine caracterizați. Factori generali de transcripție cu activități similare au fost izolați din celule umane în cultură, ficat de șobolan, embrioni de *Drosophila* și drojdii.

Genele care codifică pentru aceste proteine la drojdii au fost secvențializate o dată cu secvențializarea completă a genomului de drojdii ca și în cazul ADNc uman și a celui de la *Drosophila*. În toate cazurile, factorii de transcripție echivalenți de la diferitele tipuri de eucariote sunt foarte bine conservați. Deși factorii generali de transcripție permit polimerazei II să inițieze

transcripția *in vitro* la nivelul aceluiași situsuri prezente *in vivo*, protein adiționale sunt necesare pentru inițierea transcripției *in vivo*.

### **Etapele procesului de asamblarea a complexului Pol II de inițiere a transcripției *in vitro*** (SLIDE 65, 66)

În figură sunt prezentate schematic etapele asamblării complexului de inițiere a transcripției pornind de la ARN polimeraza II și factori de transcripție izolați.

În momentul în care întreg complexul de inițiere a fost asamblat, se produce separarea catenelor ADN la situsul de START pentru a forma complexul deschis, proces care necesită hidroliza ATP. Pe măsură ce transcripția începe să se deruleze de la nivelul promotorului, domeniul CTD al ARN Polimerazei II se fosforilează și factorii generali de transcripție disociază de complexul de legare TBP de la nivelul promotorului. Numeroase alte proteine participă la inițierea transcripției *in vivo*.

În majoritatea studiilor biochimice asupra asamblării complexului de inițiere al ARN Pol II, cercetătorii au folosit mai degrabă TBP (TATA Binding Protein) decât complexul TFIID care este format din multe subunități și care este dificil de purificat. Legarea TBP și a altor factori generali de transcripție la promotorii cu secvențe consens TATA a fost analizată prin tehnica de “foot-printing” cu DN-aza I și prin tehnica electroforetică de întârziere în gel. Aceste studii demonstrează că complexul Pol II de inițiere este asamblat *in vitro* în conformitate cu secvența de reacții determinată și prezentată în imagine.

*In vitro*, TBP este prima proteină care se leagă la promotorul TATA box. Toate proteinele eucariote TBP analizate prezintă domenii C-terminale foarte similar, de aproximativ 180 de aminoacizi. Secvența genică care codifică aceste regiuni prezintă o omologie de 80% de la drojdia la om, majoritatea diferențelor fiind substituții conservative. Acestea conservă funcțiile domeniului C-terminal, ca și întreaga lungime a proteinei în urma transcripției *in vitro*. Domeniul N-terminal al TBP, care variază foarte mult ca secvență și lungime la diferite eucariote, funcționează în transcripția genelor care codifică pentru ARNsn.

Factorii de transcripție generali se leaga secvențial la ARN polimeraza II purificata (Pol II), la nivelul secvenței TATAbox pentru a forma complexul de pre-initiatie. Hidroliza ATP furnizeaza energia pentru desfacerea moleculelor de ADN la situsul de START prin legarea subunitatii TFIID. Inițierea transcripției de către Pol II conduce la formarea complexului deschis, iar polimeraza se deplasează de la nivelul promotorului și domeniul CTD este fosforilat. *In vitro*,

factorii generali de transcripție (cu excepția TBP) disociază de complexul TBP și promotor, dar nu se știe încă care dintre factori rămân asociați cu regiunea promotoare pentru fiecare rundă de inițiere a transcripției *in vivo*.

### **Concluzii (SLIDE 69)**

ARN polimeraza II necesită mai mulți factori generali de transcripție pentru a localiza clar punctul de START la nivelul matriței ADN și a iniția transcripția. Dintre acești factori, TFIID care se leagă la TATA-box prin intermediul TBP.

Transcripția genelor care codifică pentru proteine este controlată de către ARN polimeraza II care poate fi inițiată *in vitro* prin legarea secvențială a factorilor: TBP, care se leagă la TATA box; TFIIB; un complex Pol II și TFIIF; TFIIE și final TFIIF;

Activitățile helicazice ale celor două subunități TFIIF separă catenele la nivelul situsului de START la majoritatea promotorilor, un proces care necesită hidroliza ATP. După ce ARN polimeraza II începe transcrierea dincolo de situsul de START, elementele CTD sunt fosforilate de către o altă subunitate a TFIIF.

Inițierea de către Pol II *in vivo* necesită un complex multimediator multiproteic, care se asociază cu elementele CTD nefosforilate ale Pol II, formând un mare complex holoenzimatic care include și majoritatea factorilor de transcripție. Se crede că această holoenzimă preasamblată se leagă la ADN promotor într-o singură etapă *in vivo*.

Complexul de inițiere a transcripției care se assemblează la nivelul promotorilor *in vivo* poate cuprinde 60-70 polipeptide, cu o masă totală similară cu cea a unui ribozom.

**Inițierea transcripției realizată de Pol I și Pol III este analogă celei realizată de Pol II (SLIDE 92, 94, 95).**

ARN polimeraza I (Pol I) este dedicată sintezei pre-ARNr, iar ARN polimeraza III (Pol III) transcrie genele ARNt, genele ARNr 5S și genele care codifică multe alte molecule de ARN mici. Formarea complexelor de inițiere a transcripției care implică Pol I și Pol III este similară în unele privințe cu asamblarea realizată în cazul Pol II. Totuși, fiecare din cele trei ARN polimeraze eucariote nucleare necesită factori de transcripție specifici (sau de inițiere) care recunosc elemente de control diferite. Pe de altă parte, nici Pol I și nici Pol III nu necesită hidroliza ATP pentru a iniția transcripția, în timp ce Pol II folosește energia rezultată din hidroliza ATP.

### **Inițierea transcripției de către Pol I**

Genele umane pentru pre-ARNr prezintă două regiuni de control: un element **core**, care include situsul de START și care este esențial pentru transcripție și un element de control din amonte - **UCE (upstream control element)** de aproximativ 50 pb, care începe la aproximativ 100 pb fata de situsul de START, și care stimulează transcripția *in vitro* de aproximativ 10 la 100 ori. Două proteine de legare la ADN asistă Pol I pentru a iniția și transcrie corect genele pre-ARNr. Primul dintre ele este factorul de legare din amonte **UBF (Upstream Binding Factor)**, care se leagă atât la UCE, cât și la porțiunea din amonte a elementului core, chiar dacă se consideră că acestea au în aparență o secvență puțin similară. Factorul de selectivitate I (**SL1**), o proteină multimeră se leagă și stabilizează complexul. După legarea subunităților SL1, Pol I se leagă și inițiază transcripția. Una dintre subunitățile SL1 este similar TBP (TATA Binding Protein), proteină cu care joacă rolul central în inițierea transcripției de către Pol II (**Fig. a**), **SLIDE 92**).

### **Inițierea transcripției de către Pol III**

Spre deosebire de genele care codifică pentru proteine și pre-ARNr, **regiunile promotoare ale genelor pentru ARNt și ARNr 5S sunt în întregime în interiorul secvenței transcrise**. Două astfel de **elemente promotoare interne**, numite **box (cutia) A** și **box (cutia) B** sunt prezente în toate genele pentru ARNt. Aceste secvențe înalt conservate nu funcționează numai ca promotori dar codifică și două porțiuni ale moleculelor de ARNt la eucariote care sunt necesare sintezei de proteine. **La nivelul genelor ARNr 5S se află numai o regiune de control internă numită box (cutia) C**, care **acționează ca un promotor**. Factorii generali de transcripție necesari Pol III pentru a iniția transcripția genelor pentru ARNt și ARNr 5S au fost bine caracterizați la *S. cerevisiae*. Doi factori multimeri, TFIIC și TFIIIB participă atât la inițierea ARNt și ARNr 5S la drojzii. Asamblarea complexului de inițiere ale genelor ARNt începe prin legarea singulară a TFIIC la cutia A și cutia B. Interacția factorului TFIIC cu TFIIIB îl direcționează pe acesta din urmă să se lege la secvențe de aproximativ 30 pb situate în amonte de situsul de START al transcripției. Asamblarea complexului de inițiere a transcripției genelor pentru ARNr 5S începe prin legarea factorului de transcripție TFIIIA la cutia C. Apoi, TFIIC se leagă la TFIIIA și este poziționat prin intermediul interacțiilor proteină-proteină într-o poziție similară față de situsul de START ca și în cazul legării TFIIC la genele pentru ARNt. TFIIIB interacționează analog cu TFIIC și se leagă apoi la o secvență ADN situată la aproximativ -30 pb, asemeni legării la gena ARNt.

Jumătatea N-terminală a uneia dintre subunitățile TFIIB, numită BRF (TFIIB Related Factor), este similară cu secvența TFIIB (un factor al Pol II). Similaritățile sugerează că BRF și TFIIB realizează o funcție similară în inițiere, adică direcționarea polimerazei la situsul corect de START. Legarea TFIIB atât la genele pentru ARNt cât și pentru ARNr 5S determină legarea Pol III, iar aceasta poate iniția transcripția în prezența ribonucleotidtrifosfaților. Subunitatea BRF a TFIIB interacționează specific cu una dintre subunitățile polimerazice unice pentru Pol III, aceasta fiind elementul specific important pentru inițierea procesului de transcripție de către Pol III. După legarea TFIIB, TFIIC poate fi înlăturat fără a afecta inițierea realizată de către Pol III. Astfel, TFIIC se poate considera că ca fiind un factor de asamblare critic legarea factorului de inițiere TFIIB.

### **Concluzii (SLIDE 96)**

- inițierea transcripției de către Pol I este dirijată de către un element promoter “core”, care se suprapune cu situsul de start, și o regiune de control situată în amonte (UCE). Aceasta polimeraza necesită doi factori de transcripție generali, SL1 și UBF;
- inițierea transcripției de către ARN polimeraza III este cel mai adesea dirijată de către elemente promotore interne. Doi factori generali de transcripție sunt necesari pentru inițierea transcripției ARNt (TFIIC și TFIIB); un factor suplimentar (TFIIIA) este necesar pentru inițierea transcripției genelor ARNr 5S;
- inițierea transcripției de către toate cele trei ARN polimeraze nucleare eucariote necesită un factor de transcripție specific general care conține TBP ca subunitate;
- inițierea de către Pol I și Pol III nu necesită ATP, spre deosebire de inițierea realizată de Pol II care depinde de hidroliza ATP.

### **NOTE DE CURS BM4**

#### **Terminarea transcripției (SLIDE 6)**

- Există mai multe mecanisme care reglează terminarea transcripției la bacterii și celulele eucariote.
- **La bacterii** principalele mecanisme implică **ARN polimeraza**, unul dintre mecanisme necesitând și **factorul de terminare Rho**.
- **La eucariote**, mecanismele pentru terminarea transcripției diferă pentru cele trei tipuri de ARN polimeraze.

## Terminarea transcripției ADNr

Deoarece transcripții pre-ARNr se termină în proximitatea sau exact la extremitatea 3' a secvenței ARNr 28S, s-a crezut mai întâi că această regiune conținea un terminator recunoscut de ARN polimeraza I. Experiențele ulterioare au demonstrat că transcripția *in vitro* a genelor care codifică ARNr continuă dincolo de extremitatea 3' a ARNr 28S matur. La *Xenopus*, transcripția se realizează dincolo de regiunea care conține separatorul intragenic și se termină la aproximativ 200 pb în amonte de centrul promotorului pre-ARNr 35S următor.

La drojdii, transcripții care încep la nivelul centrului promotorului, se termină la nivelul separatorului intergenic, dincolo de ARNr 5S și la 300 pb în amont de începerea secvenței pre-ARNr următor.

La ADNr de șoarece, terminarea transcripției are loc între 550 și 600 pb dincolo de extremitatea 3' a ADNr 28S. Semnalul de terminare a acestui sistem este o secvență foarte conservată poziționată pe catena sens (5'-AGGTCGACCAATTANTCCG-3' numită **cutia Sal I**) și precedată, în general, de unul sau mai multe regiuni cu resturi pirimidinice, bogate în timină (T). Se găsesc 8 regiuni de acest tip între perechile de baze 500 și 1 200 plecând de la extremitatea 3' a ADNr 28S, dar în general terminarea se realizează în grupul T care precede cutia Sal I. Secvența conservată de 18 pb poate ea însăși să determine terminarea, dar aceasta este mai eficientă și mai precisă când cele 18 pb sunt precedate și urmate de regiuni bogate în T. Deleții mici și inserții sau numai schimbarea unor perechi de baze la nivelul segmentului de 18 pb, la fel ca și inversarea sa, duc la pierderea activității sale de terminator. Se consideră că un aranjament destul de asemănător de secvențe repetitive conservate (5'-GGGTTGACCA-3') furnizează probabil semnalul de terminare a pre-ARNr la om. Secvențe repetitive conservate (5'-GACTTGC-3' precedate de grupări de T) există în separatorii intergenici de la *Xenopus*, terminarea realizându-se cel mai adesea la nivelul secvenței repetitive care precede promotorul situat la extremitatea 3' a regiunii separatoare.

Există argumente indirecte care sugerează că terminarea la nivelul acestor situsuri depinde de fixarea unor proteine pe aceste situsuri specifice. Aceste indicații au fost confirmate prin demonstrația că unul sau mai mulți factori proteici din extractul nuclear se fixează de secvența de 18 pb. În acest caz, fixarea proteinei la secvența de terminare a fost evidențiată de protecția față de digestie cu endonuclează III a unui fragment marcat la 5' și care conținea acest semnal. Principiul acestei experiențe este că digestia enzimatică a unui segment este oprită de proteina



legată. Trebuie să ne amintim că endonucleaza III digeră ADN bicatenar pornind de la extremitățile 3'. În aceste condiții vor rezulta fibre marcate în 5', care vor avea lungimea determinată de locul în care digestia a fost blocată de proteina legată. Cum fiecare catenă poate fi marcată independent la capătul său 5', se vor putea determina cele două limite ale secvenței blocate.

### **Cuplarea terminării și inițierii transcripției**

O caracteristică originală a secvențelor de terminare este că acestea sunt adeseori situate imediat în amonte de promotorii genelor care codifică pentru ARNr următor. O curiozitate o reprezintă faptul că eliminarea sau modificarea acestor secvențe de terminare provoacă o reducere marcantă a transcripției care demarează la poziția + 1 a următorului promotor. Inițial, s-a presupus că un factor de terminare ar putea să reacționeze cu un factor de inițiere sau cu ARN polimeraza I, sau că activarea ar putea fi rezultatul localizării favorabile a enzimei la nivelul unor situsuri de juxtapunere apropiate promotorului ADNr. O altă explicație este că terminarea împiedică polimerazele ocupate cu transcrierea să se rupă complet de complexe de transcripție care sunt asamblate în imediata vecinătate a următorului promotor.

### **Terminarea transcripției de către ARN polimeraza III**

Dacă inițierea transcripției de către ARN polimeraza III necesită secvențe de reglare complexe și mai multe proteine, terminarea nu necesită decât enzima și un grup de resturi dezoxiadenilate la nivelul matricei. Terminarea se realizează perfect *in vitro* cu o ARN polimerază III înalt purificată și cu o matrice de ADN duplex. Experimental, au fost utilizate matrițe de transcripție formate din AND clonat special pentru a încerca definirea semnalului de terminare. Aceste experiențe au demonstrat că terminarea depinde de existența unui număr de resturi dezoxiadenilate grupate pe catenele matricei și de secvențele situate de o parte și de alta a acestui grup. De exemplu, patru dezoxiadenozine înconjurate de dezoxitimidină nu determină terminarea eficace, dar patru dezoxiadenozine flancate de o parte de CG și de cealaltă parte de GC (5'-GCAAACG-3'), cum este cazul ADNr 5S din celulele somatic de *Xenopus borealis*, sunt un semnal de terminare eficace. Mutațiile care modifică numărul de resturi adenilate reduc eficacitatea terminării. Sinteza ARN se oprește la nivelul unui rest uridilat complementar unei secvențe depărtate și nici a unor secvențe secundare. Dacă extremitatea 3' a ADNr 5S este eliminată din gena clonată, transcripția este continuată până când polimeraza III întâlnește un semnal adecvat în secvența flancantă a vectorului. Dacă o serie de dezoxiadenozine este inserată

la un situs incorect din structura matricei unei gene din clasa III, terminarea se realizează adesea la acest situs modificat.

A fost propusă o **secvența ipotetică de baze care formează un terminator eficient la *E. coli*** (SLIDE 7).

**Terminarea Rho-independentă apare la nivelul unor secvențe caracteristice ale ADN din *E. coli*** (SLIDE 8).

În figură este prezentată *Secvența de terminare a operonului triptofanului trp, un situs Rho independent.*

Operonul triptofan, compus din cinci gene (A-E) este urmat de o secvență de 36 pb la capătul căreia este localizată secvența semnal de terminare transcripției. Capătul 3' al ARNm corespunzător prezintă o structură sub formă de buclă bogată în GC care precede ultimele patru resturi U, o trăsătură caracteristică a ARNm produs de gene cu situsuri de terminare Rho independente. În general, situsurile de terminare bacteriene apar între operoni, dar nu între genele structurale din interiorul unui operon. Din acest motiv, genele din interiorul unui operon pot fi reglate coordonat, în timp ce genele din diferiți operoni sunt reglate independent.

**Terminarea prematură prin atenuare ajută la reglarea expresiei unor operoni bacterieni** (SLIDE 9).

Atenuarea furnizează un mecanism secundar de control a expresiei operonului Trp. Secvența Leader (L) (reprezintă situsul de legare a polipeptidei Leader), care este localizată între operator (O) și prima genă structurală (E) conține un situs atenuator de la nivelul căruia transcripția este terminată în funcție de concentrația citosolică a triptofanului. În figura respectivă, AUG indică codonii START pentru traducerea secvenței leader (roșu) și a ARNm pentru trpE (negru).

**Mecanismul de atenuare pentru transcripția operonului trp (SLIDE 10)**

a) ARNm corespunzător secvenței leader este constituit din 140 de nucleotide și este împărțit în patru regiuni diferite care pot forma structuri sub formă de buclă, dar numai una dintr aceste structuri poate determina atenuarea procesului de transcripție.

b) Translația secvenței trp leader începe de la capătul 5', imediat după ce ARNm corespunzător secvenței leader este sintetizat, în timp ce este realizată sinteza restului moleculei ARNm trp.

La concentrații crescute de triptofan, formarea buclei stem (între regiunile 3-4), urmată de o serie de resturi de U în 3', determină terminarea transcripției printr-un mecanism Rho-independent. La concentrații scăzute de triptofan, regiunea 3' este sechestrată în bucla stem format între regiunile

2-3 și nu se poate împerechea cu regiunea 4. În absența structurii sub formă de buclă necesară pentru terminare, transcripția secvențelor care codifică pentru triptofan continuă.

Siturile pentru **terminarea Rho dependentă** sunt prezente în unele gene ale fagului  $\lambda$  și ale *E. coli* (SLIDE 11).

- Factorul Rho factor este o proteină hexameră care se leagă specific la un segment de 70 – 80 baze al transcriptului ARN în formare.
- Factorul Rho alunecă apoi de-a lungul ARN în direcția 3' până când acesta desface hibridul ARN-ADN de la nivelul situsului activ al ARN polimerazei.
- Dacă transcripția este terminată sau dacă nu depinde de factorul Rho; ARN Polimeraza disociază numai dacă acest factor Rho este prezent la nivelul semnalelor stop dependente de factorul rho.
- Siturile Rho dependente nu posedă secvențe consens clare și mecanismul de terminare Rho dependent este caracteristic numai câtorva operoni.

**Cele 3 ARN polimerazele folosesc diferite mecanisme de terminare** (SLIDE 12).

- ARN polimeraza I realizează terminarea printr-un mecanism care necesită factori specifici de terminare care se leagă în aval de unitatea de transcripție.
- ARN polimeraza II conține semnalul de terminare la nivelul unei regiuni de 0.5-2 kb aparținând situsului de adăugare a cozii poli(A), iar terminarea este cuplată cu procesul care prin care este realizată clivarea și poliadenilarea capătul 3' al transcriptului.
- Transcripția realizată de către ARN polimeraza III este terminată după adăugarea unei serii de resturi U.

Transcripția genelor pre-ARNr de către ARN polimeraza I este terminată printr-un mecanism care necesită un factor de terminare specific polimerazei. Această proteină de legare la ADN se poziționează în aval față de unitatea de transcripție, diferit de modul de legare a factorului Rho de la *E. coli*, care este un factor de terminare care se leagă la ARN, o trăsătură a secvențelor stop dependente de factorul rho fiind că acestea se găsesc în ARN nou sintetizat și nu în tiparul ADN.

ARN polimeraza III purificată realizează terminarea după polimerizarea unei serii de U; spre deosebire de terminarea Rho independentă de la bacterii, totuși, terminarea realizată de ARN polimeraza III nu necesită o structură secundară sub formă de buclă stem situată în amonte în structura transcriptului ARN.

La majoritatea genelor de la mamifere care codifică pentru proteine, ARN polimeraza II poate termina transcripția la situsuri multiple localizate la nivelul unei regiuni de 0,5-2 kb aparținând situsului de agățare a cozii poli(A). Experimentele cu gene mutante demonstrează că terminarea este cuplată cu procesul care prin care se realizează clivarea și poliadenilarea transcriptului ARNm în formare, la nivelul unor secvențe specifice asociate cu domeniul fosforilat carboxil terminal (elementele/domeniul CTD) ale ARN polimerazei.

Acest complex de clivare/poliadenilare poate suprima terminarea, cât timp secvența care semnalizează scindarea și poliadenilarea este transcrisă de către polimerază. Studii pe mutații de drojzii indică că scindarea transcriptului în formare, mai mult decât poliadenilarea, reprezintă un eveniment critic care semnalează ARN polimerazei să termine transcripția.

Așa cum a fost prezentat anterior, terminarea transcripției la bacterii poate fi reglată prin atenuare și prin mecanisme de anti-terminare. Mecanisme analoge de control, care implică o decizie între elongarea catenei și terminare apar la unele gene transcrise de către ARN polimeraza II.

**Un exemplu de reglare a transcripției printr-un mecanism de anti-terminare este cel al transcripției genomului virusului HIV (SLIDE 14).**

În figura de pe slide este prezentat **complexul de anti-terminare** compus din **proteina HIV Tat** și alte proteine celulare eucariote. Elementul TAR al transcriptului HIV conține secvențe recunoscute de către proteina Tat și proteina celulară ciclina T. Ciclina T activează și ajută poziționarea protein kinazei Cdk9 lângă substratul său, domeniul CTD al ARN polimerazei II și astfel, transcripția virusului HIV este reglată printr-un mecanism de anti-terminare. Transcripția genomului virusului imunodeficienței (HIV) de către ARN polimeraza II furnizează cel mai bine înțeles exemplu de reglare a terminării transcripției la eucariote. Exprimarea eficientă a genelor HIV necesită a mică proteină virală codificată de locusul tat. Celulele infectate cu mutații tat- produc transcripți virali scurți care hibridizează cu fragmente de restricție care conțin regiuniile proximale promotorului ADN HIV, dar nu și cu fragmentele de restricție poziționate în aval față de promotor. În contrast, celulele infectate cu tipul sălbatic HIV sintetizează transcripți lungi care hibridizează cu fragmente de restricție corespunzătoare întregii și singurei unități de transcripție HIV. Astfel, proteina Tat funcționează ca un factor anti-terminator, care permite ARN polimerazei II să citească de-a lungul unui bloc transcripțional, mai mult decât proteina N a fagului lambda.

Deoarece mecanismul de anti-terminare coordonat de proteina Tat este necesar pentru replicarea HIV, înțelegerea în viitoare a acestui mecanism de control poate oferi posibilități de determinare efectivă a unor terapii pentru rezolvarea sindromului de imunodeficiență dobândită (AIDS, Acquired ImmunoDeficiency sindrom).

Asemănător proteinei N, proteina Tat este o proteină care se leagă la o secvență specifică din structura ARN. Ea se leagă la o copie ARN a unei secvențe numite TAR, care este localizată lângă capătul 5' al transcriptului HIV. Secvența TAR conține două situsuri de legare: unul care interacționează cu Tat și unul care interacționează cu o proteină celulară numită ciclina T.

Asa cum se observă în figură, proteina Tat a virusului HIV și ciclina T celulară se leaga cooperativ la secvența TAR a ARN. Interacția ciclinei T cu o protein kinază numită Cdk9 activează kinaza, al cărei substrat este domeniul CTD al ARN polimerazei II. Studiile de transcripție *in vitro* folosind inhibitori specifici ai Cdk9 sugerează că moleculele de ARN polimerază II care inițiază transcripția la nivelul promotorului HIV stopează procesul după transcrierea a aproximativ 50 de baze, numai dacă CTD este hiperfosforilată de Cdk9. Legarea cooperativă a ciclinei T și a proteinei Tat la secvența TAR de la capătul 5' al transcriptului HIV poziționează Cdk9, astfel încât aceasta poate fosforila domeniul CTD, prevenind astfel terminarea și permitând polimerazei să conțină elongarea lanțului. De remarcat că, Spt5 una dintre proteinele complexului ciclinei T-Cdk9, posedă o regiune a secvenței de aminoacizi omologă cu NusG, una dintre proteinele implicate în procesul de antiterminare dirijat de proteina N al bacteriofagului  $\lambda$ . Această omologie sugerează că există similitudini în mecanismul de reglare a elongării de la bacterii la animale.

## **Note de curs BM4**

### **PROCESAREA ARNm LA EUCARIOTE (SLIDE 15)**

La puțin timp după inițierea transcripției de către ARN polimeraza II la primul nucleotid al primului exon al unei gene, la capătul 5 al moleculei ARN în formare se adugă 7-metilguanină. Transcripția realizată de către ARN polimeraza II se termină la unul sau mai multe situsuri de terminare situate în aval de situsul poli(A), care este localizat la capătul 3' al exonului final. După ce transcriptul primar este scindat la nivelul situsului poli(A), se adaugă o prelungire formată din A. Coada poli(A) conține aproximativ 250 resturi la mamifere, 150 la insecte și 100

la drojdii. Pentru transcripții primari cu puțini introni, poliadenilarea, clivarea și splicingul urmează de obicei terminarea. Pentru genele mari, cu un număr mare de introni, intronii sunt adesea înlăturați din pre-ARNm în formare înainte ca transcripția genei să fie completă. De reținut că “capișonul” rămâne în moleculele de ARNm mature.

### ***Informații suplimentare***

#### *Terminarea transcripției și poliadenilarea*

Transcripția a numeroase gene eucariote se continuă dincolo de locul în care ARN este clivat pentru a se forma extremitatea 3' a ARNm matur. Terminarea are loc la situsuri diferite care se întind pe sute sau chiar mii de perechi de baze ADN. Astfel, regiunea în care se oprește transcripția  $\beta$ -globinei la vertebrate este situată la mai mult de 1 kpb de situsul de poliadenilare. La fel, transcripția primelor două gene în tandem ale  $\alpha$ -globulinei de la mamifere se oprește în regiunea de separare de aproximativ 2 kpb care separă cele două gene. Un exemplu interesant este furnizat de transcripția regiunii tardive a adenovirusurilor. În acest caz, transcripția continuă dincolo de cele cinci situsuri de poliadenilare, înainte de a se termina la capătul ADN viral.

Din contră, anumite gene eucariote par să aibă un semnal autentic semnal de terminare a transcripției. De exemplu, în gena gastrinei umane, semnalul de terminare este situat la 193 pb în aval de situsul de poliadenilare. Această secvență, care dictează terminarea transcripției, atât *in vivo*, cât și *in vitro*, este un segment bogat în T, cu următoarea secvență: 5'-T9A2T5AT4AT4AT5-3'. Transcripția unui ADN care posedă această secvență de-a lungul catenei sale se termină exact în 5' față de acest semnal. O secvență similară, dar cu câteva diferențe, este prezentă în amonte sau în aval de situsul de START la mai multe gene de vertebrate. Se presupune că acesta este un semnal de terminare și anumite secvențe bogate în T din genele de drojdii, 5'-CAATCTTG-3' și 5'-T5ATA-3', amintesc de semnalele de terminare dependente de rho dependente de la *E. coli*. Independent de prezența semnalelor specifice de terminare, transcripția depășește aproape întotdeauna situsurile de poliadenilare. În consecință, extremitățile 3' poliadenilate trebuie să fie create prin clivare endonucleotidică, urmată de polimerizarea de resturi adenilate la nivelul hidroxilului 3' astfel format.

Pentru determinarea situsului de clivare sunt necesare două secvențe situate la distanță mică. Una dintre acestea este cvasi-invariantă, AATAAA și este localizată între 10 și 30 pb în amonte de o secvență CA situată în apropierea situsului de clivare-poliadenilare. Înlocuirea lui T cu G

în AATAAA reduce puternic eficacitatea poliadenilării, dar cele câteva extremități formate sunt poliadenilate. Înlocuirea ultimului A din secvență cu G la capătul primei gene din tandemul celor două care codifică pentru  $\alpha$ -globulinei reduce producerea normală a  $\alpha$ -globulinei. Această experiență, la fel ca și altele, indică necesitatea absolută a secvenței AATAAA. Totuși, deoarece secvența AATAAA există și în regiunea codantă a numeroase gene, unde nu există poliadenilare, este evident necesară o a doua secvență pentru a provoca reacțiile de clivare și de poliadenilare. Localizarea acestui cel de-al doilea semnal a fost studiată examinând efectul delețiilor situate în vecinătatea situsului de poliadenilare. Experiențele demonstrează că secvența și poziția precisă a acestui semnal variază de la o genă la alta. În numeroase gene de vertebrate semnalul cel mai probabil este un segment bogat în GT sau T, cu secvența 5'-YGTGTGY (Y fiind o pirimidină), situat cam la 25 pb în aval de situsul de poliadenilare și adesea urmat la o distanță variabilă de o scurtă secvență de T. În anumite circumstanțe situsurile autentice de poliadenilare pot să nu fie efective. În maturarea ARNm tardiv al adenovirusului de tip 2, un singur situs de poliadenilare din cinci potențiale este utilizat pentru producerea fiecăruia din transcriptii primari. Abundența relativă a celor cinci clase de ARNm tardiv la adenovirus reflectă, în parte, frecvența cu care transcriptul primar este clivat și poliadenilat la unul din situsurile sale caracteristice. În acest caz, abundența relativă a moleculelor de ARNm produse plecând de la transcriptul primar de modifică în cursul infecției, indicând că alegerea situsului de poliadenilare este controlată.

Un al doilea exemplu este cel de trecere de la producerea formei membranare la cea a formei secretate în cazul imunoglobulinelor. Aceste două forme diferă printr-o regiune a lanțurilor lor grele. În acest caz poliadenilarea pre-ARNm declanșată de prima pereche de semnale, cea care este folosită pentru formarea ARNm care codifică forma secretată, este suprimată; molecula de ARNm este clivată conform semnalelor următoare, producând un pre-ARNm care poate fi maturat pentru a da naștere ARNm care codifică forma membranară, deci trecerea de la forma membranară la forma secretată rezultă din utilizarea reglării diferențiate a reacției de poliadenilare. Fenomenul și modalitățile sale de control nu sunt deplin elucidate.

Secvența AATAAA a fost identificată ca având rol și în terminarea transcripției. Această secvență situată la extremitatea 3' a secvenței codante a  $\beta$ -globinei majore de șoarece este necesară pentru terminarea care se realizează la o distanță de aproximativ 1,5 kb de situsul de poliadenilare. În plus, aceleași deleții și modificări care blochează poliadenilarea corectă reduc

și eficacitatea cu care se realizează terminarea transcripției în aval. Aceasta sugerează că recunoașterea situsului de poliadenilare trebuie să se realizeze înainte de terminarea transcripției la secvența adecvată situată în aval.

Ipoteza potrivit căreia complexul de transcripție nu achiziționează capacitatea de a realiza terminarea transcripției decât după ce a realizat transcrierea semnalului de poliadenilare constituie o modalitate interesantă de a asocia poliadenilarea și terminarea. Ea poate fi valabilă în cazul în care complexul de transcripție ar conține un factor care ar putea împiedica terminările inoportune în timpul transcripției matricei și l-ar pierde pe acesta la trecerea peste semnalul de poliadenilare. Dacă lucrurile s-ar petrece astfel, complexul de transcripție lipsit de acest factor s-ar putea desprinde de matrice la nivelul secvenței bogate în T pe care o întâlnește în continuare.

Acest model a fost propus ca urmare a studiilor realizate *in vitro* cu ARN polimerază II purificată, care demonstrează că transcripția mai multor tipuri de matrice ADN se termină frecvent la nivelul secvențelor bogate în T prezente în genă. Este posibil ca procesul de clivare a transcriptului la semnalul de poliadenilare să permită unei ARN nucleaze sau unei proteine de tip rho să se fixeze la polimerază și să inducă terminarea la nivelul secvenței bogate în T prezente în aval. Clivarea și poliadenilarea se pot realiza în diferite tipuri de extracte celulare, singura condiție fiind ca acestea să posedă moleculele ARN substrat purtătoare de semnale adecvate.

Studiile realizate *in vitro* cu astfel de sisteme au clarificat diferite aspecte ale acestor reacții. Astfel, clivarea și poliadenilarea se pot realiza independent una de cealaltă. În prezența unor analogi ai ATP (de exemplu cordicepina, care este 3' dezoxiadenozin trifosfat), clivarea se realizează la situsul exact, dar nu este urmată de poliadenilare. Poliadenilarea se poate realiza la nivelul radicalului –OH din extremitatea 3' terminal, care este precedată de o secvență AAUAA. Fraționarea unui extract de celule HeLa care au capacitate de clivare și de poliadenilare a unui substrat ARN furnizează multiple fracții proteice care, împreună, catalizează aceste două reacții. Una dintre componente este o poli A polimerază, o altă fracție conține una sau mai multe particule ribonucleoproteice nucleare mici (snRNP – small nuclear ribonucleic particles), iar cea de a treia este o proteină (64 kD) care se leagă la o regiune care conține secvența AAUAAA. Acești trei factori sunt necesari pentru a asigura cuplarea clivării cu poliadenilarea, în condițiile în care poli A polimeraza singură poate realiza poliadenilarea



unei catene de ARN clivată corect. Cu toate că mecanismul nu este pe deplin elucidat se poate considera că snRNP U11 împreună cu proteina 64 kD și probabil cu alte snRNP pot forma un complex în imediata vecinătate a secvenței AAUAAA care asigură clivarea ARN și permit o poliadenilare corespunzătoare. Implicarea snRNP se bazează pe descoperirea unor fragmente care conțin secvența AAUAAA în imunoprecipitatele obținute cu anticorpi anti-snRNP și pe inhibarea poliadenilării *in vitro*, realizată de acești anticorpi. Rolul snRNP în poliadenilare este complementar în maturarea ARN: U7 snRNP este implicată în crearea extremităților 3' ale ARNm de histone, U3 snRNP în producerea extremității corecte a ARNr 28S și U1, U2, U5 ȘI U4/U6 în maturarea pre-ARNm.

Cum genelor de la drojdii și probabil de la celelalte eucariote inferioare le lipsesc semnalul canonic de poliadenilare, AAUAAA, formarea cozii poli A poate fi cuplată cu terminarea transcrierii. Secvențe ca 5'-T5ATAT, care favorizează terminarea, pot fi utile în crearea extremității 3' necesare poliadenilării. Formarea extremității 3' poate fi și rezultatul clivării endonucleotidice a unui transcript mai lung, urmată de poliadenilarea realizată de poli A polimeraza. Nu s-a stabilit încă dacă la drojdii intervin particulele ribonucleoproteice de tip snRNP.

### **Procesarea post-transcripțională. Formarea “coifului” ARNm de la eucariote (SLIDE 16)**

În imagine este prezentată structura “cap” a ARNm la eucariote. Pot rezulta mai multe tipuri de structuri “cap” - **cap 0, cap 1 sau cap 2**, în funcție de pozițiile care sunt metilate.

Capătul 5' al ARNm nou format prezintă o grupare trifosfat. Inițial, un rest fosfat este îndepărtat prin hidroliză. Într-o etapă următoare capătul 5' al restului difosfat atacă grupare atomul de P din poziția  $\alpha$  al GTP, formându-se o legătură 5'-5' trifosfat, iar restul de G este metilat la N<sup>7</sup> de către **enzima guanin 7-metil transferaza**, rezultând o structura de tip cap 0. Dacă grupările 2'-OH ale primelor două nucleotide care se succed structurii cap sunt metilate enzimatic în prezența S-adenozil metioninei (SAM) și sub acțiunea enzimei 2'-O-metil transferaza, rezultă cap 1 și, respectiv cap 2.

### ***Informații suplimentare***

#### ***Coiful***

În afară de o excepție cunoscută (ARNm de la picornavirus) toate moleculele de ARNm de la eucariote, celulare și virale, posedă un coif la extremitatea 5'. Coifurile transcripturilor nucleare și citoplasmatici ai polimerazei II sunt asociate cu proteine care se leagă specific. Pentru moment

rolurile acestor proteine și ale coifului nu sunt deplin clarificate. S-a stabilit totuși că structura coif interacționează specific cu factorii de inițiere în asamblarea complexului de traducere.

Formarea coifului la extremitatea 5' a transcripturilor polimerazei II are loc imediat după inițiere, chiar și moleculele de ARN cele mai scurte posedând acest coif.

Unul dintre substratele enzimei care realizează adăugarea coifului, și anume guanin 7 metil transferaza este GTP; celălalt este extremitatea difosfat creată prin eliminarea fosfatului terminal al trifosfatului primului nucleotid al ARN în formare. În reacția de adăugare a coifului, fosfatul terminal al pre-ARNm este înlocuit cu gruparea guanil a GTP și este eliminat P<sub>Pi</sub>.

Restul de guanidină adăugat este metilat în poziția 7 a ciclului purinei de către guanin-7-metil transferaza; grupările -OH din poziția 2' ale primelor două nucleotide sunt apoi metilate de către 2'-O-metil transferaza. Gruparea 7-metil-GTP nu poate constitui substrat pentru reacția de adăugare a coifului.

Reacțiile de inițiere ale procesului de adăugare a coifului realizate de către polimeraza II sunt probabil cuplate obligatoriu deoarece transcriptii nucleari formați de către ARN polimeraza III, care posedă o extremitate trifosfat (ex. ARN U6, ARNr 5S și pre-ARNt), nu suferă procesul de adăugare a coifului.

Rolul coifurilor în timpul transcripției și maturării transcripturilor primari în ARNm nu este clar. De exemplu, rezultatele obținute privind implicarea coifului în maturare sunt contradictorii. Este posibil ca coiful să aibă un rol în formarea complexelor nucleare de ribonucleoproteine (nRNP), în care moleculele de pre-ARNm sunt sechestrate în timpul transportului ARNm în afara nucleului sau în asamblarea particulelor ribonucleoproteice citoplasmice (cRNP). Există indicații care sugerează că structurile de tip coif și proteinele asociate acestora asigură protecția moleculelor de ARNm de o degradare care ar putea să demareze la extremitatea 5'.

Mai bine stabilită este necesitatea coifului pentru atașarea ARNm la ribozomi înainte de demararea traducerii. Unul dintre factorii de inițiere, eIF-4 E, este un factor de recunoaștere specific coifului în timpul asocierii subunității 40S a ribozomului la extremitatea 5' a ARNm. Se consideră că eliminarea coifului unei molecule de ARNm împiedică traducerea sa *in vivo*. Prin experimente *in vitro* s-a dovedit că moleculele de ARNm care nu posedă coif pot fi totuși traduse. Singura excepție cunoscută de la această exigență este ARN genomic de la picornavirus (de ex. virusul poliomieltic) care funcționează ca ARNm pentru toate proteinele codificate de virus. În plus, în celulele infectate de către virusul poliomieltic se oprește sinteza

proteinelor gazdei deoarece o proteină virală inactivează proteinele celulare prin legare la nivelul coifului.

**Capătul 5' -cap este adăugat moleculei de ARNm în formare după inițierea transcripției de către ARN polimeraza II (SLIDE 17).**

În figura de pe slide sunt prezentate reacțiile prin care este realizată adăugarea “coifului” la transcripția ARNm în formare. Primele două reacții sunt catalizate de o enzimă responsabilă de adăugarea coifului care se asociază cu domeniul CTD al ARN polimerazei II la scurt timp după inițierea transcripției. Două metil-transferaze diferite catalizează reacțiile 3 și 4. S-adenozil metionina (SAM) este sursa de grupări metil (-CH<sub>3</sub>) pentru cele două etape de metilare ulterioare; restul guanilat (G) este metilat primul, apoi hidroxilul din poziția 2' al primului sau al primelor două nucleotide (N) din transcript.

**Moleculele de pre-ARNm sunt scindate la situsuri specifice 3' și rapid poliadenilate (SLIDE 21).**

A fost propus un model de clivare și poliadenilare a ARNm în celulele de mamifere conform căruia un factor specific de poliadenilare și clivare (CPSF) se leagă la o regiune ce conține semnalul de poliadenilare AAUAAA situată în amonte de situsul de poliadenilare. Un factor proteic CStF interacționează atât cu o secvență din aval bogată în GU sau U, cât și cu CPSF legat formând o buclă în ARN; legarea ulterioară a factorilor CFI și CFII ajută stabilizarea complexului. Legarea poli(A) polimerazei (PAP) stimulează clivarea la nivelul situsului poli(A), care este de obicei poziționat la 10-35 nucleotide în 3' față de semnalul de poliadenilare. Factorii de clivare sunt eliberați, ca și produsul de clivare din aval care este rapid degradat. Legarea PAP determină adăugarea a aproximativ 12 resturi A la o rată scăzută la gruparea hidroxil 3' generată prin reacția de clivare. Legarea proteinei de legare PABII (**P**oli**A** **B**inding protein) la coada poli(A) inițială scurtă accelerează rata de adăugare catalizată de PAP. După adăugarea a 200-250 resturi, PABII semnalizează oprirea polymerizării.

**Etapa finală de maturare: intronii sunt înlăturați, iar exonii sunt sudați împreună (SLIDE 22).**

*Informatii suplimentare*

Secvențele codante ale genelor eucariote (exonii) sunt frecvent întreruși de către regiuni necodante (intronii). Această descoperire curioasă a ridicat numeroase întrebări fundamentale: unde, când și cum au apărut intronii la nivelul genelor și care este rolul lor în evoluție?; cum

transcripții primari ai acestor gene mozaic sunt transformați în molecule ARN mature, lipsite de introni? care este influența procesului de excizie-sudare (episaj) asupra reglării expresiei genelor?; care este funcția, dacă există vreuna, a intronilor în multitudinea operațiilor care afectează genoamele contemporane; sunt intronii vestigii ale evoluției genelor și au vreun rol în viața organismelor contemporane?

Roberts și Sharp au obținut premiul Nobel pentru Fiziologie și Medicină în 1993, pentru descoperirea structurii discontinue a genelor.

### **Frecvența intronilor**

Frecvența genelor în mozaic variază foarte mult în funcție de speciile eucariote. Aceștia sunt foarte frecvenți la vegetale și la animale și la virusurile care le infectează. Intronii sunt mult mai rari în genele nevertebratelor (ex. *Drosophila*, nematodele și ariciul de mare). Totuși, mai multe gene care guvernează morfogeneza la *Drosophila* prezintă aranjamente complexe de introni și exoni. Cea mai mare parte a genelor *S. cerevisiae* sunt lipsite de introni (genele pentru actină și pentru alte câteva molecule de ARNt reprezentând excepții), dar genele în mozaic sunt mult mai numeroase la o specie înrudită numită *Schizosaccharomyces pombe*. Cu toate că intronii sunt rari în genele nucleare, ei sunt mult frecvenți în genele mitocondriale ale acestei drojdii.

La descoperirea lor s-a crezut că genele mozaic sunt o caracteristică proprie genelor eucariote. Evidențierea lor în gena timidilat sintetazei bacteriofagului T4 a demonstrat că această idee era falsă. De atunci, au mai fost descoperiți introni și în gena care codifică ARNt<sup>Ser</sup> și în gena pentru ARNr de la archeobacterii, etc. Acum nu mai este surprinzător că cercetări mult mai avansate evidențiază alți introni în genele procariotelor. Au fost descoperiți introni și la anumite bacterii.

Intronii există în genele nucleare care codifică pentru ARNm, ARNr (numai în genele lungi pentru ARNr de la eucariotele inferioare) și pentru un grup de gene care codifică ARNt. De asemenea, intronii sunt frecvenți în genele mitocondriale și din cloroplaste, care codifică pentru moleculele ARN mesager, ribozomic și de transfer din aceste organite. Totuși, în genele mamiferelor unde prezența intronilor este o regulă, există și excepții. Cea mai mare parte a genelor pentru histone nu conțin introni, ca și cele două familii de gene care codifică pentru interferon  $\alpha$  și  $\beta$ , în timp ce gena pentru interferon  $\gamma$  conține foarte mulți. Nu se cunosc încă exemple privind prezența intronilor în genele pentru ARNr 5S și 5,8S și nici pentru ARN U, 7SL și 7SK.

## **Diferite tipuri de introni**

Toți intronii sunt transcriși ca părți integrante ale moleculelor precursorare de ARN și sunt apoi eliminați printr-un proces de tăiere-legare numit excizie-sudare (episaj). Structura și mecanismele de alipire stau la baza clasificării diferitelor tipuri de introni.

### **“Splicing-ul” se produce la nivelul unor secvențe conservate, scurte (SLIDE 25)**

Au fost identificate **secvențele consens din jurul situsurilor de “splicing” 5’ și 3’** din structura pre-ARNm de la vertebrate. Cu toate că bazele 5’(GU) și 3’(AG) sunt aproape invariabile în structura intronului, totuși bazele care le flanchează pe acestea sunt prezente la frecvențe mai mari decât cele așteptate în cazul unei distribuții aleatoare. O regiune bogată în pirimidină situată în apropierea capătului 3’ al intronului este prezentă în majoritatea cazurilor. Punctul de ramificare adenozinic, de asemenea invariant este format de obicei de 20 la 50 baze în direcția 3’. Regiunea centrală a intronului, care poate fi cuprinsă între 40 pb și 50 kb lungime nu este necesară procesului de splicing.

### ***Informații suplimentare - Intronii genelor nucleare pentru ARNm***

Intronii genelor nucleare care codifică pentru proteine, primii care au fost descoperiți, au o mărime de la 10 pb la 10kpb. Intronii genelor omologe de la diferite specii de vertebrate pot să difere prin lungime și secvență asemeni intronilor din genele neînrudite. Caracteristicile cele mai comune ale intronilor sunt secvențele din 5’ (secvențe situate în amonte sau donatoare) și în 3’ (în aval sau acceptoare), adică joncțiunile exon-intron sau situsurile de excizie-sudare.

Secvențele nucleotidice ale fiecărei joncțiuni exon-intron sunt conservate remarcabil în aproape toate genele nucleare pentru ARNm la aproape toate speciile studiate. Secvența care flanchează cel mai frecvent situsul de sudare din 5’ este CRG (unde R reprezintă o purină), cea care formează capătul din 3’ fiind adesea reprezentată de un singur G. Cele mai mari variații se întâlnesc în secvențele care înconjoară intronii, însă mutații la nivelul acestor situsuri nu împiedică niciodată episajul, chiar dacă există cazuri în care poate fi afectată viteza. Structurile cele mai puțin variabile se găsesc în intron. Primele două nucleotide ale extremității 5’ ale intronului în ARN sunt aproape întotdeauna GU (GC în cazuri excepționale); următoarele cinci nucleotide nu sunt invariabile, cu toate că secvența AGAGU pare să fie consensus. Modificările nucleotidelor G sau U prezente la nivelul joncțiunii împiedică, în general, realizarea episajului,

în timp ce modificările pozițiilor adiacente afectează diferit episajul. Cele șase nucleotide de la extremitatea 5' a intronului sunt suficiente pentru situsul de episaj. Chiar și joncțiunile criptice, care nu sunt folosite decât rar sau când situsurile dominante sunt pierdute sau modificate, posedă secvența GU la extremitatea 5' a fragmentului care trebuie eliminat. Capătul 3' al intronului se termină invariabil prin AG, foarte adesea precedat, în intronii de mamifere, de o secvență bogată în pirimidină (YnNYAG). Și în acest caz mutațiile care înlocuiesc invariabilele A sau G printr-o altă bază împiedică episajul acestei joncțiuni.

Un rest A apropiat de extremitatea 3' a intronului este un participant esențial la episajul moleculelor pre-ARNm. În intronii mamiferelor, poziția acestui A nu este invariabilă deoarece poate fi utilizat orice A dintre nucleotidele situate în pozițiile 18 - 37 în amonte de situsul de episaj. Totuși, mutații la nivelul secvenței vecine acestui A utilizat provoacă o importantă diminuare a eficacității episajului *in vitro*; astfel, cu toate că nucleotidul A esențial nu apare într-un context invariant, secvența vecină influențează folosirea sa. Din contră, în cazul moleculelor de pre-ARN nucleare de la drojdie, restul A este furnizat de un heptanucleotid, 5'-UACUAAC-3', situat între nucleotidele 6 și 59 în amonte de situsul de episaj.

Prezența secvențelor consensus ale situsurilor de excizie-episaj nu semnifică întotdeauna că intronul poate fi excizat. În anumite cazuri, una sau două secvențe ale acestui situs sunt situate în exoni și în introni în locuri unde nu se produce în mod normal procesul de excizie-episaj. Totuși, astfel de situsuri criptice pot funcționa în anumite circumstanțe (ex., când situsurile autentice sunt modificate sau absente).

Ocazional, intronii posedă mai mult de o joncțiune 5' sau 3', permițând **proces de excizie-episaj alternativ**. De exemplu, într-o regiune precoce a SV40 apar doi introni care au amândoi același situs 3', dar joncțiunile 5' sunt diferite. Rezultatul acestor procese de **episaj alternativ** duc la formarea a două molecule de ARNm plecând de la același transcript primar. În anumite circumstanțe, alegerea episajului este reglată în timp în funcție de țesut. Adesea, situsuri de episaj există, dar nu sunt folosite. Astfel, genoamele retrovirale provin din transcripti de ADN proviral care nu au suferit excizie-episaj, dar producerea de ARNm care codifică pentru anumite proteine virale necesită excizie. În acest caz, același transcript poate suferi o modificare prin excizie-episaj sau nu.

Cum s-a menționat anterior, intronii pot conține alte elemente genetice, cum sunt elementele activatoare (enhancers) ale altor gene, semnale de replicare și de împachetare a cromozomului sau secvențe necesare asamblării pre-ARNm în particulele ribonucleoproteice (RNP).

### **Etapele procesului de producere a ARNm matur (SLIDE 26)**

În figură este prezentată *secvența etapelor de producere a ARNm eucariot pentru gena ovalbuminei de la pui*.

În urma transcripției, transcriptului primar i se adaugă “capul” și “coada” poliA. Intronii sunt apoi excizați și exonii sudați împreună pentru a forma structura ARNm matur.

#### *Informatii suplimentare*

#### *Splicingul intronilor moleculelor pre-ARNm nucleare*

Studiul mecanismelor de splicing în cazul pre-ARNm nuclear a fost facilitat de existența genelor mozaic clonate și, în particular, a formelor modificate natural sau deliberat ale acestor gene. Substrate special preparate, prin fuziune controlată de exoni și introni, au ajutat mult la identificarea structurilor necesare pentru un splicing fidel.

#### *Caracteristici generale*

**Splicingul** moleculelor de pre-ARNm nuclear **are loc în nucleu**, în același timp cu transcripția pentru anumite gene sau după transcripție pentru altele. Există indicații care sugerează că **adăugarea coifului la extremitatea 5' a transcriptului are implicații asupra procesului de maturare**. Din punct de vedere conceptual și practic este foarte important să se cunoască cum o moleculă pre-ARNm care conține mai mulți introni poate fi maturată, astfel încât exonii vecini să poată fi asociați corect. Deoarece se știe că situsul 5' al unui intron poate fi clivat în același timp cu situsul 3' al altuia, este foarte important să se înțeleagă cum, în cursul unei reacții normale, acest lucru poate fi evitat pentru a se putea realiza un **splicing alternativ**.

#### **Mecanismul de “splicing” presupune două reacții de transesterificare (SLIDE 28)**

În prima reacție, legătura ester dintre gruparea 5'-fosfat a intronului și oxigenul din 3' al exonului 1 este schimbată pentru o legătură ester cu oxigenul 2' al restului A de ramificare. În cea de a doua reacție, legătura ester dintre fosforul din 5' al exonului 2 și oxigenul 3' al intronului este schimbată pentru o legătură cu oxigenul 3' al exonului 1, eliberând intronul sub forma unei structuri “în laso” și resudând cei doi exoni. Săgețile arată unde atomii de oxigen din grupările hidroxil activare reacționează cu atomii de fosfor.

Secvența reacțiilor de transesterificare implicate în unirea exonilor (SLIDE 29) este prezentată în figura de slide în cazul unei molecule de pre-ARNm:

- 1) Gruparea 2'-OH a unui intron specific A atacă nucleofil gruparea 5'-fosfat de la capătul 5' al capătului intronilor pentru a ajunge la o structură sub formă de laso;
- 2) Gruparea 3'-OH eliberată formează o legătură 3'-5'-fosfodiester cu restul terminal 5' a exonului 3', determinând astfel unirea celor doi exoni și eliberarea intronului sub formă de laso.

### **Moleculele de “small nuclear RNAs (snRNAs) asistă reacția de “splicing” (SLIDE 30)**

În figura este prezentată *Diagrama interacțiilor dintre pre-ARNm, U1ARNsn și U2ARNsn în procesul de splicing timpuriu.*

- Regiunea 5' a U1ARNsn prezintă o secvență complementară inițial cu cea a capătului 5' al intronului și capătul 3' al exonului din 5' al pre-ARNm (situsul de scindare).
- U2ARNsn formează perechi de baze cu o secvență care include punctul de ramificare A, deși acest rest nu este împerecheat.

**Spliceozomul este un complex ribonucleoproteic compus din multe molecule de snRNPs (SLIDE 31).**

### ***Informatii suplimentare***

#### **Asamblarea spliceozomilor**

Deoarece splicingul are loc la nivelul spliceozomilor, apare necesitatea cunoașterii structurii, asamblării și mecanismelor de acțiune ale acestora. Particulele de splicing izolate dintr-un amestec de reacție în care clivarea în 3' a fost inhibată, sunt de formă elipsoidală și cu dimensiuni de 25 x 50 nm. În particular, intronul este asociat fiecăruia dintre snRNP U1, U2, U4/U6 și U5. Spliceozomii funcționali conțin și alte proteine și în particular pe cele care sunt implicate în legarea snRNP la țințele lor care sunt normal fixate pe structura pre-ARNm nuclear.

O secvență minimală a exonului este necesară legării snRNP la intron. După schema actuală de asamblare a spliceozomilor, se leagă primele particulele snRNP U1 la extremitatea 5', chiar dacă există sau nu situsuri funcționale pentru fixarea într-o etapă ulterioară a snRNP U2 și U5; dar această singură legătură este insuficientă pentru a provoca clivarea situsului din 5'. Legarea snRNP U1 este apoi urmată de asocierea snRNP U2. Legarea particulelor U5 și U4/U6, necesită ATP.



Legarea snRNP U2 la situsul său de pe intron se bazează pe împerecherea de baze, în timp ce încorporarea complexului snRNP U4/U6 și U5 în spliceozom depinde de interacțiile dintre snRNP.

O schemă asemănătoare este valabilă pentru asamblarea particulelor de splicing pentru pre-ARNm de la drojdii. După sudarea exonilor evoluția spliceozomului este incertă, dar snRNP sau sub-ansamble ale acestora sunt probabil reciclate atunci când ARN intronic este degradat.

### **Ciclul de acțiune la nivelul spliceozomului (SLIDE 32)**

Particulele de tip RNPs<sub>sn</sub> implicate în “splicing” (U1, U2, U4, U5 și U6) sunt asociate cu pre-ARNm și între ele într-o anumită ordine secvențială în scopul formării “spliceozomului”. Acest complex “ribonucleoproteic” catalizează apoi cele două reacții de transesterificare care au ca rezultat “splicingul” (sudarea) exonilor și eliberarea intronului sub forma unei structuri “laso”. Cu toate că pentru reacțiile de transesterificare nu este necesară hidroliza ATP, se consideră că acesta este necesar pentru a furniza energia necesară rearanjamentele structurii care apar în timpul ciclului. De reținut că proteinele snRNP din structura “spliceozomului” sunt distincte față de RNPhn discutate anterior. La eucariotele superioare asocierea U2ARN<sub>sn</sub> cu pre-ARNm este asistată de o proteină RNPhn numită U2AF, care se leagă o regiune bogată în pirimidină situată lângă situsul 3’ de splicing. De asemenea U2AF acționează probabil și cu alte proteine necesare pentru splicing prin intermediul unui domeniu care conține repetiții dipeptidice serină-arginină (motivul SR).

### **Auto-”splicing-ul” intronilor de grup II furnizează indicia pentru evoluția snRNPs (SLIDE 33)**

În figura sunt prezentate diagrame schematice care compară structurile secundare de auto-splicing ale intronilor de grup II (a) și UARN<sub>sn</sub> prezent în spliceozomi (b).

Similitudinile acestor structuri sugerează că structurile ARN<sub>sn</sub> evoluează din introni de grup II, cu aceste ARN<sub>sn</sub> care acționează în trans și este funcțional analog cu domeniile corespunzătoare intronilor de grup II.

#### *Informații suplimentare*

#### *Splicing autocatalitic sau catalizat de spliceozomi*

Rezultatul final al splicingului intronilor din grupa II și a pre-ARNm este același: intronul eliberat sub formă de laso și cei doi exoni vecini uniți între ei. În plus, mecanismul reacțiilor celor două procese este asemănător. În cele două cazuri, gruparea –OH din poziția 2’ a

intronului servește drept centru nucleofil pentru realizarea clivării situsului 5' gruparea –OH din poziția 5' a exonului situat în amonte este centrul nucleofil care clivează situsul 3' pentru a forma legătura exon-exon. Diferența cheie este că anumiți introni din grupa II sunt automatizați *in vitro*, în timp ce splicingul pre-ARNm nucleari necesită o mașinărie elaborată formată din mai multe snRNP și protein accesorii. Totuși, anumiți introni din grupa II necesită maturaze pentru reacția *in vitro* cu toate că nucleotidele și centrele catalitice implicate în cele două procese sunt aceleași. Presupunând că splicingul autocatalitic necesită o repliere foarte specifică a ARN, maturazele ar fi necesare pentru stabilirea și stabilizarea unei structuri care permite transesterificări secvențiale adecvate. Astfel, singura diferență între cele două mecanisme ar fi maniera în care intronul achiziționează conformația sa cataliticactivă.

Aceleași condiții sunt valabile pentru splicingul catalizat de către particule snRNP. Este posibil ca snRNP și factorii asociați să creeze «un eșafodaj» pe care intronul să poată fi corect pliat, astfel încât cele două joncțiuni și punctul de ramificare să fie juxtapuse și active. În acest sens, implicarea spliceozomilor în splicingul moleculelor de pre-ARNm nuclear este analogă celei a maturazei pentru intronii din grupa II. Ne putem întreba de ce replierea pre-ARNm nucleari necesită un astfel de complex în timp ce numai maturaza este suficientă pentru anumiți introni din grupa II. Răspunsul se poate găsi în complexitatea maturării care trebuie să se realizeze. Intronii mitocondriali din grupa II se aseamănă, fiecare posedând mai multe secvențe foarte conservate. Una sau două proteine sunt probabil suficiente pentru a realiza și a menține conformația activă a intronilor. Problema este evident mult mai complexă pentru moleculele de pre-ARNm nuclear, dată fiind diversitatea de mărime și de secvență a acestora și, deci, și numărul de introni. O singură proteină nu ar putea determina adoptarea unei conformații active pentru toți intronii. Particulele snRNP, foarte conservate pot interacționa specific cu două sau trei secvențe conservate prezente în toți intronii, pot interacționa între ele și pot impune o conformație care favorizează cele două transesterificări catalizate de către ARN.

#### *Alt tip de auto-splicing*

Intronii din grupa II din structura moleculelor de pre-ARNm mitocondriale de la drojzii posedă și ei secvențe conservate și structuri secundare definite, dar acestea sunt distincte de cele caracteristice intronilor grupei I. Și intronii din grupa II sunt supuși procesului de auto-splicing, dar există puține informații cu privire la structurile secundare și terțiare ale acestor substraturi, și

privind detaliile moleculare ale reacției de maturare. Sunt clare două puncte ale acestui mecanism: 1) contrar reacțiilor de automaturare pe care le suportă intronii din grupa I, maturarea intronilor din grupa II nu necesită un nucleotid inițiator și 2) produsul de reacție are o structură de ‘laso’.

Ca și în cazul intronilor din grupa I, procesul de auto-splicing implică două etape de transesterificare, una în urma clivării la nivelul situsului 5’, și cealaltă pentru clivajul situsului 3’ și legarea celor doi exoni. Totuși, diferența fundamentală constă în natura atacului nucleofil: guanozina pentru intronii din grupa I și gruparea –OH din poziția 2’ aparținând unui nucleotid al intronului în cazul intronilor din grupa II. Structura în laso rezultă din realizarea unei noi legături 2’-5’ fosfodiesterice în mijlocul secvenței ARN.

Structura tridimensională a intronului, spontană sau facilitată de proteine asociate, trebuie să aducă gruparea –OH din poziția 2’ a lasoului în apropierea situsului de splicing 5’ și să activeze transesterificarea. Alegerea legăturii internucleotidice care este clivată depinde probabil de secvențele, specifice grupei II, GUGCG prezentă la extremitatea 5’ și secvența YAU prezentă la extremitatea 3’ a intronului. Acest mecanism se aseamănă cu cel folosit la maturarea moleculelor pre-ARNm nuclear. Necesitatea particulelor ribonucleoproteice care acționează în trans pentru realizarea splicingului intronilor din pre-ARNm este datorată faptului că procesul, în acest caz, este mult mai complex și specific diferitelor tipuri de pliere a diferitelor specii de pre-ARNm, necesare asigurării splicingului corect la nivelul joncțiunilor exon-intron.

#### **Auto-splicing-ul intronilor de grup 1 - ARN catalitic (SLIDE 35)**

În intronii din grupul I, un cofactor guanozinic (G) care nu face parte din catena ARN se asociază cu situsul activ. Gruparea –OH din poziția 3’ al acestei guanozine participă la o reacție de transesterificare cu gruparea fosfat de la capătul 5’ al intronului. Această reacție este analogă cu cea care implică gruparea –OH din poziția 2’ a situsului de ramificare A din intronii de grup I și din intronii pre-ARNm matisați în spliceozomi.

Transesterificarea următoare care leagă capetele 5’ și 3’ ale celor doi exoni este similară în toate cele trei mecanisme de splicing. De reținut că intronul de grup I se eliberează sub forma unei structuri lineare spre deosebire de producția ramificați din celelalte două cazuri.

#### **Autosplicingul intronilor din grupa I.**

Splicingul intronului pre-ARNr de la *Tetrahymena*, care reprezintă prototipul grupei I, se realizează printr-o serie de transesterificări concertate în cursul cărora fosfoesterii sunt

schimbați fără hidroliză intermediară. Cu excepția etapei de inițiere, favorizată de prezența guanozinei libere sau a nucleotidelor sale, toate grupele reactive implicate în transesterificare sunt conținute în secvența intronului. În plus, specificitatea de schimb a legăturilor este o consecință a organizării tridimensionale a intronului, rezultată din împerecherea între secvențe distante dar conservate ale intronului.

***Prima etapă a splicingului este legarea guanozinei la o secvență a intronului.***

Perechea de electroni liberi ai grupării –OH din poziția 3' ai guanozinei poate provoca un atac nucleofil asupra fosfatului joncțiunii exon-intron 5' (-UpA-) provocând clivarea acestui situs. Radicalul –OH din poziția 3' creat la situsul de clivare (extremitatea 5' a exonului) reacționează cu gruparea fosfat din poziția 3' al joncțiunii, provocând legarea celor doi exoni și eliberarea intronului linear, de 413 nucleotide. Segmentul de intron linear suportă apoi alte două transesterificări intramoleculare și hidrolize, cu eliberarea mai întâi a 15 nucleotide și apoi formarea intronului final prin eliberarea altor 4 nucleotide. Trebuie să reținem că reacției globale nu îi este asociată nici o modificare netă de energie: fiecare rupere a unei legături fosfodiester fiind compensată prin formarea concomitentă a unei alte legături fosfodiester.

Guanozina sau unul dintre derivații ei 5' fosforilați este specifică inițierii reacției de splicing. Modificarea grupărilor –OH din poziția 2' sau 3' sau a potențialului de împerechere a guanozinei cu un rest de citozină modifică capacitatea de inițiere a splicingului. Evident, împerecherea guanozinei cu un rest complementar al intronului este importantă pentru poziționarea corectă a radicalului –OH din poziția 3' al guanozinei și pentru reacția sa cu gruparea fosfat din poziția 5' a joncțiunii. Cele două joncțiuni exon-intron sunt probabil apropiate una de alta pentru a permite legarea celor doi exoni după clivarea inițială realizată de guanină. Acest proces este ușurat de replierea intronului astfel încât cele două joncțiuni să se poată suprapune, probabil, în vecinătatea situsului de legătură al guanozinei. O posibilitate este ca, în intron, o anumită secvență să fie capabilă să se împerecheze cu o secvență a exonului situată la fiecare joncțiune. Prezența unei astfel de secvențe «ghid intern» ar permite reacția grupării hidroxil 3', creată prin clivarea joncțiunii exon-intron 5' de către guanozină, cu fosfatul 3' al situsului de splicing pentru a forma produsul matisat. Restul G care termină intronul și nucleotidele adiacente acestuia determină situsul de splicing 3'. Modificările conformaționale în intronul linear eliberat se presupune că facilitează circularizarea și eliberarea de mici oligonucleotide.

Mecanismul de autosplicing evidențiat pentru gena ARNr de la *Tetrahymena* este aplicabil și altor introni din grupa I. Aceștia posedă cu toții 4 secvențe conservate (A, B, 9L și 2) și alte două secvențe neconservate 9R și 9R'. Aceste 6 elemente apar întotdeauna în aceeași ordine: 5'-9R'-A-B-9L-9R-2-3'. În plus, cea mai mare parte a intronilor din grupa I posedă și o secvență care poate servi drept ghid intern. Mutații care împiedică splicingul au fost evidențiate și caracterizate la drojdii și ciuperci. Astfel de mutații modifică în general posibilitatea împerecherii bazelor. Astfel, de exemplu, inactivarea splicingului determinată de modificarea secvențelor 9R sau 9R' este reversibilă prin schimbări compensatorii în altă secvență.

Splicingul *in vitro* al anumitor introni din grupa I aparținând genelor mitocondriale de drojdie pentru pre-ARNm depinde de aceleași secvențe conservate similar celor care sunt necesare splicingului pre-ARNr de la *Tetrahymena*, produșii de reacție fiind analogi. Totuși, aceste molecule de pre-ARNm nu suferă procesul de auto-splicing *in vitro*. Splicingul acestor introni depinde de proteine, care acționează în trans, codificate de aceleași gene sau de gene înrudite, de către secvențe de lectură deschise constituite din exoni și introni. Aceste proteine, maturaze, nu sunt prezente decât pentru o perioadă scurtă de timp și par să favorizeze replierea intronului din structura pre-ARNm într-o conformație care permite splicingul. Aceasta explică de ce mutațiile non-sens sau alunecarea fazei de lectură în secvența exonului care codifică pentru maturază, împiedică splicingul intronului respective, ca și pe cel al intronilor din grupa I caracteristici altor gene mitocondriale. Acest model explică și de ce mutațiile supresoare sau care conțin o alunecare de fază, care readuc la normal producerea maturazei restabilesc și splicingul corect. Plierea corectă a anumitor introni din grupa I este spontană și stabilă *in vitro* în timp alții necesită proteine pentru formarea și/sau stabilizarea structurii active autocatalitic. Există și posibilitatea ca diverse interacții proteine-proteine să aibă rol în pliere.

### **Splicingul în trans**

Reacțiile de splicing trecute în revistă până în prezent sunt intramoleculare și sunt deci reacții în *cis*. Se pune problema existenței unor procese de splicing intermolecular, deci despre care putem considera că se desfășoară în trans. Mai specific, se pune problema dacă doi exoni situați pe molecule de ARN separate se pot asocia cu eliminarea concomitentă a intronilor pe care îi conțin. Existența acestui proces de splicing intermolecular a fost demonstrat *in vitro* utilizând substrate ARN special preparate. S-a stabilit că splicingul *in trans* este o etapă esențială în

producerea celulară a tuturor moleculelor de ARNm la *Trypanosoma* și a trei dintre cele patru tipuri de ARNm rezultate în urma transcrierii genelor actinei de la *C. elegans*.

Primul exemplu de splicing *in trans* în care erau implicate unități transcripționale situate pe cromozomi diferiți a fost descris în cazul proto-oncogenei *c-myb* (Vellard et al, *Oncogene*, 1991). În fiecare caz, situsurile de clivare sunt de tip clasic, GU la joncțiunea 5' și AG în poziția 3', produsul eliminat fiind și el o structură ramificată (laso).

Formarea ARNm al actinei funcționale și a multor alte molecule de ARNm de la *C. elegans* se realizează prin splicing *in trans* (trans-splicing). În cazul actinei, coiful și primele 22 nucleotide ale ARNm provin dintr-un transcript de 100 nucleotide obținut plecând de la ADNr 5S situat pe un cromozom diferit.

Exemplul cel mai bine studiat de splicing *in trans* este cel de formare a ARNm de la *Trypanosoma*. Toate moleculele de ARNm de la *Trypanosoma* conțin în 5' o secvență identică, de 35 nucleotide, netradusă, care precedă o secvență codantă întreruptă. Acești mini-exoni 5' provin de la extremitatea 5' a unui ARN de 137 nucleotide transcris plecând de la sute de copii în tandem ale unui segment de ADN de 137 pb. Un splicing *in trans* asociază mini-exonul 5' de 35 nucleotide cu un ARN al cărui exon codifică pentru o proteină cu un alt ARN formând o moleculă de ARN ramificată. În etapa următoare, situsul de maturare 3' este clivat și cei doi exoni sunt reuniți. Segmentul ramificat, conținând secvențele «intronilor» de la două molecule de ARN separate, este apoi scindat de o enzimă care realizează deramificarea, rezultatul fiind separarea intronilor care erau asociați celor două molecule de ARN.

În exemplul analizat mai sus trebuie remarcat: 1) secvența situată imediat în aval de mini-exonul de la *Trypanosoma* este conform cu secvența consens 5(GUAUGA);

2) joncțiunea intronului asociat cu exonul codant este analogă secvenței 3' a situsurilor de excizie ([C/Un NNAG) 3) un nucleotid al intronului reprezintă punctul de ramificare, sugerând că mecanismele de splicing *in trans* și *in cis* sunt analoge.

### **Concluzii (SLIDE 41)**

- Precursorii ARNm de la eucariote sunt procesați prin adăugarea structurii “cap”, scindare în 3' și poliadenilare, înlăturarea intronilor și sudarea exonilor (splicing) înainte de transportul ARNm în citoplasmă unde urmează a fi tradus de către ribozomi.

- Structura “cap” este adăugată la capătul 5’ al pre-ARNm, în formare, de către o enzimă specifică care este asociată cu domeniul CTD fosforilat al ARN polimerazei II, la puțin timp după inițierea transcripției.
- Transcripții pre-ARNm, în formare, sunt asociați cu o clasă de proteine de legare a ARN numite hnRNP.
- La majoritatea genelor care codifică pentru proteine, un semnal de poliadenilare conservat (AAUAAA) este situat la 10-30 nucleotide în amonte de situsul poli(A) unde apare clivarea și poliadenilarea. Un complex multiproteic care include poli(A) polimeraza (PAP) realizează scindarea și poliadenilarea pre-ARNm. O proteină nucleară de legare la poli(A), PABII, stimulează adăugarea de resturi A de către PAP și operește procesul odată ce “coada” poli(A) atinge 200-250 resturi.
- Splicingul ARN este realizat de către un complex ribonucleo-proteic, numit “spliceozom”, care este asamblat prin interacțiunile a cinci particule RNPs diferite, între ele, și de asemenea cu pre-ARNm. Spliceozomul catalizează două reacții de transesterificare care sudează exonii și înlătură intronul sub forma unei structuri în “laso”, care este ulterior degradată.
- Intronii autocatalitici de grup II, care au fost evidențiați în genele cloroplastelor și mitocondriilor de la plante și fungi, prezintă o structură secundară înalt conservată, care este necesară pentru auto-splicing. Se consideră că particulele RNPs din spliceozomi că au o structură secundară similară cu cea a intronilor de grup II.
- Majoritatea transcripției și procesarea ARN din nucleii celulelor eucariote se realizează la nivelul unui număr limitat de domenii. O rețea proteică fibroasă din interiorul nucleului formează o matrice nucleară. Aceasta poate servi la organizarea unor centre de transcripție și procesare a ARN.

### **Procesarea ARNr și ARNt**

**Genele pre-ARNr sunt similare la toate eucariotele (SLIDE 66).**

În figură este prezentată *Structura generală a unităților transcripționale pre-ARNr*. Cele trei regiuni codante din structura acestor unități codifică pentru ARNr 18S, 5,8S și 28 S prezente în ribozomii eucariotelor superioare sau a echivalenților lor de la alte specii. Ordinea acestor trei regiuni codante în genom este întotdeauna în direcția 5’-3’. Variațiile de lungime ale regiunilor

*spacer* transcrise (tan) sunt responsabile de diferența majoră de lungime a unităților de transcripție de la diferite organisme.

**Moleculele de snoRNA (small nucleolar RNAs) asistă procesarea ARNr și asamblarea subunităților ribozomale (SLIDE 68).**

În figura este prezentată *procesarea pre-ARNr și asamblarea ribozomilor la eucariote*, fiind evidențiate intermediarii majori și timpul necesar pentru diferitelor etape de procesare a pre-ARNr la eucariotele superioare.

a) Proteinele ribozomale și nucleolare asociate cu pre-ARNr 45S imediat după sinteza sa, formează o moleculă pre-RNP de 80S. Sinteza ARNr 5S apare în afara nucleolului. De reținut că ARNr constituie aproape 2/3 din masa subunităților ribozomale și proteinele asociate numai 1/3.

b) Calea de procesare a transcriptului primar pre-ARNr de 6,6 kb (35S) la *Saccharomyces cerevisiae*. Regiunile *spacer* transcrise (tan) care sunt înlăturate în timpul procesării, separă regiunile corespunzătoare formelor mature de ARNr 18S, 5,8S și 25S.

**Moleculele de pre-ARNt suferă scindare și modificarea bazelor (SLIDE 69).**

Procesarea pre-ARNt pentru tirozină implică patru tipuri de modificări. Un intron de 14 nucleotide din bucla anticodon este îndepărtat prin *splicing*. O secvență de 16 nucleotide de la capătul 5' este clivată de către RNaza P. Resturile U de la capătul 3' sunt înlocuite de către secvența CCA care se găsește în toate moleculele de ARNt mature. Numeroase baze din bucla stem sunt transformate în baze modificate caracteristice (D=dihidouridină; Ψ=pseudouridină).

Nu toate moleculele de ARNt conțin introni care sunt matisați în timpul procesării, dar suferă celelalte tipuri de modificări.

**“Splicing-ul” moleculelor pre-ARNt diferă de celelalte mecanisme de “splicing” (SLIDE70).**

Molecula de pre-ARNt este clivată în două locuri, de o parte și de alta a intronului, excizând astfel intronul. Mecanismul de clivare generează un fosfomonoester 2'-3' ciclic la capătul 3' al exonului din 5'. Reacția de joncțiune dintre doi exoni este o reacție în multe etape care necesită participarea a doi nucleozid trifosfați: GTP, care contribuie cu gruparea fosfat pentru joncțiunea 3'-5' din molecula finală ARNt și o moleculă ATP care formează un intermediar ligază-AMP activat. Gruparea fosfat din poziția 2' a exonului 5' este înlăturată în etapa finală.â

***Informatii suplimentare***

***Splicingul ARNt***



Modul în care sunt eliminați intronii moleculelor de ARNt a fost cel mai bine studiat și înțeles la drojdii, dar anumite informații au fost oferite în urma studiului altor eucariote inferioare și plante. Toate enzimele implicate sunt cunoscute și purificate și toți intermediarii sunt caracterizați. O endonuclează specifică pentru ARNt taie pre-ARNt la nivelul joncțiunii 5' a intronului, formând un 2', 3'-fosfodiester ciclic la extremitatea regiunii 5' a ARNt. Aceeași enzimă taie joncțiunea celui alt intron producând o grupare 5' -OH la extremitatea părții 3' a ARNt. Fosfodiesterul ciclic este apoi deschis de către o fosfodiesterază ciclică pentru a forma un 2' fosfomonoester. După fosforilarea extremității sale 5' -OH, exonul 3' este adenilat de către o ligază specifică pentru ARNt; această ultimă reacție este identică cu cea catalizată de către ADN ligaze. Legarea celor două jumătăți de molecule prin intermediul extremităților lor activate creează o legătură neobișnuită 2' fosfat, 3', 5' fosfodiester. Eliminarea fosfatului din 2', de către o fosfatază duce la formarea ARN matur. De reținut că gruparea fosfat a noii legături diester provine din ATP. Aceasta este o caracteristică care distinge maturarea ARNt de la drojdii de cea de la vertebrate.

În această serie de reacții, clivarea endonucleazică inițială determină formarea de extremități 5' -OH și 3' fosfomonoester, acesta din urmă fiind convertit în 2', 3' fosfodiester de către o ciclază ATP dependentă. O ligază specifică pentru ARNt unește apoi cele două jumătăți fără a necesita activarea extremităților.

Fracționarea sistemului de splicing de la drojdii a dus la obținerea unor preparate înalt purificate a două enzime. Una catalizează clivarea endonucleotidică la nivelul joncțiunilor intronului, iar cealaltă posedă, la nivelul unei singure polipeptide, activitatea fosfodiesterazică, kinazică, de adenilare și ligazică. Cele două enzime sunt foarte specifice pentru reacțiile de splicing ale ARNt; ele acționează totuși fără a face distincție asupra tuturor intronilor ARNt și unesc oricare două fragmente de ARNt. Astfel au putut fi construite molecule de ARN hibride foarte interesante, formate din două jumătăți de molecule de la ARNt diferiți.

### **Concluzii (SLIDE 72)**

- Asemeni moleculelor pre-ARNm, transcripții primari produși din genele pentru ARNr și ARNt suferă procesări extensive.
- Sinteza unui precursorului pre-ARNr (45S în eucariotele superioare) de către ARN polimeraza I și procesarea acestuia apare în nucleol; clivarea, digestia exonucleotidică, și modificarea bazelor pre-ARNr 45S conduce la molecule mature 28S, 18S și 5,8S ARNr,

care se asociază cu proteinele ribozomale în subunitățile ribozomale. O serie de snoARN (small nucleolar RNA) participă la procesarea ARNr.

- ARNr 5S, care este sintetizat în nucleoplasmă de către ARN polimeraza III, nu este produs extensiv înaintea asamblării cu alte molecule de ARNr și proteine pentru a forma subunitatea ribozomală;
- Primele molecule de ARNr catalitic (ribozime) descoperite au fost intronii de grup I din ARNr de la *Tetrahymena*. Auto-splicingul intronilor de grup I și II, și splicingul pre-ARNm la nivelul spliceozomilor se realizează prin intermediul a două reacții de transesterificare;
- Transcripția genelor ARNt de către ARN polimeraza III și procesarea transcriptelor primari apare în nucleoplasmă. În toate moleculele de pre-ARNt, secvența capătului 5' este îndepărtată de RN-aza P, o ribonucleoproteină care conține un ARN catalitic; CCA este adăugat la capătul 3'; o serie de baze interne sunt modificate;
- Unele molecule de pre-ARNt conțin un intron scurt în interiorul buclei anticodon. Acesta este înlăturat de enzime printr-un mecanism distinct de mecanismul de splicing al pre-ARNm și auto-splicingul intronilor.

## TRADUCEREA (NOTE DE CURS BM5)

### - Caracteristici și etape

### - Codul genetic

Producersii transcripției (ARNt, ARNr și ARNm) participă toți la transferul informației genetice în proteine, dar numai ARNm este depozitarul informației. Termenul de **traducere** desemnează ansamblul mecanismelor care transformă informația secvenței ARNm în proteine. Transferul informației de la secvența nucleotidică la secvența proteică presupune existența unui sistem de codificare între cele două tipuri de secvențe, numit **cod genetic**. **Ribozomii** și **ARNt** sunt celelalte două componente implicate puternic în mecanismul de traducere. Pe lângă acestea, intervin mai multe zeci de alte proteine și molecule diferite (aminoacil-ARNt transferazele, nucleozid trifosfații (ATP și GTP), proteinele de inițiere (IF), de alungire (EF) și de oprire (RF).

### Codul genetic

Codul genetic girează corespondența între secvența nucleotidică a unui ARNm și secvența aminoacizilor unei proteine. Secvențele nucleotidice sunt citite secvențial sub forma de triplete. Fiecare triplet este numit **codon** și nu există suprapunere în lectura codonilor succesivi. Un nucleotid care a servit pentru lectura unui codon nu poate servi pentru lectura altui codon. Astfel, primul codon al unei secvențe este foarte important deoarece el definește **cadrul de lectură** pentru restul secvenței. Deoarece nu există suprapunere, sunt posibile doar trei cadre de lectura. În funcție de cadrul de lectura secvența în aminoacizi va fi diferită. Descifrarea codului genetic a fost realizată în urma experimentelor *in vitro* cu polinucleotide de sinteză. Nirenberg și Matthaei (1961) au sintetizat poliribonucleotide, cum este de exemplu poli U și au furnizat condițiile necesare pentru traducerea lor. Polipeptida obținută este polifenilalanina. Aceeași experiență, repetată cu polimeri de A și C a determinat sinteza polipeptidelor poliLys și polyPro, AAA și CCC codifică respectiv pentru lizină și prolină. Ceilalți codoni au fost elucidați prin traducerea poliribonucleotidelor cu secvențe dinucleotidice sau trinucleotidice alternative. De exemplu, un poliAC poartă doi codoni posibili ACA și CAC indiferent de cadrul de lectură adoptat. Traducerea sa furnizează un amestec echimolecular de histidină și de treonină. Identificarea codonilor acestor aminoacizi a fost realizată printr-o altă experiență. Este vorba de sinteza unui polimer format din A și din C în rapoarte molare de 5:1. În acest polimer codonul AAA va fi cel

mai reprezentat statistic, iar codonul CCC cel mai puțin reprezentat. Amestecurile codonilor  $(A)_2C$  (AAC, ACA și CAA) vor fi mult mai frecvente decât  $A(C)_2$  (CCA, CAC și ACC). Analiza compoziției în aminoacizi a produșilor de traducere indică că numai histidina poate fi codificată de  $A(C)_2$ , rezultând astfel că CAC codifică histidina și ACA treonina.

În alte experiențe, au fost utilizați, într-un sistem de traducere, polimeri formați din trei sau patru nucleotide. Un polimer format din trei nucleotide ( $[UUG]_n$  de exemplu) furnizează trei tipuri de peptide: poliLeu, poliCys și poliVal, în funcție de cadrul de lectură utilizat. Un polimer format din patru ( $[UUAC]_n$  de exemplu) furnizează un tetrapeptid Leu-Leu-Thr-Tyr. În final, aceste experiențe și altele au permis stabilirea în întregime a codului genetic.

Inițierea traducerii utilizează adesea codonul AUG și uneori codonii GUG și UUG. Aminoacidul încorporat este în toate cazurile Met (sau formil-Met la bacterii). Aceeași codoni codifică pentru Met, Val și Leu când nu sunt folosiți ca semnal pentru începerea traducerii.

Încheierea (terminarea) traducerii se realizează când ribozomul întâlnește unul dintre codonii “non sense” (UAA, UAG, UGA) indicați în tabel prin “Stop” (Tabel 4.2, **SLIDE 13**). Aminoacizii sunt desemnați prin trei litere și respectiv o literă de cod. Știind că codonii sunt triplete și că există patru baze în structura ARN (G, A, U, C), sunt posibile teoretic 43 adică 64 de triplete diferite. 61 dintre acestea au sens, adică codifică pentru aminoacizi și trei sunt numiți “non sens” “(fără sens) (UAA, UGA, UAG) deoarece nu există ARNt care să posedă anticodoni complementari. Acești codoni „fără sens” sau “Stop”, sunt codonii care marchează terminarea sintezei proteice. Același aminoacid poate să fie codificat de mai mulți codoni diferiți (Tabel 4.2, **SLIDE 13**). Astfel, histidina este codificată de doi codoni, CAU și CAC, iar arginina de codonii: CGU, CGC, CGA, CGC, AGA și AGG. În medie, unui aminoacid îi corespund trei codoni. Din această cauză codul genetic este numit **degenerat** sau **redundant**. Codul genetic este în același timp numit **universal** deoarece aminoacizii au aceeași codoni indiferent de sistemul biologic care asigură traducerea. Totuși există câteva excepții ale codului universal, mai ales în cazul sintezei proteice mitocondriale. **Codul genetic** folosit în mitocondriile animale și ale fungilor este diferit de codul standard folosit de toate genele procariote și eucariote; de remarcat că acest cod diferă chiar și pentru mitocondrii de la specii diferite. Cum a apărut acest fenomen în cursul evoluției este misterios. De exemplu, UGA, este în mod normal un codon STOP, dar este citit ca triptofan în mitocondriile umane și de fungi; totuși în mitocondriile plantelor UGA este încă un codon

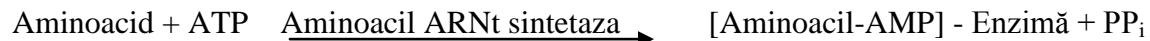
STOP. AGA și AGG, codonii nucleari standard pentru arginină codifică de asemenea pentru arginină în mitocondriile fungilor și plantelor, dar reprezintă codoni STOP pentru ADNmt de la mamifere și pentru serină la ADNmt de la *Drosophila*. Mitocondriile plantelor par să utilizeze codul genetic standard. Totuși, comparațiile secvențelor în aminoacizi ale proteinelor mitocondriale de la plante cu secvența nucleotidică a ADNmt sugerează că CGG poate codifica fie pentru arginină (aminoacidul standard) sau triptofan. Această nespecificitate aparentă a codului mitocondrial este explicată prin editarea transcripțiilor ARN mitocondriali, care pot transforma resturile de citozină în uracil. Dacă o secvență CGG este editată în UGG, codonul specifică triptofanul, aminoacidul standard pentru UGG, în timp ce codonii CGG care nu sunt editați codifică pentru arginină. Astfel sistemul de translație în mitocondriile plantelor utilizează codul genetic standard.

### **Moleculele de ARNt izoacceptoare**

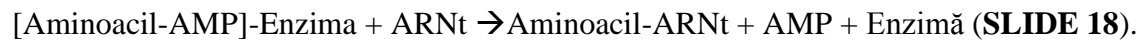
Aminoacizii nu au afinitate specifică pentru ARN<sub>m</sub>. Ei se atașează matricelor de ARN<sub>m</sub> prin intermediarul adaptatorilor, moleculele de ARN de transfer (ARNt). Moleculele de ARNt sunt molecule de mărime mică (masa lor este de aproximativ 25 000, 75 de baze). Ele prezintă particularitatea de a fi modificate după sinteză. Astfel, anumite baze sunt metilate, altele transformate în baze rare (cum sunt dihidrouridina, inozina), etc. Moleculele de ARNt adoptă o configurație spațială compactă în formă de L inversat. Din comoditate, structura este frecvent reprezentată depliată sub forma unei frunze de treflă cu patru tije formate din baze împerecheate prin legături de hidrogen și trei regiuni în formă de buclă (limburile frunzei) ne-împerecheate (**SLIDE 15**). Extremitatea 5' a ARNt este tot timpul ocupată de către G, în timp ce extremitatea 3'OH se termină prin secvența CCA a cărei funcțiune 3'OH a adenzinei este esterificată cu gruparea COOH a aminoacidului. Bucla anticodonului poartă o secvență de trei baze, încadrată în 3'de o purină și în 5' de un U. Aceasta recunoaște și hibridizează codonul complementar existent pe ARN<sub>m</sub>. Astfel, există o corespondență specifică între aminoacidul fixat la ARNt și anticodon. Anumite molecule de ARNt conțin nucleotidul **inozină** (nucleotid a cărui bază azotată este hipoxantina). Inozina poate forma legături de hidrogen cu U, C și A. În concluzie, același ARNt se poate împerechea cu trei codoni diferiți care corespund aceluiași aminoacid (**SLIDE 16, 17**).

Înainte de a participa la traducerea ARN<sub>m</sub>, moleculele de ARNt fixează aminoacidul corespunzător anticodonului lor (**SLIDE 18**). Fiecare aminoacid este mai întâi activat, în

prezență de ATP, de către enzima **aminoacil ARNt sintetaza** care catalizează într-o primă etapă formarea **aminoacil-AMP**:



Enzima rămâne legată de complexul aminoacil-AMP până la întâlnirea unui ARNt specific aminoacidului. În a doua etapă, aminoacidul este transferat pe extremitatea 3'OH a ARNt și formează un **aminoacil-ARNt** numit și **ARNt încărcat**:



Asamblarea proteinelor se realizează plecând de la acești aminoacil-ARNt. În funcție de aminoacidul transportat, aminoacil-ARNt va fi numit alanilARNtAla (sau Ala-ARNtAla), Glicil ARNtGly(sau Gly-ARNtGly), etc.

Evidențierea relațiilor codon: anticodon: aminoacid transportat de ARNt a fost realizată în urma experiențelor lui Nirenberg și Leder (1964). Ribozomii incubati cu ARNm poliU și Phe-ARNtPhe fixează cei doi acizi nucleici. Dacă Phe-ARNtPhe este înlocuită de către un alt aminoacil-ARNt, nu are loc fixare de către ribozomi. Deci, Phe-ARNtPhe recunoaște specific codonul UUU grație anticodonului său. Prin modificarea secvenței ARNm, a fost posibilă determinarea speciilor (moleculor) de ARNt care recunosc diferenții codoni. Deci, într-o celulă se găsesc douăzeci de aminoacizi diferiți care sunt transportați de aproximativ șaiszeci de molecule de ARNt la încărcarea cărora participă același număr de ARNt sintetaze corespunzătoare. În consecință, anumiți codoni pot fi recunoscuți de mai mulți ARNt care transportă același aminoacid, ARNt izoacceptori, în timp ce alții, codonii Stop, nu sunt recunoscuți de nici unul.

### **Ribozomii (SLIDE 19, 20)**

După încărcarea cu aminoacidul corespunzător, ARNt se combină cu ribozomii pentru a asigura sinteza proteică. Ribozomii, particulele citoplasmatiche constituite dintr-o subunitate mare și o subunitate mică, ambele formate din ARNr și proteine, au compoziții diferite la procariote și la eucariote. Ribozomii sunt constituenți majori în celulă. Colibacilul conține aproximativ 15000 ribozomi, reprezentând aproximativ 25% din masa uscată a celulei. Compararea secvențelor proteinelor ribozomale de la mai multe organisme relevă o foarte mare conservare în cursul evoluției. Moleculele de ARNr sunt și ele la fel de bine conservate.

Cele două subunități ribozomale sunt adesea libere și nu se asociază decât în momentul traducerii ARNm. Funcția principală a ribozomilor este de a orienta corect ansamblul aminoacil-ARNt în

raport cu ARNm cu scopul de a crea legăturile peptidice. Mai mulți ribozomi pot începe traducerea succesivă a aceleiași molecule de ARNm, ei găsindu-se astfel legați sub forma unui șirag de mărgele care constituie un **polizom** sau un **poliribozom**.

### **Sinteza proteică (SLIDE 21)**

Traducerea ARNm în proteine se realizează întotdeauna în sensul 5'-3'. La procariote, traducerea începe chiar înainte de terminarea transcripției. La eucariote, transcripția se realizează în nucleu, în timp ce traducerea are loc în citoplasmă. Deci, moleculele de ARNm sunt transcrise și apoi maturate în nucleu. Ele migrează apoi în citoplasmă pentru a fi traduse în proteine.

Traducerea poate fi împărțită în trei etape esențiale: **demararea** sau **inițierea**, **alungirea** sau **elongarea** și **oprirea** sau **terminarea**.

### **Demararea sau inițierea (SLIDE 21, 22)**

Inițierea traducerii corespunde legării ribozomului la ARNm, căutării codonului de inițiere (orientării după codonul de inițiere) și fixarea primului aminoacil-ARNt. Demararea sintezei proteice la bacterii începe deci cu formarea unui complex între subunitatea mică, 30S, primul aminoacil-ARNt și o moleculă de ARNm. Subunitatea mare, 50S, se leagă la acest complex pentru a forma ribozomul 70S funcțional. În cursul sintezei tuturor peptidelor bacteriene, primul aminoacid încorporat este N-formil-metionina (metionină modificată care posedă o grupare formil fixată de radicalul amino-terminal), al cărui codon principal este AUG, adesea GUG. Datorită faptului că gruparea amino a formil-Met (fMet) este blocată, un astfel de aminoacid nu poate fi inserat decât la începutul catenei polipeptidice.

**Factorii** denumiți **de inițiere (IF** pentru Initiation Factors) necesari sunt proteinele IF1, IF2 și IF3 (la eucariote au fost evidențiați cinci astfel de factori). Acești factori se fixează la subunitatea 30S în timpul primei etape de inițiere, iar GTP legat la IF2 este elementul stabilizator al acestei fixări. În acest stadiu, factorul IF3 împiedică asocierea subunităților 30S și 50S.

Cea de a doua etapă este asocierea fMet-ARNt<sub>fMet</sub> și a ARNm la ansamblul IF-30S-GTP. Asocierea presupune intervenția unei secvențe mici din structura ARNm, AGGAGGU, numită situs **de fixare la ribozom** sau **rbs** (ribosome binding site) sau **secvență Shine-Dalgarno** complementară unei secvențe prezentă în structura ARNr care participă astfel direct la inițiere, selecționând situsul de demarare a traducerii: codonul AUG inițiator este definit prin hibridizarea dintre secvența rbs și o secvență apropiată de extremitatea 3' a ARNr 16S.

Eliberarea IF3 permite asocierea celor două subunități, hidroliza GTP și eliberarea altor doi factori de inițiere. Complexul obținut este numit **complex de inițiere 70S**.

La eucariote, ribozomii recunosc coiful (capișonul) metilat grație unor factori numiți **proteine de fixare a coifului (cap binding protein)**. Aceștia alunecă apoi de-a lungul ARNm plecând de la extremitatea 5' până la întâlnirea primului codon AUG inițiator. Factorii adiționali de inițiere favorizează, în această etapă, desfacerea structurilor secundare susceptibile de oprirea procesului.

#### **Alungirea sau elongarea (SLIDE 24)**

Elongarea implică formarea unei legături peptidice între aminoacizi și deplasarea ribozomului de-a lungul ARNm. Lectura ARNm începe de la extremitatea 5' și se realizează în direcția 5'-3'. Fiecare ribozom posedă două situsuri de fixare a ARNt. Acestea sunt situsurile **P (Peptidil)** și **A (Aminoacil)**. Fixarea ribozomului la ARNm se realizează astfel încât fMet-ARN<sup>f</sup>Met să fie poziționat în situsul P pentru a se putea împerechea cu codonul AUG de pe ARNm. Situsul A disponibil poate astfel primi aminoacil-ARNt al cărui anticodon corespunde celui de al doilea codon din structura ARNm. O legătură peptidică se stabilește între gruparea COOH al primului aminoacid și gruparea NH<sub>2</sub> a celui de al doilea grație enzimei numită peptidil-transferază, componentă a subunității 50S. Formarea legăturii peptidice transferă gruparea COOH a aminoacidului din situsul P grupării amino a aminoacidului care se găsește în situsul A. Astfel, extremitatea COOH a catenei în formare se termină întotdeauna cu o moleculă de ARNt. Lanțul peptidic este sintetizat începând de la extremitatea NH<sub>2</sub> către extremitatea COOH. În urma formării legăturii peptidice ARN<sup>f</sup>Met descarcă aminoacidul și părăsește situsul P. La momentul respectiv, dipeptida se găsește legată la ARNt încărcat cu cel de-al doilea aminoacid și este poziționată în situsul A. În continuare, acest peptidil-ARNt se deplasează în situsul P prin mișcarea relativă a ARNm cu o lungime de trei baze. În acest moment situsul A liber poate primi un nou aminoacil-ARNt. O dată ajuns în situsul A, o altă legătură peptidică asociază aminoacidul cu cel care îl precede în situsul P. Elongarea catenei se realizează prin repetarea acestor mișcări. Alungirea catenei polipeptidice necesită două molecule de GTP care sunt hidrolizate pentru fiecare aminoacid adăugat.

**Factorii de elongare (EF** pentru Elongation Factors) sunt și aceștia indispensabili. EF-Tu reacționează cu GTP și ARNt încărcat pentru a forma complexul aminoacil-ARNt-GTP-EF-Tu care plasează aminoacil-ARNt în situsul A de pe ribozom folosind energia eliberată prin hidroliza GTP în GDP. Deplasarea peptidil-ARNt din situsul A în situsul P este dependentă de



factorul de elongare G (EF-G) numit și translocază. EF-G reacționează prin asocierea GTP la ribozom. Acesta permite translocarea și eliberarea ARNt din situsul P. Reutilizarea acestui factor de elongare necesită de asemenea hidroliza unei molecule de GTP în GDP și P<sub>i</sub>.

Mecanismele de alungire (elongare) a catenei polipeptidice la eucariote s-au dovedit similare cu participarea factorilor de elongare eEF-1 și EF2 care au roluri similare cu EF-Tu, EF-Ts și EF-G.

### **Oprirea sau terminarea (SLIDE 24)**

Terminarea implică detectarea codonului de oprire, ruperea legăturii ester dintre ultimul ARNt și catena polipeptidică și eliberarea acesteia.

Sunt necesare două condiții:

- prezența unui codon care specifică (codifică) oprirea elongării;
- intervenția unui factor de disociere (**RF** sau Release Factor) capabil să recunoască semnalul de terminare a catenei și să favorizeze ruperea legăturii cu ARNt și eliberarea catenei.

În componența codului genetic există trei codoni de oprire recunoscuți (citiți) de proteine specifice care la *E. coli* sunt RF1 (recunoaște UAG și UAA) și RF2 (recunoaște UGA și UAA). Aceștia favorizează interacția catenei în formare cu o moleculă de apă și eliberarea lanțului. Când mărimea unei catene polipeptidice atinge douăzeci și cinci de aminoacizi, codonul de inițiere AUG de pe ARNm este complet liber și se poate iniția o nouă demarare. Fenomenul se reproduce de mai multe ori până când ARNm este acoperit de ribozomi distanțați între ei prin aproximativ douăzeci și cinci de nucleotide. Această unitate de traducere mare, numită **poliribozom** sau **polizom**, permite celulei reproducerea mai multor copii ale unei proteine plecând de la o moleculă de ARNm. Dacă ARNm este policistronic și dacă al doilea codon de inițiere nu este situat (poziționat) prea departe de primul codon de oprire (de terminare), ribozomul se deplasează alunecând și se leagă la codonul de inițiere al cistronului următor.

La eucariote, nu există reinițiere a sintezei proteice după întâlnirea codonului de terminare.

### **Traducerea informației genetice la eucariote**

Sinteza proteică sau traducerea este activitatea de sinteză cea mai complexă din celulă. În timp ce sinteza altor molecule celulare este rezultatul unor reacții enzimatiche relativ directe, asamblarea unei proteine necesită prezența tuturor moleculelor de ARNt și a tuturor aminoacizilor corespunzători atașați la aceștia, de ribozomi, de molecule de ARNm, de mai multe proteine cu funcții diverse, de cationic și de GTP. Această complexitate nu trebuie să ne mire dacă ținem cont că sinteza proteică necesită incorporarea a peste 20 aminoacizi în funcție de o secvență

precisă dictată de un mesaj codificat scris într-o limbă care folosește simboluri diferite. Procesul de traducere de la eucariote seamănă remarcabil cu cel de la procariote. Principală diferență este că în celula eucariotă traducerea implică un mare număr de factori proteici solubili (neribozomici).

### **Inițierea (SLIDE 25)**

Etapele fundamentale de inițiere a traducerii depind de atașarea corectă a ribozomului la o secvență bine determinată din structura ARNm pentru ca mesajul să poată fi citit într-un cadru de lectură corect. Aceasta se realizează deoarece ribozomul începe traducerea la nivelul unui situs specific de pe structura mesagerului numit **codon de inițiere**. ARNm nu se fixează la ribozomii constituiți (întregi). Acesta se unește la cele două subunități în momente diferite. Prima etapă esențială este unirea subunității ribozomale mici la prima (sau la una din primele) secvențe AUG ale mesajului care funcționează ca un codon de inițiere. Care este modalitatea prin care subunitatea mică alege codonul AUG inițial și nu un alt codon intern. Așa cum am arătat anterior, procariotele posedă o secvență specifică de nucleotide numită Shine-Dalgarno care este complementară cu o secvență nucleotidică apropiată de extremitatea 3' a ARN ribozomal 16S din subunitatea mică.

ARNr \_\_ACCUCCUUA3'

ARNm \_\_GGAGGA\_\_5'

Recunoașterea codonului de inițiere AUG este consecința unei interacții între aceste secvențe complementare de pe ARNm și de pe ARNr când subunitatea mică recunoaște mai întâi extremitatea 5' a mesajului, care poartă un capșon de metilguanozină. După această subunitatea mică baleiază secvența ARNm până când întâlnește o secvență de nucleotide (de obicei 5'-CCACCAUGC-3') care aproximativ 90% din moleculele de ARNm eucariot, tripletul AUG este codonul de inițiere. Dacă ținem cont de atribuirea codonilor putem spune că AUG este mai mult decât un codon de inițiere: este singurul codon pentru metionină. De fapt metionina este întotdeauna primul aminoacid incorporat la extremitatea amino a catenei polipeptidice în formare (la procariote metionina din poziția 1 poartă întotdeauna o grupare formil). În cea mai mare parte a cazurilor metionina (sau formil-metionina) este ulterior înlăturată enzimatic, cu toate că metionina rămâne aminoacidul terminal în aproximativ 50% din catenele polipeptidice mature. Celulele posedă două molecule de ARNt pentru metionină: una este folosită pentru inițierea sintezei proteice (ARNt<sup>i</sup>Met), iar cealaltă (reprezentată de ARNt<sup>t</sup>Met) intervine în incorporarea

resturilor de metionină în interiorul polipeptidului. Trebuie să reținem că mitocondriile și cloroplastele celulelor eucariote folosesc N-formil-metionină în locul metioninei pentru a iniția traducerea, ceea ce dă cele mai bune argumente în favoarea originii procariote a acestor organite. Mai multe dintre etapele de inițiere necesită prezența unor proteine solubile numite **factori de inițiere** (reprezentate de **IF** la procariote și de **eIF** la eucariote). De exemplu, la eucariote, se consideră că eIF4E se atașează coifului la extremitatea 5' a mesagerului și înlătură regiunile bicatenare care interferează cu deplasarea ribozomului în căutarea codonului de inițiere. În plus, pe lângă factorii proteici este necesară și prezența GTP pentru a realiza o inițiere reușită. O moleculă de GTP se unește cu ARN<sup>t</sup><sub>Met</sub> înainte de cuplarea sa cu subunitatea ribozomală mică. Imediat ce ARN<sup>t</sup> inițiator este atașat, subunitățile mari care se anexează complexului, factorii solubili sunt eliberați și GTP este hidrolizat. Hidroliza GTP poate fi cauza unei modificări a conformației ribozomului necesară inițierii traducerii. Fiecare ribozom posedă două situsuri de asociere a moleculelor de ARN<sup>t</sup>. Aceste două situsuri, situsul A (aminoacil) și a situsului P (peptidil) joacă roluri diferite în traducere. Aminoacil-ARN<sup>t</sup> intră în complexul ribozom-ARN<sup>m</sup> la nivelul situsului A, în timp ce la nivelul situsului P, ARN<sup>t</sup> livrează aminoacizi catenei polipeptidice în creștere. De fapt ARN<sup>t</sup> de inițiere (ARN<sup>t</sup><sub>Met</sub>) apare mai întâi la nivelul P al ribozomului.

### **Elongarea**

O dată ce ARN<sup>t</sup> de inițiere este plasat în situsul P, ribozomul este gata pentru fixarea celui de al doilea aminoacil-ARN<sup>t</sup> în situsul A vacant, prima etapă de elongare (etapa 1). Înainte ca cel de al doilea aminoacil-ARN<sup>t</sup> sau că următorii să se poată uni efectiv la situsul A de pe ARN<sup>m</sup>, acesta trebuie să se combine cu unul dintre factorii proteici de elongare (acești factori de elongare particulari sunt numiți Tu la procariote și eEF1 la eucariote) uniți cu GTP. Cu toate că orice complex aminoacil-ARN<sup>t</sup>-Tu-GTP poate intra în situsul A, va fi blocat la nivelul ARN<sup>m</sup> printr-o specifică numai cel al cărui anticodon este complementar codonului de pe ARN<sup>m</sup> situat în situsul A. Imediat ce complexul aminoacil-ARN<sup>t</sup>-Tu-GTP adecvat recunoaște codonul de pe ARN<sup>m</sup>, GTP este hidrolizat și se eliberează complexul Tu-GDP. Complexul ARN<sup>t</sup>-Tu-GTP este regenerat. Cea de a doua etapă a ciclului de elongare este formarea unei legături peptidice între aminoacizii atașați la cele două molecule de ARN<sup>t</sup>. Reacția se realizează prin transferarea metioninei (sau a N-formil metioninei) de pe ARN<sup>t</sup> inițiator din situsul P pe aminoacidul atașat

la ARNt care este complementar celui de al doilea codon prezent în situsul A. Reacția este catalizată de către peptidil transferază, element au subunității mari a ribozomului. Mult timp s-a considerat că peptidil-transferaza era una dintre proteinele ribozomului. Mai târziu, când capacitatea catalitică a ARN a devenit evidentă, atenția s-a îndreptat către ARN ribozomal ca fiind posibilul catalizator al legăturii peptidice. În ultimul s-a demonstrat că activitatea peptidil-transferazică este efectiv localizată pe molecula mare de ARN ribozomal din subunitatea mare a ribozomului. Cu alte cuvinte, peptidil transferaza este o ribozimă. Formarea primei legături peptidice lasă o extremitate a moleculei de ARNt în situsul A încă atașat la codonul său complementar de pe ARNm în timp ce cealaltă extremitate este atașată la o dipeptidă. ARNt prezent în situsul P nu mai posedă acum un aminoacid atașat. Cea de a treia etapă a ciclului de elongare implică eliberarea ARNt neîncărcat prezent în situsul P și deplasarea ribozomului cu trei nucleotide (un codon) de-a lungul ARNm în direcția 3'. Această ultimă etapă, numită **translocare**, este însoțită de deplasarea ARNt-dipeptidei, încă unită prin legături de hidrogen la al doilea codon al ARNm, în situsul P al ribozomului. Translocarea necesită un factor de elongare numit G la procariote și eEF2 la eucariote și hidroliza GTP. Pentru fiecare ciclu de elongare sunt hidrolizate cel puțin două molecule de GTP, una (sau mai multe) în timpul selecției aminoacil-ARNt și una în timpul translocării. Îndată ce peptidil-ARNt se deplasează în situsul P, situsul A este din nou disponibil unui nou aminoacil-ARNt al cărui anticodon este complementar cu al treilea codon. Când cel de al treilea ARNt încărcat este asociat la ARNm în situsul A, dipeptida legată de molecula de ARNt prezentă în situsul P este transferată aminoacidului ARNt din situsul A. ARNt din situsul P este din nou eliberat de aminoacid. Formarea unei noi legături peptidice este urmată de eliberarea ARNt din situsul P și de translocarea ribozomului către cel de al patrulea codon, și ciclul este gata să reînceapă.

## **PRECIZIA TRADUCERII**

Interacția între tripletele din structura ARNm și ARNt nu este una foarte puternică pentru a asigura ea însăși precizia bine cunoscută a sintezei proteice (mai puțin de un aminoacid din 10000 este incorect incorporat). Au fost propuse mai multe ipoteze care să explice frecvența foarte scăzută a erorilor. O ipoteză sugerează că numai codonii și anticodonii complementari sunt capabili să se insereze la nivelul fosei formate de către ARN ribozomal. Legătura peptidică se poate forma numai în cazul în care fosa mare este ocupată. O altă ipoteză pune accent pe intervalul apparent care separă unirea complexului ARNt-Tu-GTP la ARNm și momentul în care

aminoacidul este adăugat catenei polipeptidice în creștere. Cu cât codonul și anticodonul neadekvat (necomplementari) sunt mai mult în contact, cu atât este mai ipoteză face apel la faptul că hidroliza GTP poate asigura fidelitatea traducerii. Cu toate că până în prezent se presupunea că pentru fiecare ciclu de elongare se realizează hidroliza a două molecule de GTP există din ce în ce mai multe argumente în sprijinul ideii că numărul real de molecule de GTP hidrolizate este mult mai mare (cel puțin trei). Este posibil ca energia suplimentară să fie folosită pentru eliberarea moleculelor de ARNm ce nu sunt bine împerecheate. Streptomicina, care este unul dintre cele mai prescrise antibiotice acționează unindu-se selectiv la subunitatea ribozomală mică a celulelor bacteriene provocând o lectură eronată a anumitor codoni ai ARNm, crescând sinteza proteinelor aberante. Acest antibiotic nu acționează asupra ribozomilor de la eucariote și deci nu are efect asupra traducerii din celulele gazdă. Se pare că streptomicina se unește cu proteina ribozomală numită S12 care intervine într-un fel sau în altul în controlul preciziei interacției dintre moleculele de ARNt specifice și complexul ribozom-ARNm. Rezistența bacteriilor la streptomicină se poate explica prin modificări la nivelul proteinelor ribozomale și în particular la nivelul S12. Multe antibiotice acționează interacționând cu diferite etape ale sintezei proteice în celulele bacteriene (tabel).

Eu, Pro: celule eucariote, procariote; G, P: subunitatea mare sau mică; R: recunoaștere; P: transferul polipeptidului; T: translocare; I: inițiere;

### **Terminarea**

Cum s-a remarcat anterior trei din cei 64 codoni (de trei nucleotide) potențiali sunt codoni stop ale căror funcție este semnalizarea opririi asamblării polipeptidelor. Deci, nu există anticodoni în structura ARNt care să fie complementare codonilor stop. Când ribozomul ajunge la unul dintre acești codoni, UAA; UAG sau UGA, semnalul oprește orice nouă elongare și eliberează polipeptidul asociat ultimului ARNt. Terminarea necesită prezența factorilor de eliberare care reacționează direct cu codonii stop; bacteriile posedă doi factori de eliberare (RF1 care recunoaște UAA și UAG și RF2 care recunoaște UGA) în timp ce celule eucariote nu au decât un eRF. Factorul de eliberare poartă un GTP legat care este apoi hidrolizat. Imediat ce traducerea se oprește, legătura polipeptidului la ARNt este scindată și factorul de eliberare, ca și ARNt dezacilat, sunt eliberați de pe ribozom. Hidroliza legăturii între polipeptid și ultimul ARNt poate fi realizată de către aceeași regiune a ARNr care este responsabilă de formarea legăturii peptidice în timpul elongării. Imediat ce se realizează terminarea, ribozomul se separă de mesager și se

disociază în subunități până la inițierea unui nou ciclu de traducere. Cum cei trei codoni de terminare pot fi ușor obținuți prin modificarea bazelor plecând de la alți codoni, ne-am putea aștepta la apariția unor mutații care dau codoni stop la interiorul unei gene. Aceste mutații ar putea provoca o terminare prematură a catenei polipeptidice în creștere. Mutațiile de acest tip sau mutațiile non-sens, au fost studiate timp de ani de zile și se știe că acestea sunt responsabile de diferite maladii ereditare la om. Mutațiile non-sens antrenează de obicei formele de maladii ereditare cele mai severe deoarece se realizează numai sinteza unei porțiuni a proteinei și aceasta este întotdeauna nefuncțională. Terminarea prematură a creșterii catenei poate fi determinată și de adăugarea unui antibiotic, cum este puromicina. Această moleculă seamănă cu extremitatea 3' a unui ARNt încărcat și poate intra în situsul A al ribozomului (figura). Când aceasta se plasează la acest nivel, polipeptida din situsul P este transferată puromicinei din situsul A cu formarea unei legături covalente. Transferul catenei polipeptidice în formare, pe puromicină determină formarea unui capișon la extremitatea carboxil, astfel încât alți aminoacizi nu se mai pot atașa complexului peptidil-puromicină, iar acesta se desprinde de pe ribozom.

### **Formarea de poliribozomi**

Dacă observăm la microscopul electronic o moleculă de ARNm în curs de traducere vedem întotdeauna un anumit număr de ribozomi atașați de-a lungul fragmentului de ARNm. Acest complex format din ribozomi și ARNm numit poliribozomi. Mai întâi, toți ribozomii se assemblează din subunitățile lor la nivelul situsului de inițiere, apoi ei se deplasează de la acest punct către extremitatea 3' a ARNm până când ating codonul de terminare. Imediat ce primul ribozomi se află la o distanță destul de mare de codonul de inițiere, un al doilea ribozom se atașează la ARNm și demarează o altă activitate de traducere. Traducerea simultană a aceluiași ARNm de mai mulți ribozomi crește puternic viteza de sinteză a proteinelor celulare.

Chiar dacă procesele se aseamănă ele sunt diferențiate la eucariote și la procariote deoarece la eucariote procesul de transcripție și cel de traducere sunt compartimentate și se desfășoară fiind separate în citoplasmă și în nucleu. În celulele bacteriene, sinteza proteică debutează la nivelul moleculelor de ARNm chiar înainte de terminarea transcripției acestora. Sinteza unei molecule de ARNm progresează în același sens ca și deplasarea ribozomilor care traduc mesajul, adică în direcția 5'-3'. În concluzie, o moleculă de ARNm la procariote va fi tradusă pentru traducere imediat ce extremitatea sa 5' este disponibilă pentru fixarea ribozomilor. În figură este reprezentat filamentul de ADN pe care transcripția este în curs. Se pot observa molecule de

ARNm din ce în ce mai lungi la nivelul cărora sunt atașați ribozomi pentru a forma poliribozomii. Catenele proteinelor în formare nu sunt vizibile în micrografiile realizate în cazul traducerii unor proteine normale dar se pot observa la nivelul singurului poliribozom izolat dintr-o celulă a glandelor sericigene de la viermele de mătase. Proteina de mătase în curs de sinteză este vizibilă din cauza mărimi sale și a maturii sale fibroase. Posibilitatea vizualizării transcripției și traducerii, oferită de Oscar Miller de la Ridge National Laboratory, a constituit o etapă în vizualizarea unui proces care până în acel moment numai în termeni biochimici.