

## CAP.2.1. CROMOZOMUL BACTERIAN

### A. Structura fizică și chimică a cromozomului bacterian

#### Modelul de împachetare

Imensa majoritate a bacteriilor dețin ca material genetic esențial o moleculă ADN dublu catenar circular covalent închis, suprarăsucit negativ și împachetat. Deși complexarea cu proteine și împachetarea nu este identică cu cea din cromozomii de tip eucariot, totuși această moleculă este denumită tot "cromozom" sau nucleoid. Cromozomul bacterian este atașat la membrana plasmatică a celulei bacteriene în aproximativ 20 de puncte, dar o funcție deosebită pare să aibă punctul de atașare de lângă regiunea *ori C*.

Cei mai mici cromozomi de tip procariot măsoară mai puțin de 1 Mpb și se întâlnesc mai ales la bacteriile fără perete celular (bacteriile din genurile *Mycoplasma* și *Ureaplasma* au cromozomi cu dimensiuni cuprinse între 600 și 800 kpb). La cealaltă extremă se află bacteriile din genurile *Myxococcus* și *Calothrix*, care au cromozomi foarte mari (aproximativ 12-13000 kpb). *Escherichia coli* are un cromozom de dimensiune intermediară: 4700 kpb.

În general, molecula ADN ce formează cromozomul bacterian este de aproximativ 1000 de ori mai lungă decât celula bacteriană. Această moleculă este complexată cu proteine, suprarăsucită și împachetată formând o structură conformă cu modelul elaborat de **Pettijohn** în 1974 (Fig.2.1.). Conform acestui model, prin asocierea ADN cu proteine și cu ARN (de regulă, ARN nascent) se formează aproximativ 50 de domenii topologice, semi-independente, per genom de *E.coli*. Fiecare domeniu (buclă) este suprarăsucit separat de celelalte domenii și poate fi relaxat independent de celelalte prin introducerea unei rupturi monocatenare. Gradul de suprarăsucire negativă este dat de balanța dintre activitatea a doua enzime - ADN giraza și ADN topozomeraza I - amândouă reglând densitatea helicală a moleculelor de ADN.

Una dintre cele mai importante specii proteice cu rol în stabilirea și menținerea arhitecturii cromozomului bacterian este ADN giraza (enzimă ce face parte din clasa ADN topozomerazelor II). Această proteină se atașează la molecula cromozomului bacterian într-un mod situs-specific (la secvențele **REP**), atașarea fiind stimulată de interacțiunea cu **proteina HU** (*Helix Unwinding*). Moleculele de ADN girază sunt distribuite la intervale de aproximativ 100 kpb, ceea ce corespunde la o moleculă de ADN girază per domeniu topologic.

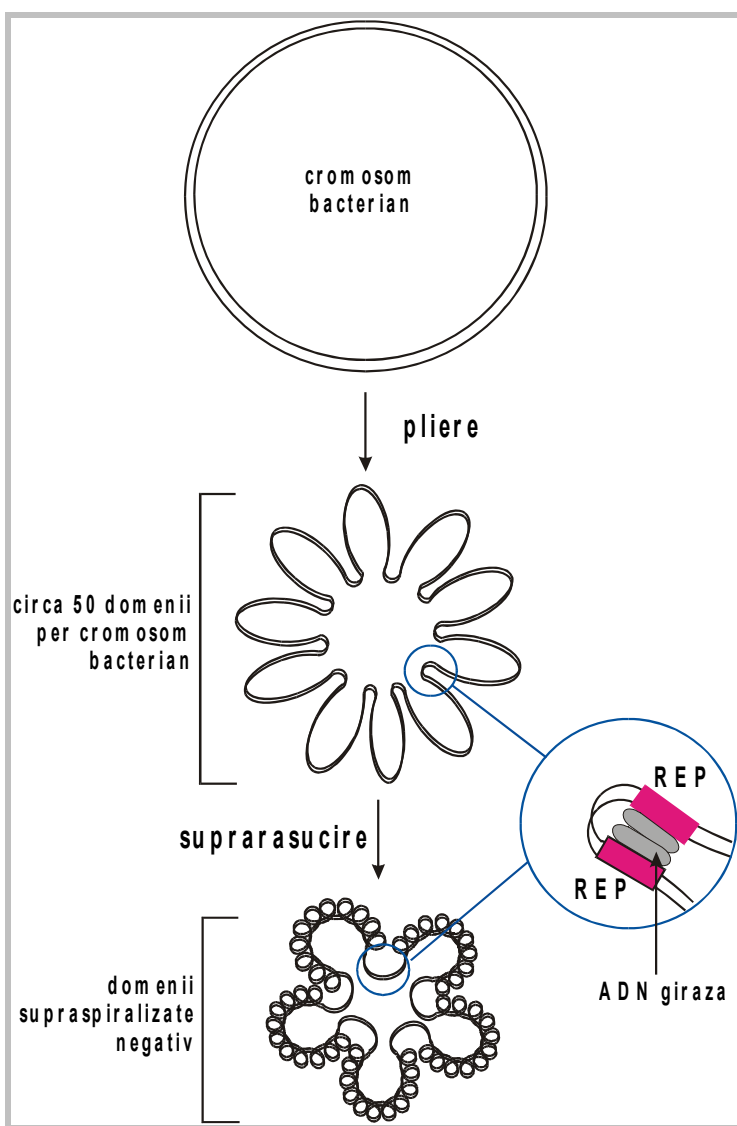


Fig.1.1. Modelul de împachetare a cromozomului bacterian [Pettijohn, 1974].

### Configurații ADN în cromozomul bacterian

Principalul tip de structură secundară a ADN ce formează cromozomul bacterian este **forma B de dreapta**, dar această configurație nu se găsește uniform de-a lungul întregului cromosom bacterian. Alte configurații, prezente pe distanțe scurte, sunt ADN-Z (în dublu helix de stânga), structuri cruciforme (în regiunile cu secvențe invers repetate) și regiuni triplu-catenare (în zone extrem de bogate în purine sau pirimidine).

Este de reamintit faptul că, în general, expresia genică este afectată nu numai de secvența primară a ADN (succesiunea de nucleotide), ci și de structura secundară și terțiară a acestuia.

### Compoziția în nucleotide

În majoritatea cazurilor, compoziția globală în nucleotide a unei molecule de ADN se exprimă prin procentul molar de guanină + citozină (%mol GC), restul până la 100% fiind, evident, reprezentat de adenină + timină (datorită criteriului de complementaritate între cele două catene ADN, respectiv între bazele purinice și cele pirimidinice). Compoziția în nucleotide a cromozomului bacterian variază în limite foarte largi: 25 %mol GC la *Mycoplasma capricolum* și 75 %mol GC la *Micrococcus luteus*.

Procentul molar de guanină + citozină este corelat și cu compoziția în codoni a genomului. Aceasta variază în sensul preferinței pentru 1-2 codoni dintr-un grup de codoni sinonimi. Se constată astfel că la *Mycoplasma capricolum* este favorizată prezenta A/T în poziția 3-a a codonilor sinonimi.

S-a mai constatat că în cromozomul bacterian există și o serie de gene a căror activitate poate afecta compoziția totală în nucleotide a unei molecule de ADN. Astfel, la *E.coli* au fost descrise două gene (**mut T** și **mut Y**) care afectează frecvența transversilor A-T / C-G și, respectiv, C-G / A-T. Echilibrul între exprimarea acestor două gene afectează compoziția globală în nucleotide a moleculei de ADN ce reprezintă cromozomul de *E.coli*. În mod evident, codonii favorizați se corelează cu abundența moleculelor de ARNt cognate din microorganismul respectiv. Astfel, există codoni "optimi" ce sunt cel mai corect și eficient recunoscuți de către cele mai abundente molecule de ARNt.

O altă problemă o reprezintă contextul de citire a unui anumit codon, în speță, configurația codonilor adiacenți. Numărul teoretic posibil de codoni (sens) adiacenți este foarte mare ( $612 = 3721$ ), dar, examinând 237 de gene de la *E.coli* s-a constatat că perechile de codoni adiacenți nu sunt distribuite randomizat, ci anumite perechi de codoni sunt mai abundente decât altele. S-a dedus astfel că, compoziția în nucleotide a unui codon este corelată și cu compoziția codonilor adiacenți, corelat probabil cu procesele de atașare la situsurile A (*Aminoacil*) și P (*Peptidil*) ale ribozomilor. Se pare deci, că aparatul de traducere a informației genetice ar fi putut să determine evoluția unor anumite trăsături ale matriței genetice.

## B. Principalele clase proteice asociate cu cromozomul bacterian

În afară de ARN polimeraza, care datorită ratei înalte de transcriere la procarote, rămâne asociată cvasi-permanent cu moleculele de ADN, cromozomul bacterian este asociat cu o serie de proteine bazice, similare cu histonele de la organisme eucariote și denumite **proteine histone-like**. La *E.coli* au fost descrise 9 specii moleculare de proteine bazice, cu g.m. între 9 și 28 kd, care se atașează la ADN într-o manieră situs-nespecifică și care au funcții similare cu histonele de la organisme eucariote.

Dintre cele 9 specii moleculare proteice, cea mai importantă este o proteină denumită **HU** care are caracter bazic și o g.m. de 9.7 kd. Proteina HU funcționează ca dimer, iar un monomer de HU este format din două polipeptide, denumite **HU- $\alpha$**  și **HU- $\beta$**  și codificate de gene diferite: **hup A** și, respectiv, **hup B**. Proteina HU are proprietăți fizice și compoziție totală în aminoacizi similară cu histonele de la eucariote, dar secvența de aminoacizi este diferită de acestea. Ca și histonele de la eucariote, proteina HU are un grad înalt de conservare în lumea bacteriană. Există aproximativ 25000 de molecule HU per genom de *E.coli*, dar aceste molecule nu sunt dispuse uniform de-a lungul cromozomului bacterian, ci sunt mai dense la periferia nucleoidului (în zonele extrem de active transcripțional).

Din complexarea moleculelor HU cu ADN cromozomal bacterian se formează structuri similare cu nucleosomii de la eucariote, denumite **structuri nucleosom-like** (chiar și în condiții *in vitro*), în care sunt cuprinse circa 200 pb/nucleosom. Structurile **nucleosom-like** nu sunt statice, ci se află într-un echilibru dinamic. Studii mai ample au demonstrat faptul că proteina HU nu este o

simplă proteină structurală a arhitecturii cromozomului bacterian, ci are rol foarte important în diverse procese din celula bacteriană – în mod special în procese ce implică interacțiuni ADN-proteine, situație în care moleculele HU favorizează atașarea altor proteine la ADN, favorizând inclusiv interacțiunea girazei cu situsurile REP.

Alte proteine *histone-like* de la *E.coli* sunt **proteina H** (similară cu histona H2A de la eucariote), **proteina H1**, **proteina H-NS**, **proteina IHF**. Proteina IHF (*Integration Host Factor*) a fost descoperită în studiile privind integrarea genomului fagului  $\lambda$  în cromozomul de *E. coli*. Ulterior, s-a constatat că această proteină intervine și într-o serie de alte procese din celula bacteriană. Molecula de IHF (g.m.=20 kd) este formată din 2 subunități codificate de 2 gene distincte: *him A* și *him B*. Spre deosebire de proteina HU și chiar de alte proteine *histone-like*, moleculele de IHF se atașează la ADN într-o manieră situs-specifică. Moleculele de IHF nu sunt necesare pentru recombinarea bacteriană omoloagă, dar par să intervină în fenomene de recombinare neomoloagă, specializată, cum sunt de altfel, procesele de transpoziție și de conjugare bacteriană.

Deși proteinele *histone-like* au fost studiate extensiv la *E.coli*, totuși au fost identificate proteine cu funcții omoloage și la alte genuri și specii bacteriene. Pe de altă parte, s-a constatat că o serie de microorganisme din grupul **Archaea** prezintă proteine *histone-like* cu structură intermediară între cele de la **Bacteria** și histonele de la **Eukarya**.

### C. Secvențe ADN repetate

În cromosomul bacterian au fost identificate secvențe ADN repetate, dintre care unele nu codifică proteine/ARN (secvențe repetate necodificatoare), iar altele codifică o serie de proteine sau specii moleculare de ARN (ARNr, ARNt) – secvențe repetate codificatoare.

#### 1) Secvențe repetate necodificatoare = ADN repetitiv

Acestea sunt reprezentate de secvențe scurte ce apar de regulă în afara ORF-urilor (*Open Reading Frames*). În mare majoritate a cazurilor, asemenea secvențe servesc ca situsuri de interacțiune ADN-proteine (inclusiv în procese de recombinare, inversie, excizie, transpoziție) și ca situsuri modificate ce identifică catena matriță în timpul replicării semiconservative a cromozomului bacterian. Dintre cele mai cunoscute asemenea secvențe sunt secvențele *REP*, *Chi*, *Dam*.

#### Secvențe REP (Repeated Extragenic Palindromes)

Secvențele REP sunt formate din palindroame repetate, majoritatea aflate în regiuni extragenice. O secvență REP conține o secvență consensus de 38 bp. Clusteri de 2-4 secvențe REP separate între ele prin 20 pb formează un **element REP (REPE = REP Element, Fig.1.2.)**. La *E.coli* și *Salmonella typhimurium* există 100-200 structuri REPE (reprezentând deci, aproximativ 0.5% din cromozomul bacterian), localizate însa diferit.

La secvențele REP se leagă molecule de proteină HU și de ADN girază, aceste secvențe având deci rol în împachetarea nucleoidului. Tot la aceste secvențe se leagă și ADN polimeraza intervenind în procese de replicare și reparare ADN.

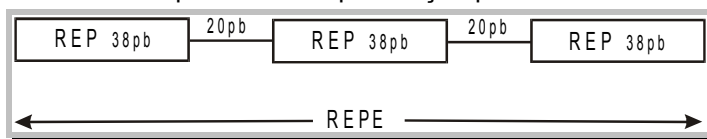


Fig.1.2. Organizarea elementelor REP la *Escherichia coli*.

#### Situsuri Chi Un situs *Chi* are o lungime de 8 pb: 5' GCTGGTGG 3'

Datorită faptului că reprezintă situsul de recunoaștere și tăiere a moleculelor ADN de către **endonucleaza RecBCD**, secvențele *Chi* stimulează recombinarea omoloagă mediată de **RecA** și **RecBCD**. S-a mai constatat că efectul recombinativ al situsului *Chi* este polar: stimulează recombinarea la capatul 5' al lui *Chi*, și nu la 3'.

Situsuri *Chi* au fost identificate atât pe cromosom, cât și pe unele plasmide de la *E.coli*. Mai mult chiar, s-a constatat că secvențele *Chi* reprezintă situsuri de stimulare a recombinării genetice la toți membrii familiei *Enterobacteriaceae*. În contrast, la *Pseudomonadaceae*, complexul enzimatic RecBCD nu acționează la situsuri *Chi*.

În cromosomul de *E.coli* au fost identificate aproximativ 950 de situsuri *Chi*, ce par să fie distribuite relativ uniform (1 situs *Chi* la 5 kbp), cu excepția zonei adiacente regiunii *oriC*, unde există un cluster de 22 de situsuri *Chi*.

#### Situsuri Dam sunt formate din 4 bp (5' GATC 3') și reprezintă situsurile de metilare a adeninei (adenina este metilată la poziția N6). Cromosomul de *E.coli* conține foarte multe situsuri *Dam* (peste 18.000) care, dacă ar fi dispuse randomizat ar exista 1 situs *Dam*/250

bp. Și în acest caz s-a constatat un număr foarte mare de situsuri *Dam* în, și în jurul, regiunii *oriC*: în cele 245 bp ce definesc *oriC* la *E.coli* există 8 situsuri *Dam*, iar în cele 350 bp ce flanchează *oriC* există încă 12 situsuri *Dam*.

**2) Secvențe repetate codificatoare** au dimensiuni mult mai mari decât cele necodificatoare. **operoni *rrn*** există la toate bacteriile și cuprind secvențe ce codifică pentru molecule de ARN ribozomal (ARNr). La cele mai multe dintre speciile bacteriene, linkage-ul în cadrul unui operon *rrn* este 16S – 23S – 5S.

*E.coli* și *S.typhimurium* au câte 7 loci (operoni) *rrn*, localizați în poziții echivalente. Cei 7 operoni sunt notați ***rrnA*....*rrnG***, fiecare cuprinzând secvențe 16S-23S-5S cotranscrise într-o moleculă ARN de 30S. Numărul operonilor *rrn* variază de la un gen/specie bacteriană la alta: *Bacillus subtilis* are 10 operoni *rrn*, *Mycobacterium* sp. are 1-2 operoni *rrn*.

Chiar în cadrul aceleiași tulpini bacteriene, locii *rrn* pot diferi între ei în ceea ce privește prezența/absența genelor pentru ARNt în regiunile “*spacer*” dintre genele ARNr. Astfel, toți cei 7 operoni *rrn* de la *E.coli* au 1-2 gene pentru ARNt între secvențele pentru 16 și 23S, și 0-2 gene ARNt după 5S.

Distribuția locilor *rrn* pe cromosom variază la diverse specii bacteriene: la *E.coli*, majoritatea locilor *rrn* sunt localizați în jumătatea cromosomală ce are în centru regiunea *oriC*; la *B.subtilis*, locii *rrn* sunt grupați într-o zonă ce reprezintă 30% din cromosom.

### **genele pentru ARNt**

Cromosomul de *E.coli* conține 41 de gene/operoni pentru ARNt distribuiți în tot cromosomul. Unii din acești operoni codifică pentru o moleculă de ARNt, alții pentru mai multe molecule de ARNt, iar alții au și secvențe ce codifică diverse proteine.

**secvențe *rhs*** (“*rearrangement hot-spots*”) sunt capabile să genereze duplicații genice prin procese de crossing-over inegal între secvențe repetate. La *E.coli* K-12 au fost identificate 4 secvențe ***rhs***, notate ***rhs A, B, C și D***. Fiecare din aceste 4 secvențe au dimensiuni între 8 și 9 kbp și sunt compuse dintr-o regiune “*core*” (*miez*) (cu secvență conservată de aproximativ 3700 bp) și din segmente flancatoare. Prin exprimarea regiunilor “*core*” ale secvențelor *rhs A, B și C* sunt sintetizate 2 proteine cu funcție deocamdată necunoscută. În contrast cu *E.coli* K-12, la *E.coli* B și *E.coli* C, precum și la *S.typhimurium*, nu au fost descrise secvențe *rhs*.

## **D. Numărul de copii genice**

La bacterii, în afară de genele sau operonii aflați în copii multiple, numărul de copii ale unei gene cromosomale poate varia în funcție de următorii factori:

1. *poziția genei pe cromosom, de-a lungul unui gradient pornind de la oriC spre regiunea ter*. În timpul replicării cromosomului bacterian, în funcție de poziția lor pe cromosom, unele gene sunt deja replicate, în timp ce altele încă nu. Astfel, unele gene se găsesc într-un număr mai mare de copii decât altele.
2. *copii ale unei gene cromosomale pot exista și pe plasmide*. În această situație se folosește termenul de ***meroploidie parțială***, care se referă la o ploidie parțială ce afectează anumite gene, fără să existe multiplii de cromosomi întregi (este cazul genelor prezente și pe cromosom și pe un plasmid).
3. *într-o celulă bacteriană pot exista mai multe copii ale întregului cromosom*.

La anumite specii bacteriene există mai multe copii cromosomale în aceeași celulă: *Desulfovibrio vulgaris* (4 copii), *D.gigas* (17 copii), *Azotobacter chroococcum* (20-25 de copii), *A.vinelandii* (40-80 de copii cromosomale).

Este de subliniat faptul că la majoritatea speciilor bacteriene cu copii cromosomale multiple există procese de inactivare cromosomală ce asigură funcționarea unui singur cromosom și inactivarea celorlalți.

## E. Densitatea informației în cromosomul bacterian

Densitatea informației biochimice, adică numărul de reacții biochimice catalizate per kb de material genetic, este un parametru extrem de variabil în genomul bacterian, în primul rând datorită faptului că nu întotdeauna genele au o relație 1-la-1 cu reacțiile biochimice. În mod uzual, este valabilă relația 1 genă => 1 polipeptidă => 1 reacție biochimică. Există însă și foarte multe excepții, de exemplu:

- *enzime formate din mai multe lanțuri polipeptidice*: de exemplu, enzima succinat dehidrogenaza este formată din 4 polipeptide codificate de 4 gene diferite – **sdh A, B, C și D**. În acest caz, relația este 4 gene => 1 reacție biochimică;
- *enzime polifuncționale, ce catalizează mai multe reacții biochimice*: de exemplu, complexul enzimatic FAD ce oxidează acizii grași este format din 2 polipeptide codificate de 2 gene diferite – **fadA și fadB**. Acest complex enzimatic catalizează 5 reacții biochimice (din care 4 sunt catalizate de **FadA**). În acest caz, relația este 1 genă => 4 reacții biochimice.

În cadrul genomului bacterian variază și densitatea de citire a informației genetice prin transcriere. Astfel, deși marea majoritate a genelor bacteriene sunt contigue, totuși cromosomul bacterian cuprinde și gene cu diverse grade de suprapunere:

- *gene cu promotor plasat în regiunea terminală a genei "upstream"*: **trpA-trpB, ilvA-ilvD**
- *gene cu promotor plasat în interiorul genei "upstream"*: **mioA-mioD**
- *gene cu grad ridicat de suprapunere*:

**1. gene suprapuse total**, codificate pe cele 2 catene ale ADN; de exemplu, locusul **CysE** (codifică serin-acetil-transferaza), ce este implicat în biosinteza cisteinei, are 2 ORF-uri (**cys X** și **cys E**) codificate pe cele 2 catene ale locusului CysE (Fig.1.3.).

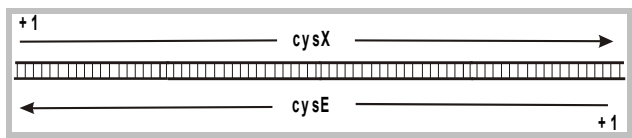


Fig.1.3. Organizarea ORF-urilor în locusul CysE.

**2. secvențe codificatoare ce sunt folosite de mai multe ori**, pornind transcrierea din poziții diferite; de exemplu, locusul **McrB** (restricția ADN la 5-metil-citozină) conține 3 ORF-uri pe aceeași catenă ADN, pornind din 3 situsuri diferite de inițiere a transcrierii (Fig.1.4.). Traducerea celor 3 transcripse duce la formarea a 3 proteine (de 51, 53 și, respectiv, 54 kdal) cu secvență carboxi-terminală identică, dar diferită în regiunea amino-terminală.

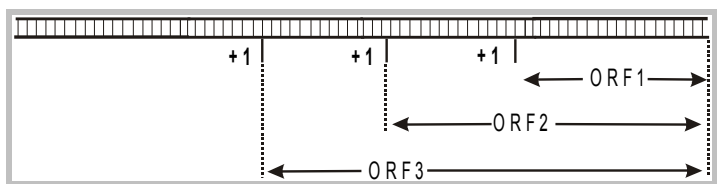


Fig.1.4. Structura locusului mcrB.

**3. gene suprapuse total**, cu molecule transcript identice, dar traduse diferit prin procese de "frameshift" translational (Fig.1.5.): de exemplu, gena **trpR** (codifică 2 proteine, dintre care una este **TrpR** ce are funcție de represor multiplu al exprimării unor diverse gene, inclusiv pentru operonul **trp** – implicat în biosinteza triptofanului); alt exemplu îl constituie gena **dnaX**, care codifică 2 subunități (**gamma** și **tau**) ale ADN polimerazei III.

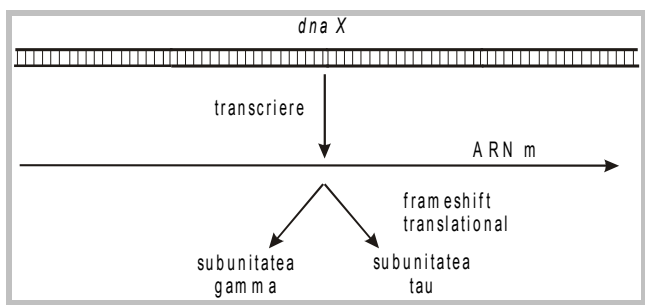


Fig.1.5. Frameshift translational în gena dnaX.

Asemenea gene – cu grad ridicat de suprapunere – sunt foarte bogate în informație genetică (unele au chiar informație dublă). Multe din ele (de exemplu, genele *mcrB* și *dnaX*) nu sunt tipice pentru cromosomul de *E.coli*, în ceea ce privește procentul molar de guanină + citozină. Astfel, gena *mcrB* are 48% mol GC, iar gena *dnaX* are 58 % mol GC, în timp ce cromosomul de *E.coli* are, *in toto*, 51-51.5 % mol GC. S-a emis ipoteza că asemenea gene care prezintă un procent molar de guanină + citozină extrem de diferit față de media cromosomală, ar fi fost achiziționate de *E.coli* din alte genomuri, prin procese de transfer de material genetic pe orizontală (transformare, conjugare, transducție).

Un alt parametru legat de densitatea informației biochimice în cromosomul bacterian este determinat de **redundanța genelor și a produselor genetice**. Astfel, la *E.coli*, pentru foarte multe funcții celulare, există câte 2 gene, ca și cum programul genetic al acestor bacterii ar avea sisteme de “backup” (copii de siguranță). Această redundanță ar putea fi considerată ca fiind o densitate redusă a informației genetice. Cu toate acestea, cel puțin în unele cazuri, această informație este reglată diferit și folosită în scopuri metabolice diferite. Astfel, în cromosomul de *E.coli* există 58 de perechi (2 gene) sau clusteri (mai mult de 2 gene), ce codifică 125 de produse genice. Din acestea, în 56 de perechi/clusteri reacțiile catalizate sunt identice. În cazul anumitor grupuri genice, reglajul exprimării lor este diferit, de exemplu:

- genele *aroF*, *aroG* și *aroH* codifică, toate trei, o aceeași enzimă: 3-deoxi-D-arabinoheptulonat – 7 – fosfat-sintetaza (DAHPS). Enzima codificată de *aroF* este sensibilă la concentrația de tirozină, cea codificată de *aroG* este sensibilă la concentrația de fenil-alanină, iar cea codificată de *aroH* este sensibilă la concentrația de triptofan ;

- genele *glpA* și *glpD* codifică amandouă glicerol-3-fosfat-dehidrogenaza. Enzima codificată de *glpA* este sintetizată și folosită în condiții de anaerobioză, iar cea codificată de *glpD* – în condiții de aerobioză.

Unele din aceste gene pereche au secvență similară una cu cealaltă (și proteinele corespunzătoare au secvență similară în aminoacizi) și este probabil că au apărut prin procese de duplicație genică, urmată de o oarecare divergență funcțională.

Altele însă, deși enzimele desfășoară aceeași funcție, totuși au secvență diferită (atât ca proteine, cât și genele corespunzătoare). Este posibil ca asemenea gene să fi apărut fie prin evoluție convergentă, fie prin transfer lateral de la alți repliconi bacterieni (această ultimă situație poate fi identificată prin determinarea procentului molar de guanină + citozină).

## F. Situsuri de inserție fagică în cromosomul bacterian

Mulți bacteriofagi își inseră genomul în cromosomul bacterian, fie prin transpoziție în poziții randomizate în cromosom (bacteriofagul *Mu*), fie prin recombinare situs-specifică numai în anumite poziții (fagii *lambda*).

Situsurile cromosomale în care fagul *lambda* (și fagii *lambda*) își inseră genomul, sunt denumite situsuri *attB* (“attachment on *Bacteria*”) și corespund unor situsuri *attP* (“attachment on *Phage*”) pe genomul fagic. Fagul *lambda* necesită un situs *attB* de minimum 21 bp, din care 15 sunt identice cu 15 bp din *attP* (234 bp). *Lambda* are un situs preferat *attB* în cromosomul de *E.coli* K-12 (dar și situsuri secundare) ce e situat intergenic. Alți fagi *lambda* (**21**, **e14**, **P22**) se inserează în situsuri *attB* cu localizare intragenică.

Pe cromosomul de *E.coli* K-12 au fost cartate situsurile în care se integrează genomurile a 15 fagi *lambda* și s-a constatat că ele sunt grupate între pozițiile 6 min și 44 min (Fig.1.6.).

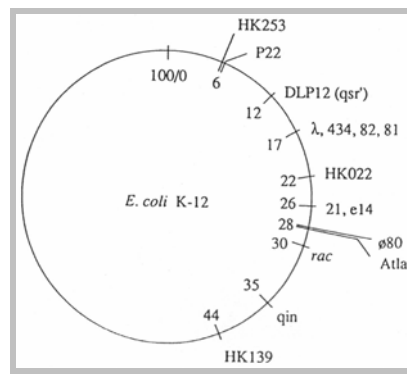


Fig.1.6. Localizarea situsurilor de integrare pentru fagi *lambda* pe cromosomul *E.coli* K12.

Această organizare a situsurilor în care se inseră genomul fagilor *lambda* reprezintă încă un argument în favoarea ipotezei conform căreia cromosomul de *E.coli* (și nu numai) ar avea o origine himerică, unul din segmente derivând dintr-o gazdă ancestrală a fagilor *lambda*.

Pe de altă parte, s-a constatat că foarte multe elemente genetice cu caracter “mobil” se inseră în genele pentru ARNt, fapt pentru care s-a sugerat că acestea ar fi reprezentat situsurile folosite de fagii ancestrali.

