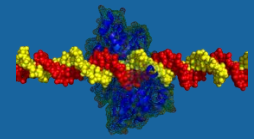


Tehnici de Biologie Moleculară

10.12.2024

Curs 4 – Amplificarea enzimatică *in-vitro* a acizilor nucleici. Secvențierea acizilor nucleici

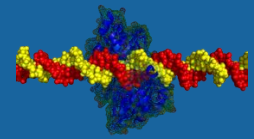


Ce este reacția PCR?

Folosind oricare din metodele descrise în capitolul anterior, purificarea acizilor nucleici din diverse surse devine o sarcină relativ simplă, ADN-ul obținut putând fi folosit direct în aplicații diverse. În unele cazuri însă, metodele descrise duc la obținerea unor cantități foarte mici de ADN, făcând etape ulterioare extrem de dificil de realizat. În asemenea situații se face uz de o altă tehnică deosebit de importantă în laboratorul de biologie moleculară – **amplificarea enzimatică *in vitro* a acizilor nucleici**. Tehnica este cunoscută sub denumirea de reacția PCR, *polymerase chain reaction* – **reacția în lanț a polimerazei**.

Amplificarea enzimatică *in vitro* a ADN-ului sau PCR-ul este așadar o **reacție de replicare enzimatică repetată a unei molecule de ADN matriță**, reacție ce este mediată de un set oligonucleotide și realizată de o ADN-polimerază ADN-dependentă. Pentru realizarea acestei reacții sunt necesare următoarele componente:

1. O **moleculă de ADN matriță** ce conține secvența ce se dorește a fi amplificată;
2. o **ADN polimerază** – enzimă capabilă să polimerizeze o nouă catenă de ADN pe baza complementarității cu ADN-ull matriță; **Cel mai frecvent se utilizează enzime termostabile** precum DNA polimeraza Taq din *Thermus aquaticus* sau DNA-polimeraza Pfu din *Pyrococcus furiosus*;
3. **2 oligonucleotide amorsă (primers)** – complementare cu capetele 3' ale catenelor sens și anti-sens ale ADN-ului matriță;
4. **4 deoxiribonucleozid-trifosfat** (dNTPs: dATP, dGTP, dCTP; dTTP) – **care este rolul lor?**
5. o **soluție tampon** în care ADN-polimeraza este activă și stabilă;
6. **Ioni bivalenți**, cel mai frecvent Mg^{2+} sau Mn^{2+} ; și **ioni monovalenți**, cel mai frecvent K^+ ;

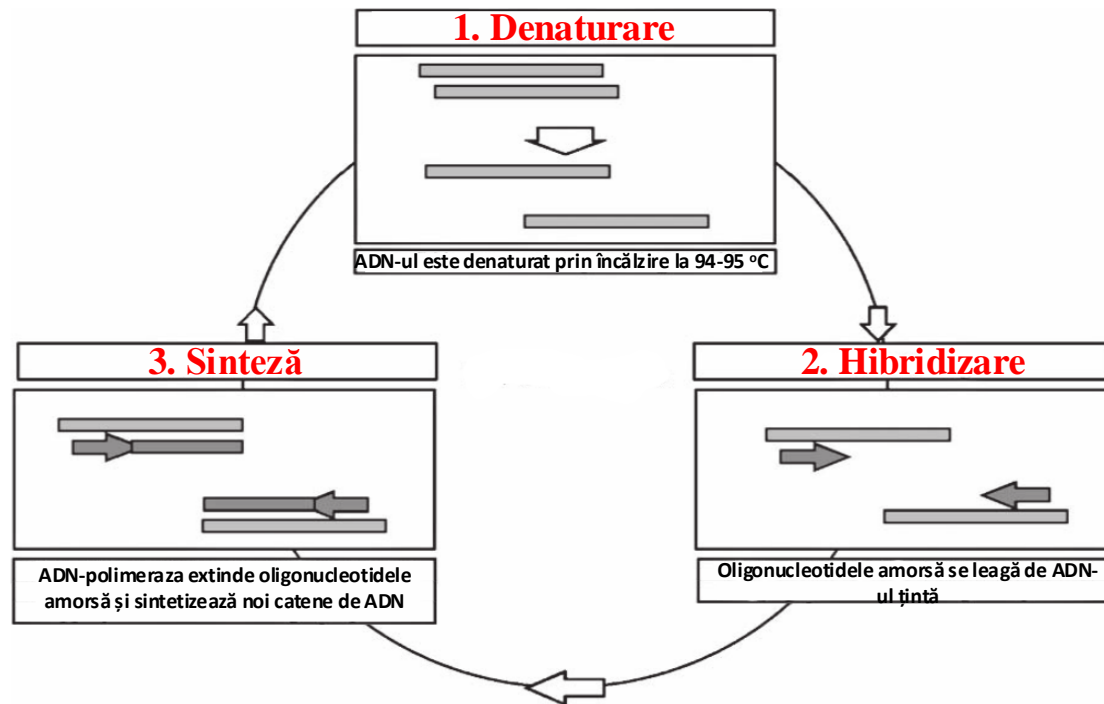


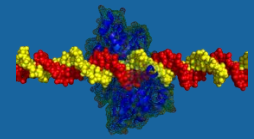
Etapele unei reacții PCR

Reacția PCR are la bază:

1. **proprietatea oligonucleotidelor amorsă (primer) de a se hibridiza** (de a se lega pe bază de complementaritate), la o anumită temperatură, de molecula ADN matrită. Oligonucleotidele amorsă hibridizate vor expune un capăt 3`-OH;
2. **proprietatea ADN-polimerazei de a încorpora nucleotide pe bază de complementaritate** în direcția 5`-3`, începând cu capătul 3`-OH al oligonucleotidelor amorsă (primer).

.....și constă în replicarea ciclică a **3 etape**:

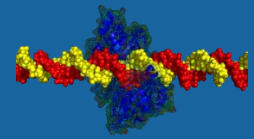




Etapele unei reacții PCR

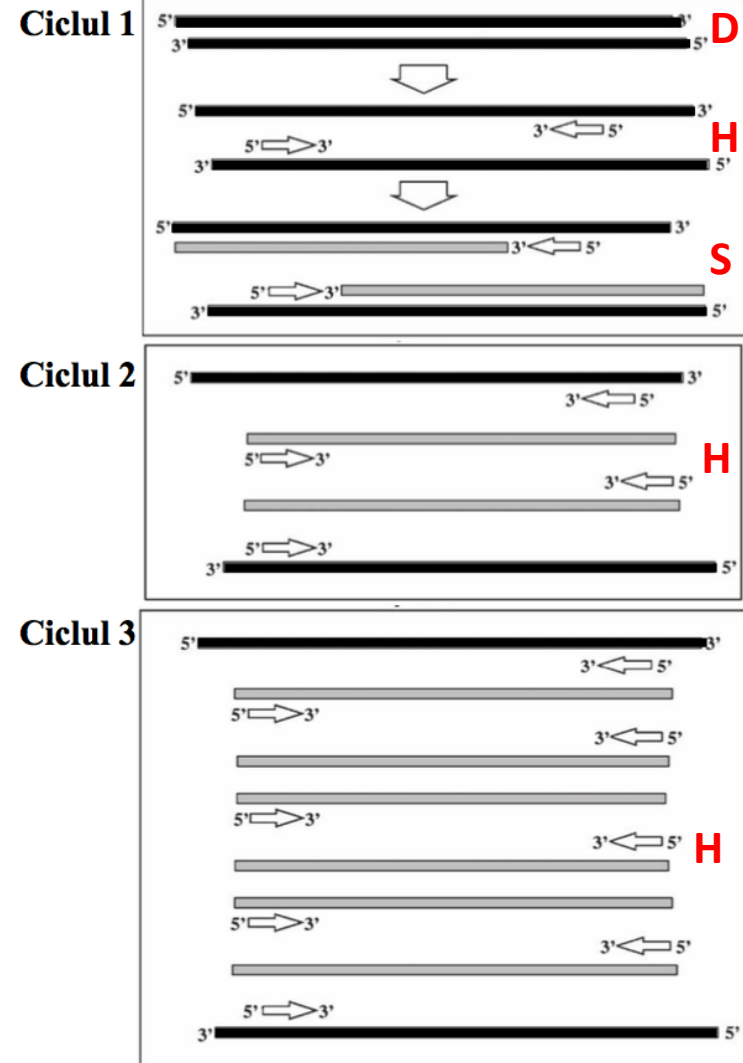
- 1. Denaturarea** constă în incubarea amestecului de reacție alcătuit din sursa de ADN-matriță, polimeraza și oligonucleotidele amorsă și cele 4 deoxiribonucleozide-trifosfat la temperatura de 94-95 °C timp de un minut. Prin încălzire, ADN-ul dublu-catenar din amestec este denatura, separându-se în cele 2 catene componente. La finalul acestei etape tot ADN-ul din amestec este mono-catenar;
- 2. Hibridizarea** presupune incubarea amestecului de reacție la o temperatură dictată de temperatura de topire a oligonucleotidelor amorsă. În această etapă are loc legarea oligonucleotidelor amorsă de ADN-ul monocatenar obținut în etapa anterioară, pe bază de complementaritate la nivel de secvență;
- 3. Sinteza** implică incubarea amestecului din etapa anterioară la temperatura de 72°C pentru un interval de timp variabil, cuprins între 0 secunde până la 2 minute. ADN-polimeraza devine activă și adaugă nucleotide la capătul 3'-OH al oligonucleotidelor amorsă legate, extinzându-le. Nucleotidele sunt adăugate în ordinea dictată de complementaritatea cu ADN-ul matriță. În acest mod are loc *replicarea sau sinteza* unei noi molecule de ADN dublucatenar. Una dintre catenele acestei molecule este catena originală, matriță, iar a doua este catena nou sintetizată, care încorporează oligonucleotida amorsă (primer).

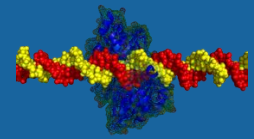




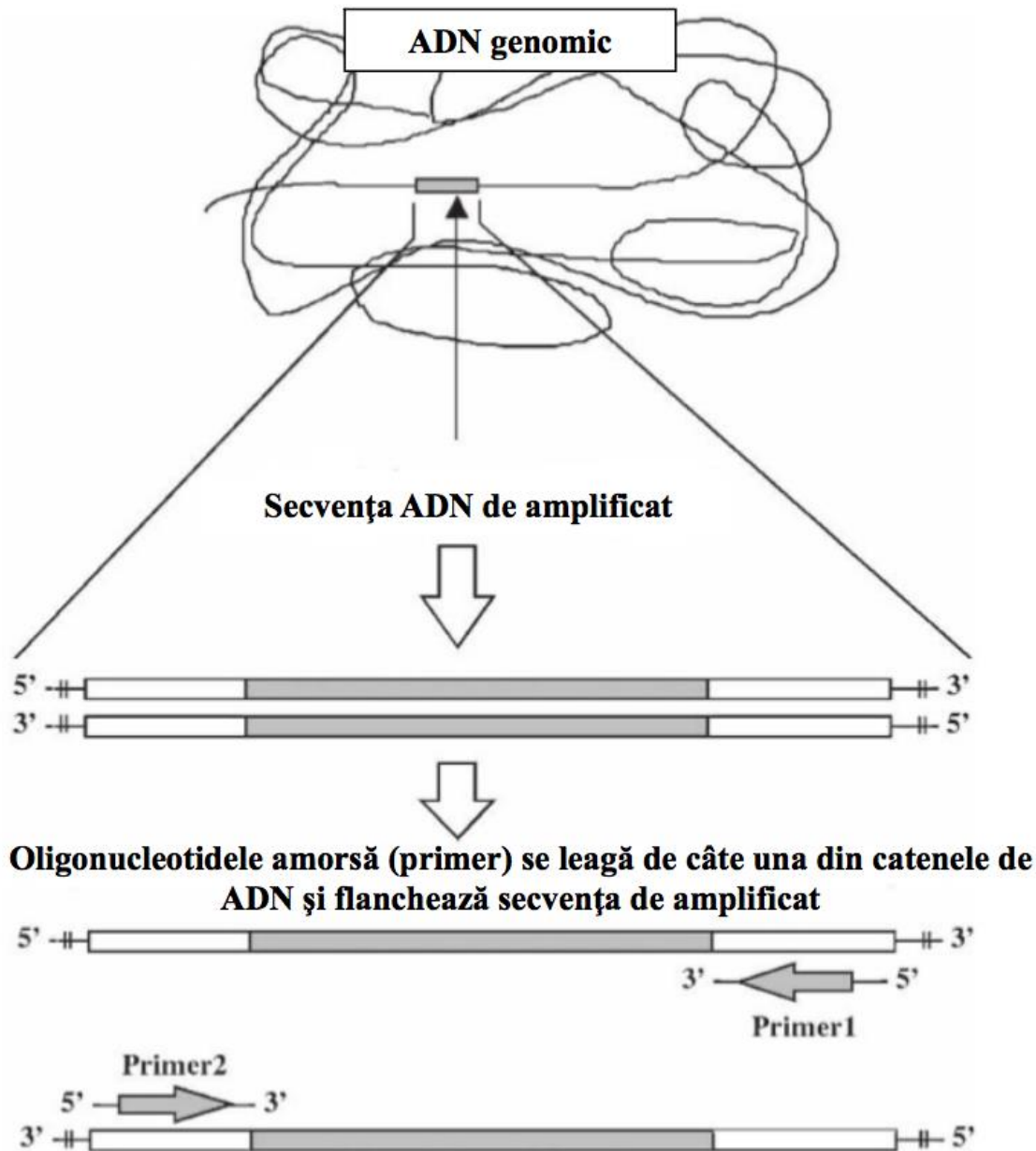
Etapele unei reacții PCR

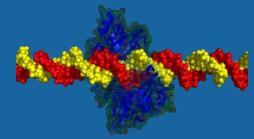
Cele trei etape alcătuiesc un **ciclu** care este repetat în mod obișnuit într-o reacție PCR de **30-35 ori**, întregul proces durând în jur de 3-4 ore. În etapa de *legare* a primului ciclu, oligonucleotidele amorsă (primer) se leagă doar de catenele separate ale ADN-ului sursă, iar catenele nou sintetizate sunt clar definite doar la capătul 5` (cel reprezentat de oligonucleotida amorsă). Poziția capătului 3` depinde de durata etapei de sinteză și poate fi variabilă. După primul ciclu PCR se obține un amestec de molecule de ADN sursă și ADN nou sintetizat de dimensiuni variabile. În ciclurile următoare sunt utilizate ca matriță și catenele nou sintetizate, ceea ce face ca pe măsură ce procesul înaintază, moleculele nou sintetizate să încorporeze treptat, la ambele capete, oligonucleotidele amorsă (primer). Faptul că la fiecare ciclu nou se utilizează ca matriță și catenele sintetizate în etapele anterioare face ca sinteza ADN-ului pornind de la noile catene să fie un proces exponențial, înlănțuit. La sfârșitul celor 30-35 cicluri, produsul rezultat în urma amplificării (numit *amplicon*) va fi format predominant din catene sintetizate în ultimele cicluri. **În cursul unei reacții de amplificare enzimatică in vitro a ADN-ului (PCR), are loc multiplicarea exponențială, înlănțuită a fragmentului de ADN cuprins între cele două oligonucleotide amorsă.**





Etapele unei reacții PCR





Etapele unei reacții PCR

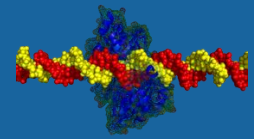
Cele 3 etape obligatorii ale reacției PCR pot fi la ora actuală complet automatizate prin intermediul **termocicloarelor**. Acestea sunt dispozitive termice cu unul sau mai multe elemente Peltier care au capacitatea de a modifica în mod controlat, cu precizie și viteză mare temperatura unui bloc de incubare în care este amplasat amestecul de reacție. În mod curent, la programarea unui termociclor, pe lângă cele trei etape obligatorii pentru realizarea amplificării *in vitro* a ADN-ului, se adaugă trei etape suplimentare care nu se repetată:

A. denaturarea inițială – constă în menținerea timp de 5 minute a amestecului PCR la temperatura de 95°C pentru denaturarea completă a ADN-ului matriță sau a celulelor, atunci când acestea sunt utilizate direct ca sursă de ADN;

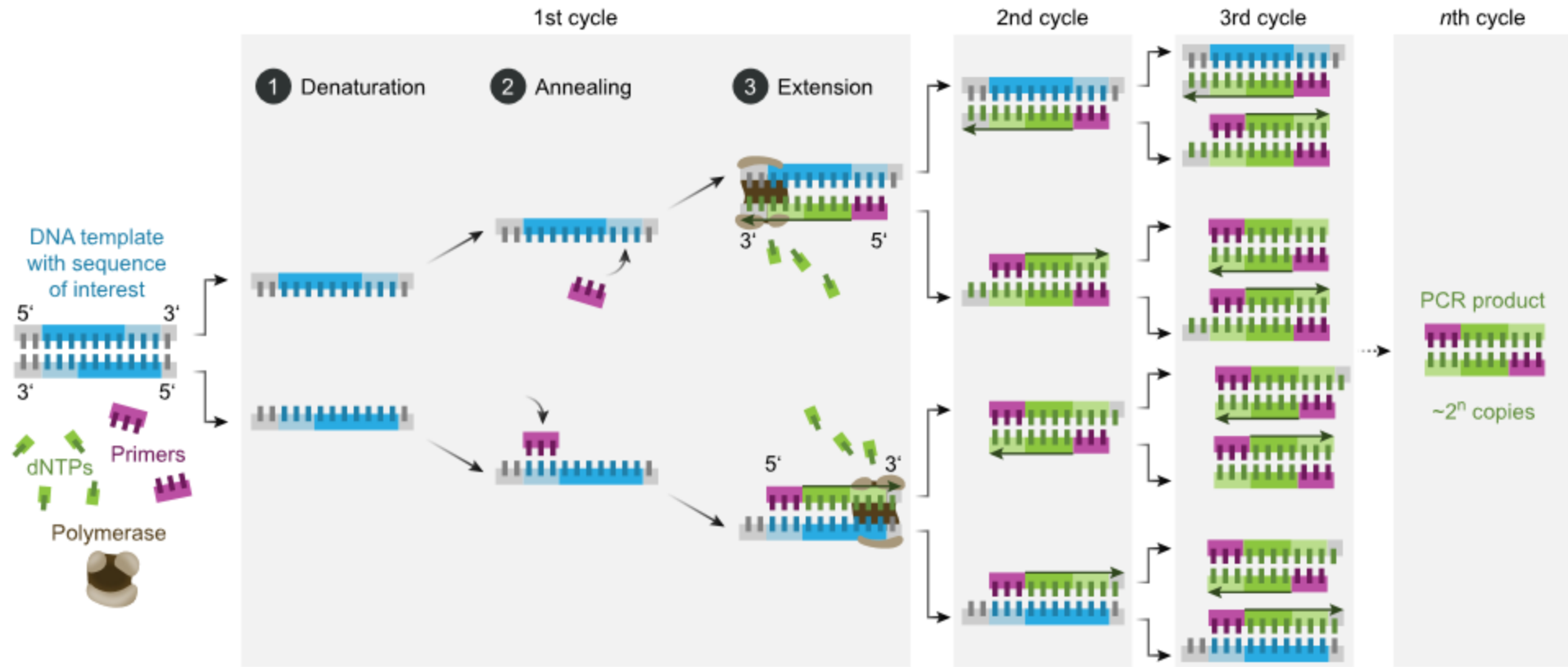
B. finalizarea sintezei – după realizarea celor 30-35 cicluri, menținerea timp de 10 minute la temperatura de 72°C are rolul de a permite finalizare sintezei tuturor catenelor la care s-a început extensia;

C. menținerea amestecului de reacție la temperatura de 4-12°C este etapa finală, în general fără limită de timp, care are rolul de a păstra probele la o temperatură optimă până la intervenția operatorului.

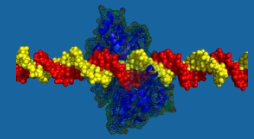




Privire generală asupra etapelor reacției PCR



Simultan cu mărirea numărului de molecule ADN disponibile dintr-o probă, prin PCR se realizează și o izolare a fragmentului de ADN cuprins între cele 2 oligonucleotide amorsă. Tehnicile de purificare ale ADN-ului total enunțate anterior permit obținerea ADN-ului genomic total (sau plasmidial, după caz) sub forma unor fragmente de dimensiuni variabile. Amplificarea enzimatică in vitro a acizilor nucleici permite izolarea unui fragment ADN specific a cărei dimensiune și secvență pot fi controlate de către cercetător (cum?)

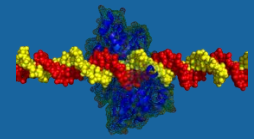


Reguli generale pentru alegerea oligonucleotidelor amorsă

Întregul proces de amplificare *in vitro* a ADN-ului se bazează pe capacitatea oligonucleotidelor amorsă de a se lega de ADN-ul matriță. Din acest motiv alegerea corectă a secvenței acestora este elementul cheie ce poate face diferența dintre un PCR reușit și un eșec total. Specificitatea și randamentul reacției PCR nu depind totuși doar de acest factor, ci și de compoziția amestecului de reacție prin conținutul de Mg^{2+} precum și de temperatura și durata fiecărei etape a ciclului de amplificare. Cunoscând secvența ADN-ului de amplificat, secvența oligonucleotidelor amorsă poate fi stabilită ținând cont de următoarele reguli:

1. **Secvența oligonucleotidelor amorsă trebuie să fie complementară și unică.** Cel mai important parametru pentru selecția secvenței celor două oligonucleotide amorsă capacitatea acestora de a se lega pe bază de complementaritate de secvență de molecula ADN de amplificat și de a forma un duplex stabil cu aceasta. Situs-urile de legare trebuie să flancheze regiunea ADN de amplificat, să fie specifice pentru fiecare oligonucleotidă și, de asemenea, **să fie unice (să nu se repete în molecula de ADN matriță).**

Complementaritatea de secvență cu molecula ADN-matriță este esențială la capătul 3'. Faptul că la capătul 5' există una până la trei baze care nu sunt complementare cu secvența de amplificat nu modifică randamentul reacției de amplificare. Acest lucru este exploatat în sensul modificării specifice a secvenței de amplificat cu ajutorul unor **oligonucleotide degenerate**. Prin introducerea unor situs-uri de restricție pentru enzime în secvența oligonucleotidelor amorsă (primer), acestea vor fi încorporate în secvența amplificată, iar ampliconul va conține deci situsurile de restricție.



Reguli generale pentru alegerea oligonucleotidelor amorsă

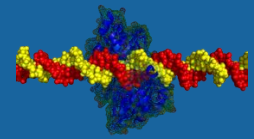
2. Să aibă o lungimea cuprinsă între 18 și 28 nucleotide. Cu cât lungimea unei oligonucleotide este mai mare, cu atât specificitatea reacției PCR este mai mare, însă legarea la ADN-ul matriță este mai puțin eficientă și în consecință cantitatea de ADN obținută în urma reacției de amplificare este mai mică;

3. Să prezinte temperatura de topire cuprinsă între 52°C și 60°C. Temperatura de topire (T_m) reprezintă temperatura la care jumătate din moleculele de oligonucleotid amorsă (primer) sunt asociate cu ADN-matriță și sunt o indicație a tăriei duplexului oligonucleotid-ADN format. T_m este un parametru care se calculează pentru fiecare oligonucleotid amorsă (primer) în parte și este influențat de secvența specifică a acestuia. Sunt descrise mai multe modalități de calcul ale valorii T_m , cea mai corectă folosind noțiunile de entalpie și entropie-funcții de stare ale unui sistem termodinamic. În mod obișnuit, o modalitate simplă care permite aproximarea rapidă a valorii T_m pentru o oligonucleotidă dată folosește formula:

$$T_m(^{\circ}\text{C}) = 4 \times (\text{GC}) + 2 \times (\text{AT})$$

unde G, C, A, T reprezintă numărul de G, C, A și respectiv T din secvența oligonucleotidei amorsă (primer).

În general, valoarea optimă pentru T_m este cuprinsă între 52°C și 60°C și trebuie să aibă valori apropiate pentru ambele oligonucleotide amorsă (primer) utilizate. Valoarea T_m este utilă pentru stabilirea temperaturii optime din etapa de hibridizare a ciclului PCR. Aceasta trebuie să fie mai mică cu câteva grade decât cea mai mică temperatură T_m a celor două oligonucleotide amorsă.

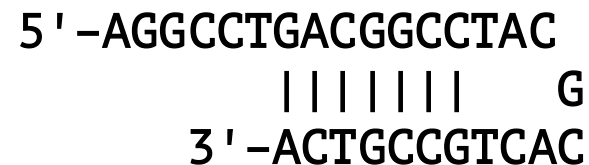


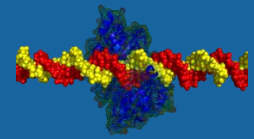
4. **Conținutul de GC** trebuie să fie între 40% și 60%.

5. **Să formeze o așa-zisă clemă GC.** Prezența de până la trei nucleotide G sau C printre ultimele cinci nucleotide de la capătul 3' duce la asocierea mai puternică a oligonucleotidului amorsă cu ADN-ul matriță, formându-se o așa-numită clemă GC, ceea ce conduce la creșterea randamentului reacției de amplificare. Mai mult de trei nucleotide GC în ultimele cinci face foarte dificilă disocierea duplexului rezultat, ceea ce scade randamentul reacției.

6. **Să nu formeze structuri secundare specifice.** Prezența în oligonucleotidele amorsă a structurilor secundare produse prin legături inter- sau intra-moleculare între nucleotide poate duce la scăderea drastică a randamentului de amplificare sau chiar la stoparea completă a reacției. Cele mai importante tipuri de structuri secundare care trebuie evitate sunt:

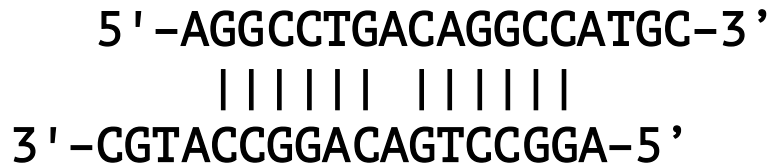
a. **ace de păr**, precum cel exemplificat mai jos;





Reguli generale pentru alegerea oligonucleotidelor amorsă

b. **auto-dimeri** – se formează prin asocierea pe bază de complementaritate a două molecule ale unui oligonucleotid de același tip;



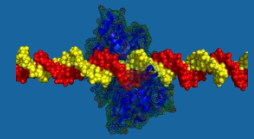
c. **hetero-dimeri** – se formează prin asocierea pe bază de complementaritate a două molecule ale celor două oligonucleotide utilizate pentru amplificare;



7. **Să nu conțină repetiții.** O repetiție este o grupare de două nucleotide care se repetă succesiv, precum **ATATATATAT**. Numărul maxim de repetiții acceptat este patru.

8. **Să nu conțină mai mult de patru repetiții ale aceleiași baze, de forma:**



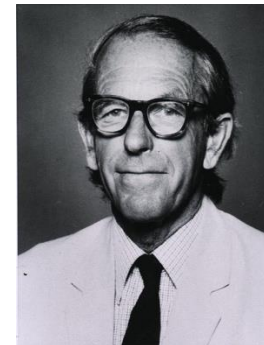


Secvențierea acizilor nucleici

Secvențierea ADN-ului - procesul de identificare a ordinii precise a nucleotidelor (ATCG) într-o moleculă de ADN – structura primară a ADN-ului.

În prezent există numeroase metode de secvențializare a ADN-ului clasificate în:

- metode derivate de la **metoda chimică** a lui **Maxam și Gilbert**;
- metode derivate de la **metoda enzimatică** a lui **Sanger**;
- **metode de "nouă generație (next-gen)"** sau mai corect **metode de secvențiere în masă (high throughput)** – **pirosecvențiere, ion-torrent, etc.**



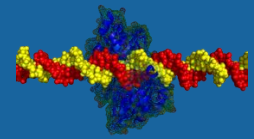
Frederick Sanger - 13 August 1918 – 19 November 2013

Metoda enzimatică Sanger

- una din primele metode de secvențializare a fost descrisă în 1977 de Frederick Sanger care a primit cel de-al doilea premiu Nobel pentru descoperirea sa;
- se bazează tot pe reacția de replicare a ADN-ului catalizată de o **polimerază ADN-dependentă**;

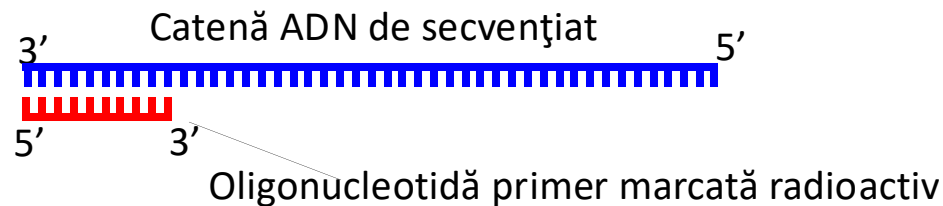
Elementele esențiale realizării unei reacții de secvențiere prin metoda Sanger

- a) **fragment ADN de secvențiat** – o moleculă de ADN monocatenară;
- b) **polimeraza ADN**;
- c) cele **4 deoxinucleotide** obișnuite (dNTP: A, T, C, G);
- c) o **oligonucleotide amorsă** (primer) complementare cu molecula de secvențiat marcată radioactiv și care oferă un capăt 3' liber;
- d) **4 di-deoxinucleotide** (ddNTPs: ddA, ddC, ddG, ddT) – nucleotide ce nu au gruparea hidroxil în pozițiile **2'** (**deoxi**) și **3'** (**di**).



Etapele reacției de secvențiere

1) Legarea oligonucleotidului primer



2) **Elongarea** primerului prin acțiunea ADN polimerazei în prezența dNTP și ddNTP cu sinteza unei catene noi. Reacția se realizează în **4 eprubete diferite**, fiecare conținând un singur tip de ddNTP (ddATP, ddGTP, ddCTP sau ddTTP). **Încorporarea unui ddNTP în noua catenă face ca sinteza acesteia să se oprească** (ddNTP sunt "terminatori" de catenă). Încorporarea unui ddNTP de către ADN-polimeraza este un proces statistic ceea ce face ca prin elongare să se genereze în fiecare eprubetă câte o colecție de catene ADN nou sintetizate de dimensiuni diferite.

ddA +
A; T; C; G;
#TCGACGGGC
Amorsa #

#ddA

#AGCTGCCCCG

ddC +
A; T; C; G
#TCGACGGGC
Amorsa #

#AGddC

#AGCTGddC

#AGCTGCddC

ddG +
A; T; C; G
#TCGACGGGC
Amorsa #

#AddG

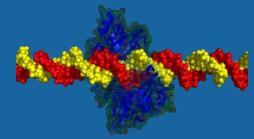
#AGCTddG

#AGCTGCCddG

ddT +
A; T; C; G
Amorsa #

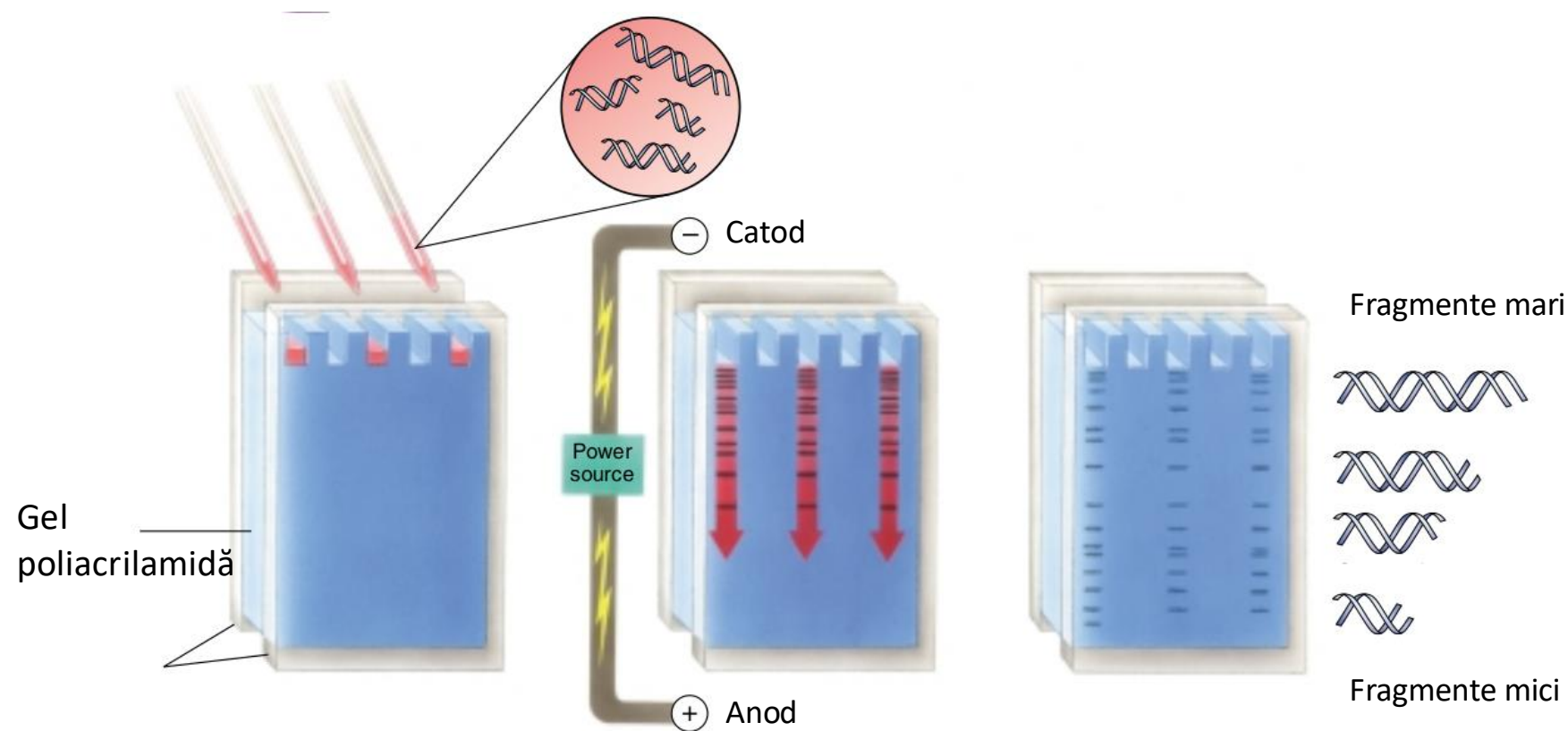
#AGCddT

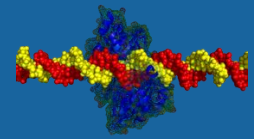
#AGCTGCCCCG



Etapele reacției de secvențiere

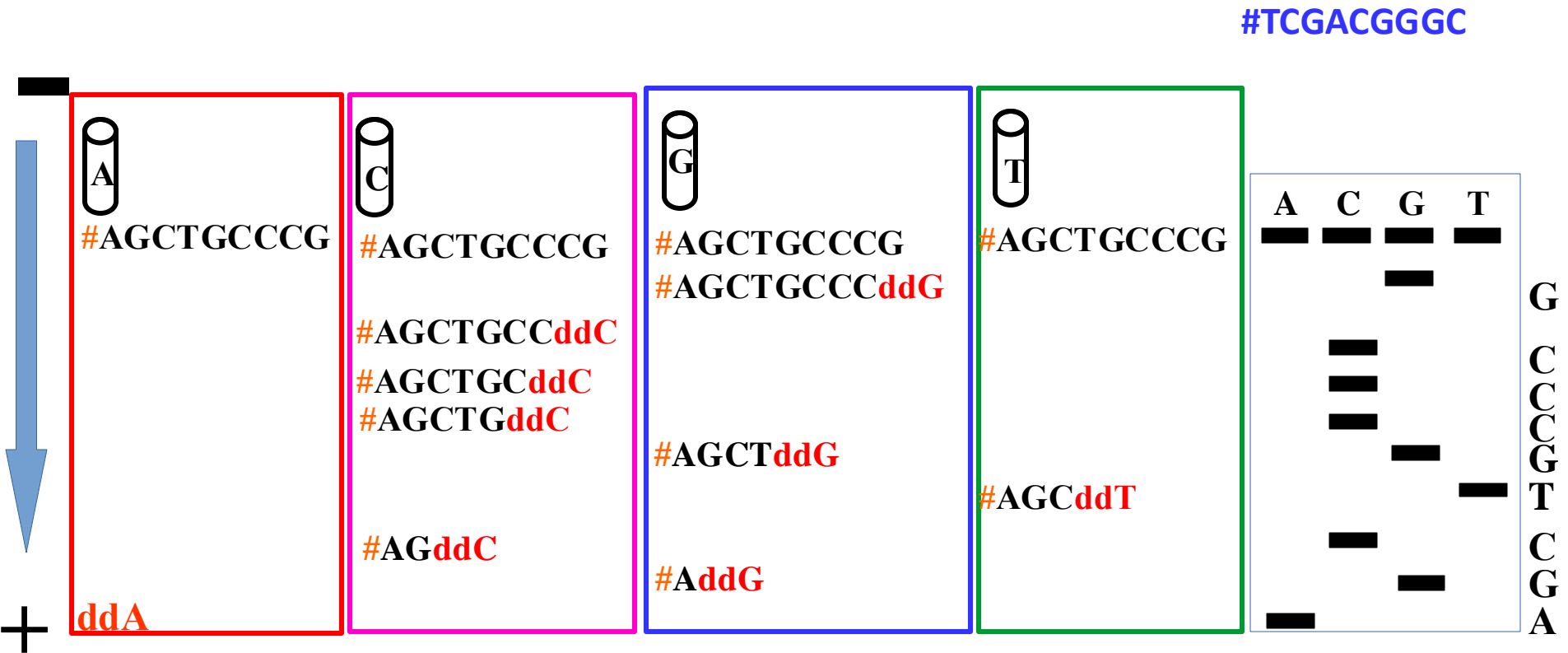
3) Separarea catenelor nou sintetizate funcție de dimensiune – se realizează prin **electroforeză** – moleculele de ADN sunt încărcate negativ și la aplicarea unui curent vor migra spre polul pozitiv. Electroforeza se realizează în geluri de poliacrilamidă ce acționează ca o sită, frânând moleculele mari, astfel încât moleculele mici vor migra mai mult iar cele mari mai puțin.



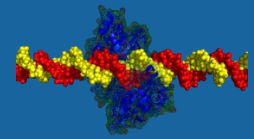


Etapele reacției de secvențiere

4) **Vizualizarea catenelor nou sintetizate** – se realizează prin **autoradiografie** – pe baza marcajului radioactiv al oligonucleotidului amorsă.



Schema autoradiografiei unui gel de secvențiere



Etapele reacției de secvențiere



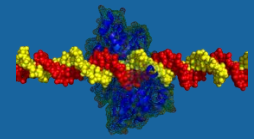
Care este secvența fragmentului de ADN prin a cărui secvențiere Sanger se obține gelul de alături?

Secvența citită de pe gel:

3' -GCAGAAATAAGTAC-5' deci secvența catenei de secvențiat era:

3' -CGTCTTTATTCATG-5' ,

respectiv 5' GTACTTATTTCTGC-3'



Etapele reacției de secvențiere

O variantă a metodei Sanger este:

- **utilizarea de "terminatori" de catenă marcați fluorescenți.** Fiecare din cele 4 ddNTP sunt marcate cu câte un fluorocrom diferit (**ddA – verde, ddT – roșu, ddG – galben, ddC – albastru**). Reacția se desfășoară într-o singură eprubetă, iar fragmentele generate sunt separate prin electroforeză. Ordinea nucleotidelor este citită cu ajutorul unui laser și reprezentată ca "peak-uri".

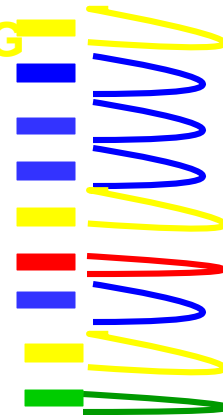


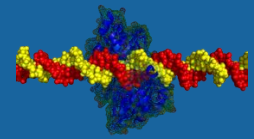
#ddA
#AddG
#AGddC
#AGCddT
#AGCTddG
#AGCTGddC
#AGCTGCddC
#AGCTGCCddC
#AGCTGCCddG
#AGCTGCCG

Separare
funcție de
dimensiune



#AGCTGCCddG
#AGCTGCCddC
#AGCTGCddC
#AGCTGddC
#AGCTddG
#AGCddT
#AGddC
#AddG
#ddA





Etapele reacției de secvențiere

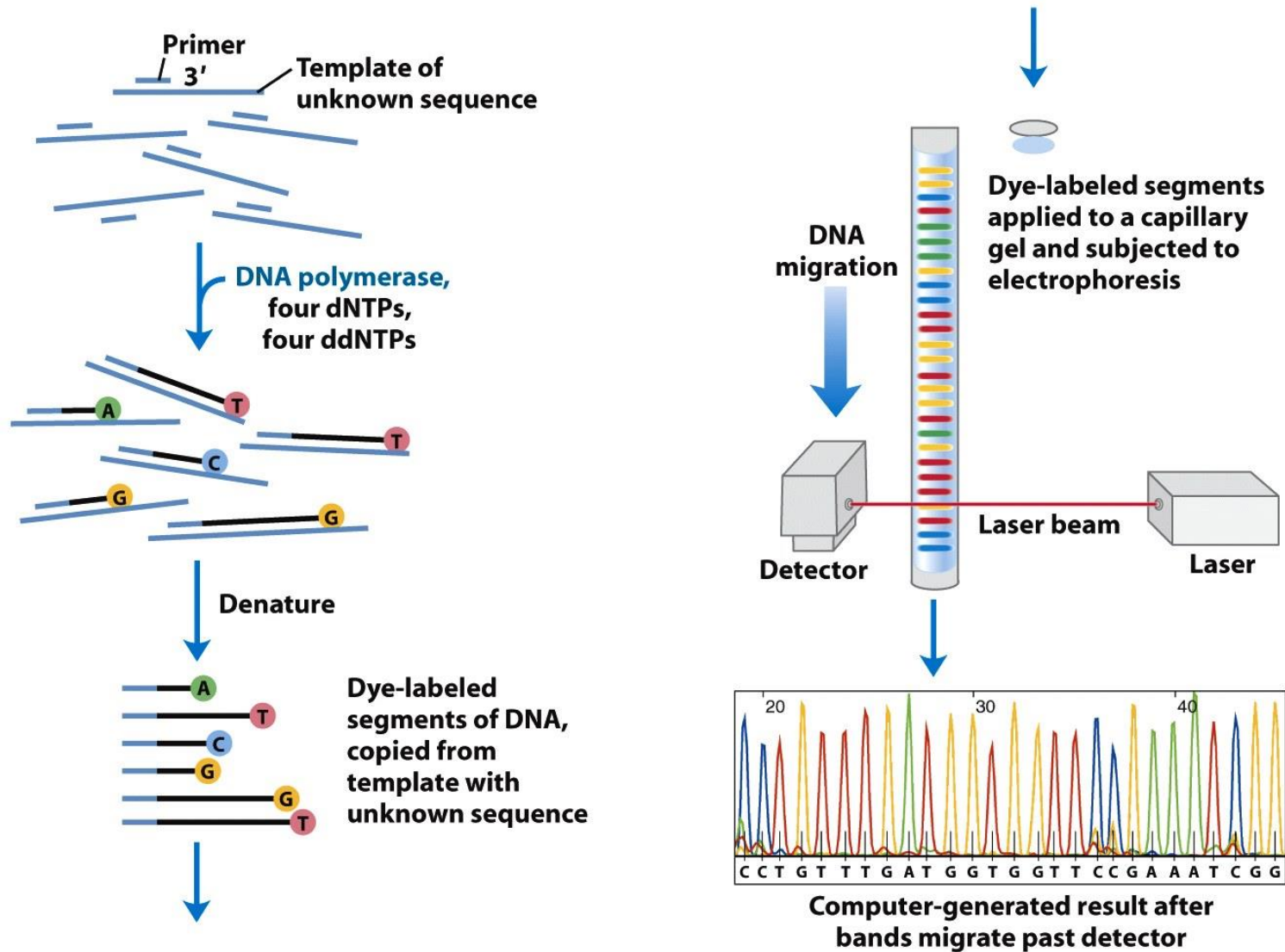
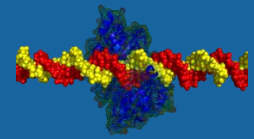


Figure 8-34
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. H. Freeman and Company

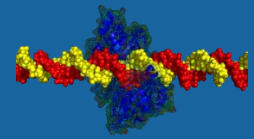


Metode de secvențiere “de nouă generație”

Nevoia de a secvenția cantități mari de ADN (**genomul uman - 6.27 Gpb, un genom bacterian 4-5 Mpb**) rapid și cu costuri reduse a dus la apariția unor noi metode de secvențiere, ce funcționează după principii diferite de metoda Sanger. Aceste metode au fost numite general **next generation sequencing / Nextgen / metode NGS / metode de nouă generație**. Primele metode Nextgen au apărut în anii 80. De atunci s-au dezvoltat succesiv două generații de metode de secvențiere și este de așteptat ca noi metode să fie dezvoltate în viitor. De aceea termenul **Nextgen** a devenit ambiguu și se preferă înlocuirea sa cu:

1. Metode de secvenție de generația a II / secvențiere masivă paralelă / metode de secvențiere a fragmentelor mici: **secvențierea prin hibridizare și secvențierea prin sinteză** (SBS – **sequencing by synthesis**)
2. Metode de secvenție de generația a III / metode de secvențiere a fragmentelor lungi: **SMRT** (Single Molecule Real Time) și secvențiere folosind nanopori proteici **Oxford Nanopore Technologies**

Toate aceste metode permit generarea unui număr semnificativ mai mare de nucleotide secvențiate comparativ cu metoda Sanger, de aceea ele mai sunt denumite și metode **metode extensive (High-throughput)**



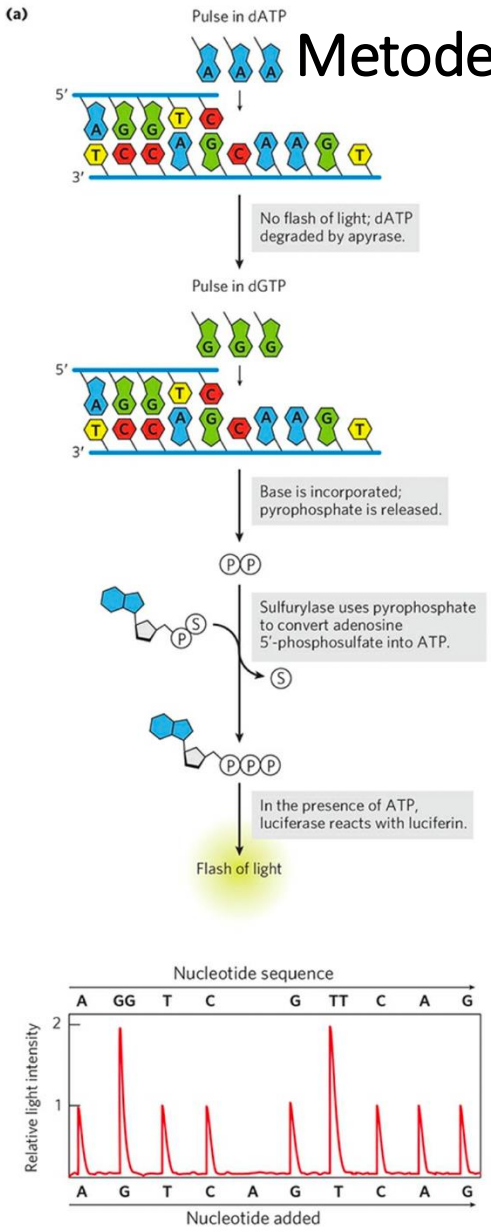
Metode de secvenție de generația a II

Secvențierea prin hibridizare - a fost dezvoltată în anii 80 și și-a găsit aplicații mai ales în diagnostic, însă impactul acestei tehnologii în prezent este redus

Secvențierea prin sinteză - SBS – moleculele de ADN sunt fragmentate enzimatic sau mecanic în bucăți cu lungimea de câteva sute de nucleotide. Aceste fragmente sunt apoi immobilizate pe suporturi solide și fiecare fragment este amplificat individual prin PCR pentru a forma grupe (cluster) de secvențe identice, copii ale fragmentului de ADN matriță. Suprafața solidă este parte a unui canal prin care circulă reactivii necesari procesului de secvențiere. Pe suprafața solidă de câțiva cm lățime sunt atașate astfel milioane de cluster de fragmente de ADN, fiecare cluster conținând copii ale aceleiași molecule. Clusterelor vor fi secvențiate simultan, fiecare cluster generând date de secvență ale fragmentului original. Funcție de modul cum se detectează secvența din aceste cluster, tehnologiile SBS se clasifică în:

A. Pirosecvențiere sau secvențierea **454** - soluții cu câte una din cele 4 dNTP spală suprafața solidă pe care sunt fixate clusterelor de fragmente alternativ. Soluția cu un dNTP stă în contact cu suprafața solidă suficient cât aceasta să fie incorporată în catene de ADN complementare cu cele din cluster de către o ADN polimerază. Excesul de nucleotide este apoi distrus de o apirază înainte de a veni pulsul cu următoarea nucleotidă. Dacă o nucleotidă este încorporată în unul din clusteri, încorporarea sa va fi însoțită de eliberarea unui rest de pirofosfat. Enzima sulfurilaza folosește acest pirofosfat pentru a transforma adenzine 5'- fosfosulfatul din mediu în ATP, iar ATP-ul este utilizat de enzima luciferază pentru a produce lumină. Adăția unei nucleotide în unul din cluster va fi deci însoțită de eliberarea de lumină. Cunoscând ce dNTP scâlde suprafața solidă la momentul eliberării de lumină, se identifică ce nucleotidă a fost încorporată.

Metode de secvenție de genera



„.....when dCTP is added to the solution, flashes occur only at clusters where G is the next base in the template and C is the next nucleotide to be added to the growing DNA chain. If there is a string of two, three, or four G residues in the template, a similar number of C residues are added to the growing strand in one cycle. This is recorded as a “flash” amplitude at that cluster that is two, three, or four times greater than when only one C residue is added. Similarly, when dGTP is added, flashes occur at a different set of clusters, marking those as clusters where G is the next nucleotide added to the sequence (where C is present in the template). The length of DNA that can be reliably sequenced in a single cluster by this method—often referred to as the read length, or “read”—is typically 400 to 500 nucleotides, and is rapidly increasing.”

Lehninger 7th edition, pg 307

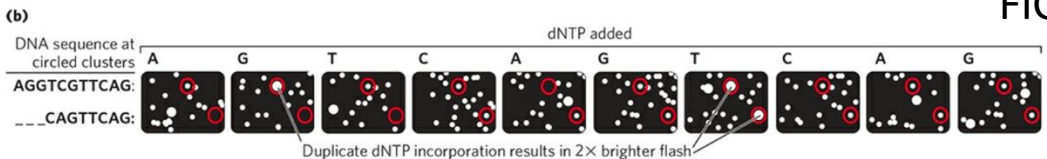
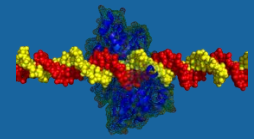


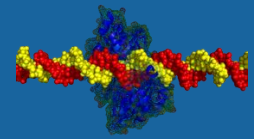
FIGURE 8-36 pg 307, Lehninger 7th edition



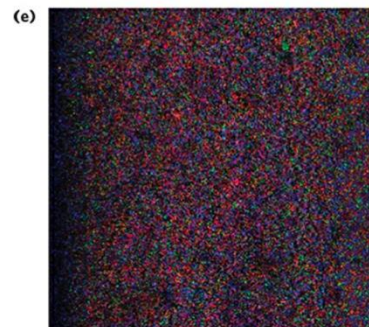
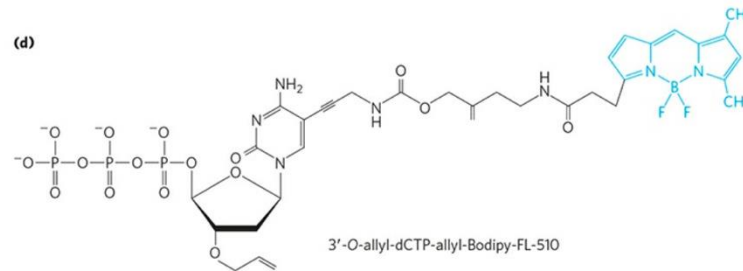
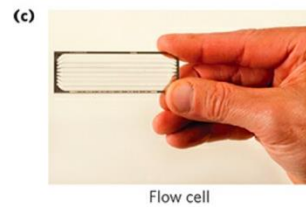
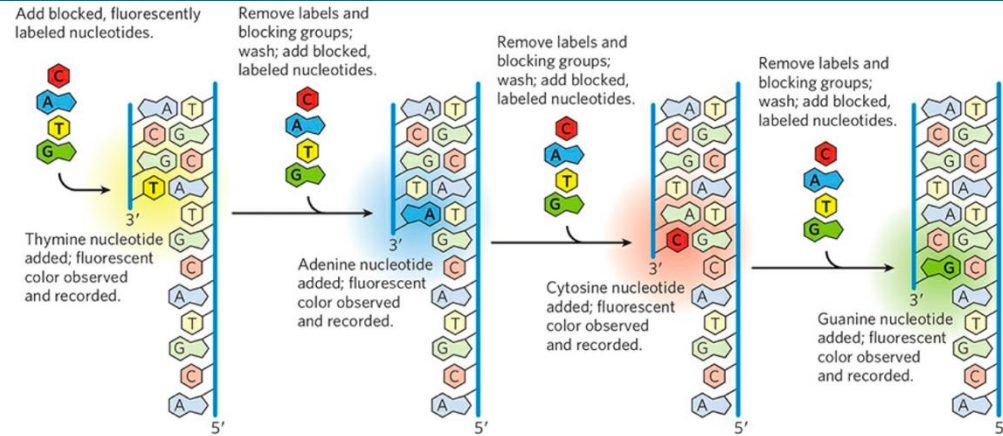
Metode de secvenție de generația a II

B. Tehnologia Ion Torrent - tehnologia are la bază faptul că în replicare a ADN-ului, adăția unei nucleotide pe catena nascentă este însoțită de eliberarea unui proton, ceea ce duce la o modificare a pH-ului soluției în care se produce sinteza. Această modificare de pH poate fi înregistrată și corelată cu nucleotida atașată. Fragmentele de ADN de secvențiat sunt atașate pe bile de dimensiuni mici și sunt amplificate pentru a genera mii de copii ale aceleiași molecule, fiecare bilă având un set de copii ale aceleiași molecule de ADN de câteva sute de nucleotide. Fiecare bilă este amplasată într-un godeu și spălată succesiv cu soluții ce conțin câte una din cele 4 dNTP. Dacă la o spălare se încorporează o nucleotidă, se va înregistra o modificare de pH în godeul corespunzător. Dacă nucleotida din soluția de spălare nu se încorporează, nu se va modifica pH-ul și concluzia este că fragmentul nascent nu conține acea nucleotidă.

C. Tehnologia Illumina sau secvențierea cu terminatori reversibili – tehnologia se bazează pe un set de cele 4 deoxinucleotide modificate: fiecare nucleotidă (A, T, G, C) conține un marker fluorescent specific (fluorofor) și la capătul 3' un terminator de sinteză detașabil. Amestecul de 4 nucleotide scaldă suportul solid conținând milioane de clustere de fragmente identice iar polimeraza încorporează nucleotidele marcate cu terminatorul atașat. În fiecare cluster se atașează o nucleotidă specifică, complementară cu catena matriță. Un laser excită fiecare cluster care emite lumină de culoare diferită, funcție de nucleotida ce a fost încorporată. Marcherii fluorescenți și terminatorii de sinteză din poziția 3' sunt apoi eliminați printr-o reacție chimică și se adaugă din nou amestecul de nucleotide peste clustere, ceea ce duce la atașarea unei noi nucleotide și emiterea de lumină funcție de nucleotida atașată.

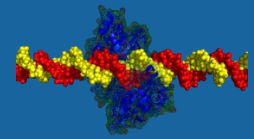


Metode de secvenție de generația a II



Lehninger 7th edition, pg 308

FIGURE 8-37 Next-generation reversible terminator sequencing. (a) The



Metode de secvenție de generația a III

Spre deosebire de metodele de generația a II, cele de generația III își propun să secvențieze moleculele de acizi nucleici fără a le mai fragmenta și fără a genera clusteri de secvențe identice. Momentan, 2 tehnologii distincte reușesc să secvențieze fragmente de ADN mai lungi de 30-50 de kpb:

A. Tehnologia SMRT - Singe Molecule Real Time – o moleculă de ADN de secvențiat este atașată împreună cu o ADN-polimerază de fundul unui godeu de dimensiuni foarte mici ce poartă numele de ZMW (**zero-mode waveguide**). Prin dimensiuni și formă, această cameră ghidează lumina spre o zonă foarte redusă spre fundul godeului. Peste ADN-polimerază și ADN-ul de secvențiat se adaugă o soluție cu cele 4 nucleotide de care sunt atașate prin legături fosforice fluorofori. ADN-polimeraza începe sinteza unei noi catene și încorporează un nucleotid și pe fundul ZMW se formează o imagine corespunzător culorii emise de fluorofor. Prin încorporare, fluoroforul este eliberat și difuzează spre partea superioară a ZMW, ceea ce face ca semnalul luminos să dispară. O nouă nucleotidă este atașată pe fundul ZMW pe bază de complementaritate și apare deci o nouă culoare corespunzătoare fluoroforului atasat de acea nucleotidă.

Real-Time DNA Sequencing from Single Polymerase Molecules

John Eid,¹ Adrian Fehr,² Jeremy Gray,² Khai Luong,² John Lytle,² Geoff Otto,² Paul Peluso,² David Rank,² Primo Baybayan, Brad Bettman, Arkadiusz Bibillo, Keith Bjornson, Bidhan Chaudhuri, Frederick Christians, Ronald Cicero, Sonya Clark, Ravindra Dalal, Alex deWinter, John Dixon, Mathieu Foquet, Alfred Gaertner, Paul Hardenbol, Cheryl Heiner, Kevin Hester, David Holden, Gregory Kearns, Xiangui Kong, Ronald Kuse, Yves Lacroix, Steven Liu, Paul Lundquist, Congcong Ma, Patrick Marks, Mark Mazham, Devon Murphy, Insil Park, Thang Pham, Michael Phillips, Joy Roy, Robert Sebra, Gene Shen, Jon Sorenson, Austin Tomoney, Kevin Travers, Mark Trulson, John Vieceli, Jeffrey Wegener, Dawn Wu, Alicia Yang, Denis Zaccarin, Peter Zhao, Frank Zhong, Jonas Korlach,¹ Stephen Turner[†]

We present single-molecule, real-time sequencing data obtained from a DNA polymerase performing uninterrupted template-directed synthesis using four distinguishable fluorescently labeled deoxyribonucleoside triphosphates (dNTPs). We detected the temporal order of their enzymatic incorporation into a growing DNA strand with zero-mode waveguide nanostructure arrays, which provide optical observation volume confinement and enable parallel, simultaneous detection of thousands of single-molecule sequencing reactions. Conjugation of fluorophores to the terminal phosphate moiety of the dNTPs allows continuous observation of DNA synthesis over thousands of bases without steric hindrance. The data report directly on polymerase dynamics, revealing distinct polymerization states and pause sites corresponding to DNA secondary structure. Sequence data were aligned with the known reference sequence to assay biophysical parameters of polymerization for each template position. Consensus sequences were generated from the single-molecule reads at 15-fold coverage, showing a median accuracy of 99.3%, with no systematic error beyond fluorophore-dependent error rates.

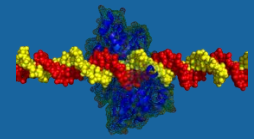
The Sanger method for DNA sequencing (*1*) uses DNA polymerase to incorporate the 3'-dideoxynucleotide that terminates the synthesis of a DNA copy. This method relies

[†]These authors contributed equally to this work.
[‡]To whom correspondence should be addressed. E-mail: jkorlach@pacificbiosciences.com (J.K.); sturne@pacificbiosciences.com (S.T.)

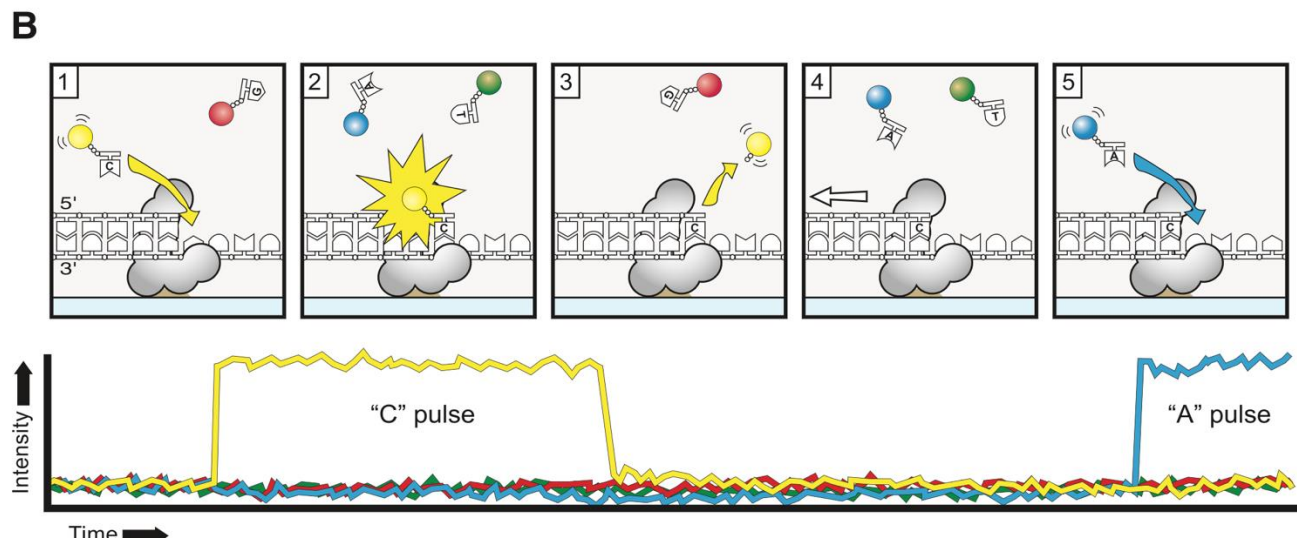
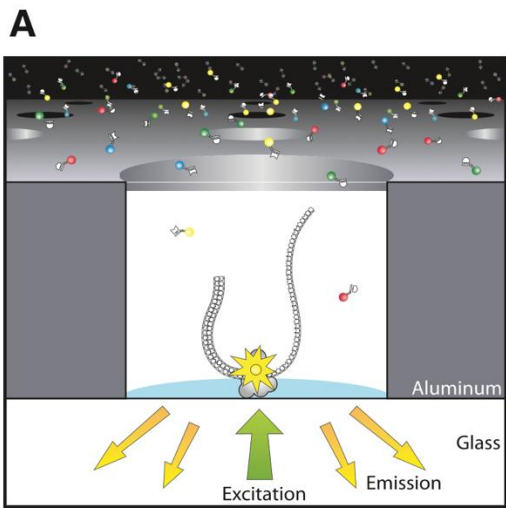
on the low error rate of DNA polymerases, but exploits neither their potential for high catalytic rates nor high processivity (*2–4*). Increasing the speed and length of individual sequencing reads beyond the current Sanger technology limit will shorten cycle times, accelerate sequence assembly, reduce cost, enable accurate sequencing analysis of repeat-rich areas of the genome, and reveal large-scale genomic complexity (*5, 6*). Alternative approaches that increase sequencing performance

have been reported [(*7–10*), reviewed in (*11, 12*)]. Several of these methods have been deployed as commercial sequencing systems (*13–16*), which have greatly increased overall throughput, enabling many applications that were previously unfeasible. However, because these methods all gate enzymatic activity, using various termination approaches, they have not yielded longer sequence reads (limited to ~400 nucleotides), nor do they exploit the high intrinsic rates of polymerase-catalyzed DNA synthesis.

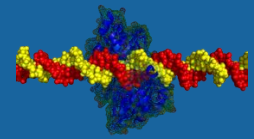
The use of DNA polymerase as a real-time sequencing engine—that is, direct observation of processive DNA polymerization with base-pair resolution—has long been proposed but has been difficult to realize (*7, 8, 17–22*). To fully harness the intrinsic speed, fidelity, and processivity of these enzymes, several technical challenges must be met simultaneously. First, the speed at which each polymerase synthesizes DNA exhibits stochastic fluctuation, so polymerase molecules would need to be observed individually while they undergo template-directed synthesis. Because of the high nucleotide concentrations required by DNA polymerases (*20*), a reduction in the observation volume beyond what is afforded by conventional methods, such as confocal or total internal reflection microscopy, directly improves single-molecule detection. Second, deoxyribonucleoside triphosphate (dNTP) substrates must carry detection labels that do not inhibit DNA polymerization even when 100% of the native nucleotides are replaced with their labeled counterparts. Third, a surface chemistry is required that retains activity of DNA polymerase molecules and inhibits nonspecific adsorption of labeled dNTPs. Finally, an instrument is required that can faithfully detect and distinguish incorporation of four different labeled dNTPs. Here, we provide proof-of-concept for an approach to highly



Metode de secvenție de generația a III

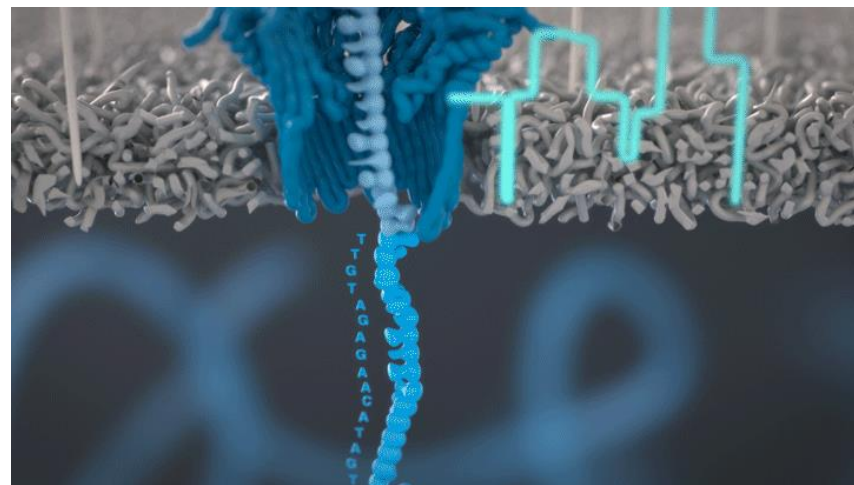
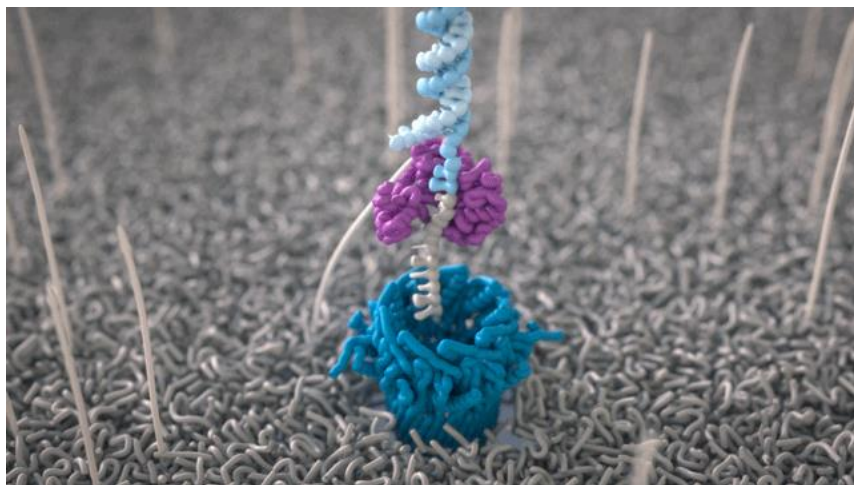


[DOI: 10.1126/science.1162986](https://doi.org/10.1126/science.1162986)

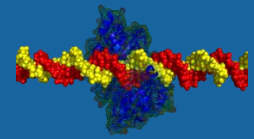


Metode de secvenție de generația a III

B. secvențiere folosind nanopori proteici – folosește proteine transmembranare ce pot fi traversate de molecule de acizi nucleici precum alfa-hemolisină și porina A din *Mycobacterium smegmatis* (MspA). Moleculele de ADN trec prin por cu o viteză de constantă controlată de ADN polimerază phi29 și duc la modificarea diametrului interior al porului. Modificările sunt dependente de forma nucleotidei ce trece și sunt înregistrate sub forma unor curenți transmembranari. Funcție de caracteristicile curentului generat de trecerea unei molecule de ADN, se poate detecta secvența acesteia.



<https://nanoporetech.com/how-it-works>



Fișiere cu secvențe

O **secvență**, fie că este ADN, ARN sau proteine **reprezintă o înșiruire de litere** (A; T; G; C pentru ADN, A; U; G; C pentru ARN, cei 20/22 de aminoacizi notați cu o literă pentru proteine) **ce poate fi foarte ușor stocată într-un fișier text**.

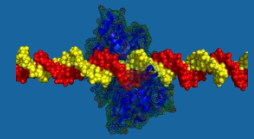
În general o secvență este însoțită de o serie **de informații accesorii precum specia de la care provine secvența, gena, cromozomul, lucrarea în care este descrisă secvența**. **Modul în care sunt organizate și stocate aceste informații alături de secvența propriu-zisă reprezintă formatul sau standardul unui fișier cu secvențe**.

De-a lungul timpului au existat un număr mare de variante de fișiere/formate create pentru a înregistra secvențe, însă fișierul/formatul FASTA este cel ce s-a impus.

Fișierul FASTA – Fast-All

```
>AJ507836.1 Arthrobacter nicotinovorans pA01 megaplasmid sequence, strain ATCC 49919
GATCCGGCGGTCCGCCGTCGTGGCGGCGGGCGGAGGTTGCCGGCGGCGGACCGCCGCCCGCACAAAGAA
GGCCTTCGGGTTCCGGCAGGTGGCGCGGCCCGGACACTGGTCTTCGCTGGGGTGGAAACGTGGGGTGCT
GGGTGTGGGCGGTTCCCGGGCGAACC GGAAAGGCGTCCACTCCTCTTCGCTTCCGTGGCCGGCGGTTG
GGGCCGTTCCCGTTCGTCGCAGAGCTCCGCCGTGCCCGGCCCGCCCGTTGTCTACGACGTCTTTTTGTGG
CTGTCCTTCGGATCATAACGGTTCGGCTCCAGCTCAAGCTCCTGGCAGTCCGCAAAGCTCCGGTCTAGCAC
ACAGCGATGCGGGTAATGATGGCGACGAATTCGTCCAGAGCGCTGGTGGGATTTCGTGGCCCATACCTG
```

```
>gi|5524211|gb|AAD44166.1| cytochrome b [Elephas maximus maximus]
LCLYTHIGRNIYYGSYLYSETWNTGIMLLLITMATAFMGYVLPWGQMSFWGATVITNLFSAIPYIGTNLV
EWIWGGFVSDKATLNRFFAFHFILPFTMVALAGVHLTFLHETGSNNPLGLTSDSDKIPFHPYYTIKDFLG
LLILLLLLLLLLLALLSPDMLGDPDNHMPADPLNTPHMKPEWYFLFAYAILRSVPNKLGGVLALFLSIVIL
GLMPFLHTSKHRSMMLRPLSQALFWTLTMDLLTLTWIGSQPVEYPYTIIGQMASILYFSIILAFPLIAGX
IENY
```



Fișiere FASTA - fișiere cu secvențe

Fișierul FASTA

- un fișier text ce poate **avea sau nu** extensia .fasta;
- conține secvențe de nucleotide sau de aminoacizi stocate conform **standardului FASTA**:

Elementele formatului FASTA:

A. primul rând de text, marcat cu “>” (mai mare) conține o serie de **informații cu caracter opțional**, - specia sau denumirea genei (proteinei);

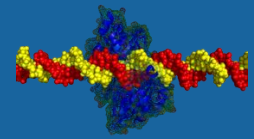
B. următoarele rânduri conțin secvența propriu-zisă, în care nucleotidele/aminoacizii sunt reprezentați folosind codul standard IUPAC cu o singură literă;

C. fiecare rând al secvenței are în general 80 de caractere (nu mai mult de 120)

A	adenozină	M	A/C (amino)
T	timină	W	A/T (weak)
C	citidină	R	G/A (puRine)
U	uracil	B	G/T/C
G	guanină	D	G/A/T
N	A/G/C/T (oricare)	H	A/C/T
K	G/T (keto)	V	G/C/A
S	G/C (strong)	-	spatiu de dimensiune intermediară
Y	T/C (pYrimidine)		

A	alanină	P	prolină
B	aspartat/asparagină	Q	glutamină
C	cistină	R	arginină
D	aspartat	S	serină
E	glutamat	T	threonină
F	fenilalanină	U	selenocisteină
G	glicină	V	valină
H	histidină	W	triptofan
I	isoleucină	Y	tirosină
K	lisină	Z	glutamat/glutamin
L	leucină	X	oricare
M	metionină	*	stop
N	asparagină	-	spațiu

Standardul IUPAC pentru notarea nucleotidelor și aminoacizilor



Fișiere FASTA - fișiere cu secvențe

Fișierul FASTA

D. O secvență de nucleotide este notată în direcția 5' → 3' : prima literă de pe primul rând corespunzător secvenței este nucleotida 1 și are o grupare PO₄ liberă, ultima literă de pe ultimul rând are gruparea OH 3' liberă. Moleculele circulare se reprezintă linear, prima nucleotidă fiind cel mai frecvent originea de replicare;

E. O secvență polipeptidică este notată în sensul sintezei, de la capătul N terminal spre cel C terminal; prima literă din primul rând reprezintă aminoacidul 1 din secvență - aminoacidul N-terminal; ultima literă reprezintă aminoacidul C-terminal;

F. Poziția literelor poate fi sau nu numerotată; în cazul în care literele din secvență sunt numerotate, numărarea se face la începutul fiecărui rând și se include un spațiu după fiecare a 10-a literă.

G. Literele ce desemnează secvența pot fi sau nu scrise cu majuscule; indiferent de tipul de scriere, semnificația este aceeași;

H. Unele programe nu acceptă caracterul '-' (spațiul în secvență) se indică cu un sir de N pentru nucleotide sau X pentru aminoacizi; spațiul între litere este ignorat;

```
>secventa peptidica necunoscuta 1 fara numere
```

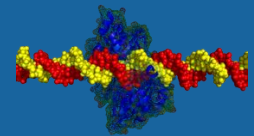
```
QIKDLLVSSSTDLDTTLVLVNAIYFKGMWKTAFAEADTREMPFHVTKQESKPVQMMCMNNSFNVATLPAE  
KMKILELPPFASGDLMLVLLPDEVSDLERIEKTINFEKLTWETNPNTMEKRRVKVYLPQMKIEEKYNLTS  
VLMALGMTDLFIPSANLTGISSAESLKISQAVHGAFMELSEDGIEMAGSTGVIEDIKHSPESEQFRADHP  
FLFLIKHNPTNTIVYFGRYWSP
```

```
>secventa peptidica necunoscuta cu numere
```

```
1  qikdllvsss tlddttlv lv naiyfk gmwk tafnae dre mpfhv tkqes kpvqmm cmnn  
61  sfnvatlp ae kmkilelp fa sgdlsmlv ll pdevsdleri ektinfeklt ewtnpntmek  
121 rrvkvylp qm kieekynlts vlmalgmt dl fipsanltgi ssaeslkisq avhgafmels  
181 edgiemagst gviedikhsp eseqfradh p flfli khnpt ntivyfgryw sp
```

Cum scriu secvențele?
- Cu font **monospațiat**

TNR **Proportional**
Courier New **Monospace**



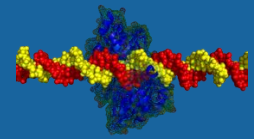
De unde pot obține secvențe?

1. Un instrument de secvențiere a ADN-ului – cel mai frecvent;
2. Un instrument de secvențiere a proteinelor bazat pe degradarea Edman (cel mai puțin frecvent) sau un spectrometru de masă – destul de rar;
3. **O bază de date cu secvențe** – cel mai ușor de accesat; create de utilizatori ce au acces la instrumentele 1 și 2 și deci pot determina experimental o secvență;

O bază de date cu secvențe reprezintă **o colecție de secvențe de acizi nucleici sau aminoacizi ce au fost stabilite experimental și care au fost depozitate în formă digitalizată pe un server central într-un format tip specific**. Fiecărei secvențe i se alocă un **identificator unic – ID** – o **combinație unică de litere și cifre ce poate fi folosită pentru a regăsi fără echivoc secvența în respectiva bază de date**.

În general, accesul la bazele de date cu secvențe este gratuit.

Bază de date	Site web	Dimensiune *
Baze de date cu secvențe		
INSDC	http://www.insdc.org/	206 293 625 secvențe
DDBJ	http://www.ddbj.nig.ac.jp	Aceste baze de date sunt sincronizate zilnic între ele și conțin aceleași secvențe.
EMBL Nucleotide Sequence Database	http://www.ebi.ac.uk/embl/	
GenBank	http://www.ncbi.nlm.nih.gov/genbank/	
European Nucleotide Archive (ENA)	http://www.ebi.ac.uk/ena/	
KEGG	http://www.genome.jp/kegg/	25 679 056 secvențe
nr	http://blast.ncbi.nlm.nih.gov	145 296 712 secvențe
UniProtKB/Swiss-Prot	http://www.uniprot.org/	556 568 secvențe
UniProtKB/TrEMBL	http://www.uniprot.org/	107 627 435 secvențe



Principalele baze de date cu secvențe

UniProtKB/TrEMBL - <http://www.uniprot.org/>

- conține toate secvențele de nucleotide din GenBank translate în secvențe de aminoacizi;
- are un nivel foarte mare de redundanță;

UniProtKB/SwissProt - <http://www.uniprot.org/>

- toate secvențele conținute sunt adnotate manual – siguranță crescută legată de funcțiile secvențelor conținute;
- nivelul de redundanță este minim;
- adăugarea unei noi secvențe în SwissProt este un proces care implică prelucrarea manuală a datelor corespunzătoare secvenței respective => SwissProt nu conține cele mai noi secvențe.

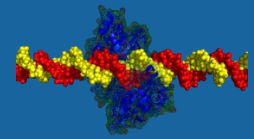
TrEMBL este completă dar incorectă, SwissProt este mai corectă, dar incompletă.

Protein Identification Resource PIR

- o colecție completă de secvențe lipsită de redundanță;
- secvențele identice sau foarte similare sunt gruparea într-o singură intrare în baza de date;

nr

- o bază de date non-redundantă completă, cuprinzând intrări din GenBank, SwissProt, PIR, Protein Research Foundation (PRF) și Protein data Bank (PDB).



Operații simple cu secvențe

<http://www.bioinformatics.org/sms2/>

https://mail.uaic.ro/~marius.mihasan/research/mirrored_sites_tools/sms2/index.html

Operații cu secvențe
selectabile în SMS2

SMS Sequence Manipulation Suite:
Translate

Translate accepts a DNA sequence and converts it into a protein in the reading frame you specify. Translate supports the entire IUPAC alphabet and several genetic codes.

Paste a raw sequence or one or more FASTA sequences into the text area below. Input limit is 200000 characters.

```
>Arthrobacter nicotinovorans orf388 sequence
ctactaacctcagcgccgcccttcaaccctttctctgaggaaaatgactaaaaca
gcgattgtccgctgcgcatgaatggcatcaccggccggatgggctaccgcagcacctg
ccgcagccggcggcttcacctcgaagacggcaccaggctccagatcgaaccgatcct
cgtaggccgaacgaagcaagatcccggaactcggcgaagcacaaggtgccgagtg
gagcacggacctggactcggctcgtcaacgacccaccgctcgacatcatcttcgacgctc
catgaccagcctccgcgccaccctgaagaaggcgtgctggccggaagcacatctt
caccgagaagccaccggaaaccctggaagaggcattgaactggccgcatcggcaa
gcaggcaggctcaccgcaggctcgtacacgacaagctgacctccgggcttggtcaa
gctccgcccctggtagcgaaggcttctcggccgcatcctgtccatcccggtgagtt
cggctactgggtcttgaaggtgagcttcaggcagcacagcgccctcctggaactaccg
caaggaagcggcggggaatgaccacggacatgtctgccactggaactacgtccttga
aggcatcatcggcaaggtcaagagcgtcaacgccaagaccgccacgacatccccaccg
ctgggacgaagccggaaggagtacaaggcaacggctgatgacgcttctacggcatctt
cgagcttgaacccccggcggcagcagctcagccagatcaactcttctgggcccgt
ccgcttaccgacgaactcgtcgaatccaggtggacggcaccacggctccgctcgt
tggcggcctgaacaagtgcgtcggcagcagcgcacacaccccccaagccggtctggaa
ccctgacctgcccgtcaccgactcctccgcaccagtgcacgaaatccccgccaccg
```

Please check the browser compatibility page before using this program.

Submit Clear Reset

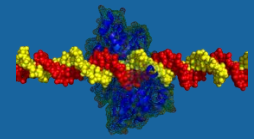
- Translate in on the strand.
- Use the genetic code.

*This page requires JavaScript. See browser compatibility.
*You can mirror this page or use it off-line.

new window | home | citation

Fri Mar 30 17:46:57 2012

Căsuță text pentru
secvență FASTA



Operații simple cu secvențe

Denumire operație*	Descriere
Convertire între diverse formate	
Combine FASTA	– permite combinarea a două sau a mai multe secvențe în format FASTA și obținerea unei singure secvențe
EMBL to FASTA	– permite convertirea unei secvențe din formatul EMBL în formatul FASTA; funcția este utilă în situația în care se dorește eliminarea rapidă a informațiilor care nu au legătură cu secvența de ADN dintr-un fișier EMBL
Filter DNA	– elimină caracterele care nu corespund codului standard IUPAC pentru ADN-ul dintr-o secvență (inclusiv spații, numere, etc)
Filter Protein	– elimină caracterele ce nu corespund cu codul standard IUPAC de o literă pentru aminoacizi dintr-o secvență (inclusiv spații, numere etc.)
GenBank to FASTA	– permite convertirea unei secvențe din formatul GenBank în formatul FASTA; funcția este utilă în situația în care se dorește eliminarea rapidă a informațiilor care nu au legătură cu secvența de ADN dintr-un fișier GenBank
One to Three	– permite convertirea unei secvențe de aminoacizi din codul standard IUPAC de o literă în cel de trei litere
Analiza secvențelor	
Codon Plot	– analizează frecvența de utilizare a codonilor dintr-o secvență de ADN și generează un grafic cu aceste frecvențe
Codon Usage	– analizează frecvența de utilizare a codonilor dintr-o secvență de ADN și poate fi utilizat pentru a identifica preferința pentru un anumit codon sinonim
DNA Molecular Weight	– calculează masa moleculară a uneia sau a mai multor secvențe de ADN
DNA Stats	– calculează frecvența fiecărei nucleotide într-o secvență dată, rezultatele fiind exprimate în procente
PCR Primer Stats	– analizează și calculează proprietățile specifice ale unui set indicat de primeri, inclusiv temperatura de topire și procentul de GC
PCR Products	– realizează amplificarea virtuală prin PCR a unei secvențe date folosind un set de primeri indicat de utilizator
Protein Isoelectric Point	– calculează valoarea teoretică a punctului izoelectric pentru o secvență dată
Protein Molecular Weight	– calculează valoarea teoretică a masei moleculare pentru una sau mai multe secvențe de aminoacizi
Protein Stats	– calculează frecvența fiecărui aminoacid într-o secvență dată, rezultatele fiind exprimate în procente
Translate	– realizează translația și transcripția virtuală a unei secvențe de ADN date, utilizatorul având posibilitatea de a alege cadrul de citire specific
Reverse Translate	– realizează procesul invers față de Translate, acceptând o secvență de aminoacizi și generând secvența ADN corespunzătoare

* corespunzătoare cu denumirea din suita de programe SMS2