

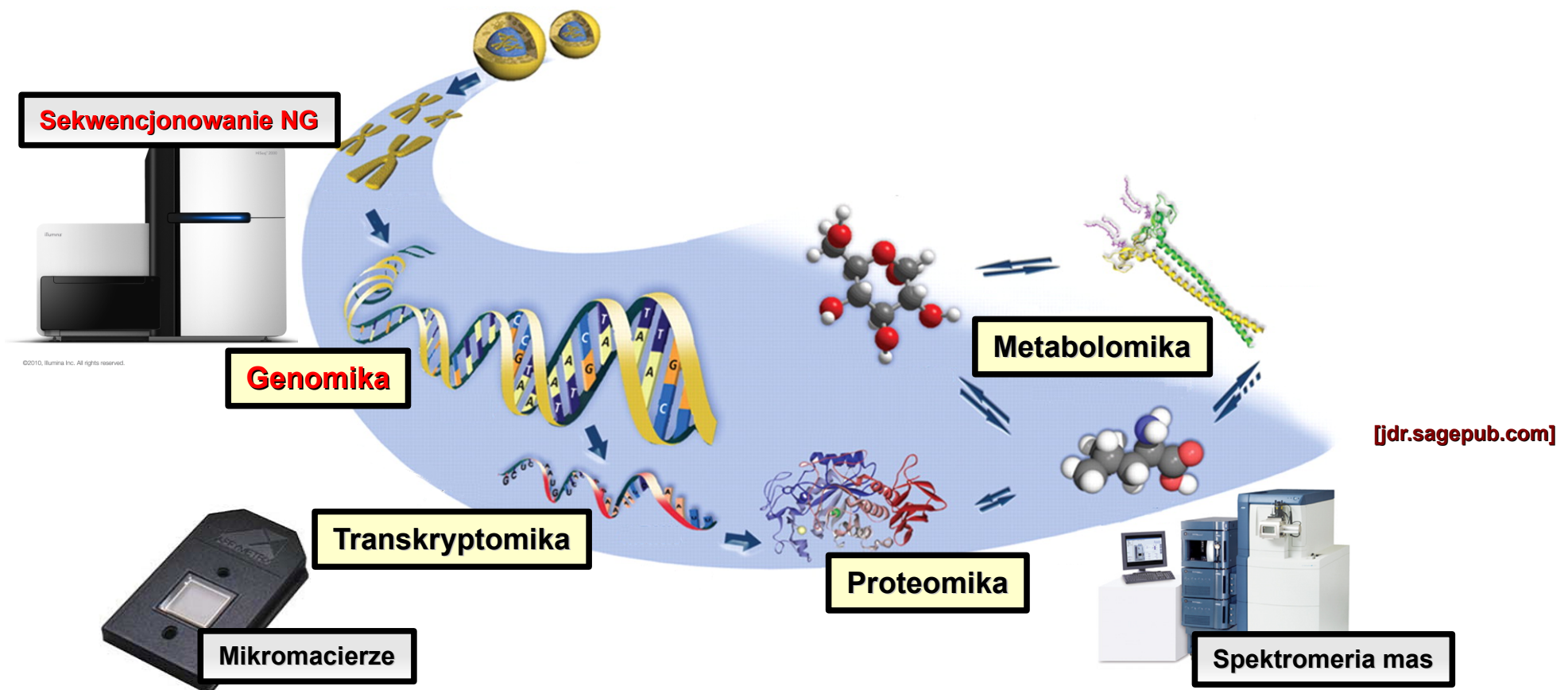
Uczenie maszynowe w bioinformatyce

Wykład 3: genomika

Tymon Rubel

Zakład Elektroniki Jądrowej i Medycznej
Instytut Radioelektroniki i Technik Multimedialnych PW

Genomika jest dziedziną wiedzy zajmującą się analizą genomów organizmów. Jej głównym celem jest poznanie kompletnej sekwencji genomu, określenie miejsc występowania i funkcji jego elementów (np. genów, sekwencji regulatorowych) oraz badanie występujących pomiędzy nimi zależności. Istotnym jej elementem jest również śledzenie zmian jakim genomy podlegają w czasie.

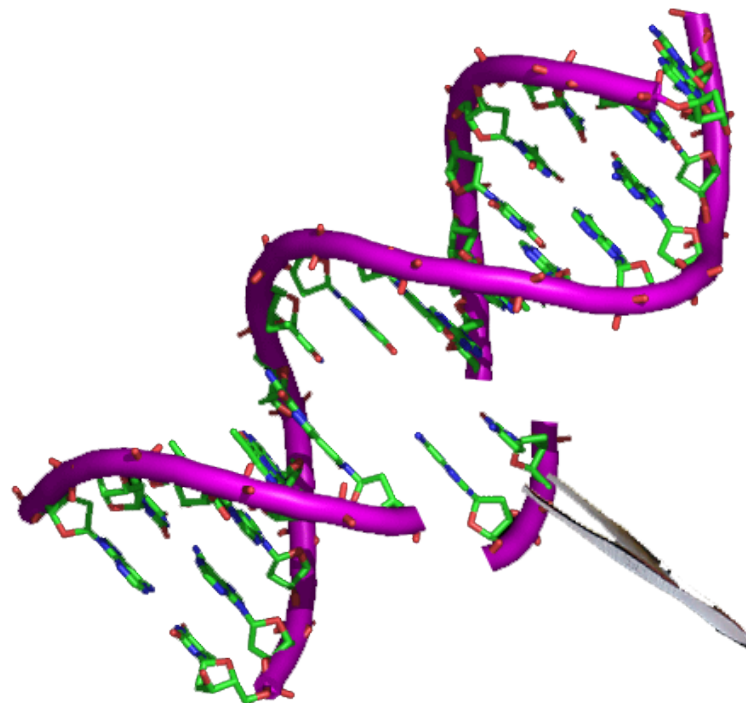


Genomika: narzędzia i techniki

Podstawowym narzędziem genomiki są metody **sekwencjonowania cząstek DNA**, czyli ustalania kolejności wchodzących w ich skład nukleotydów.

Jednak efektywne badanie obiektu tak dużego jak genom wymaga użycia szeregu dodatkowych **technik manipulacji cząstkami DNA**, służących m. in. do:

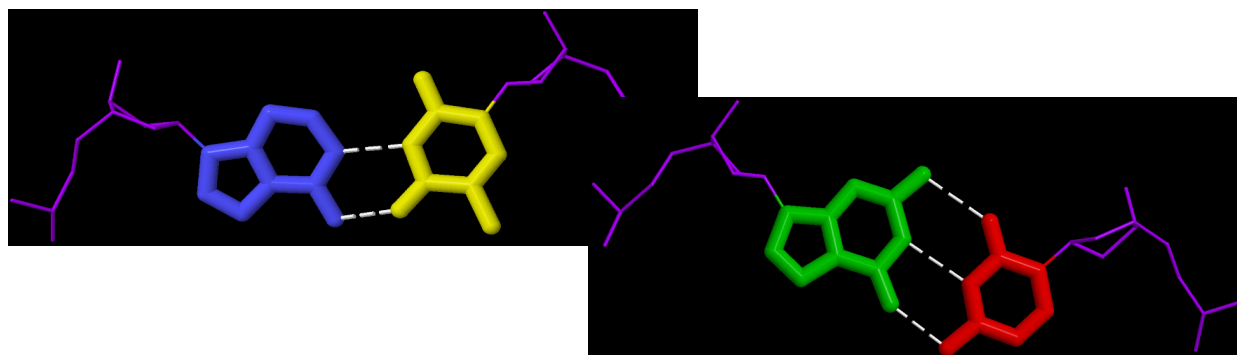
- **podziału** łańcuchów nukleotydów na krótsze odcinki;
- **rekombinacji**, czyli łączenia fragmentów pochodzących z różnych cząstek DNA;
- **amplifikacji** (zwielokrotnienia liczby kopii) wybranych odcinków;
- **separacji** pod względem długości;
- **detekcji** cząstek o zadanej sekwencji



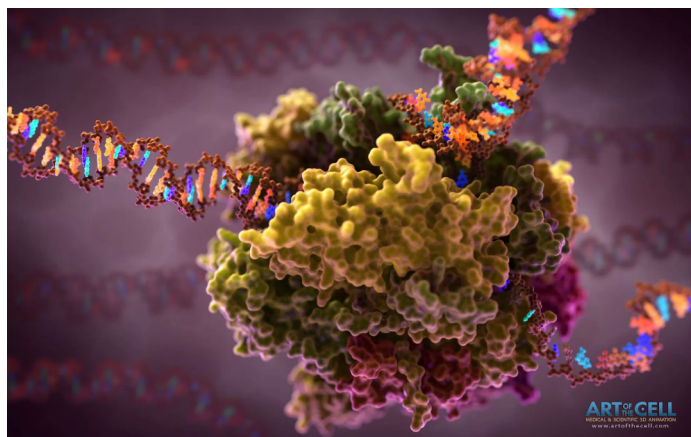
Genomika: narzędzia i techniki

Techniki manipulacji DNA są nieustannie modyfikowane i rozwijane od lat 70. XX wieku, jednak u podstaw działania większości z nich niezmiennie leżą:

- **podstawowe cechy cząstek DNA:** istnienie reguł komplementarności nukleotydów, denaturacja pasm pod wpływem temperatury i ich samoistna hybrydyzacja;



- **właściwości enzymów białkowych (polimeraz, nukleaz i ligaz)** biorących udział w procesach zachodzących naturalnie w komórkach, np. w replikacji i naprawie DNA, czy też podczas degradacji kwasów nukleinowych.



Uczenie maszynowe w bioinformatyce

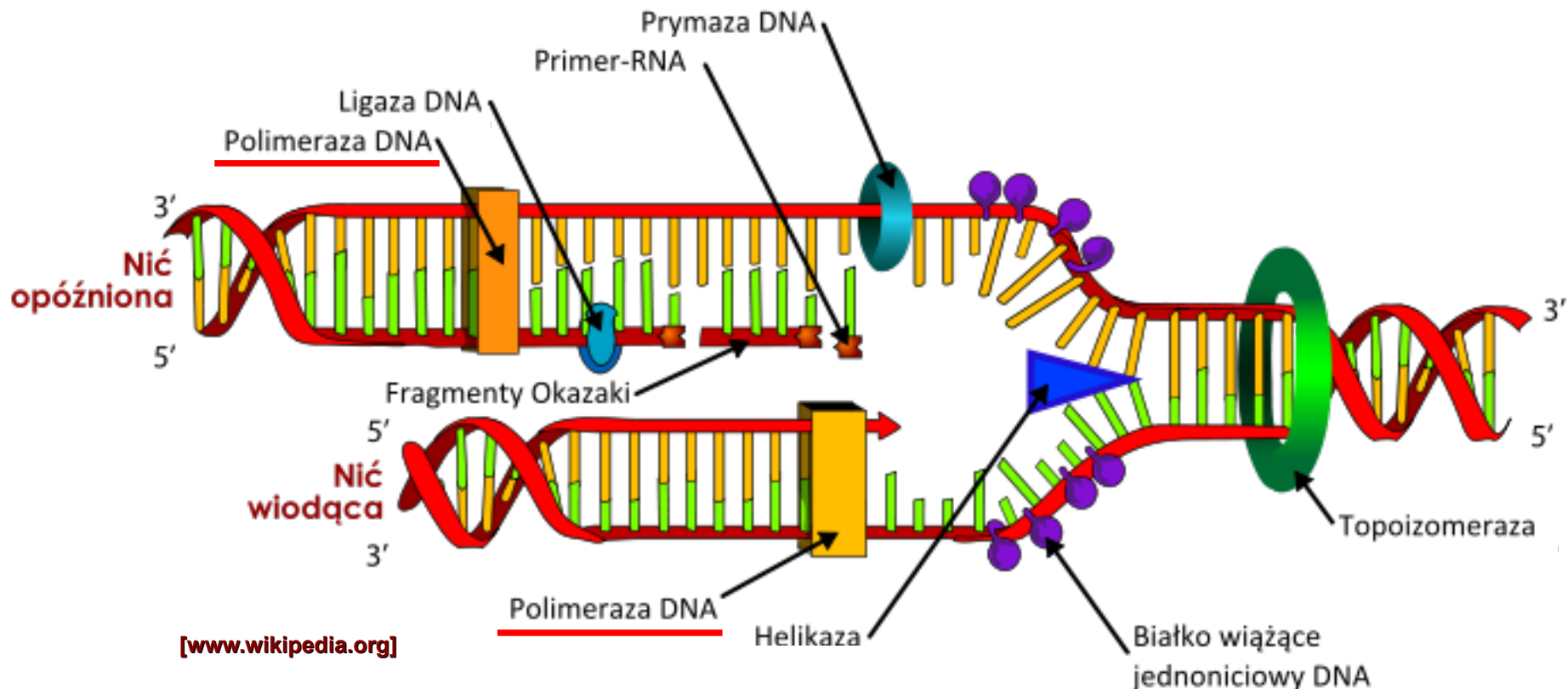
Wykład 3: genomika (enzymy używane do manipulacji DNA)

Tymon Rubel

Zakład Elektroniki Jądrowej i Medycznej
Instytut Radioelektroniki i Technik Multimedialnych PW

Enzymy służące do manipulacji DNA: polimerazy DNA zależne od DNA

Polimeraza DNA zależna od DNA to enzym syntetyzujący nić DNA komplementarną do matrycy, którą stanowi jednoniciowe DNA. Występuje naturalnie w komórkach wszystkich organizmów żywych, gdzie bierze udział w **replikacji** i **naprawie DNA**.



W biologii molekularnej zwykle używa się oczyszczonych i zmodyfikowanych wersji polimeraz bakteryjnych. Wykorzystywane są one przede wszystkim do **amplifikacji** oraz podczas **sekwencjonowania**.

Enzymy służące do manipulacji DNA: polimerazy DNA zależne od DNA

Najważniejsze właściwości polimerazy DNA:

- synteza nici DNA odbywa się w kierunku 5'→3' (kolejne nukleotydy dołączane są do końca 3' syntezowanej nici) zgodnie z regułą komplementarności wobec matrycy;



[T. A. Brown: *Genomy*. PWN, 2009]

- polimeraza DNA nie może rozpocząć syntezy na matrycy, która jest jednoniciowa na całej długości. Musi istnieć krótki odcinek dwuniciowy, powstający po dołączeniu do matrycy **startera (primera)** z wolnym końcem 3'. W komórce podczas replikacji DNA za utworzenie startera odpowiada polimeraza RNA (prymaza). W laboratorium funkcję primera pełni syntetyczny **oligonukleotydyd**, którego sekwencja decyduje o miejscu rozpoczęcia syntezy.

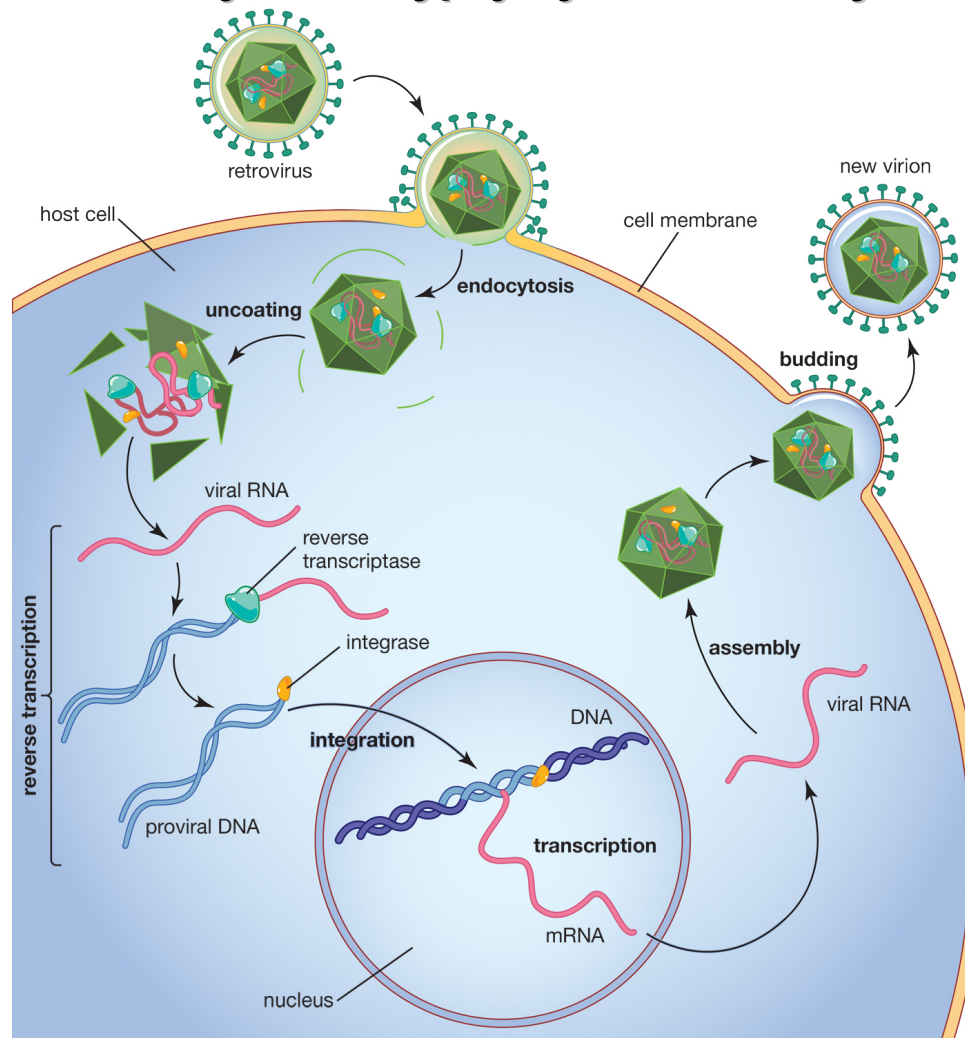


[T. A. Brown: *Genomy*. PWN, 2009]

Enzymy służące do manipulacji DNA: polimerazy DNA zależne od RNA

Polimeraza DNA zależna od RNA odpowiada za **odwrotną transkrypcję**, czyli syntezę nici DNA na matrycy RNA (stąd też jej inna nazwa: **odwrotna transkryptaza**).

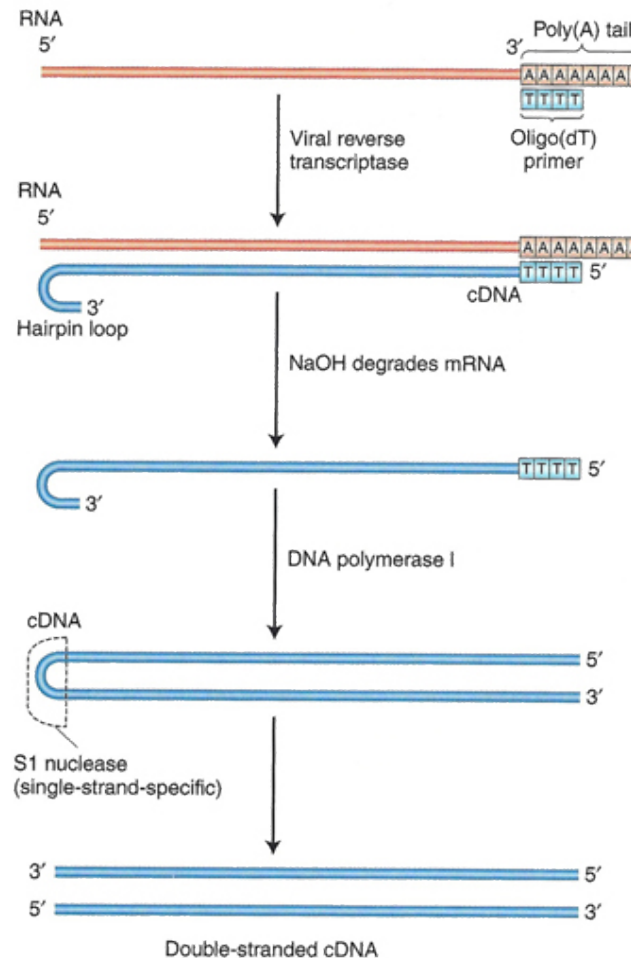
W warunkach naturalnych bierze udział w cyklach replikacyjnych retrowirusów (np. wirusa HIV), których genomy mają postać jednoniciowego RNA. Podczas infekcji RNA wirusa podlega odwrotnej transkrypcji i jest wbudowywane w DNA gospodarza.



Enzymy służące do manipulacji DNA: polimerazy DNA zależne od RNA

Odwrotna transkrypcja ma duże znaczenie laboratoryjne, gdyż może być użyta w procesie tworzenia dwuniciowych kopii DNA na podstawie transkryptów (cząstek mRNA). Kopie te nazywa się **komplementarnym DNA (cDNA, complementary DNA)**.

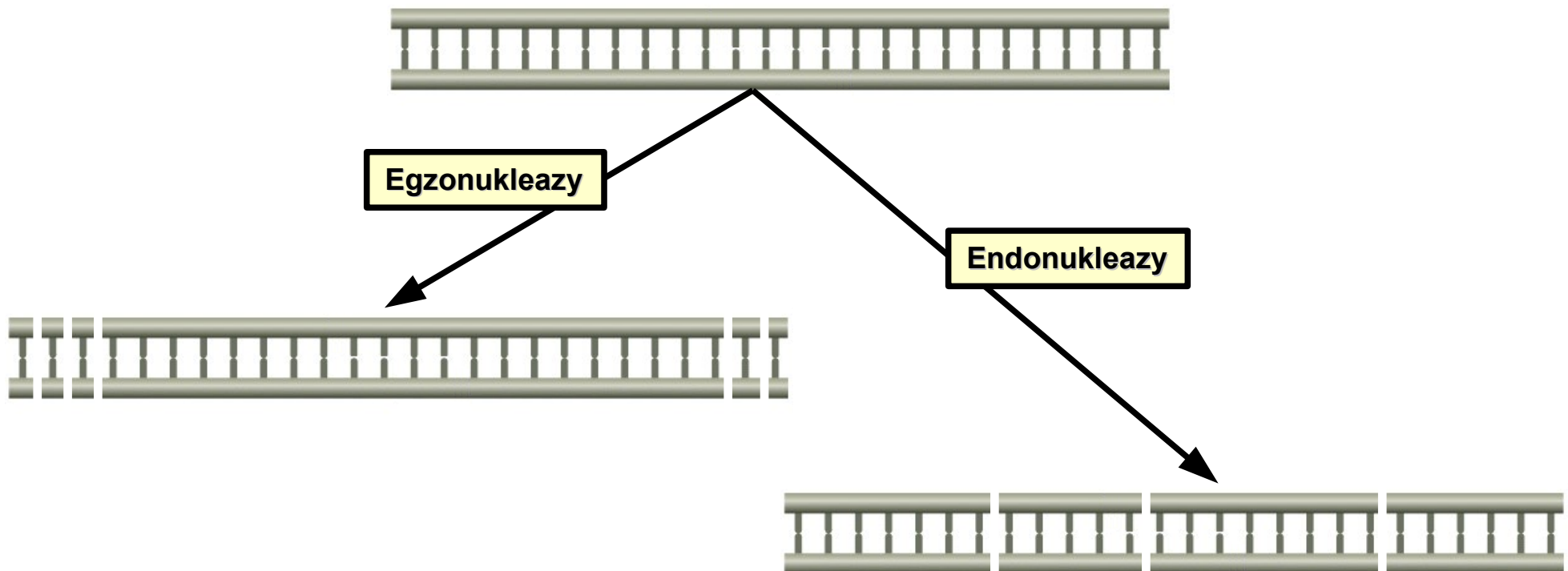
Zamiana mRNA na cDNA pozwala analizować transkrypty w taki sam sposób jak cząstki DNA. Dodatkową zaletą jest większa stabilność cDNA w porównaniu z mRNA.



Enzymy służące do manipulacji DNA: nukleazy

Nukleazy są enzymami, które degradują (trawią) kwasy nukleinowe rozrywając wiązania fosfodiesterowe łączące cukrowo-fosforanowe części nukleotydów:

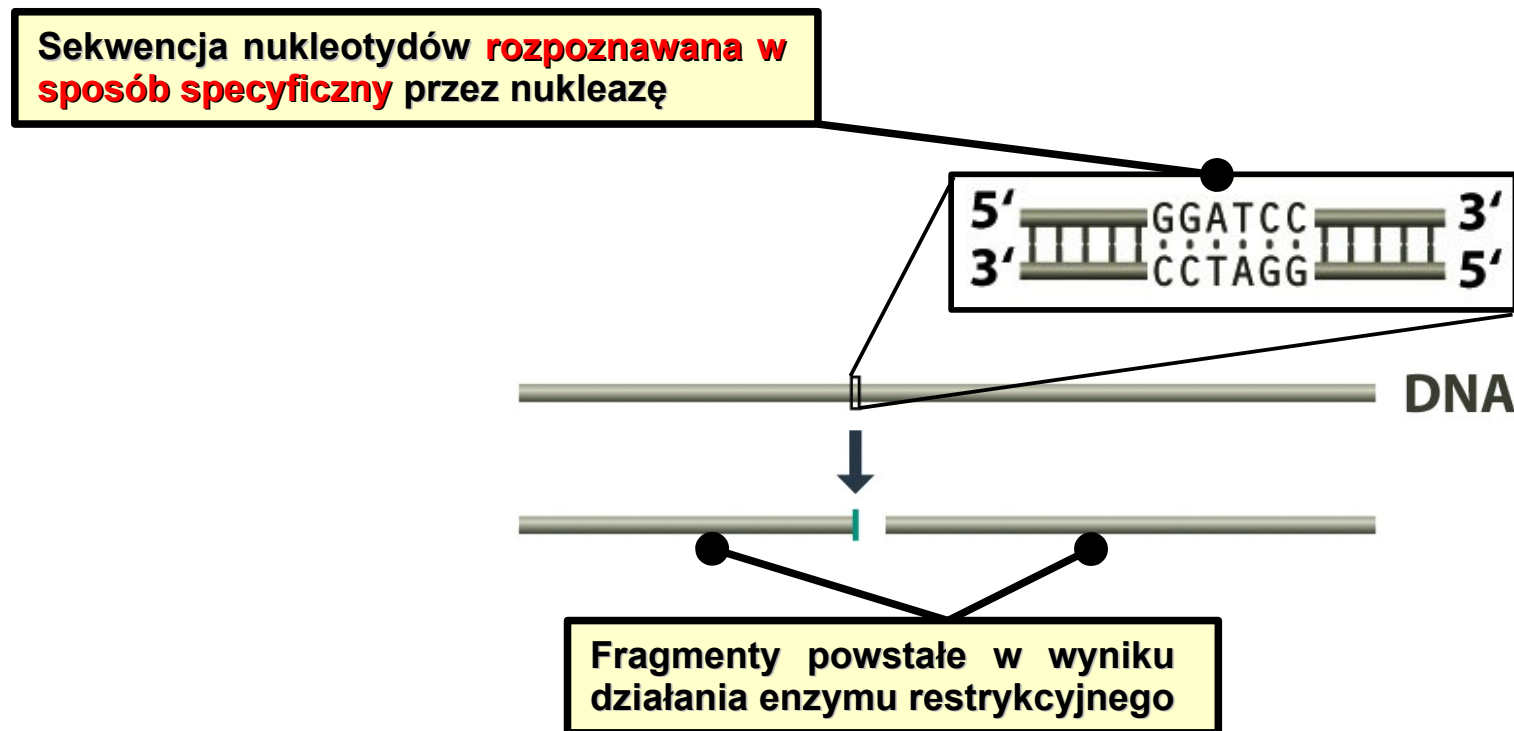
- **endonukleazy** tną wiązania wewnątrz cząstek;
- **egzonukleazy** usuwają pojedyncze nukleotydy z końców cząstek.



Enzymy służące do manipulacji DNA: nukleazy (enzymy restrykcyjne)

Istnieje wiele typów nukleaz, jednak z punktu widzenia procedur laboratoryjnych największe znaczenia mają **endonukleazy restrykcyjne**, które przecinają obie nici DNA tylko w bezpośrednim sąsiedztwie ściśle określonej sekwencji nukleotydów.

Różne rodzaje enzymów restrykcyjnych wykazują specyficzność do odmiennych sekwencji nukleotydów. Zwykle są one krótkie (od 4 do 8 nukleotydów) i często mają strukturę palindromu.



Enzymy służące do manipulacji DNA: nukleazy (enzymy restrykcyjne)

W praktyce laboratoryjnej wykorzystuje się obecnie kilkuset rodzajów endonukleaz restrykcyjnych pochodzenia bakteryjnego.

Enzym	Rozpoznawana sekwencja	Organizm	Enzym	Rozpoznawana sekwencja	Organizm
TaqI	5'-TCGA-3' 3'-AGCT-5'	Thermus aquaticus YTI	HindIII	5'-AAGCTT-3' 3'-TTCGAA-5'	Haemophilus influenzae
RsaI	5'-GTAC-3' 3'-CATG-5'	Rhodopseudomonas sphaeroides	KpnI	5'-GGTACC-3' 3'-CCATGG-5'	Klebsiella pneumoniae OK8
Sau3AI	5'-GATC-3' 3'-CTAG-5'	Staphylococcus aureus 3A	ClaI	5'-ATCGAT-3' 3'-TAGCTA-5'	Caryophanon latum
EcoRI	5'-GAATTC-3' 3'-CTTAAG-5'	Escherichia coli	BssHII	5'-GCGGCGC-3' 3'-CGCGGCG-5'	Bacillus stearothermophilus
BamHI	5'-GGATCC-3' 3'-CCTAGG-5'	Bacillus amyloliquefaciens H.	NotI	5'-GCGGCCGC-3' 3'-CGGCCGCG-5'	Nocardia otitidiscaviarum

Długości fragmentów powstałych wyniku działania enzymu restrykcyjnego zależą od jego specyficzności oraz sekwencji cząstki DNA.



Enzymy służące do manipulacji DNA: nukleazy (enzymy restrykcyjne)

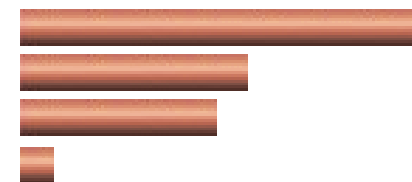
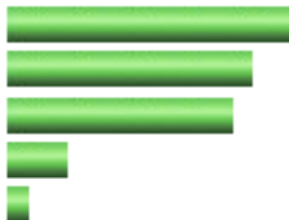
Przy zadanym enzymie wszystkie cząstki DNA jednego rodzaju zostaną pocięte na fragmenty o ściśle określonych długościach, zależnych od częstości występowania rozpoznawanej przez enzym sekwencji.



Cząsteczki DNA o odmiennej sekwencji, będą charakteryzować się innym rozlokowaniem miejsc, do których enzym wykazuje specyficzność. W efekcie wzorce cięcia w obu rodzajach cząsteczek będą się od siebie różniły.

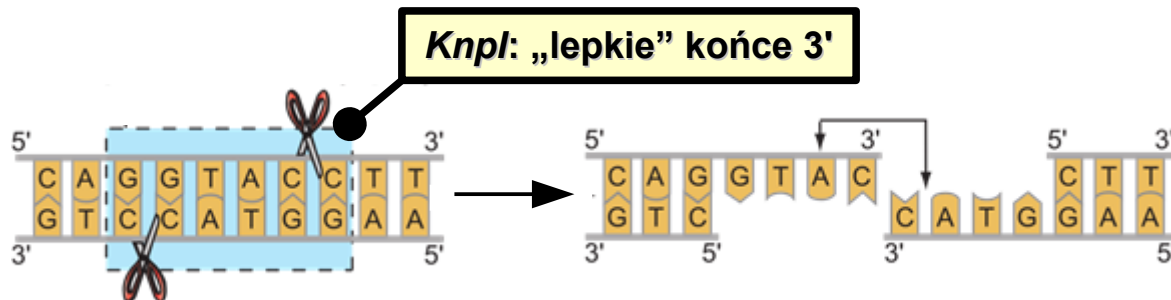
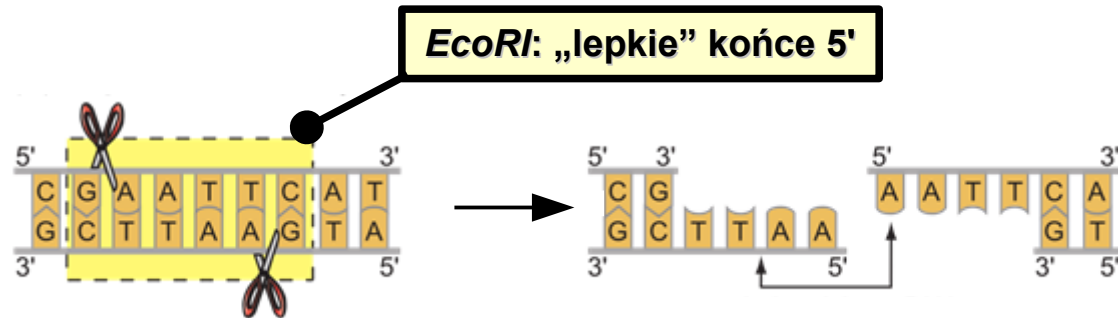
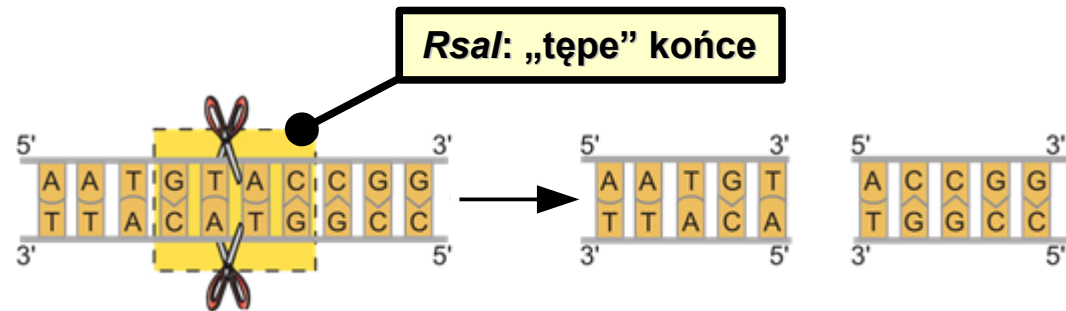


Uporządkowanie fragmentów pod względem długości (za pomocą omawianej dalej elektroforezy w żelu) pozwala rozróżnić różne rodzaje cząsteczek DNA na podstawie wzorów cięcia przez enzymy restrykcyjne.



Enzymy służące do manipulacji DNA: nukleazy (enzymy restrykcyjne)

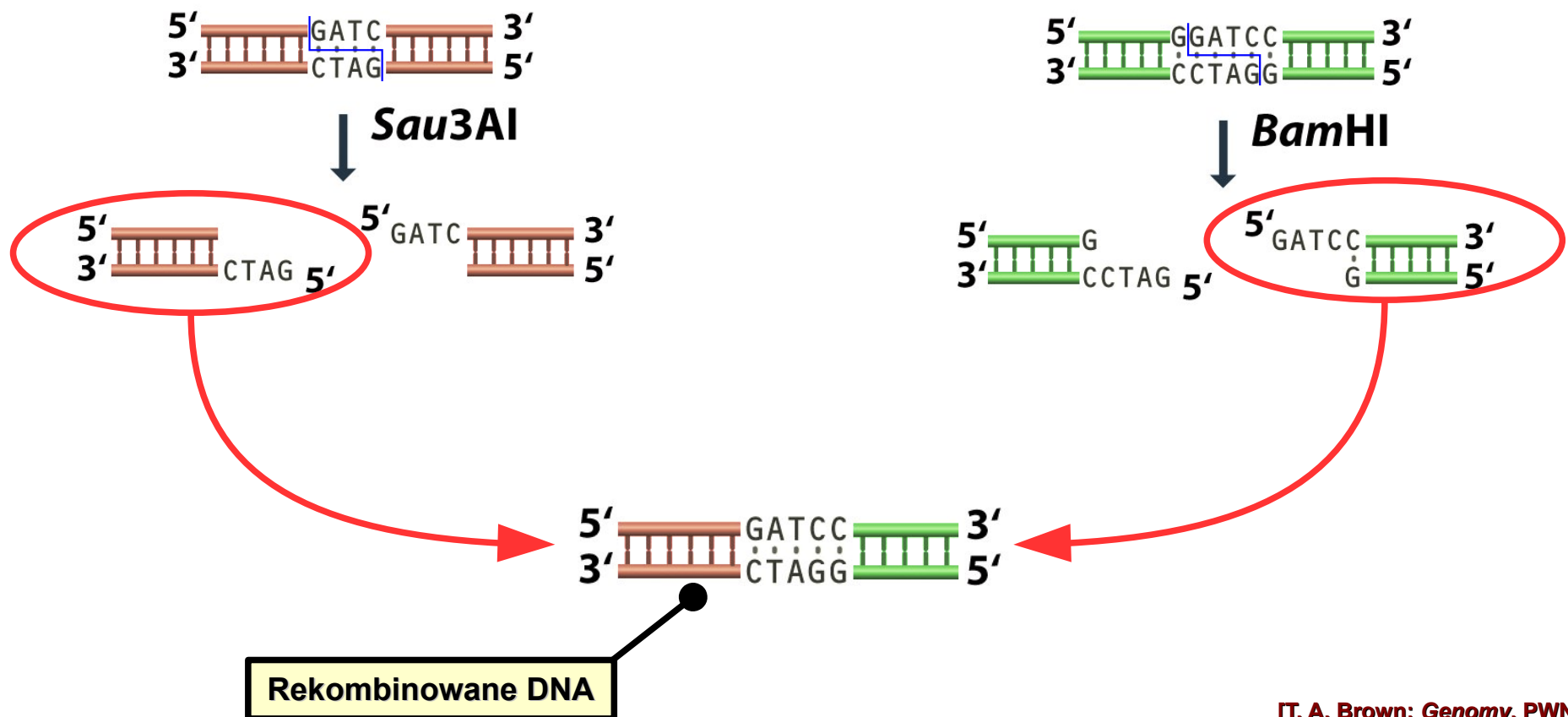
Szczególnie przydatne są enzymy restrykcyjne tnące obie nici DNA w nieco innych miejscach, powodując że powstałe fragmenty zakończone są krótkimi odcinkami jednoniciowego DNA (tzw. **lepkimi końcami**).



Enzymy służące do manipulacji DNA: nukleazy (enzymy restrykcyjne)

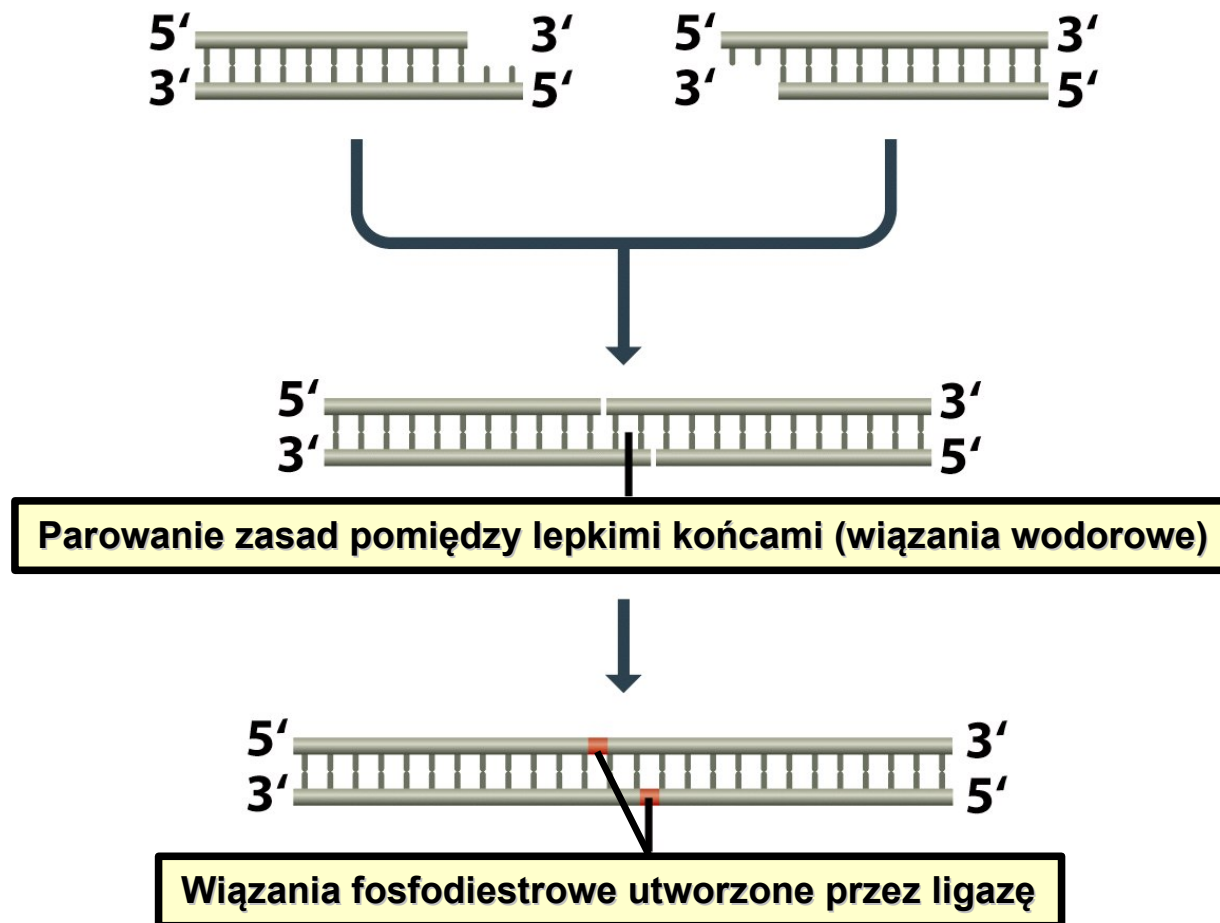
Komplementarność sekwencji lepkich końców fragmentów pochodzących z tej samej lub różnych cząstek DNA powoduje, że mogą one samoistnie łączyć się ze sobą za pośrednictwem wiązań wodorowych.

Rearanżacja DNA przez dzielenie i łączenie sekwencji różnych cząsteczek nazywana jest **rekombinacją**. Stanowi ona podstawę współczesnej inżynierii genetycznej.



Enzymy służące do manipulacji DNA: ligazy

Utrwalenie połączenia utworzonego przez lepkie końce dwóch fragmentów wymaga działania **ligazy**, czyli enzymu syntetyzującego wiązania fosfodiesterowe pomiędzy nukleotydami na końcach dwóch cząsteczek DNA.



Uczenie maszynowe w bioinformatyce

Wykład 3: genomika (techniki manipulacji cząstkami DNA)

Tymon Rubel

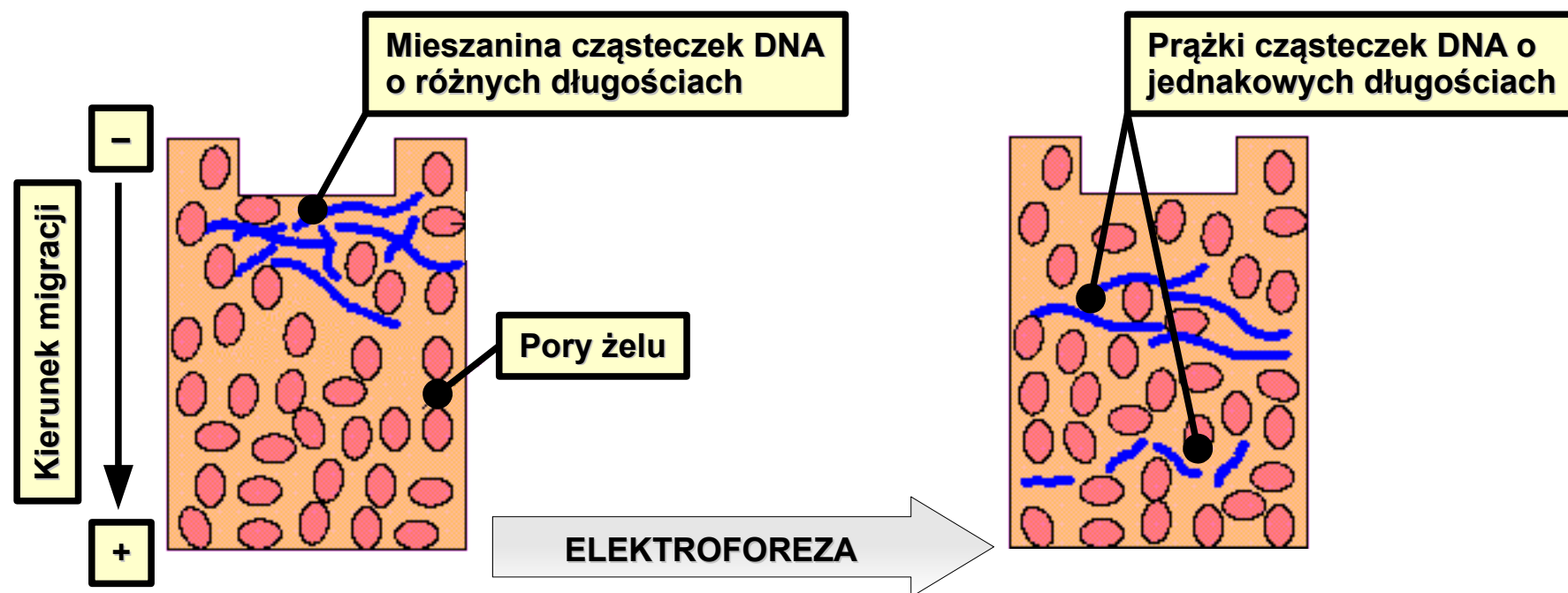
Zakład Elektroniki Jądrowej i Medycznej
Instytut Radioelektroniki i Technik Multimedialnych PW

Techniki manipulacji DNA: elektroforeza w żelu

Elektroforeza w żelu pozwala rozdzielać obdarzone ładunkiem cząstki (w tym także DNA) poprzez wymuszenie ich migracji w polu elektrycznym.

Interakcje z siecią przestrzenną żelu powodują, że szybkość migracji analizowanych cząstek jest zależna od wielkości: krótkie łańcuchy DNA swobodniej przemieszczają się w żelu, dzięki czemu wędrują szybciej od dłuższych cząsteczek.

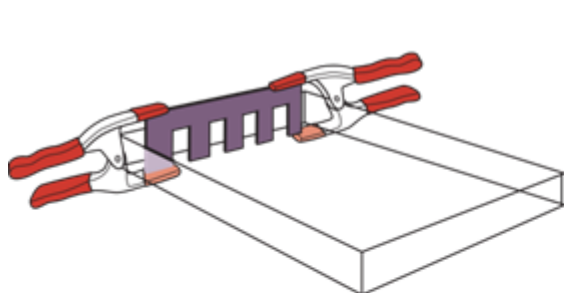
Efektem rozdziału elektroforetycznego jest powstanie w żelu szeregu prążków, z których każdy zawiera cząstki DNA o jednakowej długości (z pewną rozdzielczością, zależną od rodzaju żelu i zakresu długości rozdzielanych cząsteczek).



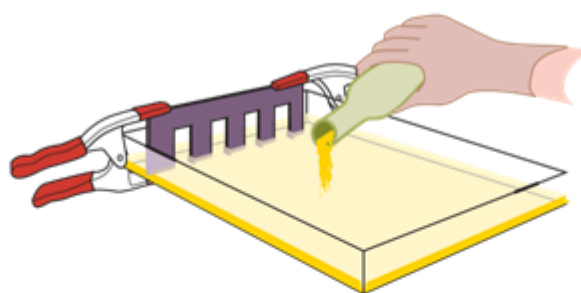
Techniki manipulacji DNA: elektroforeza w żelu

W elektroforezie cząstek DNA najczęściej stosuje się dwa rodzaje żeli:

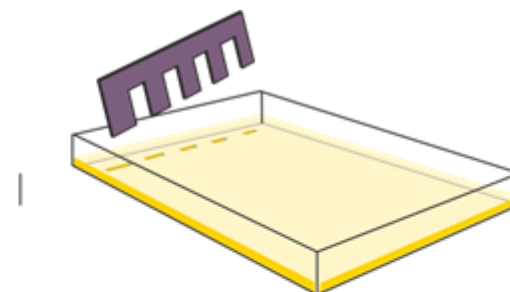
- **agarozowe** – do rozdziału długich fragmentów (niska rozdzielczość);
- **poliakrylamidowe** – do rozdziału krótkich fragmentów (wysoka rozdzielczość, która pozwala rozróżnić prążki cząstek różniących się długością o jeden nukleotyd).



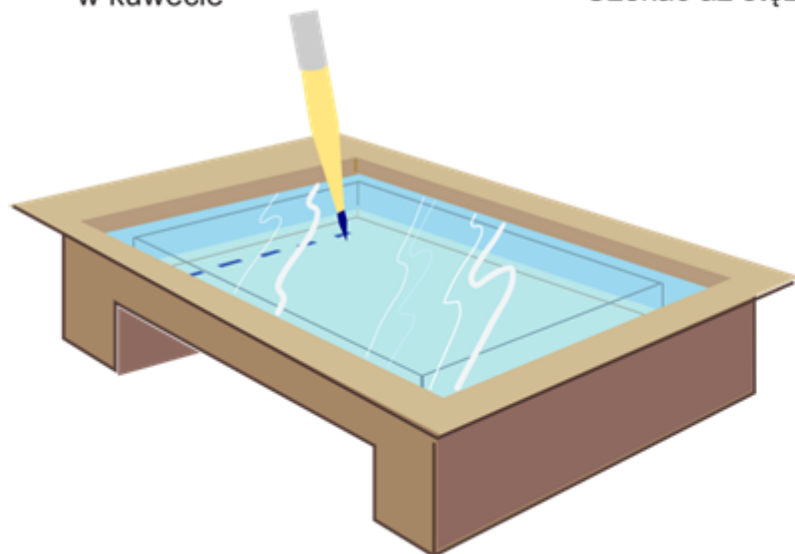
1. Ustabilizować grzebień w kuwecie



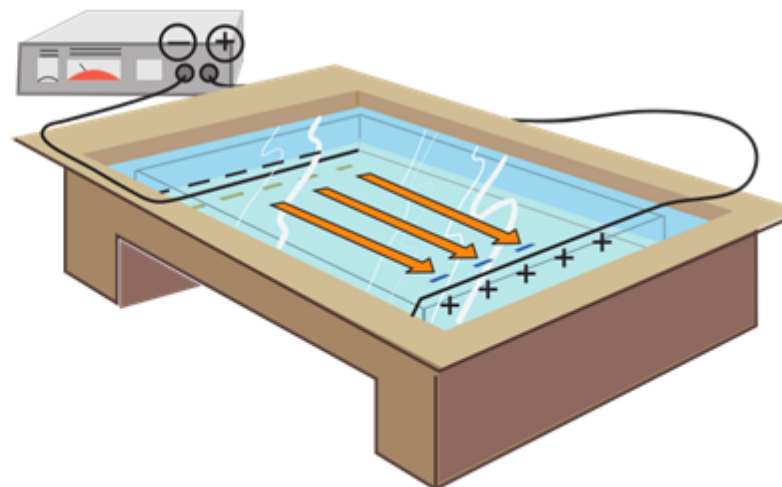
2. Wlać rozpuszczony roztwór agarozy. Czekać aż stężeje



3. Usunąć grzebień, usunąć żel z kuwety



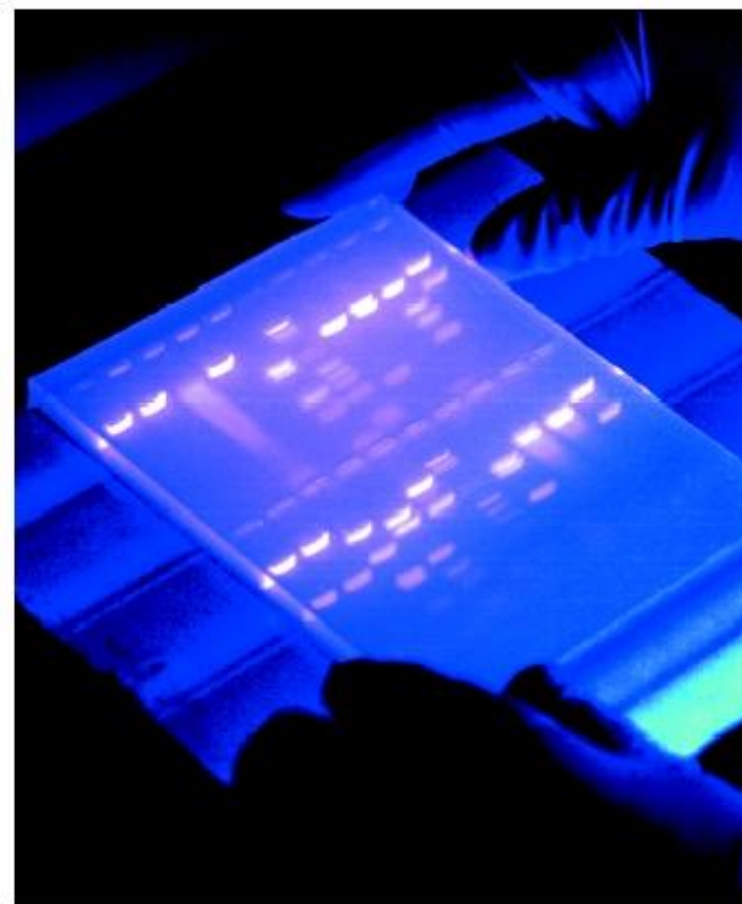
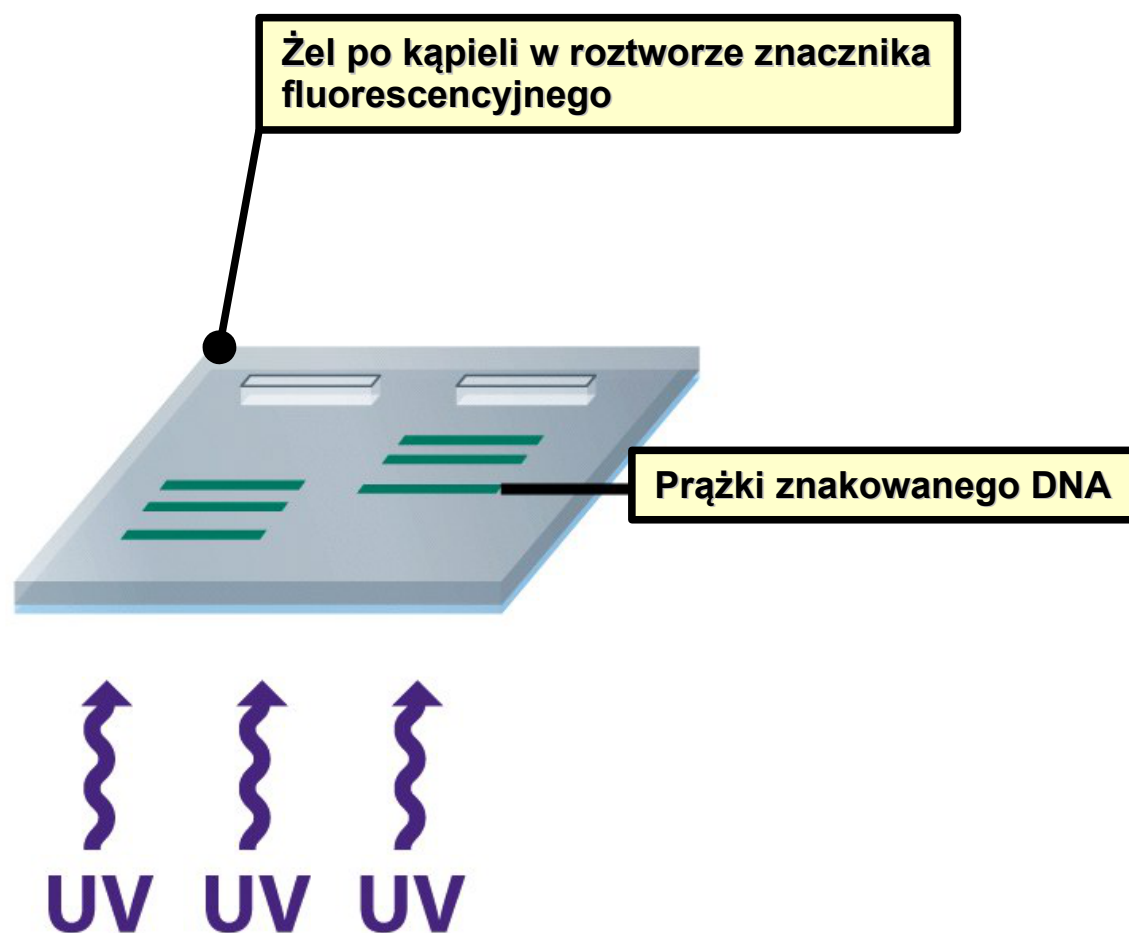
4. Zanurzyć żel w buforze, nałożyć próbkę



5. Przyłożenie napięcia - DNA jest naładowane ujemnie migruje w kierunku "+"

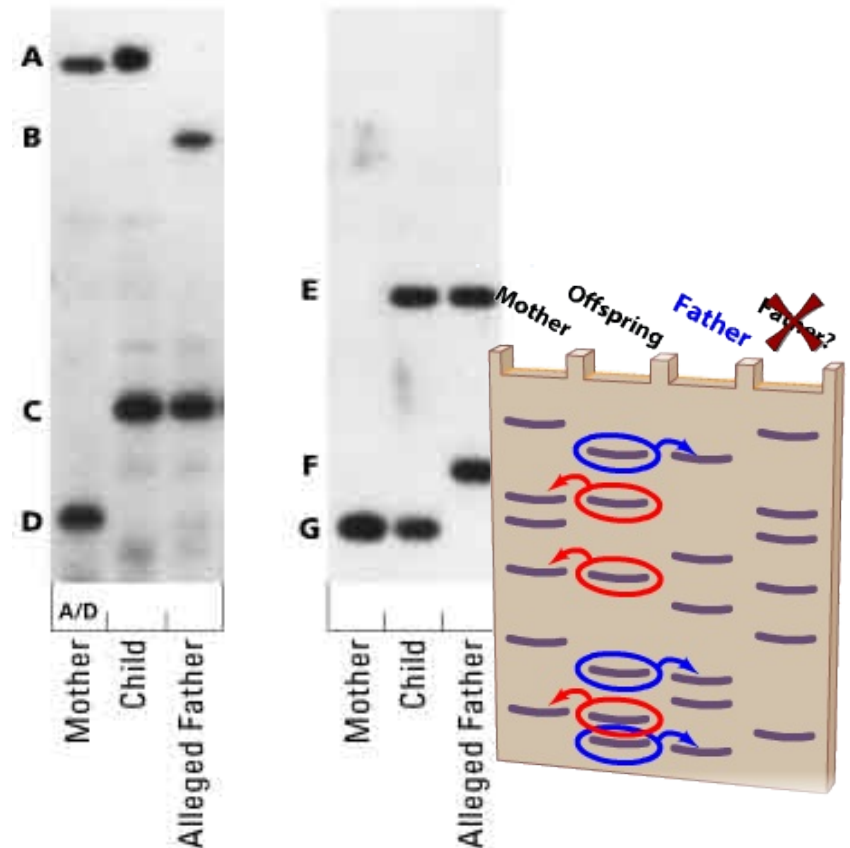
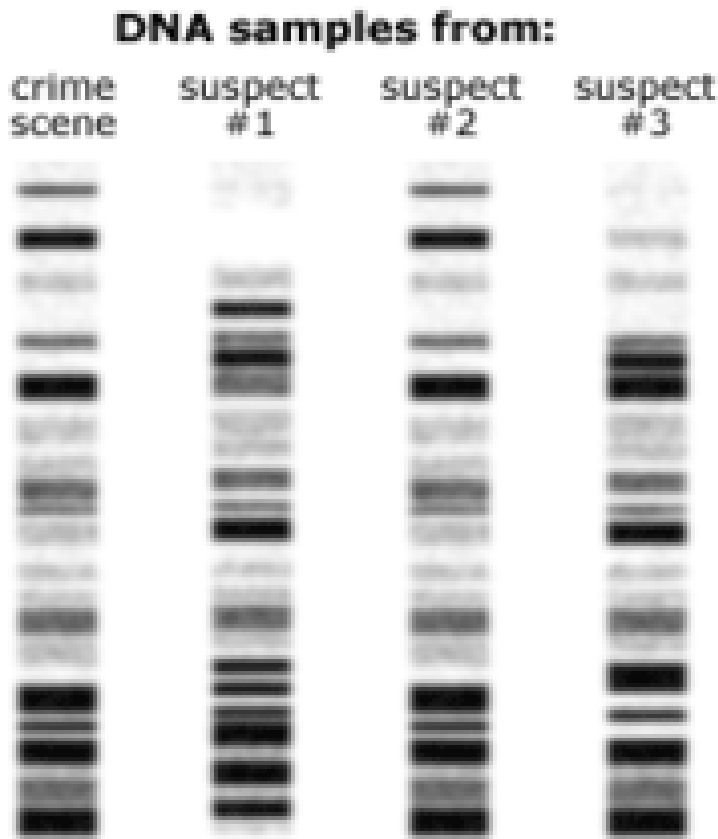
Techniki manipulacji DNA: elektroforeza w żelu

Powstałe w wyniku rozdziału prążki uwidacznia się przez odpowiednie **wybarwienie DNA za pomocą znaczników fluorescencyjnych** (np. bromku etydyny, GelGreen).



Techniki manipulacji DNA: elektroforeza w żelu (mapy restrykcyjne)

Elektroforeza w żelu fragmentów DNA powstałych w wyniku działania enzymów restrykcyjnych pozwala utworzyć tzw. **mapy restrykcyjne**, na podstawie których możliwe jest rozróżnianie cząsteczek DNA nawet bez znajomości ich sekwencji.

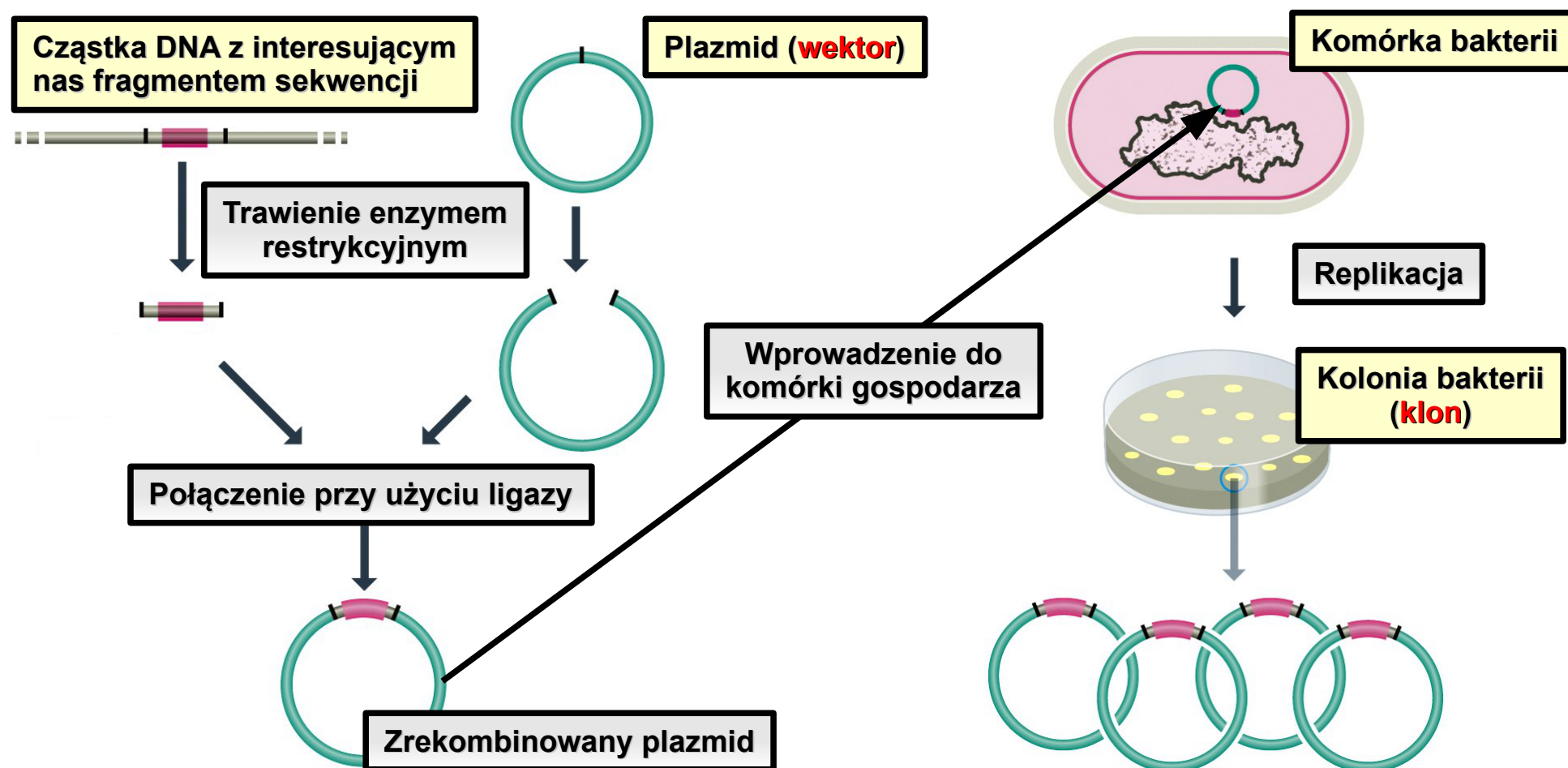


[<http://dna-fingerprinting.wikia.com>]

Do czasu upowszechnienia się technik sekwencjonowania mapy restrykcyjne były jedną z podstawowych metod identyfikacji i porównywania cząstek DNA.

Techniki manipulacji DNA: klonowanie molekularne

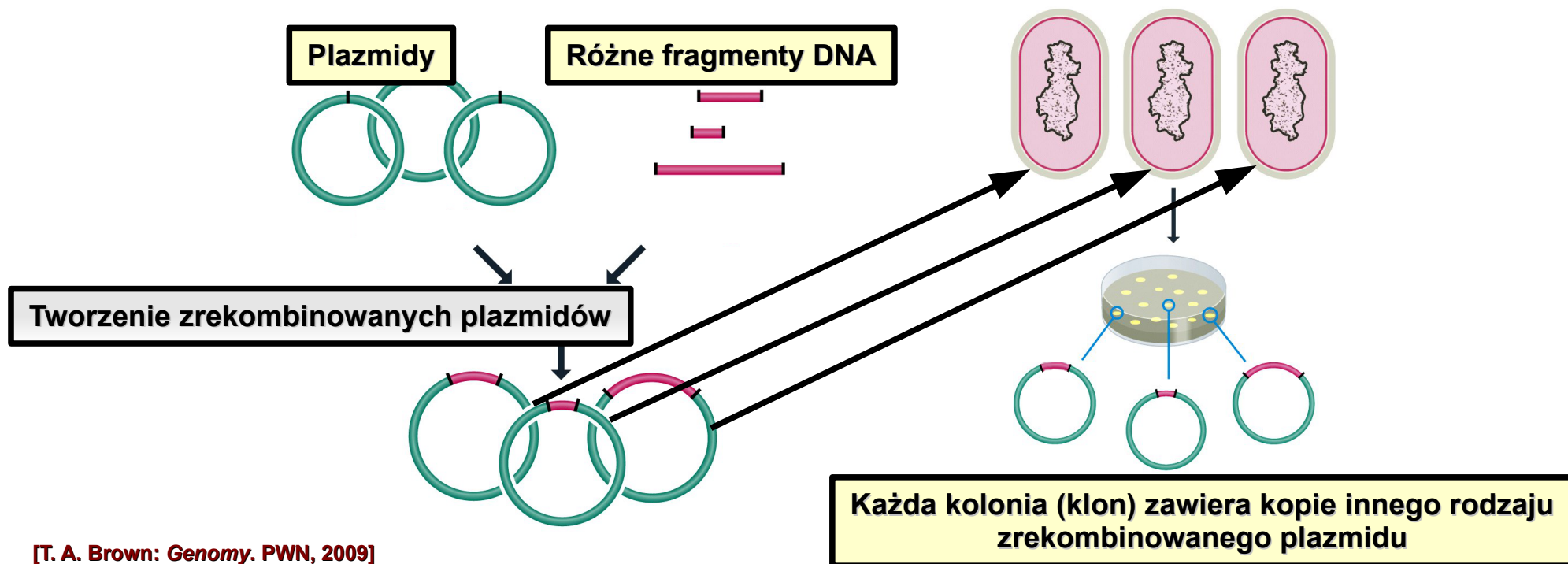
Klonowanie molekularne to procedura służąca do amplifikacji (zwiększania liczby kopii) DNA w organizmie gospodarza, najczęściej bakterii. Wymaga ona włączenia wybranego fragmentu DNA do **wektora**, czyli cząsteczki DNA posiadającej zdolność replikacji w komórce gospodarza (może nim być np. bakteryjny plazmid).



Techniki manipulacji DNA: klonowanie molekularne

Oprócz namnożenia amplifikacji DNA klonowanie są stosowane jest do:

- produkcja białka o sekwencji wynikającej z sekwencji interesującego nas DNA (korzystające przy tym mechanizmów ekspresji genów w organizmie gospodarza);
- tworzenie **bibliotek (banków) klonów**, czyli zbiorów **klonów**, będących grupami komórek zawierających różne zrekombinowane plazmidy. **Utworzenie biblioteki klonów przechowujących fragmenty genomu organizmu jest etapem wstępnym do jego sekwencjonowania.**



Techniki manipulacji DNA: reakcja łańcuchowa polimerazy

Reakcja łańcuchowa polimerazy (PCR – Polymerase Chain Reaction) jest techniką amplifikacji wybranego fragmentu DNA, której przeprowadzenie wymaga:

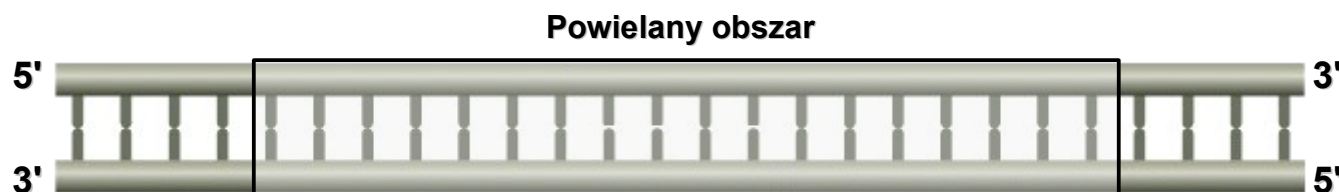
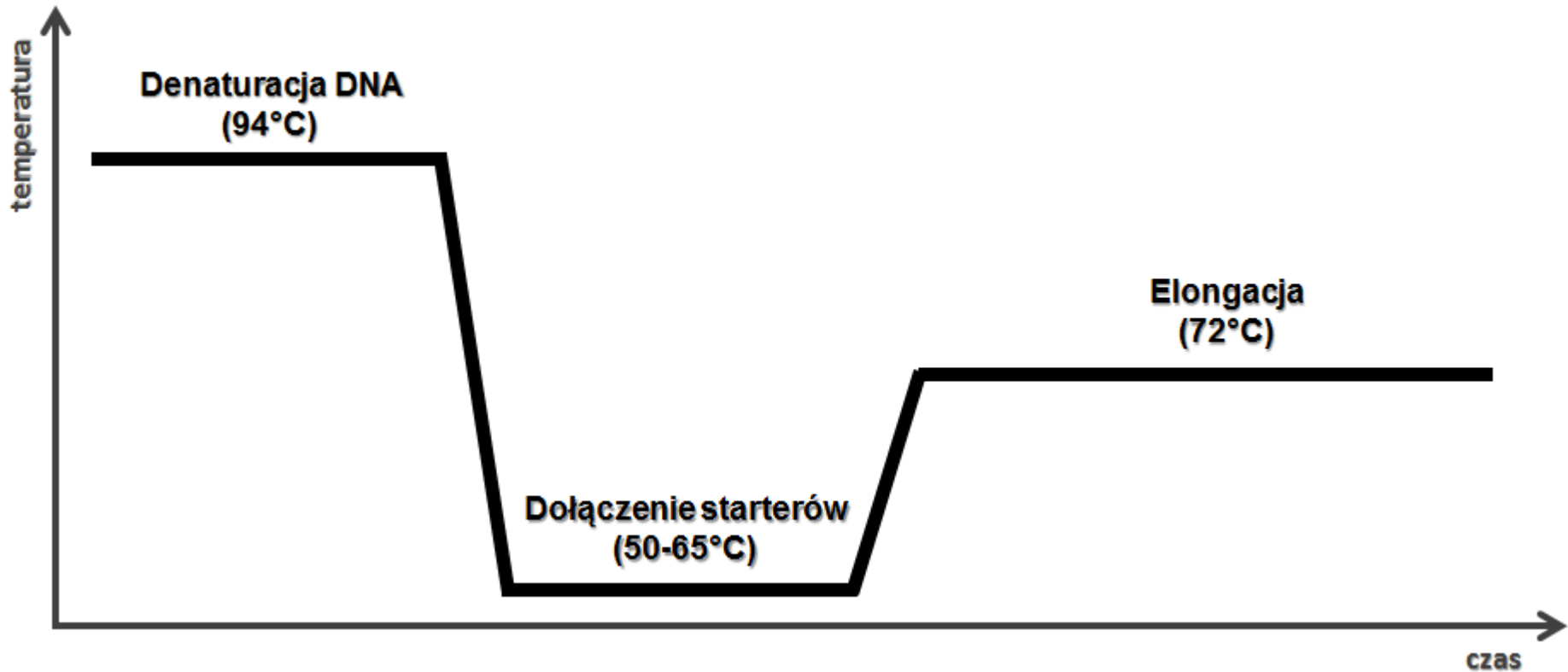
- cząstki DNA, której wybrany fragment będzie matrycą do syntezy nowych kopii;
- termostabilnej polimerazy DNA, np. z bakterii *Thermus aquaticus* (Taq);
- pary **starterów (primerów) oligonukleotydowych**, czyli dwóch krótkich odcinków jednoniciowego DNA o sekwencjach komplementarnych do rejonów otaczających powielany obszar;



- substratów reakcji, czyli czterech rodzajów trifosforanów deoksyrybonukleotydów: **dATP**, **dGTP**, **dCTP**, **dTTP**;
- termocyklera pozwalającego kontrolować i cyklicznie zmieniać temperaturę, w jakiej zachodzą poszczególne etapy PCR.

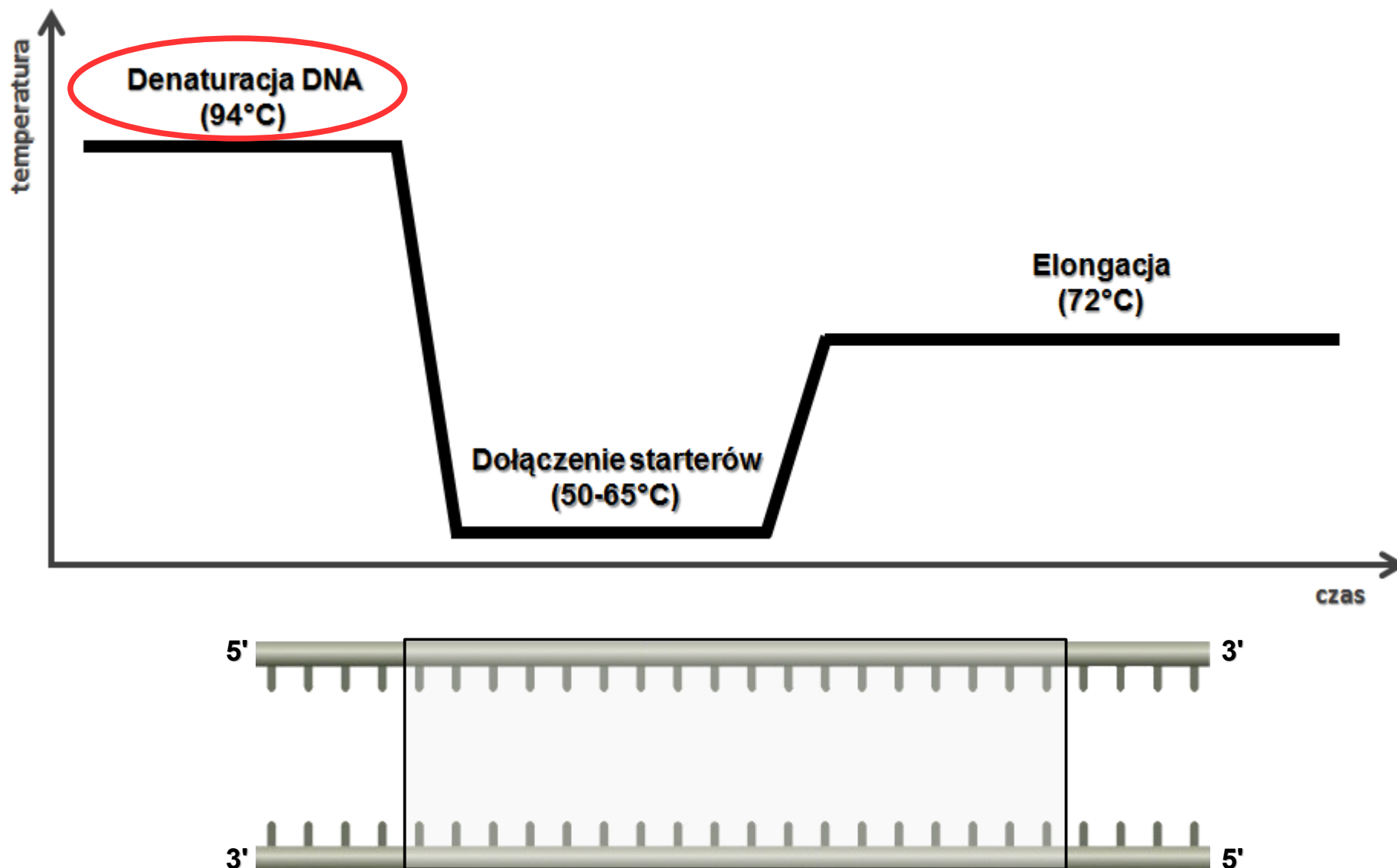
Techniki manipulacji DNA: reakcja łańcuchowa polimerazy

Powielanie następuje w cyklach, składających się z trzech kolejnych faz: **denaturacji DNA** (w temperaturze **94° C**), **dołączenia starterów** (temp. **≈ 50° C**) oraz **elongacji**, czyli syntezy nowych nici DNA przez polimerazę (temp. **72° C**).



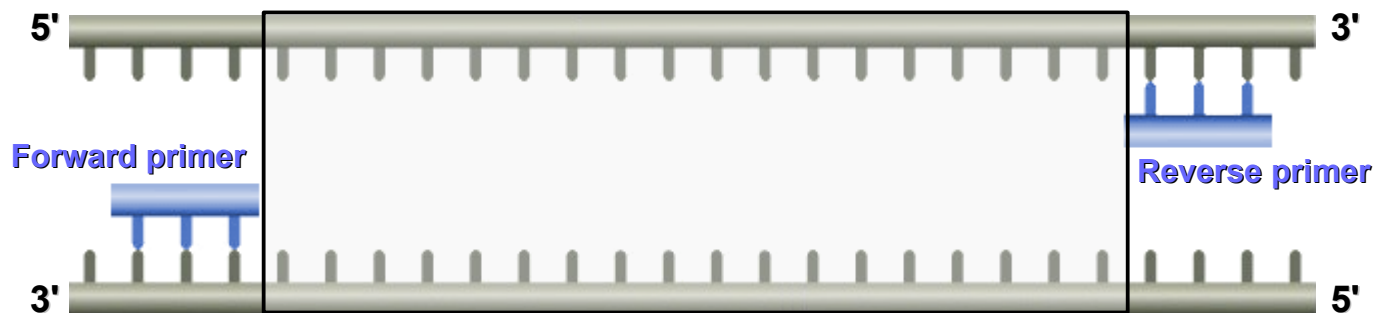
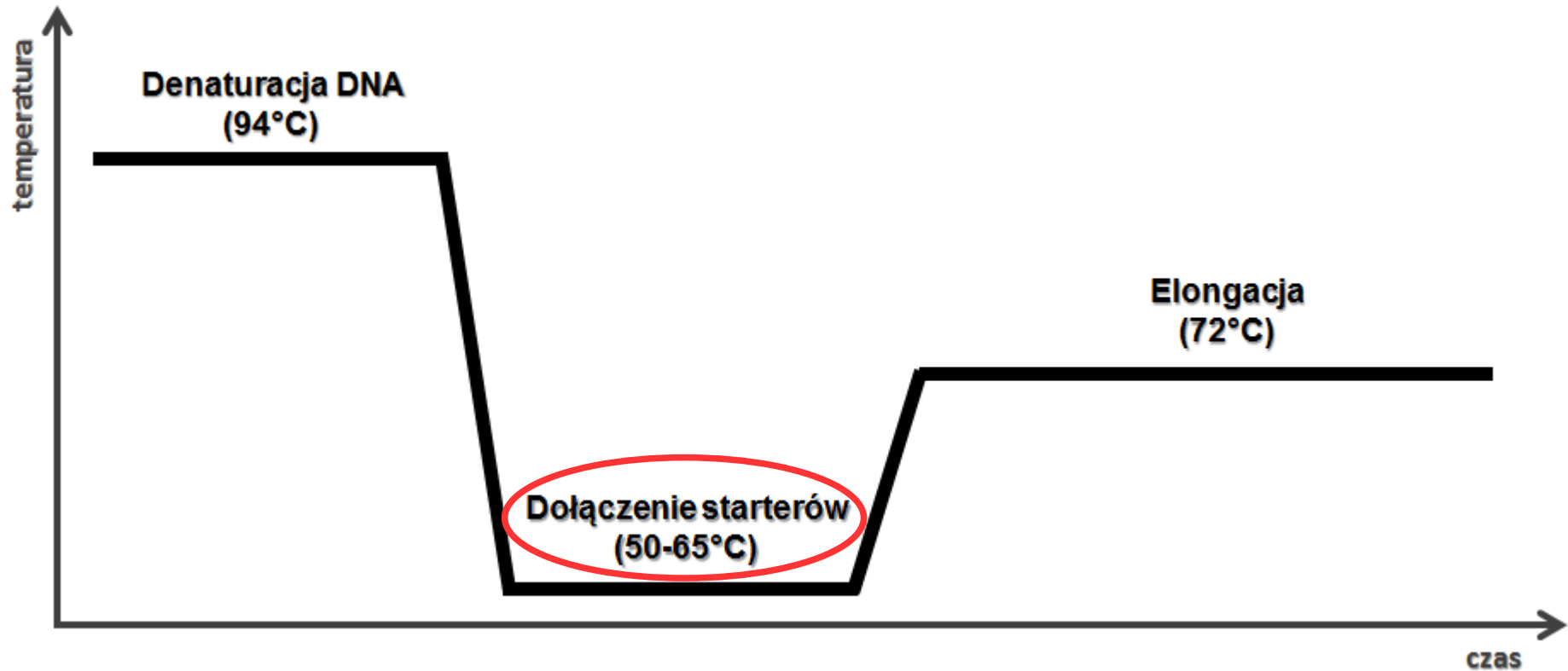
Techniki manipulacji DNA: reakcja łańcuchowa polimerazy

Powielanie następuje w cyklach, składających się z trzech kolejnych faz: **denaturacji DNA** (w temperaturze **94° C**), **dołączenia starterów** (temp. **≈ 50° C**) oraz **elongacji**, czyli syntezy nowych nici DNA przez polimerazę (temp. **72° C**).



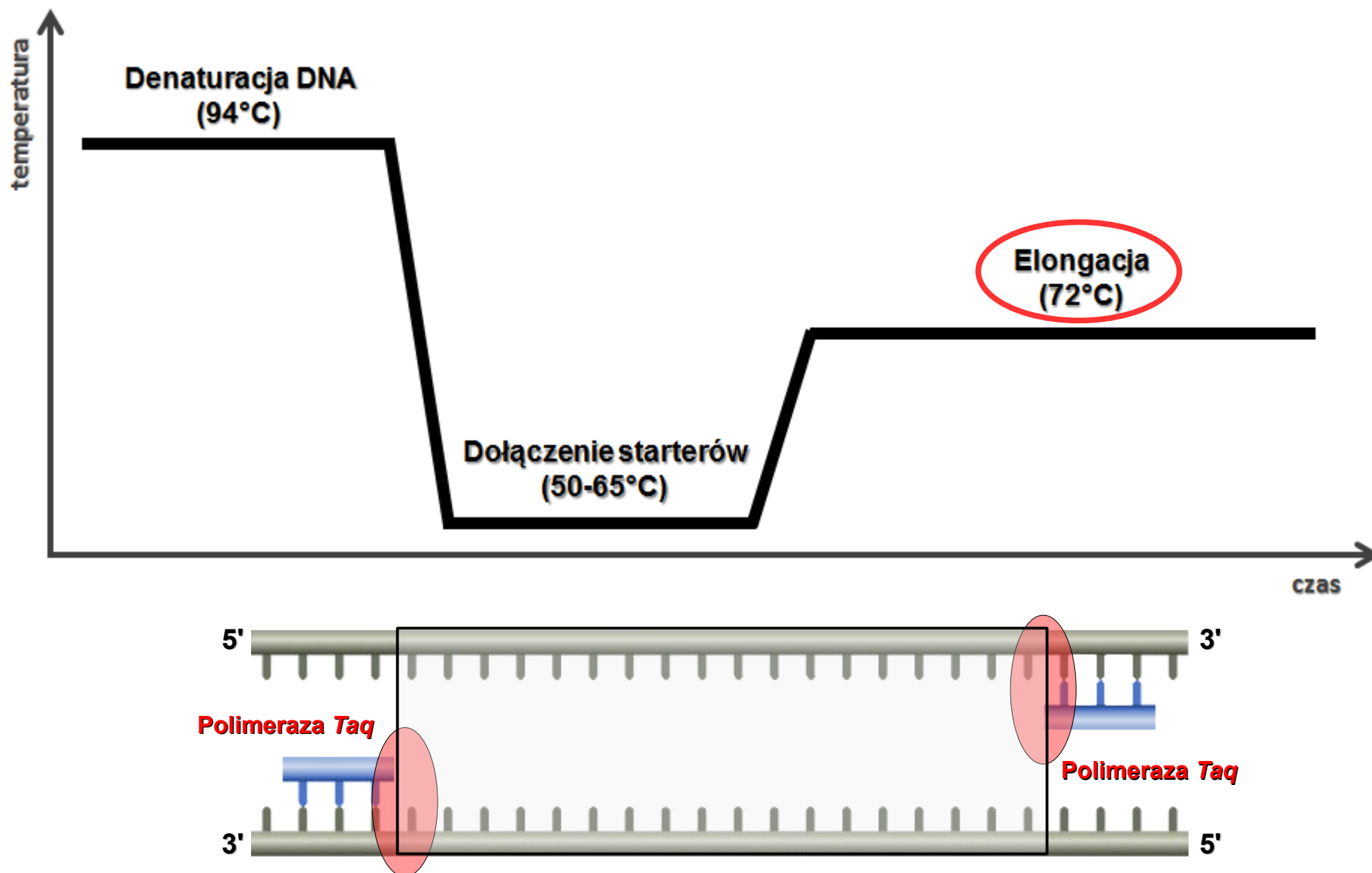
Techniki manipulacji DNA: reakcja łańcuchowa polimerazy

Powielanie następuje w cyklach, składających się z trzech kolejnych faz: **denaturacji DNA** (w temperaturze **94° C**), **dołączenia starterów** (temp. **≈ 50° C**) oraz **elongacji**, czyli syntezy nowych nici DNA przez polimerazę (temp. **72° C**).



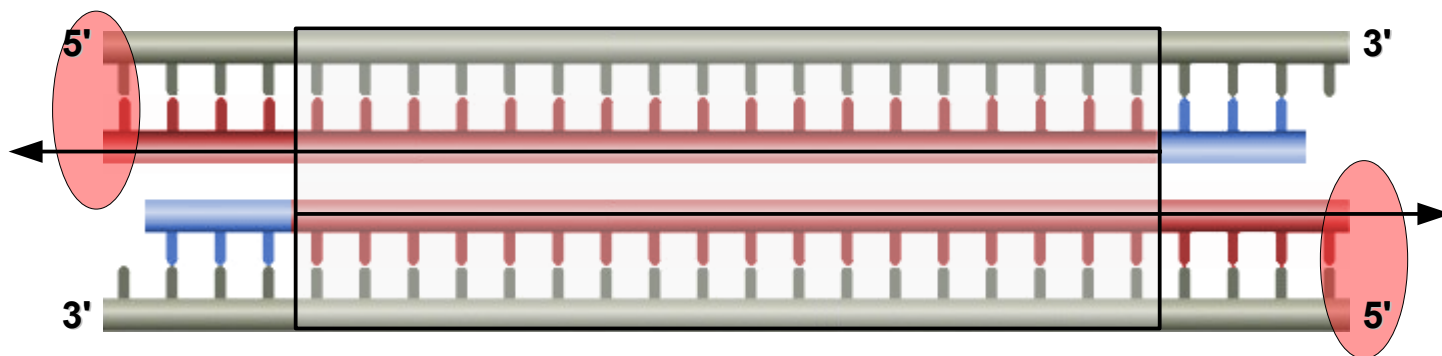
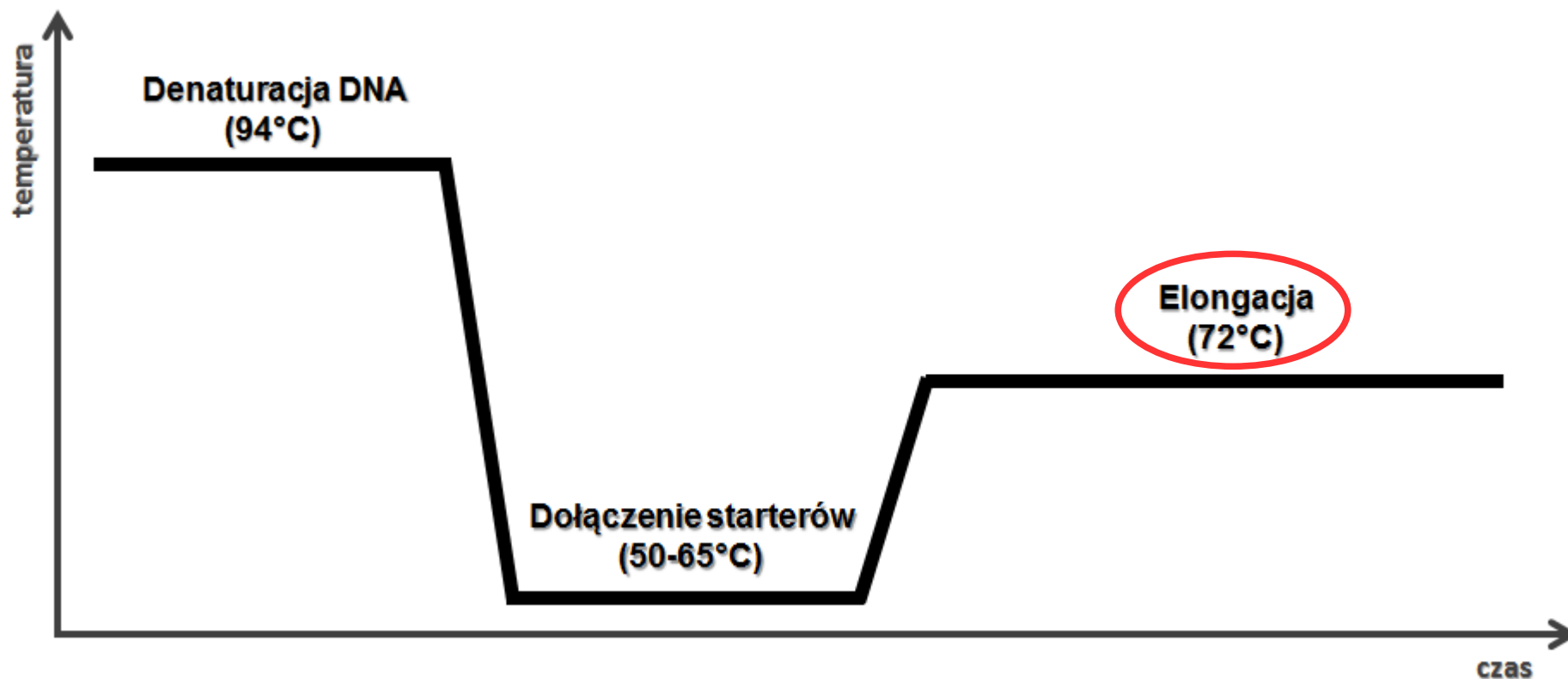
Techniki manipulacji DNA: reakcja łańcuchowa polimerazy

Powielanie następuje w cyklach, składających się z trzech kolejnych faz: **denaturacji DNA** (w temperaturze **94° C**), **dołączenia starterów** (temp. **≈ 50° C**) oraz **elongacji**, czyli syntezy nowych nici DNA przez polimerazę (temp. **72° C**).



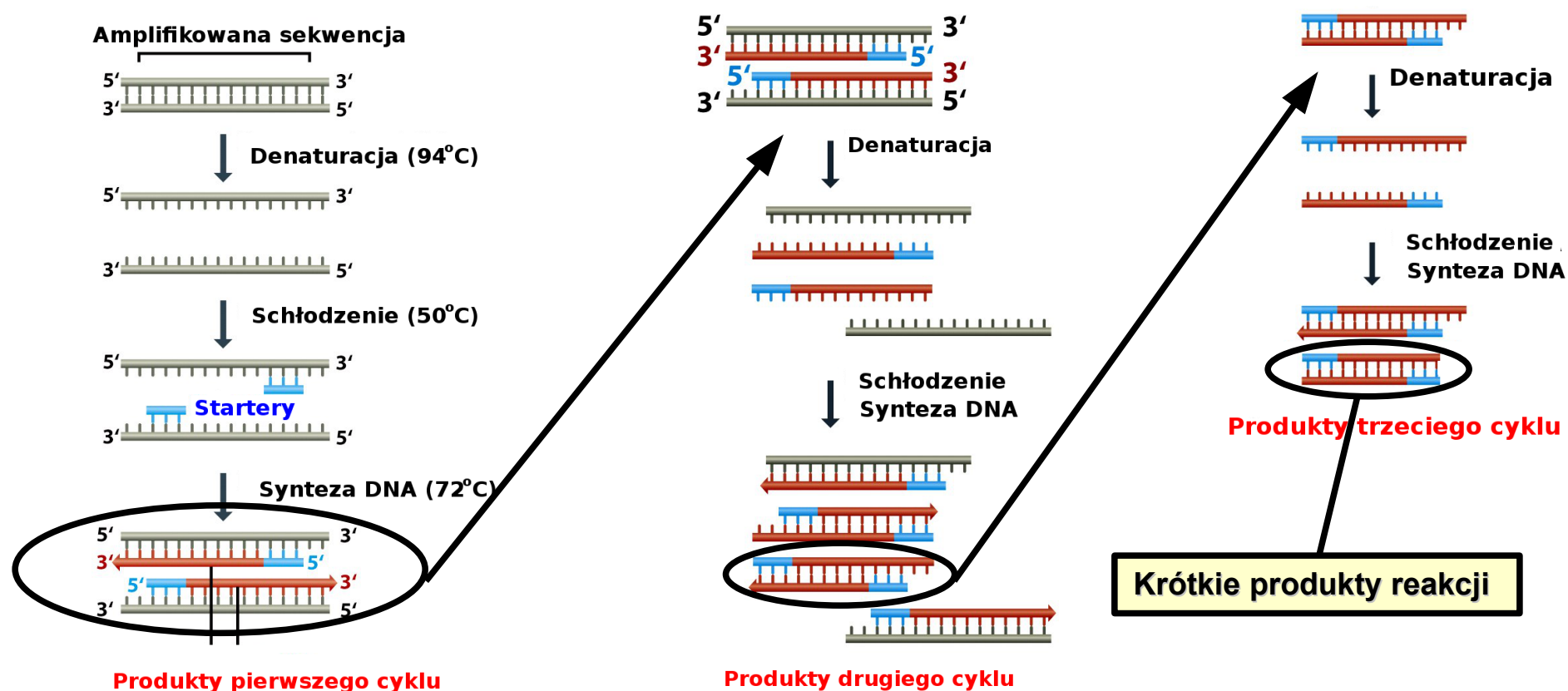
Techniki manipulacji DNA: reakcja łańcuchowa polimerazy

Powielanie następuje w cyklach, składających się z trzech kolejnych faz: **denaturacji DNA** (w temperaturze **94° C**), **dołączenia starterów** (temp. **≈ 50° C**) oraz **elongacji**, czyli syntezy nowych nici DNA przez polimerazę (temp. **72° C**).



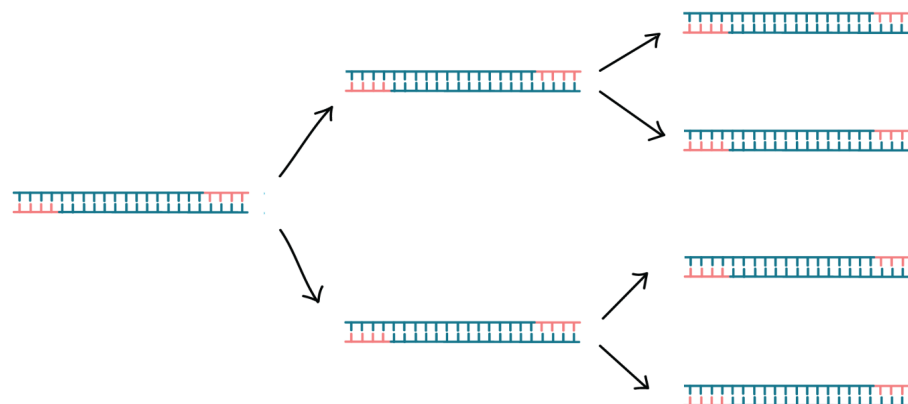
Techniki manipulacji DNA: reakcja łańcuchowa polimerazy

Cykle PCR są powtarzane wielokrotnie. W trzecim cyklu powstają **krótkie produkty reakcji**, czyli dokładne kopie interesującego nas fragmentu DNA wraz ze starterami, które w kolejnych cyklach będą powielane wykładniczo.

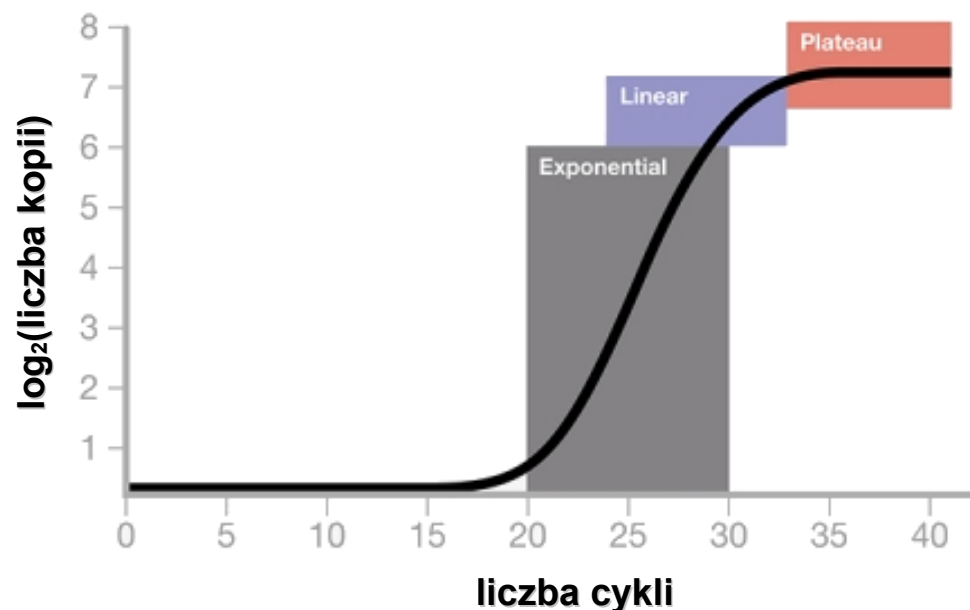


Techniki manipulacji DNA: reakcja łańcuchowa polimerazy

W przypadku krótkich produktów każda z nici jest matrycą do syntezy identycznej cząstki w kolejnym cyklu. Stąd też teoretycznie po N cyklach liczba ich kopii będzie równa 2^N (np. po 30 cyklach otrzymuje się ponad miliard kopii cząsteczki).



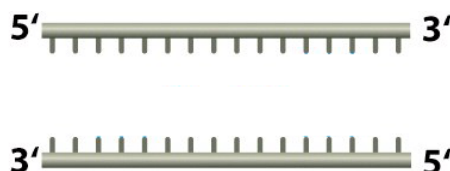
W praktyce, po początkowej **fazie wykładniczej** efektywność powielania spada (**faza liniowa**), aż wreszcie reakcja wygasa z powodu braku substratów (**faza plateau**).



Techniki manipulacji DNA: reakcja łańcuchowa polimerazy

Podstawowym celem PCR jest amplifikacja DNA, jednak może być ona również użyta do **detekcji konkretnego rodzaju cząstek DNA**, gdyż wykładnicze powielanie nastąpi tylko gdy obie nici DNA zawierają sekwencje komplementarne wobec pary starterów.

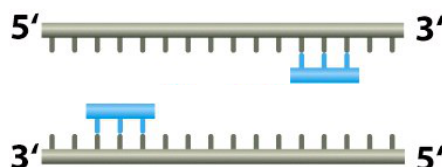
Brak miejsc komplementarnych wobec starterów → brak powielania



Startery komplementarne tylko do pojedynczej nici → brak powielania wykładniczego



Startery komplementarne do obu nici → powielanie wykładnicze



[T. A. Brown: *Genomy*. PWN, 2009]

Ta cecha PCR jest wykorzystywana m. in. w diagnostyce medycznej (do identyfikacji patogenów bakteryjnych lub wirusowych oraz przy badaniu mutacji) i kryminalistyce (do detekcji i identyfikacji DNA w materiale biologicznym z miejsca zbrodni).

Techniki manipulacji DNA: reakcja łańcuchowa polimerazy

Kluczowe znaczenie dla przebiegu PCR ma właściwy dobór syntetycznych starterów oligonukleotydowych, których zadaniem jest ograniczenie z obu stron fragmentu powielanej sekwencji DNA.

Podczas projektowania starterów należy uwzględnić następujące czynniki:

- długość w zakresie 18 – 30 nukleotydów;
- specyficzność wobec powielanej sekwencji;
- temperatura topnienia (rozdzielania się pasm DNA) w zakresie 50 – 60°C;
- procentowy udział guanin i cytozyn (tzw. *GC content*) w zakresie 45 – 55%;
- występowanie guaniny lub cytozyny blisko końca 3' (tzw. *GC clamp*);
- brak skłonności do tworzenia struktur przestrzennych (np. „spinki do włosów”);
- brak hybrydyzacji pomiędzy starterami.

Uczenie maszynowe w bioinformatyce

Wykład 3: genomika (sekwencjonowanie DNA)

Tymon Rubel

**Zakład Elektroniki Jądrowej i Medycznej
Instytut Radioelektroniki i Technik Multimedialnych PW**

Sekwencjonowanie DNA: metoda terminacji łańcucha

Sekwencjonowanie cząstki DNA to ustalanie kolejności tworzących ją nukleotydów.

Klasyczną techniką sekwencjonowania jest **metoda terminacji łańcucha** (nazywana również **metodą Sanger**, od nazwiska jej twórcy). W pojedynczej reakcji dostarcza ona informacji o sekwencji 500-700 nukleotydów z dokładnością 99,9%.

Metoda ta znalazła szerokie zastosowanie nie tylko w biologii molekularnej, ale także w medycynie i kryminalistyce. Umożliwiła również przeprowadzenie **Human Genome Project (HGP)**, w którym ustalona została sekwencja ludzkiego genomu (w roku 2000 opublikowano wstępną wersję, sukcesywnie uzupełnianą w kolejnych latach).



Frederick Sanger

Sekwencjonowanie DNA: metoda terminacji łańcucha

Metoda terminacji łańcucha ma wiele wariantów. Zamieszczony poniżej opis dotyczy jednego z nich: **sekwencjonowania cyklicznego ze znacznikami fluorescencyjnymi**.

Przebieg sekwencjonowania cyklicznego jest podobny do PCR:

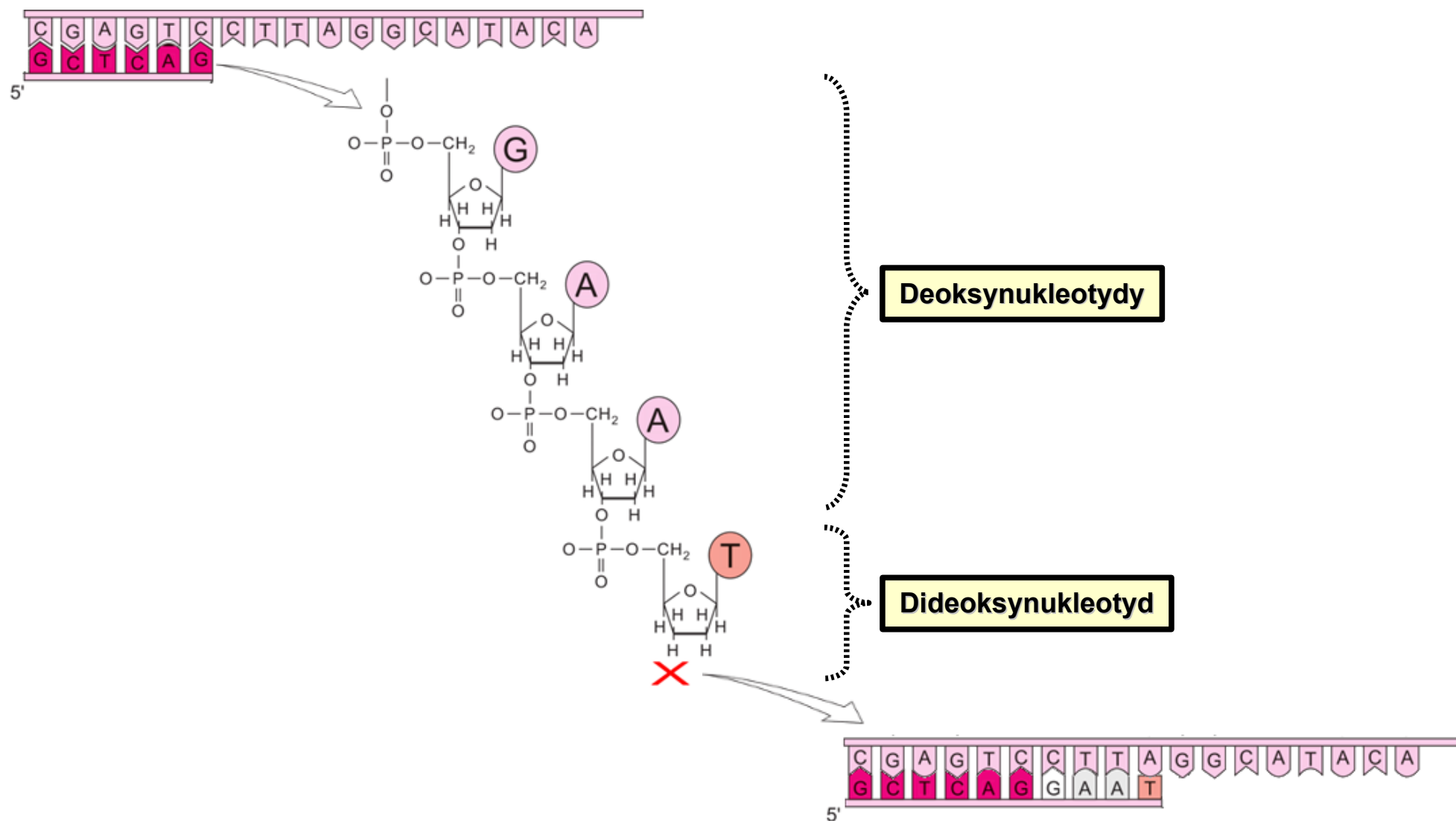
- polega na syntezie nici DNA komplementarnej do nici matrycowej;
- synteza jest katalizowana przez polimerazę DNA w obecności dużej ilości czterech rodzajów trifosforanów deoksyrybonukleotydów (**dATP**, **dGTP**, **dCTP**, **dTTP**);
- odbywa się w cyklach złożonych z faz denaturacji, dołączenia starterów i elongacji.

Różnicami wobec PCR są:

- użycie pojedynczego startera – oznacza to, że kopiowana jest tylko pojedyncza nić DNA, czyli produkty reakcji akumulują się w sposób liniowy, a nie wykładniczy;
- wprowadzenie do reakcji niewielkiej ilości czterech rodzajów trifosforanów dideoksyrybonukleotydów (**ddATP**, **ddGTP**, **ddCTP**, **ddTTP**) wyznakowanych czterema różnymi znacznikami fluorescencyjnymi.

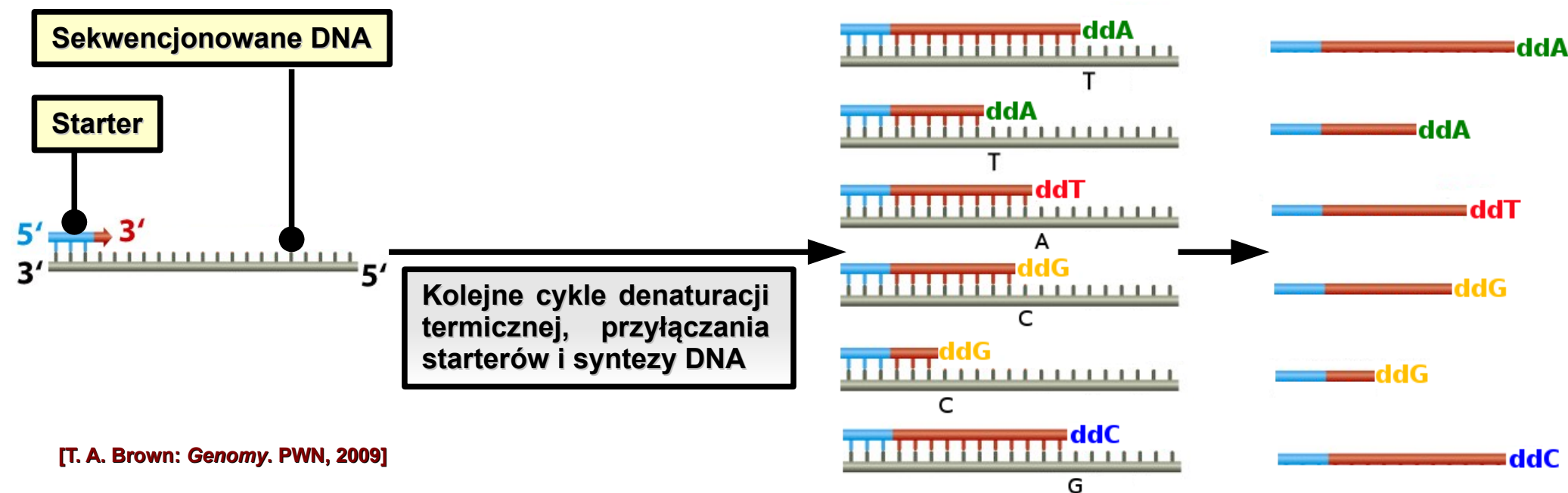
Sekwencjonowanie DNA: metoda terminacji łańcucha

W trakcie elongacji polimeraza nie odróżnia nukleotydów deoksy od dideoksy, jednak włączenie do nici DNA jednego z tych drugich uniemożliwia jej dalsze wydłużanie. W efekcie synteza kończy się szybciej niż w przypadku PCR a jej produkty zawierają wyznakowane nukleotidy na ostatniej pozycji.



Sekwencjonowanie DNA: metoda terminacji łańcucha

W rezultacie powtarzania cykli denaturacji, przyłączania starterów i syntezy DNA powstaje duża **pula cząsteczek o różnej długości, z których każda zakończona jest znakowanym dideoksynukleotydem** pozwalającym określić nukleotyd występujący na odpowiadającej mu pozycji w sekwencji nici stanowiącej matrycę.



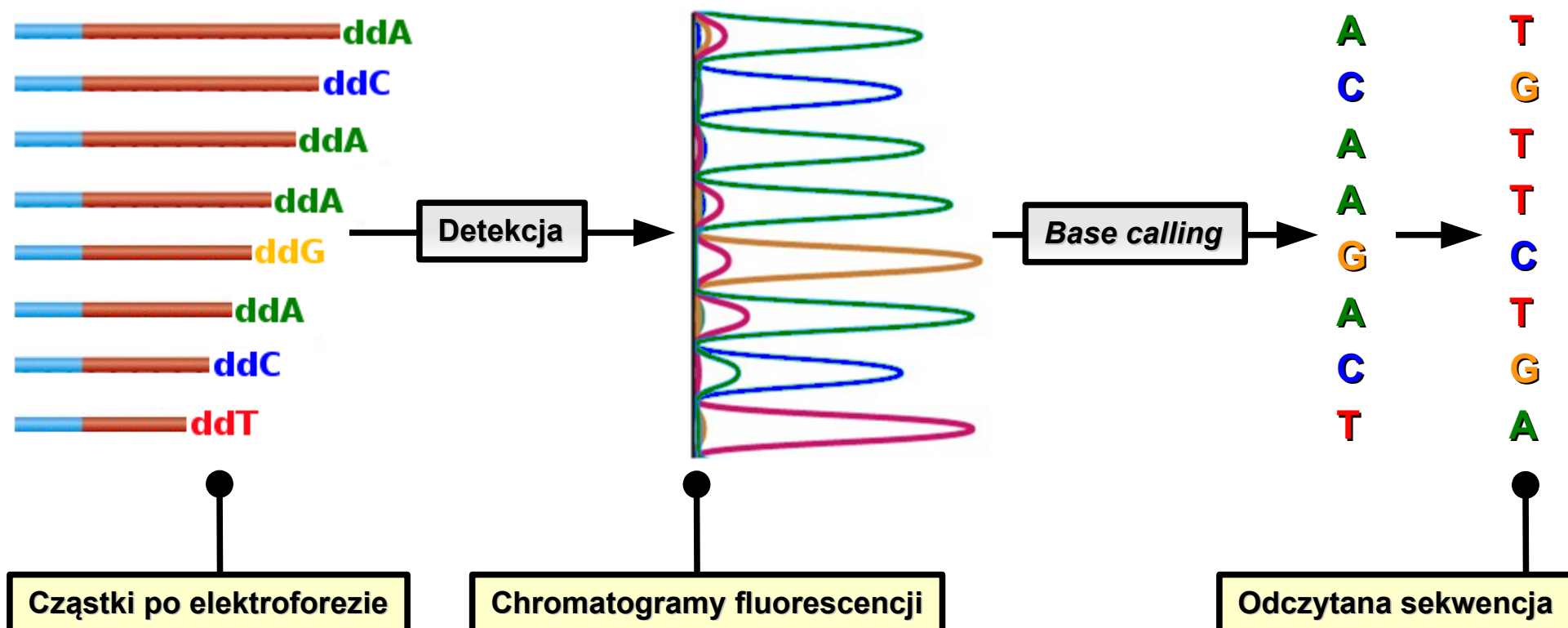
[T. A. Brown: *Genomy*. PWN, 2009]

Ustalenie kolejności nukleotydów w badanym DNA sprowadza się do identyfikacji dideoksynukleotydów znajdujących się na końcach produktów sekwencjonowania cyklicznego. Wymaga to wyodrębnienia grup cząstek o jednakowych długościach (czyli takich, których synteza została przerwana na tej samej pozycji łańcucha).

Sekwencjonowanie DNA: metoda terminacji łańcucha

Rozdział cząsteczek ze względu na ich długość odbywa się za pomocą **elektroforezy kapilarnej w żelu poliakrylamidowym**. Metoda ta ma wystarczającą rozdzielczość, aby rozdzielić cząstki DNA różniące długością się o jeden nukleotyd.

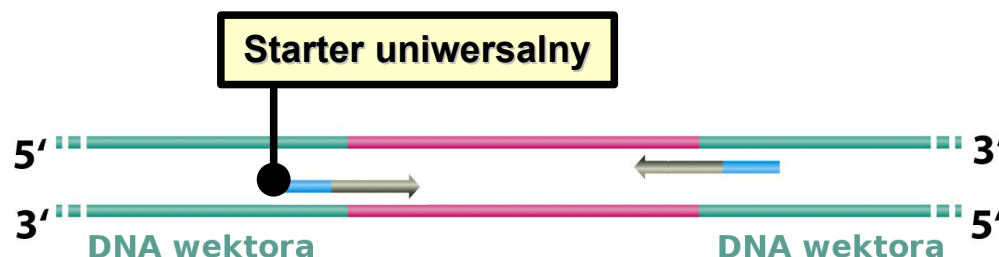
Po rozdziale następuje detekcja fluorescencji znaczników w kolejnych prążkach żelu, a zarejestrowane sygnały są podstawą dla identyfikacji sekwencji cząstki. Określenie nukleotydów na każdej z pozycji (tzw. **base calling**) odbywa się poprzez detekcję pików w **chromatogramach fluorescencji znaczników**.



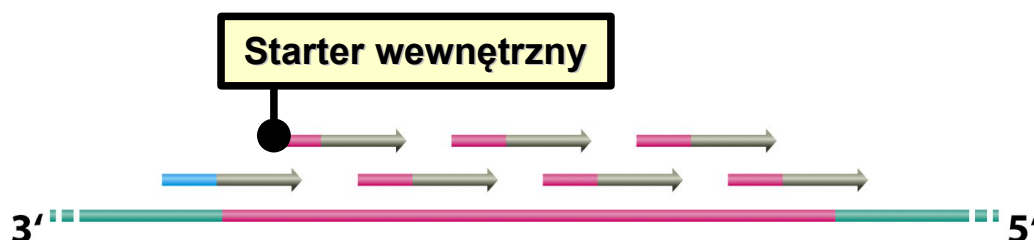
Sekwencjonowanie DNA: metoda terminacji łańcucha

Pojedynczy pomiar dostarcza informacji sekwencjach o długości do 1000 par zasad.

Dłuższe fragmenty można sekwencjonować z dwóch stron (mówi się wtedy o **parze odczytanych końców**). Używa się w tym celu pary starterów uniwersalnych, komplementarnych wobec DNA wektora do klonowania otaczającego sekwencjonowany fragment.



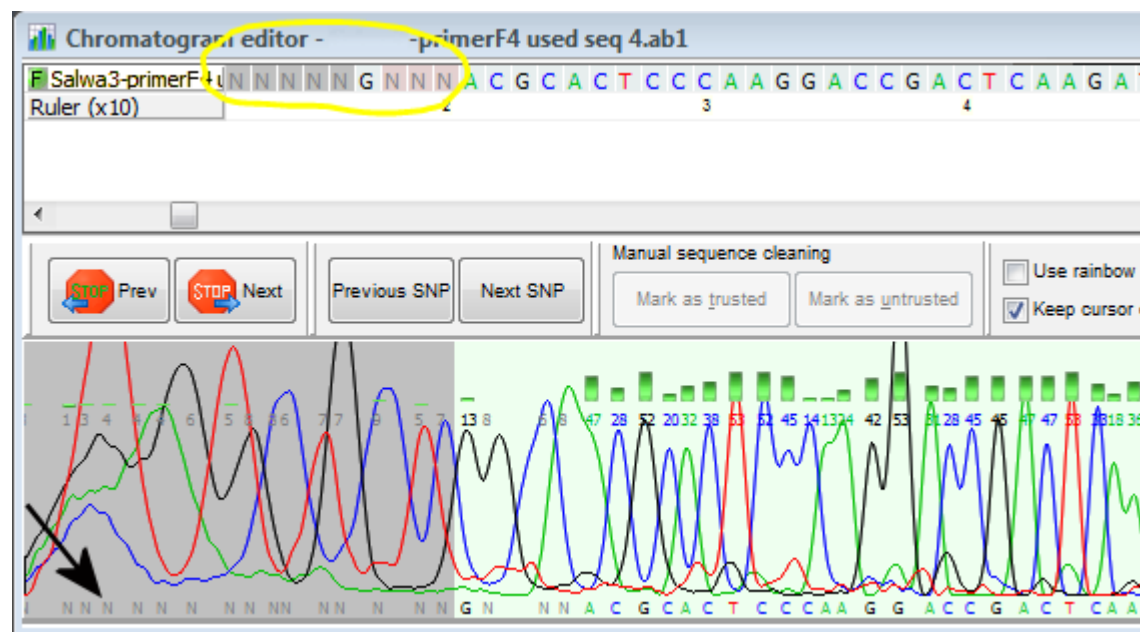
Możliwe jest również sekwencjonowanie kolejnych odcinków DNA przy użyciu **starterów wewnętrznych**, tworzonych na podstawie wcześniejszych odczytów.



W praktyce **sekwencjonowanie długich cząstek DNA odbywa się po uprzednim podzieleniu ich na krótsze odcinki**. Wiąże się to z koniecznością późniejszego **składania pełnej sekwencji** na podstawie odczytów pochodzących z fragmentów.

Sekwencjonowanie DNA: metoda terminacji łańcucha

W rzeczywistych sytuacjach można oczekiwać pewnej liczby błędnie odczytanych pozycji w sekwencjach. Wynikają one m. in. ze skończonej rozdzielczości rozdziału elektroforetycznego (szczególnie na początku i końcu sekwencji), ograniczeń detektora i występowania „przesłuchów” (*crosstalk*) pomiędzy znacznikami.



Jednak ciągły rozwój samej aparatury pomiarowej, jak i odczynników chemicznych spowodował, że nowoczesne sekwenatory Sangera charakteryzują się poziomem błędu nie większym niż $10^{-3}\%$. W efekcie metoda terminacji łańcucha jest wciąż najdokładniejszym z dostępnych sposobów sekwencjonowania.

Sekwencjonowanie DNA: techniki NGS

Metoda terminacji łańcucha – od momentu jej wprowadzenia pod koniec lat 70. XX w. do połowy poprzedniej dekady – była podstawową techniką sekwencjonowania DNA i w ogromnym stopniu przyczyniła się ona do rozwoju genomiki.

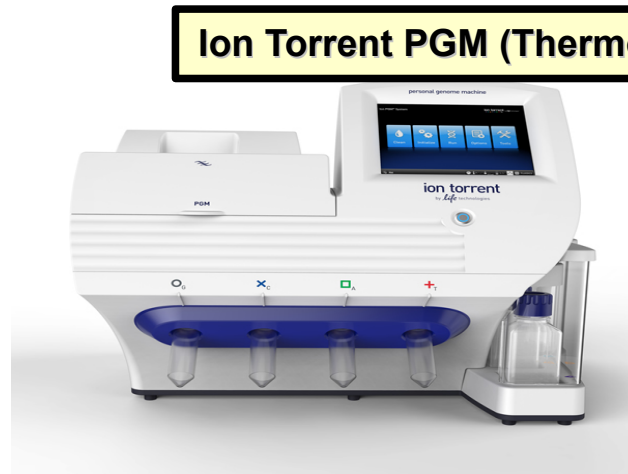
Pomimo swoich zalet jest to jednak metoda droga i powolna, dlatego też w ostatnich latach obserwuje się gwałtowny rozwój technik o większej wydajności, często określanых ogólną nazwą **sekwencjonowania drugiej (lub następnej) generacji (NGS – Next Generation Sequencing)**.

Solexa (Illumina)



©2010, Illumina Inc. All rights reserved.

Ion Torrent PGM (Thermo)



454 sequencing (Roche)



Sekwencjonowanie DNA: techniki NGS

W ogólności metody NGS opierają się na użyciu polimerazy do syntezy nici komplementarnej wobec sekwencjonowanego fragmentu DNA, ale technologie stosowane przez poszczególnych producentów mogą znacząco różnić się szczegółami tego procesu.

Techniki NGS posiadają jednak pewne cechy wspólne, wyraźnie odróżniające je od klasycznego sekwencjonowania metodą terminacji łańcucha:

- **reakcje odbywają się równolegle** dla bardzo dużej liczby (wielu milionów, a nawet miliardów) różnych fragmentów DNA;
- **w pojedynczej reakcji sekwencjonowane są krótkie odcinki DNA**, o długości od kilkudziesięciu do kilkuset par zasad. Oznacza to, że zawsze konieczna jest fragmentacja DNA i późniejsze składanie pełnej sekwencji na podstawie zbioru osobnych odczytów;
- **bezpośrednie dane pomiarowe charakteryzują się dużą złożonością** i znacznymi, dochodzącymi do terabajtów rozmiarami, a uzyskanie na ich podstawie sekwencji wymaga wieloetapowego procesu przetwarzania.

Sekwencjonowanie DNA: techniki NGS

Przykładowe techniki NGS

Technika Sangera

	Roche 454 GS FLX	Illumina HiSeq 2000	ABI Sanger 3730xl
Sequencing mechanism	Pyrosequencing	Sequencing by synthesis	Dideoxy chain termination
Read length	< 700bp	50-101bp	< 900bp
Accuracy	99.9%	98.0%	99.999%
Reads	1M	3G	—
Output data/run	0.7 Gb	600Gb	1.9~84Kb
Time/run	24 Hours	3~10 Days	20Mins~3Hours
Cost/million bases	\$10	\$0.07	\$2400

W pojedynczej reakcji sekwencjonowane są jedynie krótkie odcinki DNA.

Sekwencjonowanie DNA: techniki NGS

Przykładowe techniki NGS

Technika Sanger

	Roche 454 GS FLX	Illumina HiSeq 2000	ABI Sanger 3730xl
Sequencing mechanism	Pyrosequencing	Sequencing by synthesis	Dideoxy chain termination
Read length	< 700bp	50-101bp	< 900bp
Accuracy	99.9%	98.0%	99.999%
Reads	1M	3G	—
Output data/run	0.7Gb	600Gb	1.9~84Kb
Time/run	24 Hours	3~10 Days	20Mins~3Hours
Cost/million bases	\$10	\$0.07	\$2400

Jednocześnie wykonywanych jest wiele (miliony, a nawet miliardy) reakcji.

Sekwencjonowanie DNA: techniki NGS

Przykładowe techniki NGS

Technika Sanger

	Roche 454 GS FLX	Illumina HiSeq 2000	ABI Sanger 3730xl
Sequencing mechanism	Pyrosequencing	Sequencing by synthesis	Dideoxy chain termination
Read length	< 700bp	50-101bp	< 900bp
Accuracy	99.9%	98.0%	99.999%
Reads	1M	3G	—
Output data/run	0.7 Gb	600Gb	1.9~84Kb
Time/run	24 Hours	3~10 Days	20Mins~3Hours
Cost/million bases	\$10	\$0.07	\$2400

Koszt w przeliczeniu na liczbę odczytów jest znacząco niższy. Ta sama uwaga dotyczy czasu pomiaru.

Sekwencjonowanie DNA: techniki NGS

Przykładowe techniki NGS

Technika Sangera

	Roche 454 GS FLX	Illumina HiSeq 2000	ABI Sanger 3730xl
Sequencing mechanism	Pyrosequencing	Sequencing by synthesis	Dideoxy chain termination
Read length	< 700bp	50-101bp	< 900bp
Accuracy	99.9%	98.0%	99.999%
Reads	1M	3G	—
Output data/run	0.7 Gb	600Gb	1.9~84Kb
Time/run	24 Hours	3~10 Days	20Mins~3Hours
Cost/million bases	\$10	\$0.07	\$2400

Obecnie poziom błędu jest większy niż w przypadku sekwencjonowania Sangera.

Sekwencjonowanie DNA: techniki NGS

Przykładowe techniki NGS

Technika Sanger

	Roche 454 GS FLX	Illumina HiSeq 2000	ABI Sanger 3730xl
Sequencing mechanism	Pyrosequencing	Sequencing by synthesis	Dideoxy chain termination
Read length	< 700bp	50-101bp	< 900bp
Accuracy	99.9%	98.0%	99.999%
Reads	1M	3G	—
Output data/run	0.7 Gb	600Gb	1.9~84Kb
Time/run	24 Hours	3~10 Days	20Mins~3Hours
Cost/million bases	\$10	\$0.07	\$2400

Bezpośrednie dane pomiarowe cechują się dużymi rozmiarami i złożonością.

Sekwencjonowanie DNA: techniki NGS

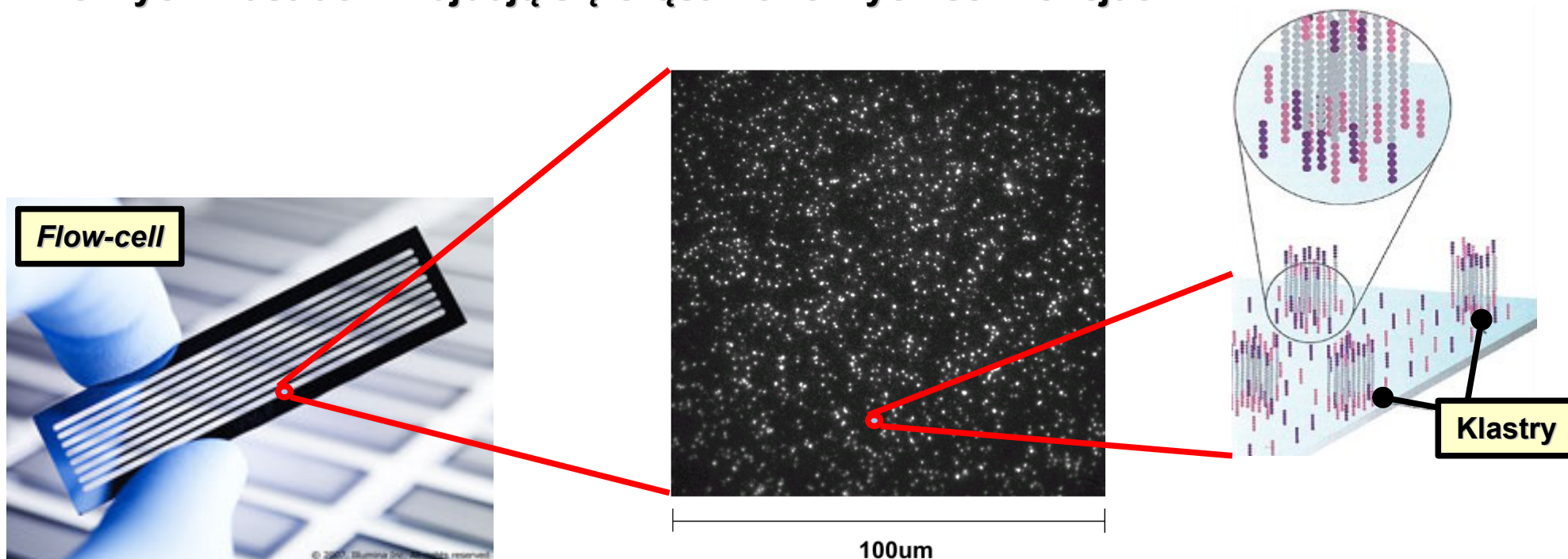
Personal Genome	Platform	Genomic template libraries	No. of reads (millions)	Read length (bases)	Base coverage (fold)	Assembly	Genome coverage (%) [*]	SNVs in millions (alignment tool)	No. of runs	Estimated cost (US\$)
J. Craig Venter	Automated Sanger	MP from BACs, fosmids & plasmids	31.9	800	7.5	<i>De novo</i>	N/A	3.21	>340,000	70,000,000
James D. Watson	Roche/454	Frag: 500 bp	93.2 ⁺	250 [§]	7.4	Aligned*	95	3.32 (BLAT)	234	1,000,000 [¶]
Yoruban male (NA18507)	Illumina/Solexa	93% MP: 200 bp	3,410 ⁺	35	40.6	Aligned*	99.9	3.83 (MAQ)	40	250,000 [¶]
		7% MP: 1.8 kb	271	35				4.14 (ELAND)		
Han Chinese male	Illumina/Solexa	66% Frag: 150–250 bp	1,921 ⁺	35	36	Aligned*	99.9	3.07 (SOAP)	35	500,000 [¶]
		34% MP: 135 bp & 440 bp	1,029	35						
Korean male (AK1)	Illumina/Solexa	21% Frag: 130 bp & 440 bp	393 ⁺	36	27.8	Aligned*	99.8	3.45 (GSNAP)	30	200,000 [¶]
		79% MP: 130 bp, 390 bp & 2.7 kb	1,156	36,88, 106						
Korean male (SJK)	Illumina/Solexa	MP: 100 bp, 200 bp & 300 bp	1,647 ⁺	35, 74	29.0	Aligned*	99.9	3.44 (MAQ)	15	250,000 ^{¶,¶}
Yoruban male (NA18507)	Life/APG	9% Frag: 100–500 bp	211 ⁺	50	17.9	Aligned*	98.6	3.87 (Corona-lite)	9.5	60,000 ^{¶,¶,¶}
		91% MP: 600–3,500 bp	2,075 ⁺	25, 50						
Stephen R. Quake	Helicos BioSciences	Frag: 100–500 bp	2,725 ⁺	32 [§]	28	Aligned*	90	2.81 (IndexDP)	4	48,000 [¶]
AML female	Illumina/Solexa	Frag: 150–200 bp ^{††}	2,730 ^{†,††}	32	32.7	Aligned*	91	3.81 ^{††} (MAQ)	98	1,600,000 ^{¶¶¶}
		Frag: 150–200 bp ^{§§}	1,081 ^{†,§§}	35	13.9		83	2.92 ^{§§} (MAQ)	34	
AML male	Illumina/Solexa	MP: 200–250 bp ^{††}	1,620 ^{†,††}	35	23.3	Aligned*	98.5	3.46 ^{††} (MAQ)	16.5	500,000 ^{¶¶}
		MP: 200–250 bp ^{§§}	1,351 ^{†,§§}	50	21.3		97.4	3.45 ^{§§} (MAQ)	13.1	
James R. Lupski CMT male	Life/APG	16% Frag: 100–500 bp	238 ⁺	35	29.6	Aligned*	99.8	3.42 (Corona-lite)	3	75,000 ^{¶,¶¶}
		84% MP: 600–3,500 bp	1,211 ⁺	25, 50						

[Michael L. Metzker: Sequencing technologies – the next generation. Nature Reviews Genetics, 2010, 11, pp. 31-46.]

Sekwencjonowanie DNA: techniki NGS (Illumina)

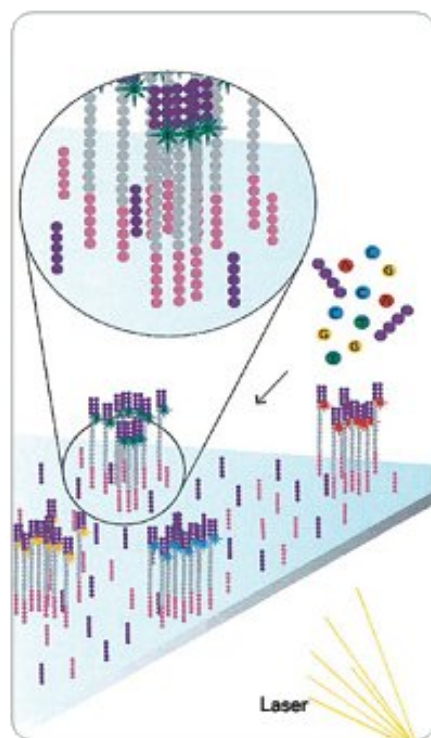
Jedną z najpowszechniej wykorzystywanych technik sekwencjonowania NG jest ta opracowana przez firmę **Solexa** (obecnie **Illumina**). Pozwala ona określić sekwencje odcinków o długości 50-100 bp z jednego lub obu końców badanej cząstki.

Przed rozpoczęciem sekwencjonowania cząstki DNA są fragmentowane za pomocą ultradźwięków, a powstałe fragmenty są denaturowane i wiązane jednym końcem z powierzchnią specjalnie przygotowanej celki (**flow-cell**). Po związaniu poddawane są one procesowi **amplifikacji na podłożu**, w wyniku którego powstaje około 1000 kopii każdego z fragmentów. Kopie te tworzą obszary zwane klastrami (**clusters**). W ramach tego samego klastra fragmenty DNA mają takie same sekwencje, natomiast w różnych klastrach znajdują się cząstki o różnych sekwencjach.

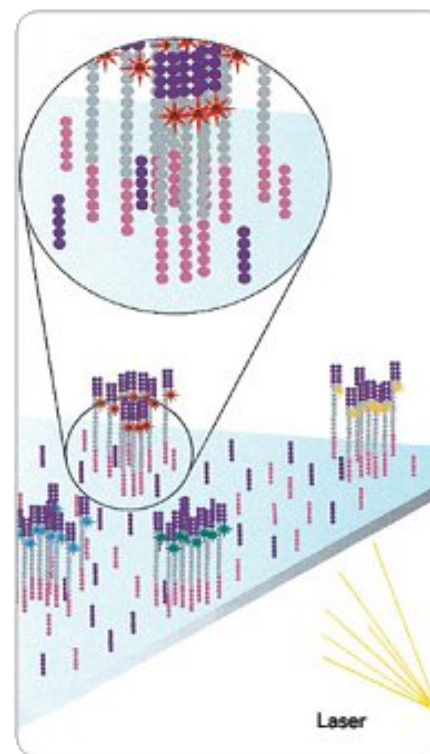


Sekwencjonowanie DNA: techniki NGS (Illumina)

Opisane powyżej **cykle pomiarowe wykonywane są równolegle dla wszystkich sekwencji z każdego klastra**, co pozwala na jednoczesne sekwencjonowanie nawet wielu milionów fragmentów DNA.



Cykl 1



Cykl 2



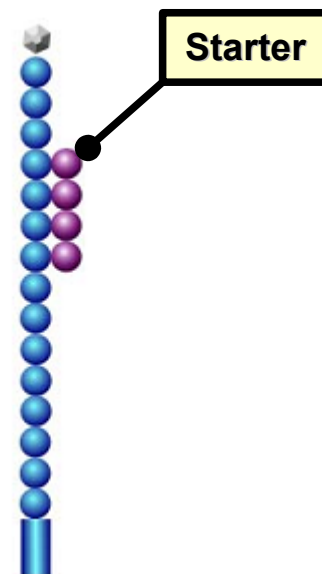
Sekwencjonowanie DNA: techniki NGS (Illumina)

W procesie sekwencjonowania biorą udział: **polimeraza DNA** oraz **deoksynukleotydy połączone ze znacznikami fluorescencyjnymi**, w taki sposób aby każdej „literze alfabetu” (**A**, **C**, **G**, **T**) odpowiadał znacznik o innej długości fali emitowanego światła.

Wszystkie biorące udział w reakcji nukleotydy są zmodyfikowane chemicznie, tak aby blokowały wydłużanie syntetyzowanej nici. Jednak w odróżnieniu od metody terminacji łańcucha, w tym przypadku modyfikacja ta może zostać usunięta.

Reakcja zaczyna się od przyłączenia starterów, po którym następują kolejne cykle, złożone z następujących kroków:

- wprowadzenie do roztworu znakowanych nukleotydów – jeden z nich (komplementarny wobec pierwszego niesparowanego nukleotydu matrycy) zostanie dołączony do syntezowanej nici, jednocześnie blokując możliwość jej dalszego wydłużania;
- wzbudzenie znacznika fluorescencyjnego światłem lasera i rejestracja pochodzącego od niego sygnału;
- usunięcie znacznika i modyfikacji terminującej, co umożliwia rozpoczęcie nowego cyklu.



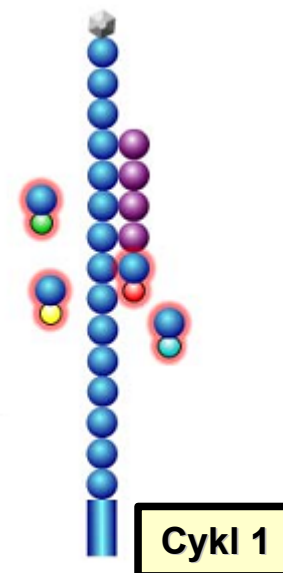
Sekwencjonowanie DNA: techniki NGS (Illumina)

W procesie sekwencjonowania biorą udział: **polimeraza DNA** oraz **deoksynukleotydy połączone ze znacznikami fluorescencyjnymi**, w taki sposób aby każdej „literze alfabetu” (**A**, **C**, **G**, **T**) odpowiadał znacznik o innej długości fali emitowanego światła.

Wszystkie biorące udział w reakcji nukleotydy są zmodyfikowane chemicznie, tak aby blokowały wydłużanie syntetyzowanej nici. Jednak w odróżnieniu od metody terminacji łańcucha, w tym przypadku modyfikacja ta może zostać usunięta.

Reakcja zaczyna się od przyłączenia starterów, po którym następują kolejne cykle, złożone z następujących kroków:

- wprowadzenie do roztworu znakowanych nukleotydów – jeden z nich (komplementarny wobec pierwszego niesparowanego nukleotydu matrycy) zostanie dołączony do syntezowanej nici, jednocześnie blokując możliwość jej dalszego wydłużania;
- wzbudzenie znacznika fluorescencyjnego światłem lasera i rejestracja pochodzącego od niego sygnału;
- usunięcie znacznika i modyfikacji terminującej, co umożliwia rozpoczęcie nowego cyklu.



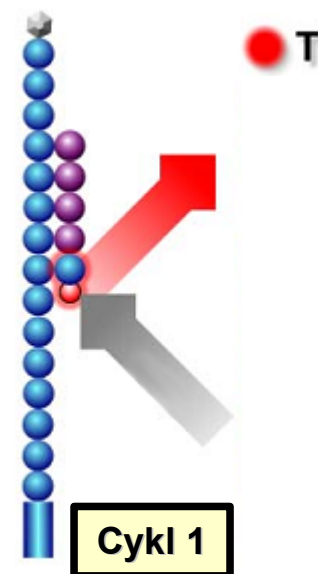
Sekwencjonowanie DNA: techniki NGS (Illumina)

W procesie sekwencjonowania biorą udział: **polimeraza DNA** oraz **deoksynukleotydy połączone ze znacznikami fluorescencyjnymi**, w taki sposób aby każdej „literze alfabetu” (**A**, **C**, **G**, **T**) odpowiadał znacznik o innej długości fali emitowanego światła.

Wszystkie biorące udział w reakcji nukleotydy są zmodyfikowane chemicznie, tak aby blokowały wydłużanie syntetyzowanej nici. Jednak w odróżnieniu od metody terminacji łańcucha, w tym przypadku modyfikacja ta może zostać usunięta.

Reakcja zaczyna się od przyłączenia starterów, po którym następują kolejne cykle, złożone z następujących kroków:

- wprowadzenie do roztworu znakowanych nukleotydów – jeden z nich (komplementarny wobec pierwszego niesparowanego nukleotydu matrycy) zostanie dołączony do syntezowanej nici, jednocześnie blokując możliwość jej dalszego wydłużania;
- wzbudzenie znacznika fluorescencyjnego światłem lasera i rejestracja pochodzącego od niego sygnału;
- usunięcie znacznika i modyfikacji terminującej, co umożliwia rozpoczęcie nowego cyklu.



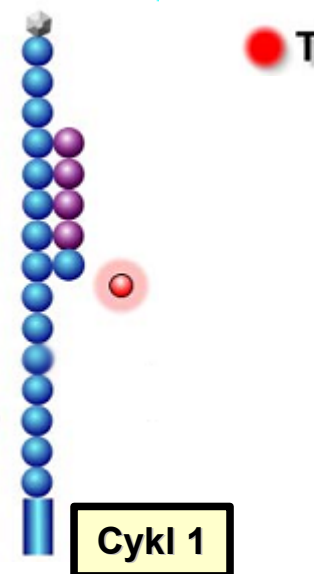
Sekwencjonowanie DNA: techniki NGS (Illumina)

W procesie sekwencjonowania biorą udział: **polimeraza DNA** oraz **deoksynukleotydy połączone ze znacznikami fluorescencyjnymi**, w taki sposób aby każdej „literze alfabetu” (**A**, **C**, **G**, **T**) odpowiadał znacznik o innej długości fali emitowanego światła.

Wszystkie biorące udział w reakcji nukleotydy są zmodyfikowane chemicznie, tak aby blokowały wydłużanie syntetyzowanej nici. Jednak w odróżnieniu od metody terminacji łańcucha, w tym przypadku modyfikacja ta może zostać usunięta.

Reakcja zaczyna się od przyłączenia starterów, po którym następują kolejne cykle, złożone z następujących kroków:

- wprowadzenie do roztworu znakowanych nukleotydów – jeden z nich (komplementarny wobec pierwszego niesparowanego nukleotydu matrycy) zostanie dołączony do syntezowanej nici, jednocześnie blokując możliwość jej dalszego wydłużania;
- wzbudzenie znacznika fluorescencyjnego światłem lasera i rejestracja pochodzącego od niego sygnału;
- usunięcie znacznika i modyfikacji terminującej, co umożliwia rozpoczęcie nowego cyklu.



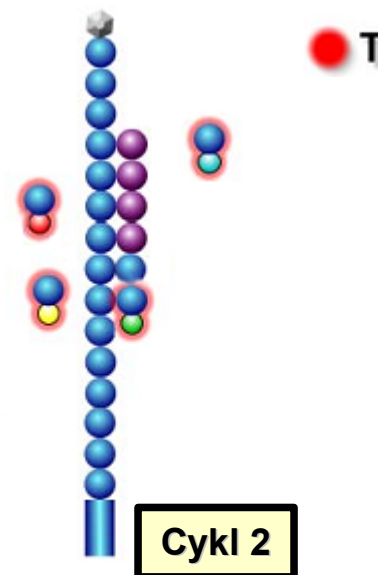
Sekwencjonowanie DNA: techniki NGS (Illumina)

W procesie sekwencjonowania biorą udział: **polimeraza DNA** oraz **deoksynukleotydy połączone ze znacznikami fluorescencyjnymi**, w taki sposób aby każdej „literze alfabetu” (A, C, G, T) odpowiadał znacznik o innej długości fali emitowanego światła.

Wszystkie biorące udział w reakcji nukleotydy są zmodyfikowane chemicznie, tak aby blokowały wydłużanie syntetyzowanej nici. Jednak w odróżnieniu od metody terminacji łańcucha, w tym przypadku modyfikacja ta może zostać usunięta.

Reakcja zaczyna się od przyłączenia starterów, po którym następują kolejne cykle, złożone z następujących kroków:

- wprowadzenie do roztworu znakowanych nukleotydów – jeden z nich (komplementarny wobec pierwszego niesparowanego nukleotydu matrycy) zostanie dołączony do syntezowanej nici, jednocześnie blokując możliwość jej dalszego wydłużania;
- wzbudzenie znacznika fluorescencyjnego światłem lasera i rejestracja pochodzącego od niego sygnału;
- usunięcie znacznika i modyfikacji terminującej, co umożliwia rozpoczęcie nowego cyklu.



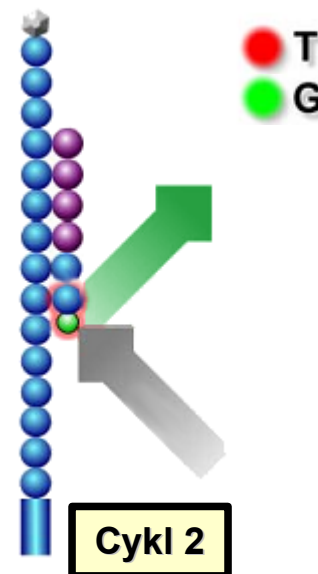
Sekwencjonowanie DNA: techniki NGS (Illumina)

W procesie sekwencjonowania biorą udział: **polimeraza DNA** oraz **deoksynukleotydy połączone ze znacznikami fluorescencyjnymi**, w taki sposób aby każdej „literze alfabetu” (**A**, **C**, **G**, **T**) odpowiadał znacznik o innej długości fali emitowanego światła.

Wszystkie biorące udział w reakcji nukleotydy są zmodyfikowane chemicznie, tak aby blokowały wydłużanie syntetyzowanej nici. Jednak w odróżnieniu od metody terminacji łańcucha, w tym przypadku modyfikacja ta może zostać usunięta.

Reakcja zaczyna się od przyłączenia starterów, po którym następują kolejne cykle, złożone z następujących kroków:

- wprowadzenie do roztworu znakowanych nukleotydów – jeden z nich (komplementarny wobec pierwszego niesparowanego nukleotydu matrycy) zostanie dołączony do syntezowanej nici, jednocześnie blokując możliwość jej dalszego wydłużania;
- wzbudzenie znacznika fluorescencyjnego światłem lasera i rejestracja pochodzącego od niego sygnału;
- usunięcie znacznika i modyfikacji terminującej, co umożliwia rozpoczęcie nowego cyklu.



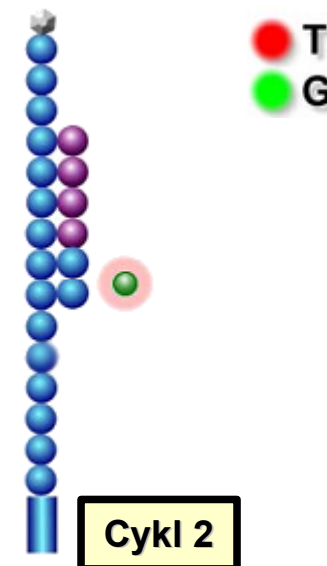
Sekwencjonowanie DNA: techniki NGS (Illumina)

W procesie sekwencjonowania biorą udział: **polimeraza DNA** oraz **deoksynukleotydy połączone ze znacznikami fluorescencyjnymi**, w taki sposób aby każdej „literze alfabetu” (**A**, **C**, **G**, **T**) odpowiadał znacznik o innej długości fali emitowanego światła.

Wszystkie biorące udział w reakcji nukleotydy są zmodyfikowane chemicznie, tak aby blokowały wydłużanie syntetyzowanej nici. Jednak w odróżnieniu od metody terminacji łańcucha, w tym przypadku modyfikacja ta może zostać usunięta.

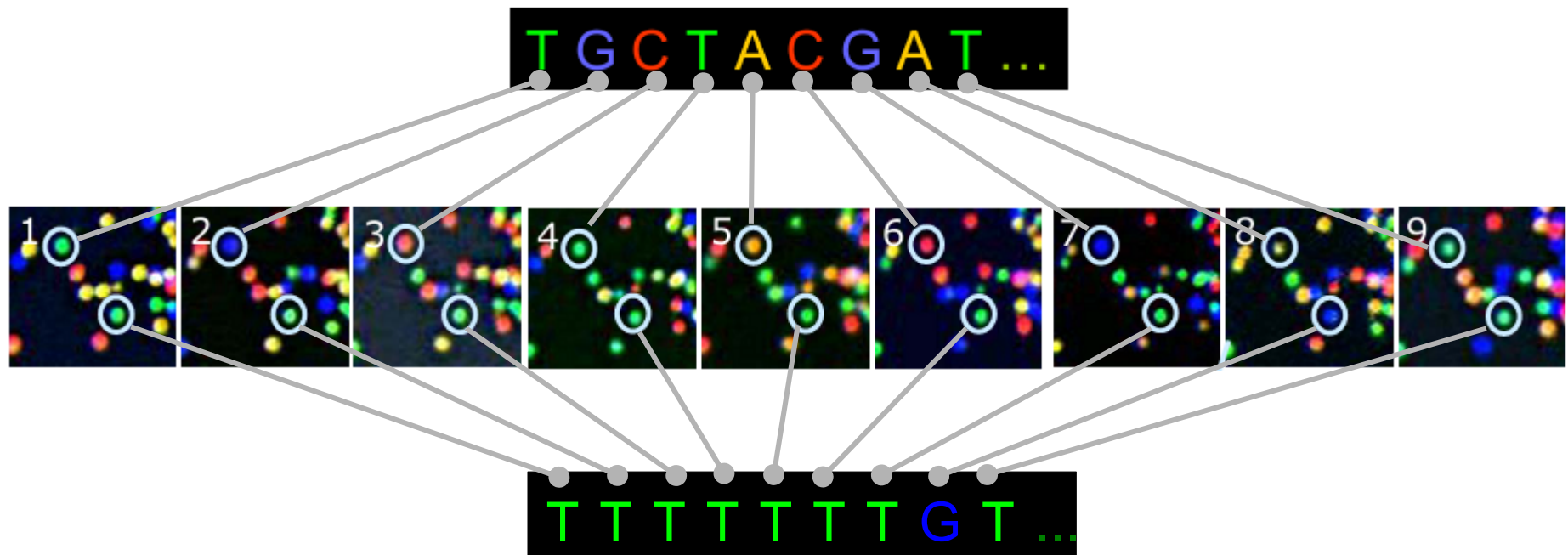
Reakcja zaczyna się od przyłączenia starterów, po którym następują kolejne cykle, złożone z następujących kroków:

- wprowadzenie do roztworu znakowanych nukleotydów – jeden z nich (komplementarny wobec pierwszego niesparowanego nukleotydu matrycy) zostanie dołączony do syntezowanej nici, jednocześnie blokując możliwość jej dalszego wydłużania;
- wzbudzenie znacznika fluorescencyjnego światłem lasera i rejestracja pochodzącego od niego sygnału;
- usunięcie znacznika i modyfikacji terminującej, co umożliwia rozpoczęcie nowego cyklu.



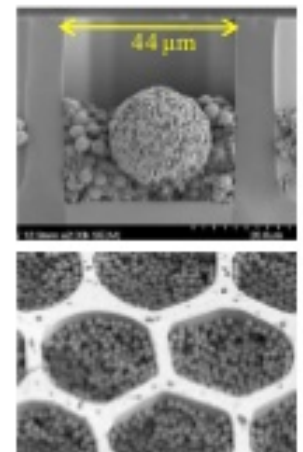
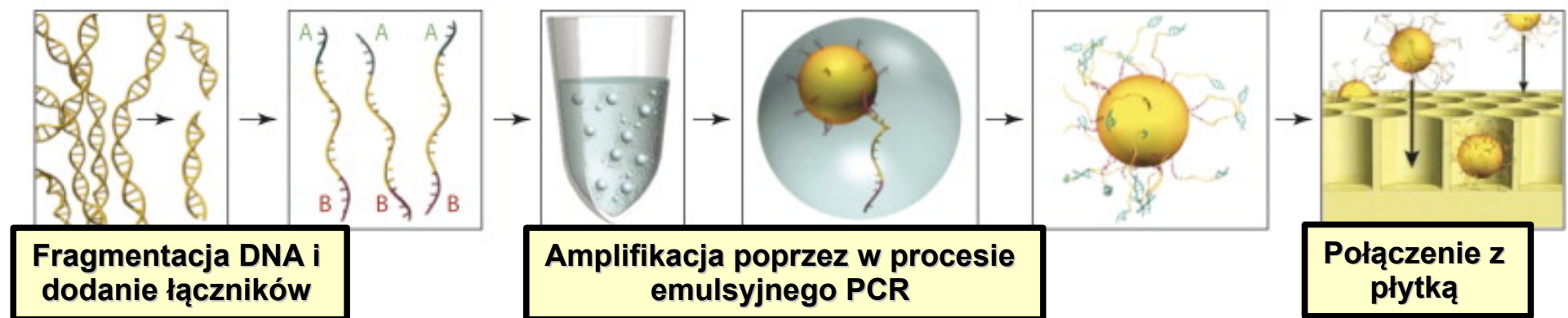
Sekwencjonowanie DNA: techniki NGS (Illumina)

Bezpośrednie dane pomiarowe mają postać serii obrazów, zarejestrowanych przez kamerę po kolejnych cyklach dodawania nukleotydów. *Base calling*, czyli określanie sekwencji nukleotydów cząstek DNA znajdujących się w poszczególnych klastrach wymaga detekcji odpowiadających im pikseli i rozpoznania rodzaju znacznika.



Sekwencjonowanie DNA: techniki NGS (pirosekwencjonowanie)

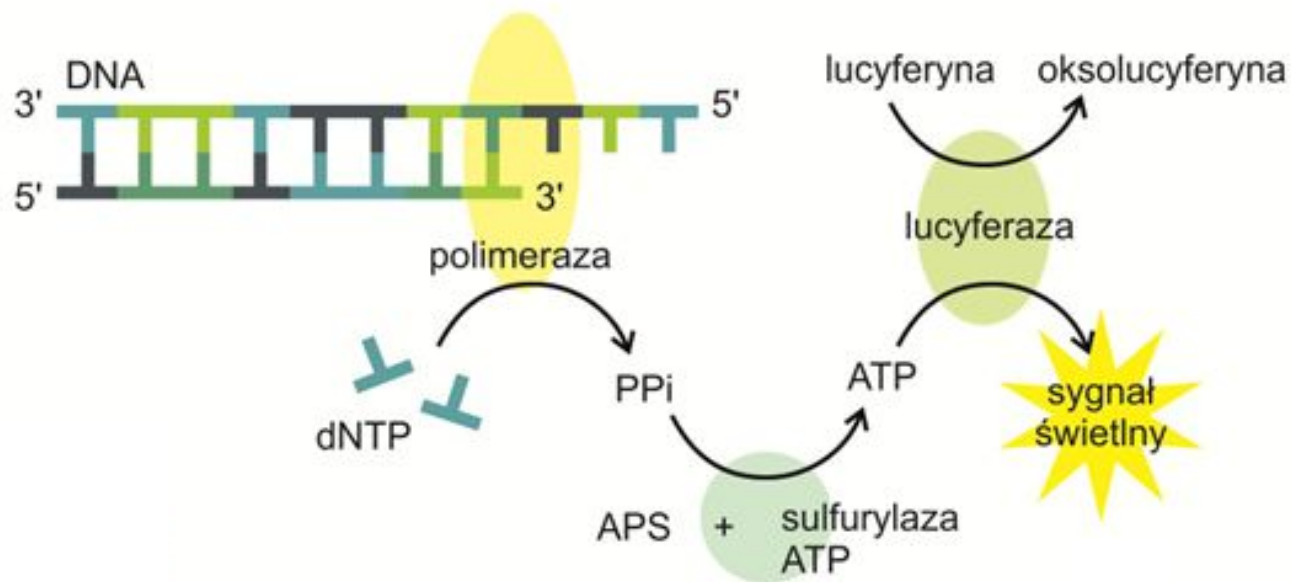
W **pirosekwencjonowaniu** (używany w urządzeniach firmy **454 Life Sciences**, obecnie **Roche**) fragmenty DNA są w pierwszej kolejności **amplifikowane w procesie zwanym emulsyjnym PCR (*Emulsion PCR – emPCR*)**, a następnie unieruchamiane na powierzchni odpowiednio przygotowanej płytki.



Sekwencjonowanie DNA: techniki NGS (pirosekwencjonowanie)

Sekwencjonowanie opiera się na detekcji pirofosforanu (PPi – cząstka złożona z dwóch grup fosforanowych) uwalnianego podczas przyłączania przez polimerazę kolejnego deoksynukleotydu do syntetyzowanej nici DNA.

Uwolnienie pirofosforanu wykrywane jest po jego konwersji na błysk świetlny, generowany w wyniku działania dwóch enzymów: sulfurylasy ATP (zamienia PPi na ATP) i lucyferazy (konwertuje ATP na błysk światła widzialnego).

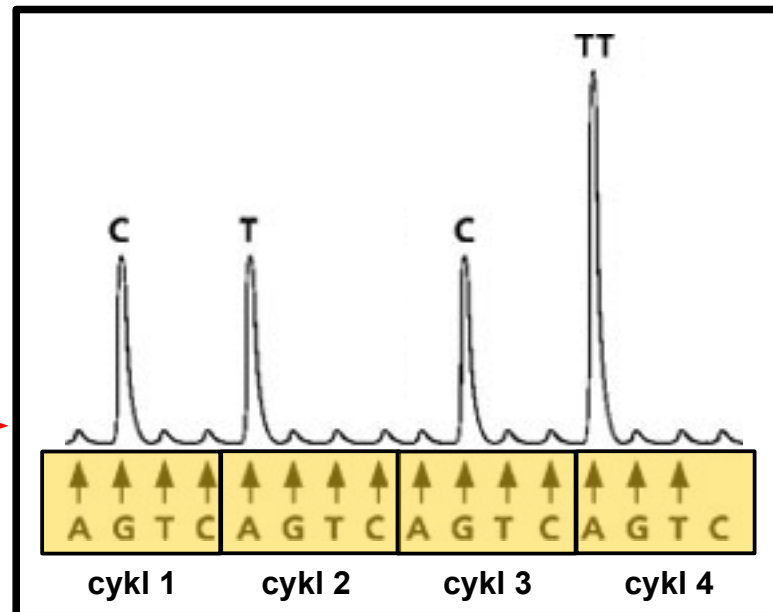
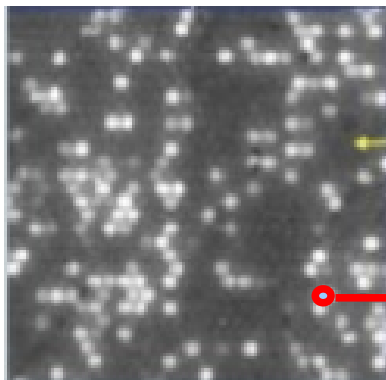


Sekwencjonowanie DNA: techniki NGS (pirosekwencjonowanie)

Proces sekwencjonowania odbywa się w **cyklach, z których każdy składa się z czterech faz**: w każdej z faz do reakcji wprowadzany jest jeden z czterech rodzajów deoksynukleotydów (jednocześnie usuwany jest nukleotyd z poprzedniej fazy).

O rodzaju deoksynukleotydu dołączonego przez polimerazę do nici DNA można wnioskować na podstawie fazy cyklu, w której nastąpił błysk. Liczbę włączonych nukleotydów (czyli liczbę kolejnych pozycji w sekwencji, na których dany nukleotyd występuje) określa intensywność błysku.

Błyski z każdego obszaru płytki w kolejnych cyklach są rejestrowane przez kamerę, a następnie zamieniane na wykresy (pirogramy), na podstawie których możliwe jest wyznaczenie sekwencji.



Sekwencjonowanie DNA: techniki NGS (IonTorrent)

Ogólny schemat sekwencjonowania w urządzeniach firmy **IonTorrent** jest podobny, do tego używanego w przedstawionym wcześniej pirosekwencjonowaniu:

- fragmenty DNA są amplifikowane i unieruchamiane w dołkach wytworzonych w krzemowym podłożu płytki;
- sekwencjonowanie odbywa się w cyklach, których kolejne cztery fazy odpowiadają wprowadzeniu do reakcji różnych rodzajów deoksynukleotydów;
- podstawową różnicą jest fakt, że **przyłączenie DNA kolejnych deoksynukleotydów (jednego lub kilku kolejnych) jest wykrywane na podstawie uwolnienia jonu (lub kilku jonów) wodoru H^+** . Detekcji podlegają sygnały elektryczne generowane w czujnikach ISFET na skutek pojawienia się w dołkach płytki naładowanych jonów.

