

**Materiały do ćwiczeń z przedmiotu
Genetyka z inżynierią genetyczną D - blok**

Inżynieria Genetyczna

ćw. 4

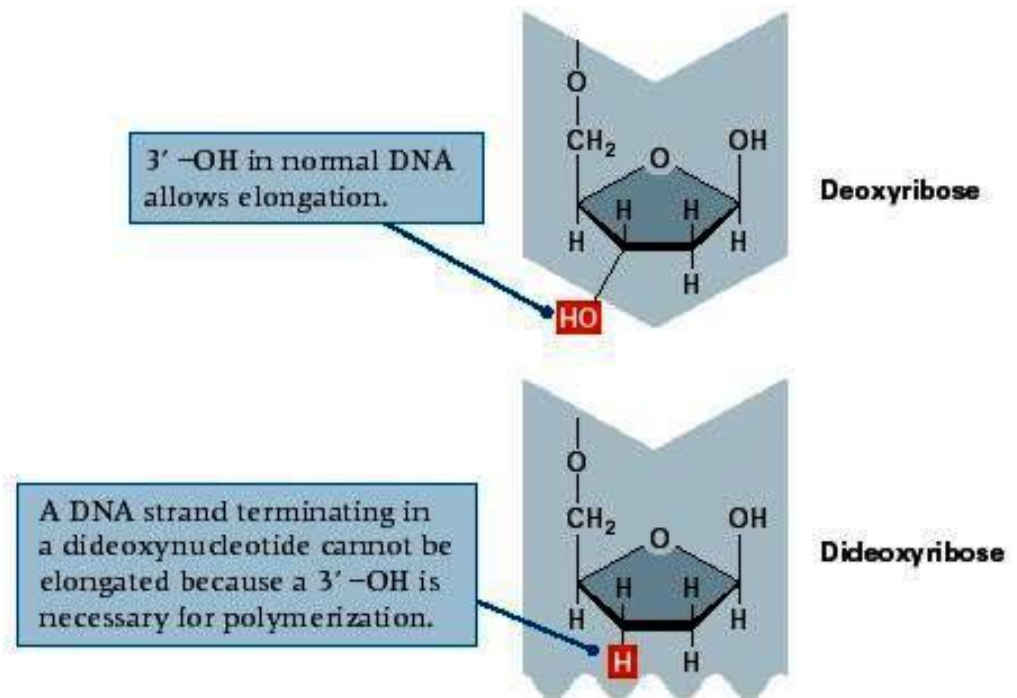
**Instytut Genetyki i Biotechnologii, Wydział Biologii,
Uniwersytet Warszawski, rok akad. 2019/2020**

*Prosimy o nierozpowszechnianie materiałów bez
naszej zgody*

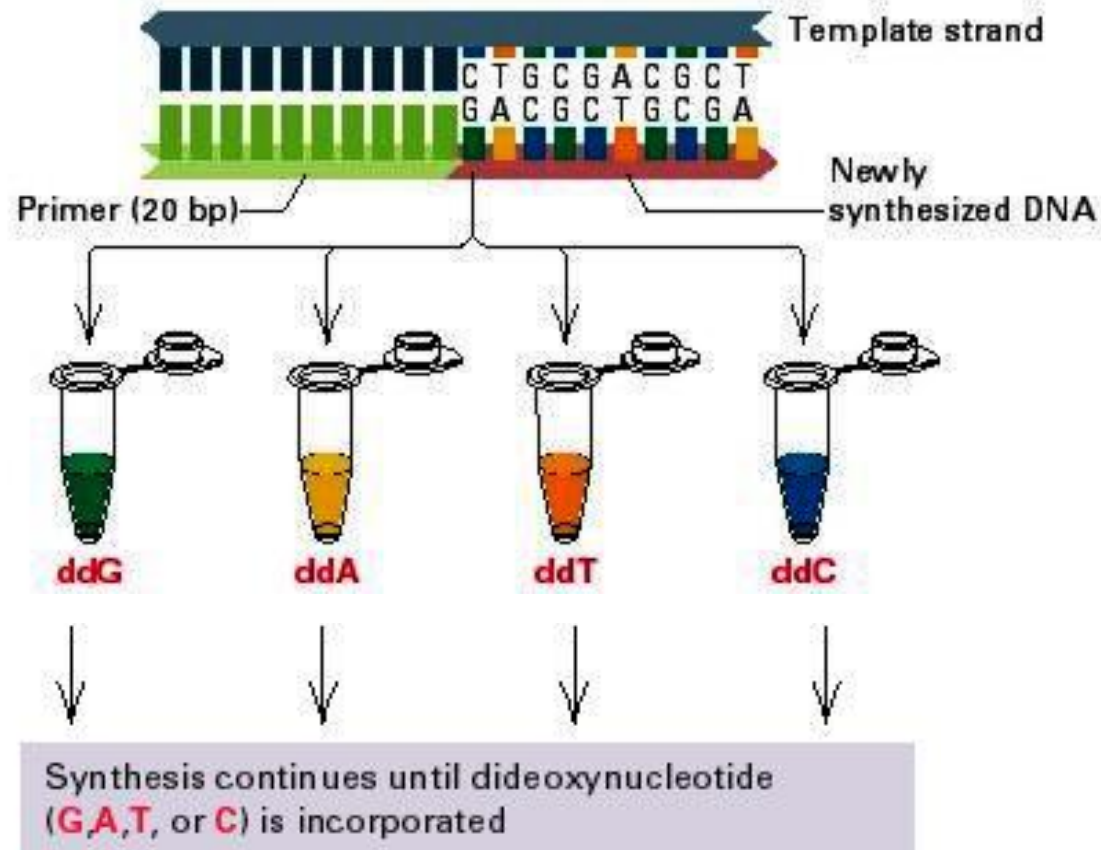
Sekwencjonowanie DNA

Sekwencjonowanie DNA: metoda Sanger

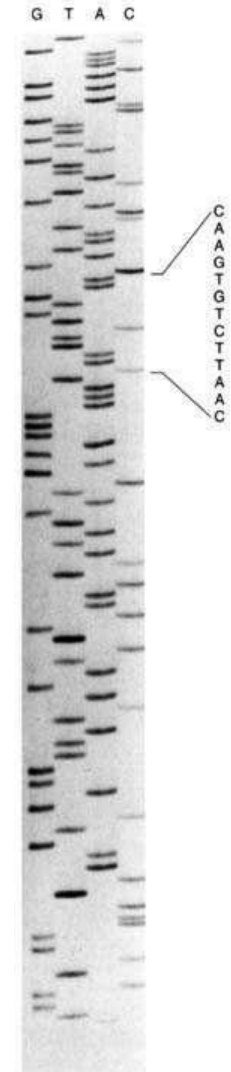
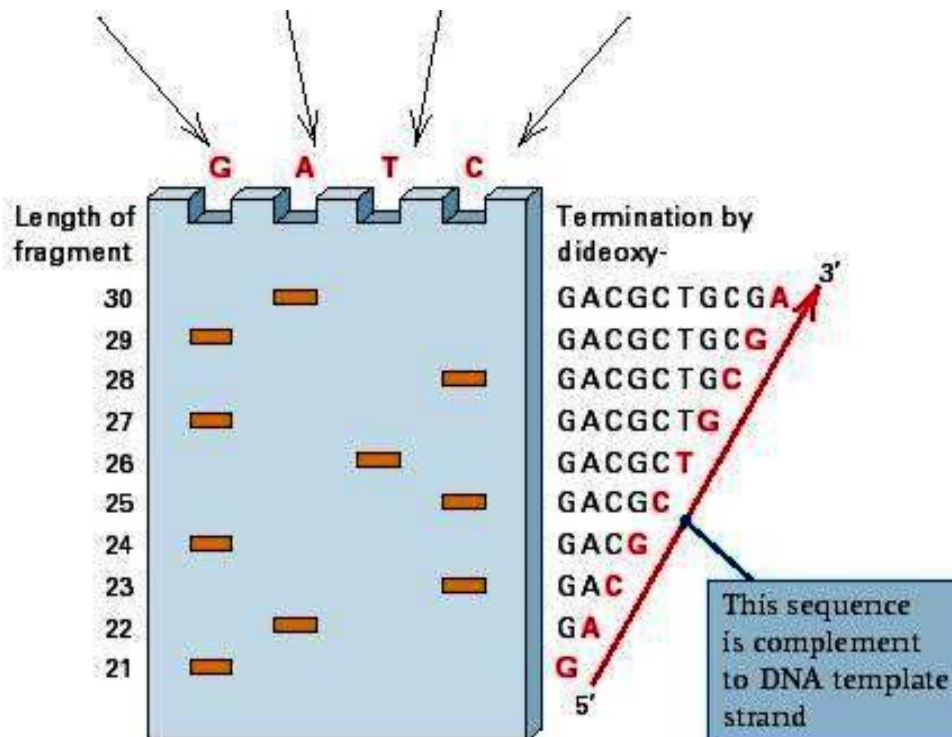
Dideoksynukleotydy
(analogi nukleotydów
bez grupy –OH w pozycji 3')
powodują przedwczesną
terminację syntezy DNA



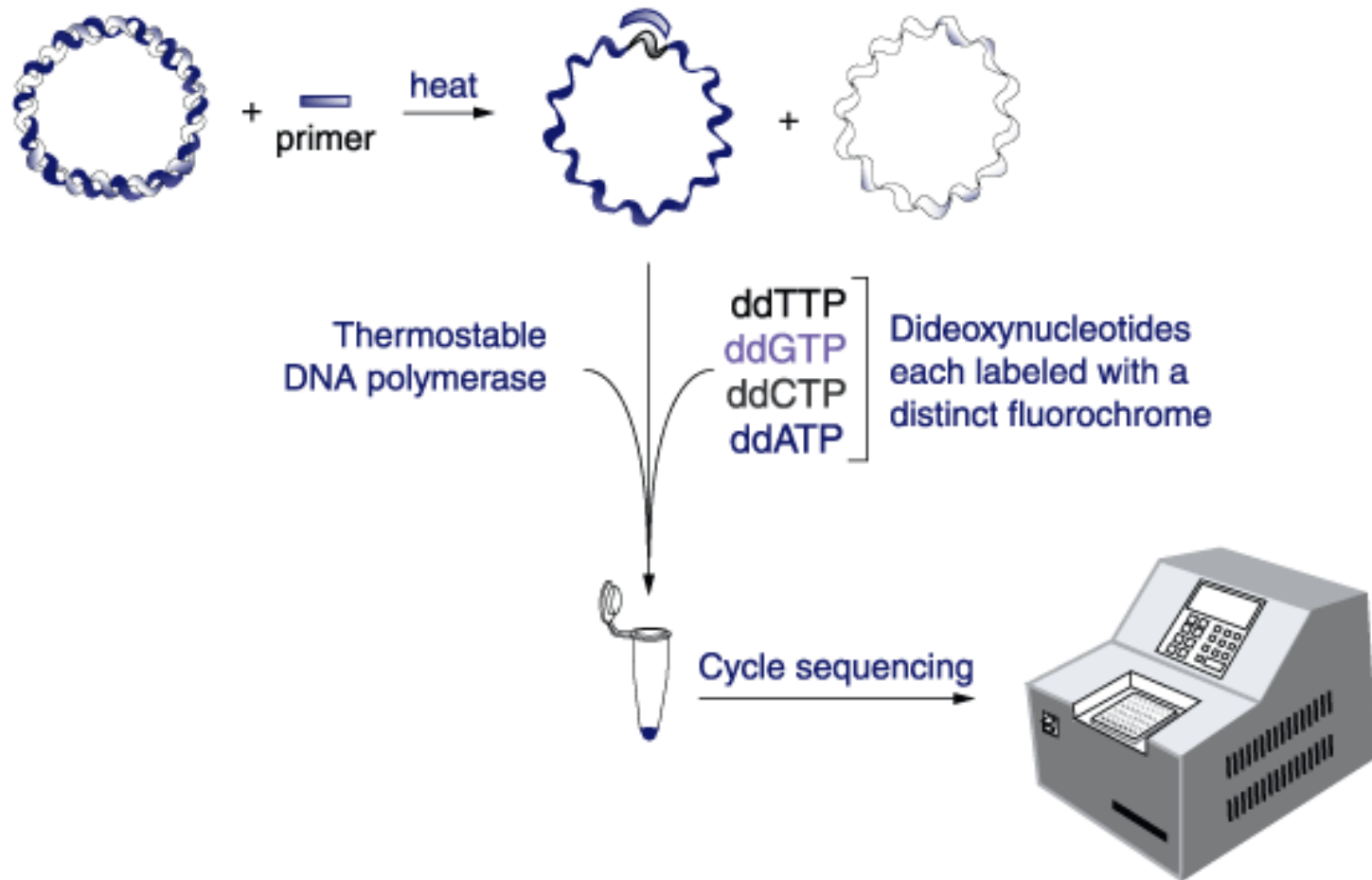
Sekwencjonowanie DNA: metoda Sanger



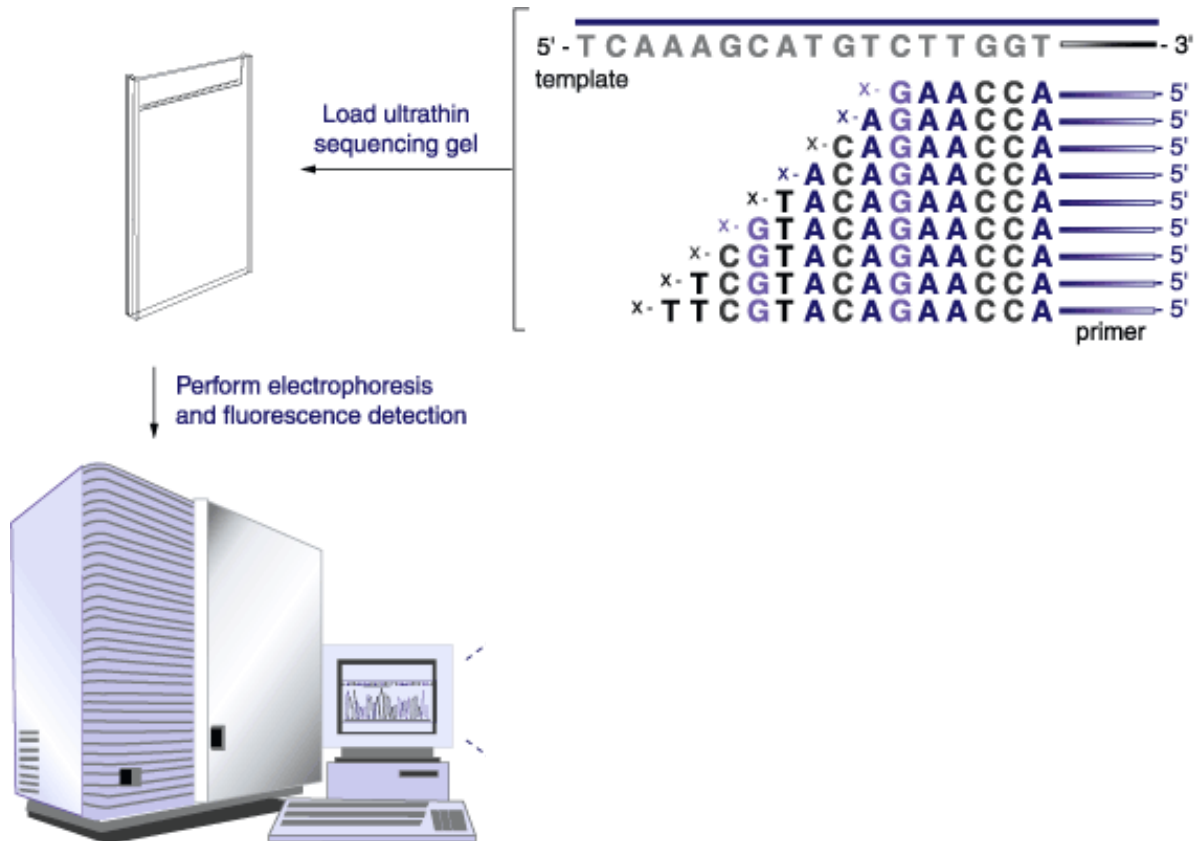
Sekwencjonowanie DNA: metoda Sangera



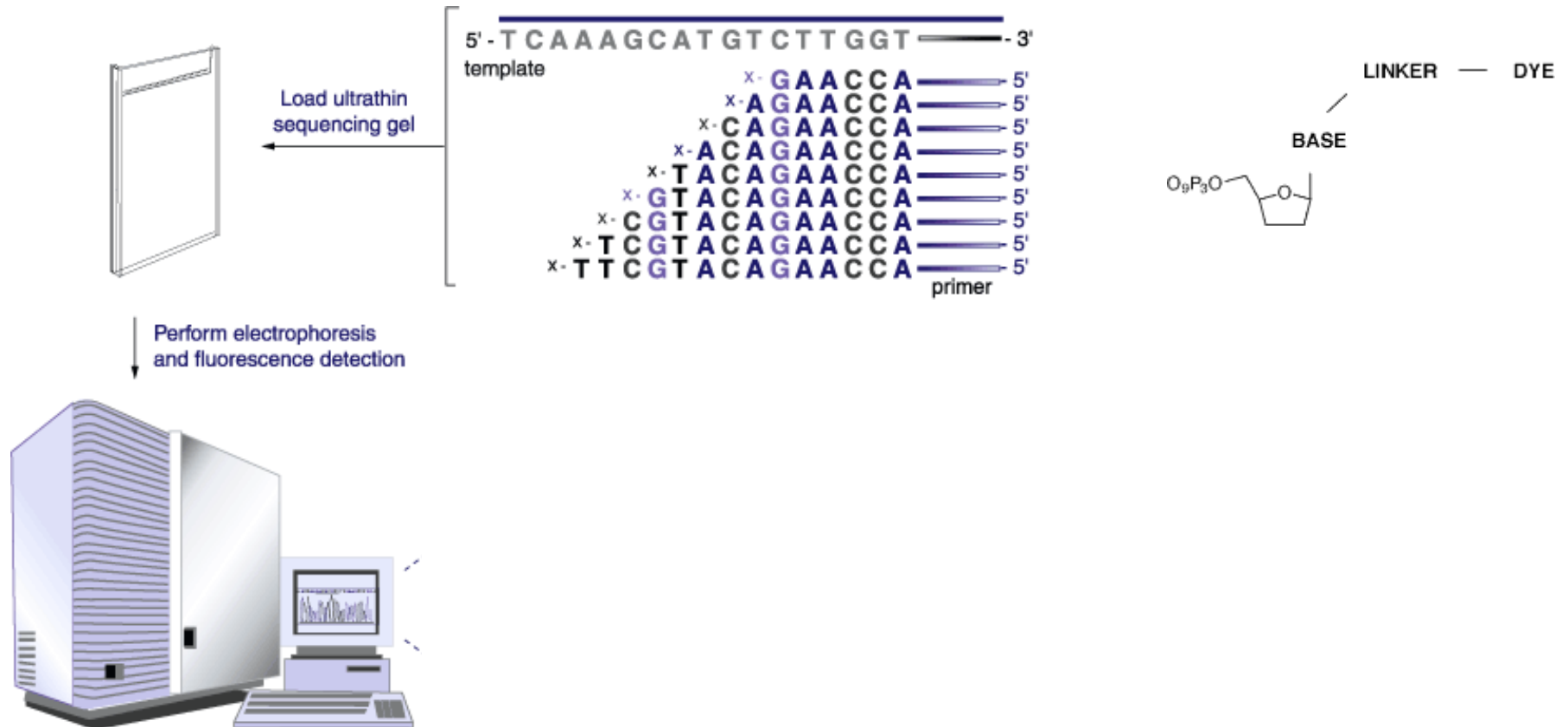
Współczesna modyfikacja metody Sangera



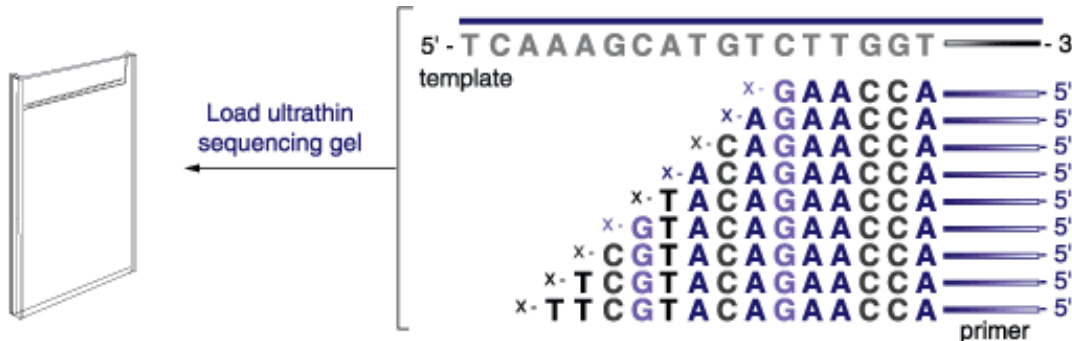
Współczesna modyfikacja metody Sangera



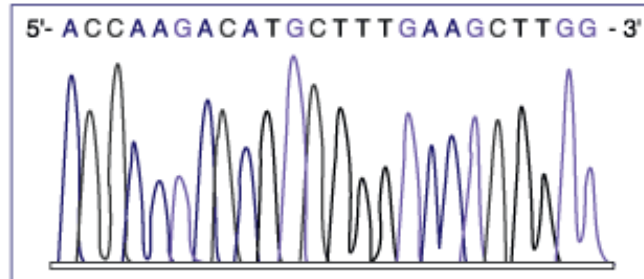
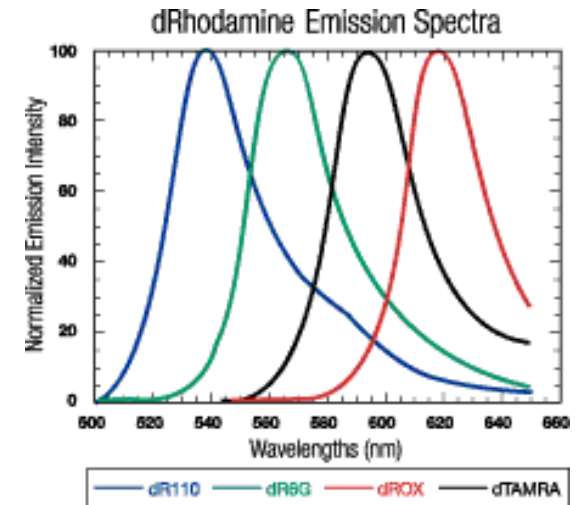
Współczesna modyfikacja metody Sanger



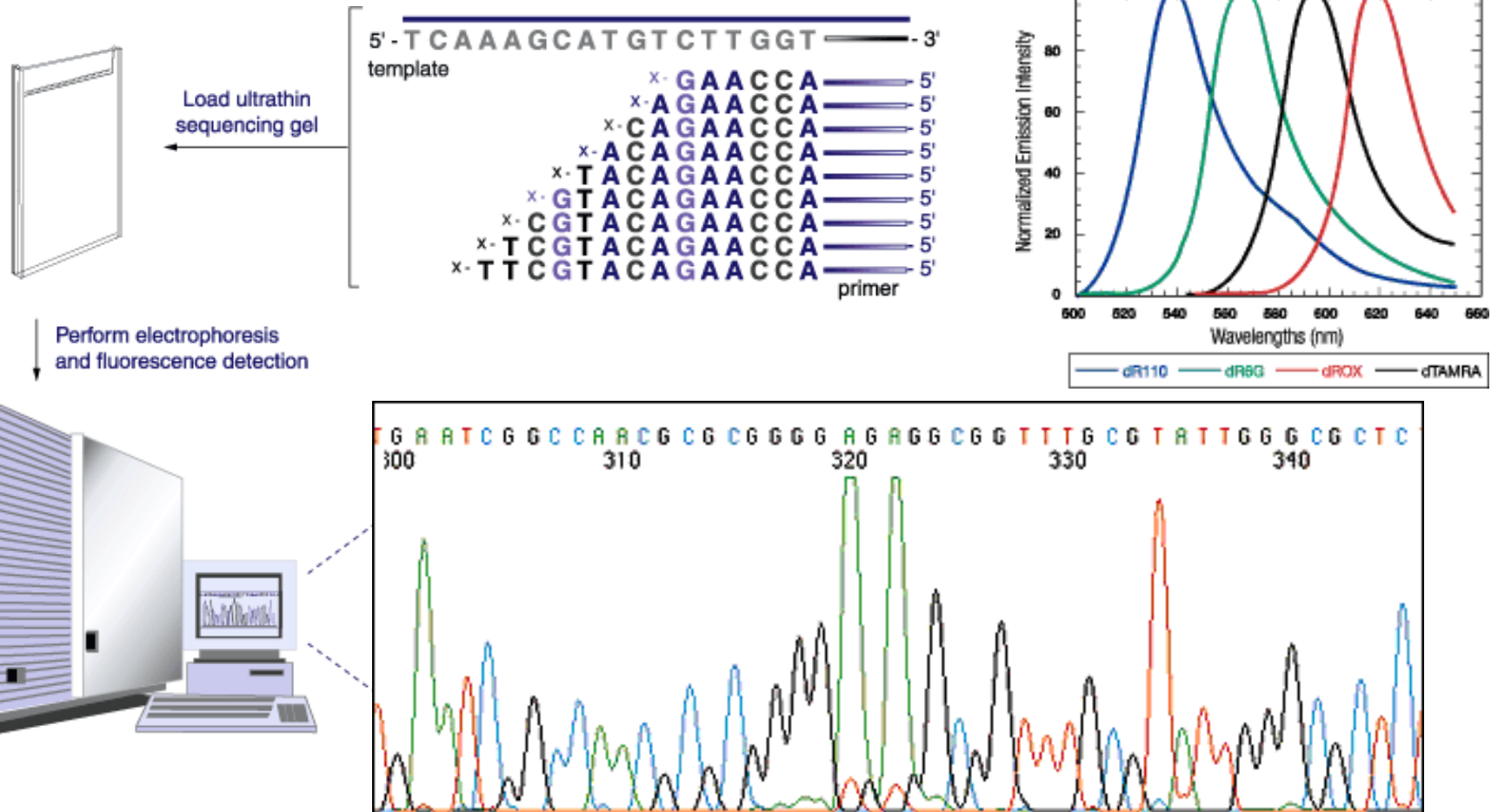
Współczesna modyfikacja metody Sangera



Perform electrophoresis
and fluorescence detection



Współczesna modyfikacja metody Sangera



Analiza sekwencji i bazy danych

- Pokaz chromatogramów: wyszukanie mutacji
- Identyfikacja sekwencji w bazach danych – blastn
- Co to jest NCBI, PubMed?
<https://www.ncbi.nlm.nih.gov/>
- Pokaz bazy specyficznej dla organizmu np. *Saccharomyces* Genome Database (SGD)
<https://www.yeastgenome.org/>

Sekwencjonowanie **nowej generacji**, wysokoprzepustowe (NGS – Next Generation Sequencing)

- **Sekwencjonowanie mostkowe**, przez syntezę z odwracalną terminacją (**Illumina**) – obecnie najczęściej stosowana i najbardziej wydajna technologia (do 2 Terra par zasad / 8 dni pracy urządzenia)
- **Sekwencjonowanie typu Ion Torrent**, przez syntezę z detekcją protonów na układzie scalonym (**Life Tech.**)
- **PacBio (Pacifyc Biosystems)** - technologia SMRT (Single Molecule, Real-Time sequencing) – długie odczyty (10-20 kb)
- **Oxford Nanopore Technologies** – sekwencjonowanie w nanoporach, najdłuższe odczyty rekord ponad 1 mln!

I wiele innych...

Sekwencjonowanie **nowej generacji**, wysokoprzepustowe – podstawowe różnice względem metody Sangera

W metodzie Sangera sekwencjonujemy jednocześnie tylko 16-96 różnych matryc, przy czym każda matryca jest mieszanina wielu cząsteczek, które nie powinny ale mogą się różnić!

Problem: przy sekwencjonowaniu produktu PCR pochodzącego od homozygoty mamy „czysty” odczyt. Gdy pochodzi on od heterozygoty – w pozycji gdzie występuje różnica między allelami, polegająca na zamianie zasad, odczyt jest wątpliwy a w przypadku delecji lub insercji wręcz nie możliwy dla całej dalszej części cząsteczki.

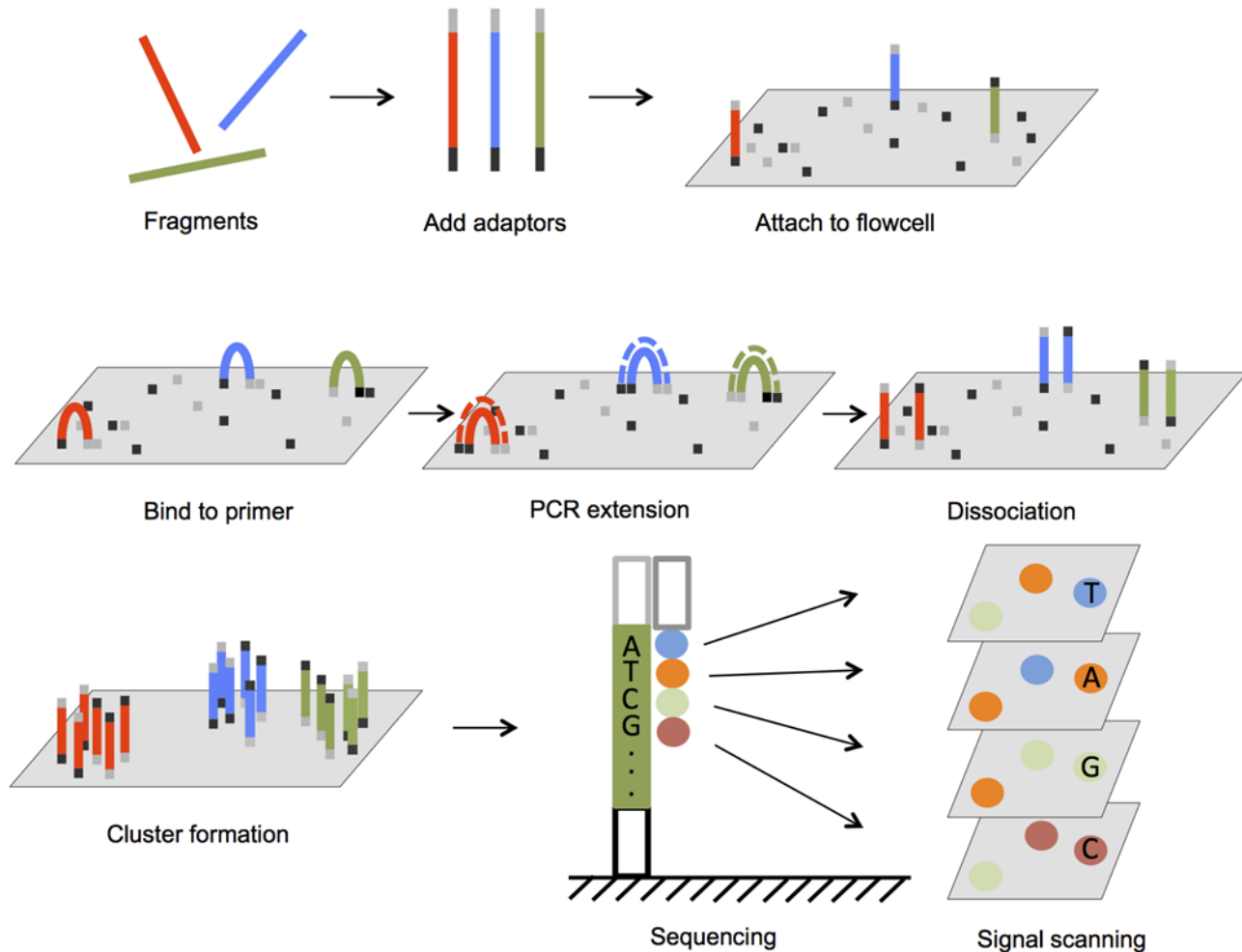
Rozwiązanie: klonowanie produktów PCR i dopiero ich sekwencjonowanie... lub sekwencjonowanie NGS.

Metody NGS opierają się o równoległe, masowe sekwencjonowanie od kilku tysięcy do kilkuset milionów różnych matryc – tzw. bibliotek.

Najpierw sekwencjonowany materiał jest losowo fragmentowany, ligowany z uniwersalnymi adaptorami i odpowiednio rozcieńczany.

Każda z finalnych matryc powstaje po amplifikacji klonalnej od pojedynczej cząsteczki poprzez **PCR na matrycy 1 cząsteczki** a nie mieszanie różnych więc jest zawsze homogenne! Każdy wariant allelu ma swoją homogenną reprezentację!

Sekwencjonowanie w technologii Illumina: mostkowe, przez syntezę z odwracalną terminacją



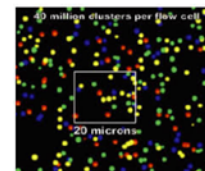
Flow cell



Sequencer



Lu et. al., „Next Generation Sequencing in Aquatic Models” in „Next Generation Sequencing - Advances, Applications and Challenges”, 2016.



<http://www.illumina.com/>

Sekwencjonowanie w technologii **Illumina**: mostkowe, przez syntezę z odwracalną terminacją

<https://www.youtube.com/watch?v=womKfikWlxM>

http://www.illumina.com/documents/products/techspotlights/techspotlight_sequencing.pdf

Illumina sequencers



MiSeq



NextSeq



HighSeq 2500

Sekwencjonowanie nowej generacji: – **wady** i **zalety** względem metody Sangera

Metoda Sangera

- Proste i szybkie przygotowanie 1 próby – nadal rutynowo używane do sprawdzenia np. poprawności konstruktu, identyfikacji pojedynczych produktów PCR!
- Niski koszt jednokrotnego uruchomienia urządzenia.
- Długość pojedynczego odczytu do ok. 500-1000 pz.
- Wysoki koszt odczytu na 1 zasadę.
- Niska przepustowość, nie możliwe są doświadczenia ilościowe.

Metody NGS

- Trudne, kosztowne i pracochłonne przygotowanie bibliotek, w zasadzie nie opłacalne dla pojedynczych prób.
- Pojedyncze odczyty bardzo krótkie, w przedziale 100-500 pz.
- Skomplikowana analiza bioinformatyczna.
- Więcej błędów sekwencjonowania.
- Bardzo duża ilość odczytów: do 2 Terra par zasad (dla 4 mld. różnych odczytów), co umożliwia dla 1 urządzenia (Illumina HiSeq 2500) odczyt 10 ludzkich genomów z ok. 100x pokryciem (powtórzeniem) dla jednego kierunku każdej z „2 nici DNA”!
- Duża ilość odczytów pozwala na prowadzenie doświadczeń ilościowych np. transkryptomicznych typu RNA-seq.
- Wielokrotnie niższy koszt odczytu na 1 zasadę.

Sekwenatory NGS – możliwości a koszty

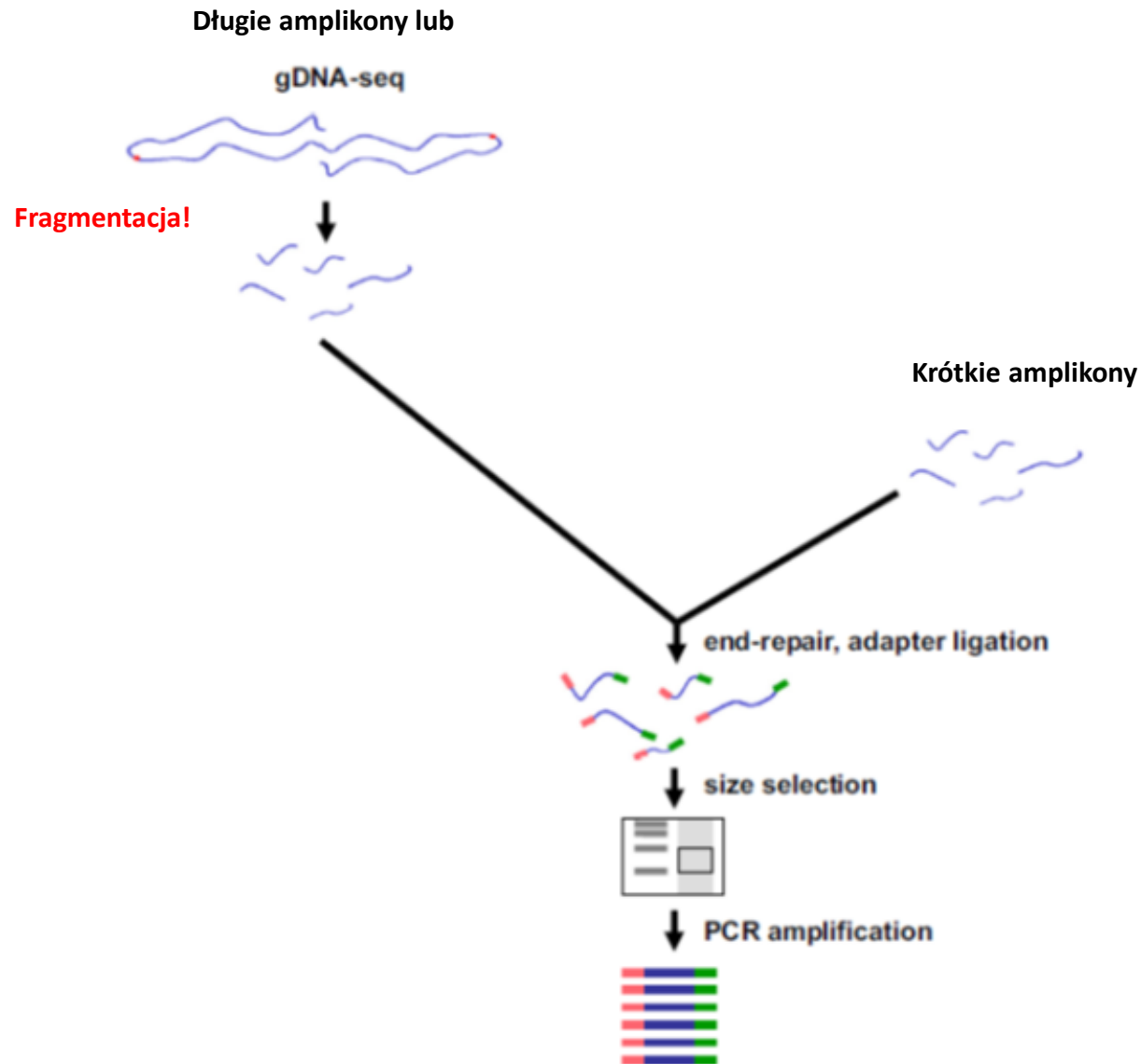
Instrument	Amplification	Run time	Millions of Reads/run	Bases / read	Reagent Cost/run	Reagent Cost/Gb	bp/run	Gbp/run	cost/Gb
Applied Biosystems 3730 (capillary)	PCR, cloning	2 hrs.	0.000096	650	\$144	\$2,307,692.31	62,400	0	\$2,307,692.31
Illumina MiSeq v3	bridgePCR	56 hrs.	25	600	\$1,530	\$102.00	15,000,000,000	15	\$102.00
Illumina NextSeq 500 - High v2	BridgePCR	29 hrs.	400	300	\$4,240	\$35.33	120,000,000,000	120	\$35.33
Illumina HiSeq 2500 - high output v3	BridgePCR	11 days	1500	200	\$15,860	\$52.87	300,000,000,000	300	\$52.87
Illumina HiSeq 2500 - high output v4	BridgePCR	40 hrs.	2000	50	\$6,240	\$62.40	100,000,000,000	100	\$62.40
Illumina HiSeq 4000	BridgePCR	3.5 days	2500	300	\$15,400	\$20.53	750,000,000,000	750	\$20.53
Illumina HiSeq X - Ten	BridgePCR	< 3 days	3000	300	\$6,375	\$7.08	900,000,000,000	900	\$7.08
Ion Torrent – PGM 314 chip v2	emPCR	4 hrs.	0.55	400	\$474	\$2,154.55	220,000,000	0.22	\$2,154.55
Ion Torrent - S5 540 chip	emPCR	2.5 hrs.	80	200	\$1,275	\$79.69	16,000,000,000	16	\$79.69
Pacific Biosciences Sequel	None - SMS	≤6 hrs.	0.385	10000	\$700	\$181.82	3,850,000,000	3.85	\$181.82

+ Illumina NovaSeq 6000
– do 20 mld. odczytów, 6 Tb

Glenn, TC (2011) Field Guide to Next Generation DNA Sequencers. Molecular Ecology Resources. 2016 Update.

<http://www.molecular ecology.com/next-gen-fieldguide-2016/>

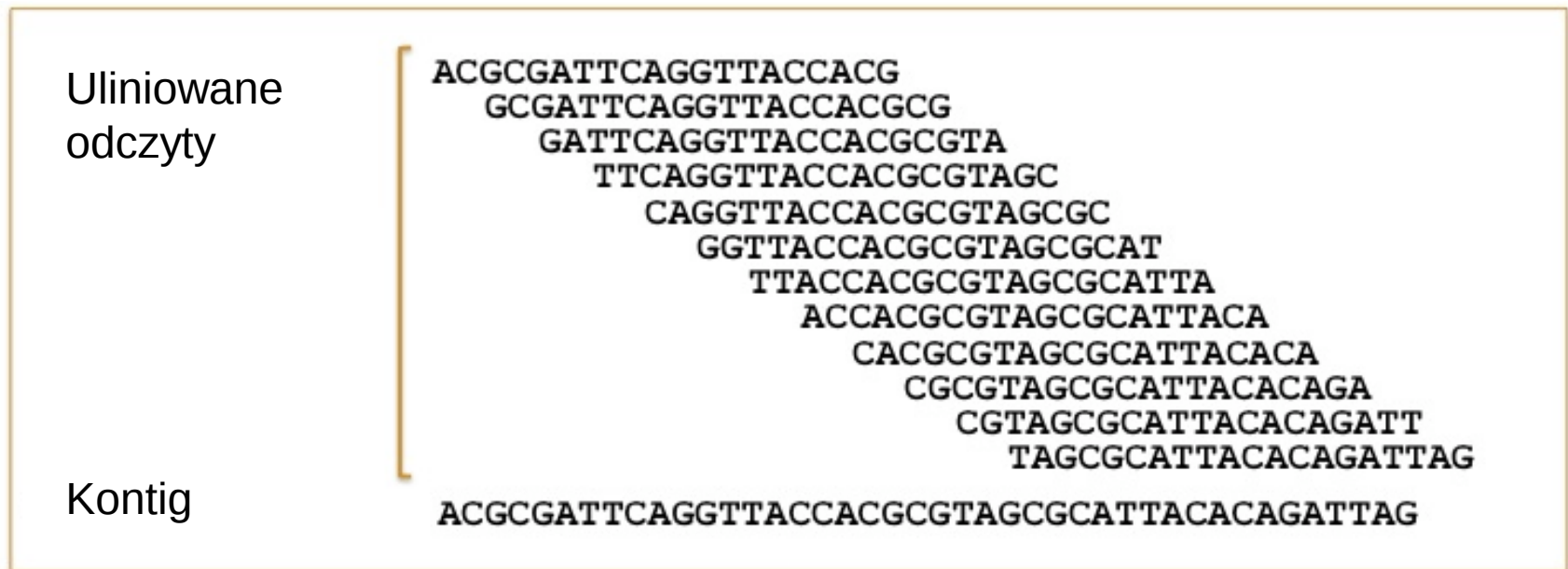
Przygotowywanie bibliotek NGS dla DNA



Sekwencjonowanie genomów – obecnie metodami NGS

Obecnie wykorzystywane jest podejście „**shot-gun**”: losowego fragmentowania gDNA, pierwotnie stosowane dla metody Sangera.

Sekwencje – pojedyncze odczyty (ang. „reads”) składamy w kontigi (ang. „contigs”) *in silico*: na zakładkę, dzięki pokrywającym się końcom odczytów.

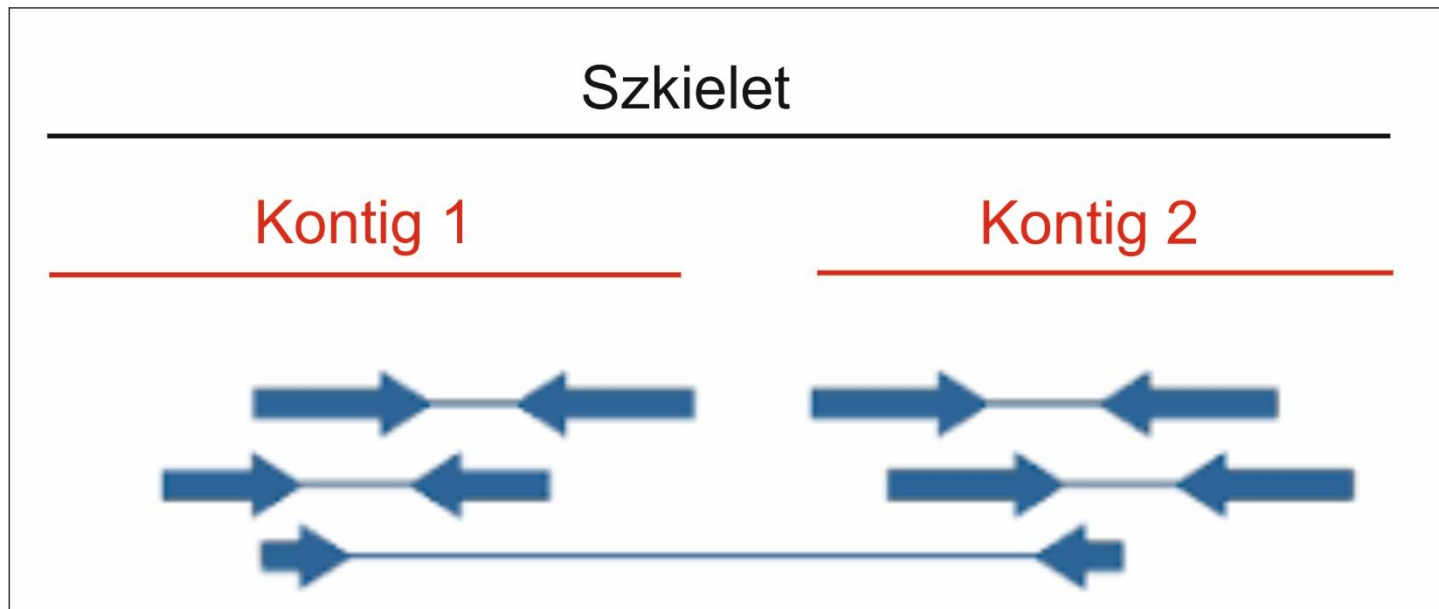


Sekwencjonowanie genomów – dwa podejścia

1. **Ponowne sekwencjonowanie** (ang. „resequencing”) – z wykorzystaniem sekwencji odniesienia (referencyjnej) już zsekwencjonowanego genomu – dla następnych osobników gatunku.
2. **Sekwencjonowanie *de novo***, gdy nie ma dobrej sekwencji odniesienia – wymaga więcej pracy i bardziej wymagające obliczeniowo.

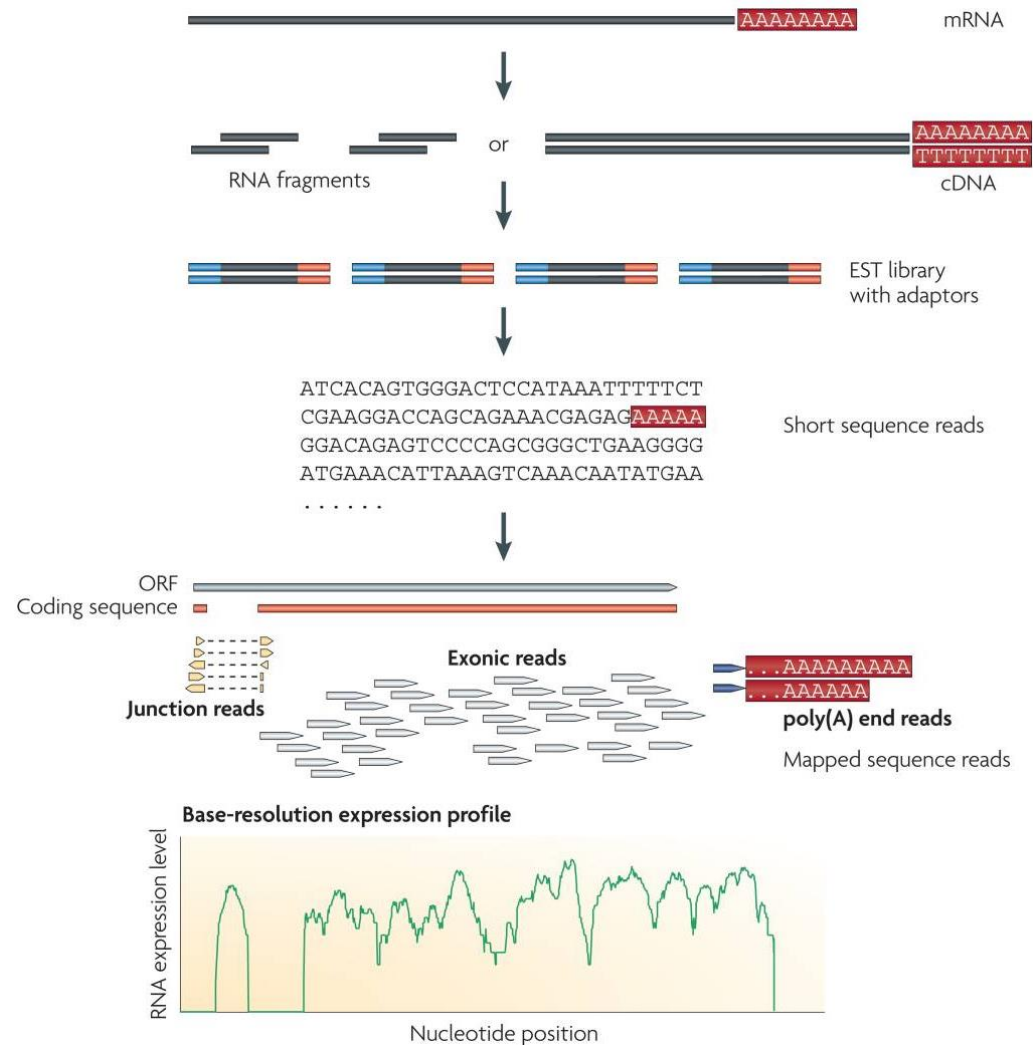
Można łączyć metody sekwencjonowania: dające długie, mniej dokładne odczyty do których przyrównuje się dużo krótkich, bardziej dokładnych odczytów.

Albo wykorzystać odczyty samych końców dla dużych fragmentów (pary „odczytów”) jako szkielety (ang. scaffold”) do układania wielu kontigów w poprawnej kolejności – znajdujemy sekwencje 2 końców dużego fragmentu w 2 różnych kontigach – więc kontigi w genomie leżą blisko siebie. Luki uzupełniamy celowanym sekwencjonowaniem.



Typowe doświadczenie mRNAseq

Pierwsza publikacja dotycząca RNAseq *Mortazavi et. al., Nature 2008*

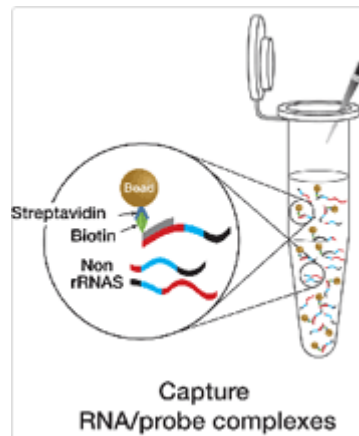
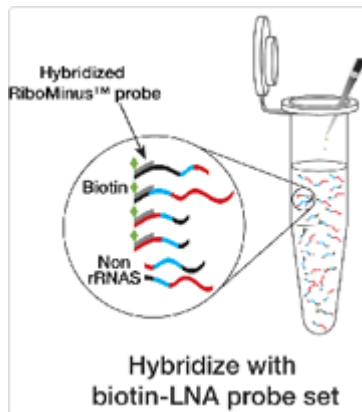


**Selekcja frakcji
PolyA+ na złożu
oligo-dT**

RNAseq dla całego transkryptomu

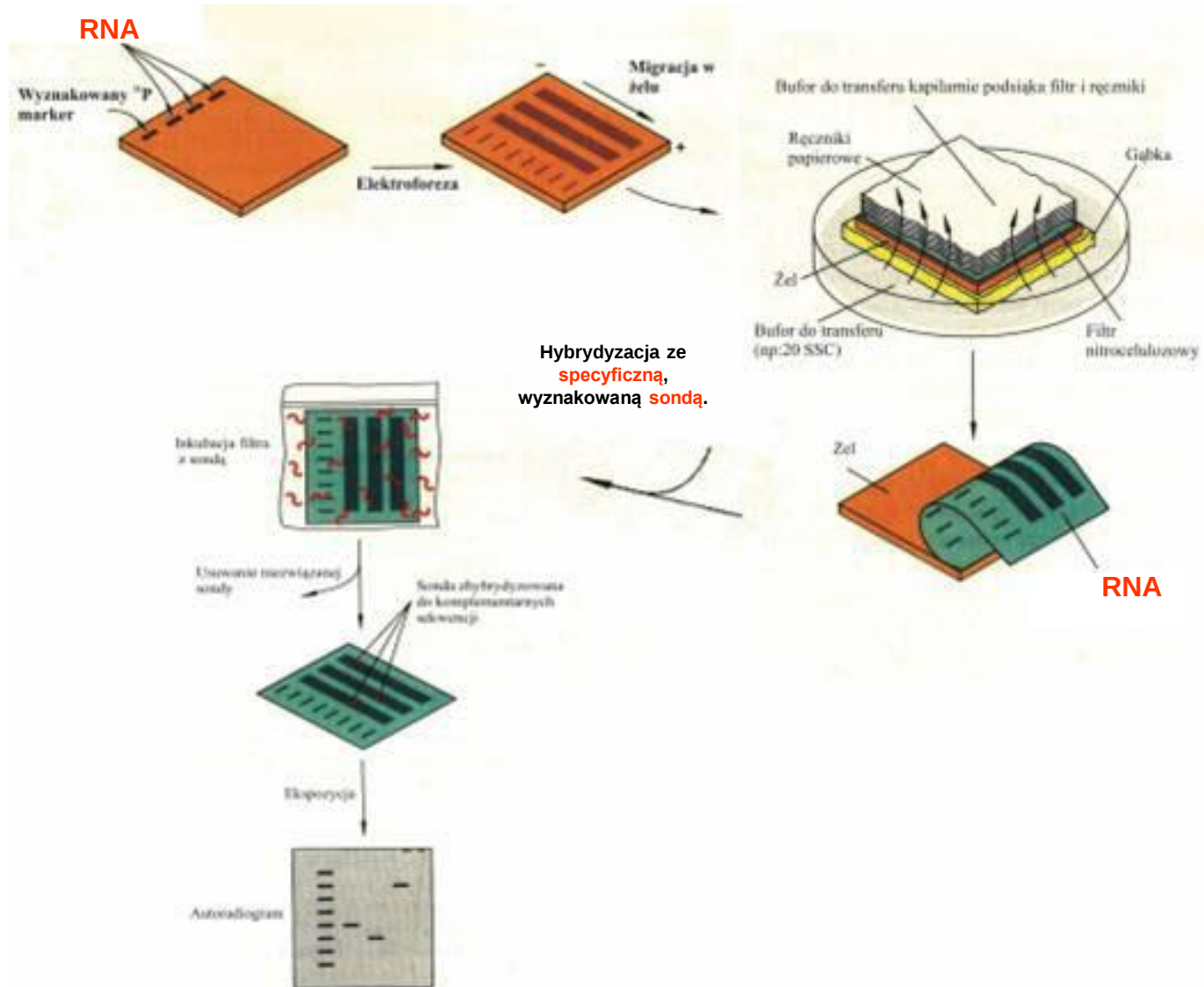
Usuwanie rRNA z całkowitego RNA

do sekwencjonowania pozostałych, wszystkich transkryptów



RiboMinus™ Technology by Life Technologies (ThermoFisher)

Technika *northern-blot*: rozdział i detekcja RNA



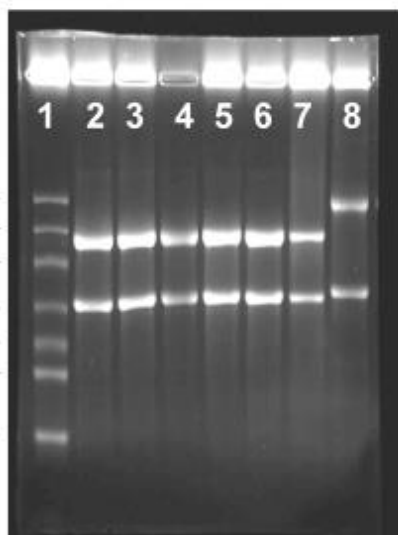
Indukcja transkrypcji genu kodującego arginazę u *A. nidulans*

Katabolizm argininy

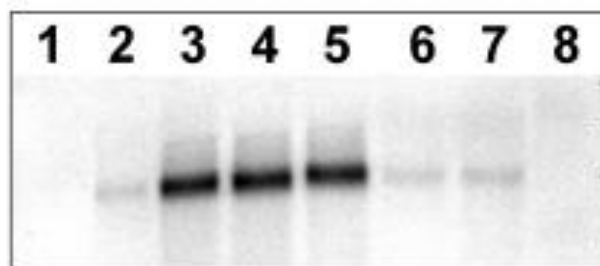
- Gen ***agaA*** koduje arginazę, rozkładającą argininę do ornityny, a gen ***otaA*** koduje transaminazę ornitynową
- Transkrypcja obu genów jest **indukowana arginina**
- Za indukcję odpowiada czynnik transkrypcyjny **ARCA**
- ***arcA^d 47*** – mutacja dominująca
- ***arcA^r 3*** – mutacja recesywna

1.	standard masy cząsteczkowej RNA (RNA Lader High Range, MBI Fermentas)	warunki hodowli	aktywność arginazy
2.	wt	MM	0,2
3.	wt	MM + arginina	1,8
4.	<i>arcA^{d47}</i>	MM	1,9
5.	<i>arcA^{d47}</i>	MM + arginina	2,1
6.	<i>arcA^{r3}</i>	MM	0,1
7.	<i>arcA^{r3}</i>	MM + arginina	0,1
8.	RNA totalne z ludzkich komórek białaczkowych Jurkat	<i>1u aktywności arginazy = aktywność enzymu dająca 1 μmol mocznika na minutę w warunkach standardowych</i>	

OBRAZ ELEKTROFOREZY



WYNIK AUTORADIOGRAFII

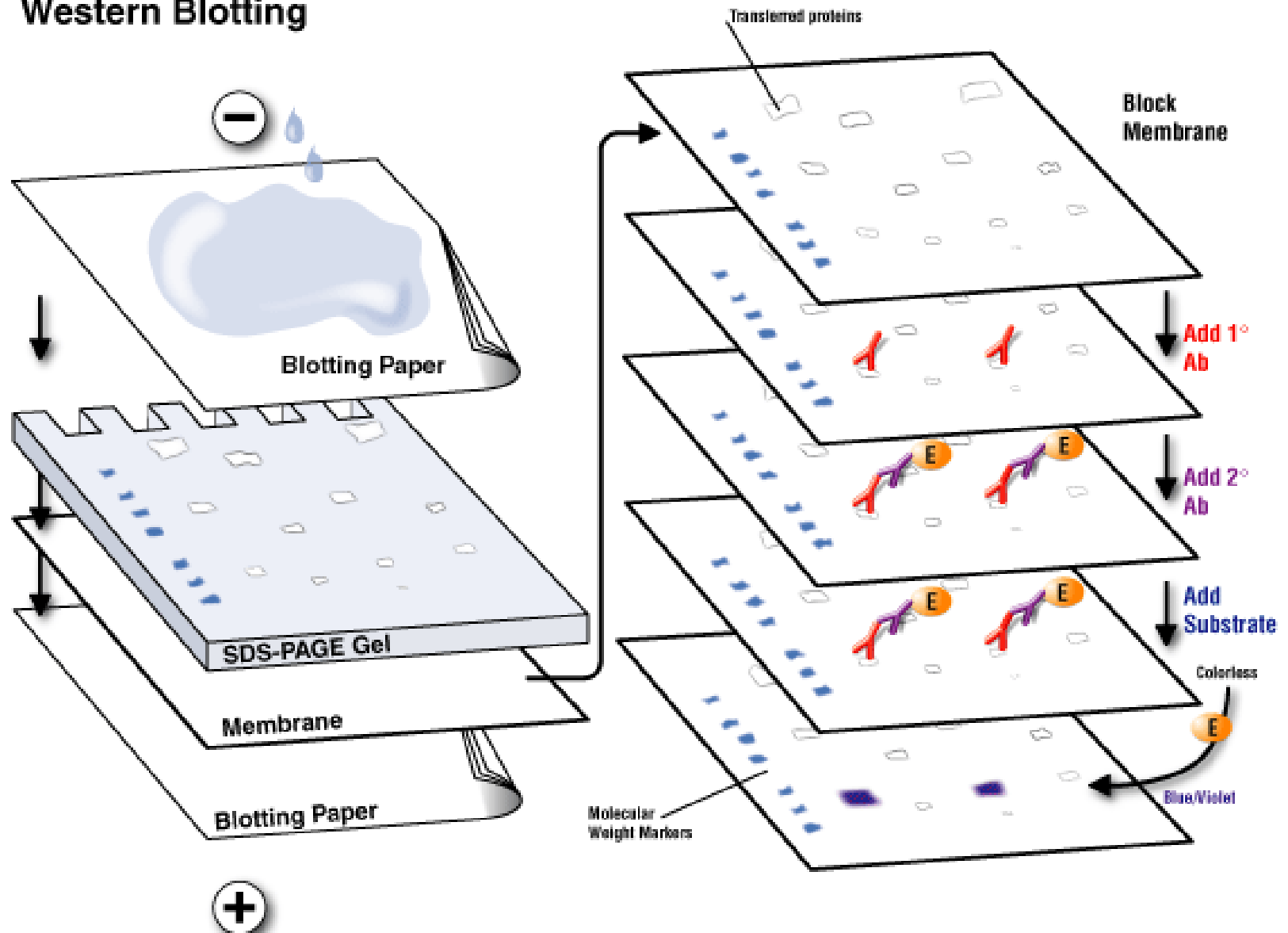


← sygnał hybrydyzacji dla
mRNA arginazy
(ok. 1300 zasad)

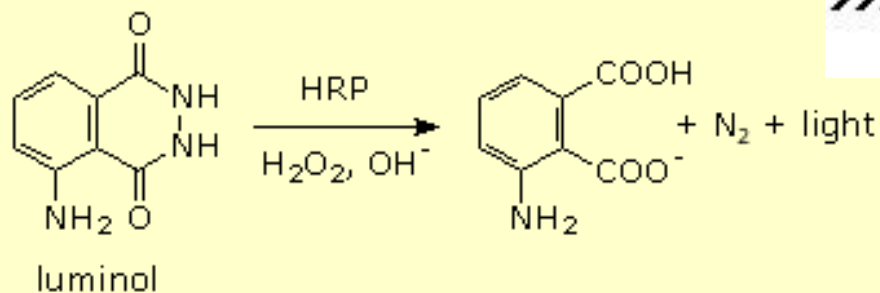
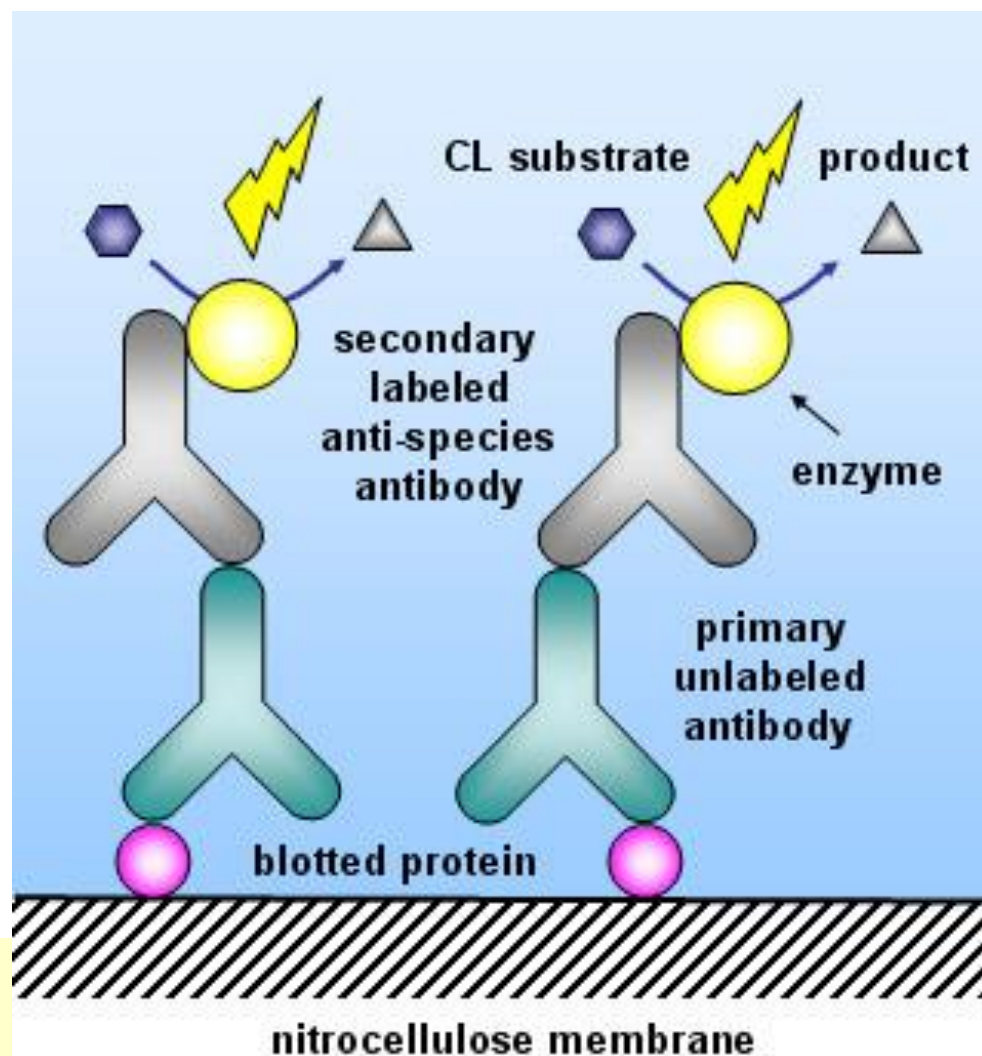
- mRNA (geny) referencyjne
- northern-blot jakościowy – alternatywny splicing

Technika *western-blot*

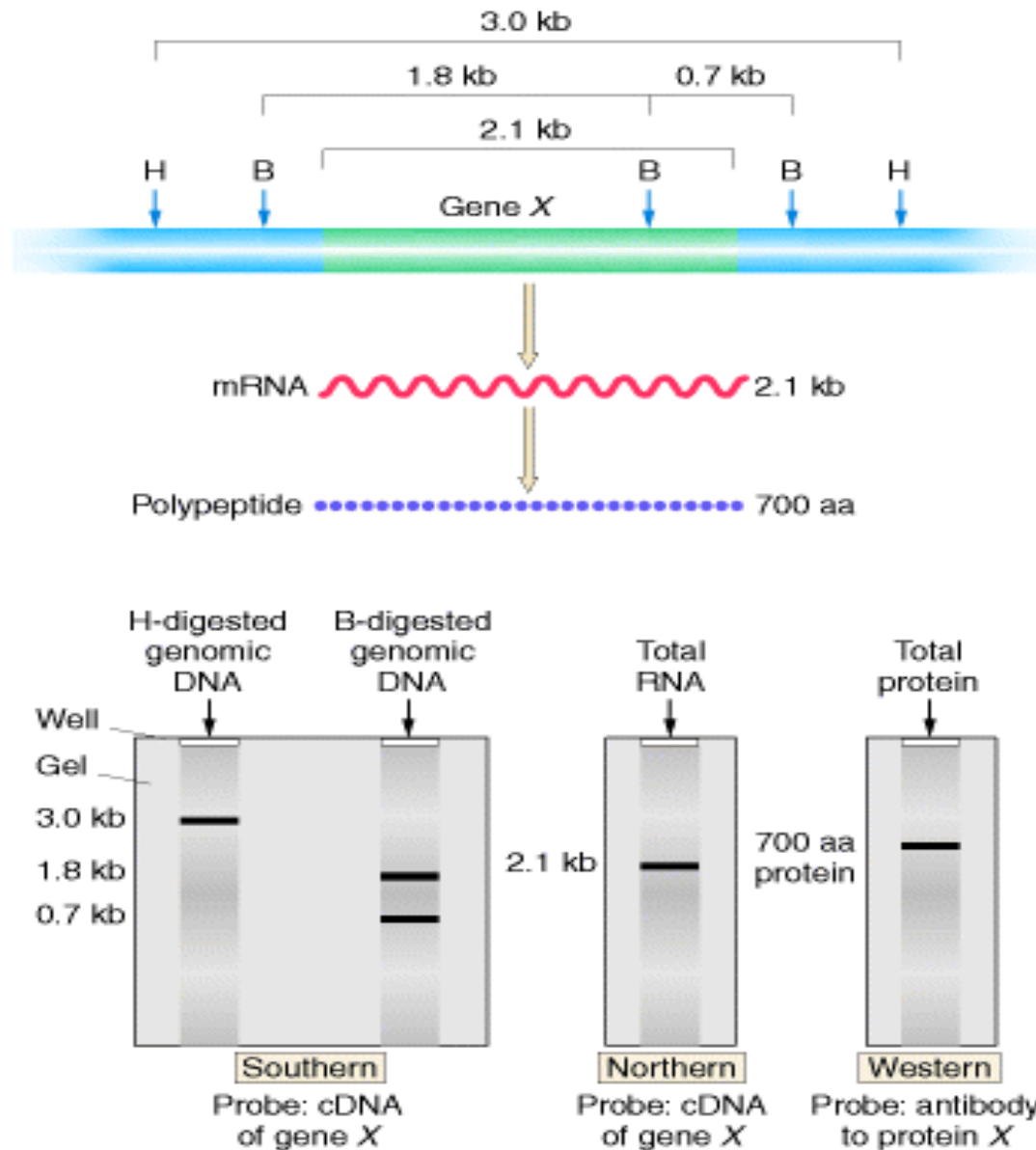
Western Blotting



Detekcja białka w western-blot

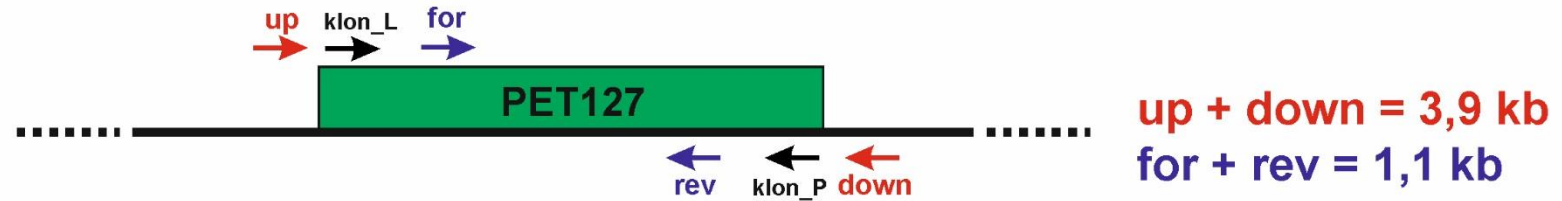


Southern, northern i western

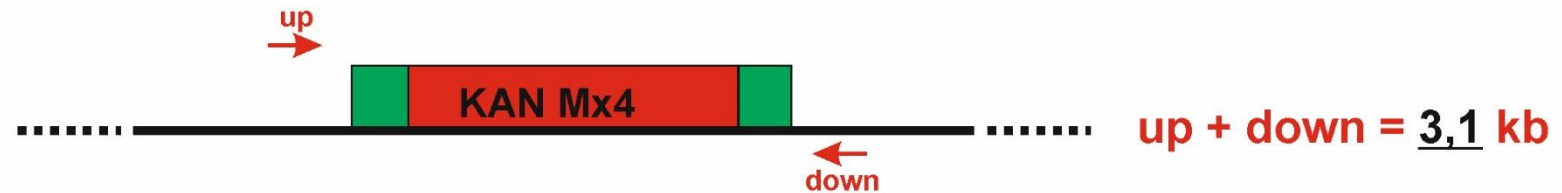


Kontrola szczepów w mutagenезie PET127 – schemat PCR

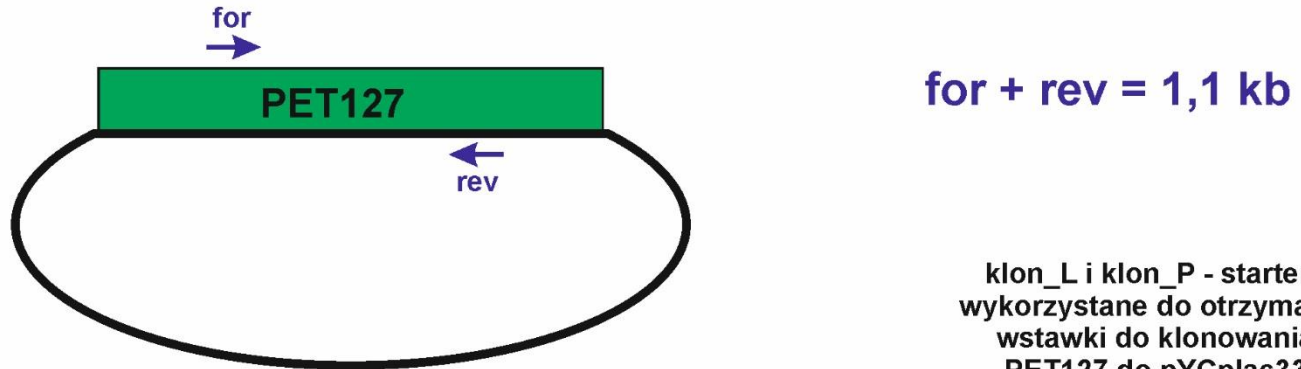
wt *locus* PET127



locus PET127 Δ



pYCplac33::PET127



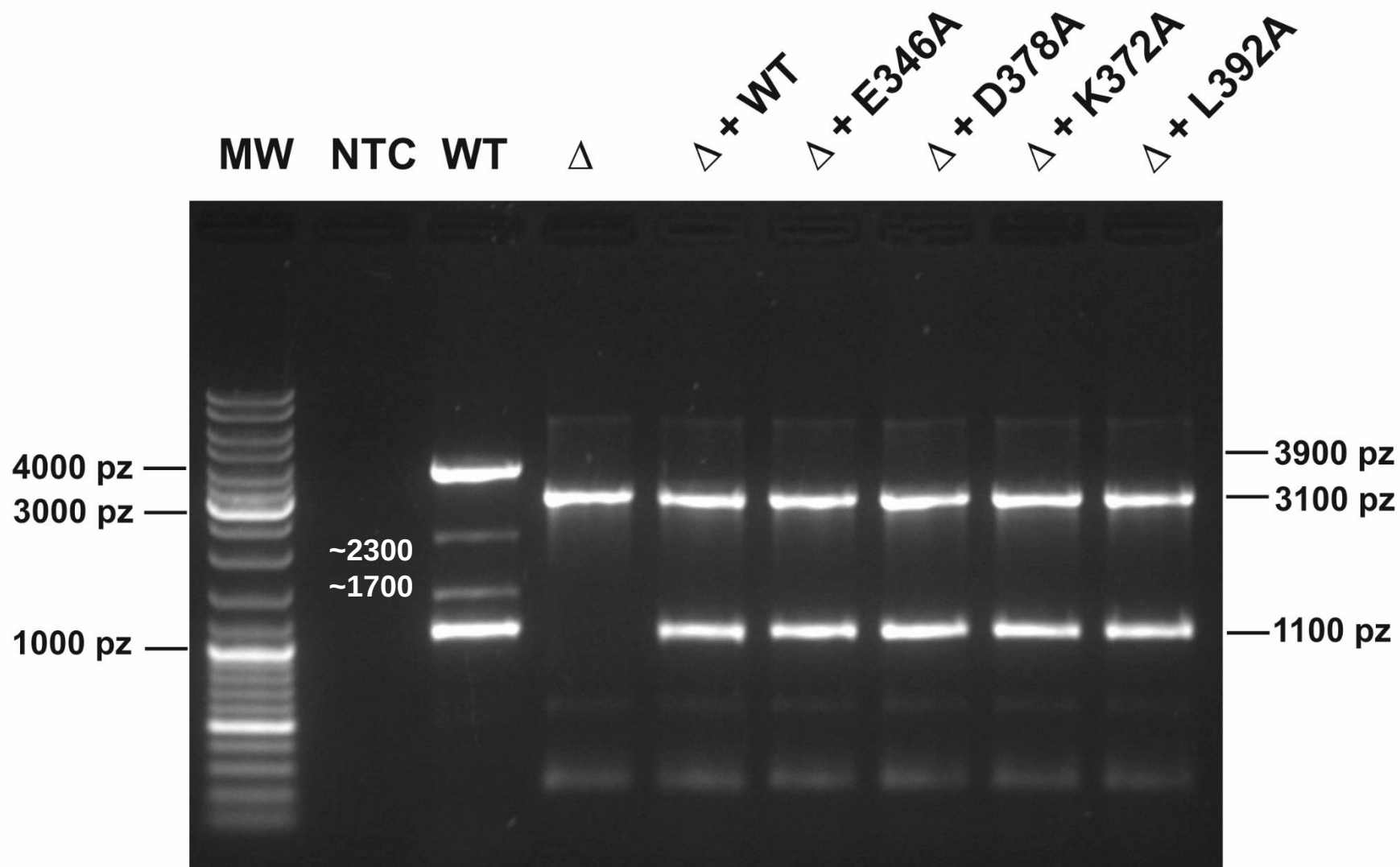
klon_L i klon_P - startery
wykorzystane do otrzymania
wstawki do klonowania
PET127 do pYCplac33

Kontrola szczepów w mutagenезie PET127 – rozdział produktów PCR w żelu agarozowym

Przygotowanie i nakładanie prób:

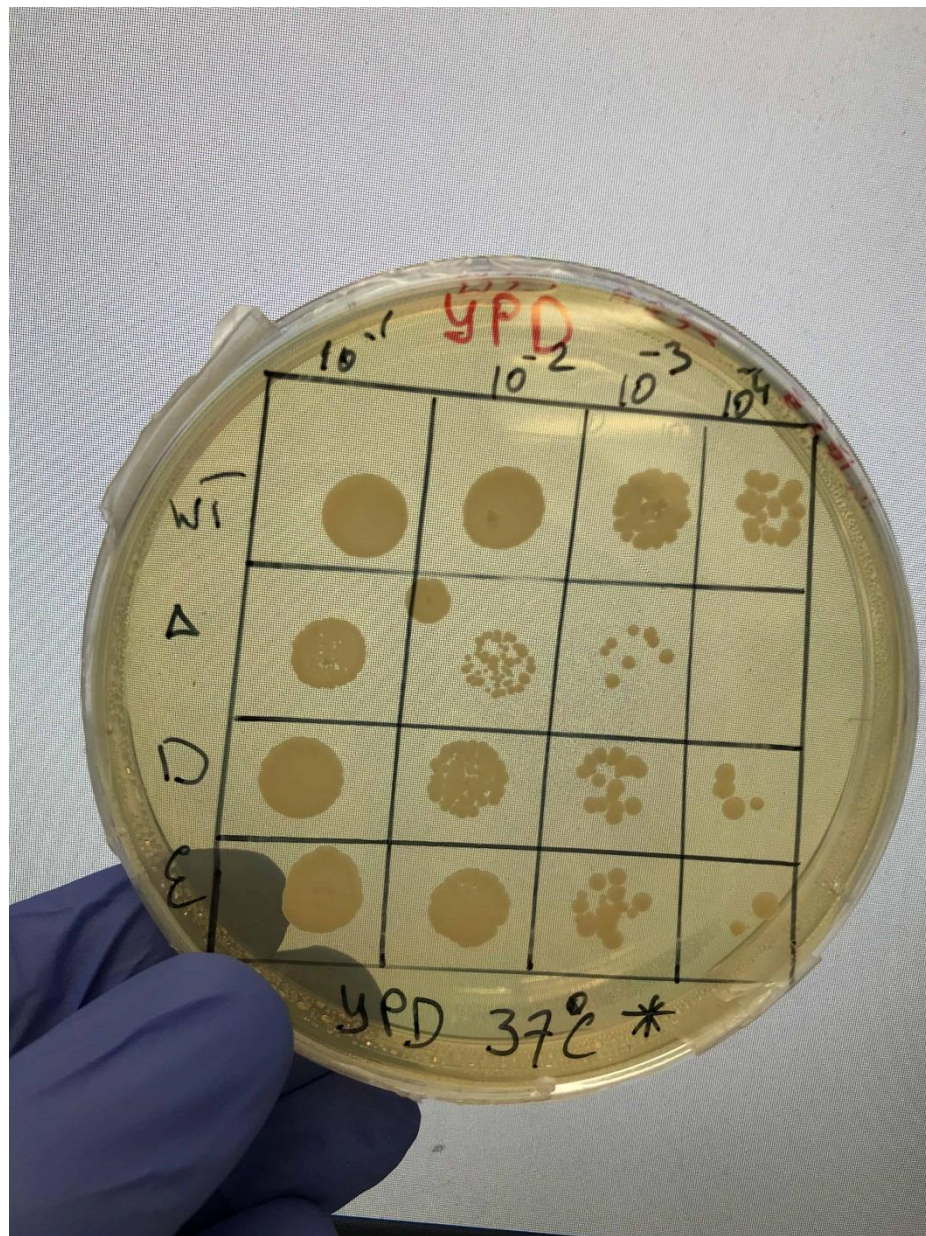
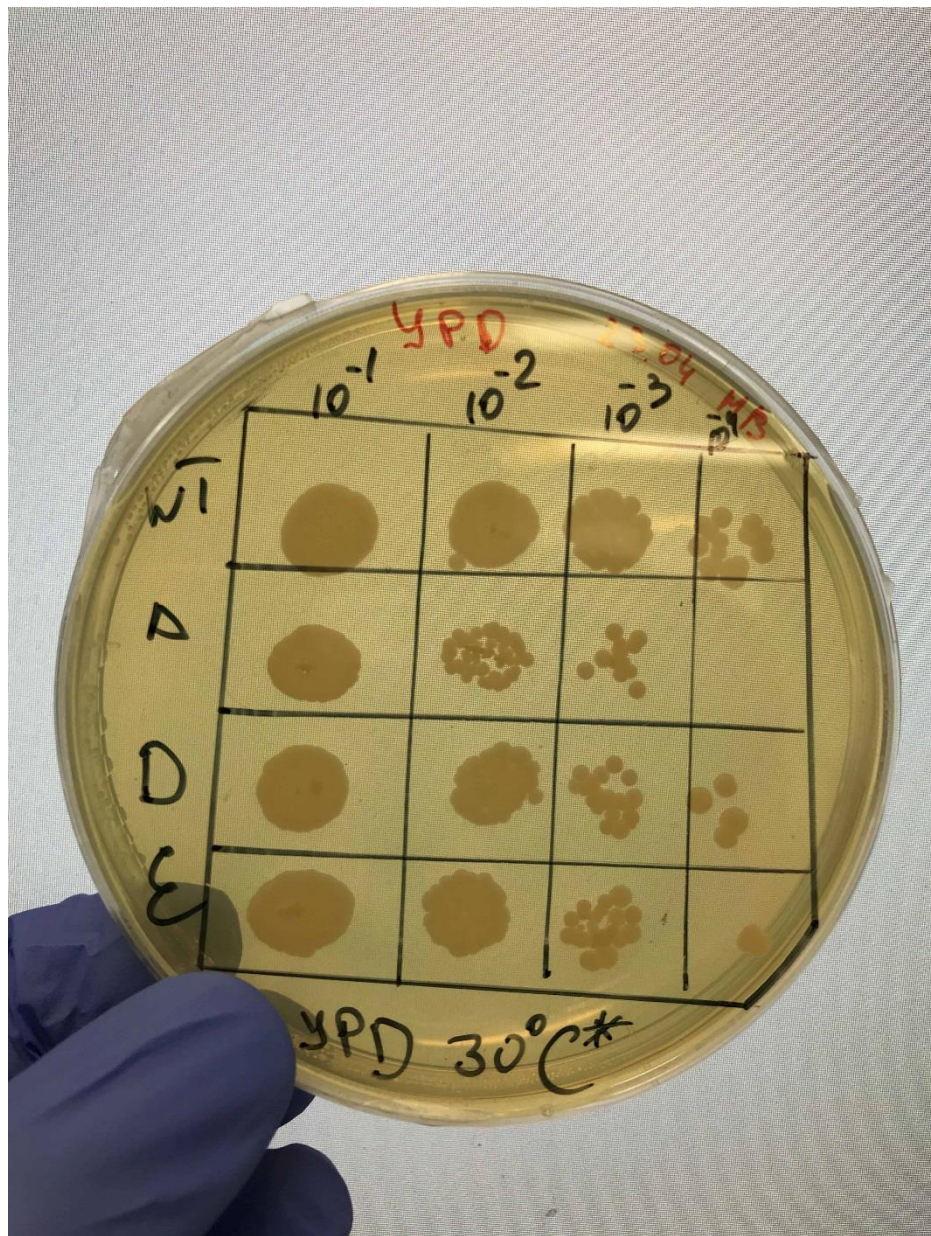
- Do reakcji PCR (20µl) dodać 4µl 6x stężonego buforu obciążającego z barwnikiem.
- Do 1-szej kieszonki nałożyć 5µl wzorca masy cząsteczkowej (gotowy, z buforem do nakładania).
- Na żel nałożyć 12µl mieszaniny.
- Elektroforezę prowadzić w 1% żelu agarozowym + EtBr, 1x buforze TAE, 70-90V do 1h.

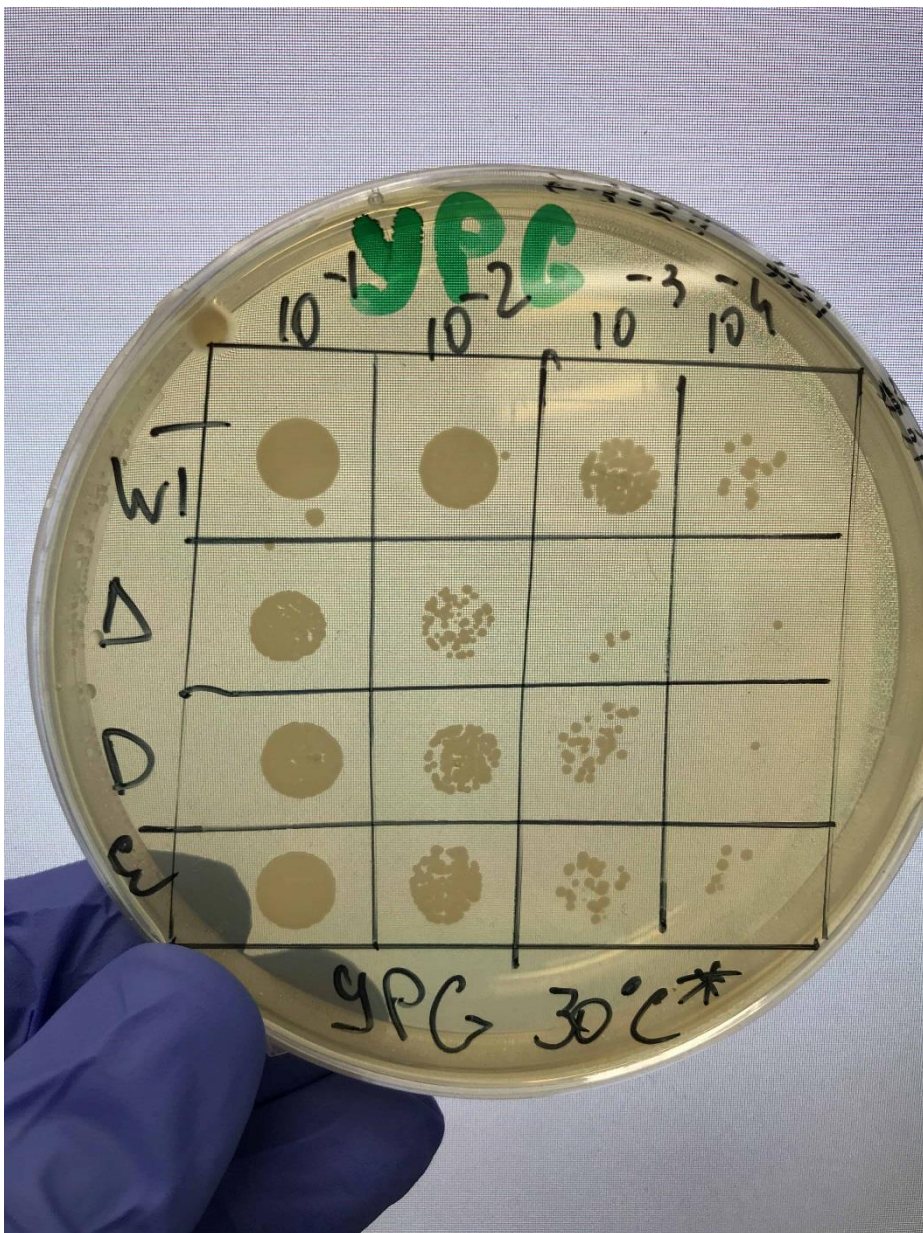
Kontrola szczepów w mutagenizie PET127 – **wynik PCR**

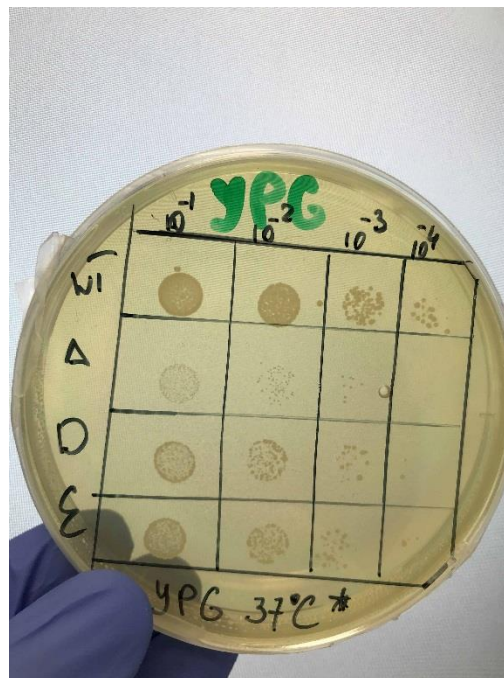
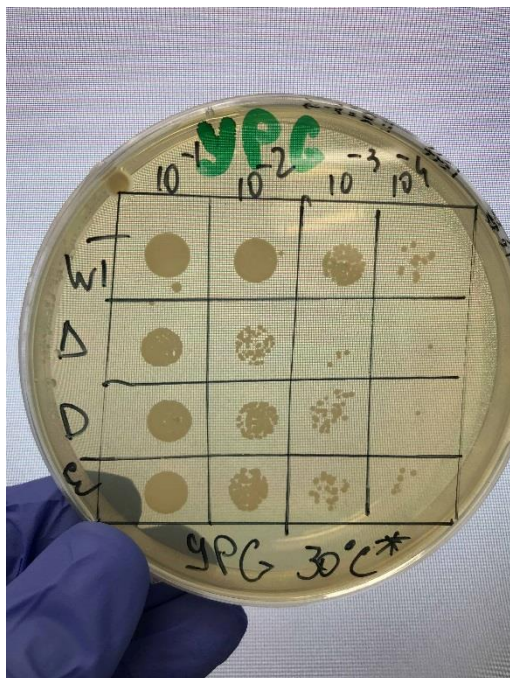
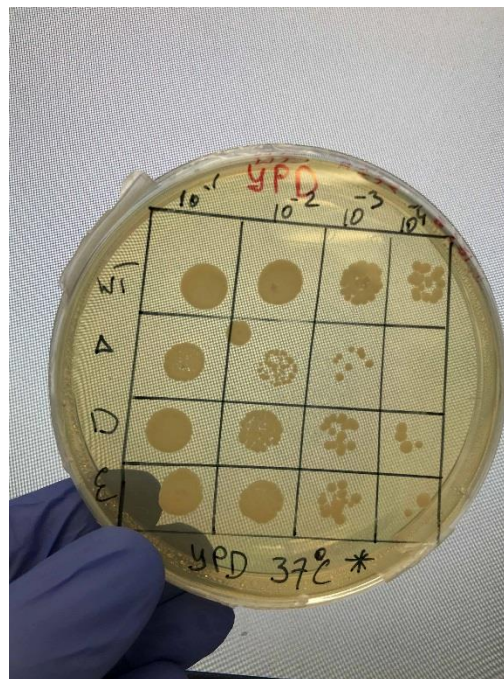
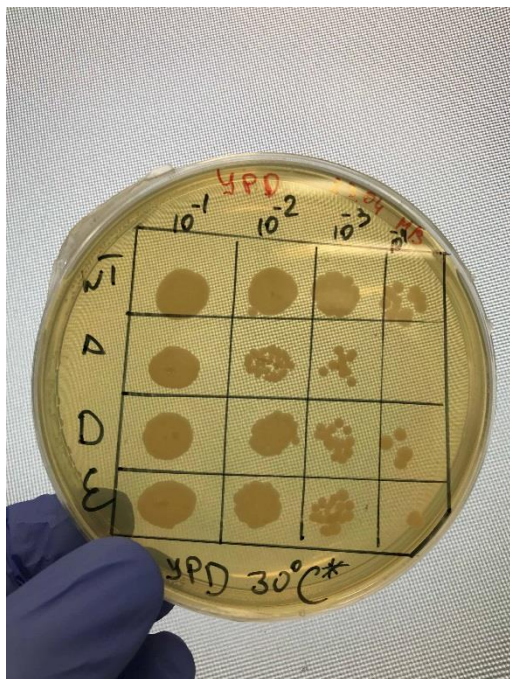


Odczyt testów wzrostowych dla mutantów w PET127

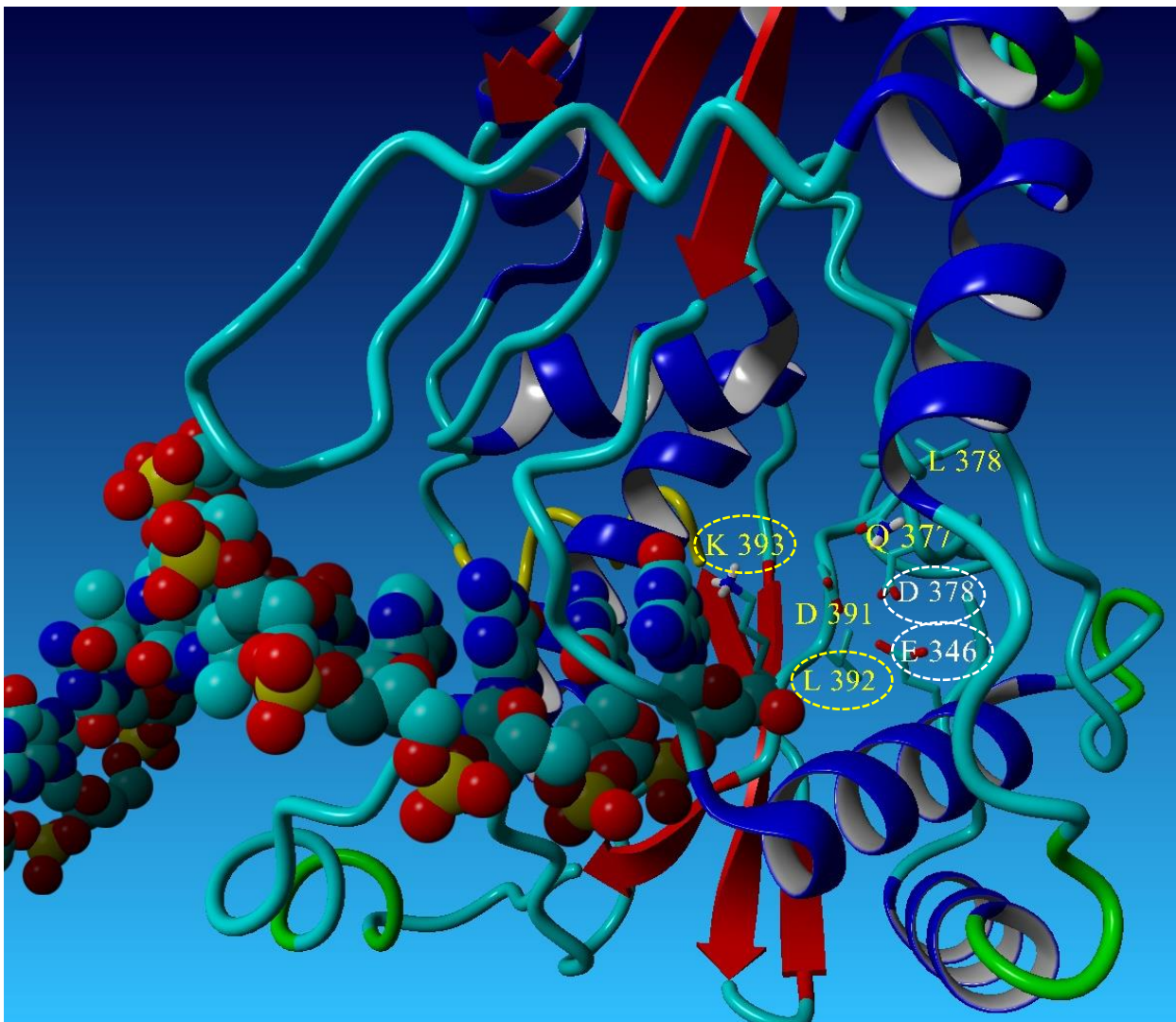
Sprawdzenie fenotypów oddechowych dla
mutantów D378A i E346A







Model struktury PET127 z *S. cerevisiae*



Na model PET127 zaznaczono:

- **struktury typu „kartka” kolorem czerwonym,**
- **helisy – ciemnoniebieskim,**
- **łańcuchy słabo ustrukturyzowane – jasnoniebieskim lub zielonym.**

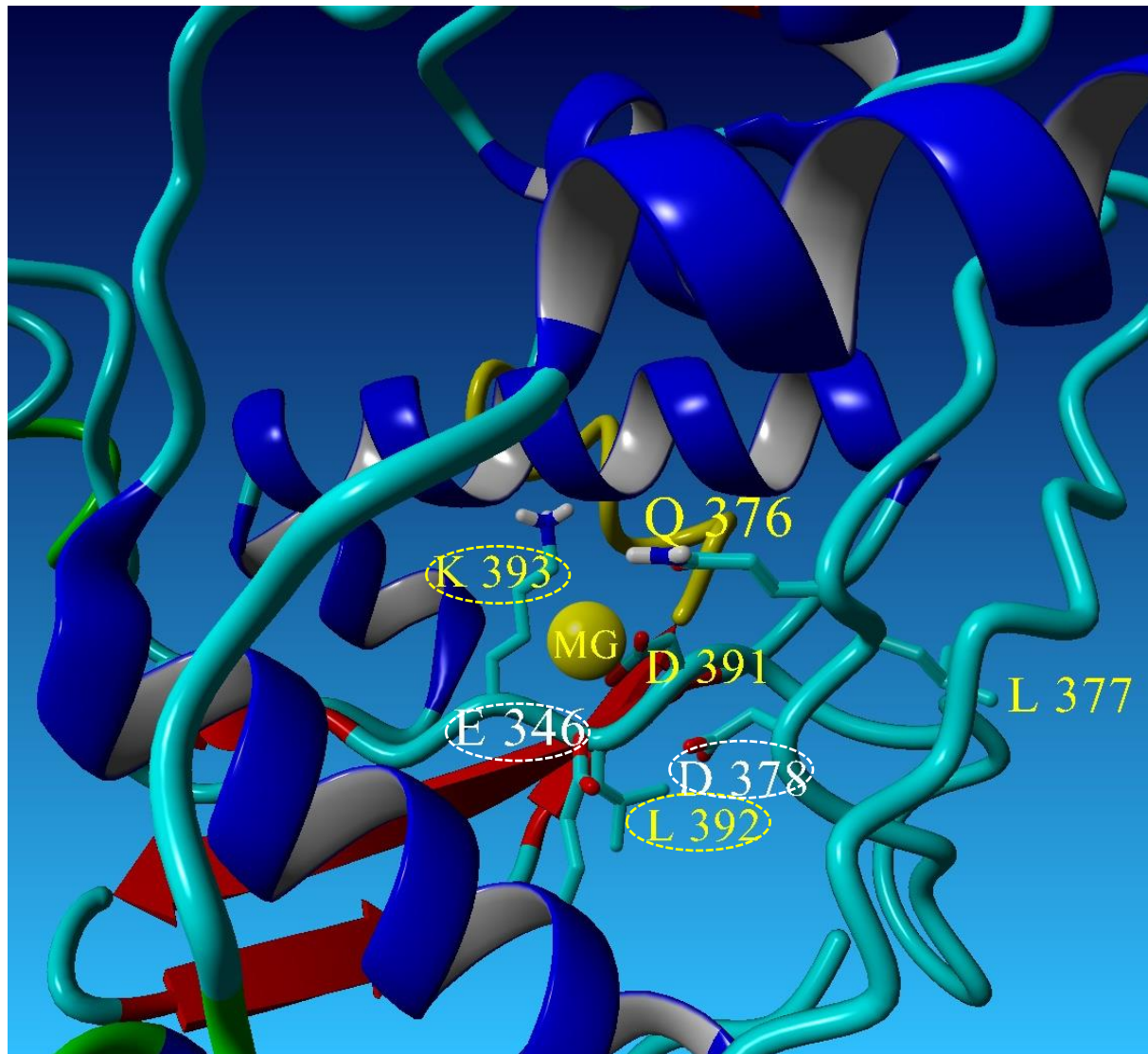
Zaznaczono mutowane podstawniki z okolic centrum aktywnego:

- **kolor biały** – podstawniki E346 i D378,
- **kolor żółty** – podstawniki L392 i K393.

Model atomowy – czaszowy: cząsteczka kw. nukleinowego (tu DNA)

Model otrzymano w oprogramowaniu I-TASSER (Yang Zhang's Lab) poprzez modelowanie na podstawie sekwencji homologicznych o znanej strukturze i zwizualizowano programem YASARA (YASARA Biosciences GmbH):
<http://zhanglab.ccmb.med.umich.edu/>
<http://www.yasara.org>

Model struktury centrum aktywnego PET127 z *S. cerevisiae*



Na model PET127 zaznaczono:

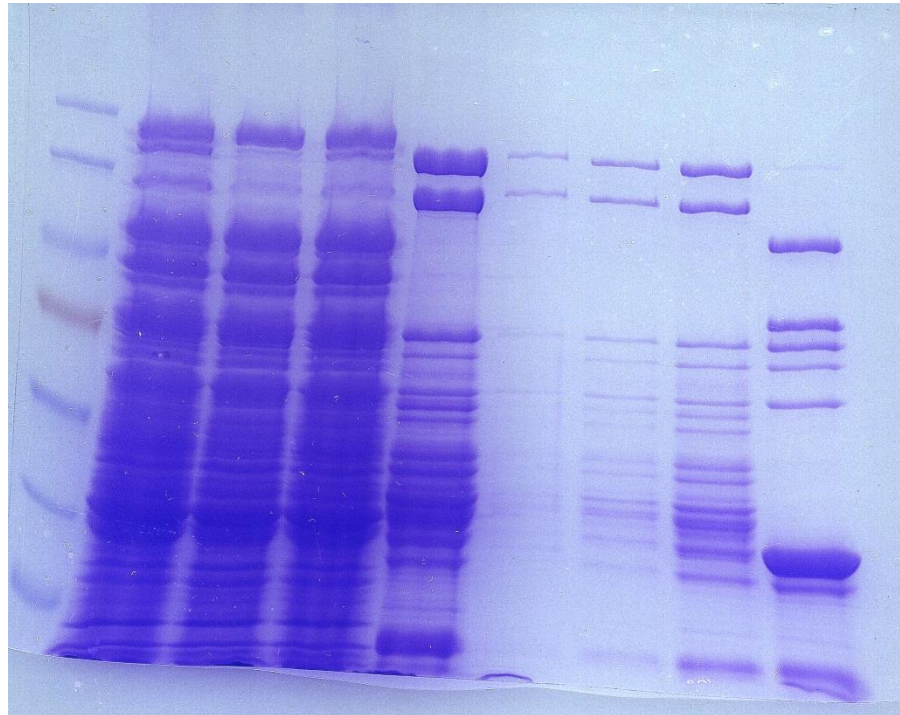
- **struktury typu „kartka”** kolorem **czerwonym**,
- **helisy** – **ciemnoniebieskim**,
- **łańcuchy słabo ustrukturyzowane** - **jasnoniebieskim** lub **zielonym**.

Zaznaczono mutowane podstawniki z okolic centrum aktywnego:

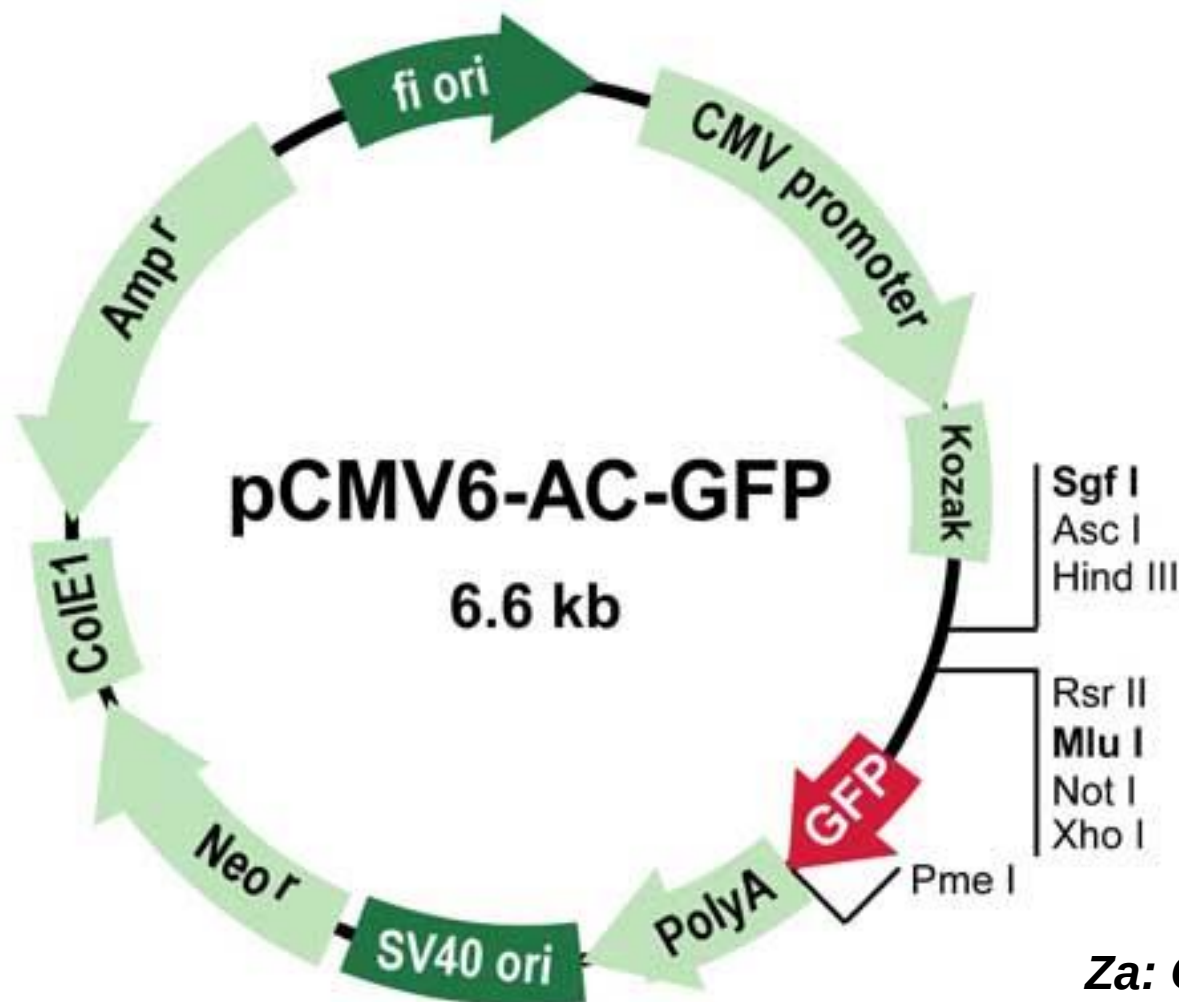
- **kolor biały** – podstawniki E346 i D378,
- **kolor żółty** – podstawniki L392 i K393.
- **MG** – skoordynowany jon Mg^{2+}

Model otrzymano w oprogramowaniu I-TASSER (Yang Zhang's Lab) poprzez modelowanie na podstawie sekwencji homologicznych o znanej strukturze i zwizualizowano programem YASARA (YASARA Biosciences GmbH):
<http://zhanglab.ccmb.med.umich.edu/>
<http://www.yasara.org>

Kilka słów o ekspresji heterologicznej



Przykład wektora bifunkcyjnego do ekspresji białek w komórkach ssaczych



Za: OriGene

Plazmid do heterologicznej ekspresji i oczyszczania białek w bakteriach

Pokaz zaindukowanych hodowli płynnych *E. coli* pod światłem UV. **Zachować ostrożność!!!**

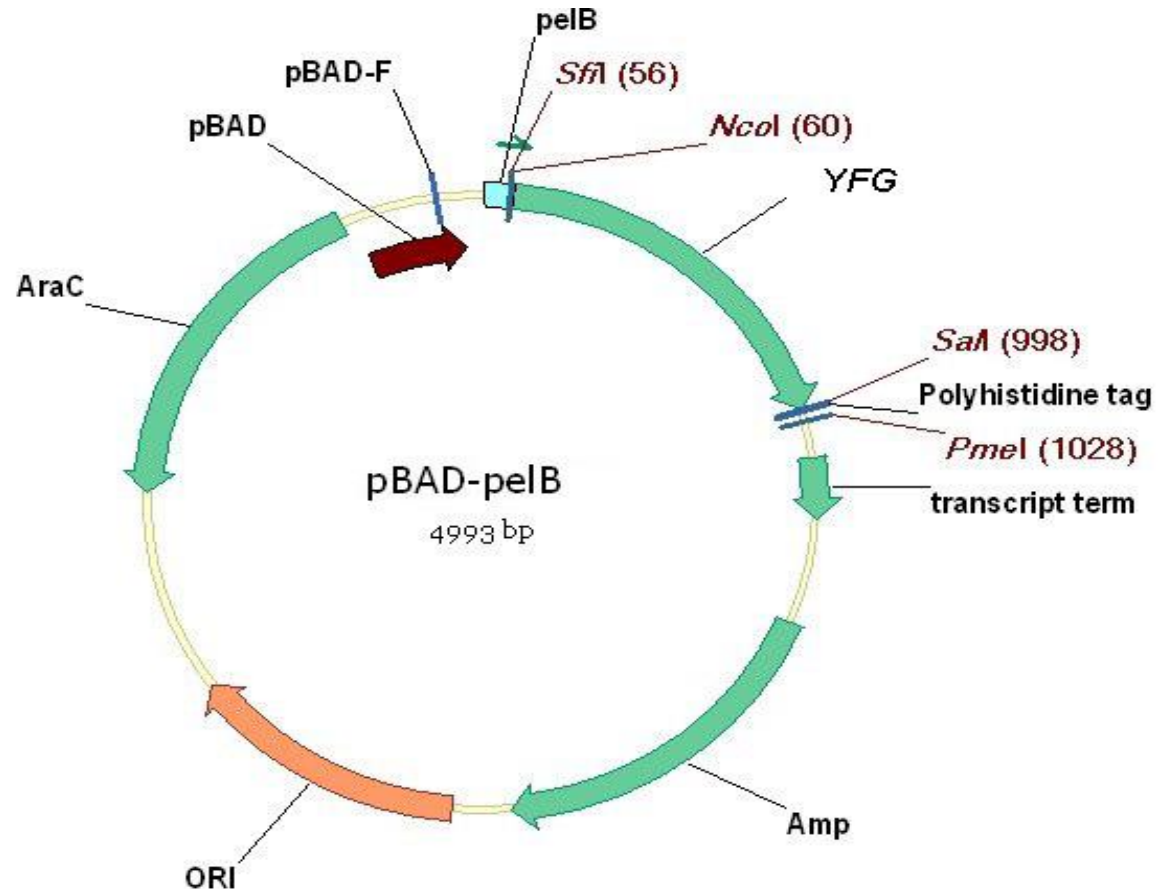
AraC – czynnik o podwójnej funkcji, wiąże się zawsze z promotorem:

- arabinoza → represor
- + arabinoza → aktywator

pBAD - promotor arabinozowy

YFG - your favourite gene

polyhistidine tag - znacznik heksahistydynowy (**6xHIS**) do oczyszczania przez chromatografię powinowactwa



Ekspresja superstabilnych („superfolderowych”) pochodnych **GFP**, **RFP** w komórkach *E. coli* – pokaz szalek – światło widzialne

Pochodne **GFP**
(„green fluorescent protein” z meduzy *Aequorea victoria*):

sfGFP - green

sfBFP - blue

sfCFP - cyan

sfYFP - yellow

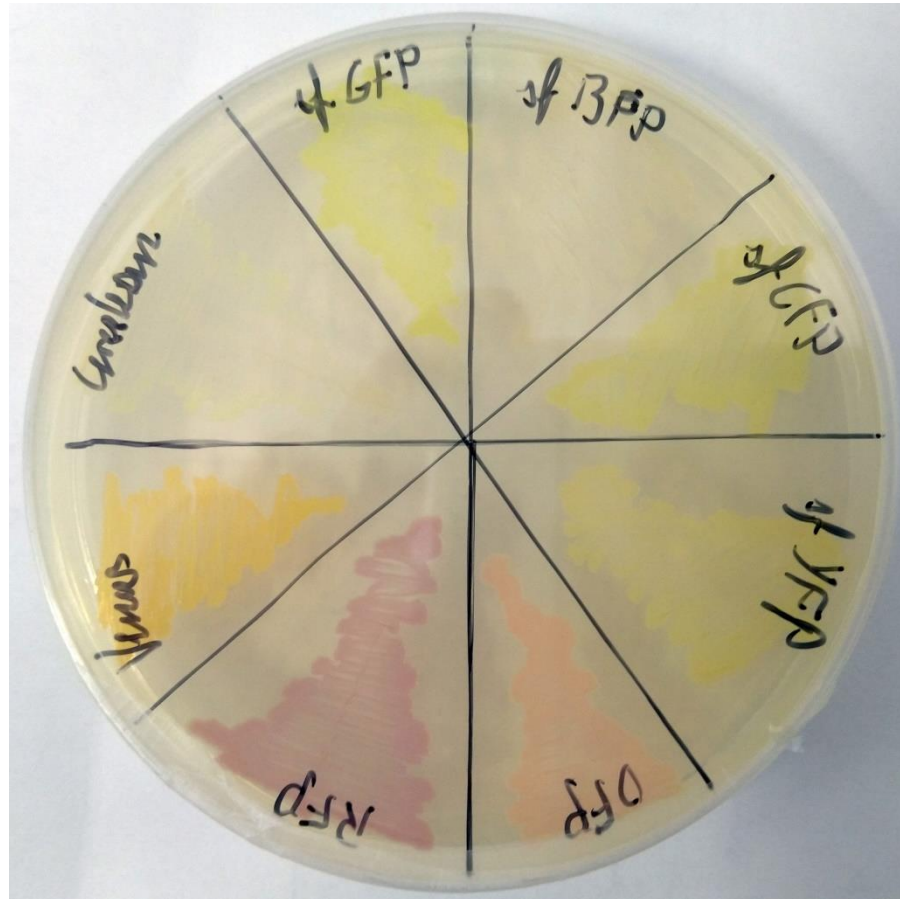
Venus

Cerulean

Pochodne **RFP**
(„red fluorescent protein” z koralowca):

OFP - orange

RFP – red



Wzorzec w
świecie
widzialnym

Wykorzystano
„Biobricks”;
iGEM

sf – „supefolder” – punktowe mutacje zwiększające stabilność białka
Konstrukty wyrażane z silnych, konstytutywnych promotorów

<https://www.nature.com/articles/nbt1172>

<http://2013.igem.org/Team:Warsaw/Parts>

https://openwetware.org/wiki/IgEM:Cambridge/2008/Improved_GFP

**Dla części widać kolor w świetle widzialnym, pełne wzbudzenie w świetle UV –
oglądamy pod trans iluminatorem, prosimy o ostrożność!!!**

Ekspresja superstabilnych („superfolderowych”) pochodnych **GFP**, **RFP** w komórkach *E. coli* – pokaz szalek – **światło UV**

Pochodne **GFP**
(„green fluorescent protein” z meduzy *Aequorea victoria*):

sfGFP - green

sfBFP - blue

sfCFP - cyan

sfYFP - yellow

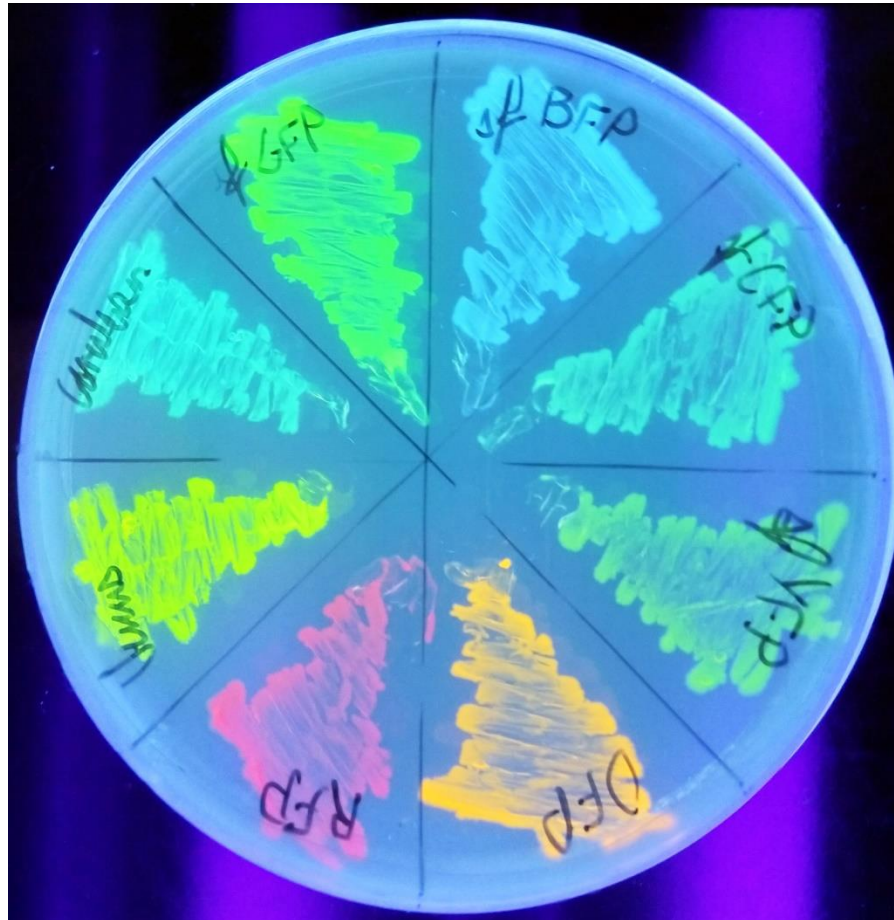
Venus

Cerulean

Pochodne **RFP**
(„red fluorescent protein” z koralowca):

OFP - orange

RFP – red



Wzorzec w
świecie
widzialnym

Wykorzystano
„Biobricks”;
iGEM

sf – „supefolder” – punktowe mutacje zwiększające stabilność białka
Konstrukty wyrażane z silnych, konstytutywnych promotorów

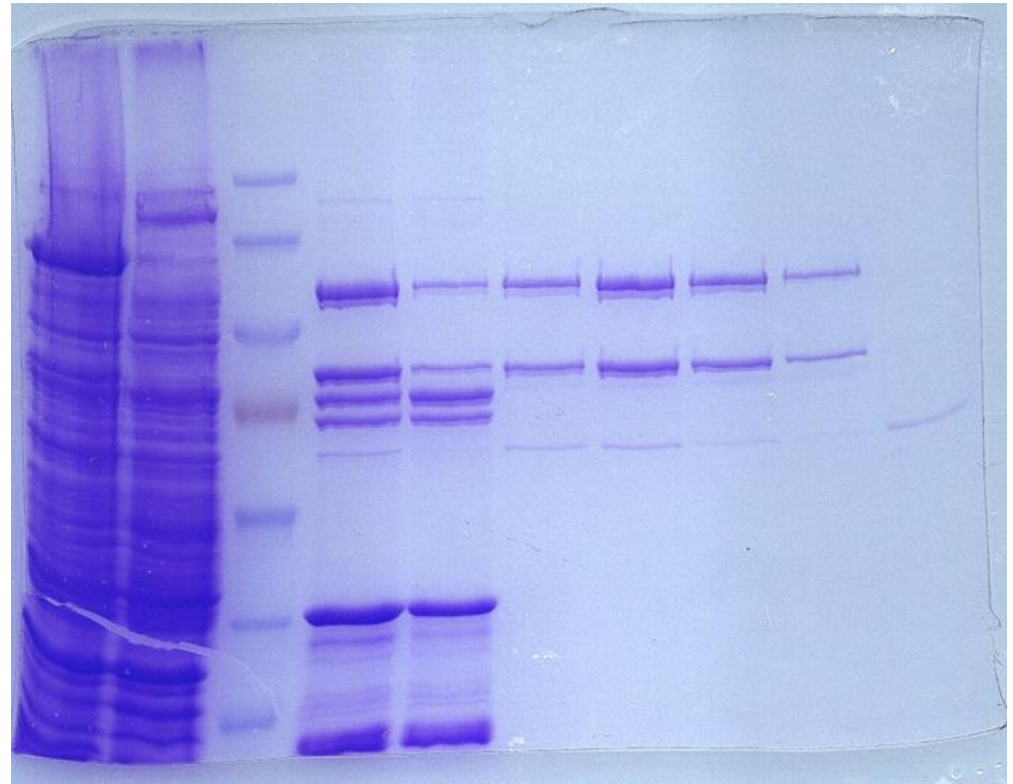
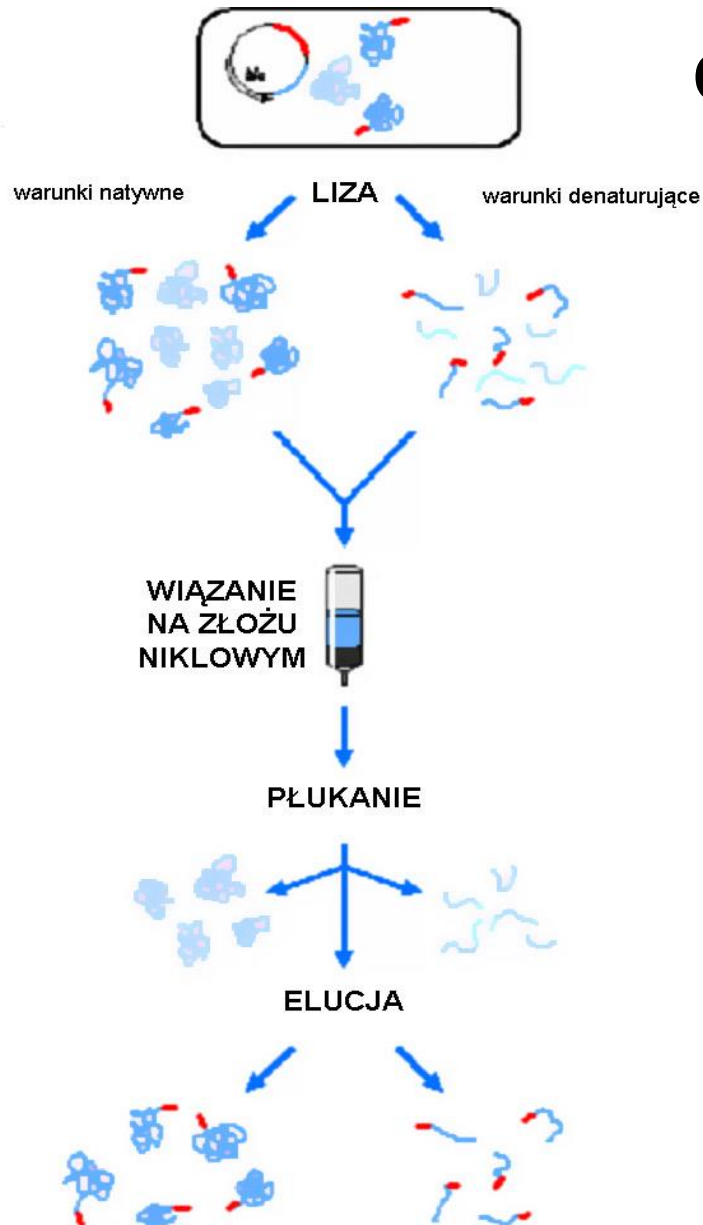
<https://www.nature.com/articles/nbt1172>

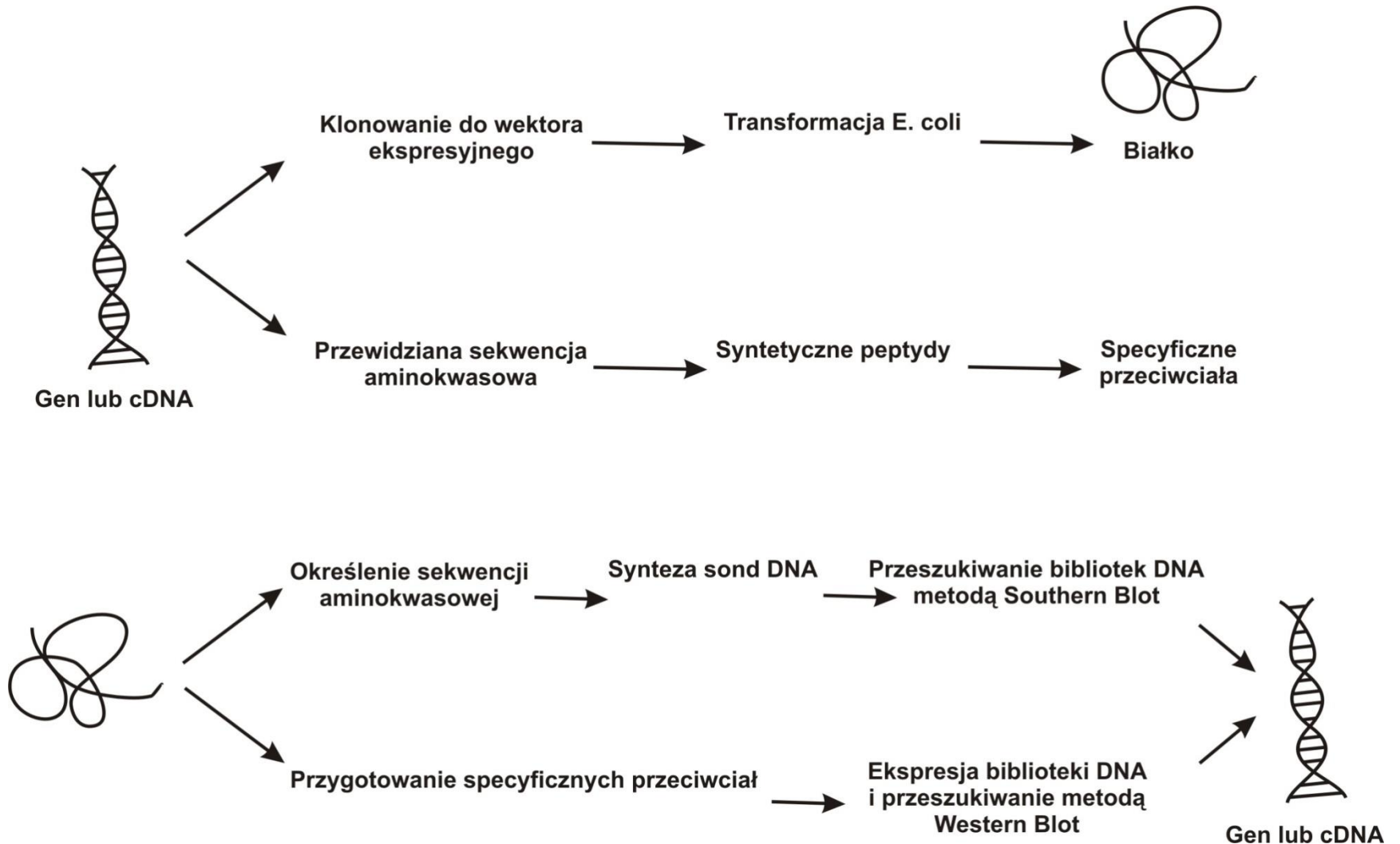
<http://2013.igem.org/Team:Warsaw/Parts>

https://openwetware.org/wiki/IgEM:Cambridge/2008/Improved_GFP

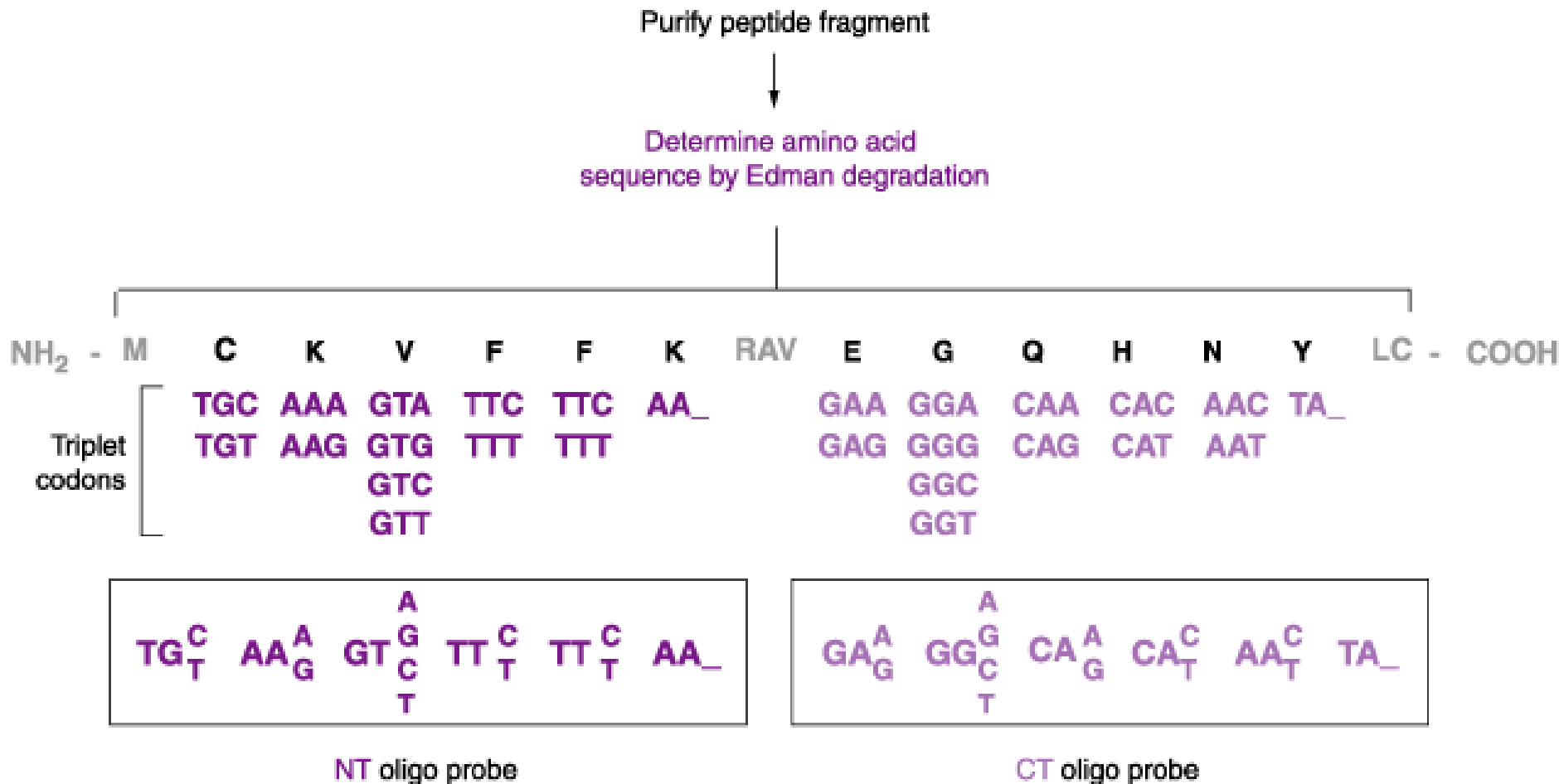
Dla części widać kolor w świetle widzialnym, pełne wzbudzenie w świetle UV – oglądamy pod trans iluminatorem, prosimy o ostrożność!!!

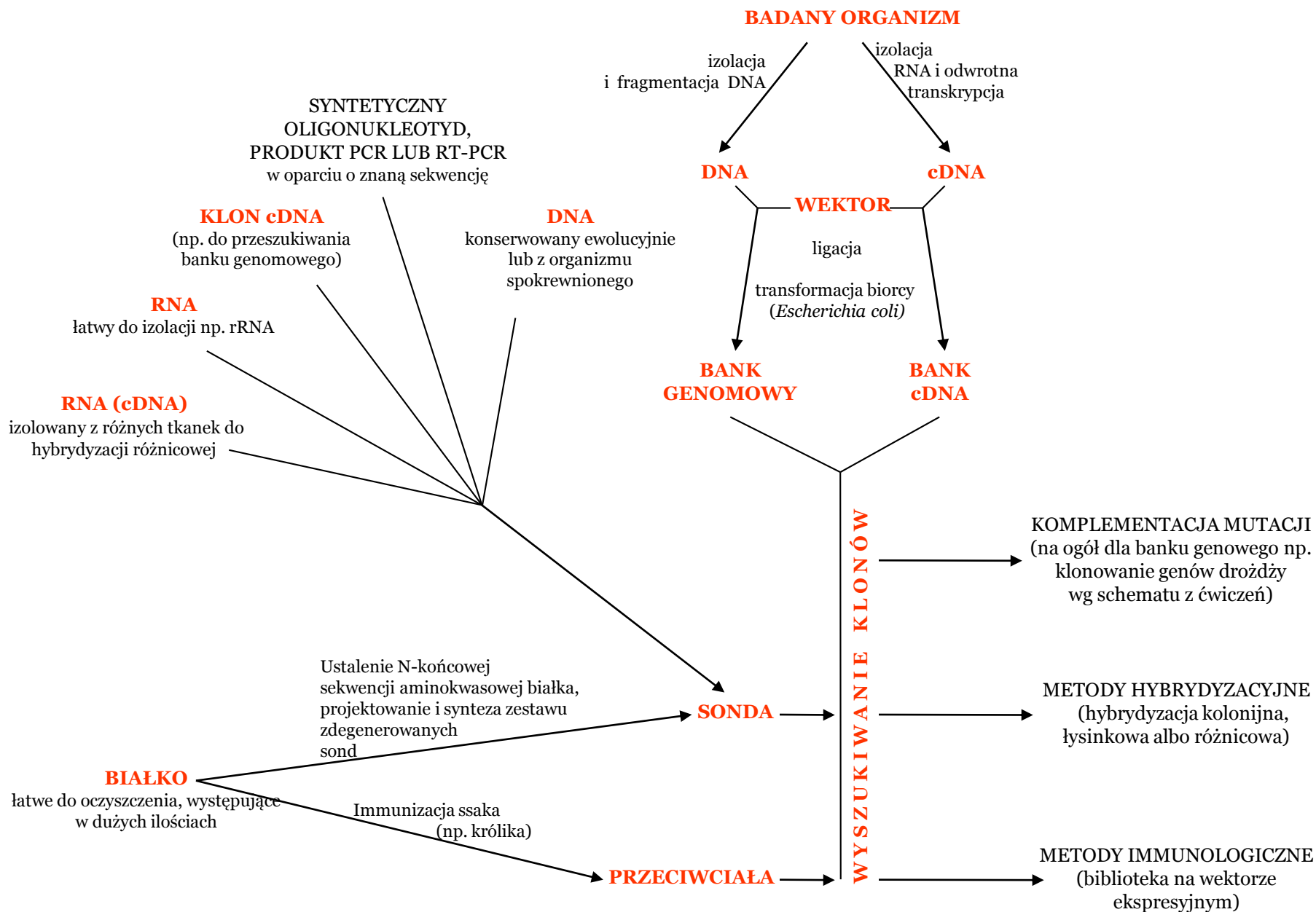
Chromatografia powinowactwa na przykładzie złoża niklowego i heksahistydyny





Konstrukcja zdegenerowanej sondy





HYBRYDYZACJI KWASÓW NUKLEINOWYCH – TECHNIKA „NORTHERN-BLOT”

WSTĘP

Technika northern-blot pozwala na badanie ekspresji genów na poziomie RNA. Hybrydyzacja typu northern polega na rozdzieleniu w żelu RNA (na ogół w warunkach denaturujących), przeniesieniu na filtr i wykonaniu hybrydyzacji z sondą, będącą zwykle znakowanym (znacznikami radioaktywnymi lub fluorescencyjnymi) fragmentem DNA, komplementarnym do badanego przez nas RNA. Poniższa, przykładowa hybrydyzacja northern pokazuje badanie różnic w ekspresji genu *agaA*, kodującego arginazę u *Aspergillus nidulans*.

KATABOLIZM ARGININY u *A. nidulans*

Arginaza rozkłada argininę do mocznika i ornityny. Reakcja ta jest elementem cyklu mocznikowego, pozwala na wykorzystanie argininy jako źródła węgla i azotu, dostarcza ornityny wykorzystywanej m.in. do syntezy poliamin. Ekspresja genu *agaA*, podlega kontroli specyficznej, związanej z indukcją arginina. Aktywność arginazy w grzybniach dzikich szczepów wzrasta ok. dziesięciokrotnie w obecności argininy. Za efekt ten odpowiada produkt genu *arcA* (ang. „arginine catabolism activator”). Białko ARCA jest czynnikiem transkrypcyjnym o wielkości 600 aa, zawiera motyw palca cynkowego typu Zn_2Cys_6 , potencjalną domenę lokalizacji jądrowej, regiony kwaśne oraz potencjalne miejsca fosforylacji. Wyizolowano mutanty w *locus* genu *arcA*: allel dominujący *arcA^{d47}* (d – „derepressed”), powoduje konstytutywnie podniesione poziomy arginazy (tzn. poziomy jak po indukcji szczepu dzikiego arginina, bez względu na jej obecność) oraz allel recesywny *arcA^{r3}* (r – „repressed”), powodujący brak indukcji arginazy, bez względu na obecność argininy.

Warunki hodowli

Grzybnie hodowano przez 19h w 37°C z wytrząsaniem, w płynnej pożywce minimalnej (MM, warunki nie indukcyjne) lub w MM z arginina (warunki indukcyjne). Oznaczono aktywności arginazy w grzybniach (patrz tabela poniżej) i wyizolowano RNA totalne.

Sonda użyta do hybrydyzacji

Jako sondę wykorzystano produkt PCR będący fragmentem sekwencji kodującej arginazy z *A. nidulans* o wielkości 568 bp. Sondę znakowano α -³²P dATP, metodą „random-priming”.

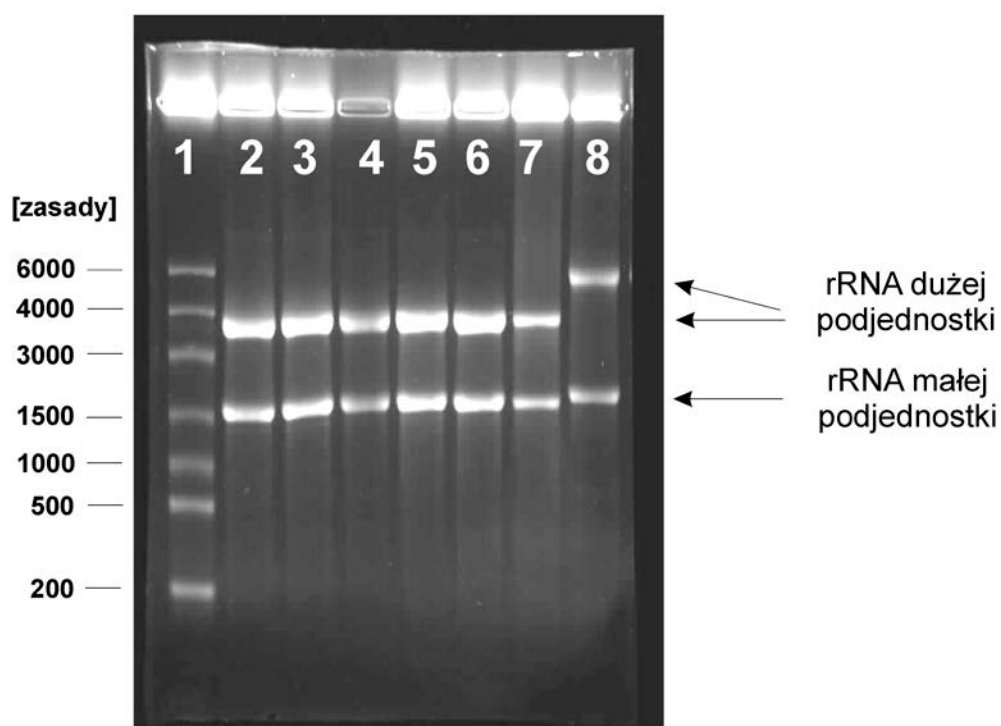
Wykonanie hybrydyzacji northern

RNA totalne rozdzielano w 1% żelu agarozowym w warunkach denaturujących (z dodatkiem formaliny, próbki przed nałożeniem denaturowano w 60°C). Rozdzielone RNA przeniesiono na filtr nitrocelulozowy wykorzystując metodę transferu kapilarnego, filtr utrwalono (związanie RNA z filtrem) „zapiękając” pod światłem UV. Prehybrydyzację prowadzono w czasie znakowania sondy, właściwa hybrydyzacja przeprowadzono w 65°C, (ostre warunki). Następnie odpłukano niezwiązaną sondę i filtr poddano autoradiografii w systemie Phospho Imager.

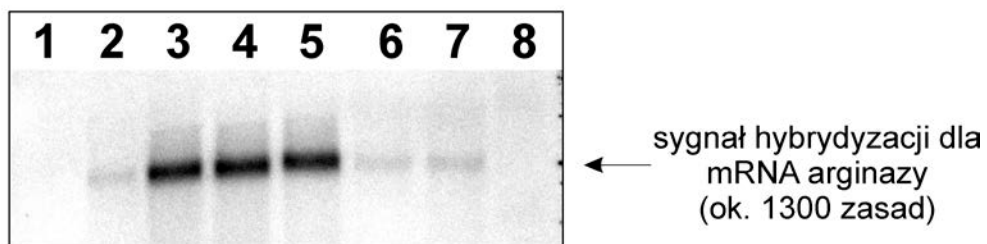
Nakładane próbki RNA:

1.	standard masy cząsteczkowej RNA (RNA Lader High Range, MBI Fermentas)	warunki hodowli	aktywność arginazy
2.	wt	MM	0,2
3.	wt	MM + arginina	1,8
4.	<i>arcA^{d47}</i>	MM	1,9
5.	<i>arcA^{d47}</i>	MM + arginina	2,1
6.	<i>arcA^{r3}</i>	MM	0,1
7.	<i>arcA^{r3}</i>	MM + arginina	0,1
8.	RNA totalne z ludzkich komórek białaczkowych Jurkat	<i>1u aktywności arginazy = aktywność enzymu dająca 1 μmol mocznika na minutę w warunkach standardowych</i>	

OBRAZ ELEKTROFOREZY

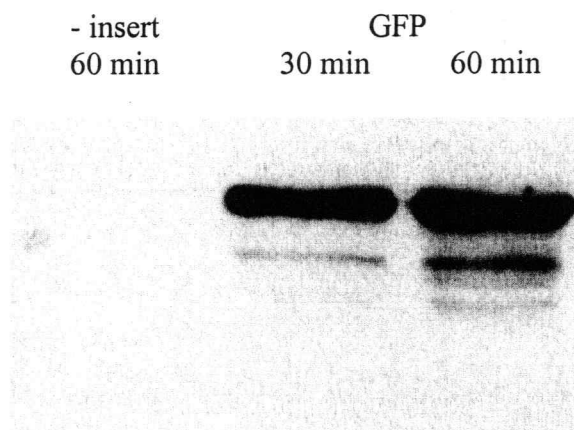


WYNIK AUTORADIOGRAFII

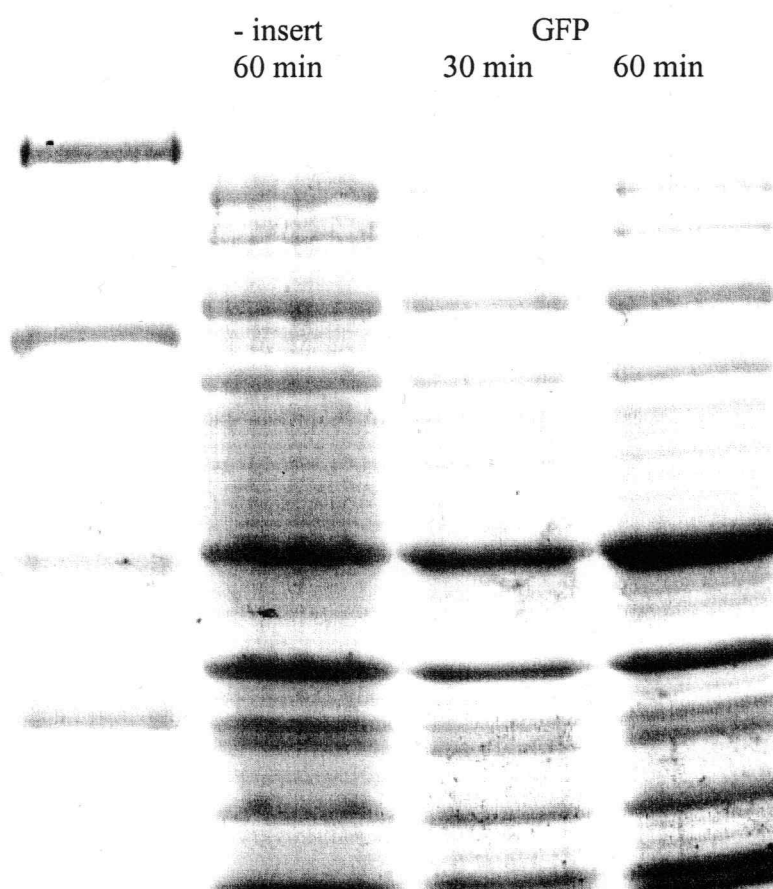


Analiza typu Western

Białko GFP zostało wyrażone w systemie pQE przez 30 i 60 min. Totalne ekstrakty bakteryjne zostały rozdzielone przy użyciu techniki SDS-PAGE, a następnie poddane analizie typu Western. Zastosowano specyficzne przeciwciała anty-GFP. Jako kontrolę wyrażano przez 60 min. plazmid nie zawierający insertu.



Rys 1. Analiza typu Western



Rys 2. Analiza typu SDS-PAGE