



# **Zaawansowane Metody Analizy Danych w Biologii Molekularnej**

Semestr letni 2018



# **Ekspresja genów**

Semestr letni 2018

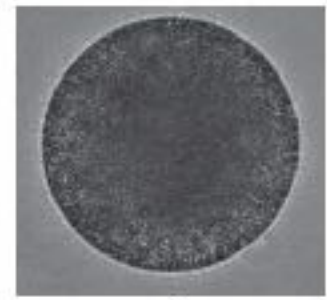
# Materiał genetyczny a zróżnicowanie komórkowe



(A) 100 μm



(C) 50 μm



(E) 50 μm



(B)



(D)



(F)



# Minimalne wymagania genetyczne

Ile genów potrzebuje żywa komórka aby realizować plan biologiczny?

- Istnieje ~60 genów wspólnych dla wszystkich organizmów
- Zakłada się, że 200-300 genów to minimum zapewniające przetrwanie
- *Mycoplasma genitalium* ma 480 genów, genom składa się z 580070 par nukleotydów

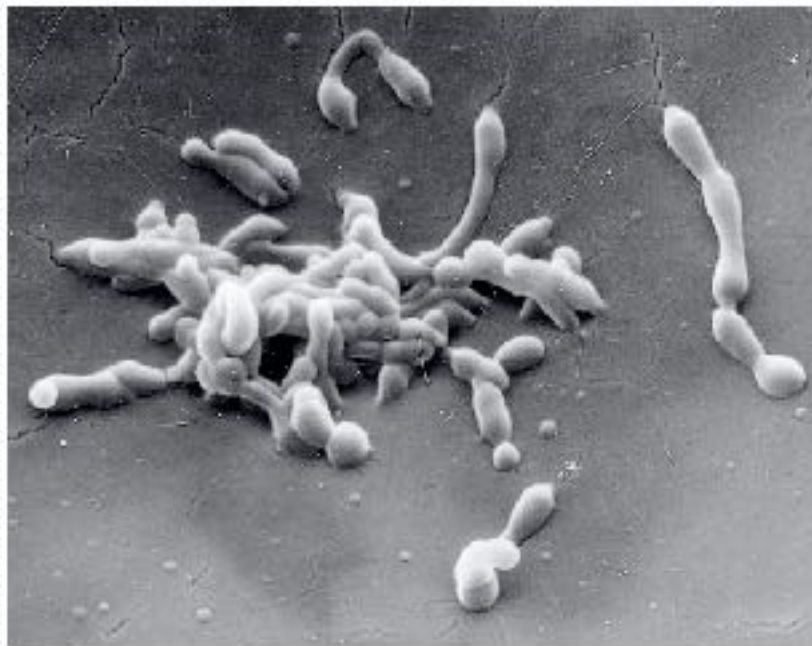


Figure 1-14a Molecular Biology of the Cell 5/e © Garland Science 2008



# Minimalne wymagania genetyczne

## *Mycoplasma genitalium*

- Liczba specyficznych genów:
  - 37 dla RNA
  - 297 koduje białka:
    - 153 związane z replikacją, transkrypcją, translacją
    - 29 współtworzących struktury błony komórkowej
    - 33 odpowiedzialnych za transport przez błonę komórkową
    - 71 związanych z przetwarzaniem energii, syntezą i degradacją drobnych molekuł
    - 11 związanych z podziałem komórkowym



Figure 1-14B Molecular Biology of the Cell 5/e (© Garland Science 2008)

0.2 μm



# Struktura DNA

## (A) building block of DNA



## (B) DNA strand

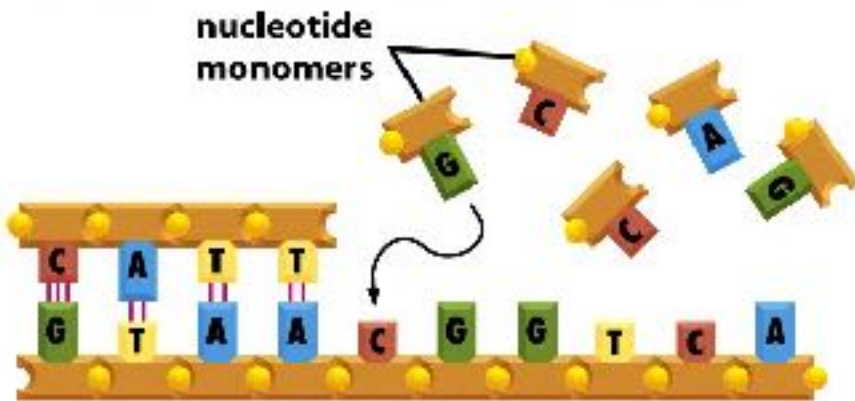
asymetria-nadaje kierunkowość



## (C) templated polymerization of new strand

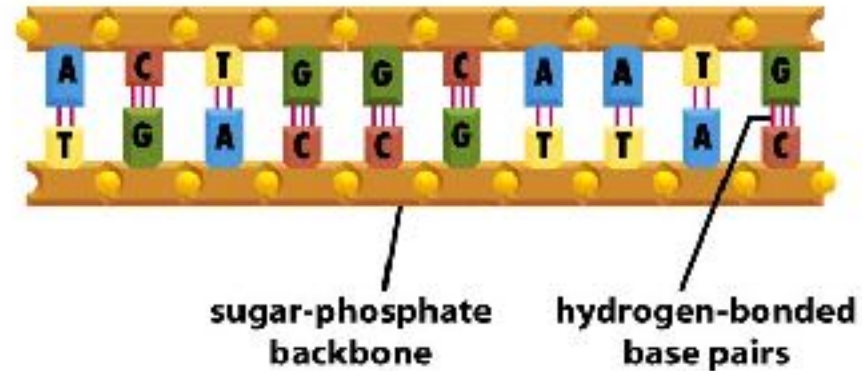
cukrowo-fosforanowy łańcuch

nucleotide monomers



## (D) double-stranded DNA

wiązania kowalencyjne i wodorowe



## (E) DNA double helix

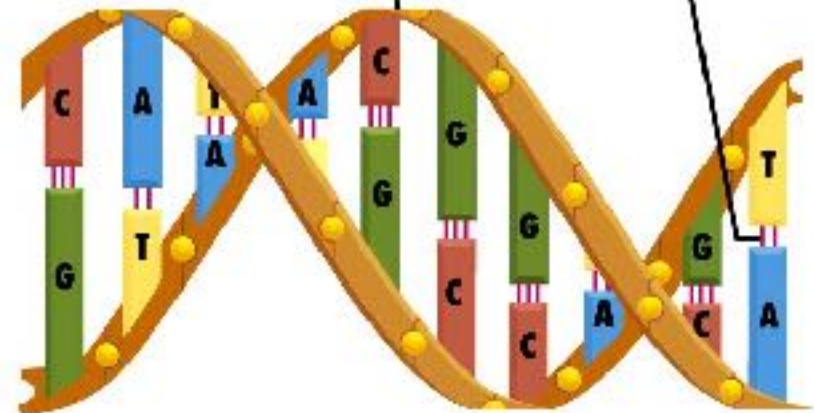
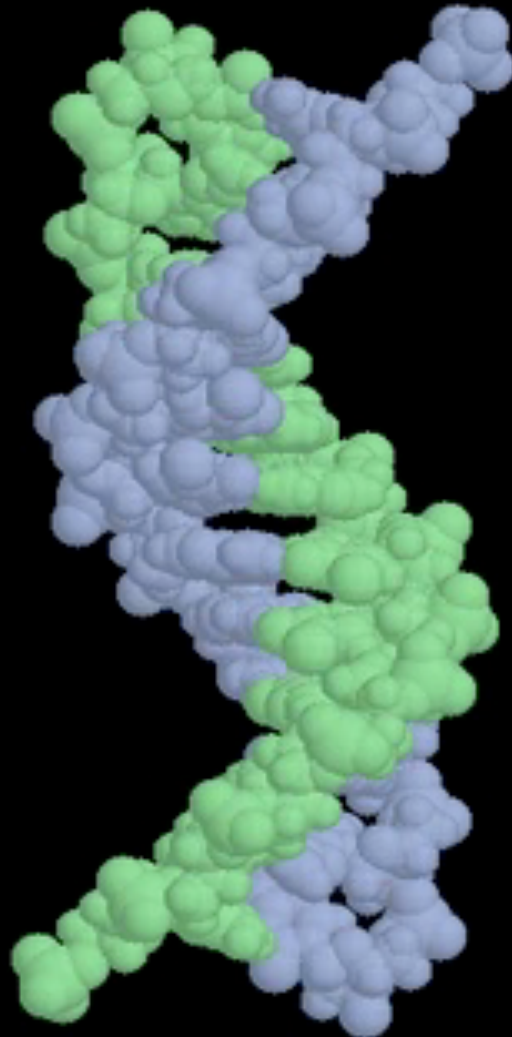


Figure 1-2 Molecular Biology of the Cell 5/e (© Garland Science 2008)

polimeryzacja matrycy - kontrola nici potomnej





# Replikacja DNA

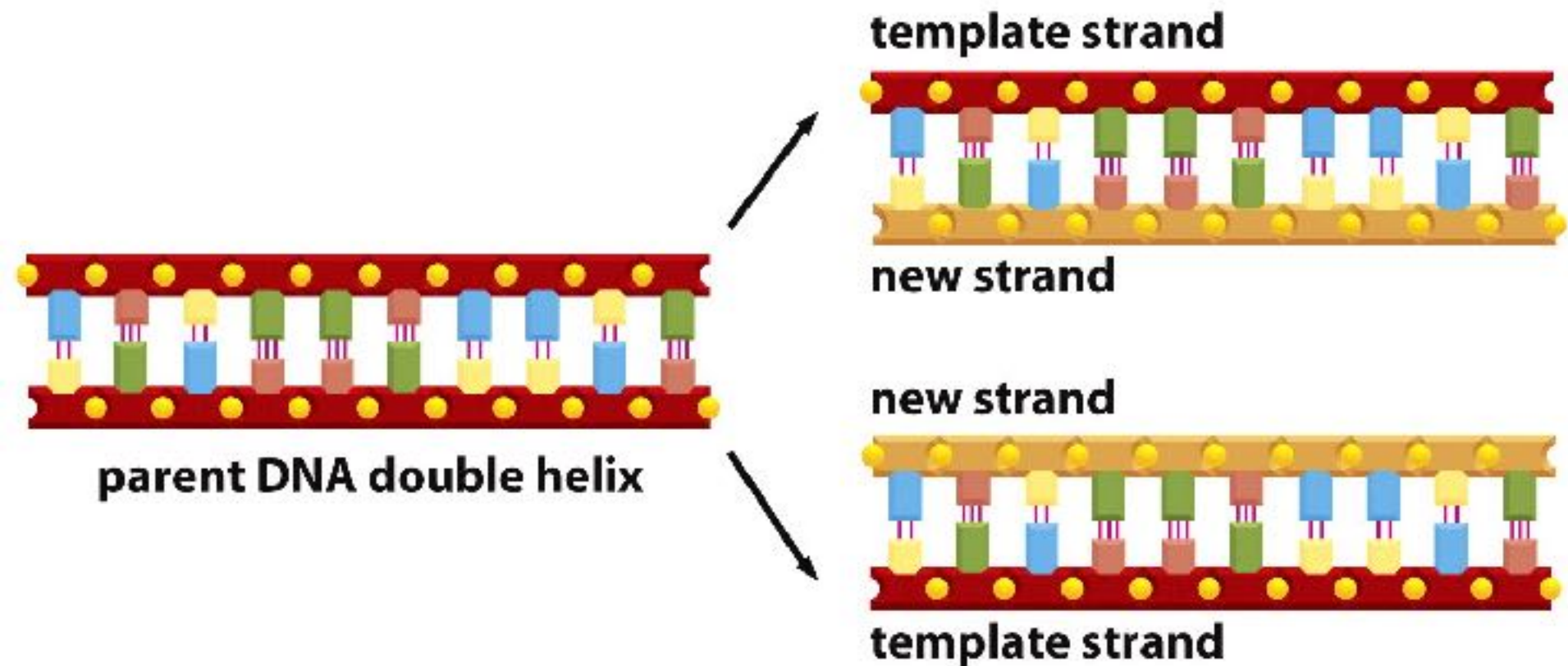


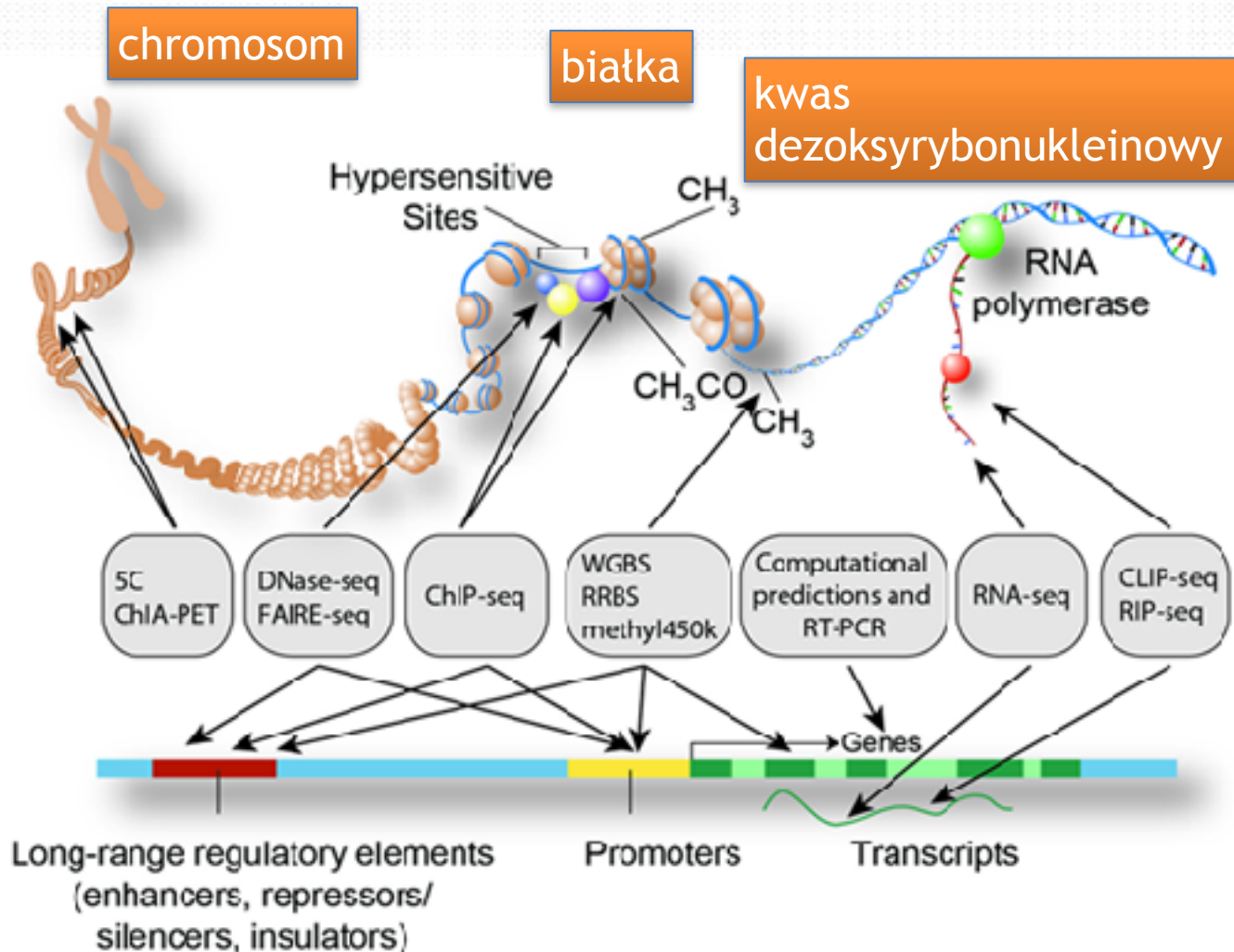
Figure 1-3 Molecular Biology of the Cell 5/e (© Garland Science 2008)

Rozerwanie podwójnej nici i użycie jako matrycy nici pojedynczych





# Materiał genetyczny

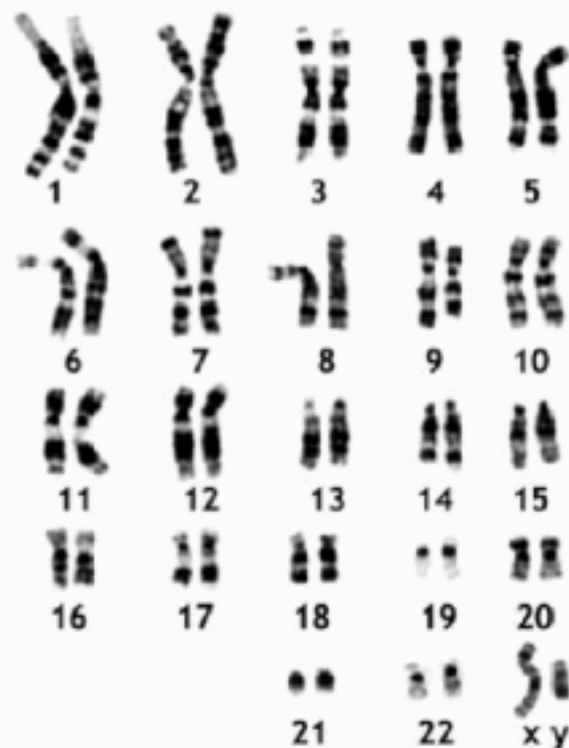


- DNA w komórce występuje w postaci bardzo długich nici
- U ludzi genom jest podzielony na 23 pary chromosomów (od ~19 do 240 milionów par zasad)
- Odcinki sekwencji zawierające informacje potrzebne do utworzenia RNA i białek nazywamy genami

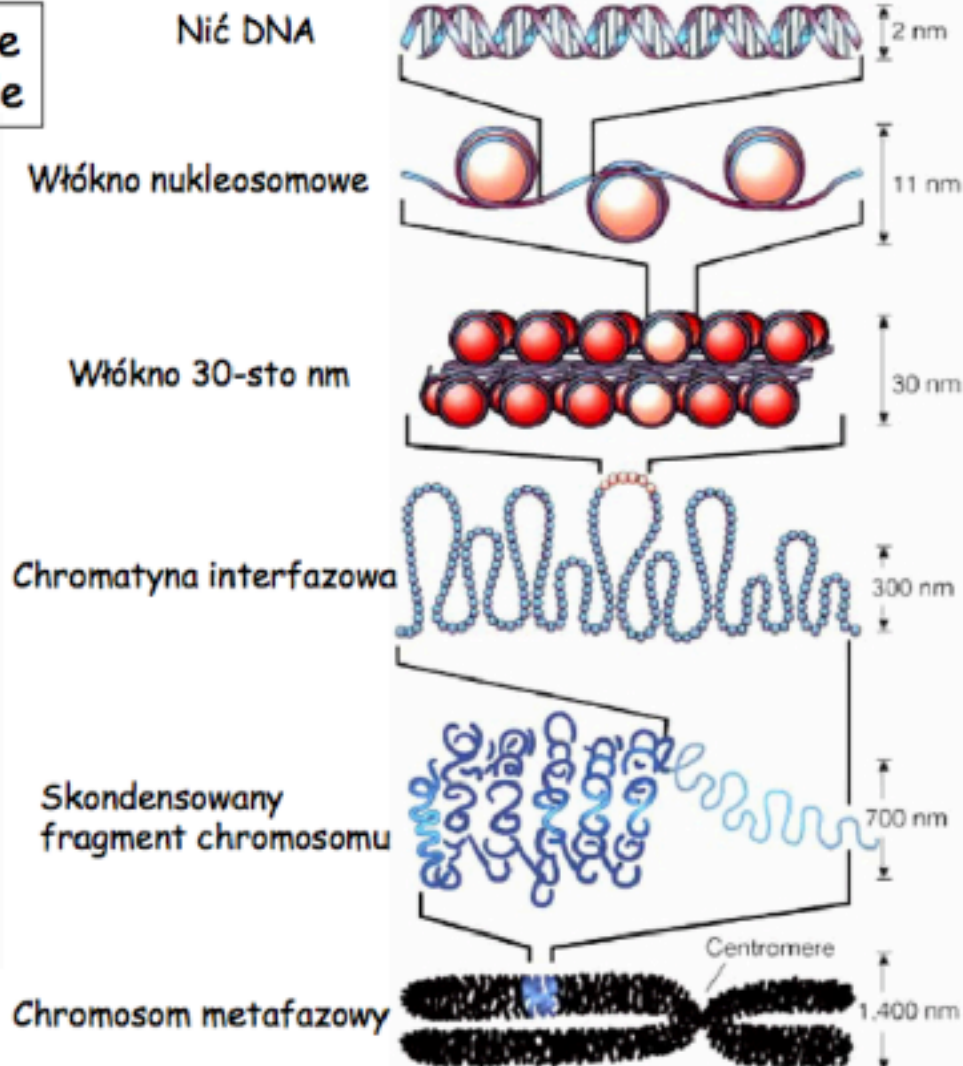


# DNA w jądrze jest ściśle upakowane

- 2 metry DNA w każdej komórce
- $5 \times 10^{10}$  km DNA w ludzkim ciele

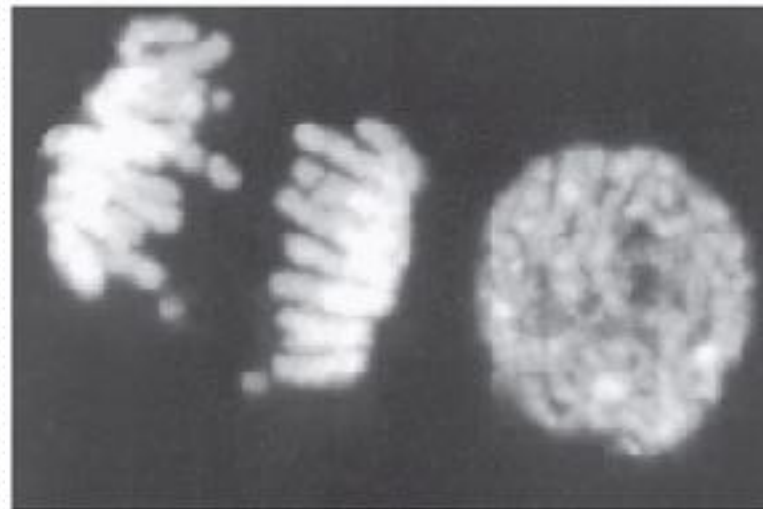


Całkowita długość 46 chromosomów metafazowych wynosi tylko 200  $\mu\text{m}$

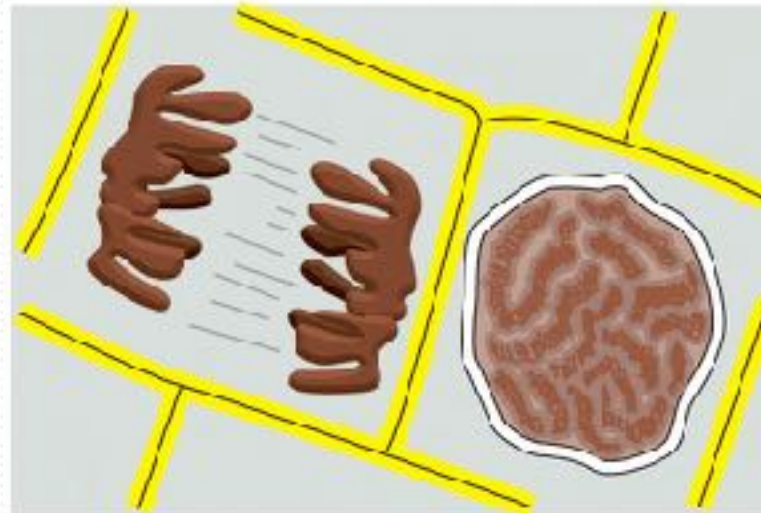




# Chromosomy w komórce



(A)      dividing cell      nondividing cell



(B)      10 μm

Figure 4-1 Molecular Biology of the Cell 5/e (© Garland Science 2008)



# Nukleosom

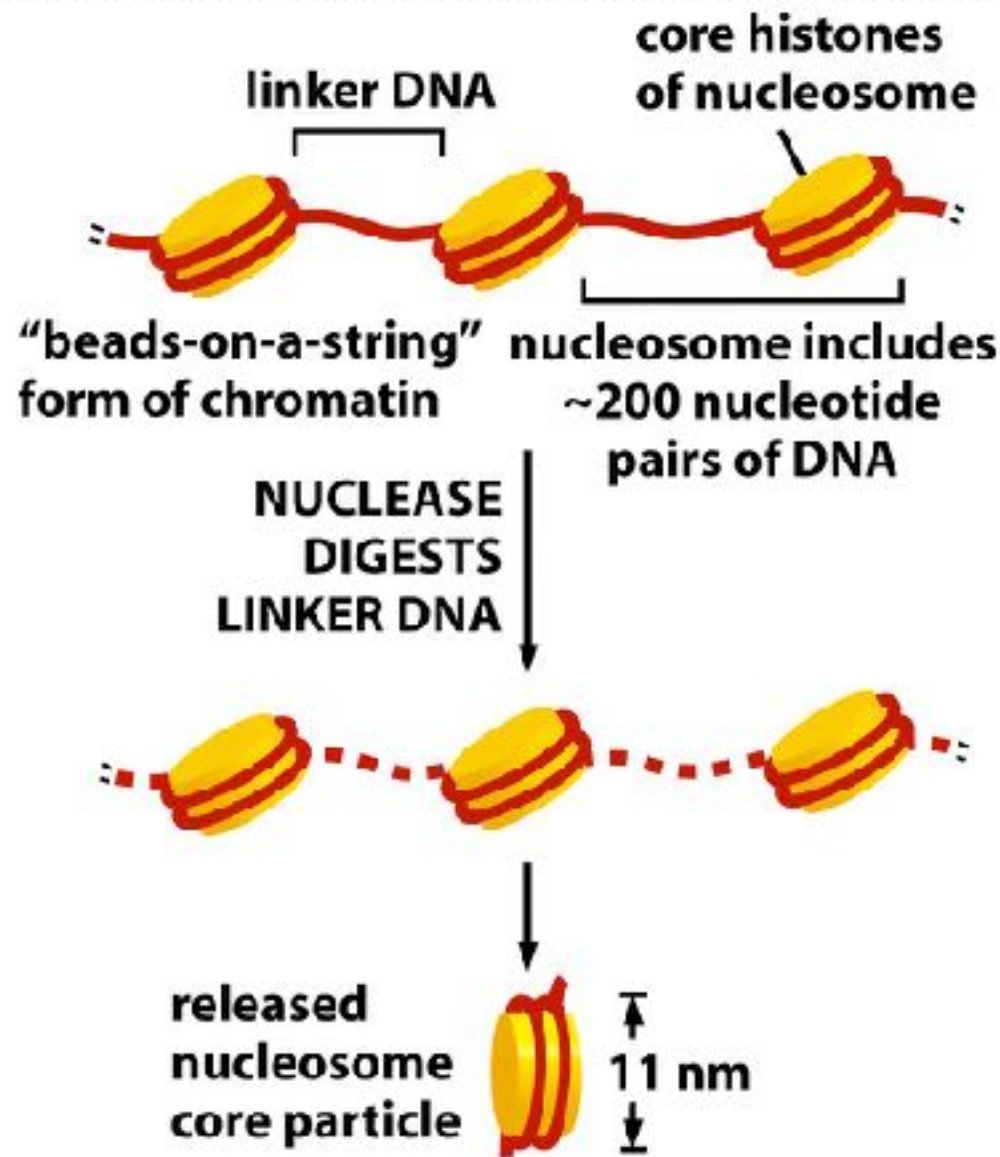


Figure 4-21 part 1 of 2 Molecular Biology of the Cell 5/e (© Garland Science 2008)





# Nukleosom

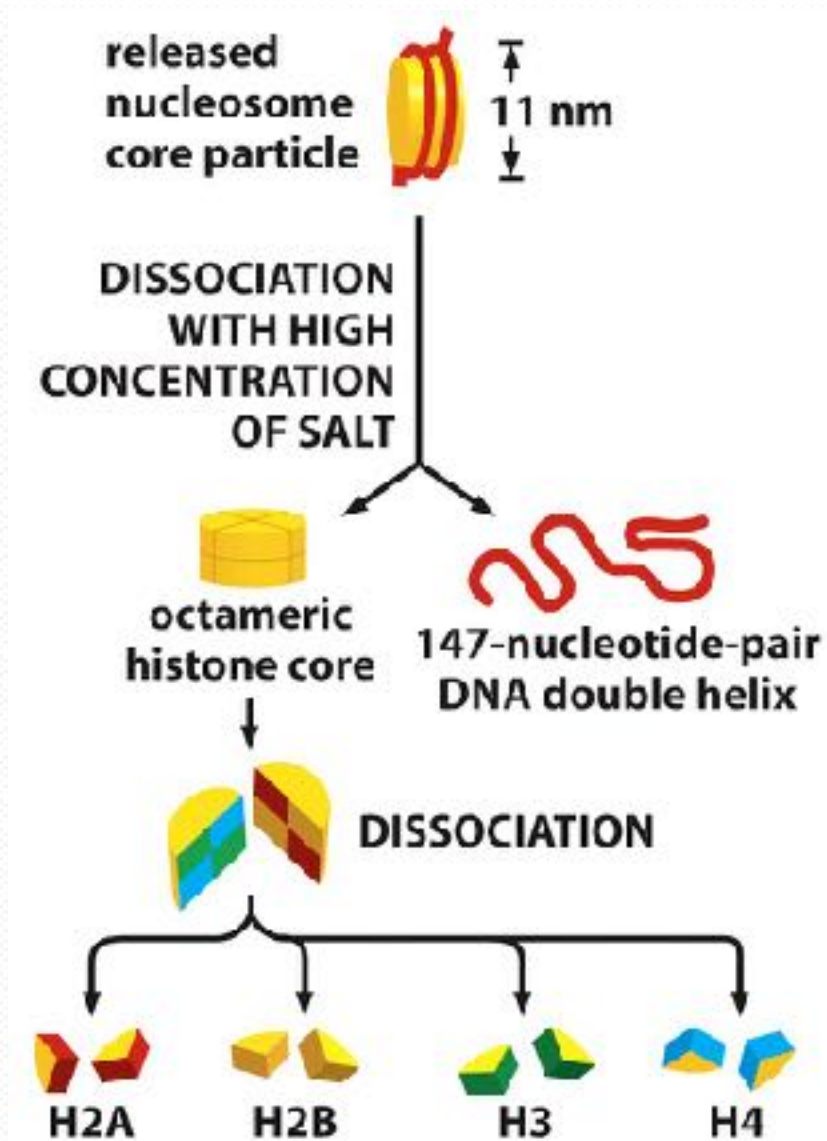


Figure 4-23 part 2 of 2 Molecular Biology of the Cell 5/e (© Garland Science 2008)



# Nukleosom

Sekwencja aminokwasów histonu H4 u groszku i krowy różni się 2 pozycjami ze 102.

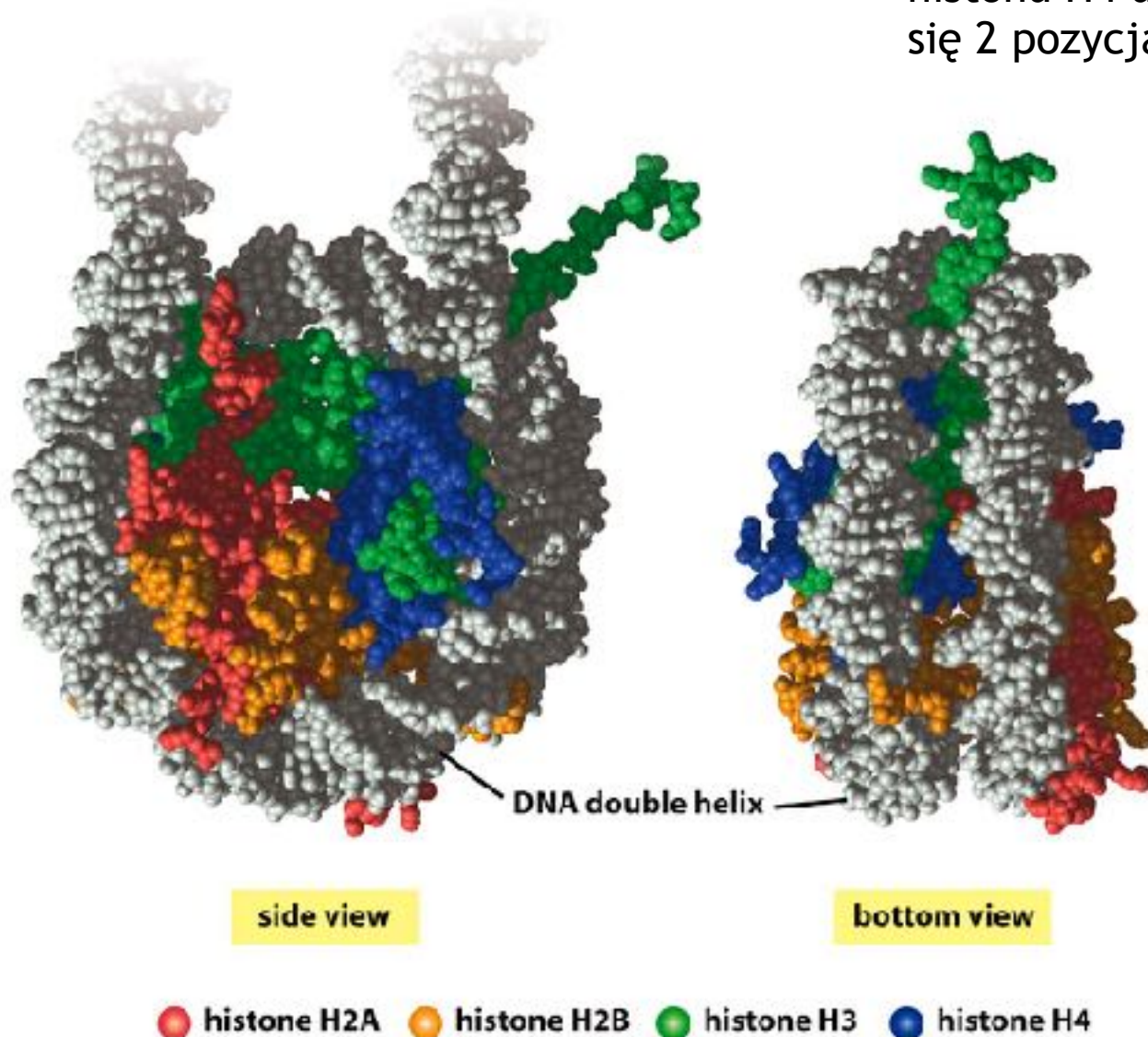


Figure 4-24 Molecular Biology of the Cell 5/e (© Garland Science 2008)





# Nukleosom

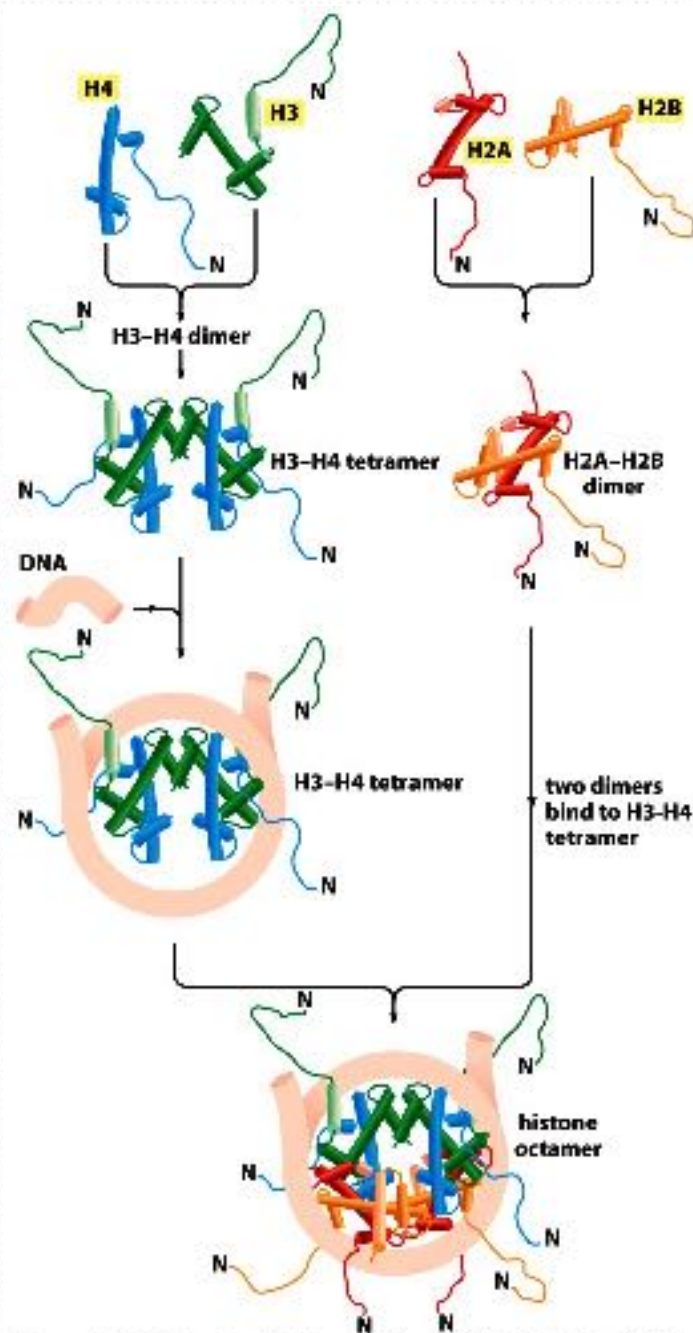
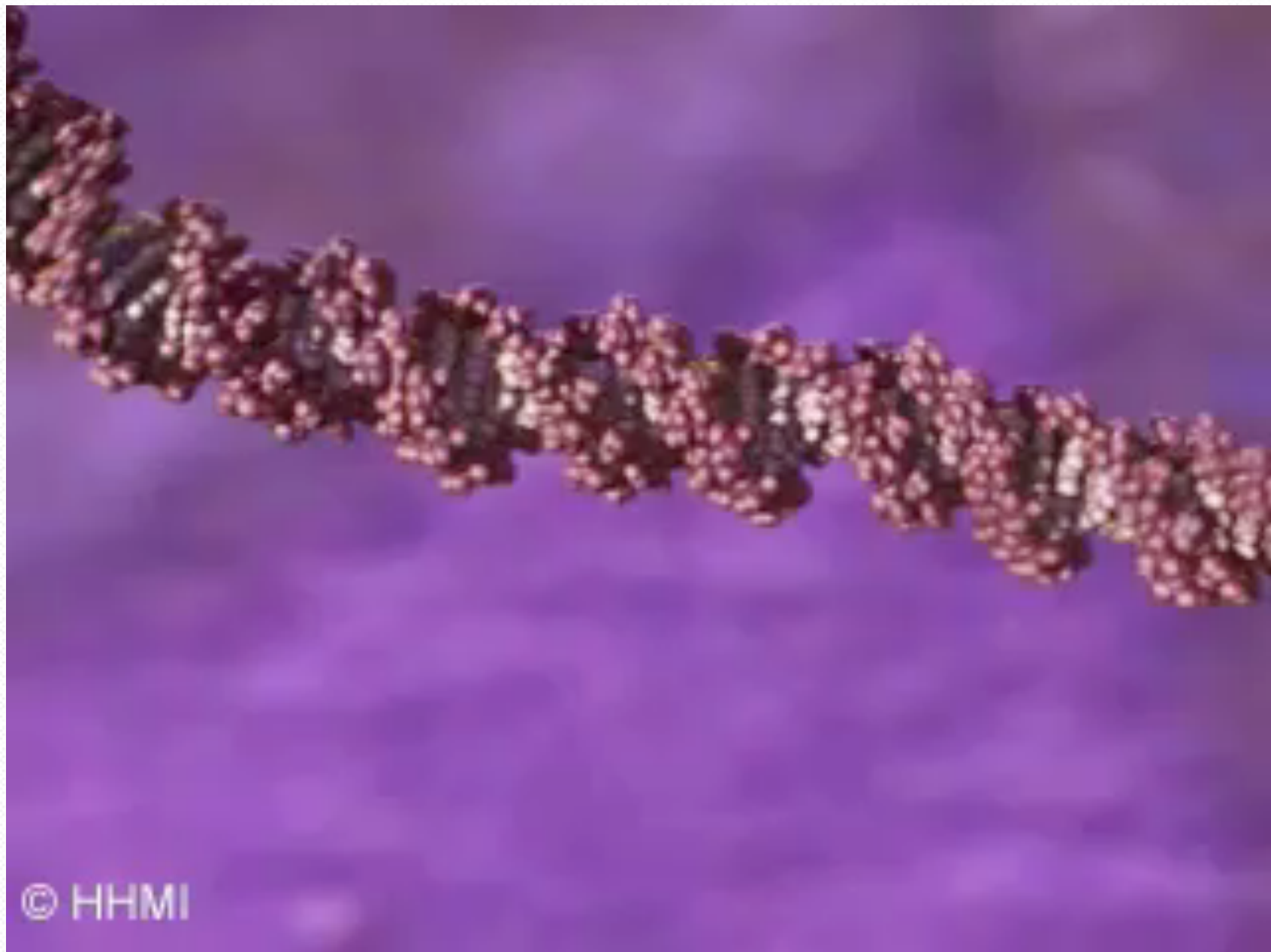


Figure 4-26 Molecular Biology of the Cell 5/e (© Garland Science 2008)





# Modyfikacje

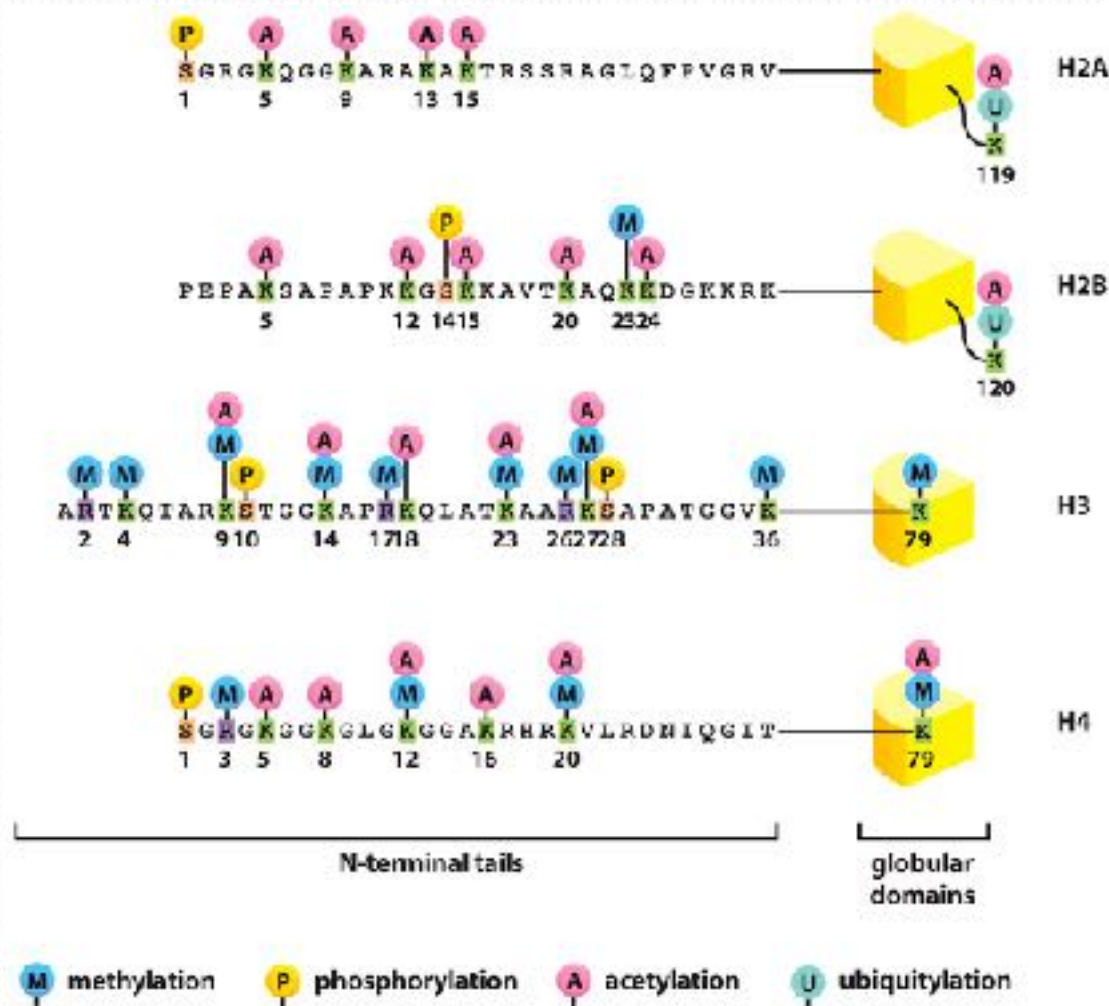


Figure 4-39b Molecular Biology of the Cell 5/e (© Garland Science 2008)

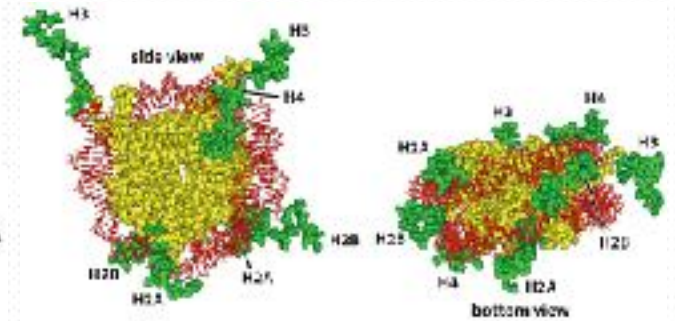
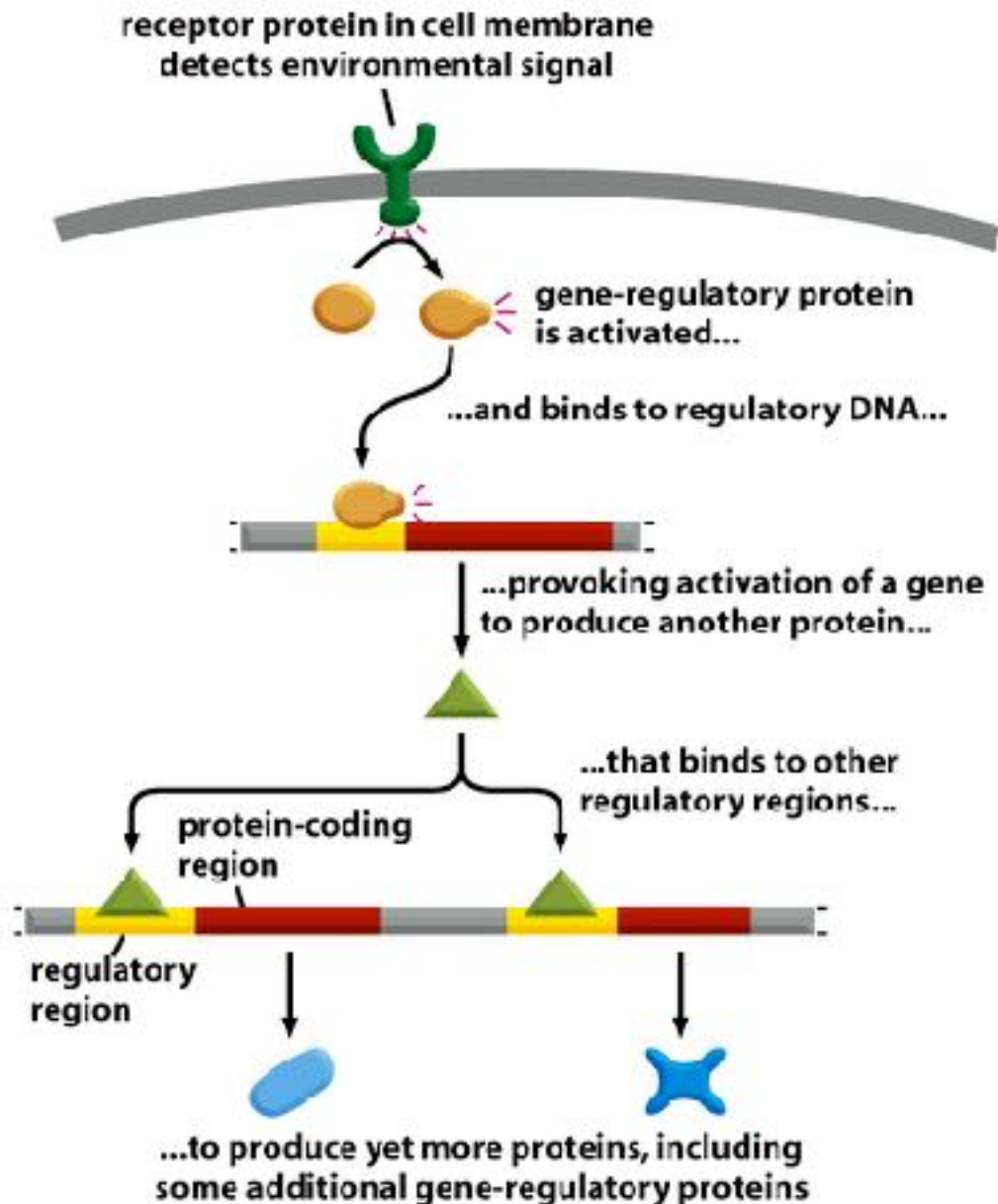


Figure 4-37a Molecular Biology of the Cell 5/e (© Garland Science 2008)



# Ścieżka sygnałowa



- Komórki wykonują instrukcje genetyczne poprzez procesy transkrypcji i translacji czyli ekspresję genów



# Od DNA do białka

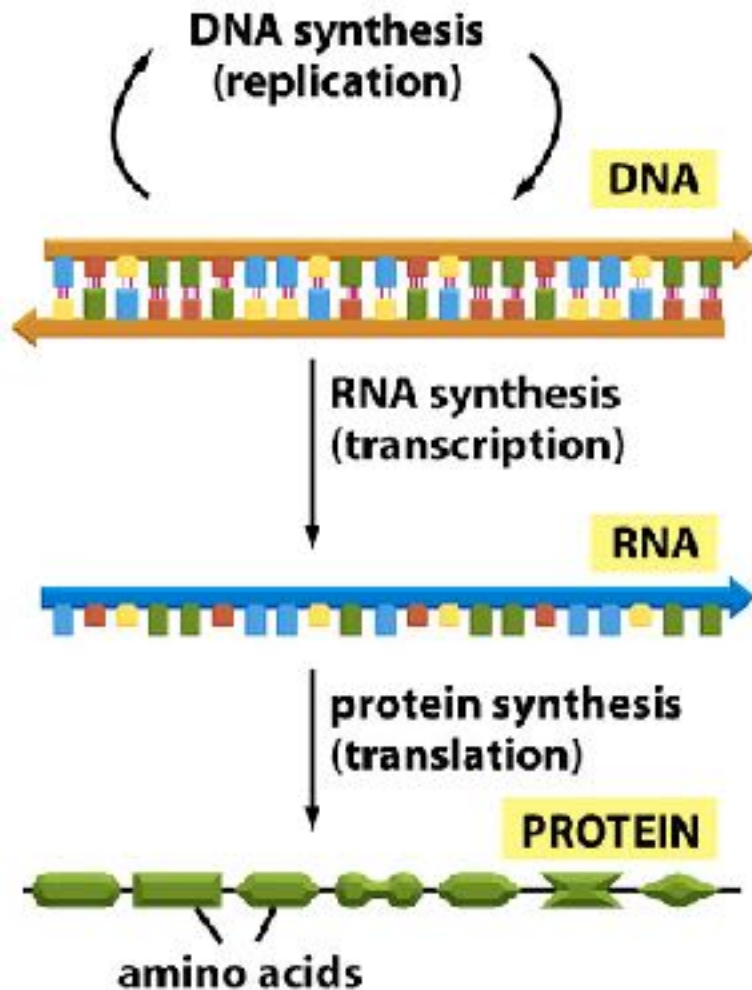


Figure 1-4 Molecular Biology of the Cell 5/e (© Garland Science 2008)

- Centralny dogmat biologii molekularnej – transkrypcja
- Informacja genetyczna przekazywana z DNA na RNA – przepisywana
- Informacja z RNA jest tłumaczona z języka nukleotydów na język aminokwasów.



# Transkrypcja

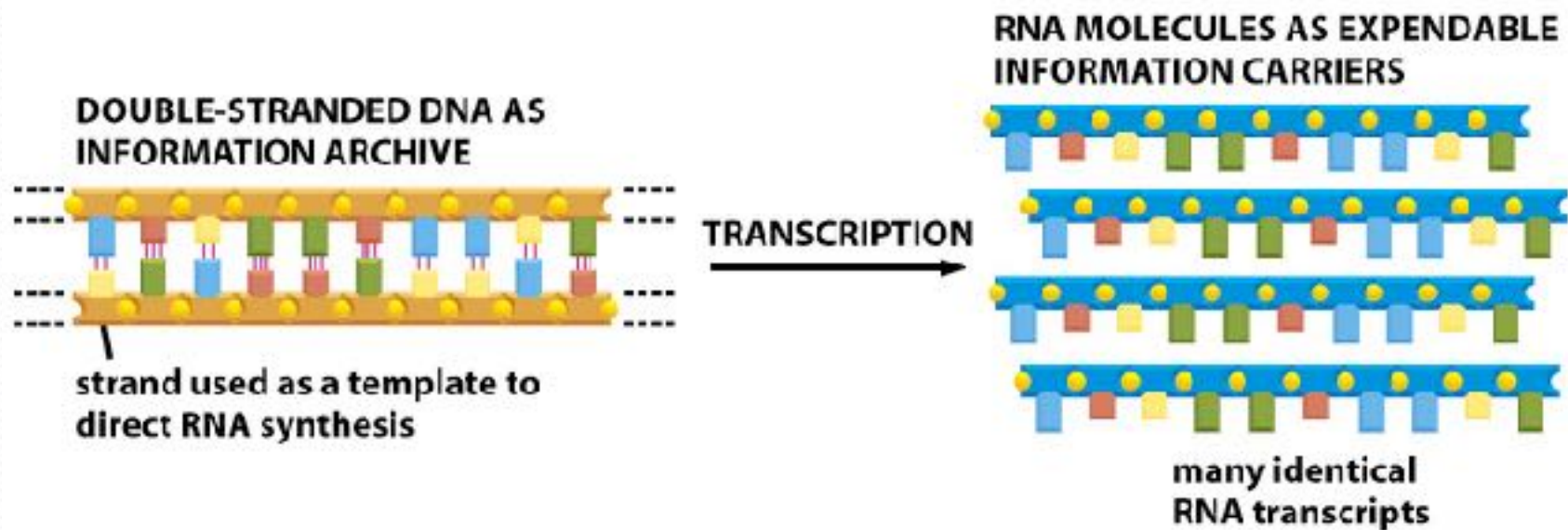
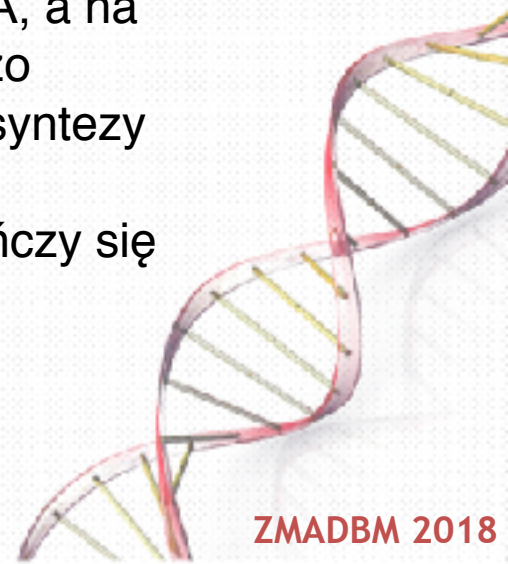


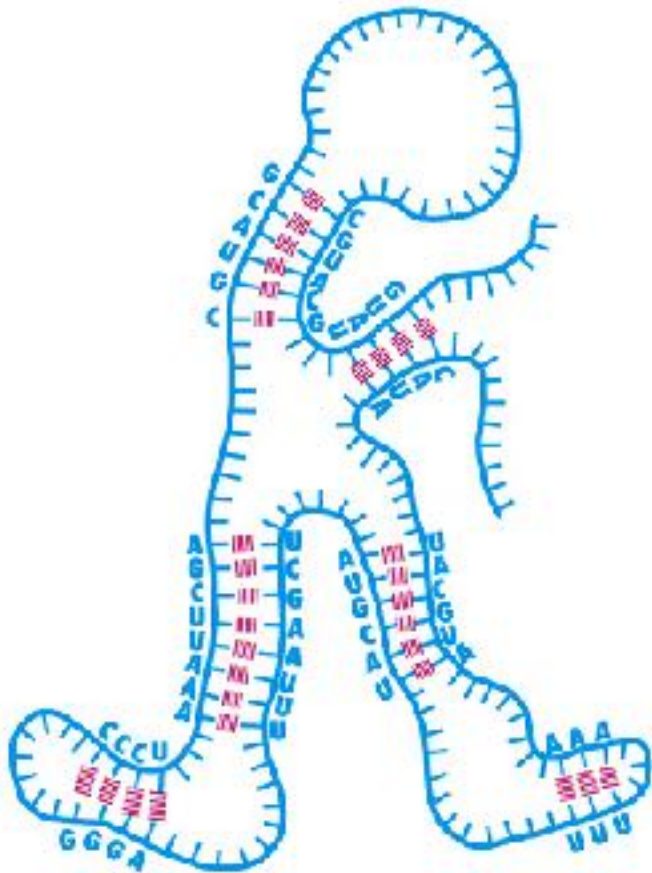
Figure 1-5 Molecular Biology of the Cell 5/e (© Garland Science 2008)

- Z jednego genu powstaje wiele identycznych cząsteczek RNA, a na bazie każdej cząsteczki RNA powstają cząsteczki białka. Dużo efektywniejsze niż gdyby DNA bezpośrednio był matrycą do syntezy białek.
- Kolejna cząsteczka RNA może podlegać syntezie zanim skończy się synteza wcześniejszej
- Geny różnią się wydajnością transkrypcji
- Transkrypcja: przepisanie bo w obębie “języka nukleotydów”
- RNA jest jednoniciowe





# RNA - wewnętrzcząsteczkowe pary zasad



**(A)**

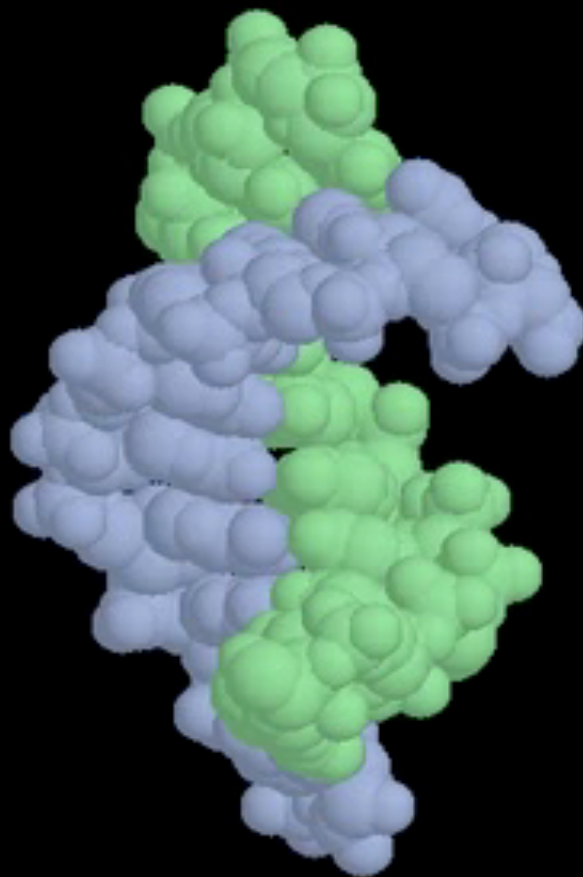


(B)

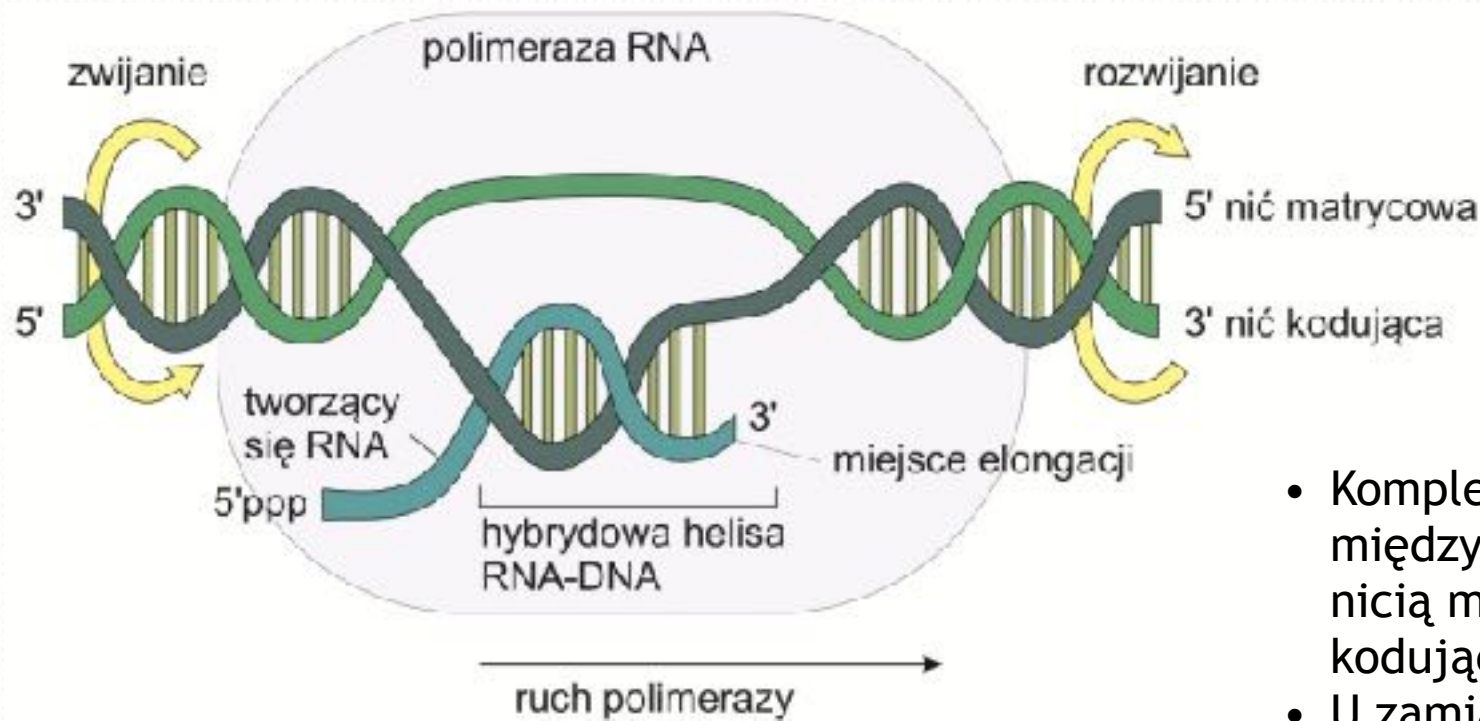
Figure 1-6 Molecular Biology of the Cell 5/e (© Garland Science 2008)



# Struktura RNA

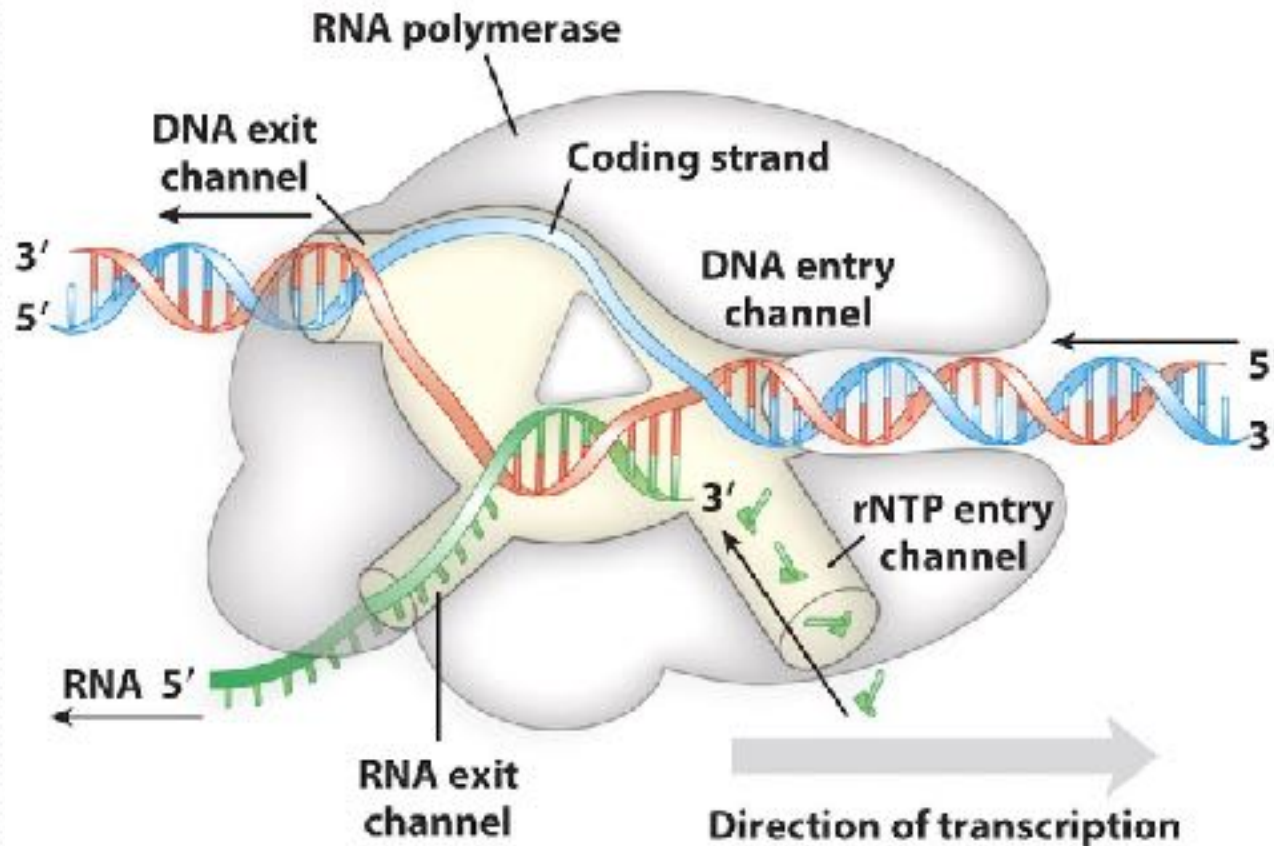






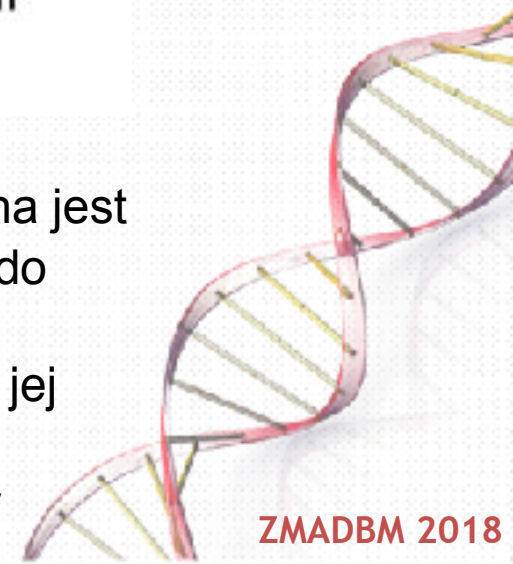
- Komplementarność między nicią RNA a nicią matrycową i kodującą DNA
- U zamiast T

- Synteza RNA jest katalizowana przez enzym polimerazę RNA (powstawanie wiązań fosfodiesterowych)
- Polimeraza wiąże się do nici DNA nazywanej matrycową
- Syntetyzowany jest fragment pomiędzy promotorem a terminatorem (jednostka transkrypcji)
- Nić DNA jest odczytywana od końca 3' do 5'
- Nić RNA powstaje od końca 5' do 3'
- Substratami są trifosforany rybonukleotydów: ATP, CTP, UTP, GTP



**Figure 15-14**  
*Molecular Biology: Principles and Practice*  
 © 2012 W. H. Freeman and Company

- Mówiąc o sekwencji DNA genu myślimy o nici kodującej, bo ona jest „niemal” identyczna co RNA które będzie stanowiło podstawę do syntezy białka.
- Polimeraza musi rozpoznać miejsce rozpoczęcia transkrypcji i jej zakończenia (promotor i terminator)
- Polimeraza RNA nie sprawdza i nie koryguje produktu syntezy





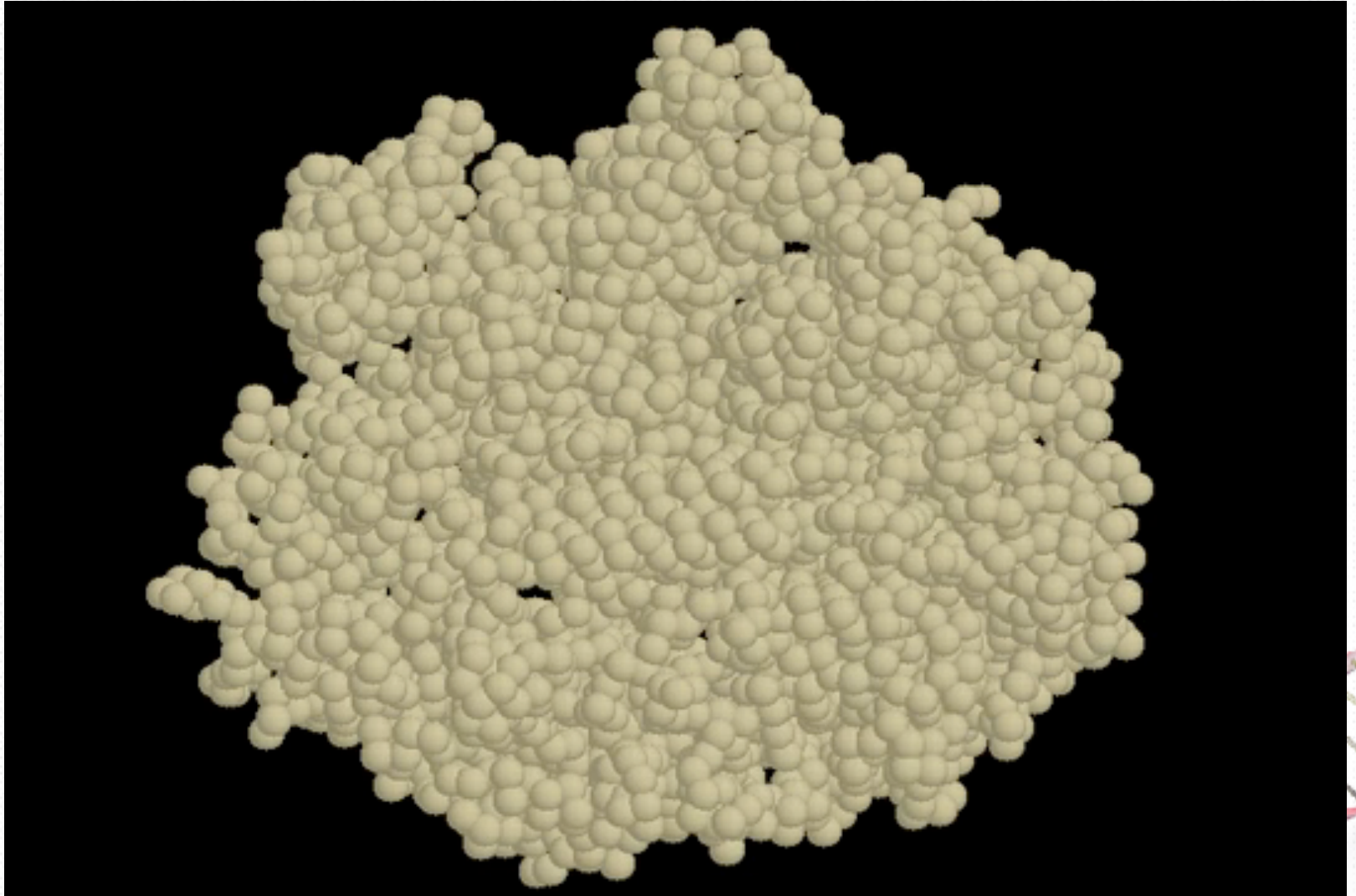
# Rodzaje RNA

Wśród kwasów rybonukleinowych wyróżnia się między innymi:

- informacyjny lub matrycowy RNA (mRNA)
- rybosomalny RNA (rRNA)
- transferowy RNA (tRNA)
- heterogeny jądrowy RNA (hnRNA lub pre-mRNA) - głównie produkty transkrypcji DNA i przetwarzania surowego transkryptu do mRNA
- antysensowy RNA albo interferencyjny RNA (siRNA i miRNA) - produkowany w celu precyzyjnej regulacji ekspresji genów kodujących białka (za pomocą mechanizmu wspólnego lub bardzo zbliżonego do systemu zwalczania wirusów RNA)
- mały cytoplazmatyczny RNA (scRNA) - odpowiedzialny za rozpoznawanie sygnału w komórce
- mały jądrowy RNA (snRNA) - pełniący funkcje enzymatyczne przy wycinaniu intronów z transkryptów
- mały jąderkowy RNA (snoRNA) - biorący udział w modyfikacji chemicznej pre-mRNA



# Polimeraza RNA II





# Co poza polimerazą RNA II

Gen odpowiada na wiele różnych sygnałów – mamy 4 podstawowe różnice w regulacji ekspresji genów u eukariotów i bakterii

- Bakterie mają jeden typ polimerazy. U eukariotów mamy trzy: RNA I, II i III. RNA I i II dokonują transkrypcji genów kodujących rybosomowe, transportujące i małe RNA (funkcje strukturalne); RNA II prowadzi transkrypcję ogromnej większości genów w tym kodujące białka i snRNA
- U bakterii polimeraza RNA może inicjować transkrypcję z u eukariotów nie i potrzebuje do inicjacji szeregu białek (czynniki transkrypcyjne)
- Ekspresją genów mogą regulować sekwencje DNA oddalone o setki tysięcy par zasad, u bakterii geny są regulowane przez sekwencje regulatorowe leżące w pobliżu promotora.



# Inicjacja transkrypcji genów eukariotycznych

Eukariotyczna polimeraza potrzebuje ogólnych czynników transkrypcyjnych. One:

- sytywiają polimerazę we właściwym miejscu,
- wspomagają rozdzielanie dwóch nici,
- rozpoczęcie transkrypcji poprzez uwolnienie polimerazy

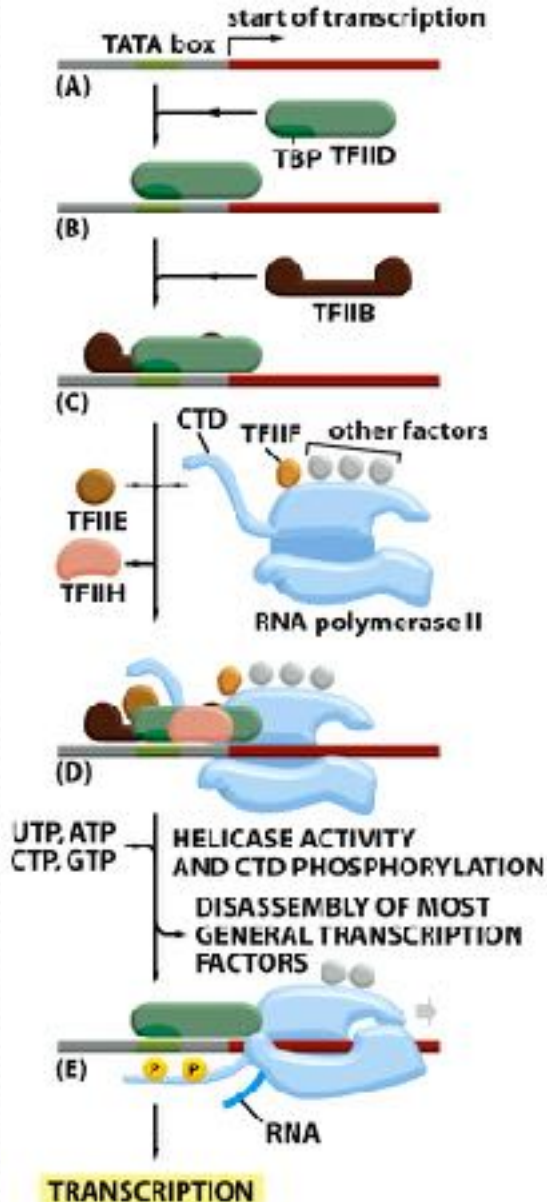
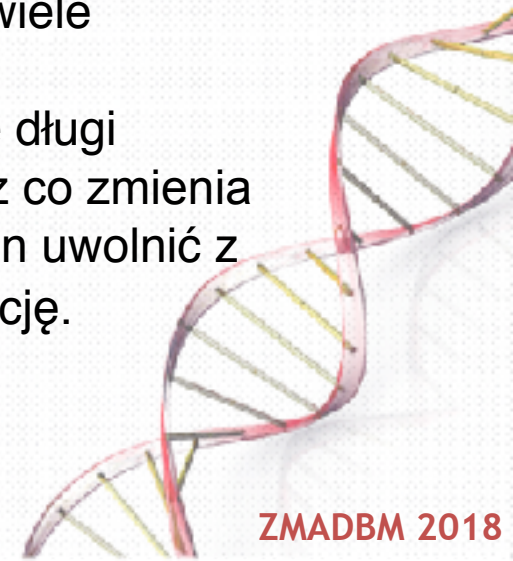


Figure 6-16 Molecular Biology of the Cell 5/e (© Garland Science 2008)



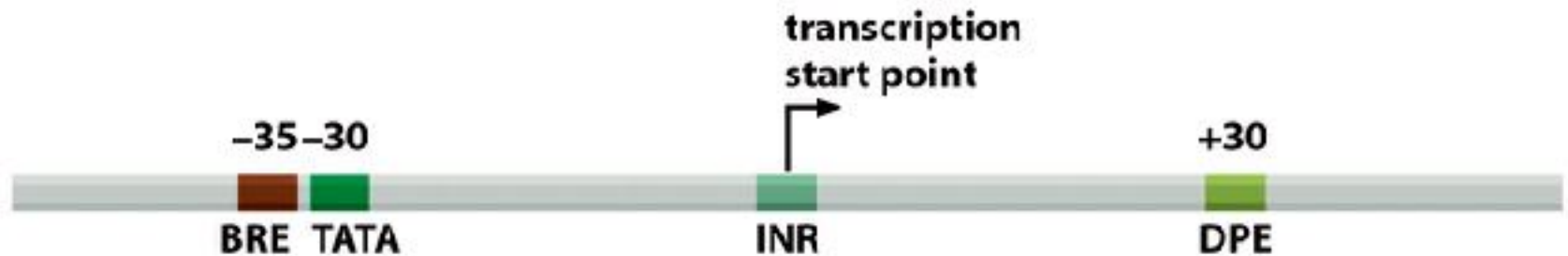




- Średniej długości gen 1500pz - transkrypcja ~ 50 sekund
- ~ 15 polimeraz może pracować jednocześnie



# Inicjacja transkrypcji



| element | consensus sequence      | general transcription factor |
|---------|-------------------------|------------------------------|
| BRE     | G/C G/C G/A C G C C     | TFIIB                        |
| TATA    | T A T A A/T A A/T       | TBP                          |
| INR     | C/T C/T A N T/A C/T C/T | TFIID                        |
| DPE     | A/G G A/T C G T G       | TFIID                        |

Figure 6-17 Molecular Biology of the Cell 5/e (© Garland Science 2008)



# TATA-binding protein - struktura 3D

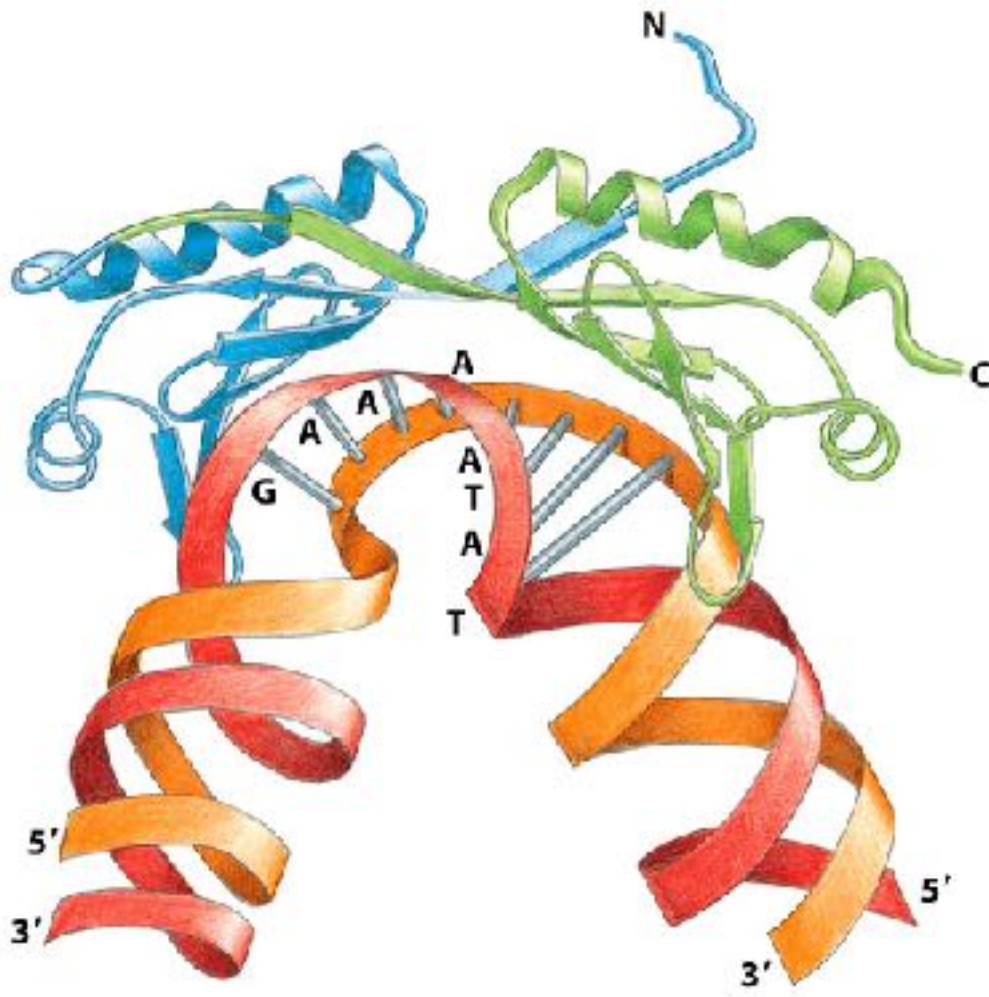


Figure 6-18 Molecular Biology of the Cell 5/e (© Garland Science 2008)



# Enhancer - sekwencja wzmacniająca

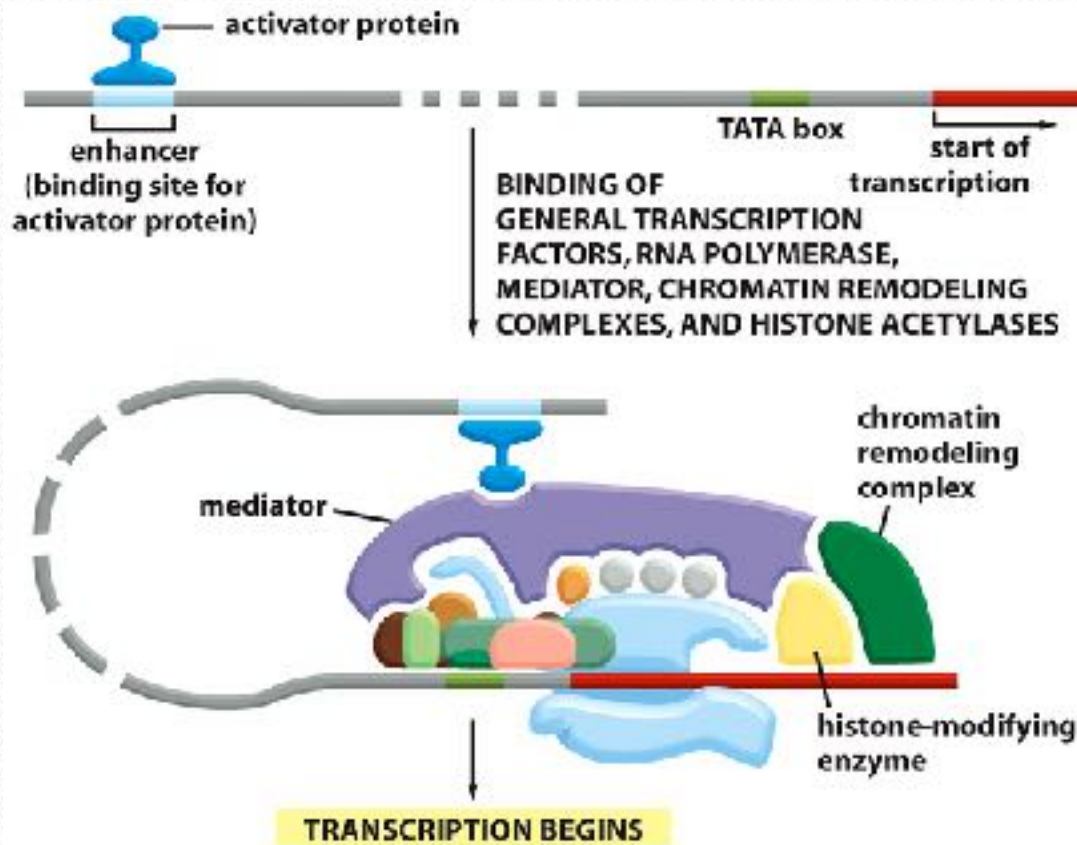


Figure 5-19 Molecular Biology of the Cell 5/e (© Garland Science 2008)

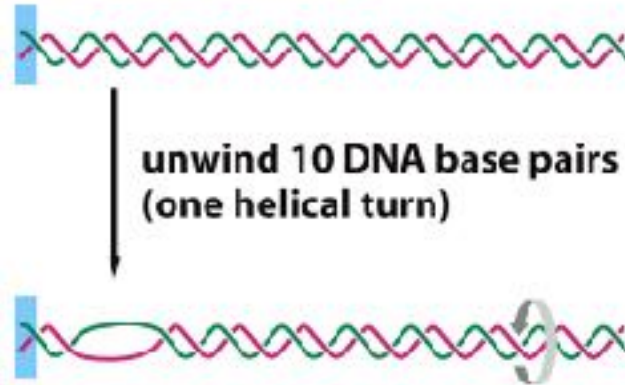
- Badania in vitro pokazały jak funkcjonuje model transkrypcji na oczyszczonym DNA. Jednak chromatyna to skomplikowany kompleks białek i DNA i rzeczywistości potrzebne są jeszcze aktywatory transkrypcji. Zwiększają one lub zmniejszają aktywność polimerazy RNA II.

- Mediator służy do właściwego oddziaływania pomiędzy aktywatorem a polimerazą i czynnikami transkrypcyjnymi
- Stan upakowania chromatyny ma istotny wpływ na inicjację transkrypcji. Nukleosomy mogą hamować jej inicjację.



# DNA supercoiling

## DNA with free end

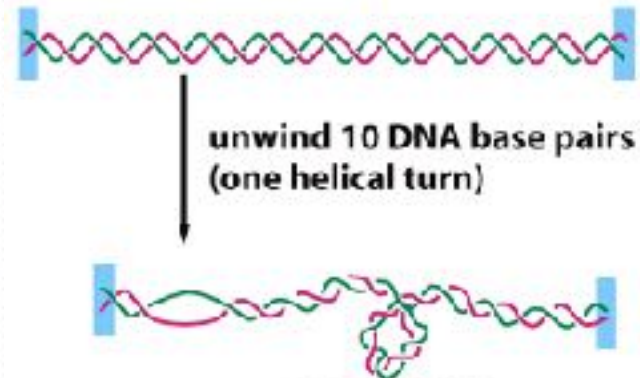


unwind 10 DNA base pairs  
(one helical turn)

DNA helix must  
rotate one turn

Figure 6-20a Molecular Biology of the Cell 5/e (© Garland Science 2008)

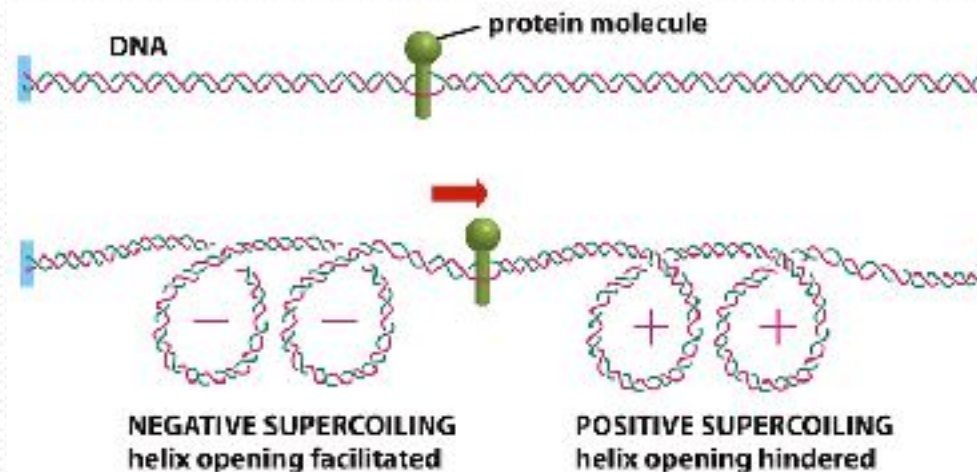
## DNA with fixed ends



unwind 10 DNA base pairs  
(one helical turn)

DNA helix forms  
one supercoil

Figure 6-20b Molecular Biology of the Cell 5/e (© Garland Science 2008)



NEGATIVE SUPERCOILING  
helix opening facilitated

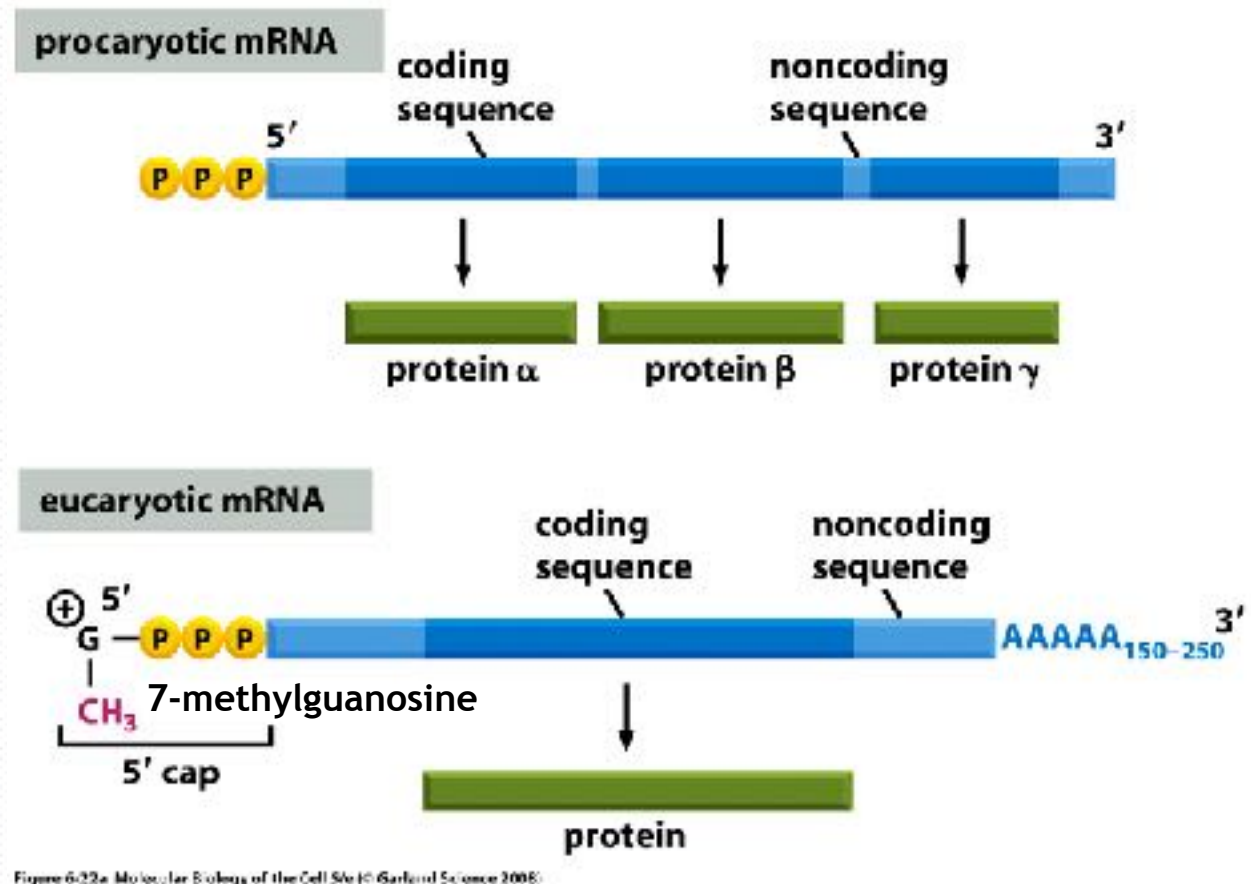
POSITIVE SUPERCOILING  
helix opening hindered

Figure 6-22 Molecular Biology of the Cell 5/e (© Garland Science 2008)



# Tempo transkrypcji - etap elongacji

- Zróżnicowane tempo, od powolnego po szalone
- Elongation factors - zmniejszają prawdopodobieństwo od dysocjowania polimerazy od DNA. Niektóre z nich powiązane są z kompleksami białek histonowych.
- Dodanie właściwych zakończeń – na końcu 5' już po syntezie ~25pz





# Organizacja dwóch genów u człowieka

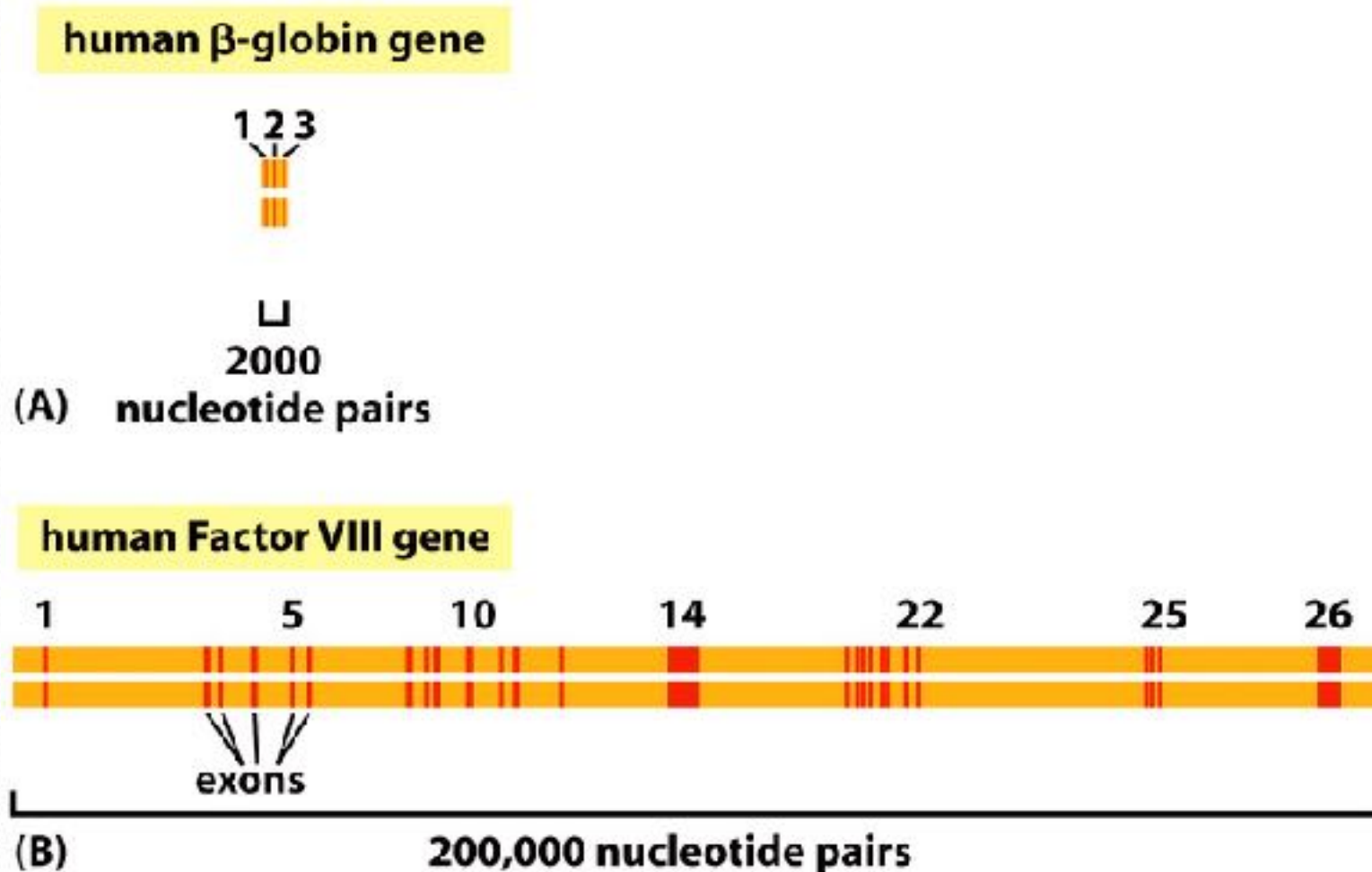


Figure 5-25 Molecular Biology of the Cell 5/e (© Garland Science 2008)

# Podumowanie

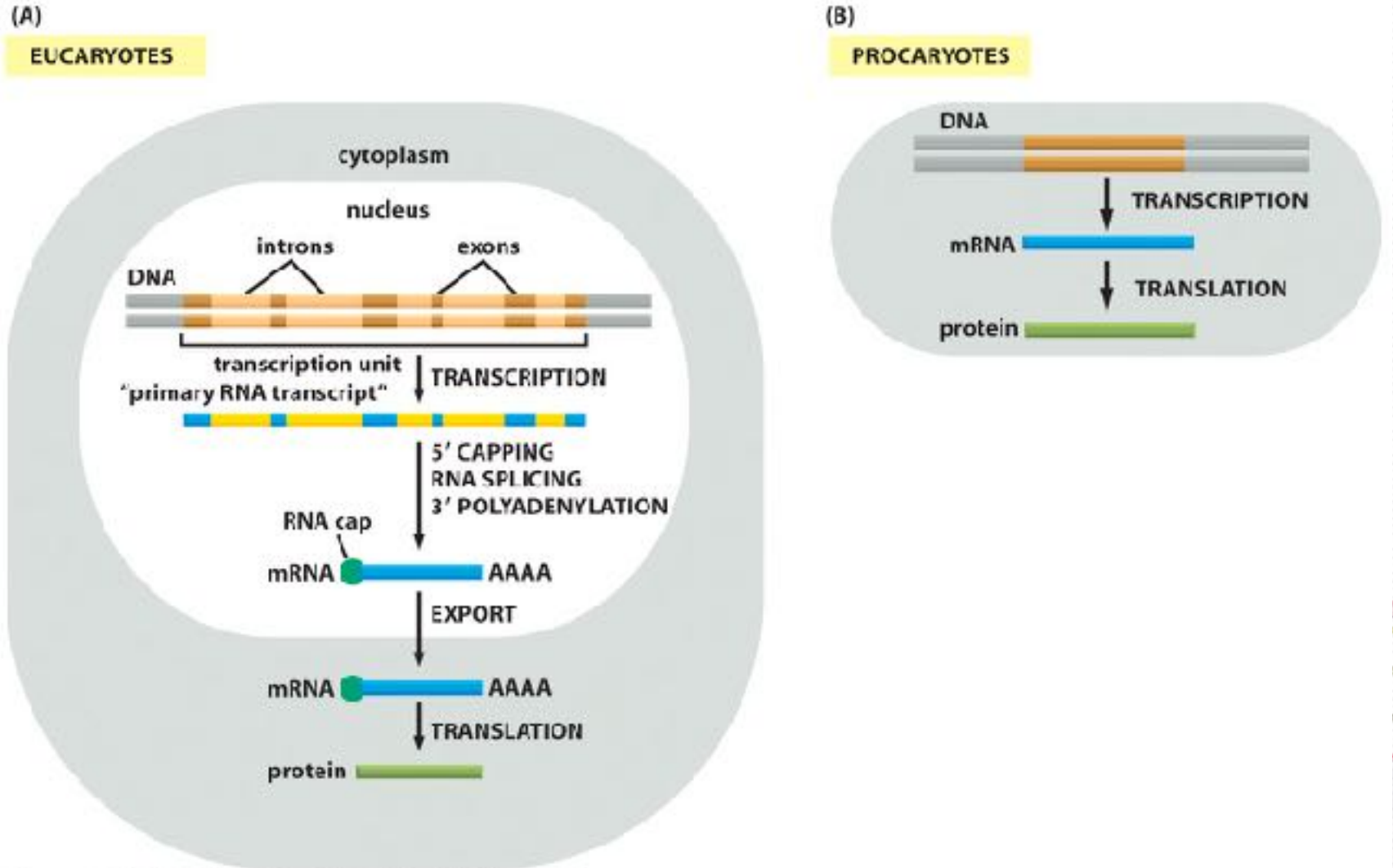


Figure 6-21 Molecular Biology of the Cell 5/e (© Garland Science 2008)



# Regulacja ekspresji genów

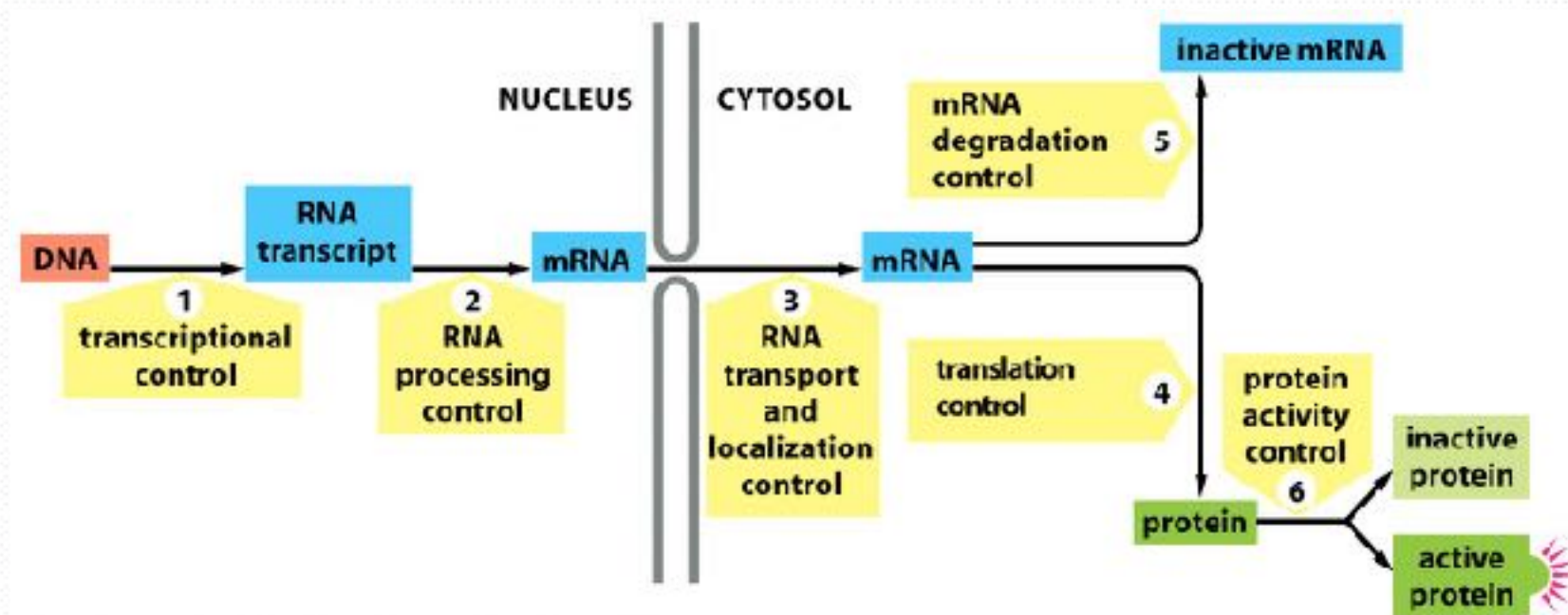


Figure 7-5 Molecular Biology of the Cell 5/e (© Garland Science 2008)

# Ćwiczenia

```
tcga <- readRDS(gzcon(url("http://zbo.ipipan.waw.pl/files/data/TCGA_KM_clinical_data.rds")))
cec <- readRDS(gzcon(url("http://zbo.ipipan.waw.pl/files/data/OrigCecRNA_en.rds")))
```

Podziel plik (OrigCecRNA\_en.rds) na dane tylko z ekspresją genów i na dane pacjentów. Zapisz jako oddzielne .rds

I. Analizujemy dane pacjentów:

- Zweryfikuj czy lista pacjentów jest unikatowa
- Zapisz \$patient dla tych którzy się powtarzają. Wyjaśnij na podstawie \$subtype\_patient i ewentualnie innych cech pacjenta, dlaczego się oni powtarzali w bazie.
- Stwórz dane (df) z unikatowymi ID pacjentów.
- Wybierz tylko tych pacjentów, dla których są dane o RNAseq.
- Będziemy analizowali śmiertelność pacjentów. Usuń te cechy kliniczne z df, które Twoim zdaniem są nieprzydatne z analityczno-matematycznego-logicznego punktu widzenia. Nazwy tych zmiennych zapisz.

II. Zaproponuj na podstawie zmiennej \$subtype\_Survival..months. podział na:

- (A)śmiertelność wczesną, średnią i późną;
- (B)śmiertelność wczesną i późną.

Czy w ustalaniu typów śmiertelności inna zmienna nie powinna być również brana pod uwagę? Zaproponuj jaką zmienną.

III. Sprawdź czy śmiertelność jest istotnie różna u pacjentów przy uwzględnieniu ich:

"subtype\_IDH.status", subtype\_IDH.codel.subtype, subtype\_TERT.promoter.status oraz innej cechy jaką sam wybierzesz – musi ona zawierać co najmniej trzy klasy.

Wyniki zobrazuj za pomocą wykresu/sów oraz wykonaj właściwe testy statystyczne.

IV. Porównaj na podstawie wybranych 4-5 zmiennych czy do dalszych analiz lepiej brać dane pacjentów z OrigCecRNA\_en.rds, TCGA\_KM\_clinical\_data.rds. Odpowiedź uzasadnij.