

wykład 10

Transkryptomika

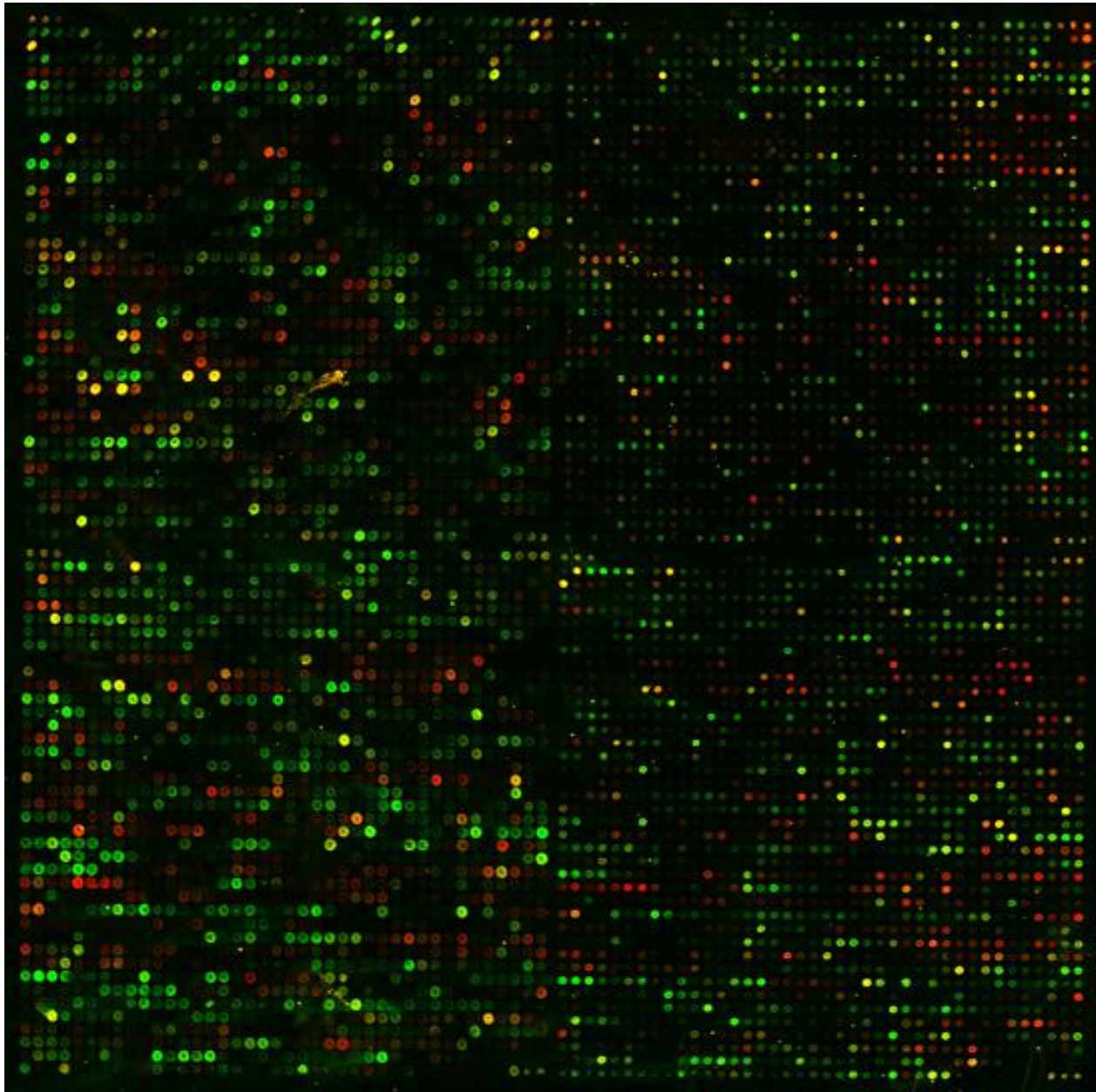
Mikromacierze

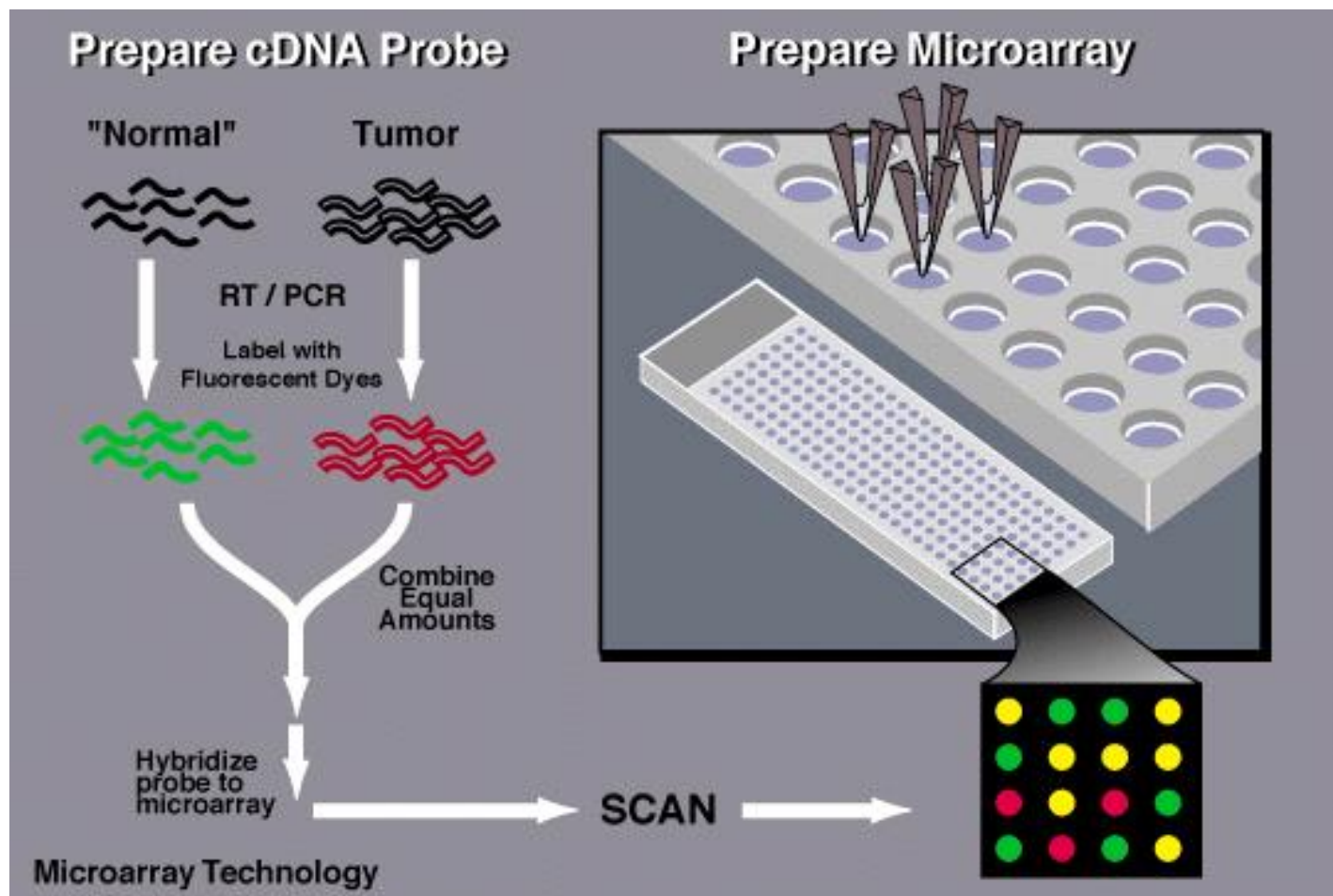
dr Jacek Śmietanski

jacek.smietanski@ii.uj.edu.pl

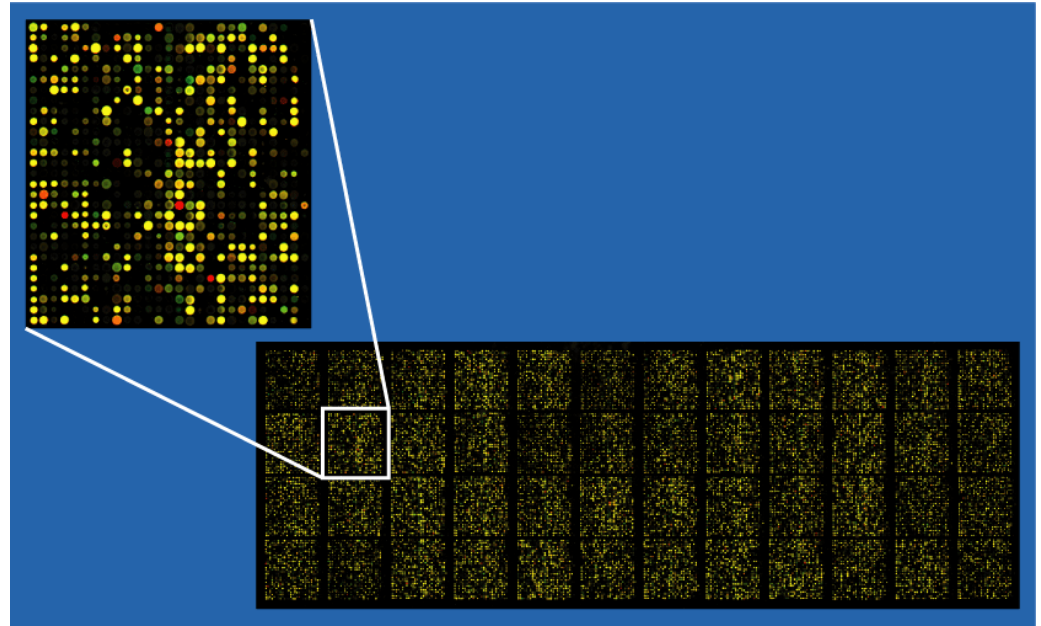
<http://jaceksmietanski.net>

Genom drożdży na mikromacierzy





Pozwalają na przeprowadzanie analiz genetycznych dla wielu (nawet tysięcy) genów jednocześnie.
Badanie ekspresji genów.



Określenie zmian w poziomie ekspresji genów. W badaniu określa się obecność i ilość cząsteczek mRNA dla poszczególnych genów w danym momencie funkcjonowania komórki.

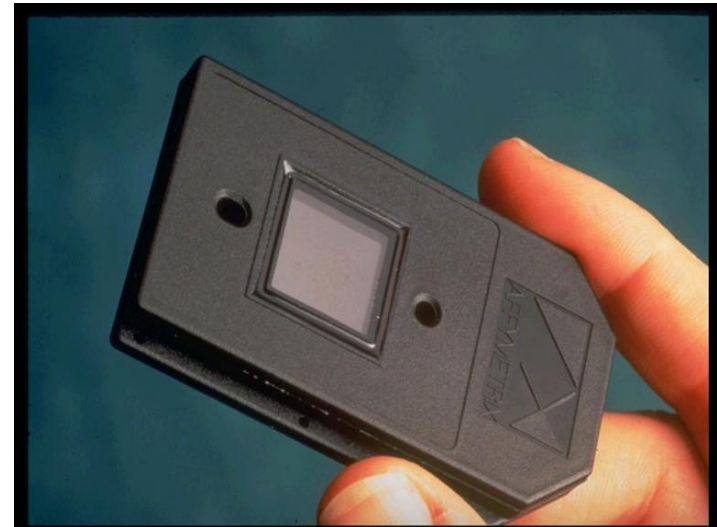
Np. jedna próbka pochodzi z komórki zdrowej, druga – z chorej.
Albo z komórek funkcjonujących w różnych warunkach środowiska (np. bakteria w warunkach tlenowych i beztlenowych).
Albo z komórek w różnych stadiach rozwoju (np. w różnych etapach mitozy).

Zwykle produkowana komercyjnie.

Każda płytka posiada tysiące ściśle określonych punktów, każdy zawiera nić pochodzącą z innego genu.

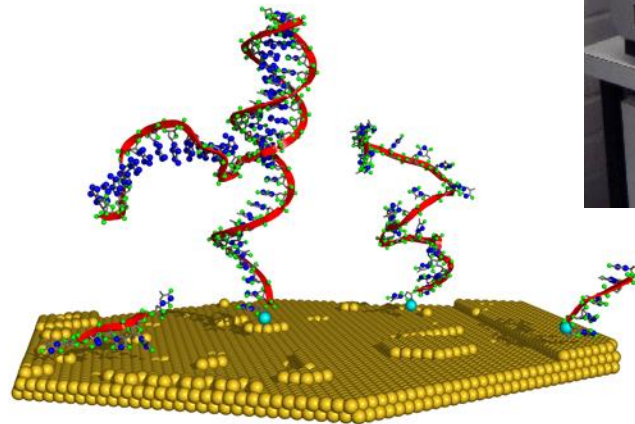
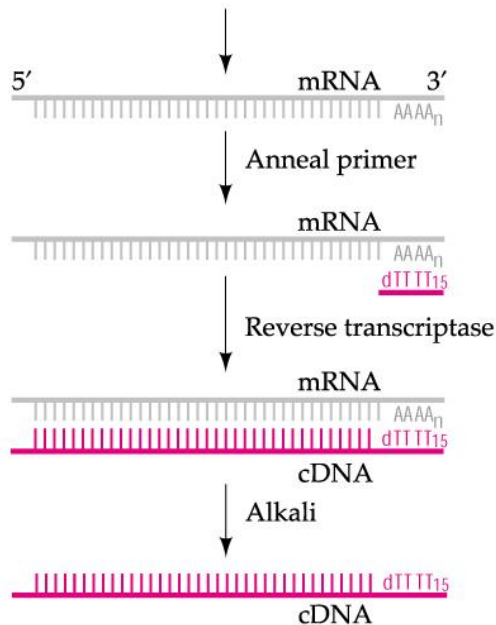
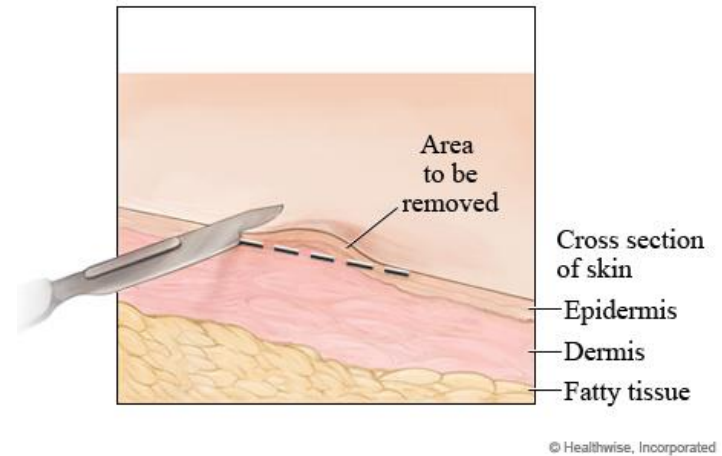
Nicią jest fragment nici cDNA lub syntetyczny oligonukleotyd.

Nici są dodawane automatycznie za pomocą igły aplikującej cDNA lub metodą podobną do tworzenia procesorów – z tego względu mikromacierze nazywane są też czipami genowymi (*Gene Chip*)

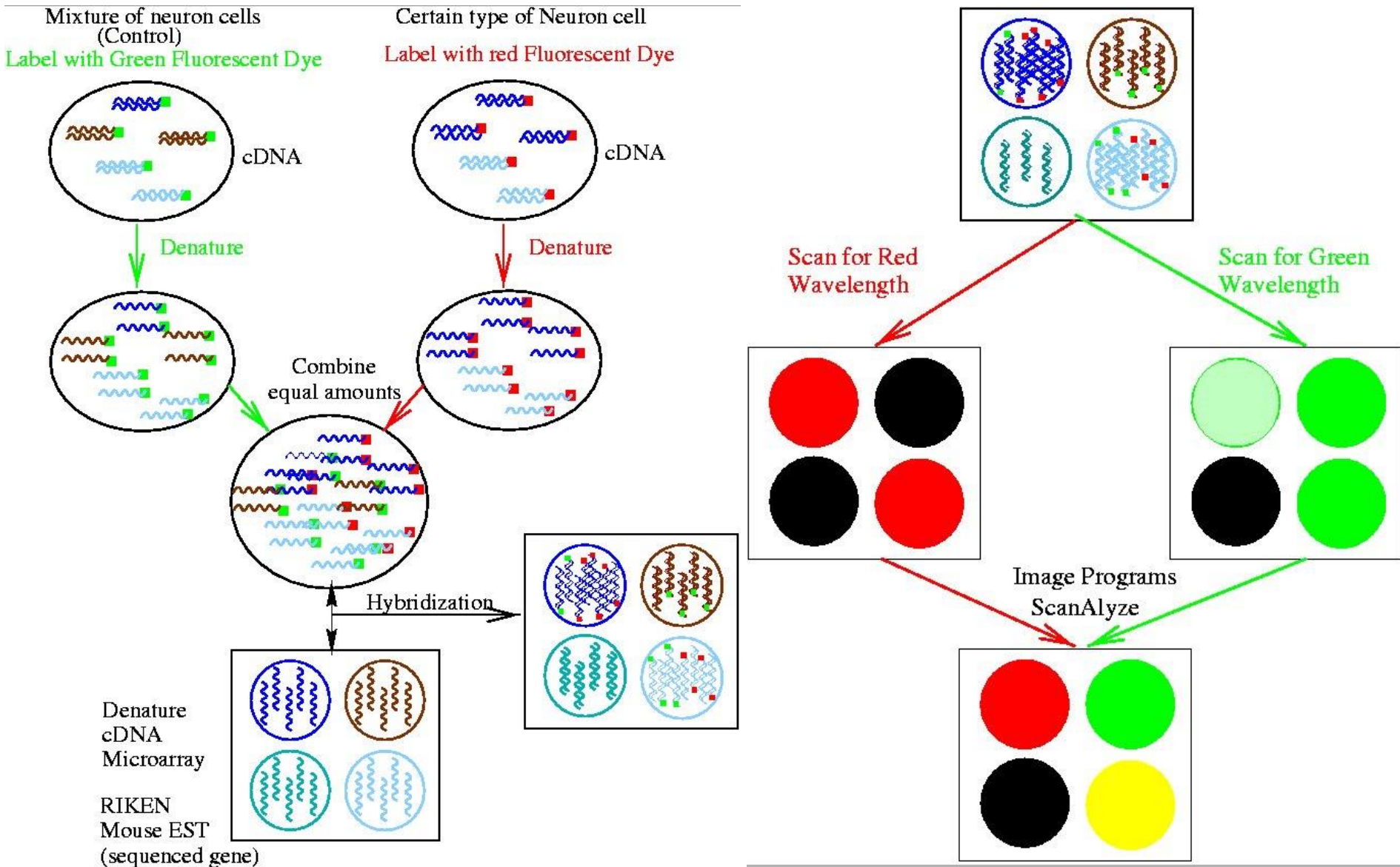


Przebieg eksperymentu

1. Pobranie próbek
2. Izolacja mRNA.
3. Odwrotna transkrypcja do cDNA.
4. Hybrydyzacja na płytce
5. Skanowanie mikromacierzy
6. Analiza danych



Eksperyment



ZIELONY - geny z próbki kontrolnej, które hybrydyzowały bardziej niż w badanej.

CZERWONY - geny z próbki badanej, które hybrydyzowały bardziej niż w kontrolnej.

ŻÓŁTY - geny z obu próbek hybrydyzowały w podobnym stopniu

CZARNY - obszary, w których żadna próbka nie hybrydyzowała do danej sekwencji DNA

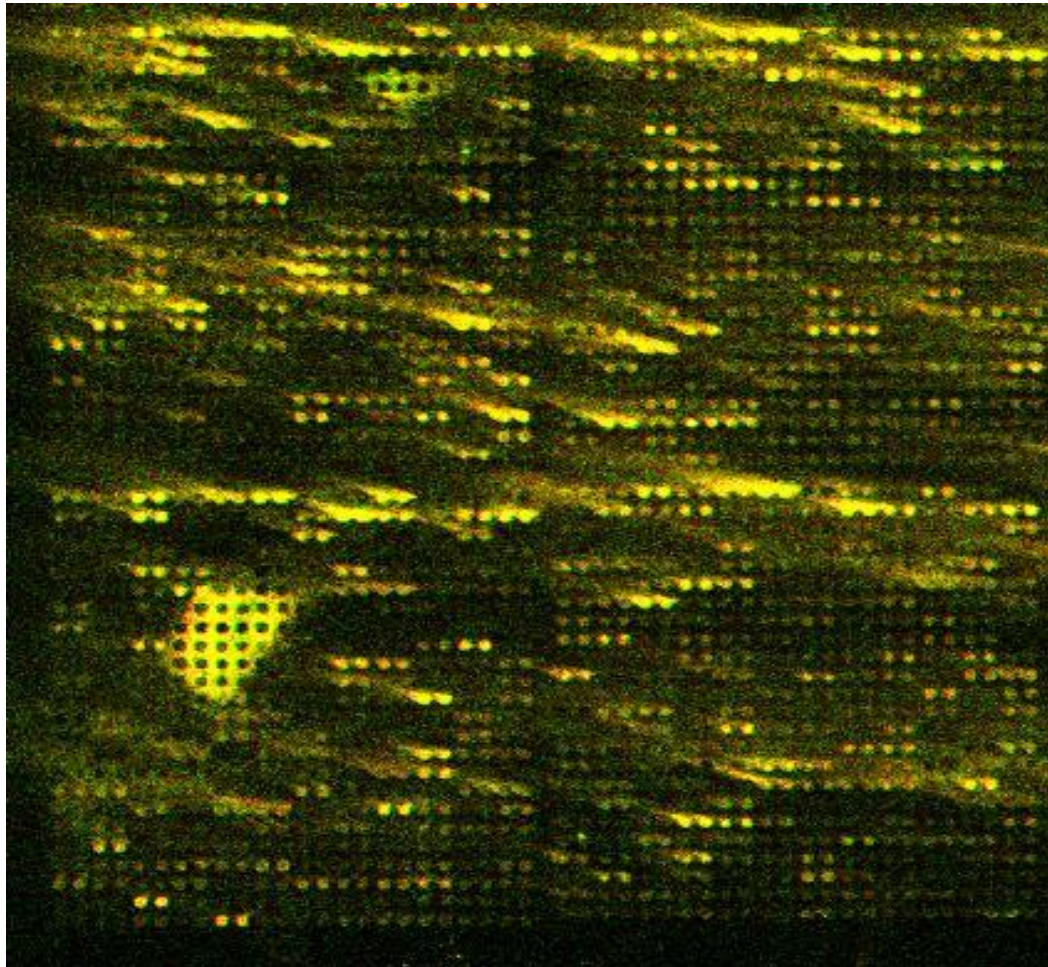
Analiza różnic w ekspresji genów pomiędzy dwiema próbkami pozwala nam zrozumieć rolę genów w poszczególnych stadiach życiowych czy w razie choroby komórki.



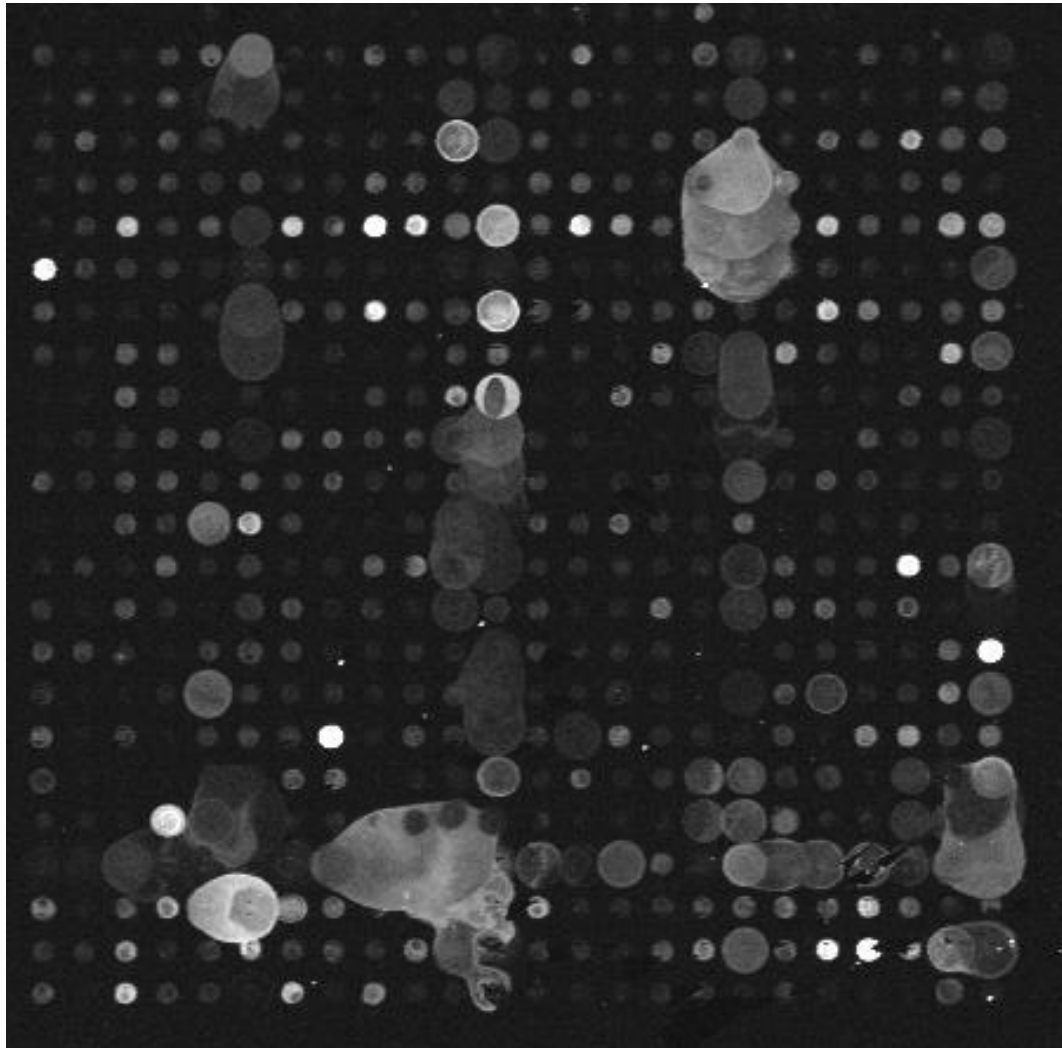
- 1) Identyfikacja punktów na obrazie odpowiadającym poszczególnym genom
- 2) Odczyt intensywności kolorów
- 3) Analizy statystyczne, itp.

Ogon komety (*comet tail*)

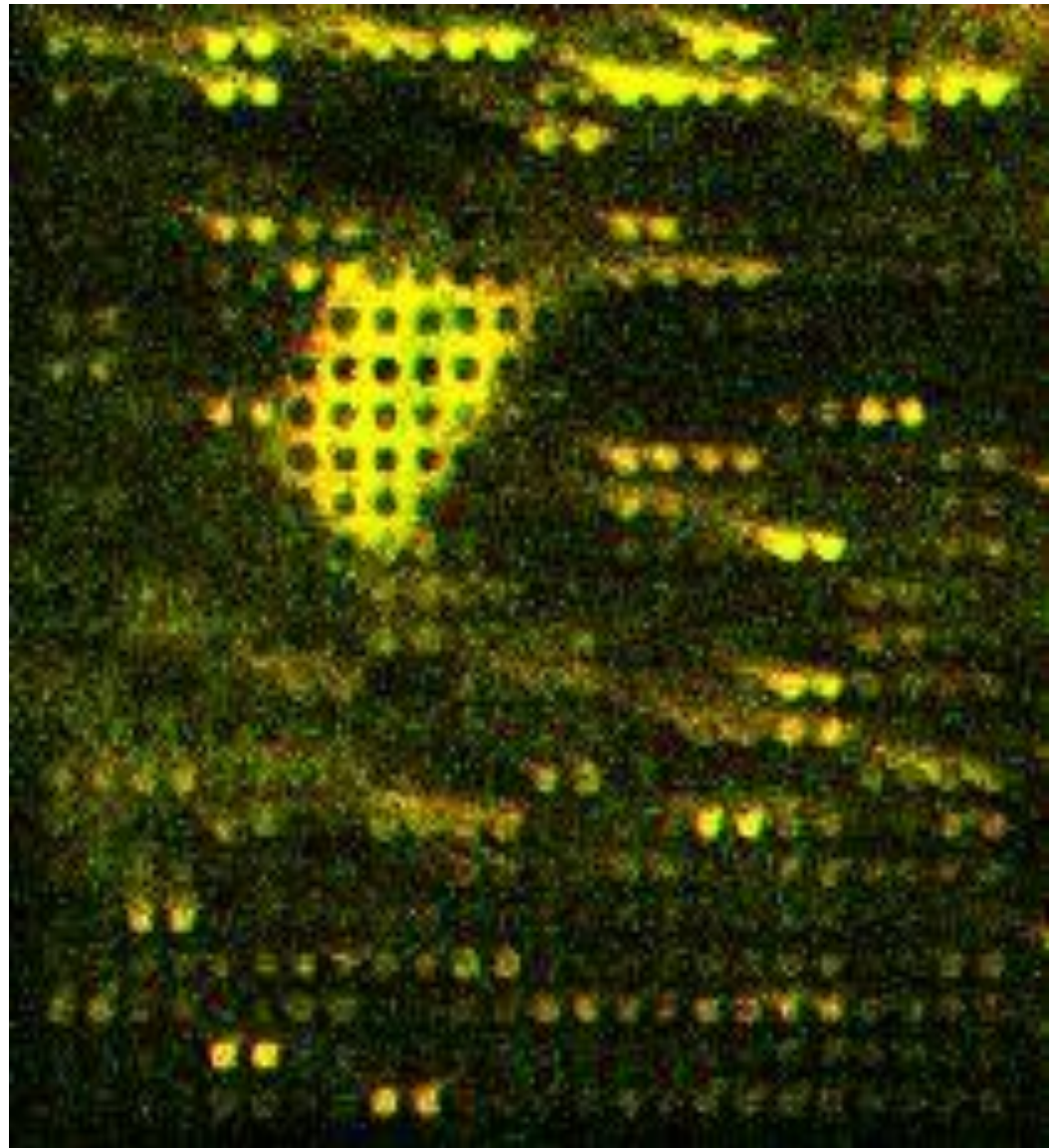
Najprawdopodobniej spowodowane niedostatecznie szybkim zanurzeniem próbki w roztworze.



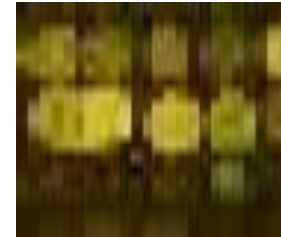
Plamy, nierówne rozmiary, nachodzące na siebie



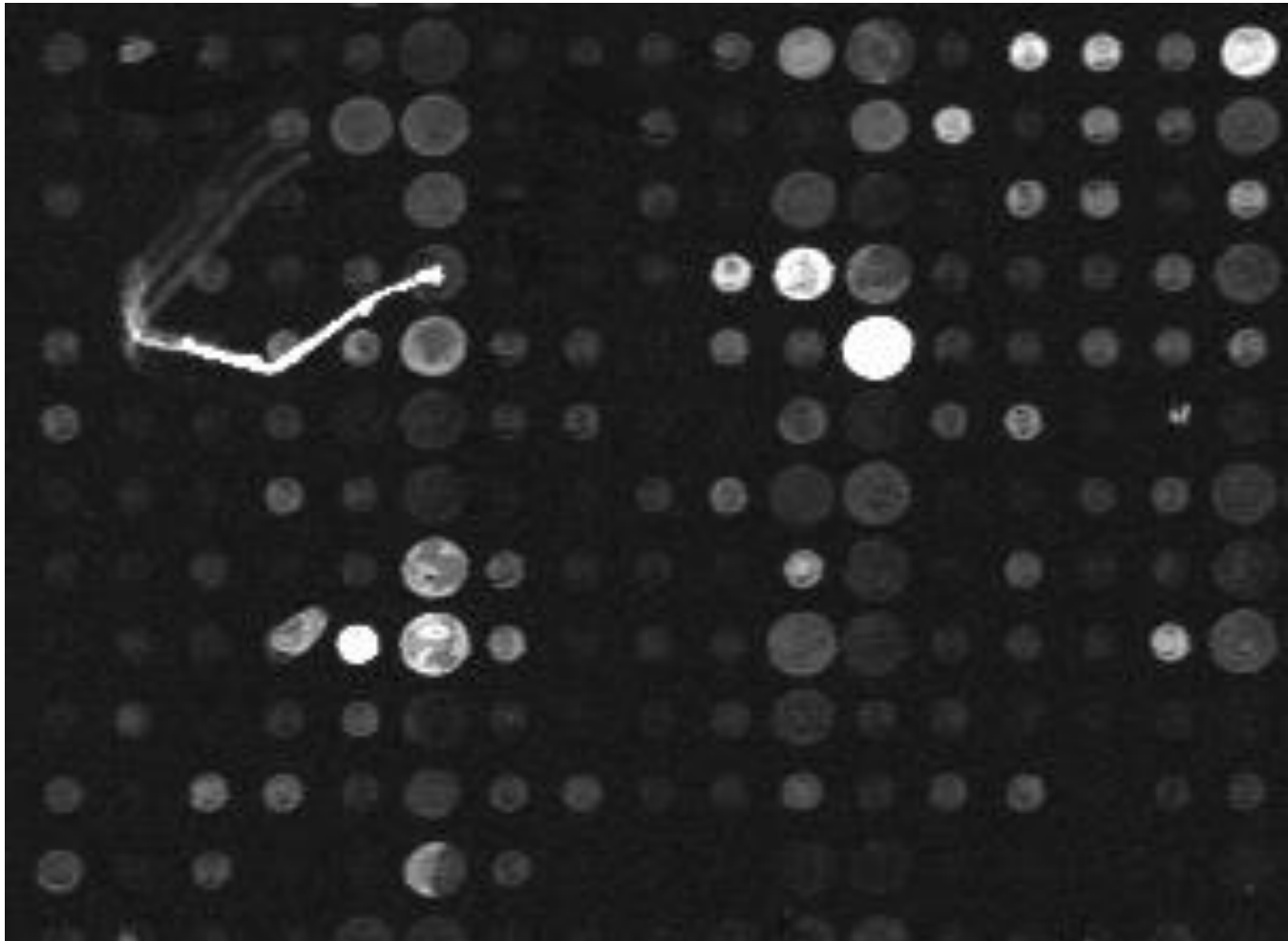
Jasne / nierówne tło.



Zachodzenie punktów.



Artefakty



Rozdzielczość: $10\ \mu\text{m}$

Standardowy rozmiar punktu: $100\ \mu\text{m}$

⇒ Średnica obiektu: 10 pixeli

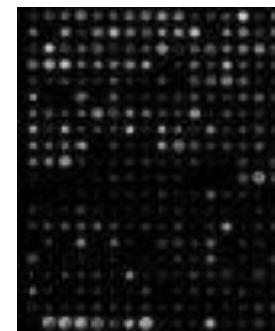
Format obrazu:

TIFF (tagged image file format) 16 bit (65,536 poziomów szarości)

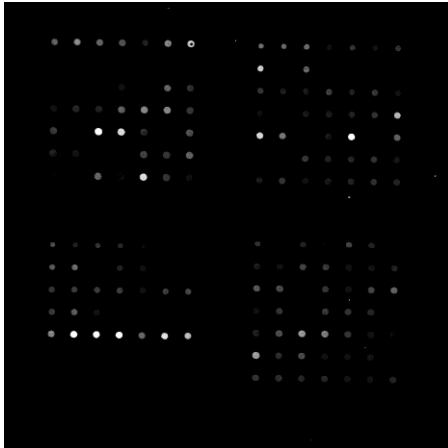
Obraz $1\text{cm} \times 1\text{cm}$ (16 bit) = 2Mb (bez kompresji)

Istnieją też inne formaty, np. SCN (Stanford)

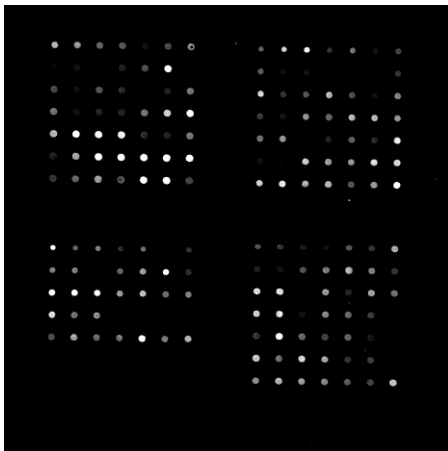
Oddzielne obrazy dla poszczególnych próbek.



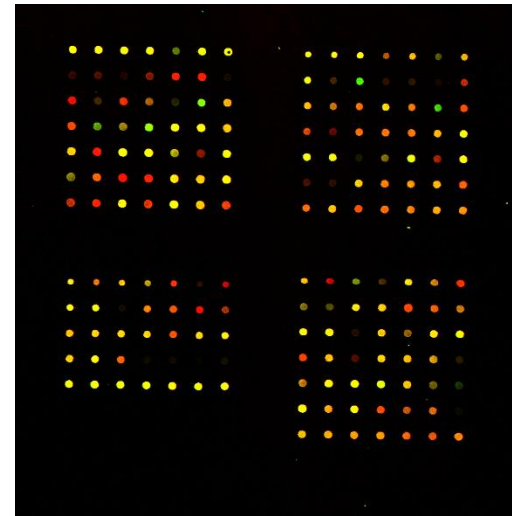
Cy3



Cy5



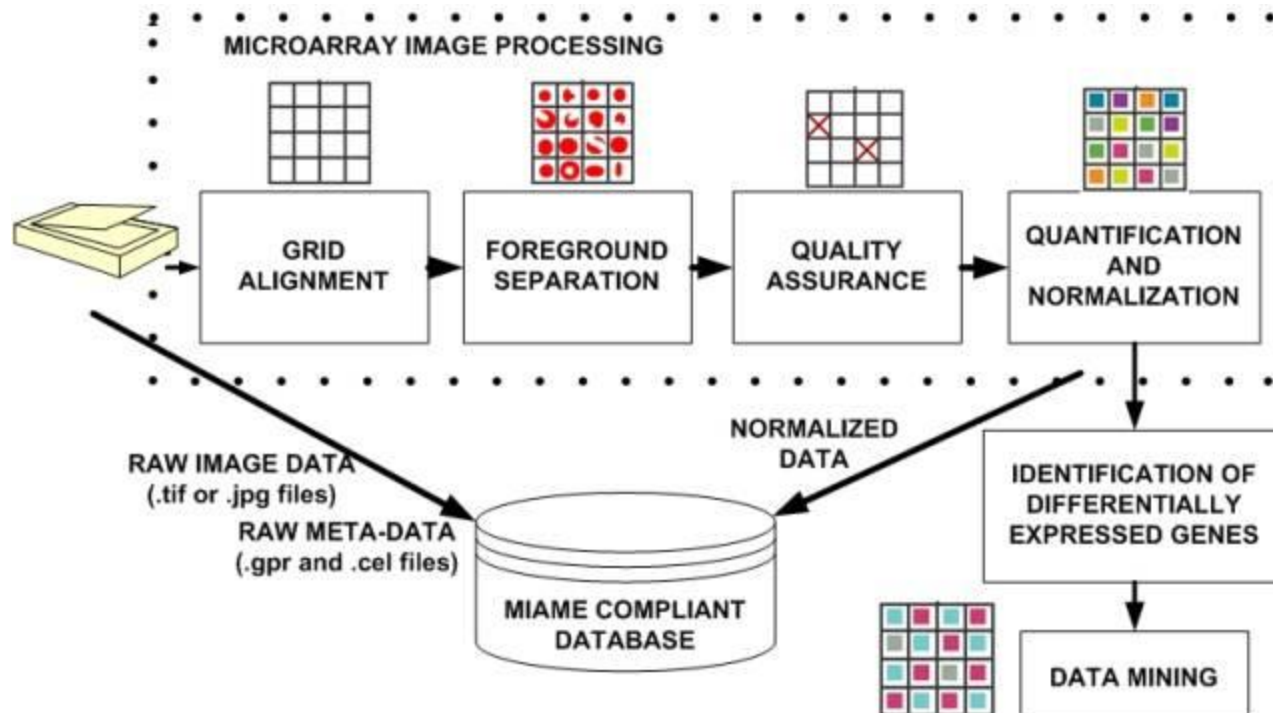
Pseudo-color overlay



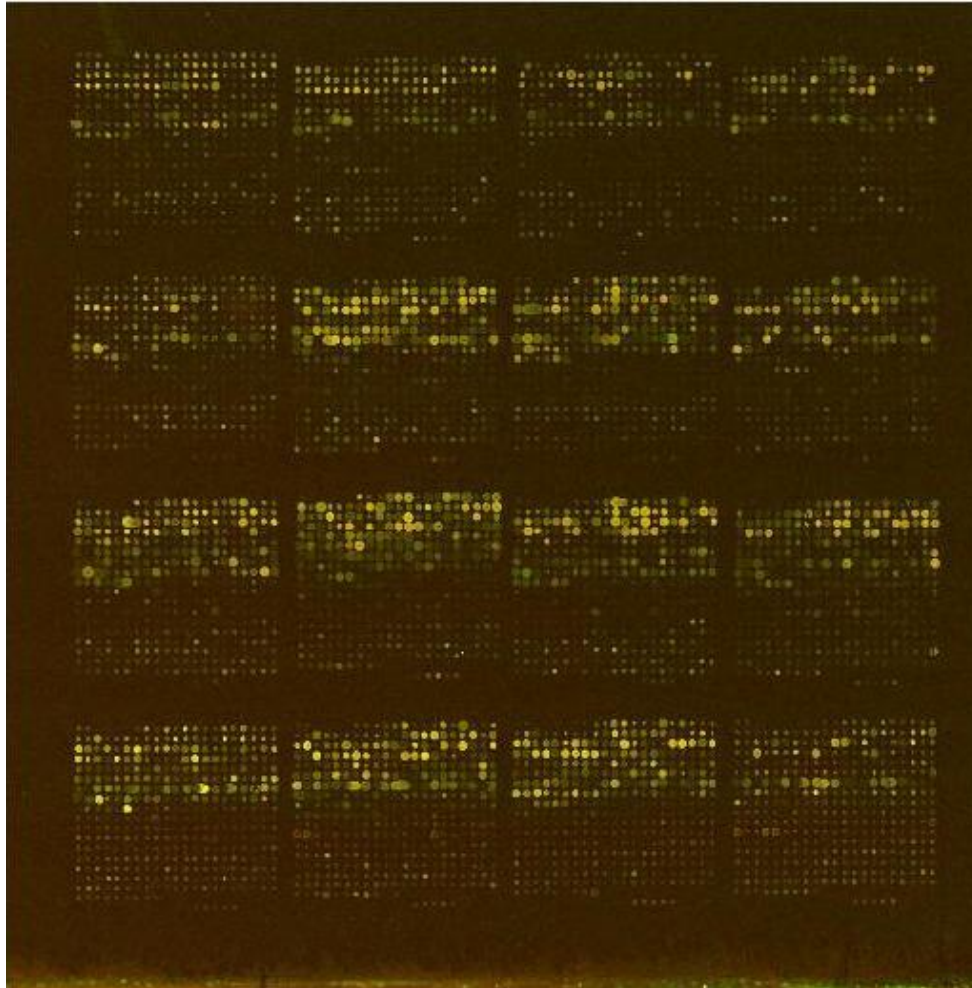
<i>Spot color</i>	<i>Signal strength</i>	<i>Gene expression</i>
yellow	Control = Treated	unchanged
red	Control < Treated	induced
green	Control > Treated	repressed

Etapy przetwarzania obrazów

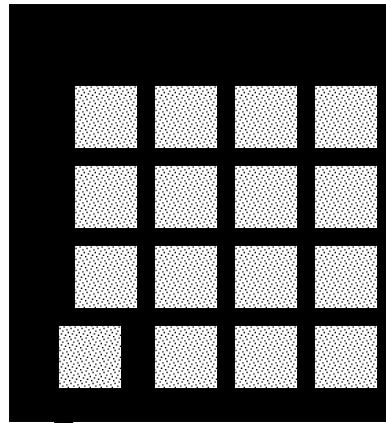
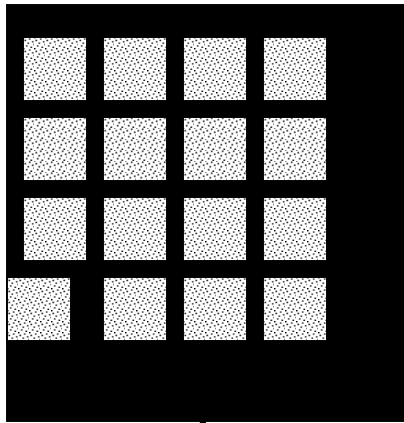
- Adresowanie (tworzenie siatki)
- Przypisanie współrzędnych do każdego punktu
- Segmentacja
- Oddzielenie sygnału od tła
- Analiza intensywności sygnału
- Analiza jakości pomiaru



This is the process of assigning coordinates to each of the spots.
Automating this part of the procedure permits high throughput analysis.



4 by 4 grids
19 by 21 spots per grid

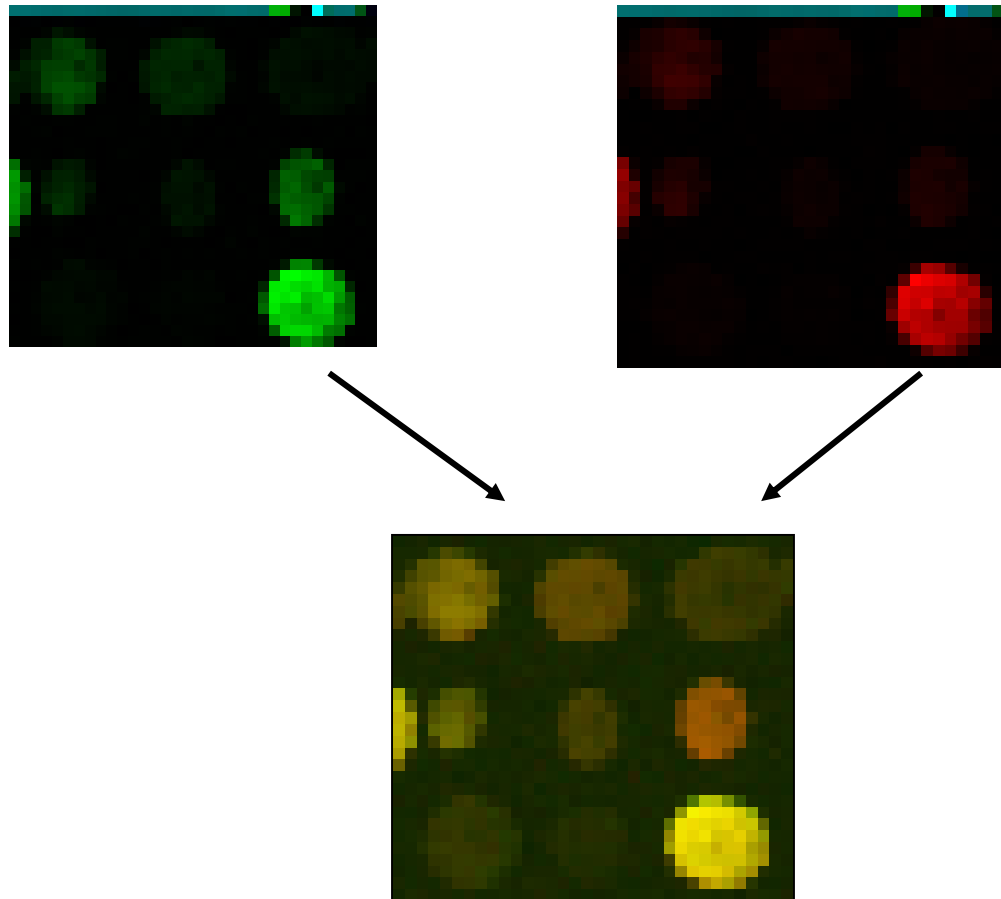


4 by 4 grids

Within the same batch of print runs; estimate translation of grids

Other problems:

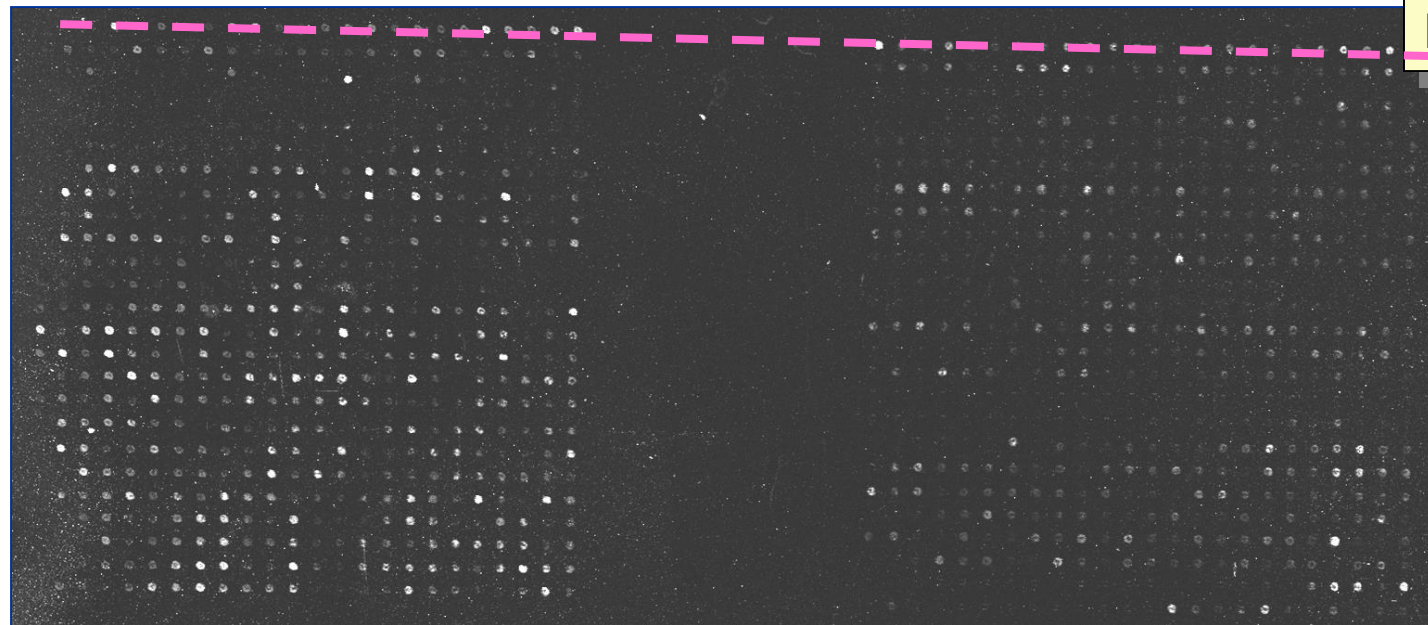
- Misregistration
- Rotation
- Skew in the array



Misregistration of the red and green channels

Rotation of the array in the image

Skew in the array



Rotation

- Analiza podstawowej struktury obrazu (determinowanej przez urządzenie)
- Pozycjonowanie macierzy na obrazie
- Rozdział na wiersze i kolumny w grupach (*grids*)
- Identyfikacja względnego przesunięcia grup
- Rozdział na wiersze i kolumny wewnątrz każdej grupy
- Korekty (przesunięcia) pojedynczych punktów

New Grid

Grids To Add

☐ 1
☒ 4
☐ 16
☐ 32

Columns

Rows

Col Spacing (um)

Row Spacing (um)

XRes (um)

YRes (um)

Tip Spacing (um)

Spot Width (pixels)

Spot Height (pixels)

First Grid Position

Left

Top

Create Cancel

ScanAlyze

Klasyfikacja pikseli do tła lub sygnału.

Metody:

- *Fixed circles*
- *Adaptive circles*
- *Adaptive shape*
- *Edge detection*
- *Seeded Region Growing* (R. Adams and L. Bishof (1994): rozrost obszarów począwszy od „punktów zasiewu” stosownie do różnicy pomiędzy jasnościami pikseli i średniej jasności sąsiadujących obszarów
- Metody bazujące na histogramie

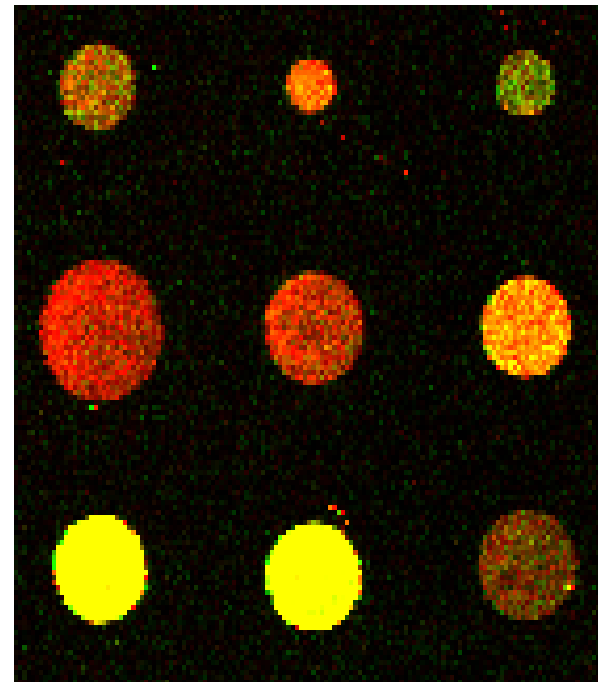
Fixed circle	ScanAlyze, GenePix, QuantArray
Adaptive circle	GenePix, Dapple, SignalViewer (uses ellipse)
Adaptive shape	Spot, region growing and watershed
Histogram	ImaGene, QuantArray, DeArray and adaptive thresholding

Dopasowuje okręgi o zadanej na sztywno średnicy do wszystkich obszarów na obrazie.

Łatwa w implementacji.

Obszary powinny mieć ten sam kształt i wielkość.

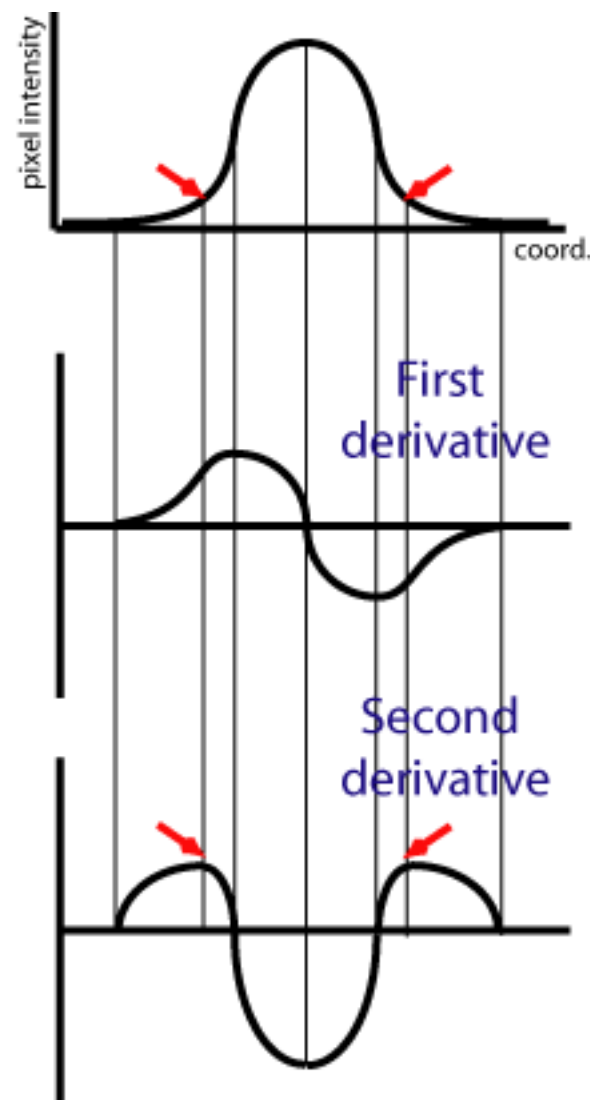
*May not be good
for this example*

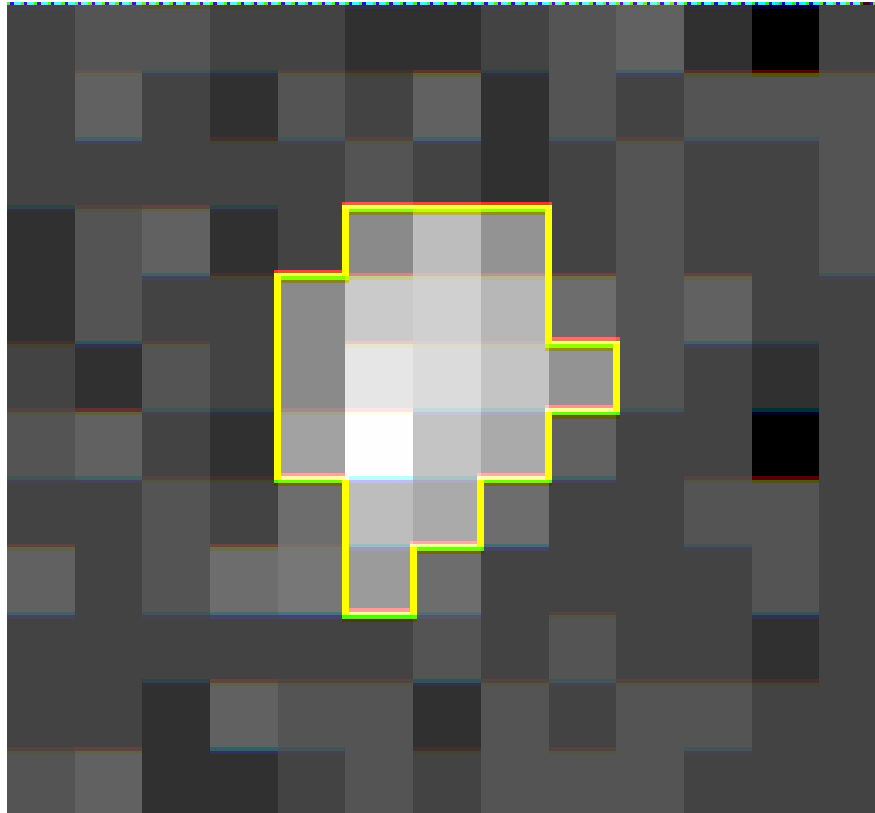


Średnica okręgu jest szacowana indywidualnie dla każdego obszaru.

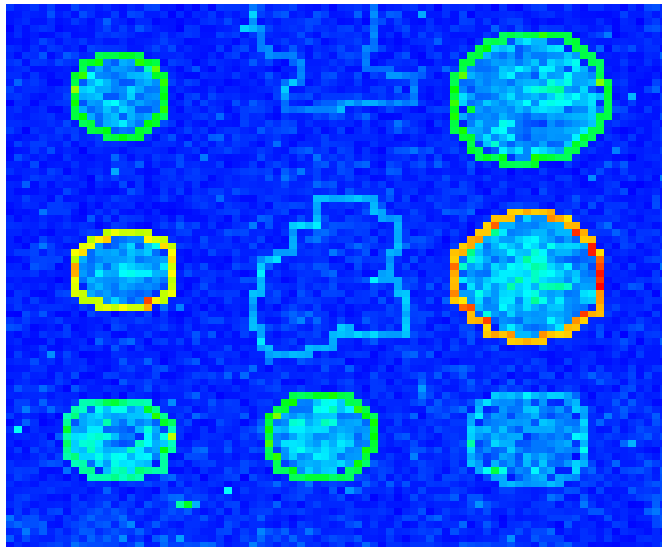
Wykorzystywane są tu metody gradientowe wykrywania krawędzi (*second derivative*)

Problematyczna dla nieokrągłych kształtów, np. owalnych.

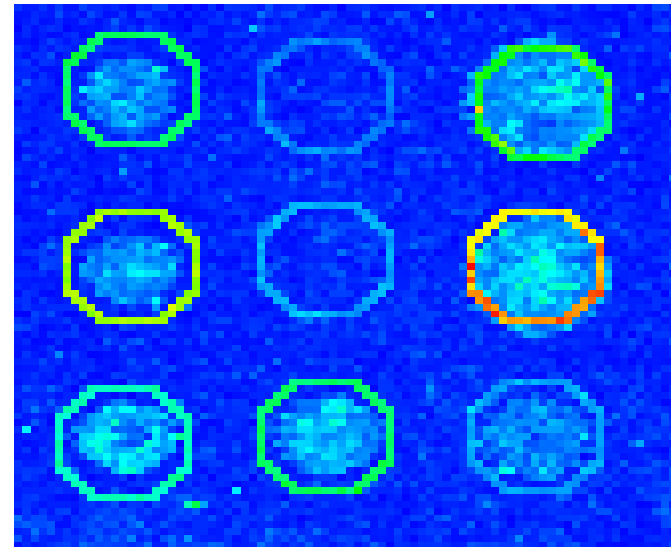




- Zbyt mały obszar
- Kształt inny niż okrąg

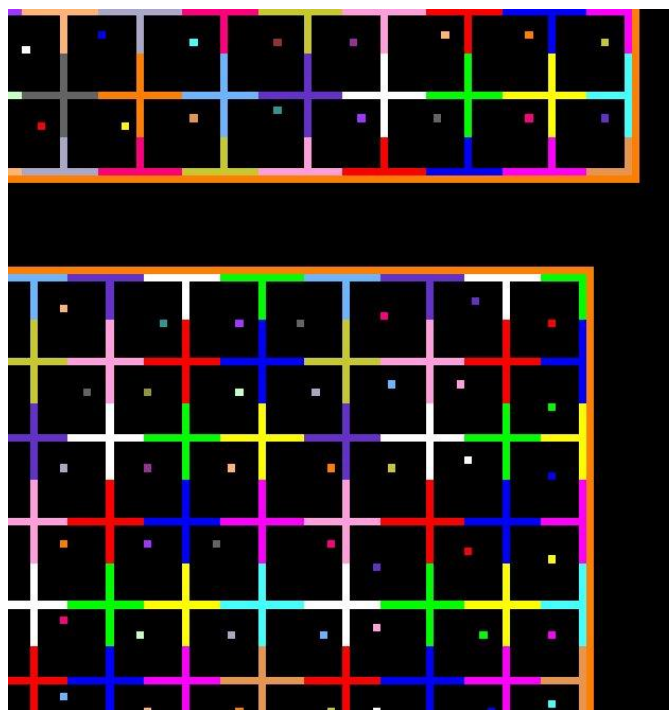
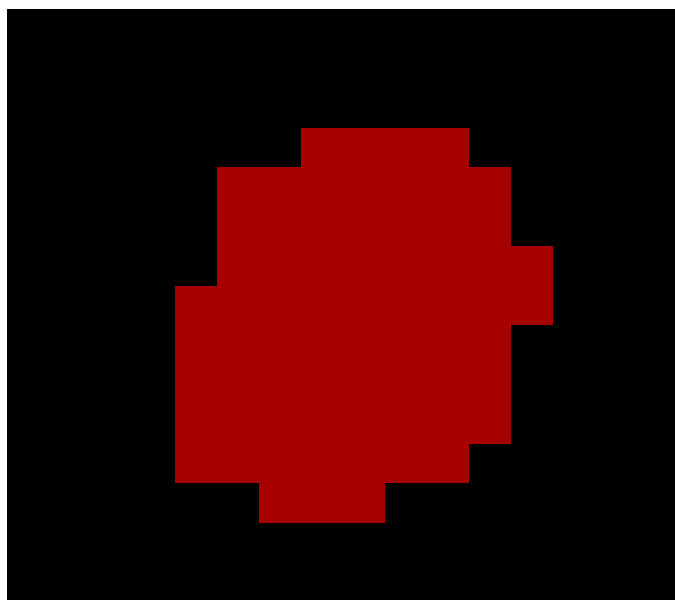


SRG



Fixed circle

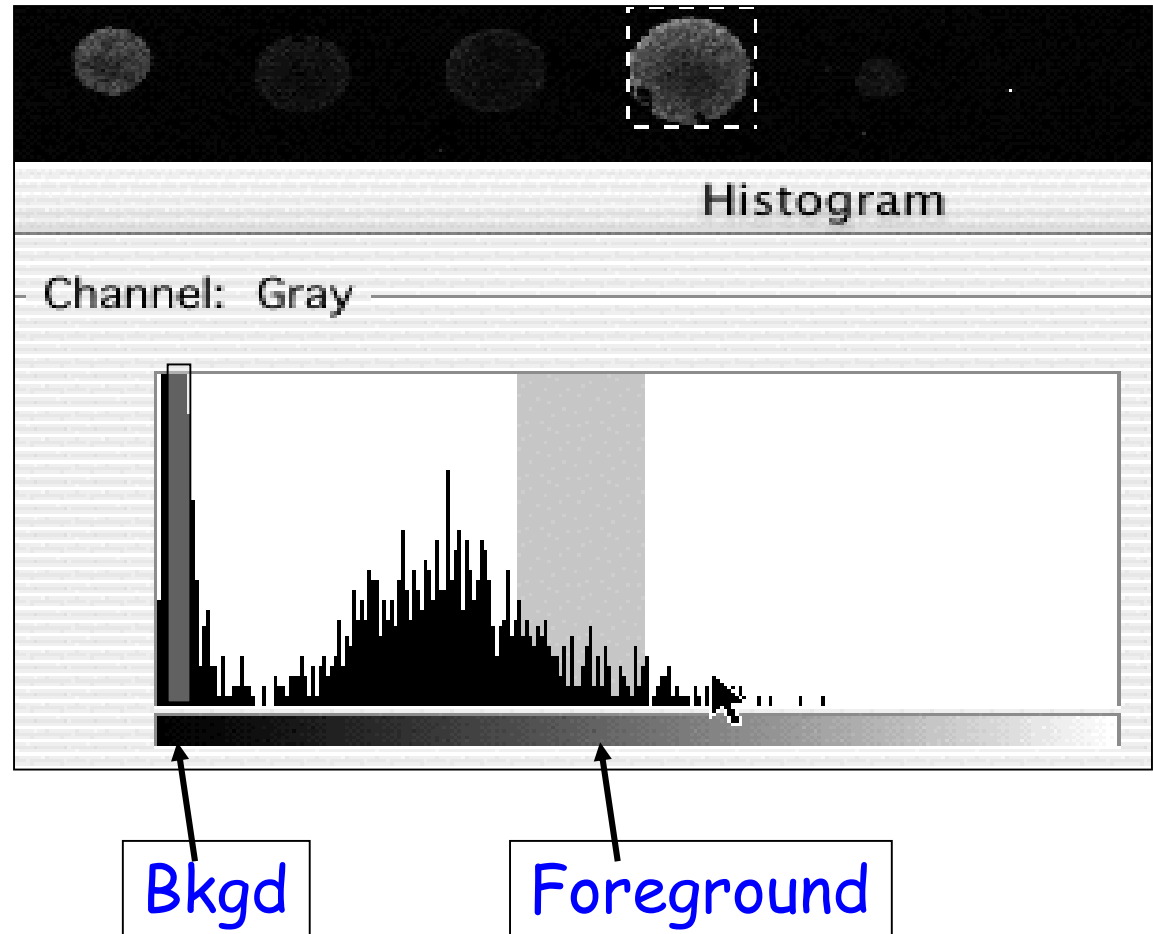
1. Ustalamy punkty startowe dla każdej pozycji w macierzy (zakładamy ich znajomość)
2. Powiększamy obszary związane z punktami startowymi na podstawie analizy jasności sąsiadujących pikseli.
3. Ostatecznie uzyskane obszary mogą mieć różne kształty.



Segmentacja na podstawie histogramu

Analizujemy obszar pokrywający w całości dany punkt. Na podstawie histogramu jasności punktów w tym obszarze ustalamy poziom jasności tła i badanej plamki.

Przykład: narzędzie QuantArray.



Przykładowe cechy:

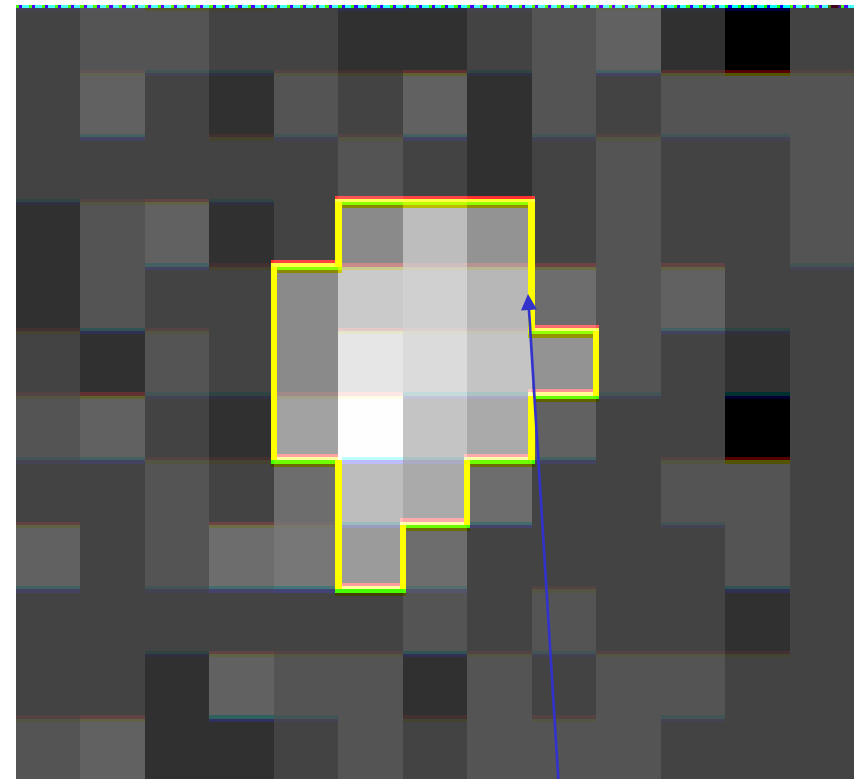
Punkty obiektu:

- średnia
- mediana
- rozkład jasności pikseli.

Jasność tła:

- nie definiujemy
- ustalana lokalnie
- wartość globalna (stała)
- operacje morfologiczne (otwarcie)

Dane jakościowe.



Take the average

Mierzony w eksperymencie poziom fluorescencji uwzględnia również wpływ niespecyficznego hybrydyzacji i zanieczyszczeń na płytce.

Poziom fluorescencji z regionów nie zawierających DNA powinien być inny niż z regionów badanych.

Rozwiązanie: stosowanie kontroli negatywnej: fragment DNA, co do którego wiemy, że nie będzie wiązał się z płytką.

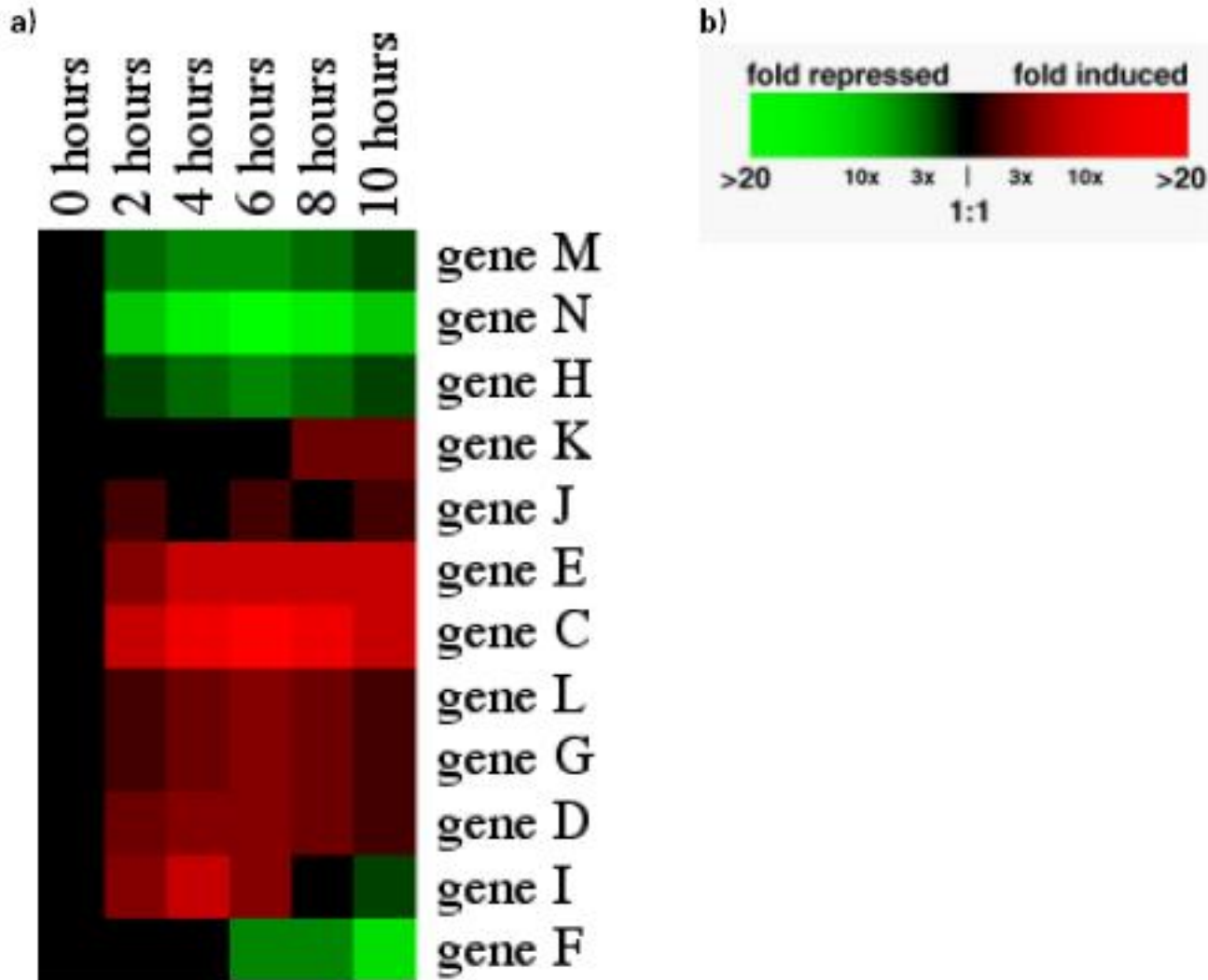
Cała macierz:

- Korelacja między intensywnościami punktów
- Stosunek liczby punktów nie dających sygnału
- Rozkład obszarów dających sygnał

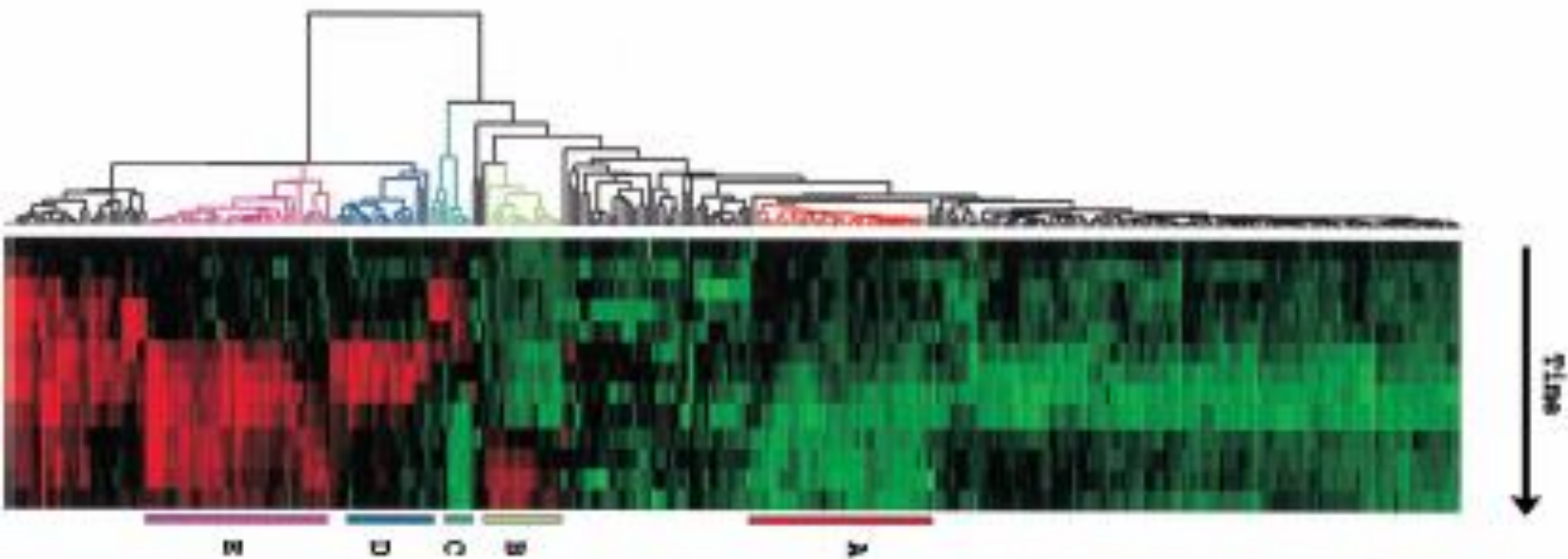
Punkt

- Stosunek sygnału do szumy
- Różnice w intensywności poszczególnych pikseli
- Identyfikacja punktów nie dających sygnału



Grupowanie genów wg stopnia aktywności w poszczególnych próbach.



Klastrowanie hierarchiczne



<http://www.ncbi.nlm.nih.gov/geo/>



Gene Expression Omnibus

[GEO Publications](#) | [FAQ](#) | [MIAME](#) | [Email GEO](#) | [Login](#)

NCBI > GEO

Gene Expression Omnibus: a public functional genomics data repository supporting MIAME-compliant data submissions. Array- and sequence-based data are accepted. Tools are provided to help users query and download experiments and curated gene expression profiles. [More information »](#)

GEO navigation

QUERY

DataSets

Gene profiles

GEO accession

GEO BLAST

DataSets

GEO accessions

Platforms

Samples

Series

Submitter login

User id:

Password:

[LOGIN](#)

[» New account](#)

[» Recover password](#)

Site contents

Public data

Platforms	9,983
Samples	726,842
Series	29,463
DataSets	2,720

Documentation

[Overview](#) | [FAQ](#) | [Find](#)
[Submission guide](#)
[Linking & citing](#)
[Journal citations](#)
[Construct a Query](#)
[Programmatic access](#)
[DataSet clusters](#)
[GEO announce list](#)
[Data disclaimer](#)
[GEO staff](#)

Query & Browse

[Repository browser](#)
[GEO2R](#)
[FTP site](#)
[GEO Profiles](#)
[GEO DataSets](#)

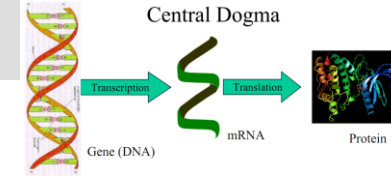
Submit

[New account](#)

RNA Seq

Analiza RNA dzięki wykorzystaniu wysokoprzepustowych technik sekwencjonowania.

Aminokwasy i białka



sekwencja -> struktura -> funkcja

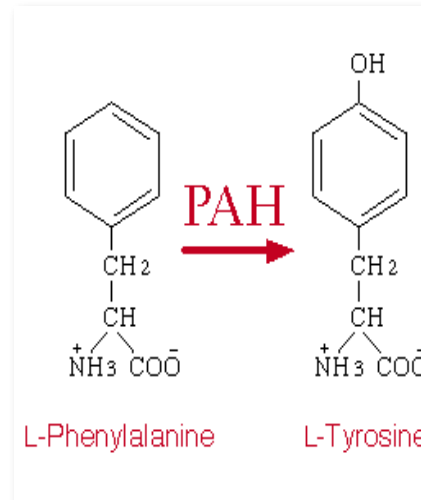
Genetic
Information

MVHLTPEEKT
AVNALWGKVN
VDAVGGEALG
RLLVVYPWTQ
RFFESFGDLS
SPDAVMGNPK
VKAHGKKVLG
AFSDGLAHL
DLNLKGTF
SQLSELHCD
KLHVDPENF
RLGNVLVC
VLARNFGKE
FTPQMQAAY
QKVVAGVAN
ALAHKYH

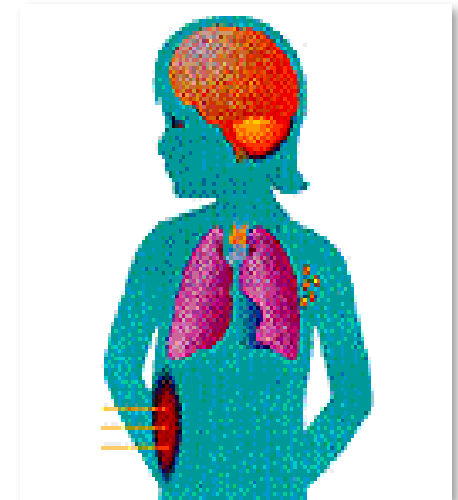
Molecular
Structure



Biochemical
Function

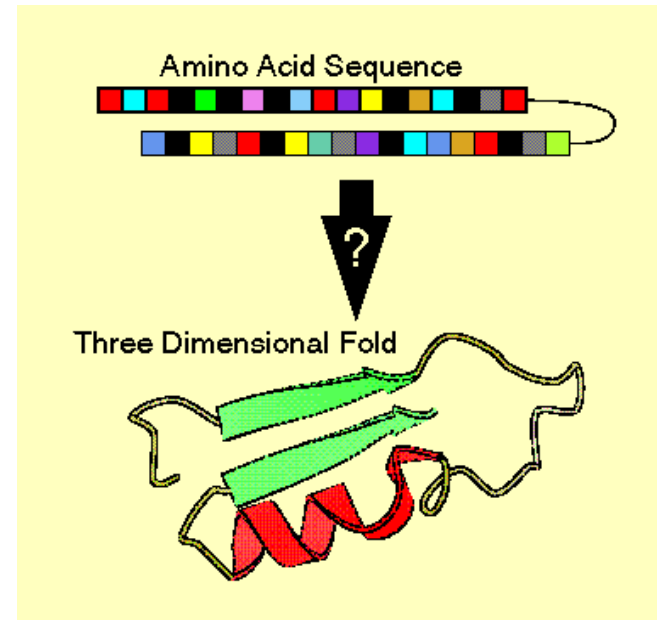


Phenotype
(Symptoms)

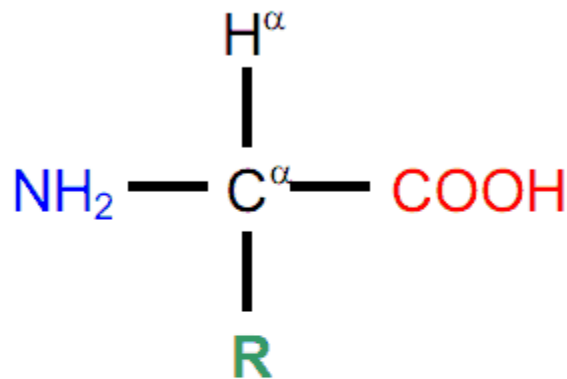


Dotyczą zarówno białek, jak i kwasów nukleinowych (głównie RNA).

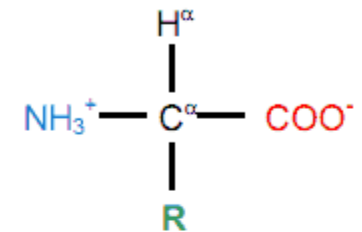
- przewidywanie struktury
- poszukiwanie miejsc wiążących
- predykcja oddziaływań
- modelowanie molekularne
- analiza relacji ewolucyjnych
- przewidywanie funkcji
- projektowanie leków (CADD)



Ogólna budowa aminokwasów:



- w neutralnym pH



grupa aminowa - NH_2

grupa karboksylowa - COOH

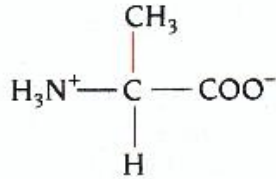
Aminokwasy białkowe

nazwa	symbol	skrót
alanina	A	Ala
arginina	R	Arg
asparagina	N	Asn
kw.asparaginowy	D	Asp
cysteina	C	Cys
glutamina	Q	Gln
kw.glutaminowy	E	Glu
glicyna	G	Gly
histydyna	H	His
izoleucyna	I	Ile

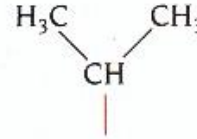
nazwa	symbol	skrót
leucyna	L	Leu
lizyna	K	Lys
metionina	M	Met
fenyloalanina	F	Phe
prolina	P	Pro
seryna	S	Ser
treonina	T	Thr
tryptofan	W	Trp
tyrozyna	Y	Tyr
walina	V	Val

Aminokwasy hydrofobowe

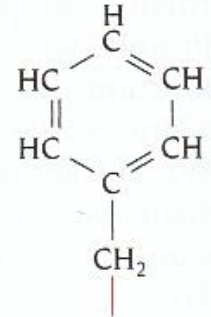
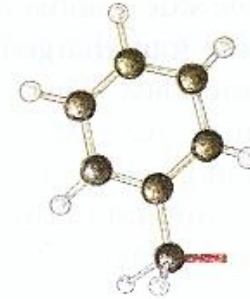
(a) Hydrophobic amino acids



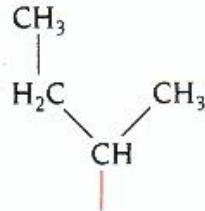
A Ala, Alanine



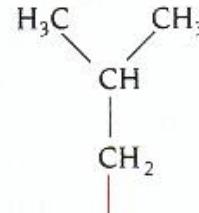
V Val, Valine



F Phe, Phenylalanine

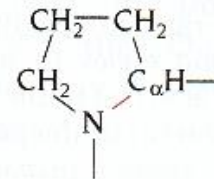
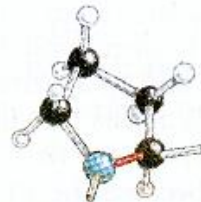


I Ile, Isoleucine

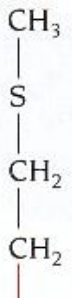


L Leu, Leucine

**A, V, F, I, L, P, M
C, G, Y, W, H, K, T**



P Pro, Proline



M Met, Methionine

Aminokwasy polarne

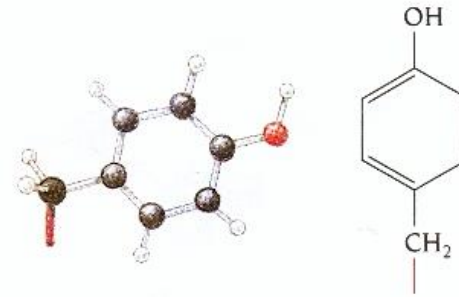
(c) Polar amino acids



S Ser, Serine



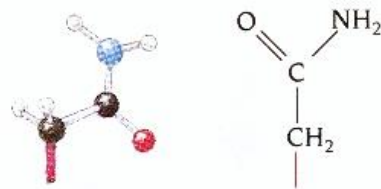
T Thr, Threonine



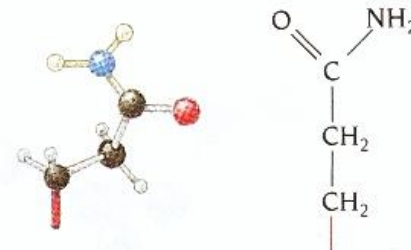
Y Tyr, Tyrosine



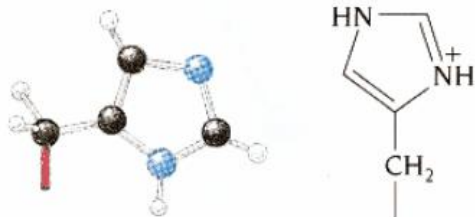
C Cys, Cysteine



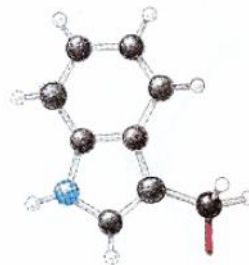
N Asn, Asparagine



Q Gln, Glutamine



H His, Histidine

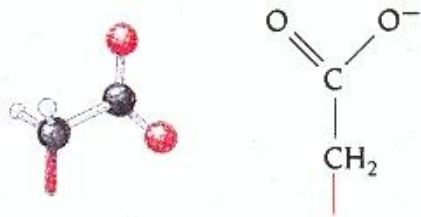


W Trp, Tryptophan

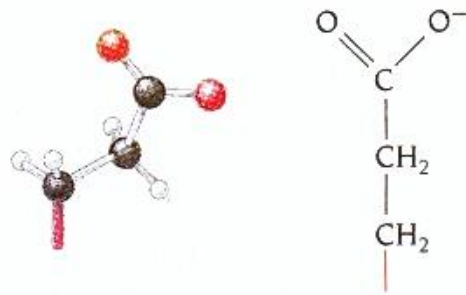
**S, T, Y, C, N, Q, H, W
D, E, K, R**

Aminokwasy naładowane

(b) Charged amino acids



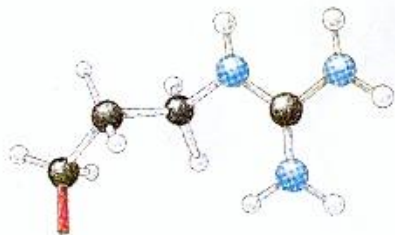
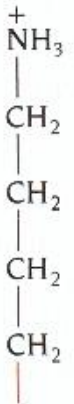
D Asp, Aspartic acid



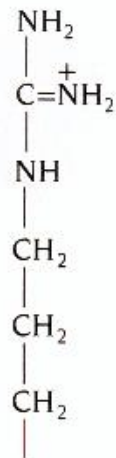
E Glu, Glutamic acid



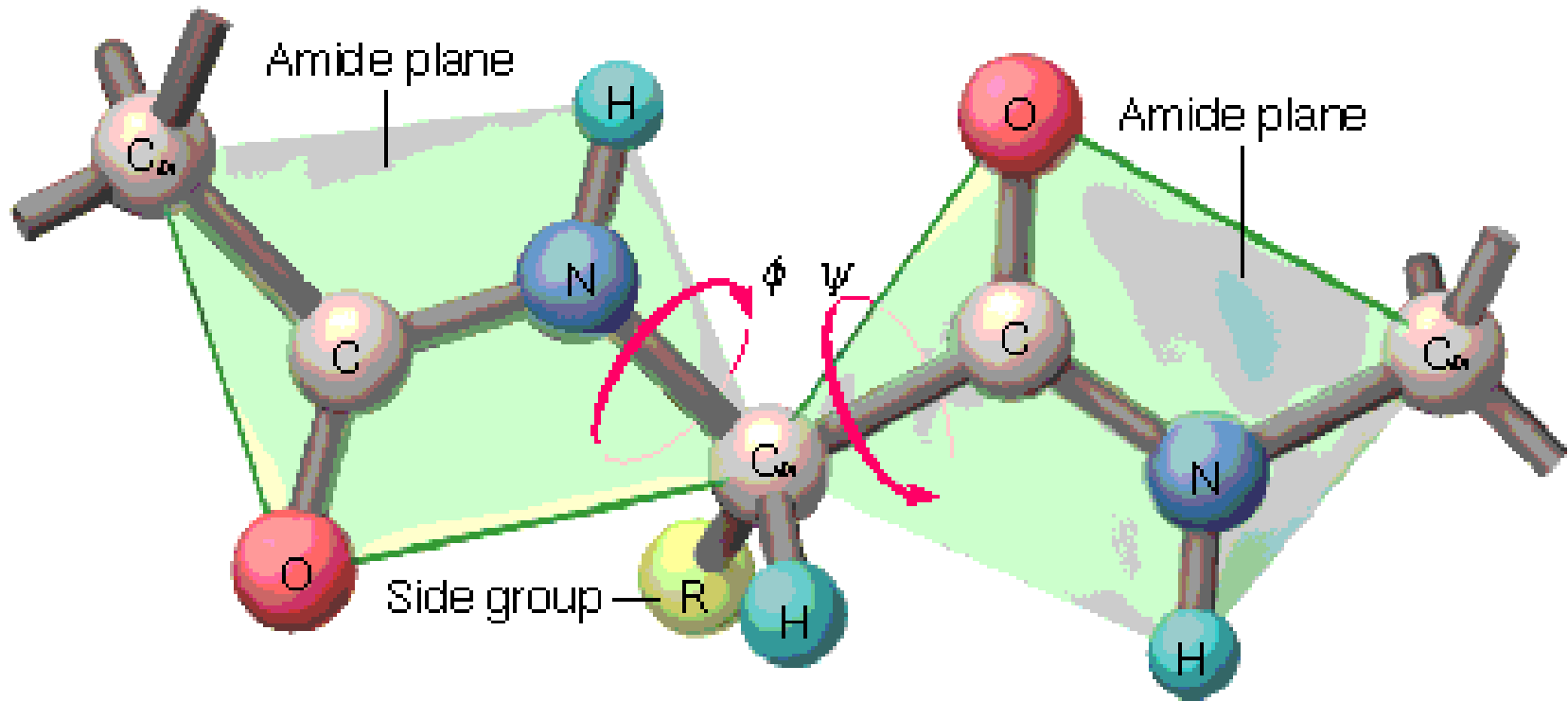
K Lys, Lysine



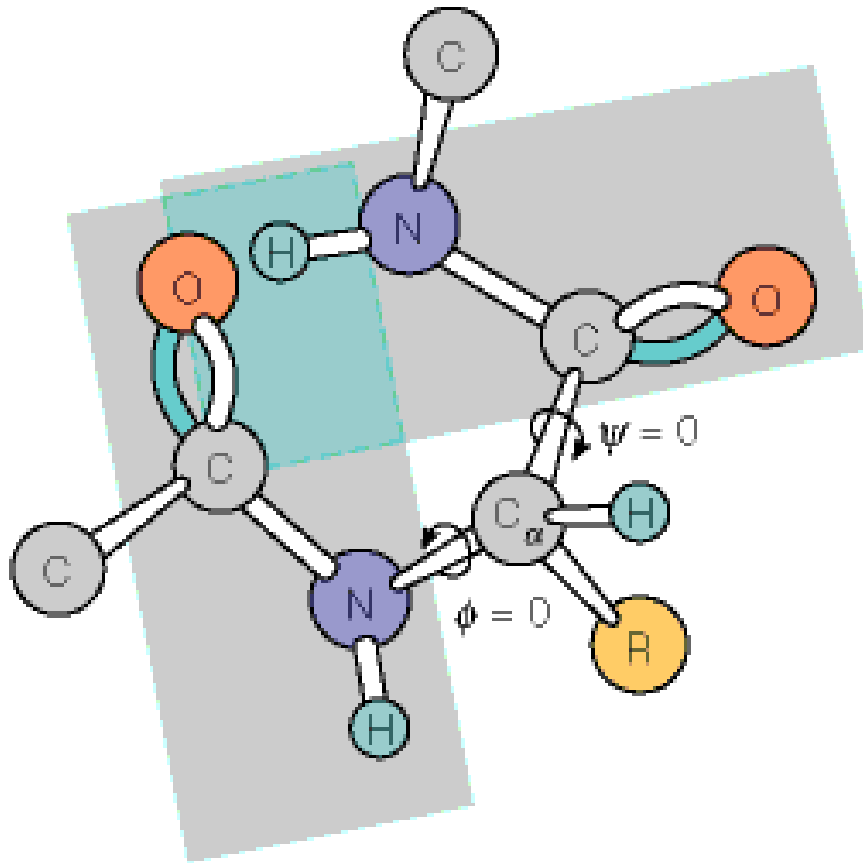
R Arg, Arginine



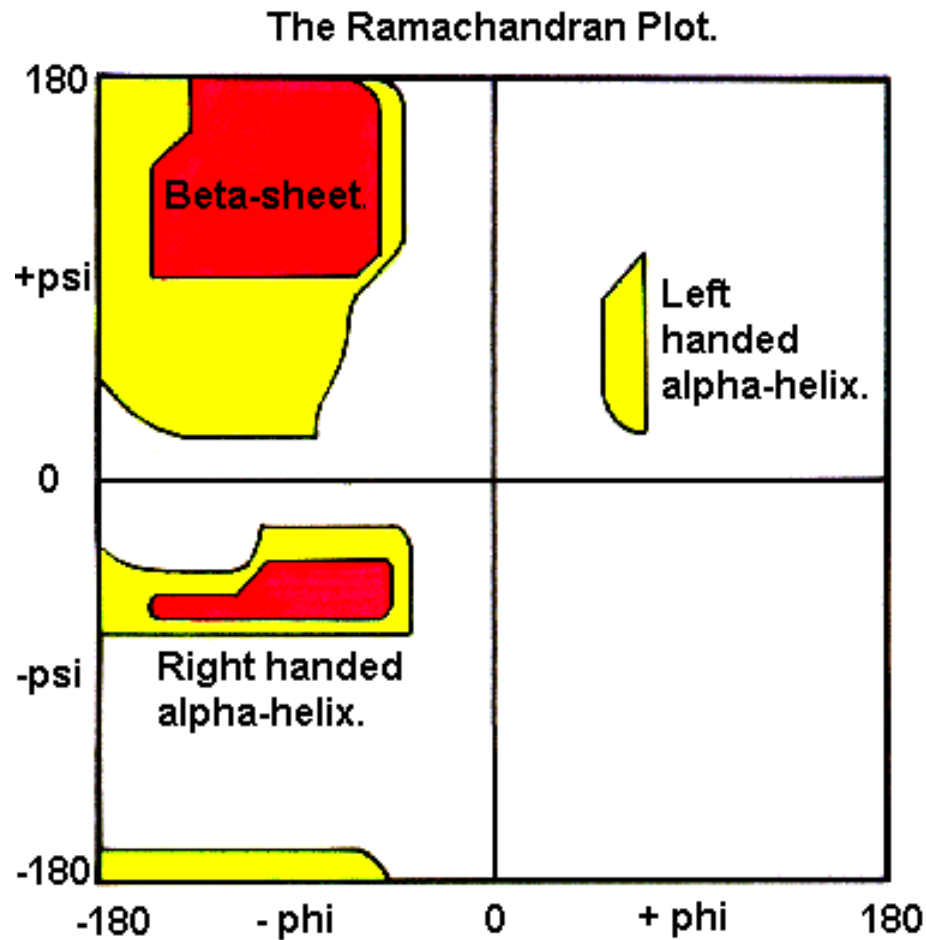
D, E, K, R, H



Ponieważ wiązanie peptydowe jest płaskie, konformacja głównego łańcucha może być wyznaczona przez kąty ϕ i ψ .



Niektóre konformacje są niedopuszczalne.



Obrazuje dopuszczalne konformacje.

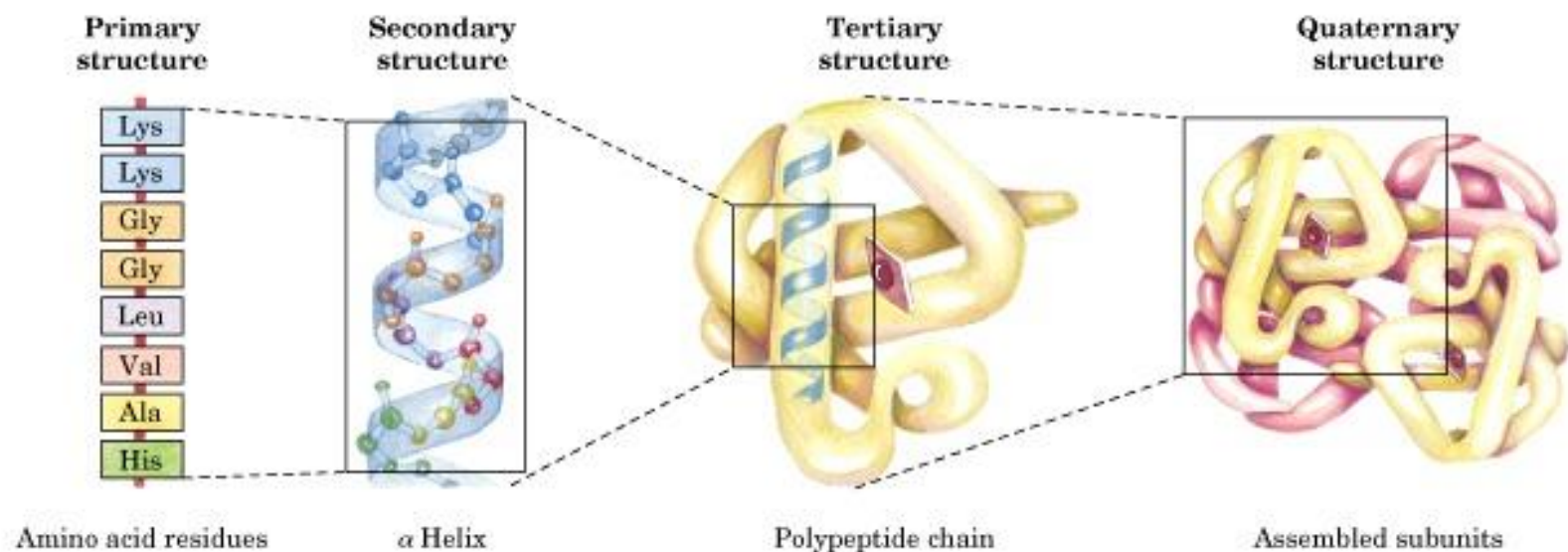
Poziomy przestrzennej organizacji białek:

I rządowa – liniowa sekwencja aminokwasów

II rządowa – opisuje lokalne pofałdowanie (α -helisy, β -kartki)

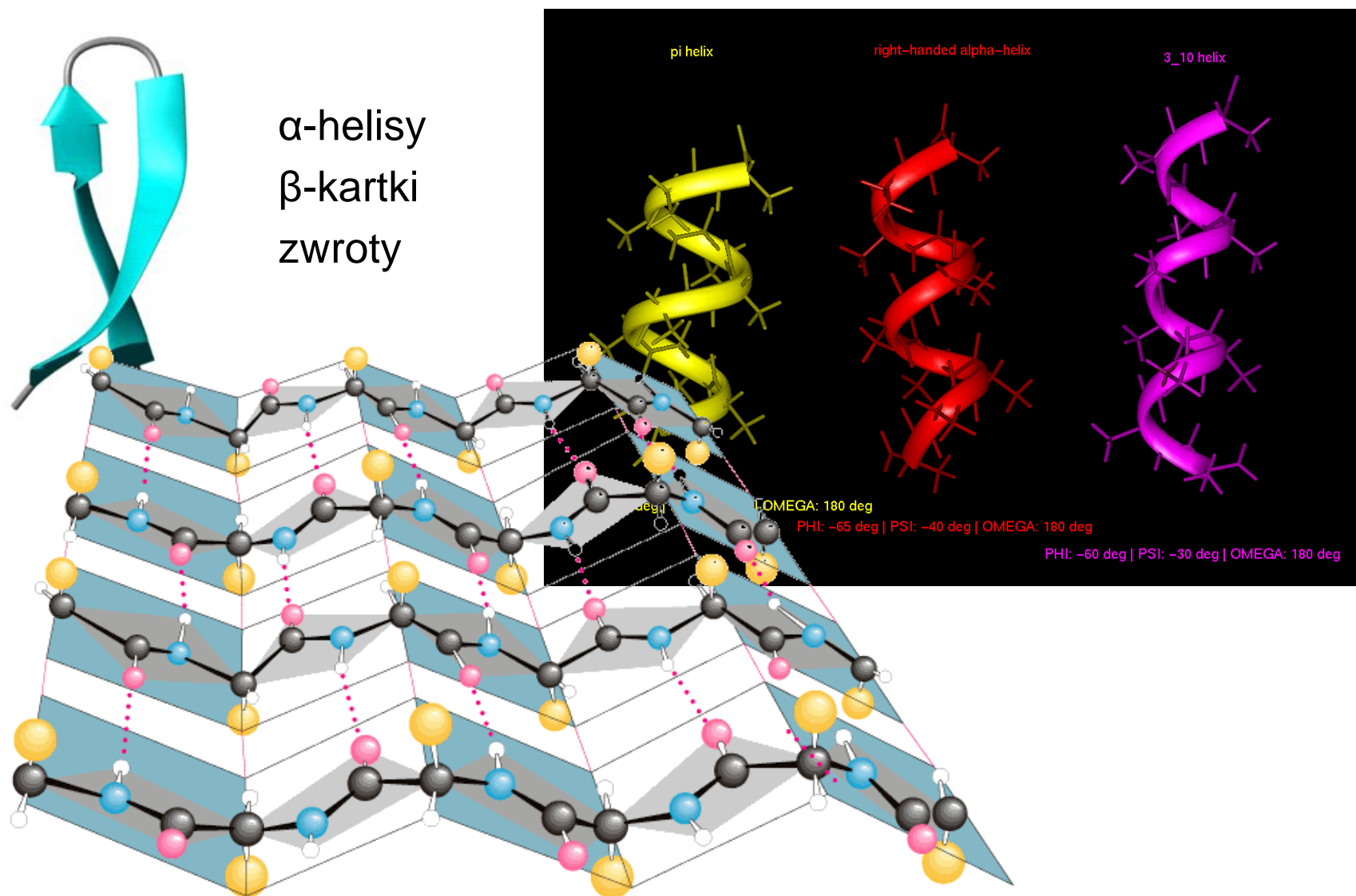
III rządowa – struktura 3D pojedynczego łańcucha

IV rządowa – struktura 3D całego białka (połączone wszystkie łańcuchy)



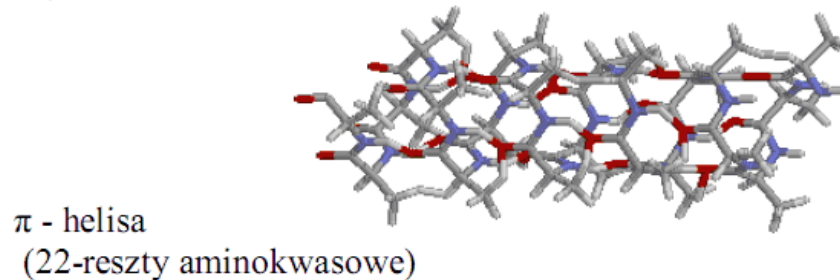
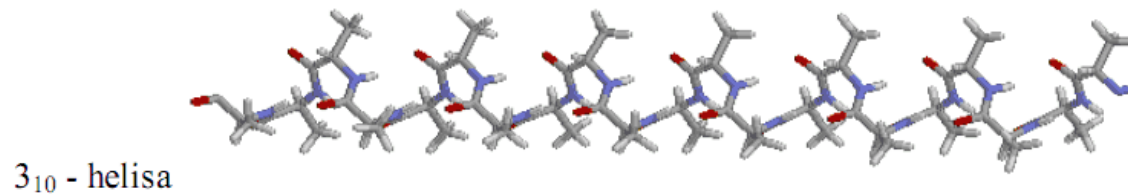
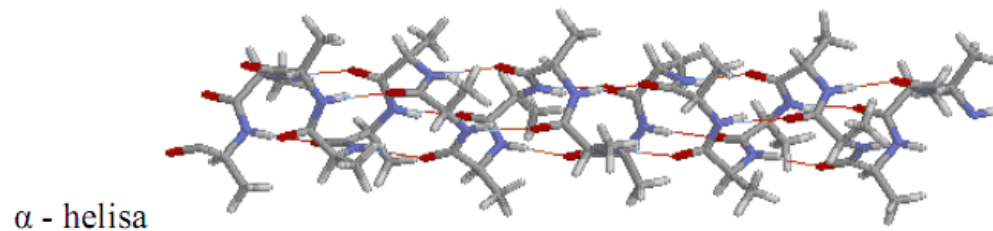
Zwijanie białka (film): <http://www.youtube.com/watch?v=fvBO3TqJ6FE>

Struktura 2-rzędowa



α -helisy

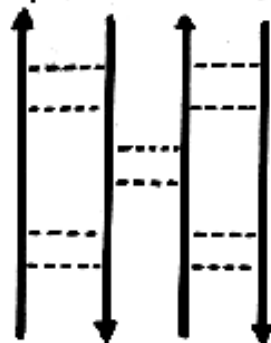
helisa	φ	ψ	ω	reszt aminkwasowych na skręt	przesunięcie na resztę (Å)	wiązania wodorowe
α helisa	-57	-47	180	3,6	1,5	i+4
3_{10} helisa	-49	-26	180	3,0	2,0	i+3
π helisa	-57	-70	180	4,4	1,2	i+5



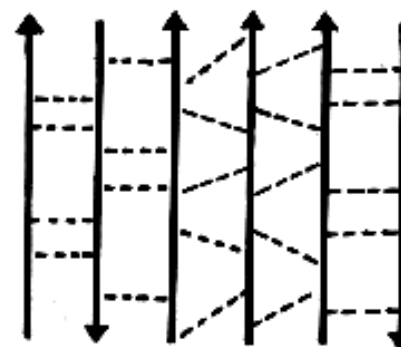
Są stabilizowane przez wiązania wodorowe i kontakt sąsiadujących łańcuchów

- równoległe
- antyrównoległe
- mieszane

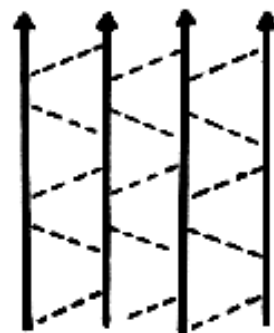
Antiparallel beta-sheet



The different types of beta-sheet. Dashed lines indicate main chain hydrogen bonds.

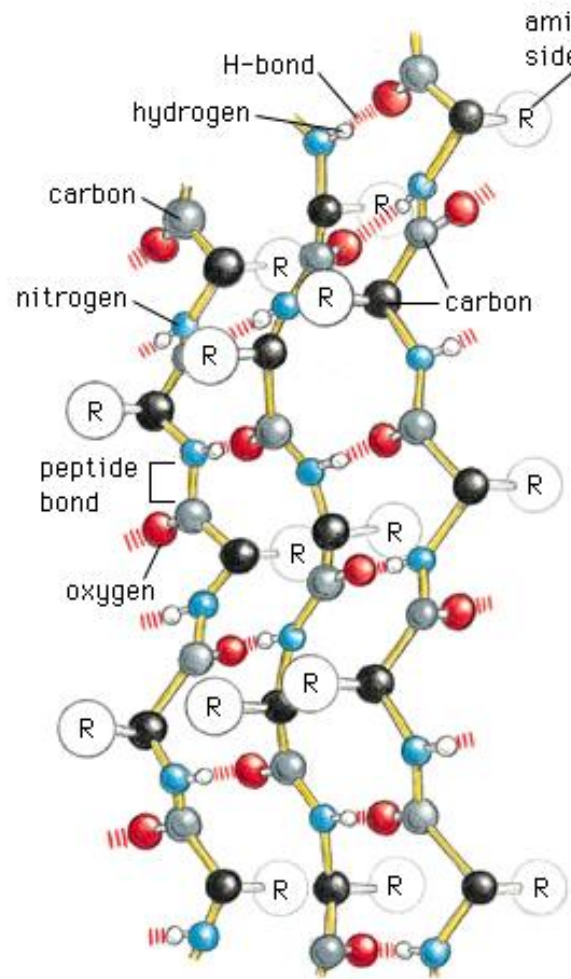


Mixed beta-sheet

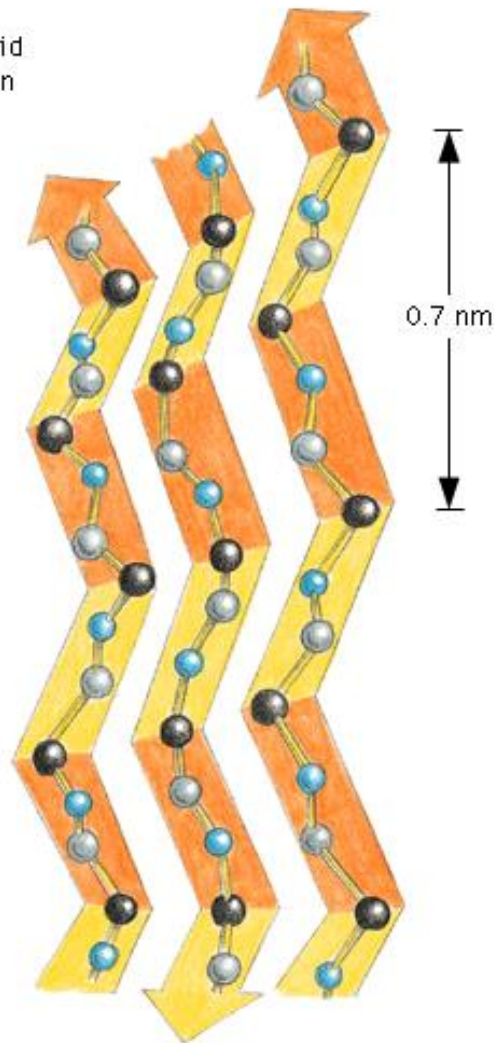


Parallel beta-sheet

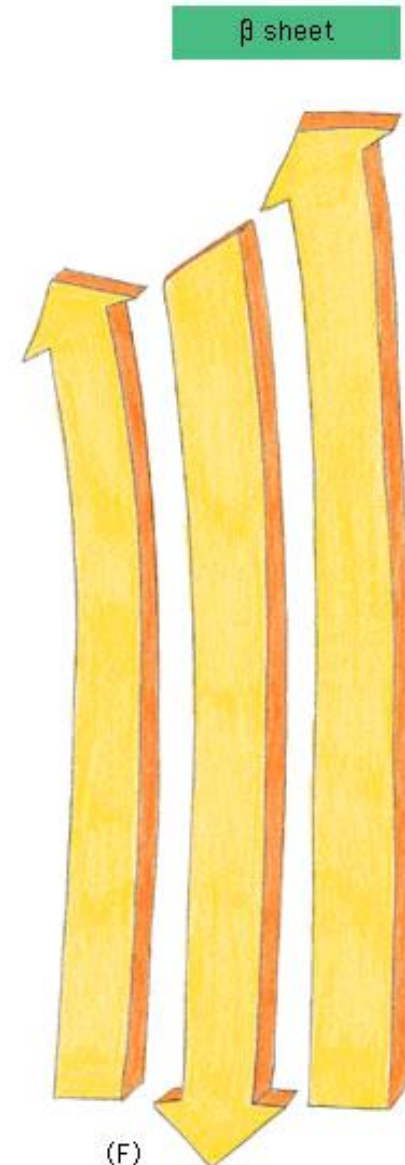
β -kartki – reprezentacja graficzna



(D)

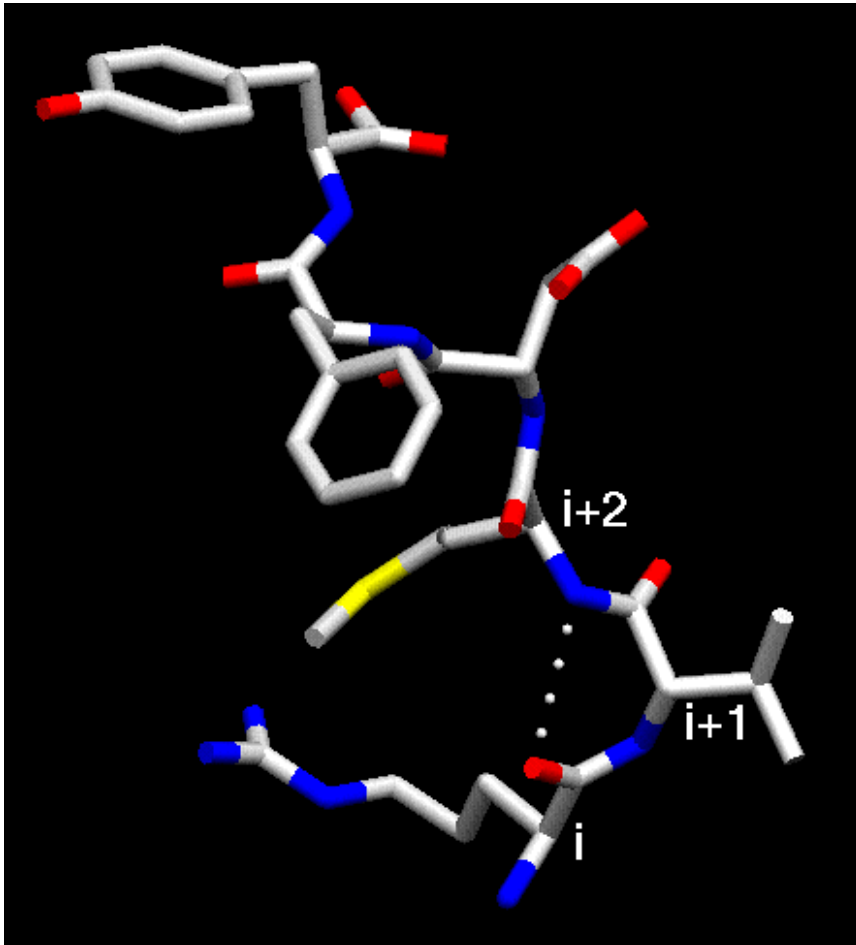


(E)

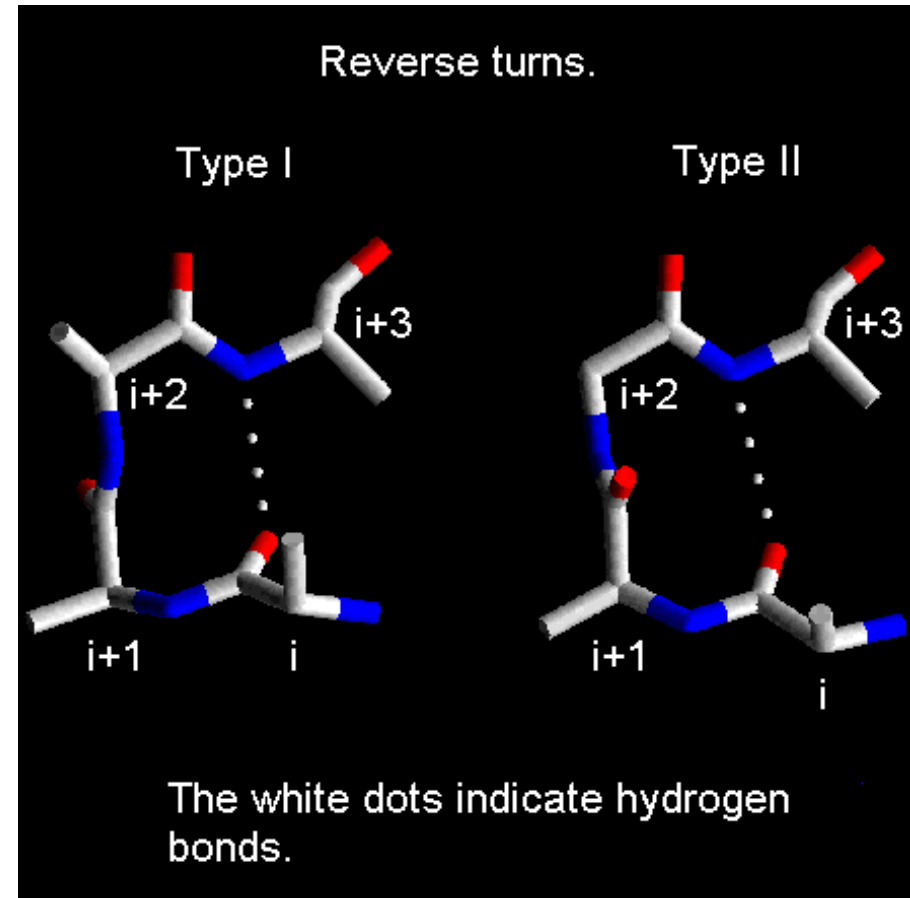


(F)

©1998 GARLAND PUBLISHING



γ -zwrot ($\phi_{i+1}=-79^\circ$, $\psi_{i+1}=69^\circ$)



β -zwroty

Typy β -zwrotów w białkach

Table 1. Frequency and mean dihedral angles for standard β -turn types

Turn type	Ramachandran nomenclature ^a	No. of turns		Mean dihedral angles ^d			
		b	c	$\phi(i+1)$	$\psi(i+1)$	$\phi(i+2)$	$\psi(i+2)$
I	$\alpha_R \alpha_R$	1,231	1,419	-64 (-60)	-27 (-30)	-90 (-90)	-7 (0)
II	$\beta \gamma_L$	405	489	-60 (-60)	131 (120)	84 (80)	1 (0)
VIII	$\alpha_R \beta$	325	451	-72 (-60)	-33 (-30)	-123 (-120)	121 (120)
I'	$\alpha_L \gamma_L$	127	142	55 (60)	38 (30)	78 (90)	6 (0)
II'	$\epsilon \alpha_R$	90	100	60 (60)	-126 (-120)	-91 (-80)	1 (0)
VIa1	$\beta \alpha_R$	15	17	-64 (-60)	142 (120)	-93 (-90)	5 (0)
VIa2	$\beta \alpha_R$	5	5	-132 (-120)	139 (120)	-80 (-60)	-10 (0)
VIb	$\beta \beta$	35	35	-135 (-135)	131 (135)	-76 (-75)	157 (160)
IV		1,666	1,241	-61	10	-53	17
TOTAL		3,899					

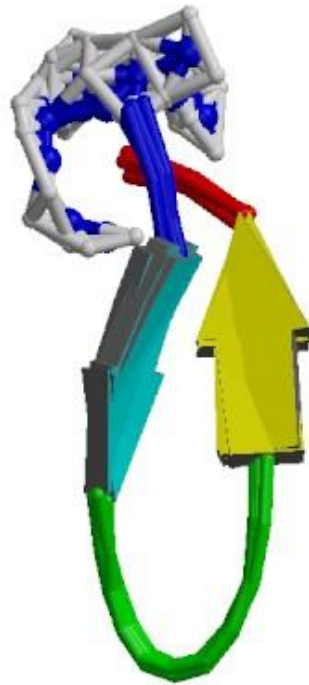
^a Ramachandran nomenclature for turn type as in Wilmot and Thornton (1990). The nomenclature describes the regions of the Ramachandran plot occupied by residues $i+1$ and $i+2$ of the turn.

^b Using normal cutoffs of 30° for deviation from standard angles, with one angle allowed to deviate by 45°.

^c Allowing up to 40° deviation from standard angles, with one angle allowed to deviate by 50°.

^d The idealized ϕ, ψ values as determined by Lewis et al. (1973) are given in parentheses after the averaged values determined from the data set. The values for the type VI turns are taken from Richardson (1981). Types VIa1 and VIa2 are the two subclasses of type VIa turns identified by Richardson (1981).

Odmiana β –zwrotu. Częsty wzór, zwykle złożony z 3 aminokwasów, łączący sąsiadujące, antyrównoległe β -kartki

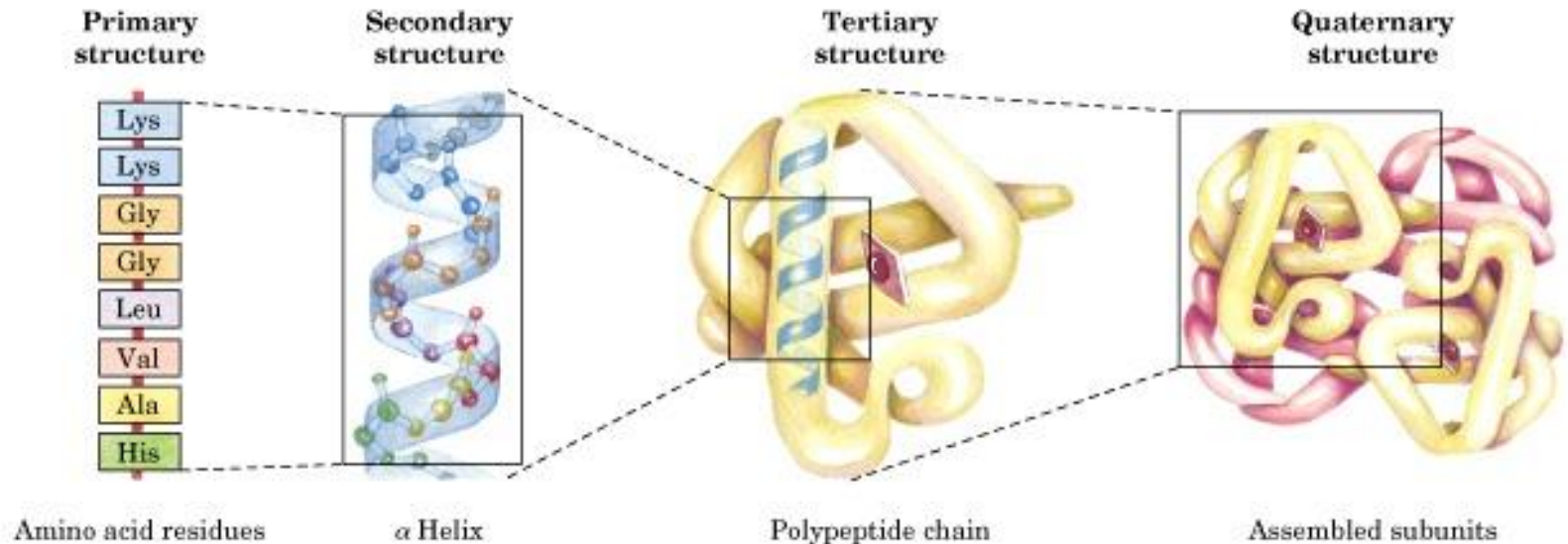


Jeśli każda reszta aminokwasowa może przybierać tylko trzy różne położenia przestrzenne. To dla białka o długości 100 aminokwasów, całkowita ilość struktur jakie może przyjąć wynosi 3^{100} , czyli ok. $5 \cdot 10^{47}$.

Jeżeli czas potrzebny na przekształcenie jednej struktury w drugą wynosi 10^{-13} s, to całkowity czas potrzebny na ustalenie optymalnej struktury wyniósłby $5 \cdot 10^{47} \cdot 10^{-13}$ s ($1,6 \cdot 10^{27}$ lat).

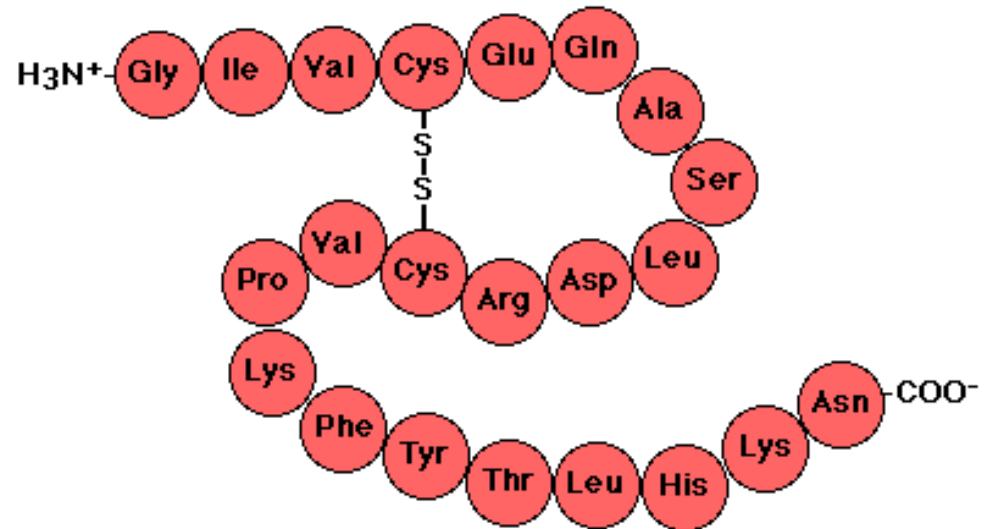
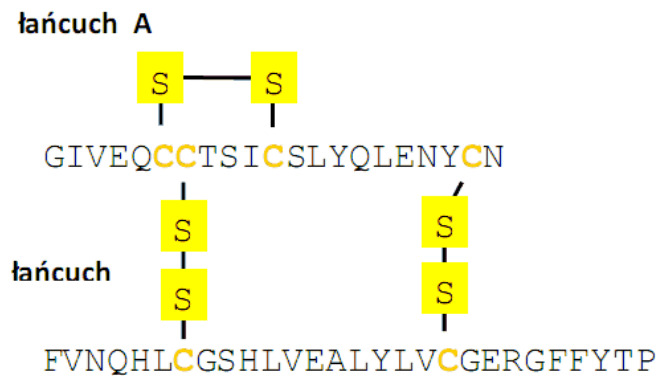
Tymczasem rzeczywisty czas potrzebny na zwinięcie się białka liczony jest w mikrosekundach.

Anfinsen (1961) – cała informacja potrzebna białku do przyjęcia ostatecznej konformacji zakodowana jest w jego strukturze pierwszorzędowej



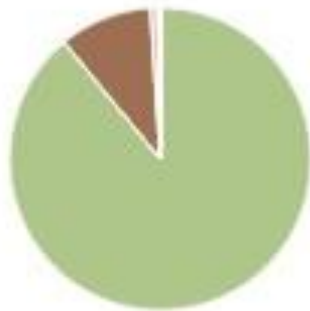
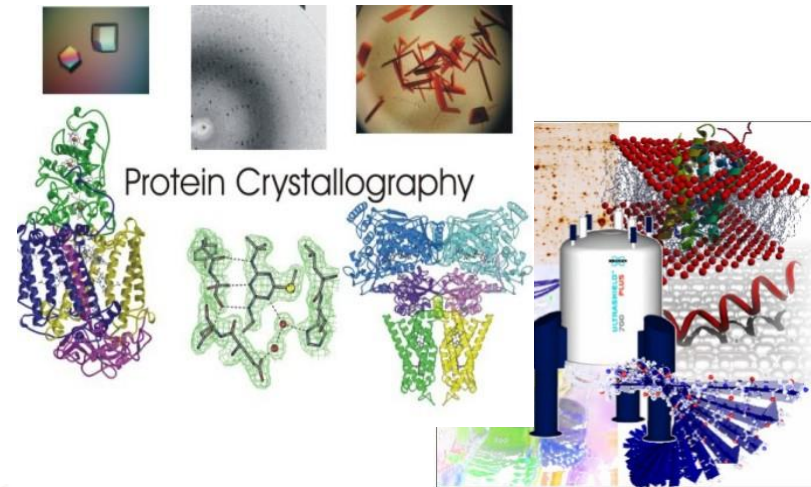
Stabilizacja struktury przestrzennej

- wiązania wodorowe;
- mostki dwusiarczkowe
- oddziaływania elektrostatyczne;
- siły van der Waalsa



- **Krystalografia rentgenowska (X-ray)**
- Jądrowy rezonans magnetyczny (NMR)
- inne

Metody drogie i czasochłonne.



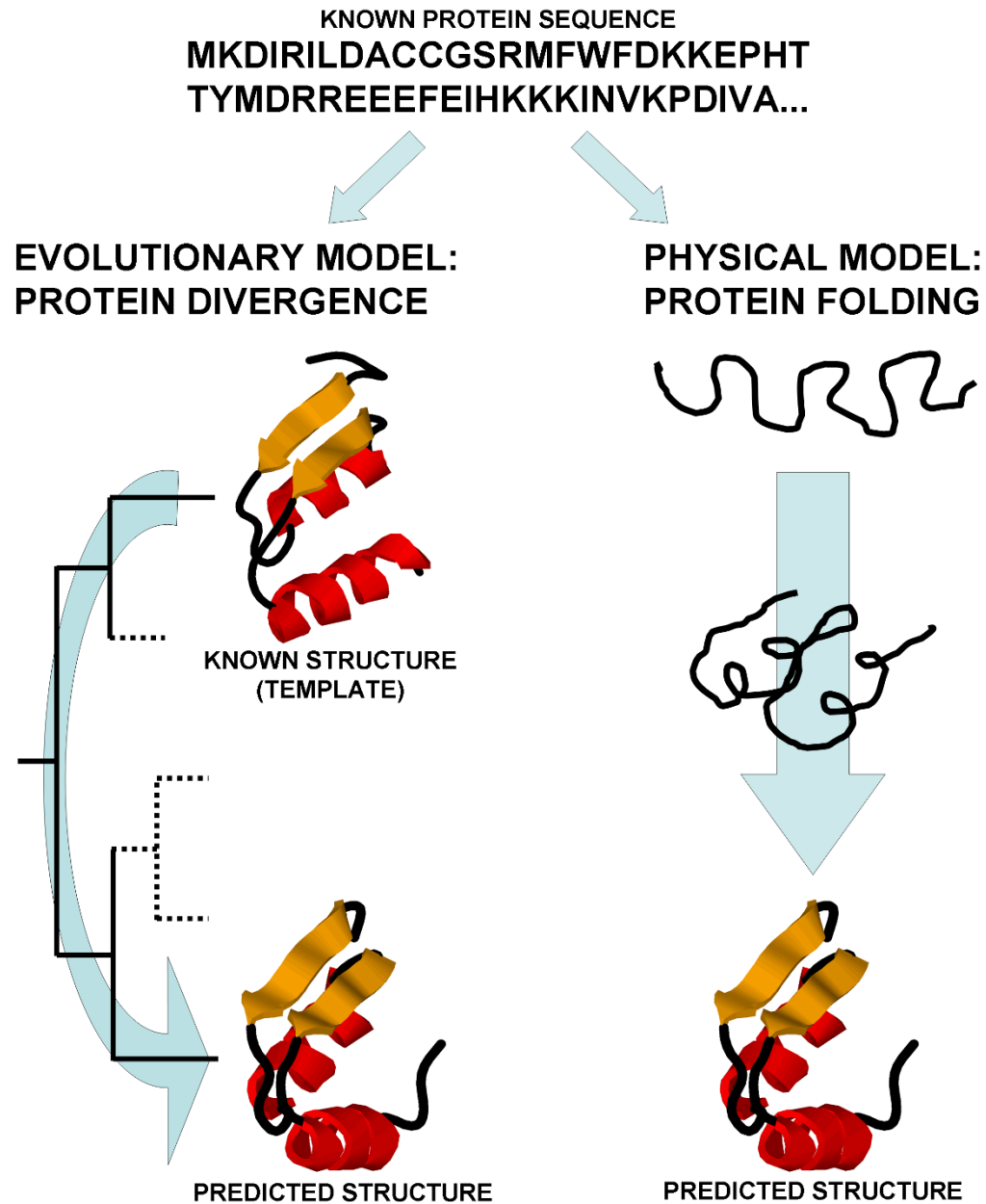
X-ray (96957)
Solution NMR (10879)
Electron Microscopy (781)
Solid-State NMR (76)
Hybrid (76)
Electron Crystallography (47)
Neutron Diffraction (47)
Fiber Diffraction (38)
Solution Scattering (32)
Other (24)

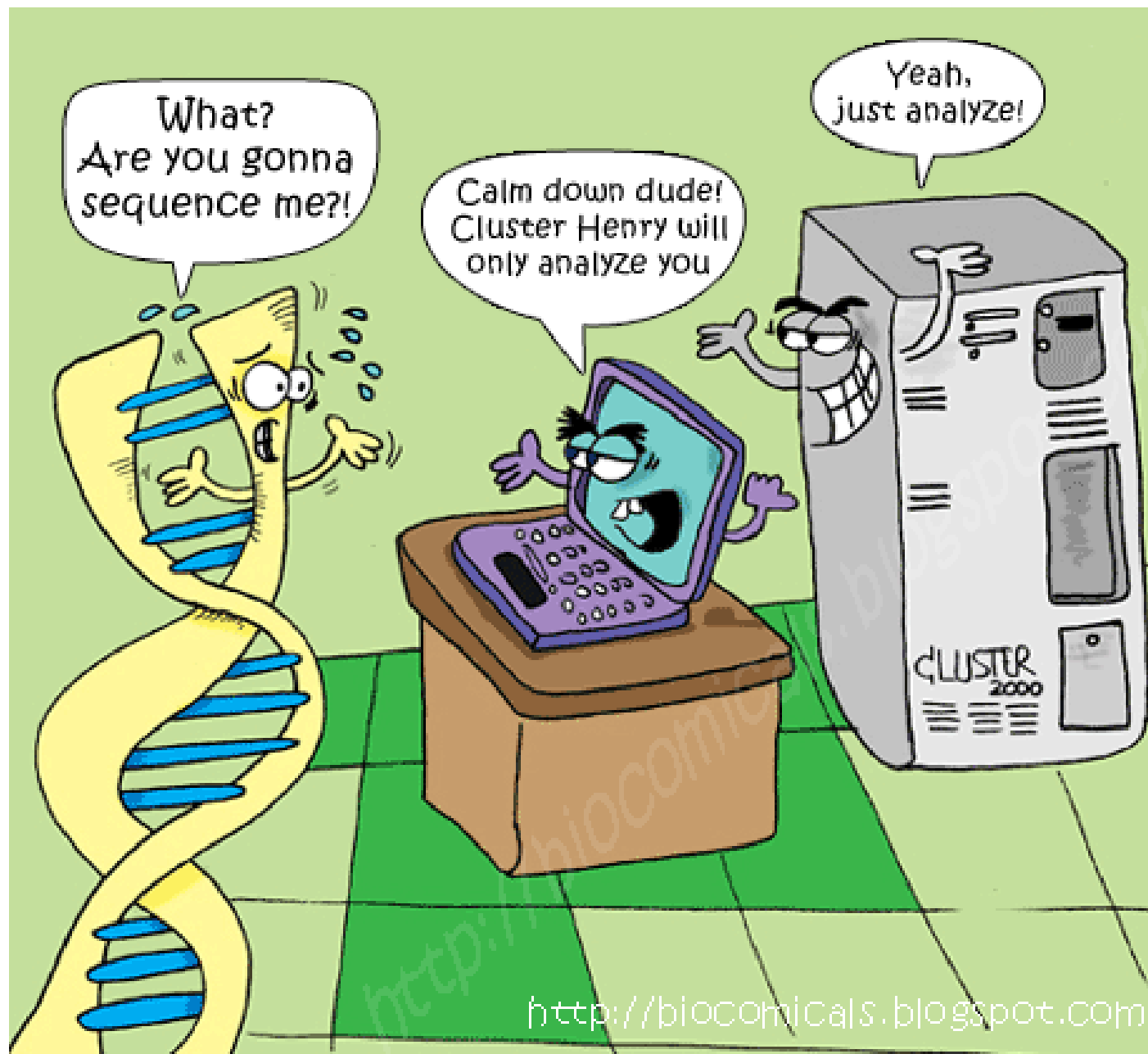
Podejście fizyczne – szkoła boltzmannowska

Modelowanie zwijania białka (procesu poszukiwania przez łańcuch konformacji o najniższej energii swobodnej, który w komórkach trwa zaledwie ułamki sekundy) korzystając z praw fizyki statycznej.

Podejście ewolucyjne – szkoła darwinowska

Rekonstrukcja procesu powstania sekwencji i struktury białka na drodze ewolucji (przyrodzie zabiera to miliony lat).





<http://biocomicals.blogspot.com>