

# Uczenie maszynowe w bioinformatyce

## Wykład 1: kwasy nukleinowe (DNA)

Tymon Rubel

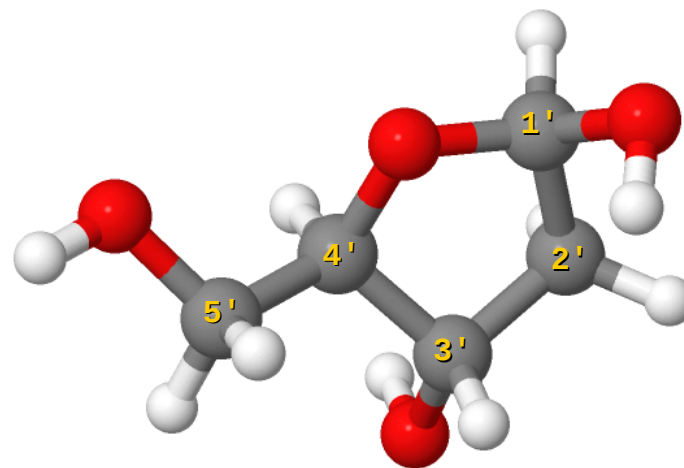
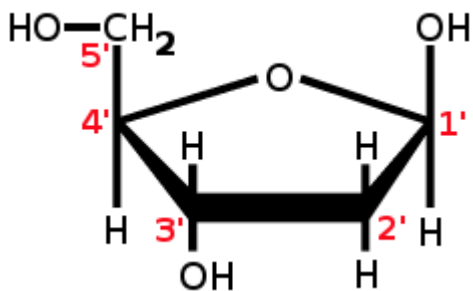
Zakład Elektroniki Jądrowej i Medycznej  
Instytut Radioelektroniki i Techniki Multimedialnych PW

# Kwas deoksyrybonukleinowy (DNA)

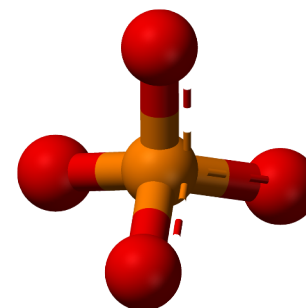
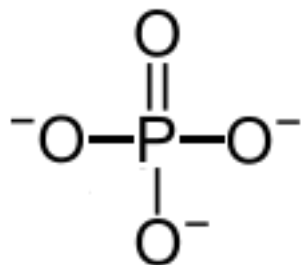
**Kwas deoksyrybonukleinowy (DNA)** jest nośnikiem informacji o budowie i działaniu wszystkich organizmów komórkowych oraz znacznej części wirusów. Jest to liniowy nierozgałęziony polimer złożony z czterech rodzajów **deoksyrybonukleotydów**.

Elementami składowymi każdego z deoksyrybonukleotydów są:

- **2'-deoksyryboza** (cukier prosty o pięciu atomach węgla, czyli tzw. pentoza);



- **grupa fosforanowa** (złożona z jednej, dwóch lub trzech reszt fosforanowych);

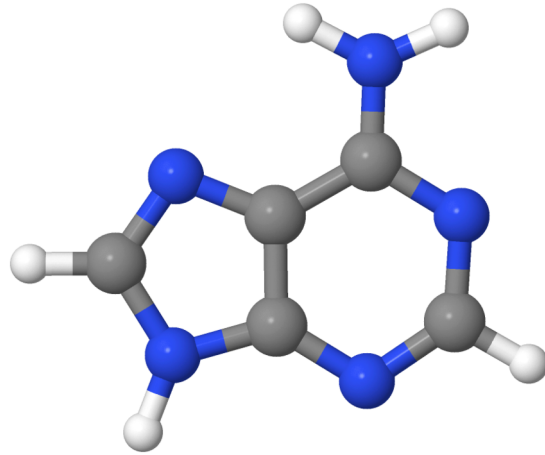
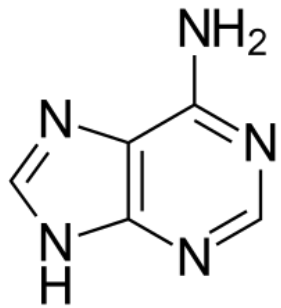


# Kwas deoksyrybonukleinowy (DNA)

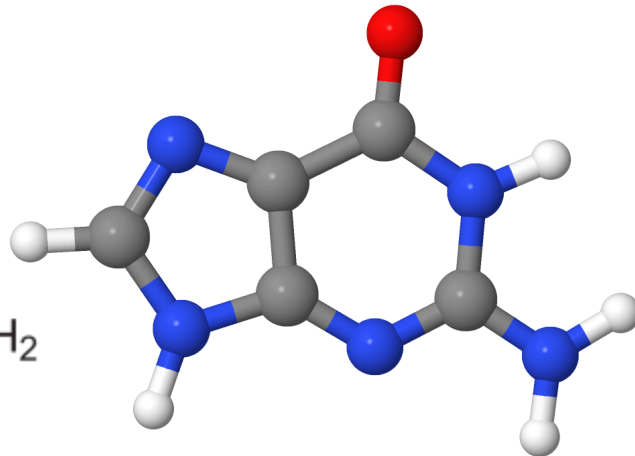
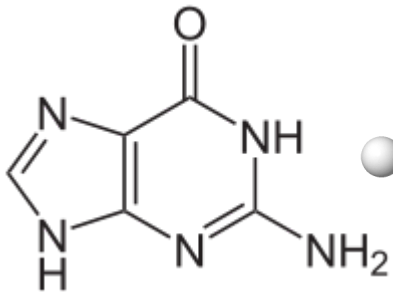
▪ jedna z czterech **zasad azotowych**.

## Zasady purynowe

### Adenina (A)

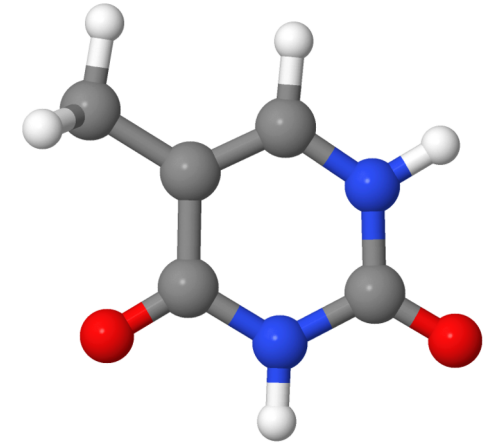
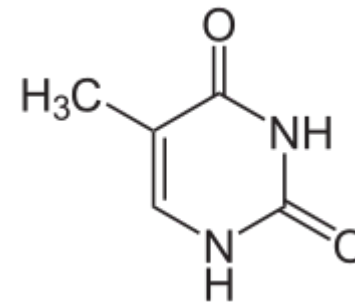


### Guanina (G)

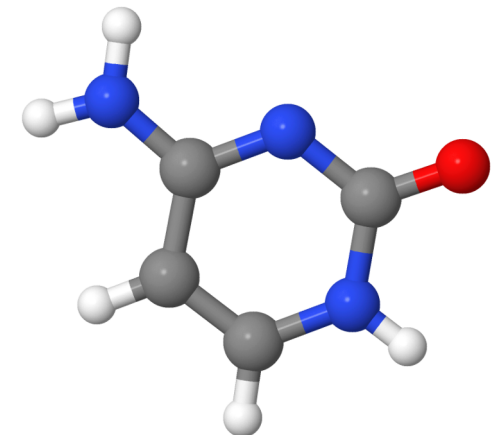
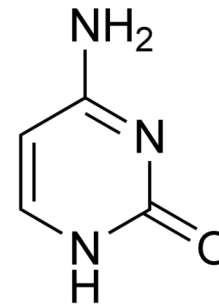


## Zasady pirymidynowe

### Tymina (T)

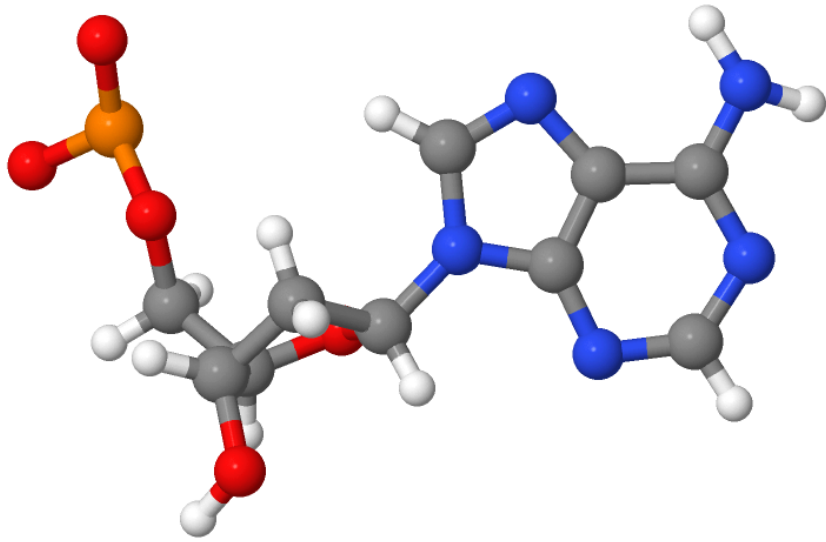


### Cytozyna (C)

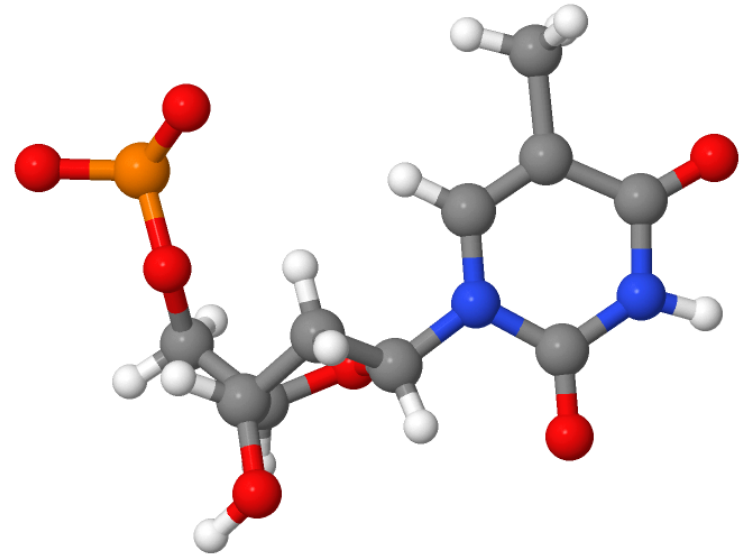


# Kwas deoksyrybonukleinowy (DNA): deoksyrybonukleotydy

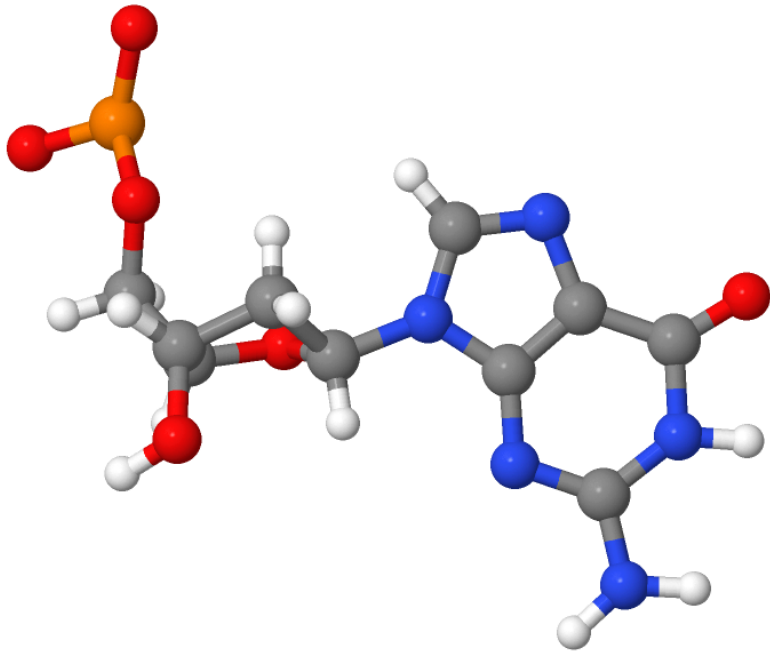
deoksyadenozyno-5'-monofosforan (dAMP)



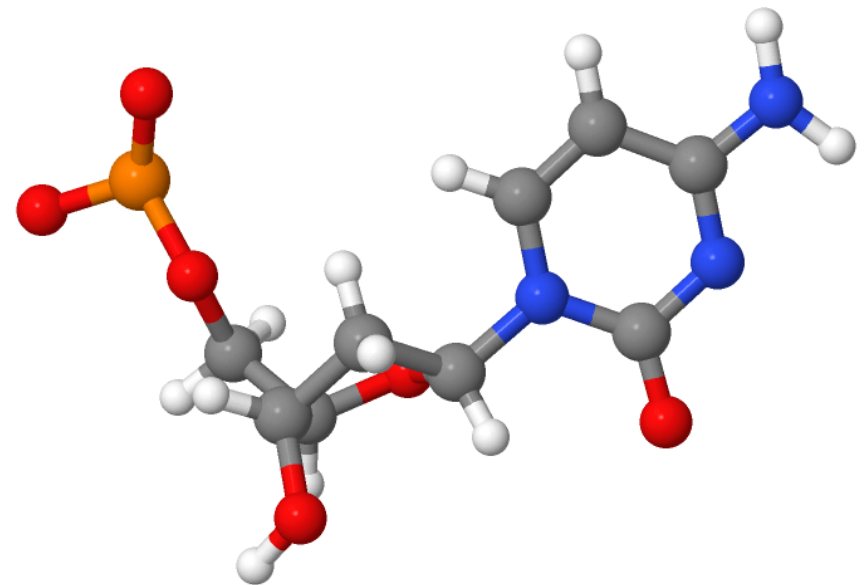
deoksytymidyno-5'-monofosforan (dTMP)



deoksyguanozyno-5'-monofosforan (dGMP)



deoksyocytydino-5'-monofosforan (dCMP)



# Kwas deoksyrybonukleinowy (DNA): łańcuchy polinukleotydowe

**Łańcuch polinukleotydowy** powstaje dzięki **wiązaniom fosfodiesterowym** łączącym atomy węgla 5' i 3' deoksyrybozy sąsiednich nukleotydów.

Łańcuch ma dwa różne końce:

**koniec 5'** (z grupą fosforanową)

**koniec 3'** (z grupą hydroksylową)

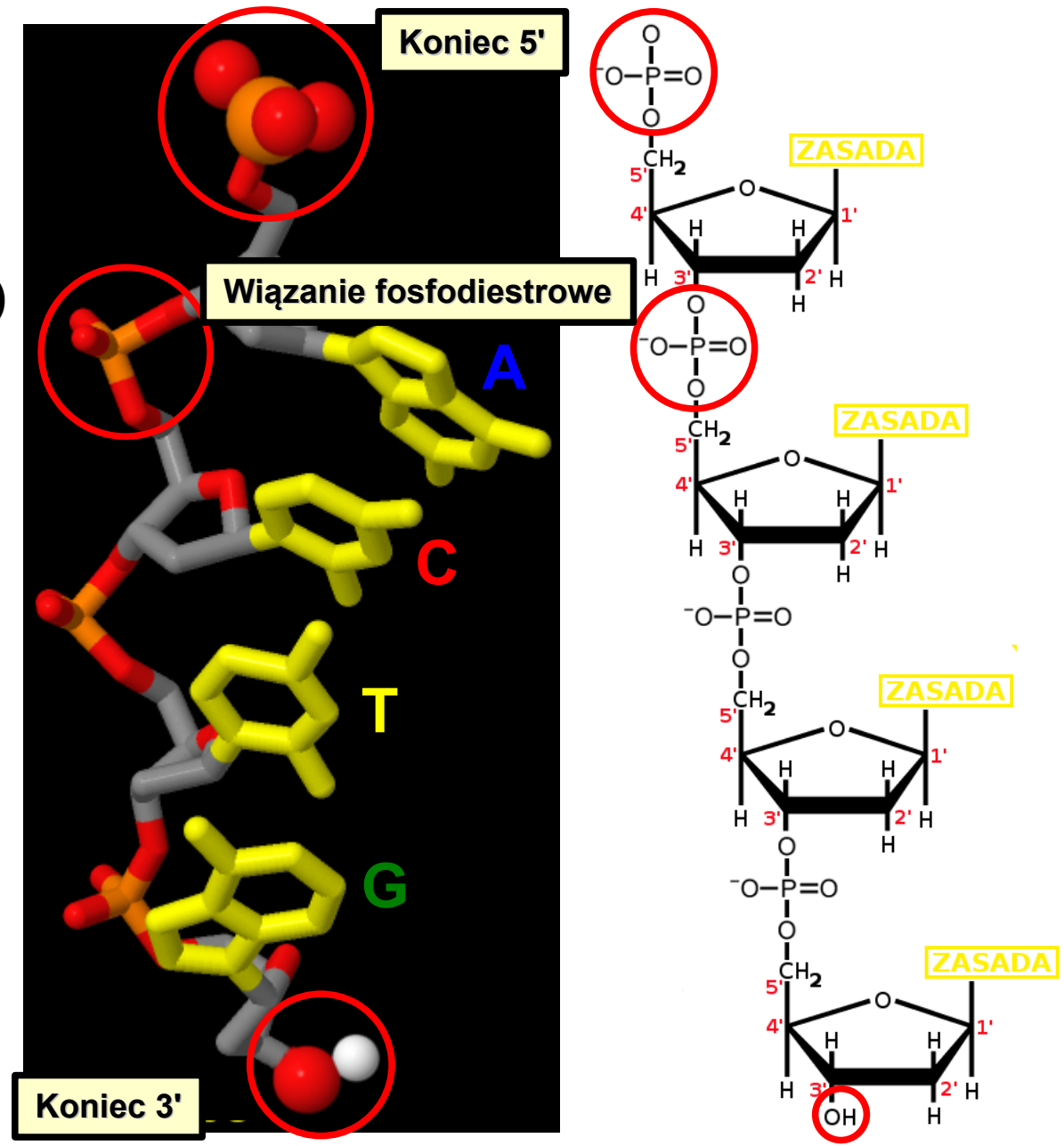
Tym samym może on mieć dwie różne orientacje:

od końca 5' do końca 3' (**5' → 3'**)

od końca 3' do końca 5' (**3' → 5'**)

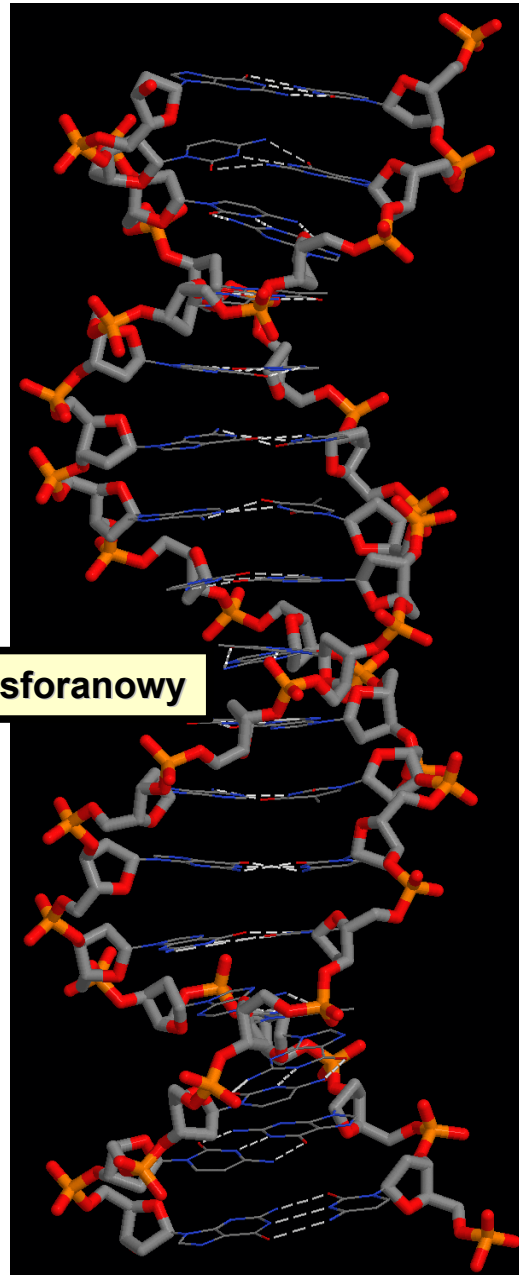
Sekwencja zasad zapisywana jest w kierunku **5' → 3'**:

(koniec 5') **ACTG** (koniec 3')

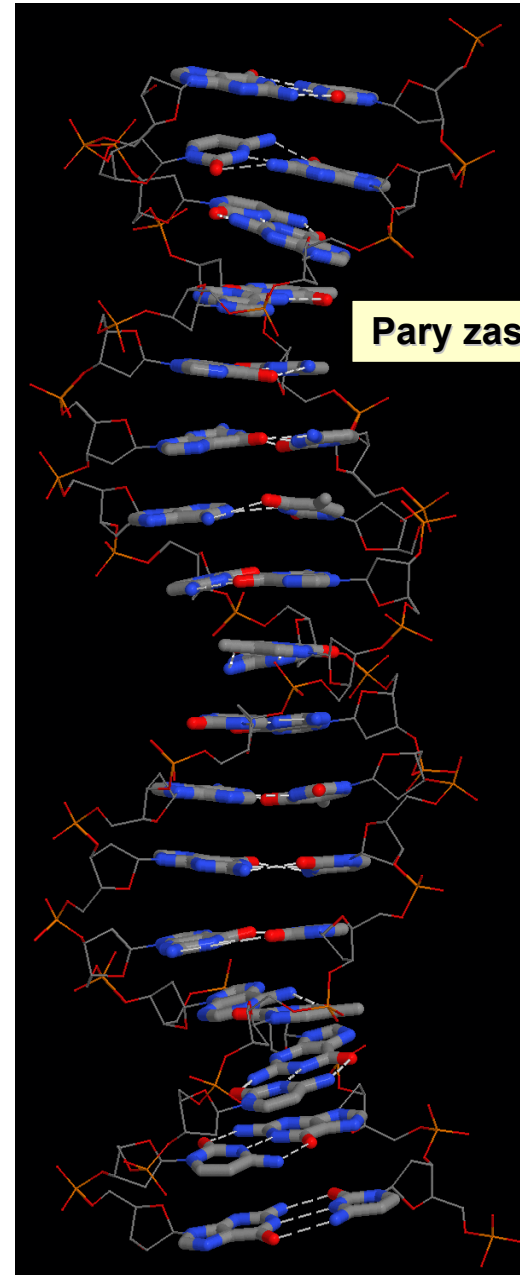


# Kwas deoksyrybonukleinowy (DNA): dwuniciowa helisa DNA

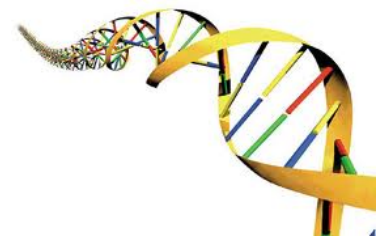
Cząstki DNA występują w postaci **podwójnej helisy**, w której nici polinukleotydowe ułożone są antyrównoległe: jedna biegnie w kierunku 5'→3', druga w kierunku 3'→5'.



Szkielet cukrowo-fosforanowy



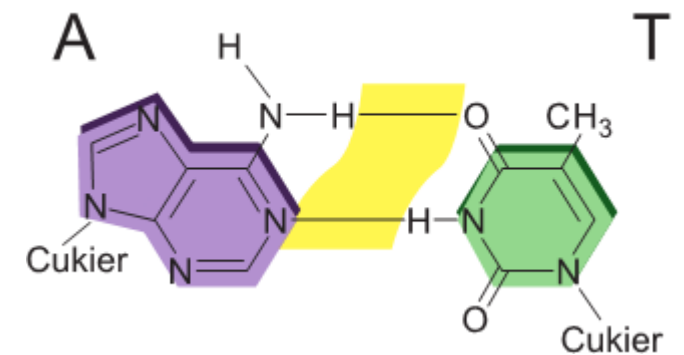
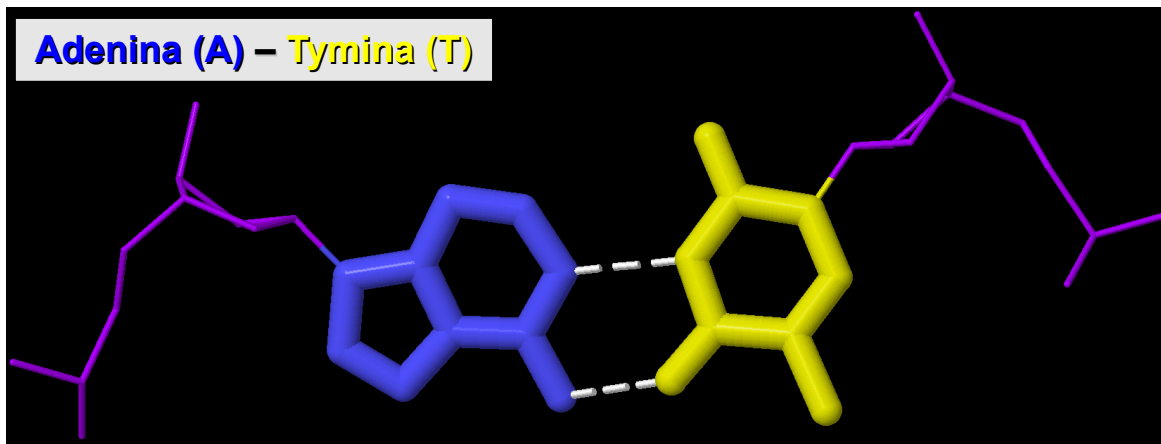
Pary zasad azotowych



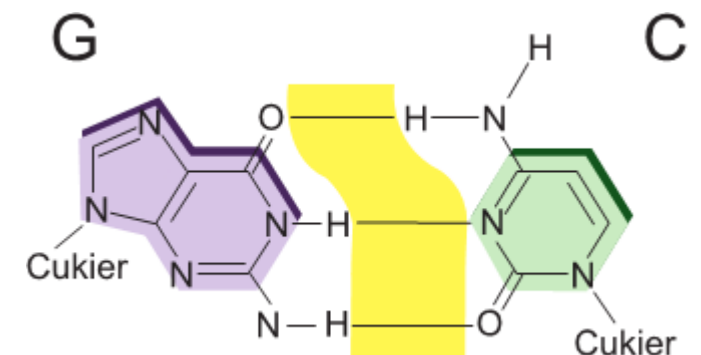
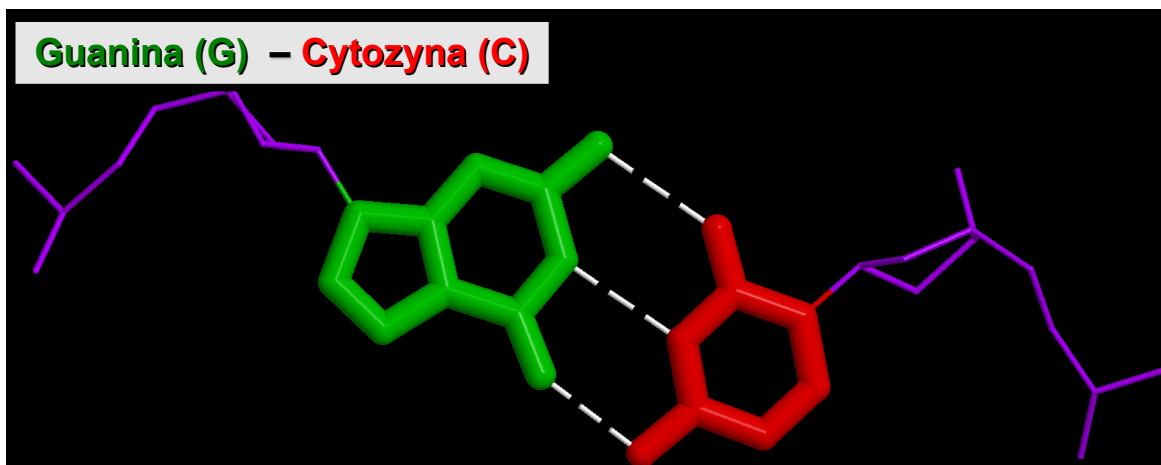
# Kwas deoksyrybonukleinowy (DNA): komplementarność zasad azotowych

Oba łańcuchy podwójnej helisy połączone są wiązaniami wodorowymi pomiędzy **komplementarnymi parami zasad azotowych**:

- **adenina łączy się z tyminą,**

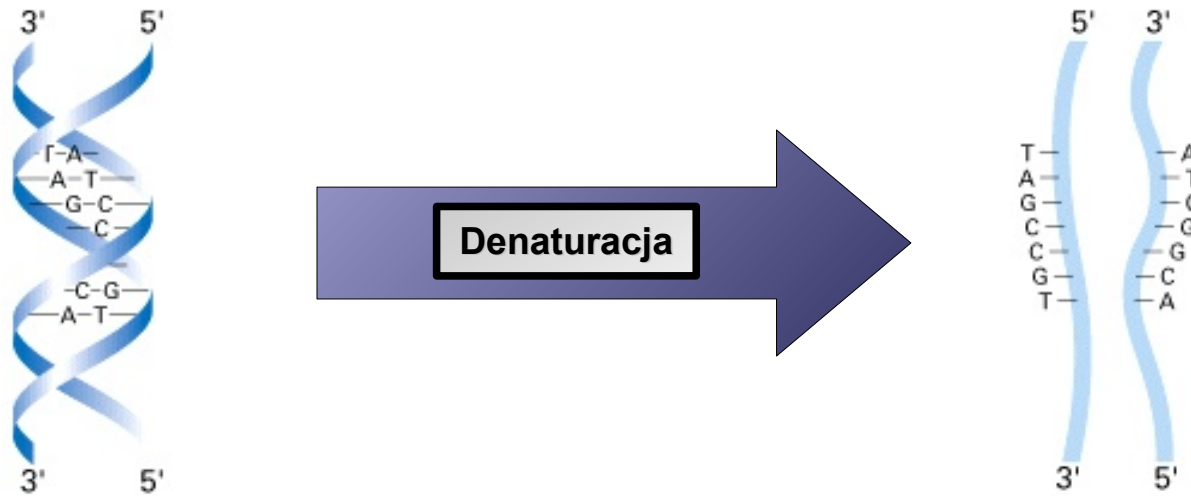


- **guanina łączy się z cytozyną.**

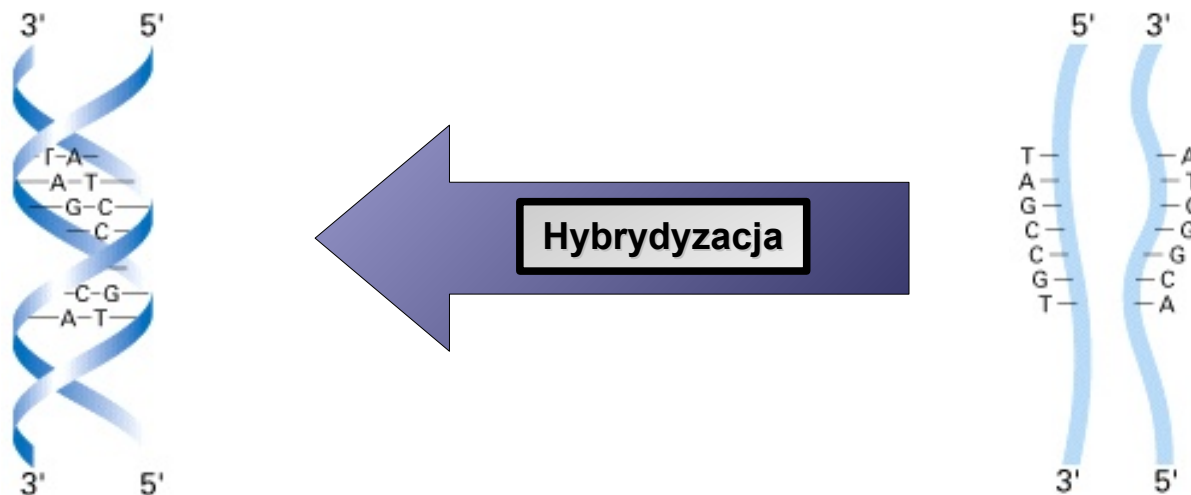


# Kwas deoksyrybonukleinowy (DNA): komplementarność zasad azotowych

Nici DNA mogą ulec rozdzieleniu (**denaturacji**) pod wpływem wysokiej temperatury (około 80-90°C) lub czynników chemicznych.



Po usunięciu czynnika denaturującego następuje **hybrydyzacja (renaturacja)**, czyli samoistne łączenie się nici, w sposób zgodny z regułami komplementarności.





# Uczenie maszynowe w bioinformatyce

## Wykład 1: kwasy nukleinowe (RNA)

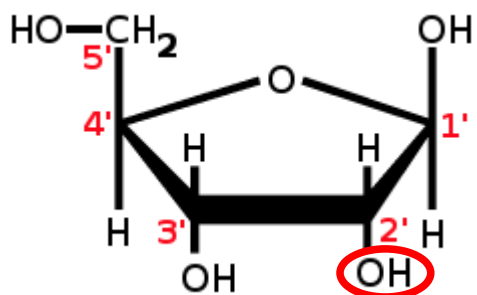
Tymon Rubel

Zakład Elektroniki Jądrowej i Medycznej  
Instytut Radioelektroniki i Techniki Multimedialnych PW

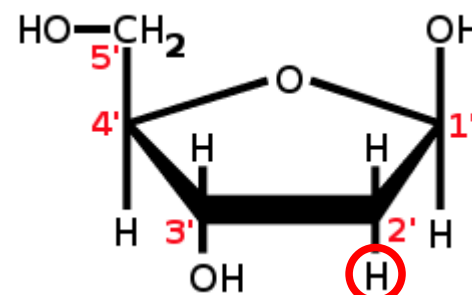
# Kwas rybonukleinowy (RNA): rybonukleotydy

**Kwas rybonukleinowy (RNA)** jest polinukleotydem o strukturze podobnej do DNA, jednak z dwoma wyjątkami:

▪ cukrem wchodzącym w skład rybonukleotydów jest **ryboza**, a nie deoksyryboza;

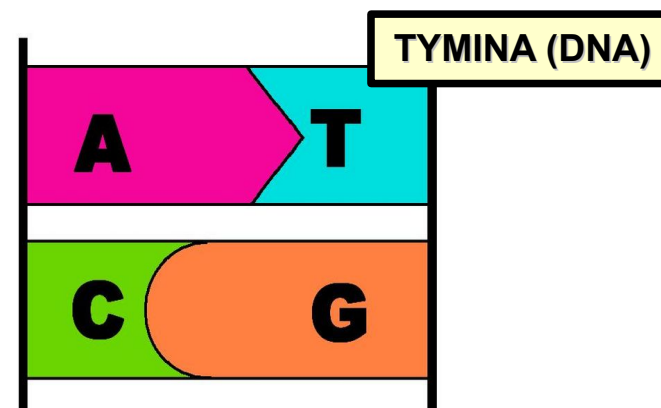
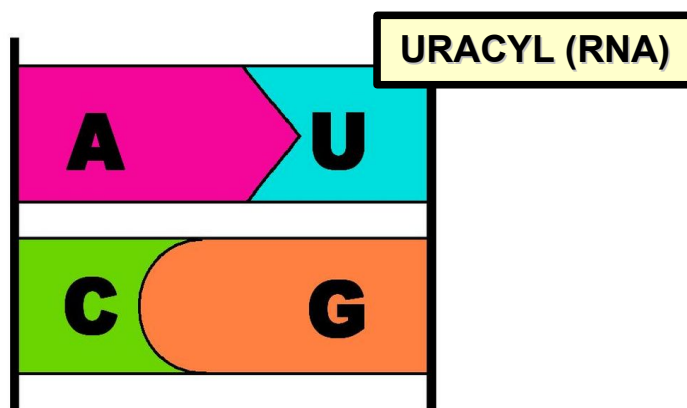


RYBOZA (RNA)



DEOKSYRYBOZA (DNA)

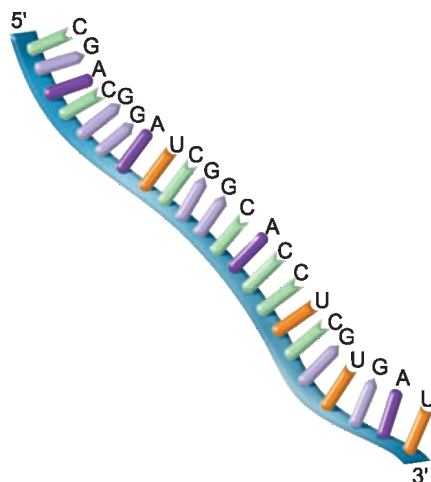
▪ występującą w DNA tyminę (T) zastępuje **uracyl (U)**, zasada pirymidynowa, która może tworzyć parę komplementarną z adeniną (A).



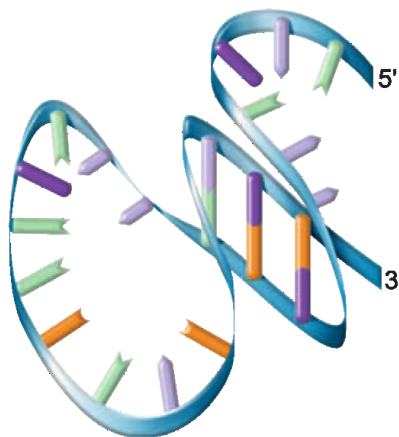
# Kwas rybonukleinowy (RNA)

Rybonukleotydy również mogą łączyć się poprzez wiązania fosfodiesterowe tworząc łańcuchy polinukleotydowe.

Cząsteczki RNA najczęściej mają postać pojedynczych łańcuchów polipeptydowych, które są mniej stabilne od podwójnej helisy DNA. Dlatego też cząstki RNA zwykle są krótsze i łatwiej ulegają degradacji.

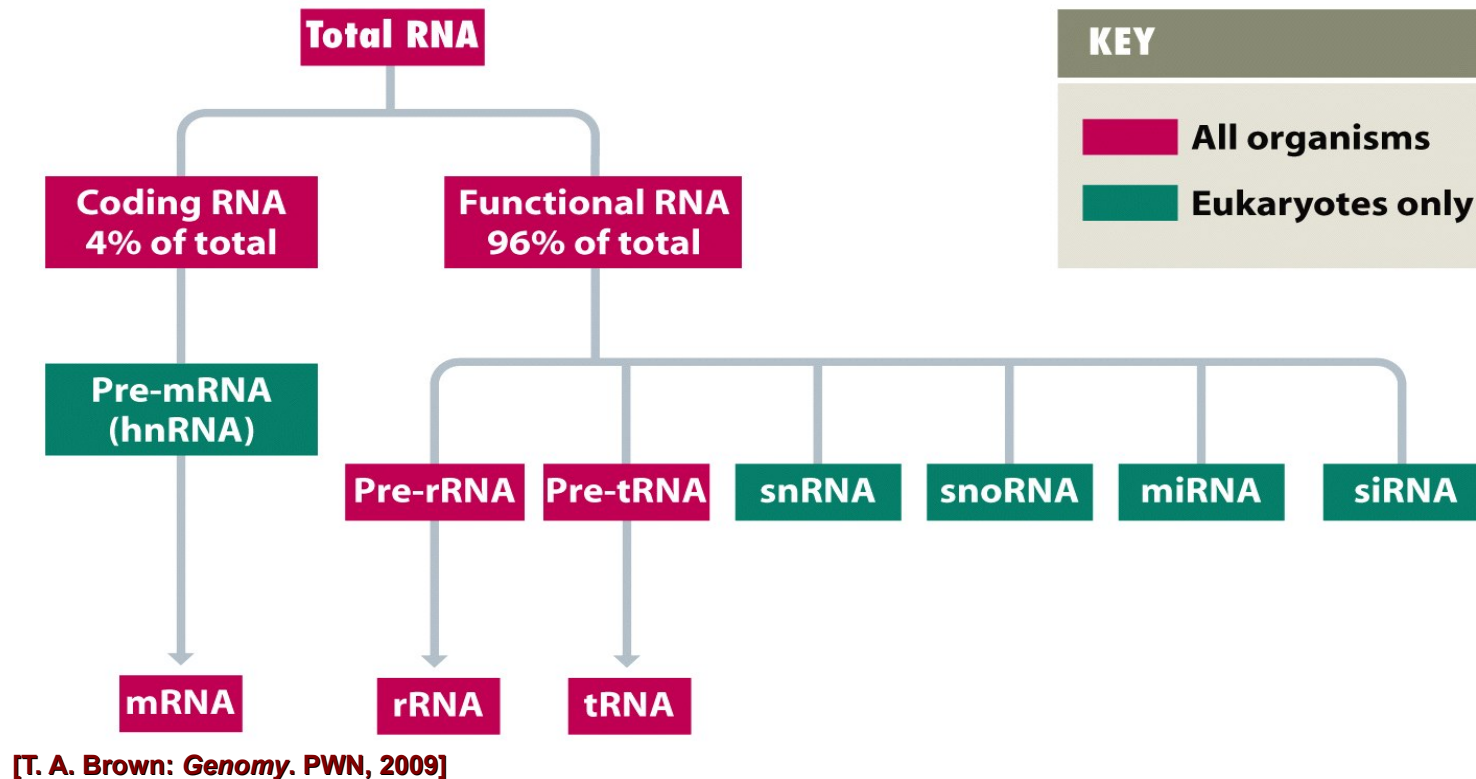


Mogą jednak posiadać strukturę przestrzenną, powstałą poprzez tworzenie regionów helikalnych pomiędzy komplementarnymi odcinkami tego samego łańcucha.



# Kwas rybonukleinowy (RNA): rola w organizmach żywych

W komórkach organizmów żywych istnieje wiele rodzajów RNA, z których większość pełni funkcje w procesie syntezy białek oraz regulacji sposobu odczytu informacji genetycznych zawartych w DNA.



U niektórych wirusów i bakteriofagów RNA (jedno- lub dwuniciowe) zastępuje DNA w roli nośnika informacji genetycznej.

# Uczenie maszynowe w bioinformatyce

## Wykład 1: białka

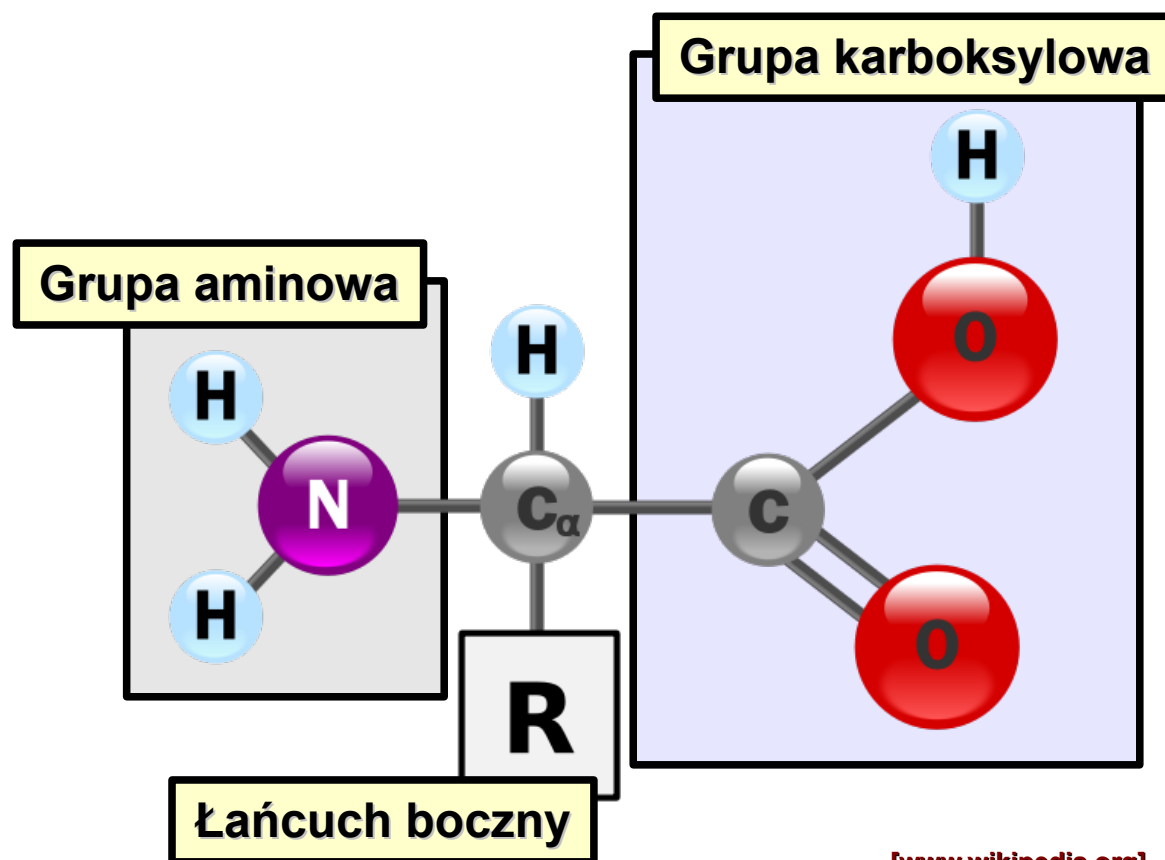
Tymon Rubel

Zakład Elektroniki Jądrowej i Medycznej  
Instytut Radioelektroniki i Techniki Multimedialnych PW

# Białka: aminokwasy

Białka są cząstkami polimerowymi, których monomerami są **aminokwasy**.

Każdy aminokwas zbudowany jest z grup karboksylowej (COOH) i aminowej (NH<sub>2</sub>), atomu wodoru (H) oraz łańcucha bocznego (R), związanych kowalencyjnie z atomem węgla, nazywanym węglem C<sub>α</sub>.



[[www.wikipedia.org](http://www.wikipedia.org)]

# Białka: aminokwasy

W skład białek wszystkich organizmów żywych wchodzi **20 rodzajów aminokwasów**.

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

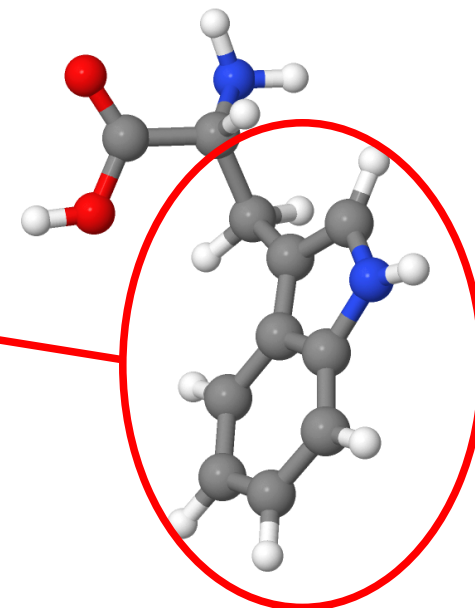
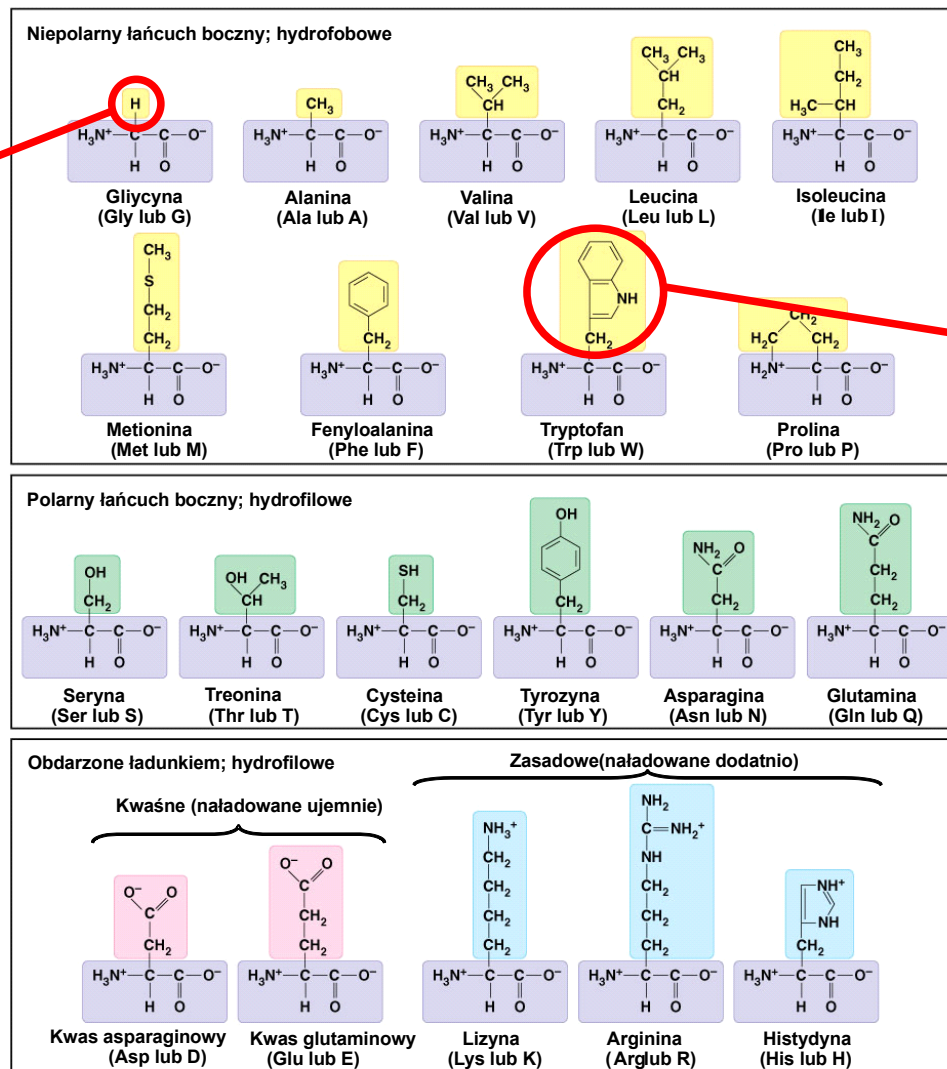
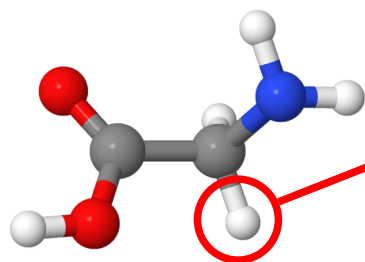
nazwa	symbol	skrót
leucyna	L	Leu
lizyna	K	Lys
metionina	M	Met
fenyloalanina	F	Phe
prolina	P	Pro
seryna	S	Ser
treonona	T	Thr
tryptofan	W	Trp
tyrozyna	Y	Tyr
walina	V	Val

Rzadziej spotykane są dwa dodatkowe aminokwasy:

- **selenocysteina (U, Sec)**, wchodzi w skład tzw. selenoprotein wielu organizmów,
- **pirolizyna (O, Pys)**, najprawdopodobniej występuje tylko w białkach archeonów.

# Białka: aminokwasy

Łańcuchy boczne aminokwasów mogą w znaczący sposób różnić się zarówno składem, jak i właściwościami chemicznymi, co przekłada się na dużą różnorodność właściwości utworzonych z nich białek.

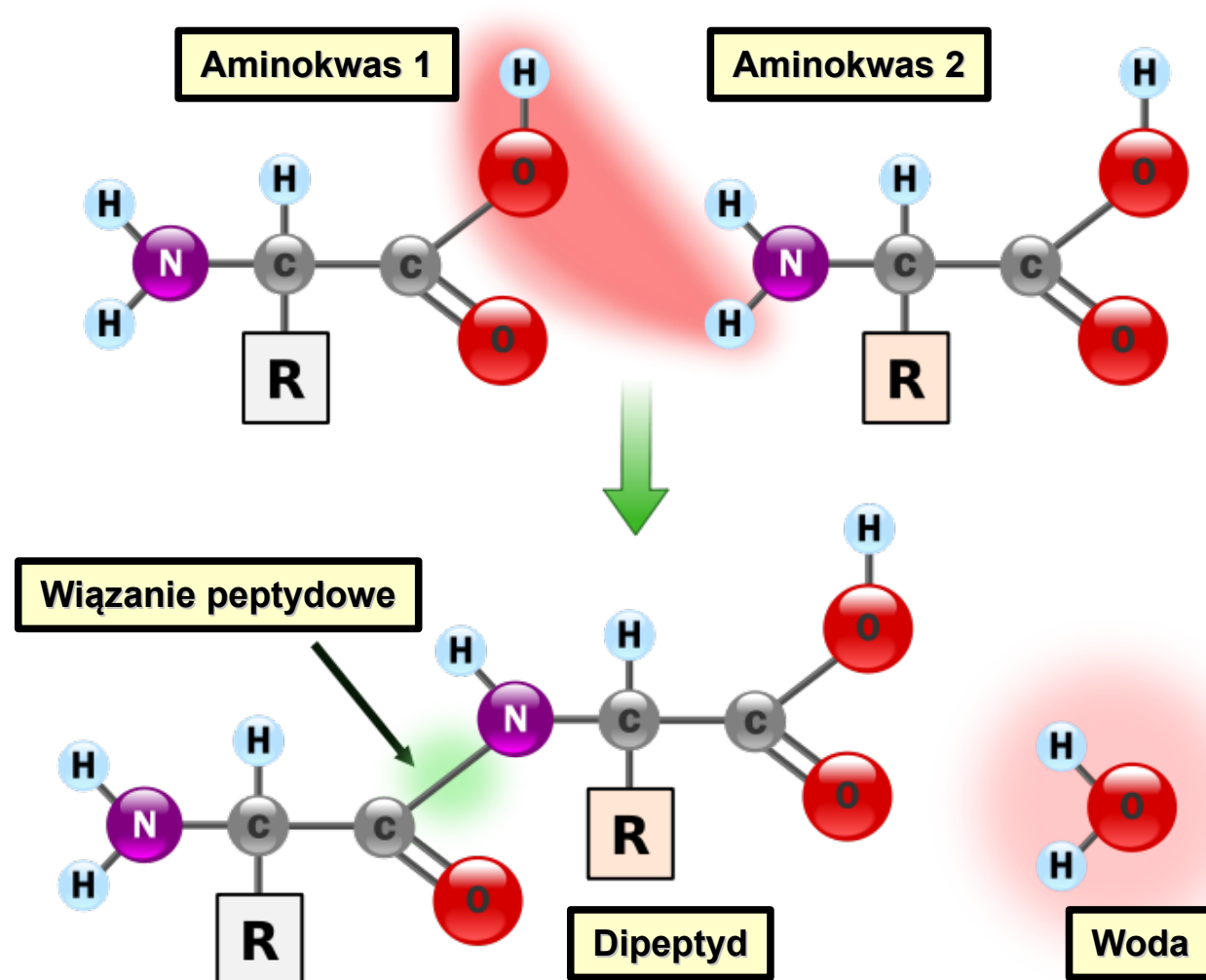




# Białka: peptydy

Podstawową cechą aminokwasów jest ich zdolność do łączenia się ze sobą za pośrednictwem **wiązania peptydowego** pomiędzy grupami aminową i karboksylową.

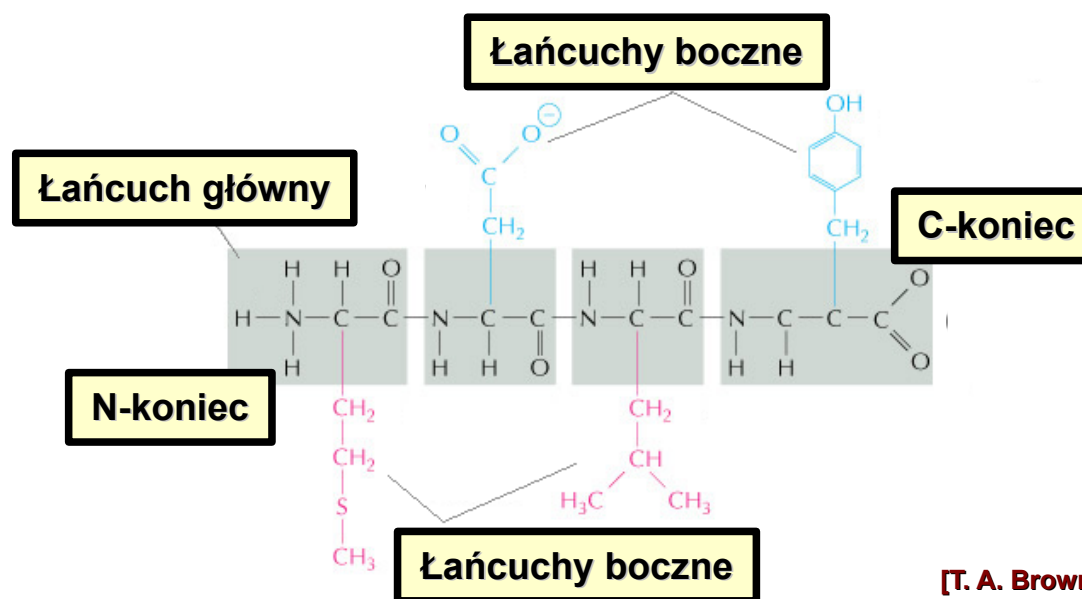
Cząstkę złożoną z dwóch lub więcej aminokwasów nazywamy **peptydem**.



# Białka: łańcuchy polipeptydowe

Wiele aminokwasów połączonych wiązaniami tworzy liniowy **łańcuch polipeptydowy**. Obecne w łańcuchu części aminokwasów nazywane są **resztami aminokwasowymi**.

Reszta aminokwasowa zawierająca wolną grupę aminową nazywana jest **N-końcem** łańcucha, podczas gdy reszta z wolną grupą karboksylową jest jego **C-końcem**.



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

Sekwencje łańcucha zapisuje się rozpoczynając od N-końca lewej strony i kończąc C-końcową resztą aminokwasową z prawej (zgodnie z kierunkiem syntezy białek).

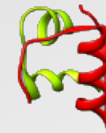
(N-końiec) M A D Y (C-końiec)

# Białka: łańcuchy polipeptydowe

**Białka** zbudowane są z jednego lub większej liczby łańcuchów polipeptydowych. Długości sekwencji typowych białek wahają się od kilkudziesięciu do nawet kilkudziesięciu tysięcy reszt aminokwasowych.

Insulina (hormon peptydowy): 51 aminokwasów

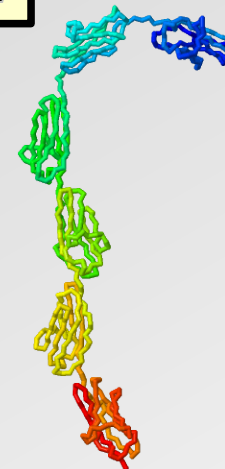
FVNQHLCGSHLVEALYLVCGERGFFYTPKTGIVEQCCTSI~~CS~~LYQLENYCN



Tytyna (białko mięśni poprzecznie prążkowanych): 34 350 aminokwasów

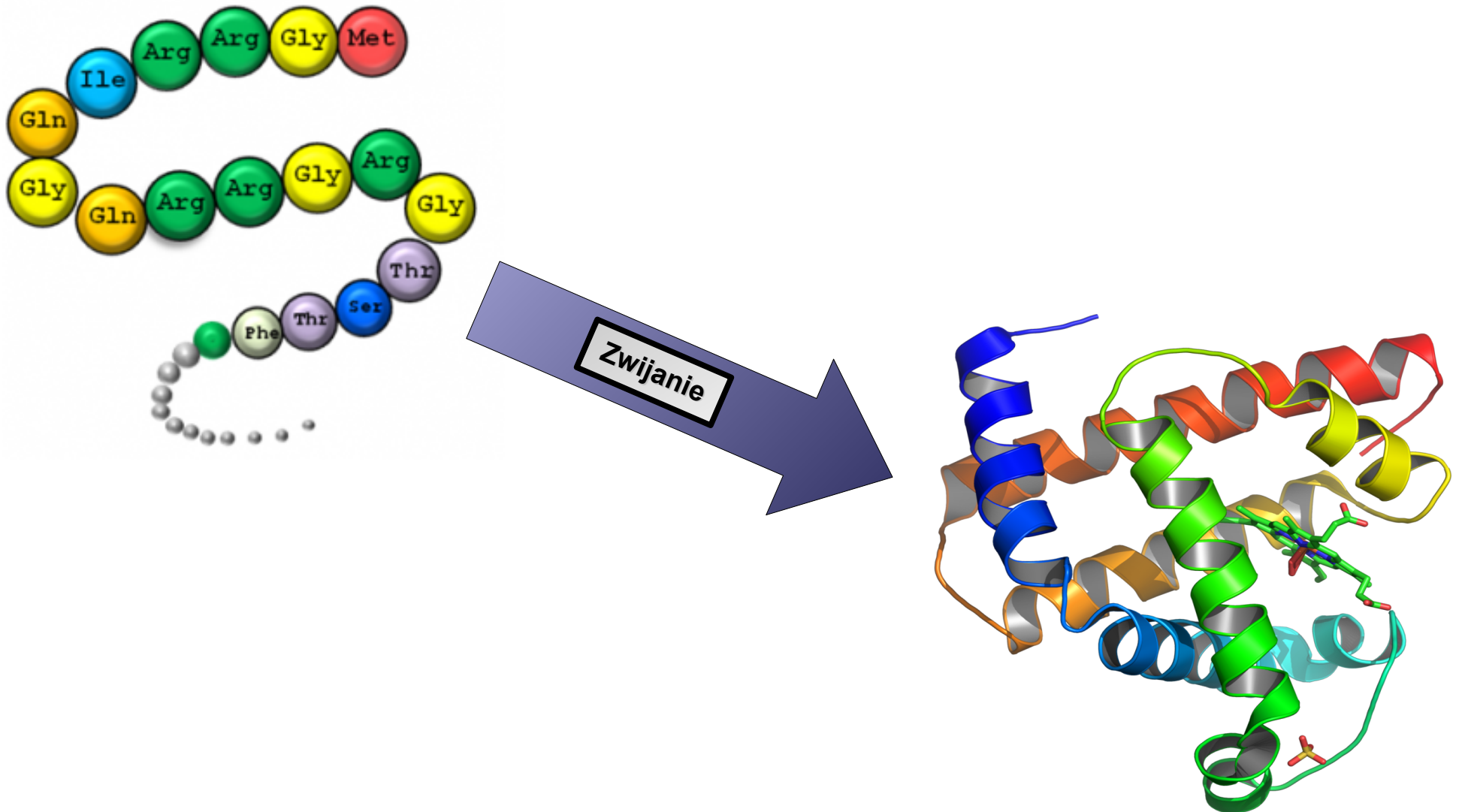
MTTQAPTFTQPLQSVVVLEGSTATFEAHISGFPVPEVSWFRDQVISTSTLPGVQISFSD  
GRAKLTIPAVTKANSGRYSLKATNGSGQATSTAELLVKAETAPPNFVQRLQSMTVRQGSQ  
VRLQVRVTGIPTPVVKFYRDGAEIQSSLDFQISQEGDLYSLLIAEAYPEDSGTYSVNATN  
SVGRATSTAELLVQGE~~EE~~VPA...

...LRTSGDTSLQGSFSSQSVQMSASKQEASFSSFSSSSAS  
SMTEMKFASMSAQSMSSMQESFVEMSSSSFMGISNMTQLESSTKMLKAGIRGIPPKIEA  
LPSDISIDEGKVLTVACAFTGEPTPEVTWSCGGRKIH~~SQ~~EQGRFHIENTDDLTTLIIMDV  
QKQDGGLYTL~~SL~~SGNEFGSDSATVNIHIRSI



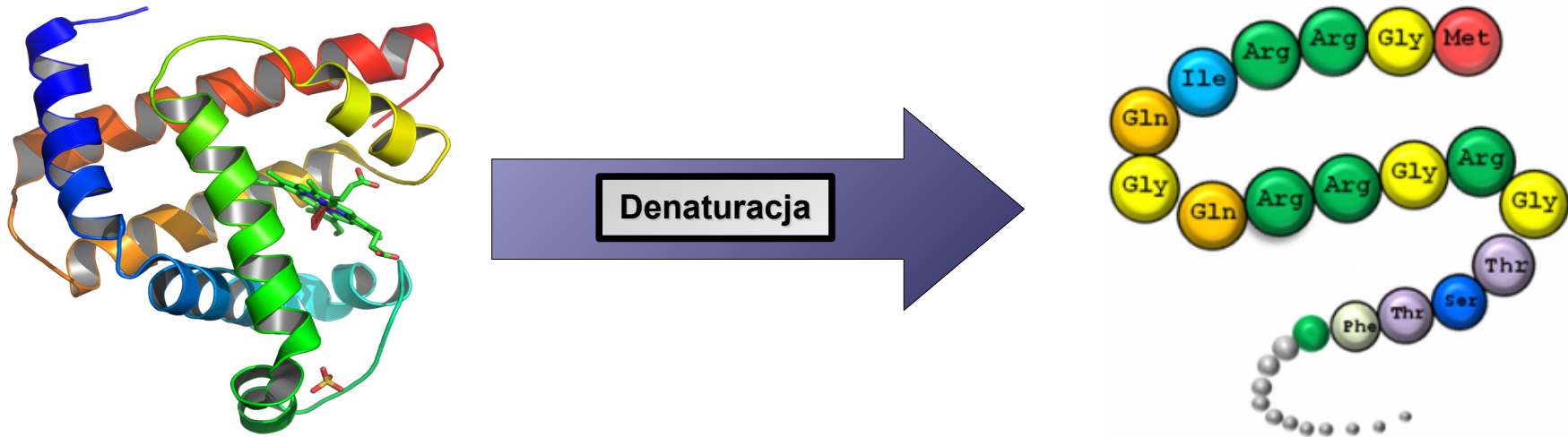
# Białka: struktura przestrzenna (konformacja)

Rozciągnięty lub ułożony przypadkowo łańcuch polipeptydowy zwykle pozbawiony jest aktywności biologicznej. Aktywność ta pojawia się dopiero wraz z odpowiednią **konformacją**, czyli przestrzennym ułożeniem łańcucha. Przyjmowanie właściwej struktury przestrzennej nazywane jest **zwijaniem** lub **fałdowaniem (*fold*ing)** białka.

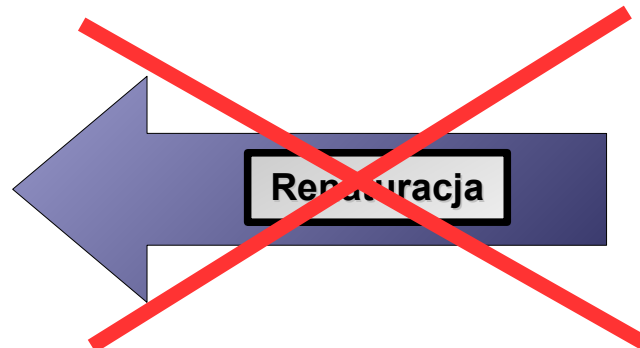
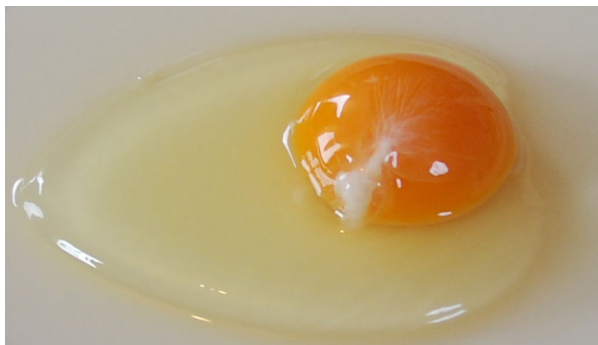


# Białka: struktura przestrzenna (konformacja)

**Właściwości biochemiczne białek ściśle zależą od ich form przestrzennych.** Dlatego **denaturacja**, czyli zniszczenie trójwymiarowej struktury skutkuje utratą aktywności biologicznej. Może ona nastąpić np. przez podgrzanie do około 50°C.

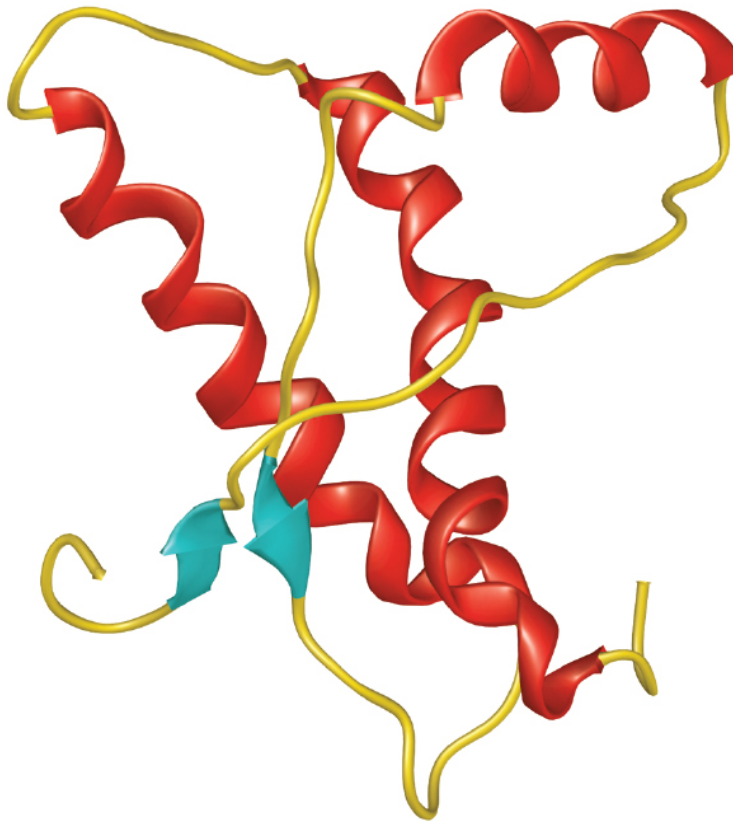


Inaczej niż w przypadku DNA, denaturacja białek najczęściej jest nieodwracalna.

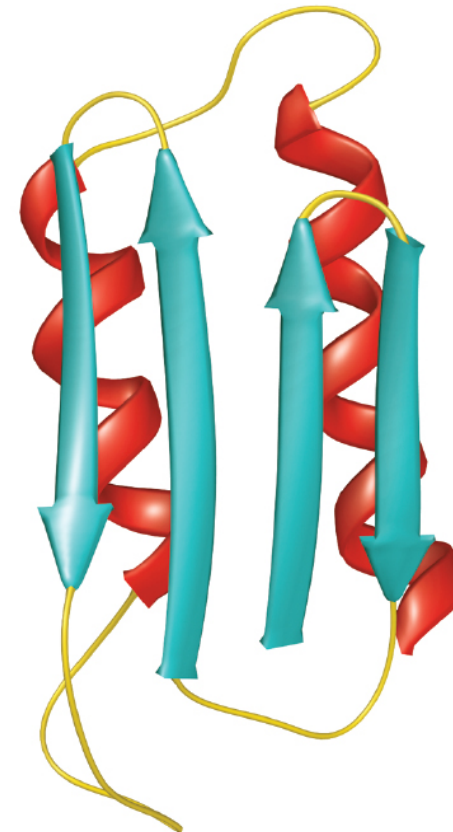


# Białka: struktura przestrzenna (konformacja)

**Zmiana konformacji prowadzi do powstania białka o odmiennych właściwościach, często szkodliwych dla organizmu:** przykładami mogą być priony odpowiedzialne za chorobę Creutzfeldta-Jakoba (CJD) oraz agregaty nieprawidłowych form białkowych obecne w wielu schorzeniach neurodegeneracyjnych (np. chorobie Alzheimera).



Prawidłowa konformacja



Nieprawidłowa konformacja (prion CJD)

# Białka: predykcja konformacji białek



Folding@home  
distributed computing

folding@home: <https://foldingathome.org/>

Rosetta@home: <http://boinc.bakerlab.org/>

[Home](#) [Download](#) [Guides](#) [FAQ](#) [Stats](#) [Science](#) [Results](#) [Awards](#) [About Us](#)

[Main](#)  
[News](#)  
[Forum \(Help\)](#)  
[Pande group](#)  
[Donate](#)

العربية



Dansk

Deutsch

English

Español

Français

Italiano

Ελληνικά



Lietuvių

Magyar

Nederlands

Norsk

Occitan

فارسی

Polski

Português

Русский

Suomeksi

## Our goal: to understand protein folding, misfolding, and related diseases

**You can help scientists studying these diseases by simply running a piece of software.**

Folding@home is a distributed computing project -- people from throughout the world [download](#) and run software to band together to make one of the largest supercomputers in the world. Every computer takes the project closer to our goals. Folding@home uses novel computational methods coupled to distributed computing, to simulate problems millions of times more challenging than previously achieved.

[Download Folding@home](#)



**Protein folding is linked to disease, such as Alzheimer's, ALS, Huntington's, Parkinson's disease, and many Cancers.**

Moreover, when proteins do not fold correctly (i.e. "misfold"), there can be serious consequences, including many well known [diseases](#), such as Alzheimer's, Mad Cow (BSE), CJD, ALS, Huntington's, Parkinson's disease, and many Cancers and cancer-related syndromes.

### What is protein folding?

Proteins are biology's workhorses -- its "nanomachines." Before proteins can carry out these important functions, they assemble themselves, or "fold." The process of protein folding, while critical and fundamental to virtually all of biology, in many ways remains a mystery.

### What have we done so far?

We have had several successes. You can read about them on our [Science](#) page, on our [Awards](#) page, or go directly to our [Results](#) page.



### Want to learn more?

Click on the links on the left for downloads or more information. You can also download our [Executive Summary](#), which is a PDF suitable for distribution. One can also help by [donating funds to the project](#), via Stanford University.

Make a gift now!

[Click here to make a gift >](#)

## AlphaFold Protein Structure Database

Developed by DeepMind and EMBL-EBI

Search for protein, gene, UniProt accession or organism

BETA

Search

Examples: [Free fatty acid receptor 2](#) [At1g58602](#) [Q5VSL9](#) [E. coli](#) Help: [AlphaFold DB search help](#)

Feedback on structure: [Contact DeepMind](#)

AlphaFold DB provides open access to protein structure predictions for the human proteome and other key proteins of interest, to accelerate scientific research.

### Background

AlphaFold is an AI system developed by DeepMind that predicts a protein's 3D structure from its amino acid sequence. It regularly achieves accuracy competitive with



AlphaFold: <https://alphafold.ebi.ac.uk/>

<https://github.com/deepmind/alphafold>



# Białka: rola w organizmach żywych

Białka odgrywają kluczową rolę w praktycznie wszystkich procesach biologicznych, zarówno w skali pojedynczej komórki, jak i całego organizmu.

Przykładowe funkcje białek:

- **katalityczne** (prawie wszystkie znane enzymy są białkami);
- **regulacyjne** (część hormonów jest białkami lub peptydami, np. insulina);
- **strukturalne i mechaniczne** (np. miozyna wchodząca w skład mięśni);
- **transportowe** (np. hemoglobina, której podstawową funkcją jest transport tlenu);
- **magazynujące** (np. ferrytyna odpowiedzialna za magazynowanie żelaza w wątrobie);
- **immunologiczne** (wszystkie przeciwciała są białkami);
- **kontrola wzrostu i różnicowania.**