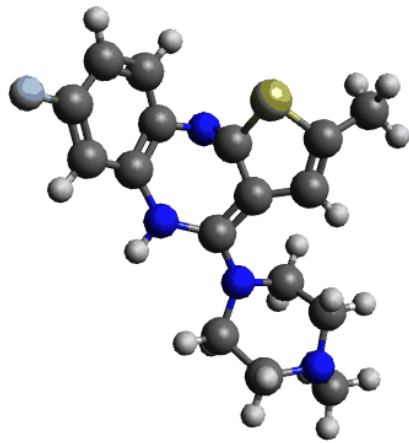


# Modelowanie Cząsteczkowe

Oto anegdota o profesorze Agassiz i o rybie.

Młody uczyony, zbrojny w dyplom i wyróżnienia, zgłosił się do prof. Agassiz po ostatnie, kończące okres nauki wskazówki. Profesor podał mu niewielką rybę i kazał opisać. Młody uczyony: „Przecież to samogłów”. Profesor: „Wiem. Lecz opisz rybę”. Po chwili młody człowiek wrócił z opisem, w którym pospolitego samogłowa ukryto pod pospolitą wiedzą za pomocą terminu *Ichthius Helioplodocus*. Opis mówił o przynależności do rodziny *Heliichtherinkus* itd., według przyjętych w podręcznikach zoologii schematów. Profesor znów poprosił o opis ryby. Młody uczyony przyniósł czterostronicową rozprawkę. Wtedy Profesor kazał przyglądać się rybie. Po trzech tygodniach samogłów znajdował się w stanie zaawansowanego rozkładu, ale badacz coś już o nim wiedział.



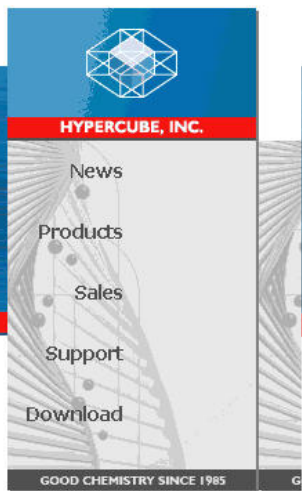
# **METODY MODELOWANIA CZĄSTECZKOWEGO**

## **Cel nauczania:**

**Zapoznanie studentów z wykorzystaniem metod modelowania cząsteczkowego w projektowaniu nowych leków. Omówienie wykorzystania metod komputerowych w wyznaczaniu zależności między budową leku a jego działaniem biologicznym.**

## **Program nauczania:**

- 1. Wprowadzenie do modelowania cząsteczkowego poprzez omówienie metod mechaniki molekularnej, metod półempirycznych i metod ab initio.**
- 2. Wykorzystanie metod modelowania cząsteczkowego do wyznaczania właściwości fizykochemicznych związków biologicznie aktywnych i określania zależności między budową leku a jego działaniem biologicznym (QSAR).**
- 3. Określanie aktywnych konformacji leków przy wykorzystaniu analizy konformacyjnej i dynamiki molekularnej.**
- 4. Poszukiwanie receptorów biologicznych, wyznaczanie budowy trójwymiarowego farmakoforu i struktur wiodących nowych leków w elektronicznych bazach danych.**
- 5. Projektowanie leków z uwzględnieniem budowy miejsca wiążącego receptora biologicznego.**



# Welcome to Hypercube, Inc.



## New HyperChem for Linux 1.0 Release Now Available

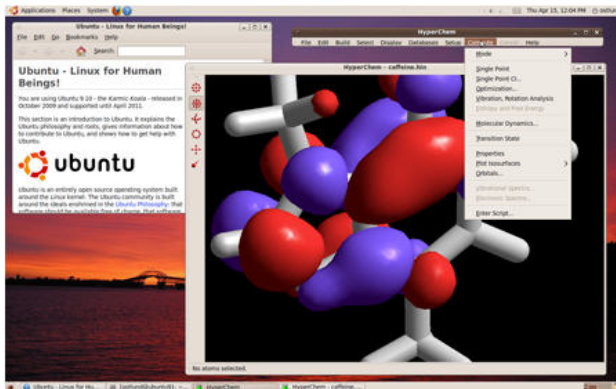
The Linux version of HyperChem is now available. Get 10% off until July 1.

This new product now means that HyperChem is available for all three major platforms - Windows, Mac, and Linux. Download a 10-day evaluation version and try it out. [HyperChem for Linux 1.0.](#)

## New HyperChem 8.0.8 Release Now Available

A new maintenance update of HyperChem 8.0 Professional is now available.

This new release is fully Vista and Windows 7 compliant and includes changes to Ray Tracing, HyperNMR, and Licensing to accommodate 64-bit versions of Windows. It also includes new client-server



## HyperChem for Linux is Shipping!

The first Release 1.0 of our latest product, HyperChem for Linux, is now available for download. HyperChem is supporting Linux with new Client-Server options. Now any of Windows, Mac, or Linux clients can farm out computations to Linux servers. Linux Clusters are coming! Read all about it [here](#).

## Facebook and Twitter

Hypercube is pleased to announce that it is now part of the social networking revolution that is Facebook and Twitter. Click [here](#) to visit our twitter page and to follow us, and click [here](#) to visit our facebook page and to become our fan.

## Enjoy our New Look and Feel

Hypercube is pleased to introduce a new design for its web site. Hopefully you will find it a considerable improvement. We expect it will be easier for our customers and friends to both find things and to interact more directly with us. Send comments to us if you find errors or have suggestions for further improvements. Since this new site is still somewhat under construction, we will keep our old site active for a short period of time. It is available at <http://www.computationalchemistry.com/>.

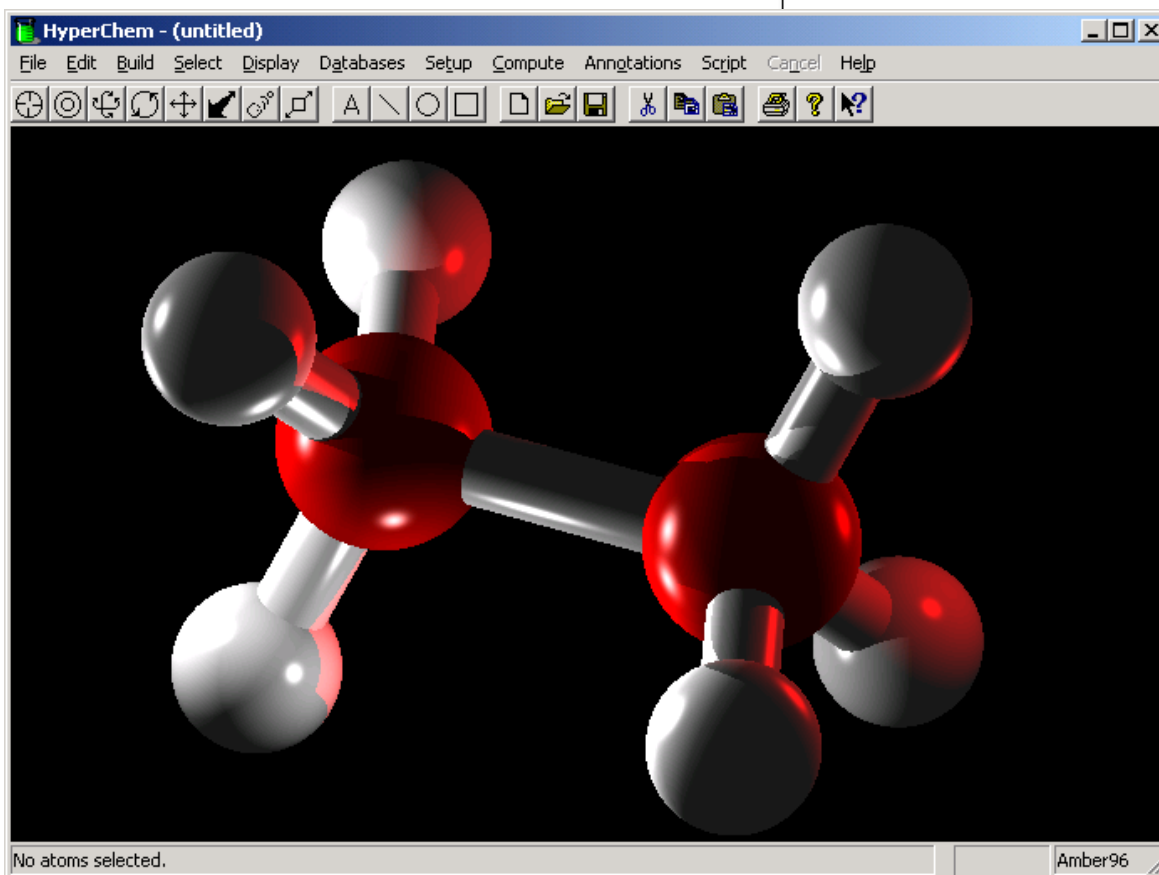
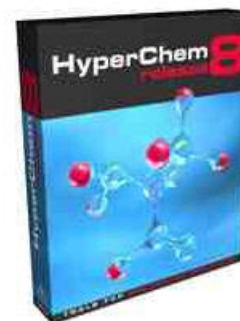
New Release of HyperChem now available **8**

HyperChem for Mac

## HyperChem Professional 8.0

HyperChem is a sophisticated molecular modeling environment that is known for its quality, flexibility, and ease of use. Uniting 3D visualization and animation with quantum chemical calculations, molecular mechanics, and dynamics, HyperChem puts more molecular modeling tools at your fingertips than any other Windows program.

Our newest version, HyperChem Release 8.0, is a full 32-bit application, developed for the Windows 95, 98, NT, ME, 2000, XP, and Vista operating systems. HyperChem Release 8.0 incorporates even more powerful computational chemistry tools than ever before, as well as supporting multiple third-party applications. Its drawing and rendering capabilities and ease of use are standards for the industry.





## Potential Applications

Quantum Chemistry  
Materials Science  
Teaching Visualization  
Drug Design

Extensions

Tools

Rendering  
Display

Colors

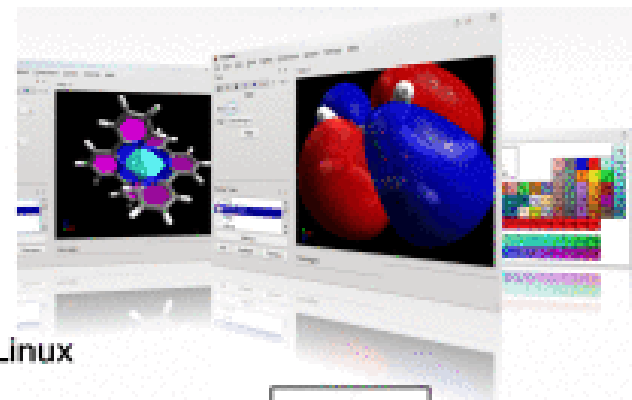
Scripting



# Avogadro

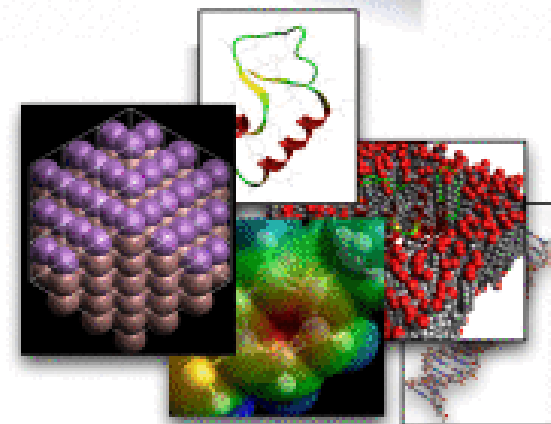
## Features

Intuitive "Drawing"  
Fast Optimization  
Results + Analysis  
20+ Languages  
Windows + Mac + Linux



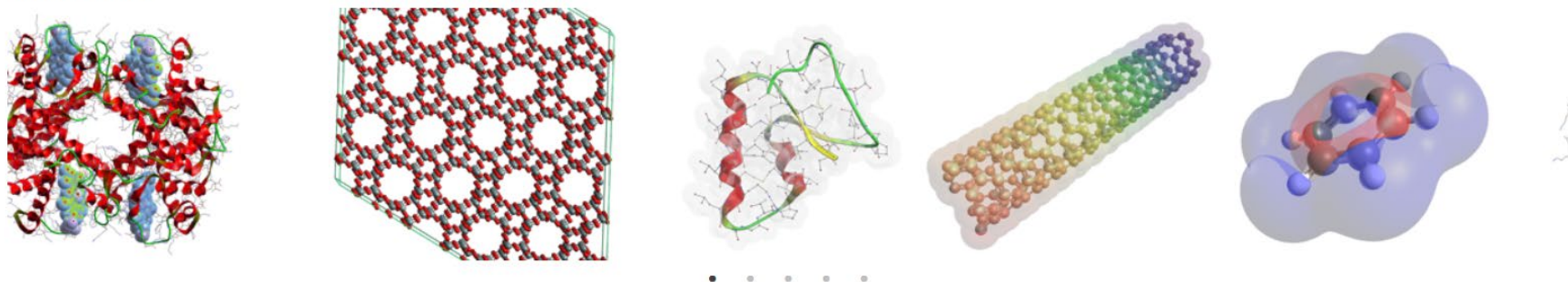
## Extendability

C++ Plugins  
Python Scripting  
Open Babel library  
Input Generation for  
simulation packages





Avogadro is an advanced molecule editor and visualizer designed for cross-platform use in computational chemistry, molecular modeling, bioinformatics, materials science, and related areas. It offers flexible high quality rendering and a powerful plugin architecture.



- **Cross-Platform:** Molecular builder/editor for Windows, Linux, and Mac OS X.
- **Free, Open Source:** Easy to install and all source code and documentation is [available to modify or extend](#).
- **International:** Translations into Chinese, French, German, Italian, Russian, Spanish, and others, with [more languages to come](#).
- **Intuitive:** Built to work easily for students and advanced researchers both.
- **Fast:** Supports multi-threaded rendering and computation.
- **Extensible:** Plugin architecture for developers, including rendering, interactive tools, commands, and Python scripts.
- **Flexible:** Features include [Open Babel](#) import of chemical files, input generation for multiple computational chemistry packages, crystallography, and biomolecules.

[Avogadro - Free cross-platform molecular editor - Avogadro](https://avogadro.cc)  
<https://avogadro.cc>

PREFACE

GETTING STARTED

Introduction  
**Drawing Molecules**  
Making Selections

TOOLS

Draw Tool  
Navigate Tool  
Bond-Centric Manipulate Tool  
Manipulate Tool  
Selection Tool  
Auto-Rotate Tool  
Auto-Optimize Tool  
Measure Tool  
Align Tool

MENUS

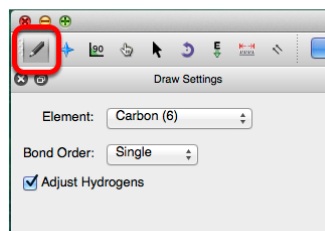
File Menu  
Edit Menu  
View Menu  
Build Menu  
Select Menu  
Extension Menu

BUILDING MOLECULES

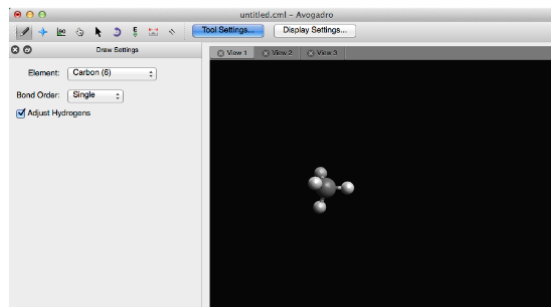
Importing Molecules  
Importing from PDB

## Drawing Molecules

Molecules are built and edited with the draw tool.



Left clicking on the black display will allow you to begin your journey into molecule creation. A left click will generate a carbon atom. A right click will delete the atom.



<https://youtu.be/WZWgoQYV1hs>

<https://youtu.be/UhSW1UugW8Q>

<https://youtu.be/fA3odPmkV-U>

<https://youtu.be/HrbahnykY6g>

<https://youtu.be/JJuho6ym14>

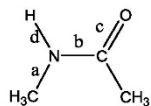
<https://youtu.be/mQtwTyclBMs>

<https://youtu.be/9JfjFGMGr-s>

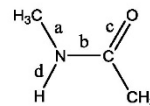
## Modelowanie cząsteczkowe (1)

Data

Porównanie trwałości i budowy geometrycznej cis- i trans- N-metyloacetamidu.



cis



trans

Ćwiczenie obejmuje narysowanie, optymalizację budowy geometrycznej oraz porównanie trwałości cis- i trans- N-metyloacetamidu przy użyciu metod półempirycznych (PM3) i mechaniki molekularnej (pole siłowe MM+).

Energia	PM3 (kcal/mol)	MM+ (kcal/mol)	Wartości doświadczalne
$E_{cis}$			X
$E_{trans}$			X
$E_{cis}-E_{trans}$			2,07

	cis PM3	cis MM+	wartości doświadczalne	trans PM3	trans MM+	wartości doświadczalne
dł. N-CH <sub>3</sub> (a)			1,445			1,447
dł. N-CO (b)			1,357			1,353
dł. C=O (c)			1,200			1,200
kąt H-N-CH <sub>3</sub> (d, a)			114,0			121,5
kąt H-N-CO (d, b)			127,4			119,4
kąt H-N-C=O (d, b, c)			0,0			180,0

Osoby:





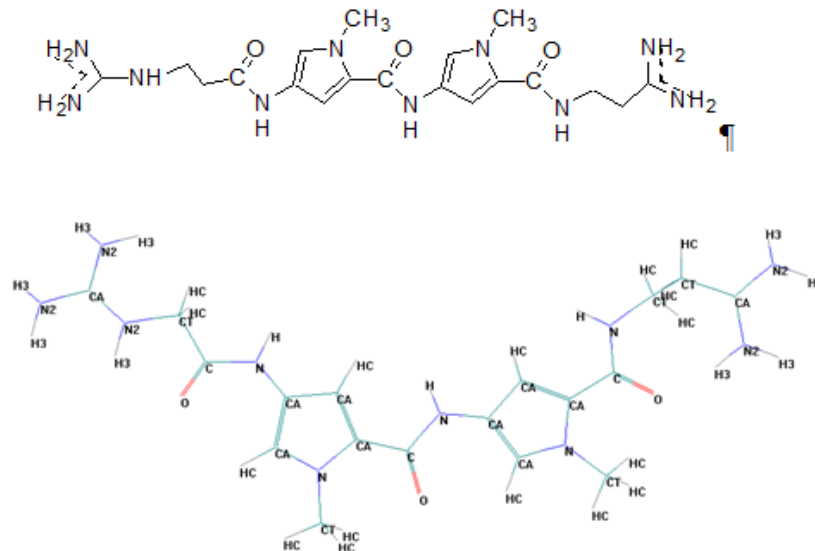
## Modelowanie cząsteczkowe (3)

### Kompleks netropsyna-DNA



Otwórz plik 101d.pdb przedstawiający kompleks antybiotyku netropsyny z dwuniciowym fragmentem DNA o sekwencji nukleotydowej  $d(\text{CGCGAATTCGCG})_2$ . Celem ćwiczenia jest wyznaczenie energii tego kompleksu. W tym celu:

1. Narysuj strukturę netropsyny (całkowity ładunek związku wynosi +2) w konformacji przedstawionej poniżej. Wyznacz cząsteczkowe ładunki atomowe poprzez optymalizację geometrii tej konformacji netropsyny w polu siłowym AM1. Następnie w polu siłowym AMBER przy parametrach Amber3 dokonaj parametryzacji poszczególnych atomów zgodnie z oznaczeniami na poniższym rysunku.



Wyznacz energię netropsyny ( $E_{\text{net}}$ ) w polu siłowym AMBER. Zapisz plik jako netr.hin.

2. Zbuduj w polu siłowym AMBER przy parametrach Amber3 fragment B-DNA o sekwencji  $d(\text{CGCGAATTCGCG})_2$ . Po dodaniu atomów sodu i optymalizacji jego geometrii wyznacz jego energię ( $E_{\text{DNA}}$ ).

3. Wprowadź cząsteczkę netropsyny z pliku netr.hin do wyznaczonego fragmentu B-DNA w obszar AATT (analogicznie jak w kompleksie 101d.pdb). Po optymalizacji geometrii kompleksu w polu siłowym AMBER wyznacz jego energię ( $E_{\text{kom}}$ ).

4. Określ energię wiązania netropsyny ( $E_w$ ):

$$E_{\text{net}} = \dots \rightarrow E_{\text{DNA}} = \dots \rightarrow E_{\text{kom}} = \dots$$

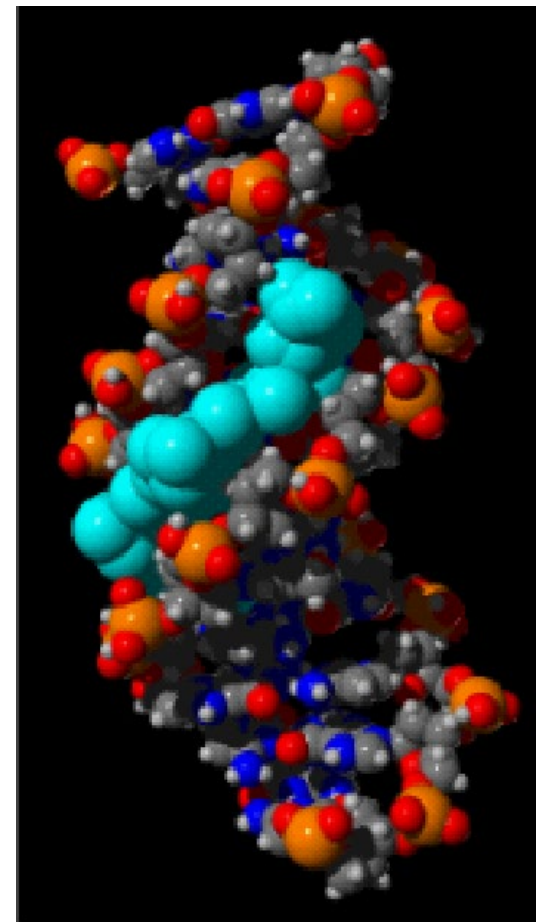
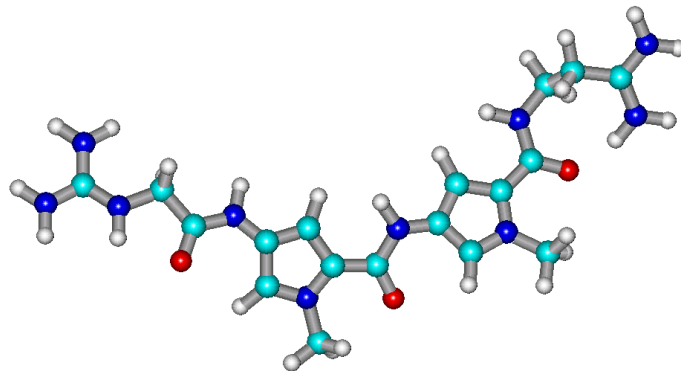
$$E_w = E_{\text{kom}} - E_{\text{DNA}} - E_{\text{net}}$$

¶

¶

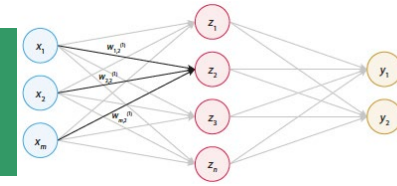
¶

Studenci ¶





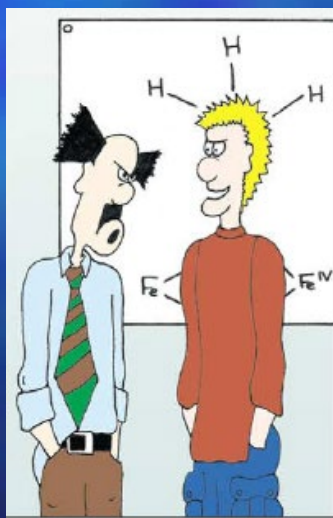
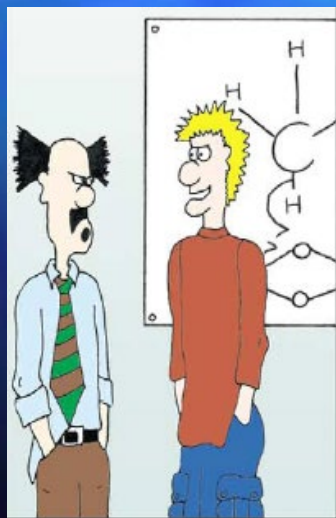
# Bioinformatyka



- ◆ Nauka zajmująca się przepływem informacji w układach biologicznych
- ◆ Bioinformatyka zajmuje się tworzeniem i wykorzystaniem biologicznych baz danych.
- Łączy osiągnięcia biologii molekularnej i genetyki z metodami statystycznymi i osiągnięciami technologii informatycznej.

# Bioinformatyka strukturalna

Dział bioinformatyki zajmujący się organizacją, przechowywaniem oraz analizą informacji strukturalnych białek, kwasów nukleinowych oraz ich kompleksów z różnymi ligandami w celu zrozumienia układów biologicznych na poziomie atomowym.





# Wprowadzenie:

- 1954: wyznaczenie struktury białka (insulina - zespół Sanger)
- 1958: pierwsza 3D struktura (X-ray) białka (mioglobina - Kendrew)
- 1972: pierwsze sekwencje DNA
- 1977: opracowanie szybkiej metody sekwencjowania DNA i białek (Gilbert i Sanger)
- 1986: PCR (szybka metoda kopiowania sekwencji DNA)
- 1992: wyznaczenie sekwencji chromosomu III drożdży ( $3 \cdot 10^5$  bp)
- 1995: wyznaczenie sekwencji bakterii *Haemophilus influenzae* ( $2 \cdot 10^6$  bp)
- 1998: odkrycie zjawiska interferencji mRNA (zjawisko wyciszania albo wyłączenia ekspresji genu przez dwuniciowy RNA)
- 1999: wyznaczenie sekwencji genomu wielokomórkowego organizmu (*Caenorhabditis elegans*) ( $10^8$  bp)
- 2001: 92% wyznaczenie sekwencji ludzkiego genomu ( $3 \cdot 10^9$  bp)
- 2003: wyznaczenie sekwencji ludzkiego genomu ( $3 \cdot 10^9$  bp)
- 2016: Kriomikroskopia elektronowa (cryogenic electron microscopy, cryo-EM) - wariant mikroskopii elektronowej pozwalający na obrazowanie makromolekuł z atomową dokładnością

Earth's heart of iron begins  
to yield its secrets p. 18

Microglia in chronic pain recovery  
and relapse pp. 33 & 86

Particle acceleration  
in a nova explosion p. 77

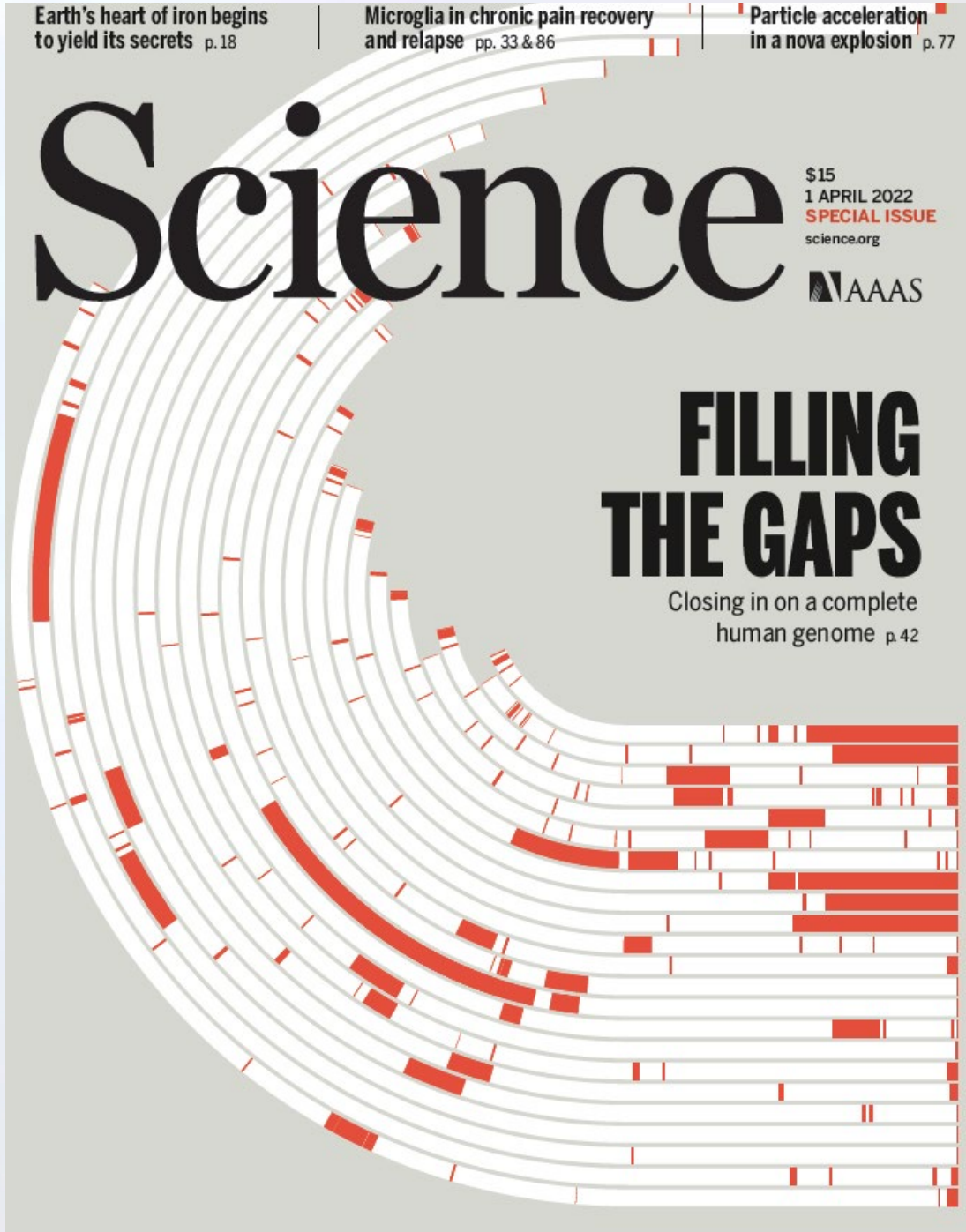
# Science

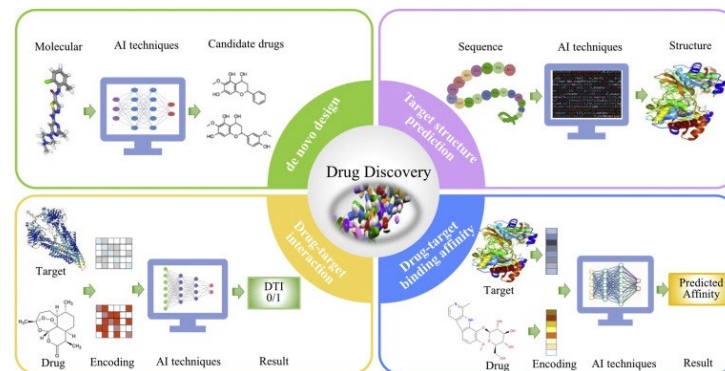
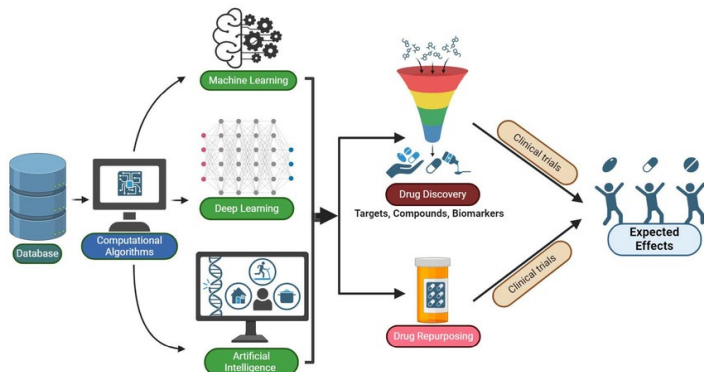
\$15  
1 APRIL 2022  
SPECIAL ISSUE  
science.org

AAAS

## FILLING THE GAPS

Closing in on a complete  
human genome p. 42



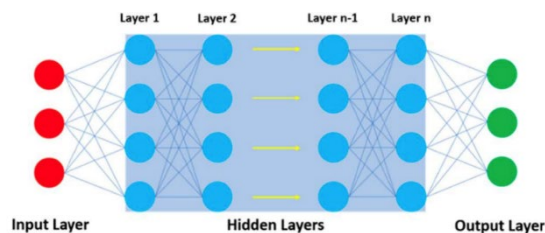


## 2017 – Attention is all you need

Artykuł "Attention Is All You Need" jest uważany za przełomowy w dziedzinie sztucznej inteligencji i przetwarzania języka naturalnego. Wprowadzenie architektury transformera zrewolucjonizowało sposób, w jaki modele językowe są projektowane i trenowane, co doprowadziło do rozwoju wielu nowoczesnych aplikacji AI, takich jak ChatGPT inne.

## 2021 – AlphaFold2

Pod koniec listopada 2020 roku odbył się 14 konkurs CASP. Startują w nim programy uczenia maszynowego. Aby ocenić wydajność ich prognoz komputerowych, każda grupa badaczy (autorów algorytmu) otrzymała sekwencje białek, których struktura została ustalona eksperymentalnie, ale nie została jeszcze opublikowana. Algorytm AlphaFold2 w przypadku większości białek uzyskał wynik zgodności (określający podobieństwo przewidzianej struktury do rzeczywistej) rzędu 90 punktów na 100 w przypadku dwóch trzecich białek. Dokładność obliczeń odpowiadała poziomom z laboratoryjnych technik eksperymentalnych.



**Sieć neuronowa**



# AlphaFold Protein Structure Database

Developed by DeepMind and EMBL-EBI

Search for protein, gene, UniProt accession or organism  BETA Search

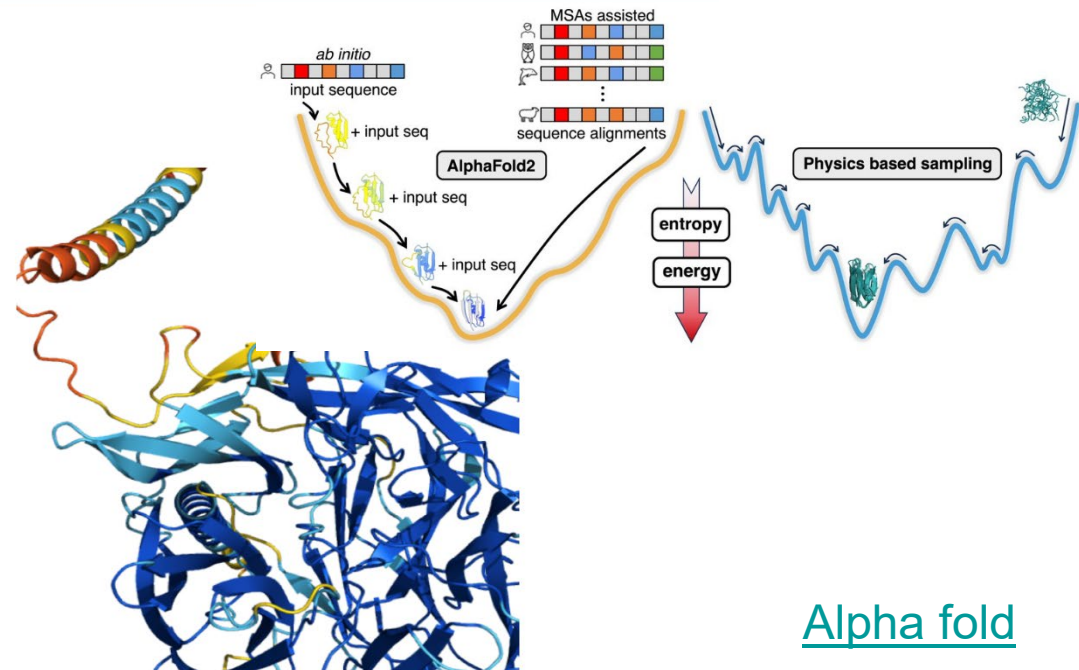
Examples: Free fatty acid receptor 2 A1g58602 Q5VSL9 E. coli [See search help](#)

AlphaFold DB provides open access to over 200 million protein structure predictions 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 experiment.

DeepMind and EMBL's European Bioinformatics Institute (EMBL-EBI) have partnered to create AlphaFold DB to make these predictions freely available to the scientific community. The latest database release contains over 200 million entries, providing broad coverage of UniProt (the standard repository of protein sequences and annotations). We provide individual [downloads](#) for the human proteome and for the proteomes of 47 other key organisms important in research and global health. We also provide a download for the manually curated subset of UniProt ([Swiss-Prot](#)).



Q813H7: May protect the malaria parasite against attack by the immune system. Mean pLDDT 85.57.

[View protein](#)

[Alpha fold](#)

# Ograniczenia sztucznej inteligencji



## 1. Brak transparentności

8 maja 2024 prestiżowe czasopismo „Nature” ogłosiło bardzo ważne wydarzenie w dziedzinie sztucznej inteligencji wspomagającej badania naukowe. Firma DeepMind (laboratorium AI należące do Google’a) zaprezentowała na łamach brytyjskiego periodyku kolejną wersję programu AlphaFold. Jest lepsza niż poprzednie, gdyż nie tylko dokładniej przewiduje struktury białek, ale również ich interakcje z innymi molekułami, takimi jak DNA, RNA i związkami małowcząsteczkowymi. Dzięki temu będzie nie tylko narzędziem służącym badaniom podstawowym, lecz także pomocą w **odkrywaniu nowych leków**.

Dlatego DeepMind oraz Isomorphic Labs (też należy Google’a i zajmuje się zastosowaniami AI w farmacji) ograniczyły transparentność publikacji o nim. W przeciwieństwie do poprzedniej wersji nie udostępniono bowiem pełnego kodu źródłowego algorytmu (czyli zestawu instrukcji napisanych w danym języku programowania), a korzystanie z narzędzia ograniczono do 10 prognoz dziennie dla jednego użytkownika. Spotkało się to ze sprzeciwem ponad 650 naukowców, którzy wysłali do redakcji „Nature” list pełen oburzenia. Zarzucają w nim nie tylko postawienie wymienionych barier, ale również naruszenie standardów naukowych, gdyż artykuł dotyczący AlphaFold3 ich zdaniem bardziej przypomina ogłoszenie firmowe niż transparentną publikację akademicką.

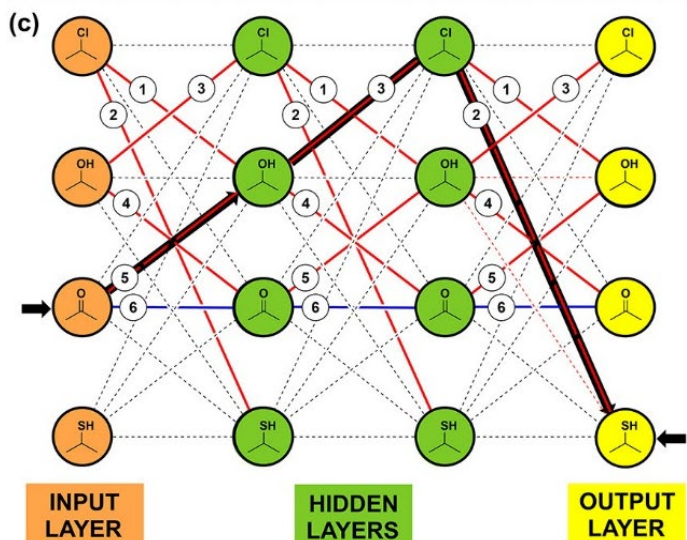
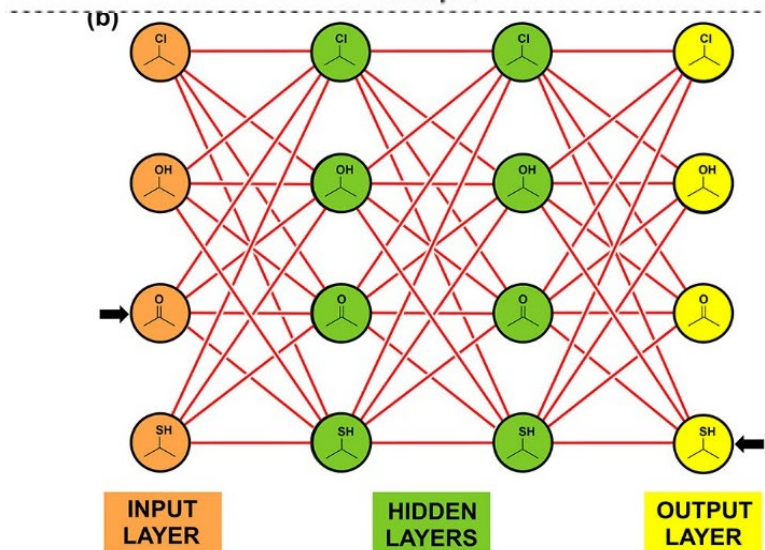
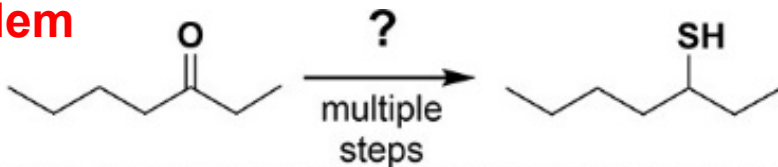
## 2. Wysokie koszty badań

Ograniczenia energetyczne stały się największą przeszkodą w budowie infrastruktury komputerowej dla AI. I nie da się temu łatwo zaradzić, gdyż przemysł energetyczny jest mocno regulowany. Przewiduje się, że infrastruktura komputerowa przeznaczona wyłącznie do trenowania i obsługi AI będzie musiała powstawać wraz z zasilającymi ją elektrowniami o mocy gigawata (największe pojedyncze reaktory jądrowe mają dziś moc 1,6 GW).

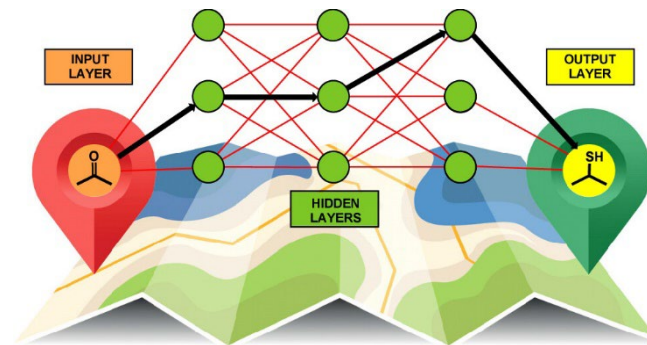
W ciekawej analizie w 2023 r. na łamach czasopisma naukowego „Joule” Alex de Vries z Vrije Universiteit Amsterdam szacuje, że w 2027 r. sektor AI może zużywać od 85 do 134 terawatogodzin (TWh) rocznie. To tyle, ile wynosi roczne zapotrzebowanie Holandii, Argentyny lub Szwecji.



# problem



- |  |  |                                 |
|--|--|---------------------------------|
| ① NaOH (S <sub>N</sub> 2) H <sub>2</sub> O (S <sub>N</sub> 1)    | ③ SOCl <sub>2</sub> (halogenation, S <sub>N</sub> 2) | ⑤ NaBH <sub>4</sub> (reduction) |
| ② NaSH (S <sub>N</sub> 2) or H <sub>2</sub> S (S <sub>N</sub> 1) | ④ PCC or PDC (oxidation)                             | ⑥ LDA, RX (enolate a            |



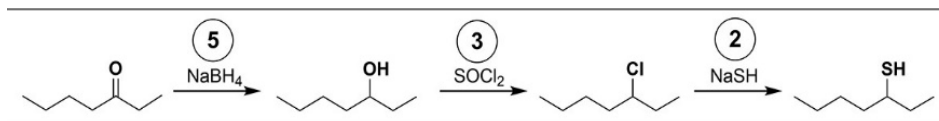
## Nowicjusz

Wszystkie reakcje równie prawdopodobne  
Wszystkie grupy funkcyjne mają jednakową wagę  
Niejasne ile warstw ukrytych jest koniecznych do rozwiązania problemu

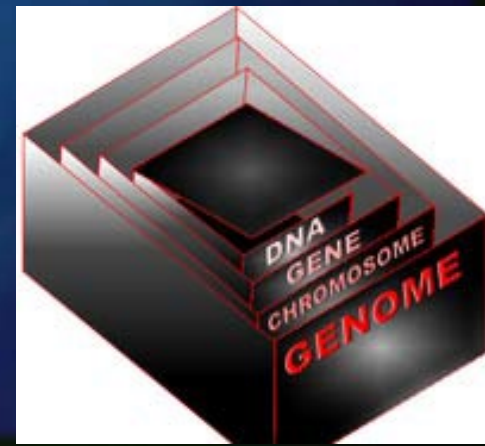
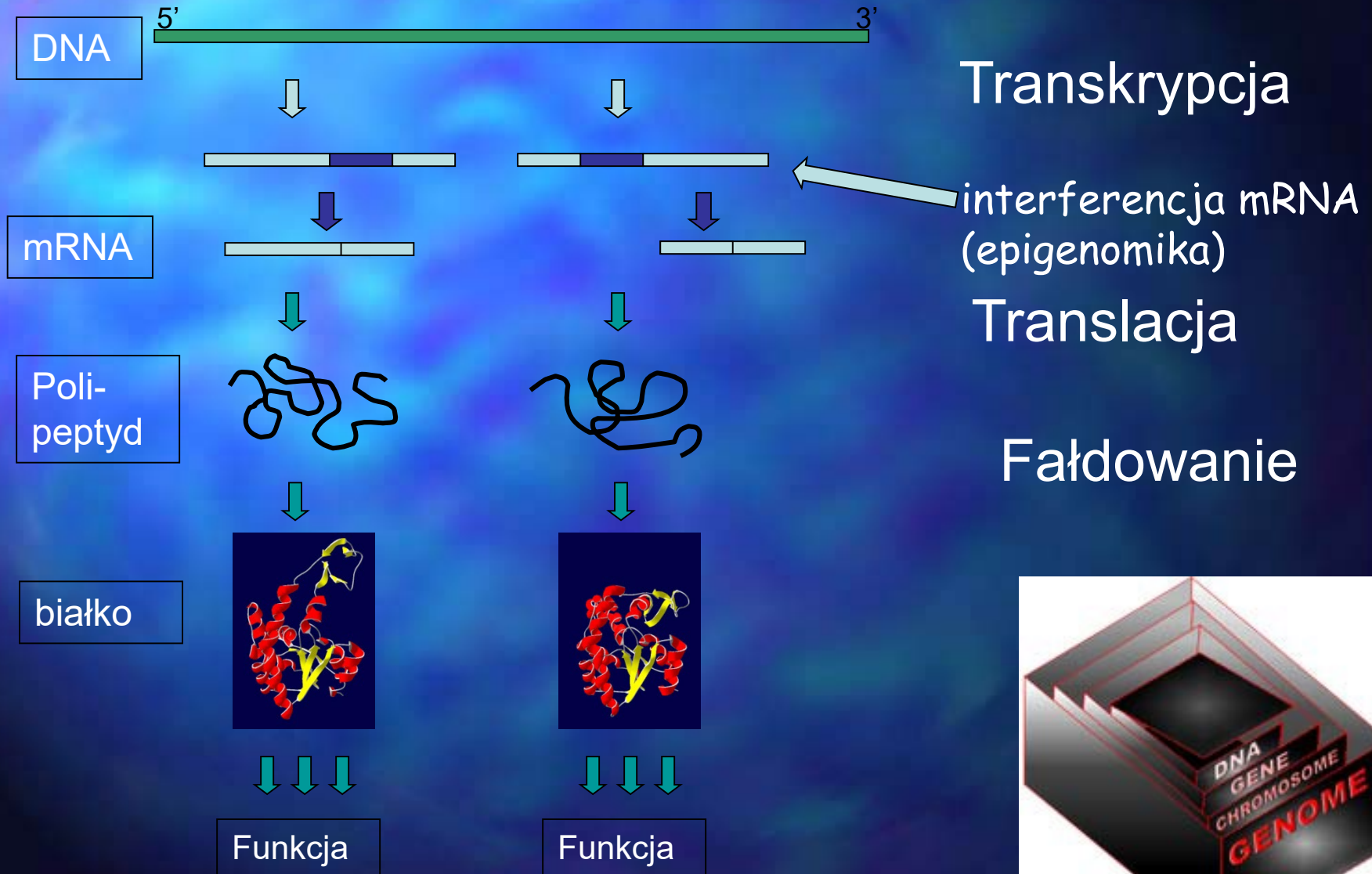
## Ekspert

Wiązanie C-C (kolor niebieski)  
Możliwe reakcje są znane (czerwone linie)  
Waga poszczególnych przekształceń odpowiada grubości linii  
Reakcje są klasyfikowane (numerowane)  
Rozległe szkolenie sieci ujawnia typowe kombinacje reakcji pomiędzy warstwami

## rozwiązanie



# Przeptyw informacji w układach biologicznych

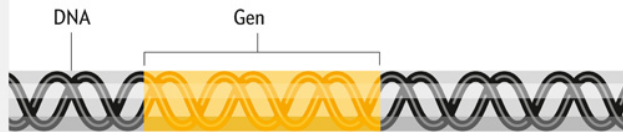


# Rewolucja RNA

Odkąd ponad 70 lat temu odkryto słynną strukturę podwójnej helisy DNA, jej najważniejsze zadanie wydawało się bardzo proste: przechowywanie kodu białek, wołów roboczych naszego organizmu. DNA przepisuje ten kod na tzw. cząsteczki matrycowego RNA, które przenoszą instrukcje do kompleksów odpowiadających za produkcję białek w komórkach. Jednak ostatnio biologowie odkryli, że DNA wytwarza także mnóstwo „niekodującego RNA” (ncRNA), które nie produkują białek. Niektóre cząsteczki ncRNA mogą włączać i wyłączać geny. Inne ich funkcje nadal są przedmiotem gorących dyskusji.

## SPOJRZENIE TRADYCYJNE

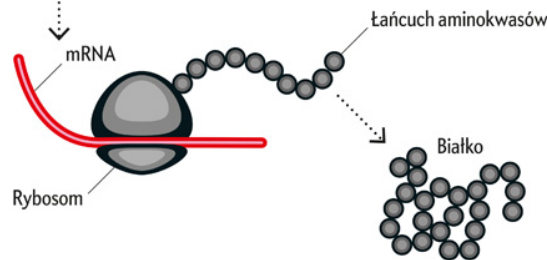
Przez dziesięciolecia uważano geny za oddzielne segmenty DNA, zapoczątkowujące proces tworzenia białek.



Te tzw. kodujące geny prowadzą do powstawania RNA matrycowego (mRNA) w procesie nazywanym transkrypcją.

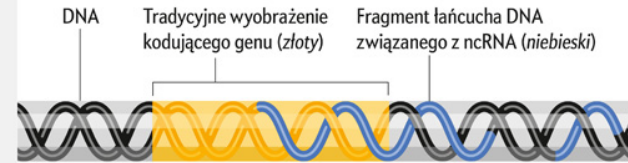


Następnie mRNA przemieszcza się do rybosomów – fabryk wytwarzających białka – i prowadzi do wytwarzania z aminokwasów białek w procesie nazywanym translacją. Proces ten przebiega jednokierunkowo: od DNA, przez RNA, do białka.

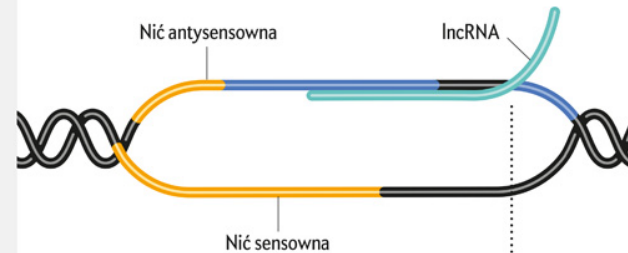


## NOWE SPOJRZENIE

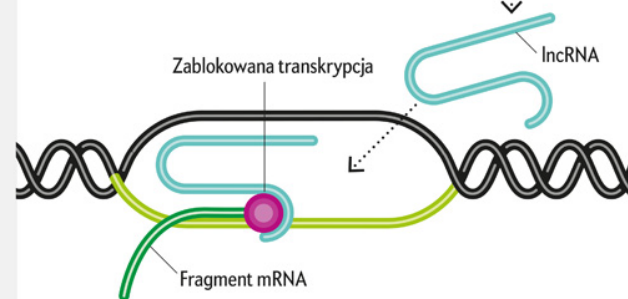
W ostatnich latach naukowcy odkryli, że wiele części genomu prowadzi do ekspresji cząsteczek RNA, które nie ulegają translacji do białek. Te niekodujące RNA (ncRNA) mogą powstawać z sekwencji położonych pomiędzy genami kodującymi białka albo się na nie nakładać.



Jedną z takich cząsteczek jest długie niekodujące RNA (lncRNA). Części genomu, które w jednej z dwóch nici DNA („sensownej”) kodują mRNA, mogą w drugiej, komplementarnej nici („antysensownej”) kodować lncRNA.



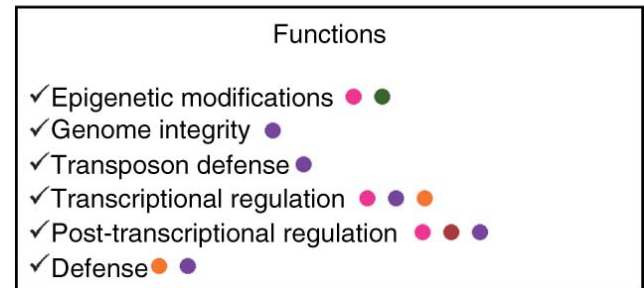
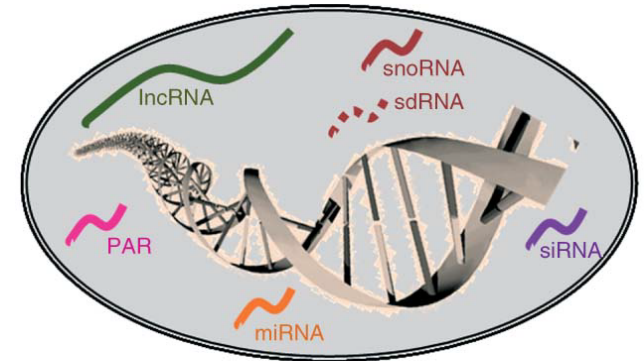
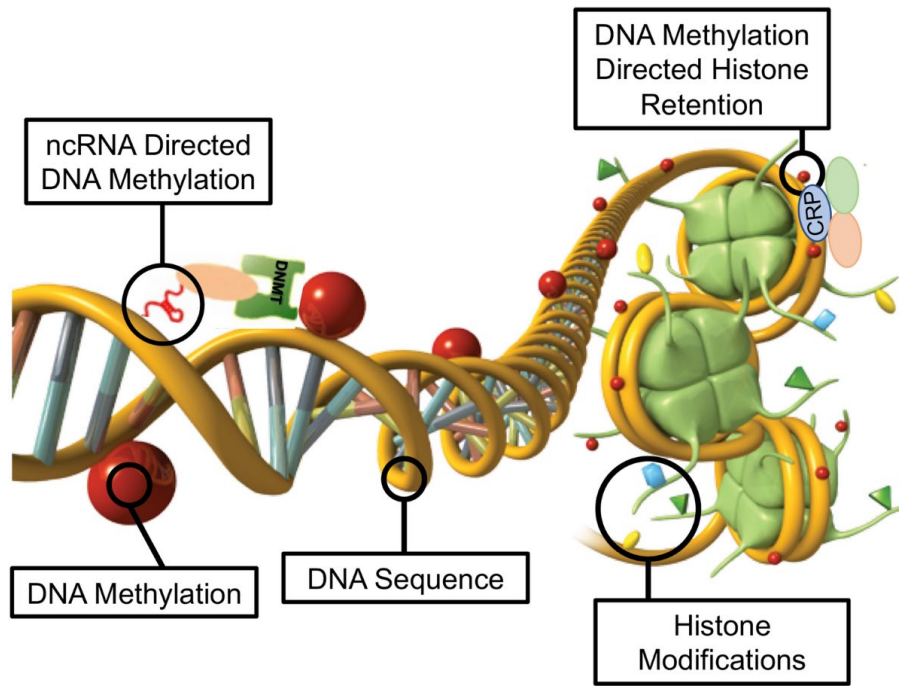
W przykładzie poniżej genom zawiera szablon dla lncRNA o działaniu regulującym. To lncRNA nie prowadzi bezpośrednio do powstawania białka. Wraca do genomu i wchodzi w interakcję z genem kodującym białko, ograniczając jego aktywność albo wręcz go wyłączając. W ten sposób lncRNA może kontrolować ilość powstającego ostatecznie białka.



Nie wszystkie ncRNA pełnią funkcje regulacyjne. Niektóre mają działania „porządkujące”. Jedną z nich jest na przykład rusztowanie, na którym dochodzi do składania wielu białek w większą jednostkę. Jednak wiele z nich nie poddaje się klasyfikacji. Na przykład transportujące RNA (tRNA) – element kompleksu syntetyzującego białka – często klasyfikuje się jako porządkujące, ale pełni ono również funkcje regulacyjne.



Co najmniej 75% genomu ulega transkrypcji do RNA. Skoro tylko 1–2% tego RNA koduje białka, to czemu ma służyć cała reszta?

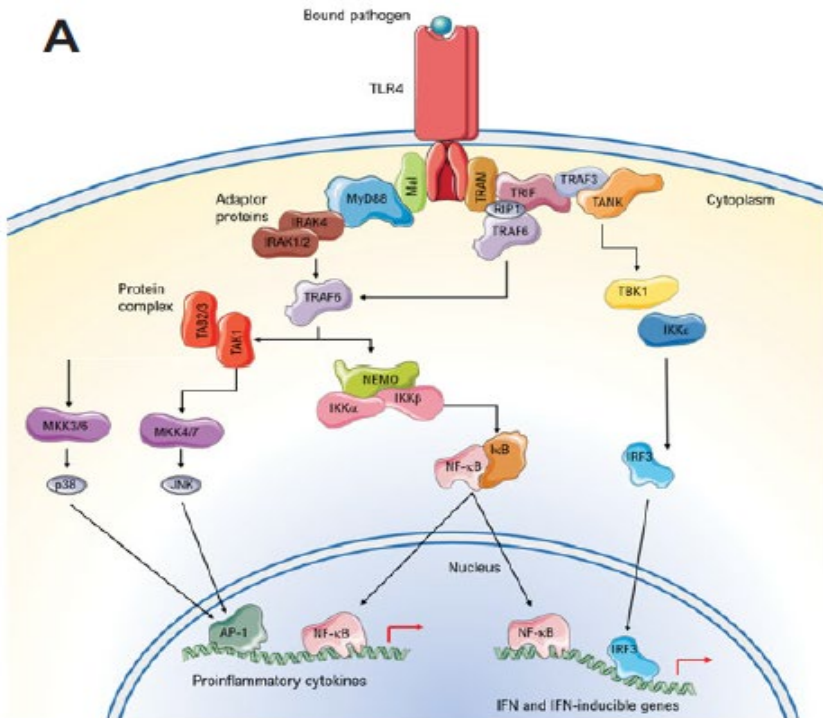


2. W ramach projektu ENCODE zidentyfikowano około 37 600 niekodujących genów czyli fragmentów DNA zawierających instrukcje dla cząsteczek RNA, które nie kodują białek. To niemal dwa razy więcej niż liczba genów kodujących białka.

3. Rewolucyjność tych odkryć wynika z funkcji tego niekodującego RNA, nazywanego w skrócie ncRNA (od noncoding RNA). Wydaje się, że w znacznej części odpowiada ono za regulację aktywności genów – nie tylko ich zwykłe włączanie lub wyłączanie, ale także precyzyjne sterowanie tą aktywnością.

4. Nils Walter z Center for RNA Biomedicine z University of Michigan napisał w artykule z 2024 roku, że rozkwitający potencjał terapii opartych na RNA „sprawia, że potrzebujemy poznać funkcje ncRNA”. Osiągnięcie tego celu „pozwoili wreszcie zrealizować w pełni zamierzenia Human Genome Project”.

# Zależności sieciowe w bioinformatyce



**C**

	TAK1	Mal	MyD88	IRAK1/2	IRAK4	IRF3	JNK	p38	TAB2/3	MKK4/7	AP-1	MKK3/6	NF-κB	NEMO	IKB	IKKa	IKKb	TLR4	TRAF6	IKKe	TBK1	TANK	TRAF3	TRIF	RIP1	TRAM
TAK1	0	0	0	0	0	0	0	0	1	0	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
Mal	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0
MyD88	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
IRAK1/2	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0
IRAK4	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0
IRF3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0
JNK	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
p38	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TAB2/3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
MKK4/7	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AP-1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MKK3/6	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NF-κB	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0
NEMO	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0
IKB	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0
IKKa	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0
IKKb	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0
TLR4	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	1	1	1
TRAF6	1	1	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	1	1	1	1	1	1
IKKe	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	1	0	0	0
TBK1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	1	1	1
TANK	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	1	1	1	1
TRAF3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	1	1
TRIF	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	1	1	1
RIP1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	1	1	1
TRAM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	1	1	1

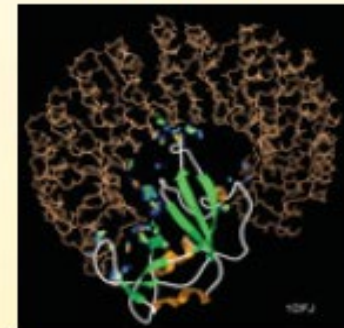
Interaction

Yes (1)

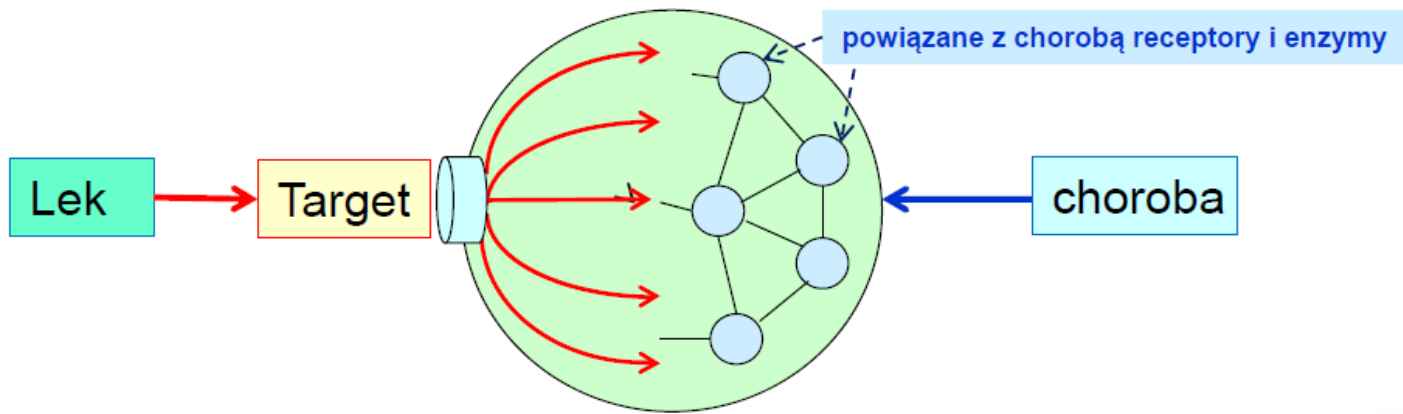
No (0)



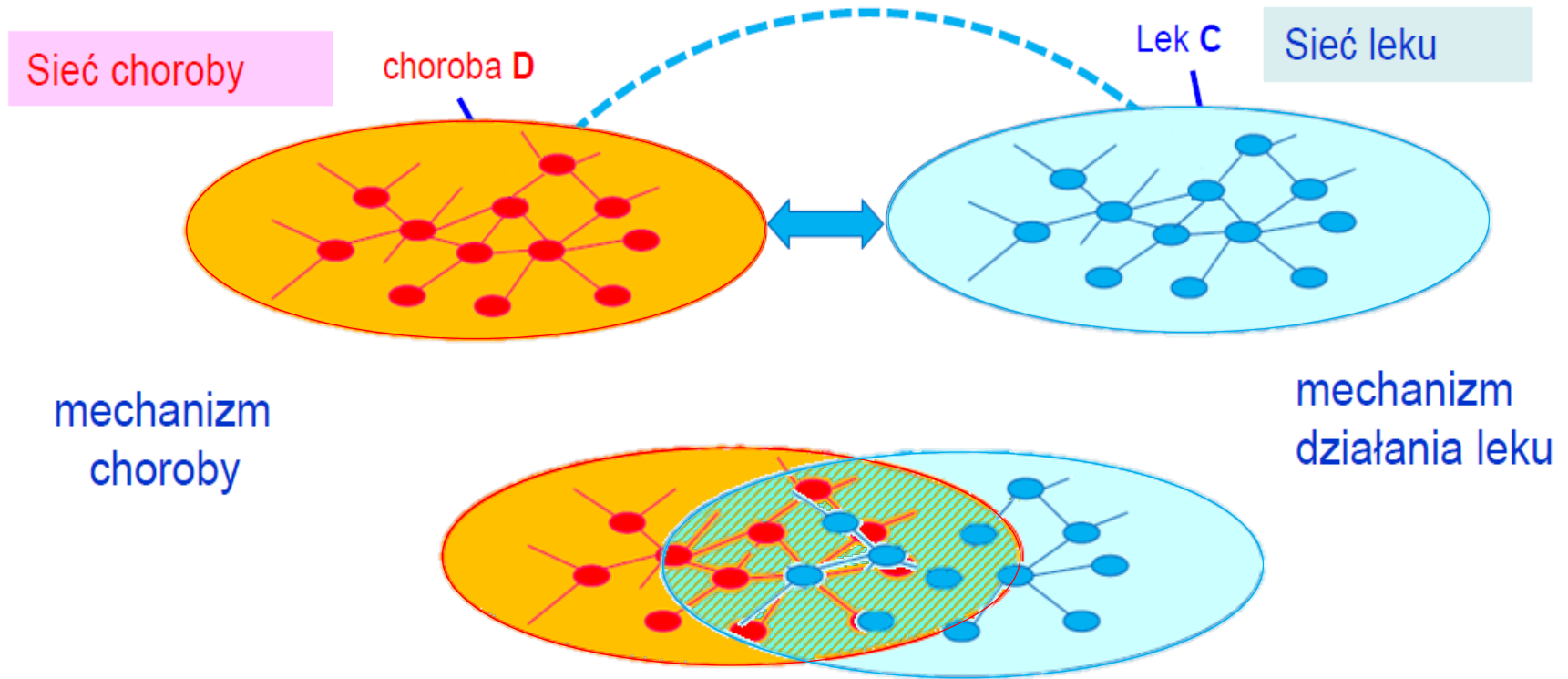
## Protein-Protein interaction network







Lek C posiada terapeutyczny efekt na chorobę D

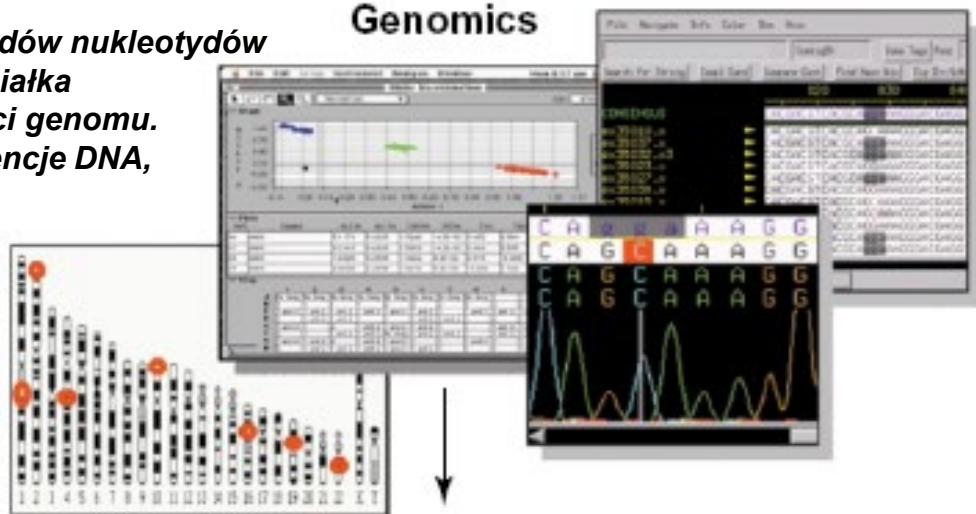


W jakim stopniu pokrywają się szlaki sygnałowe leku i choroby?  
Które spośród nich odpowiadają za efekty niepożądane?

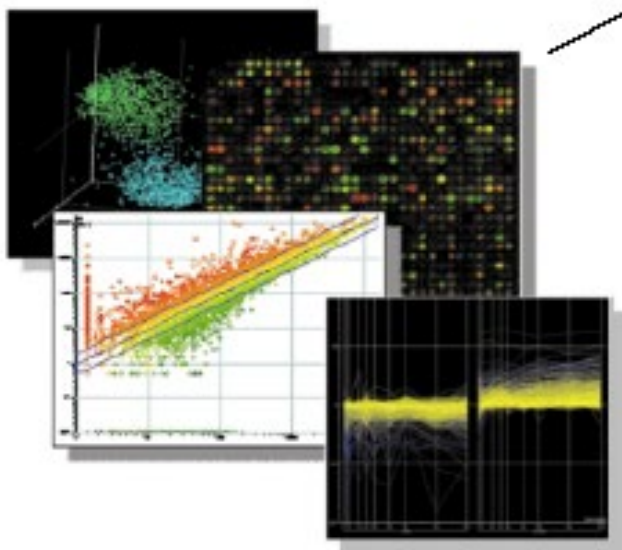
**Ludzki genom zawiera około 3 miliardów nukleotydów i prawie 20 000 genów kodujących białka – szacunkowo 1% całkowitej długości genomu. Pozostałe 99% to niekodujące sekwencje DNA, które nie wytwarzają białek**

**Mała pchła wodna Daphnia ma najwięcej genów ze wszystkich zwierząt, około 31 000.**

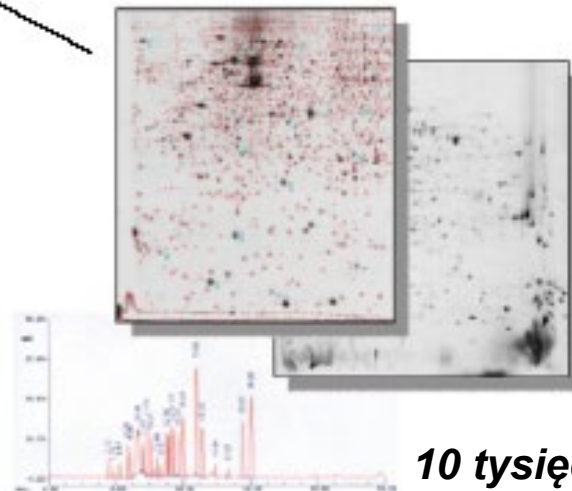
## Genomics



## Transcriptomics



## Proteomics



**Clinical data**

**10 tysięcy białek  
1300 enzymów**

Welcome

Deposit

Search

Visualize

Analyze

Download

Learn

## A Structural View of Biology

This resource is powered by the Protein Data Bank archive-information about the 3D shapes of proteins, nucleic acids, and complex assemblies that helps students and researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease.

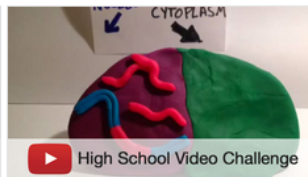
As a member of the wwPDB, the RCSB PDB curates and annotates PDB data.

The RCSB PDB builds upon the data by creating tools and resources for research and education in molecular biology, structural biology, computational biology, and beyond.

### Structure and Health Focus: HIV

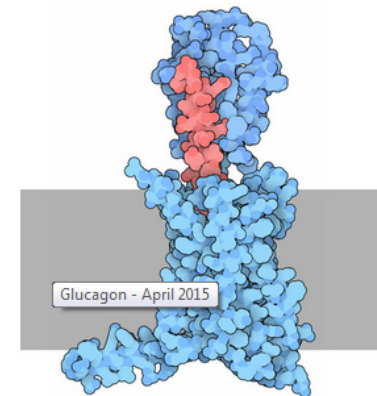


HIV Resources



High School Video Challenge

## April Molecule of the Month

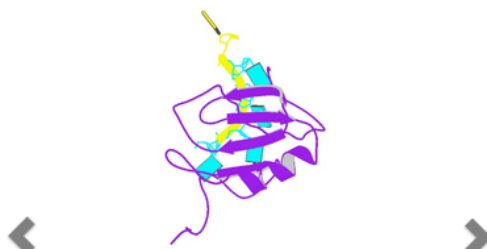


Glucagon - April 2015

Glucagon

## Latest Entries

As of Tuesday Apr 07



2MZ1

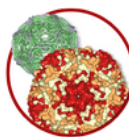
PDB Entry

Solution structure of hnRNP C RRM in complex with 5' UTR of RNA

olecule\_of\_the\_month/current\_month.html

## New Features

### December 2014 Release



Improved Representation of Large Structures  
Updates to PDB Archive, Visualization, New Queries for Ribosomes and Viruses

### November 2014 Release



Redesigned Home Page  
Simple. Clean. Usable. Tools & Functions More Visible.

## News

Publications



Spring Newsletter Published  
Read about the new Molecular Machinery viewer, released just as the PDB reached 100,000 protein entries, large structures in the PDB, a paper describing the Chemical

Component Dictionary, BioJava, and more. The Education Corner highlights a new online resource



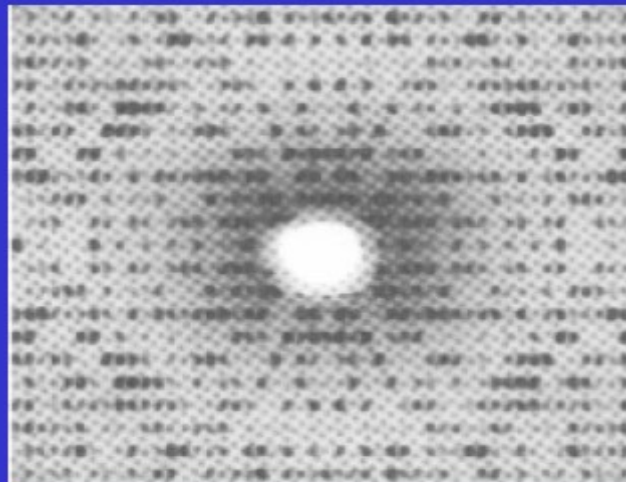
# Piękno nauki



*Visitors at the reception for "The Art of Science" exhibit that displayed images from the Protein Data Bank. The exhibit turned out to be one of the more popular shows held at The Gallery, a space dedicated to art exhibits at Rutgers University.*

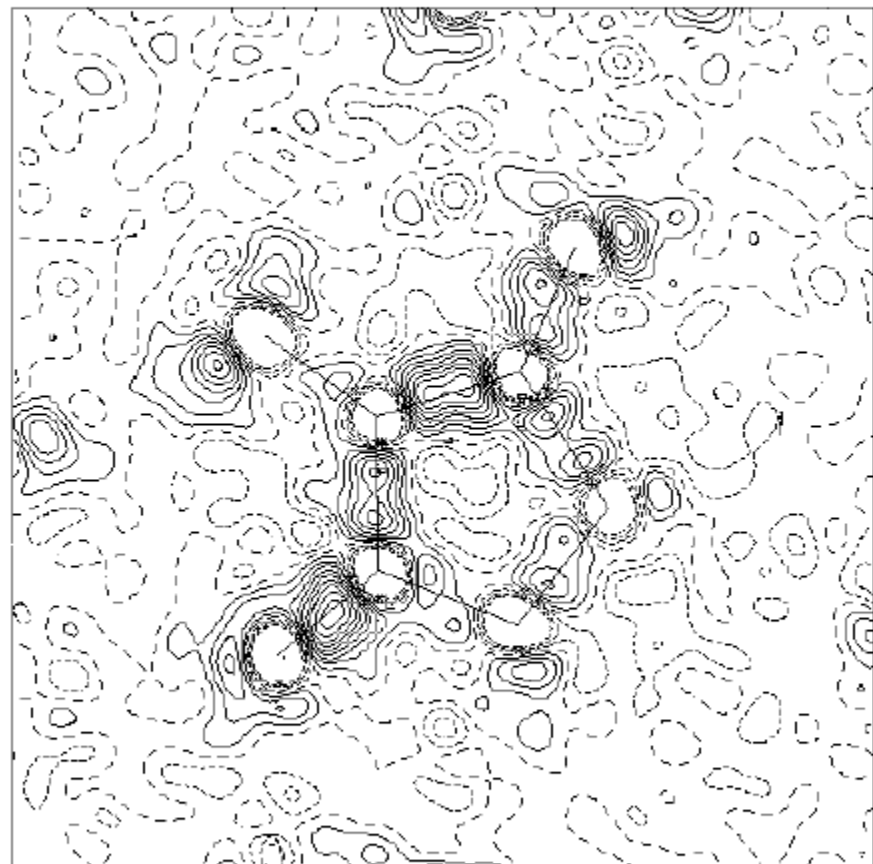
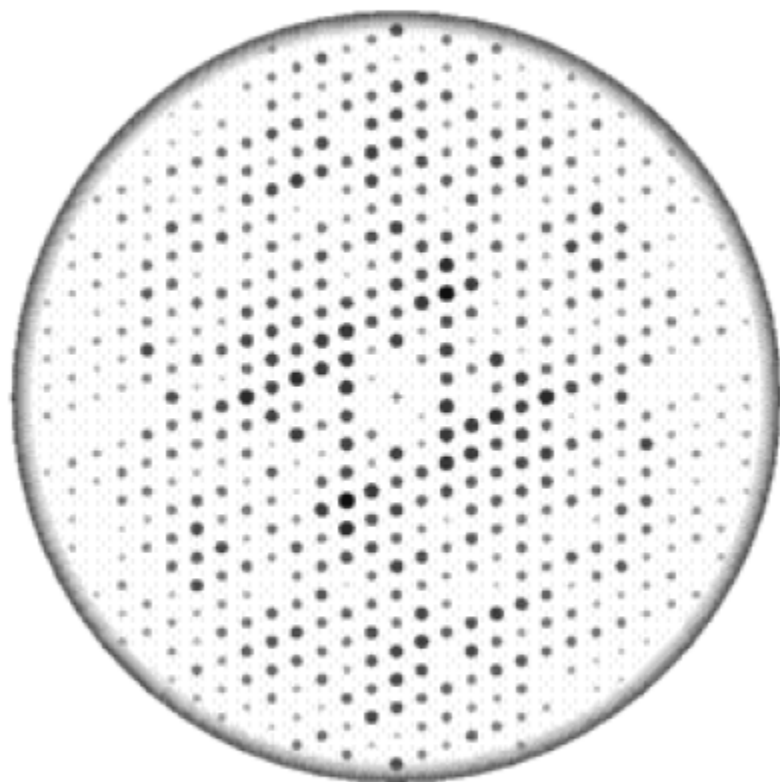
# Otrzymywanie struktur trójwymiarowych kwasów nukleinowych i białek

- Krystalografia rentgenowska
- Badania NMR
- mikroskopia elektronowa
- modelowanie cząsteczkowe (?)





# Krystalografia rentgenowska



X-ray diffraction pattern



Electron density map

# Poszukiwanie leków in silico

### Psi-Blast Sequence Search

Search Sequence

```
>MM000002 ENSMUSP00000019897  
Gene:ENSMUSG00000019763 Clone:10.1-5000000  
Contig:10.1-5000000 Chr:10 basepair:1420486  
MPSRLHLTLASLHSISAKAHQCQLTLWTLRPLRKSESTTCHNGL  
TIQQHLHSFFPSITA
```

Browse... Search Sequence Length: 177

Search Criteria

Max Hits  50

Max E\_Value

Max Gaps

### Search DS AtlasStore

Find: Protein Structure

- Organism
- Data source
- Ligand
  - IUPAC Name
  - Compound Name
  - 2D Structure
  - Registration Number
  - Molecular Weight
  - Molecular Formula
  - Autotyping
- X-Ray
  - Resolution(High)

Logic	Attribute	Operator	Value
1	POB Header:Classifica...	Fuzzy	kinase
2	Sequence:Blast Seque...	is	Blast...
3	Ligand:2D Structure	Match: Substruc...	CEH7N

- Match: Equal
- Match: Exact
- Match: Substructure
- Similarity: Tanimoto
- Similarity: Euclid
- Similarity: EDS

### Select Cell Dimensions

Dimensions

A  100

B  100

C  100

Angles

Alpha  90

Beta  90

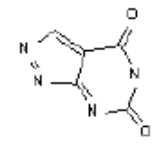
Gamma  90

Range

Resolution (%)  5.0

OK Cancel

### Query Chemistry



Formula  
C5H2N4O2

SMILES

OK Cancel Recall List Chemistry Similarity Cutoff

BLAST searching

X-ray metadata

Ligand 2D chemistry

Wybór receptora biologicznego

Struktury trójwymiarowe  
ligand – receptor biologiczny

Budowa modelu receptor – ligand

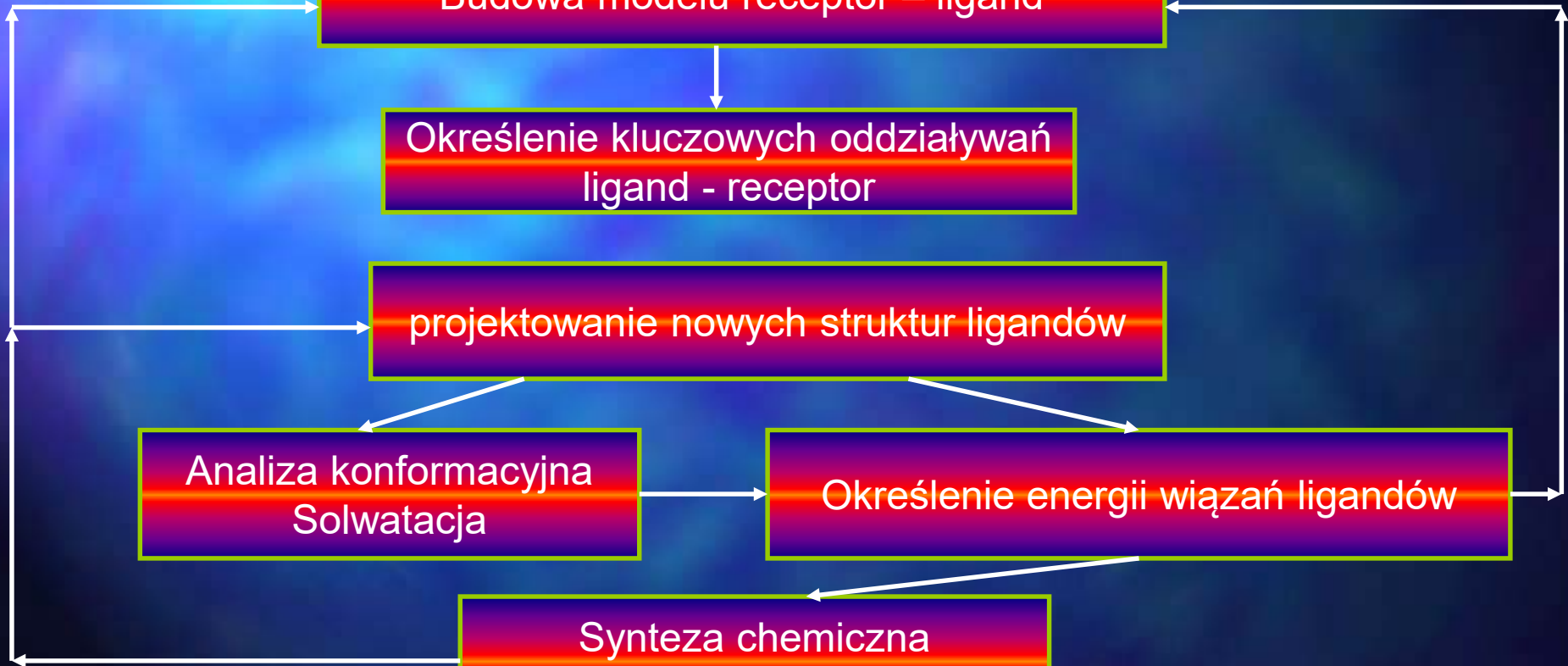
Określenie kluczowych oddziaływań  
ligand - receptor

projektowanie nowych struktur ligandów

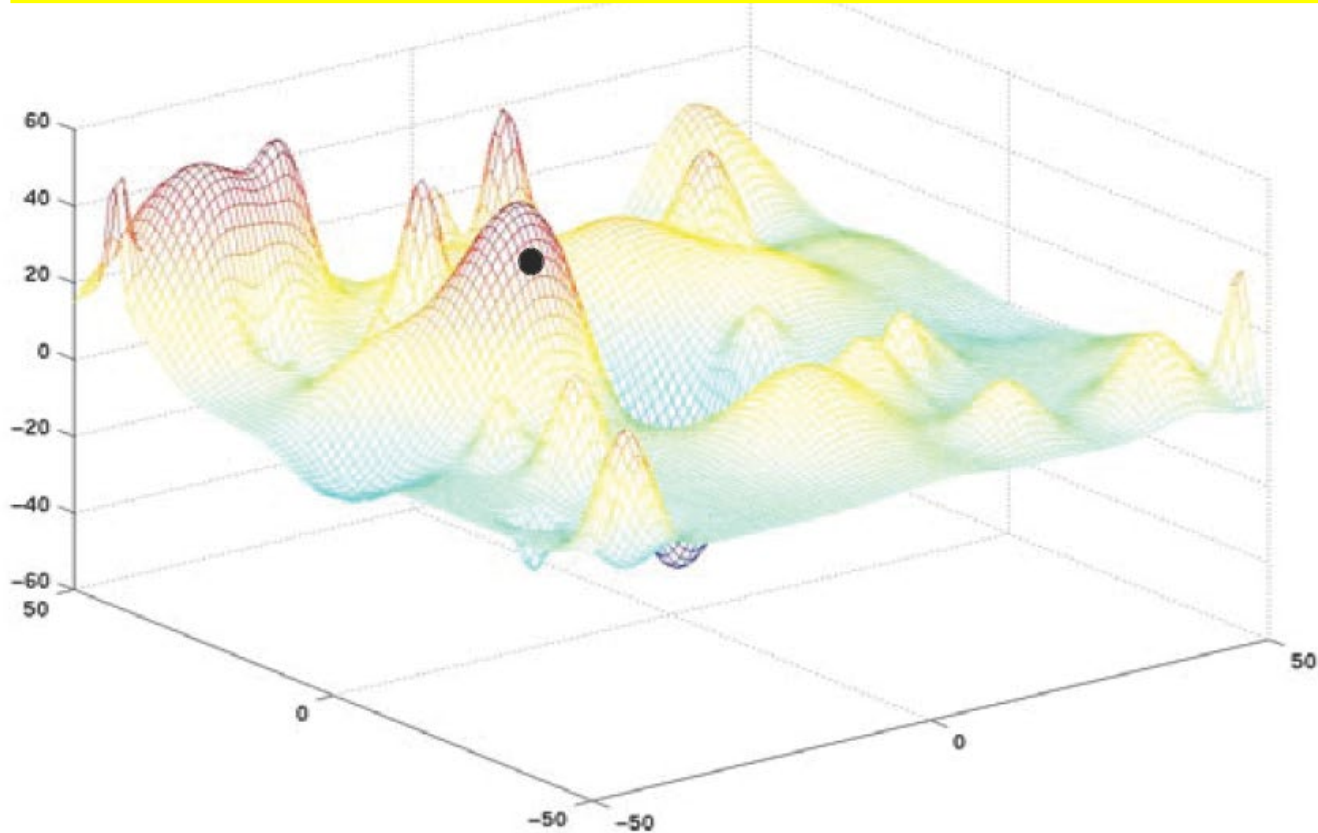
Analiza konformacyjna  
Solvatacja

Określenie energii wiązań ligandów

Synteza chemiczna  
Testy biologiczne



# Hyperpowierzchnia energii potencjalnej

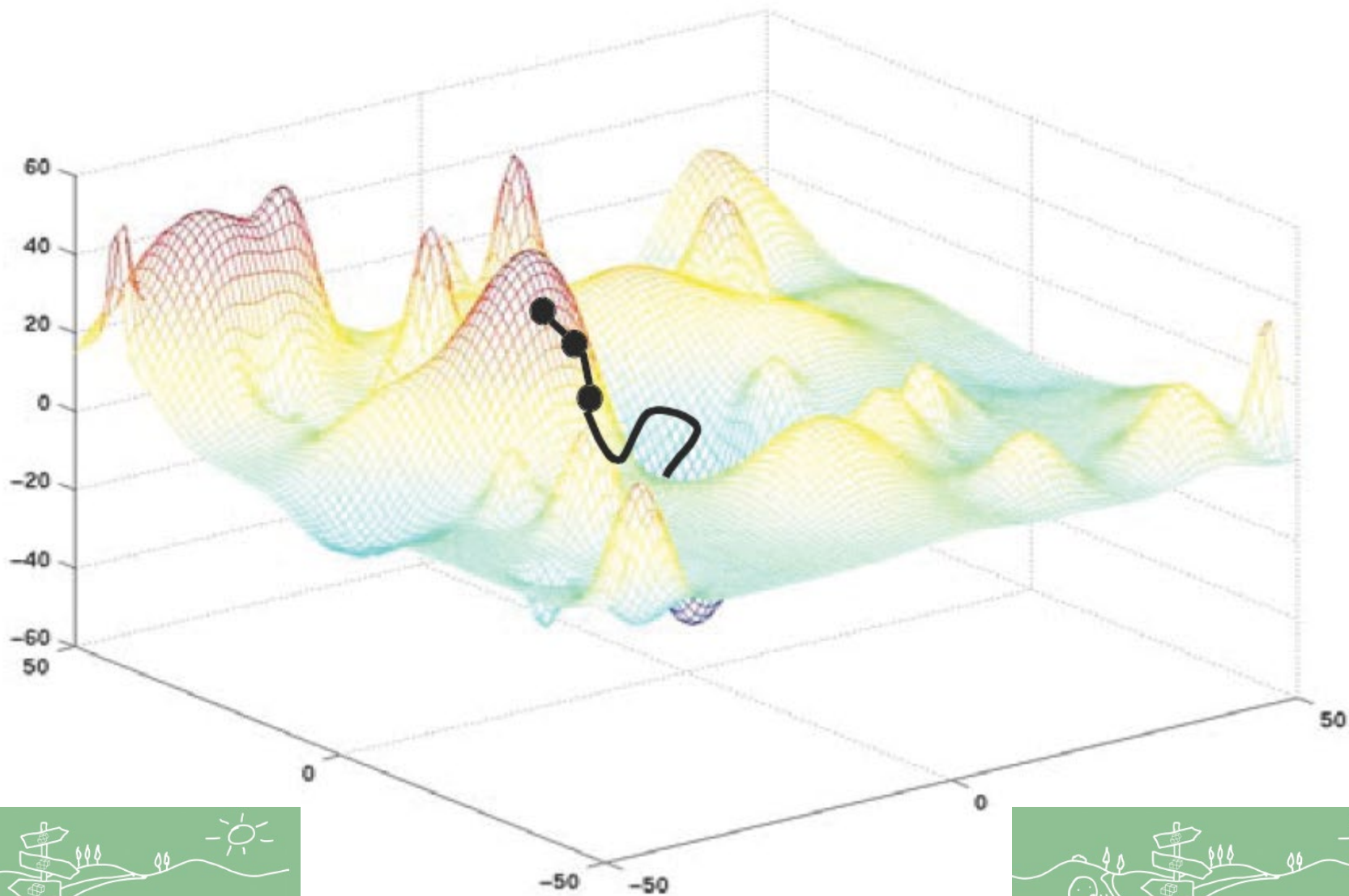


***Optymalizacja geometrii = minimalizacja energii***

**Potrzebujemy zależności funkcyjnej wiążącej  
położenie atomów z energią:**

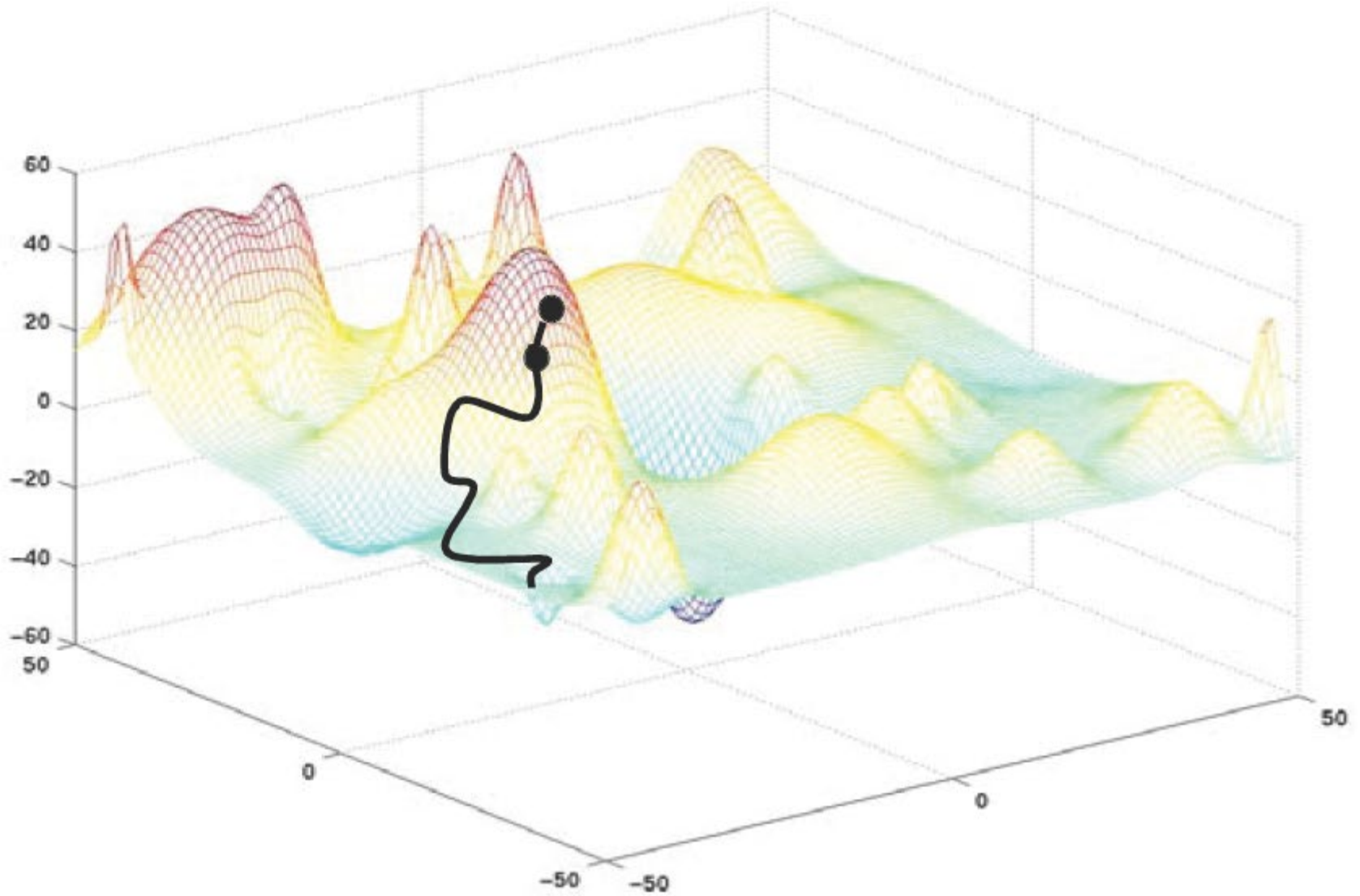
$$E = f(Q)$$

# Hyperpowierzchnia energii potencjalnej



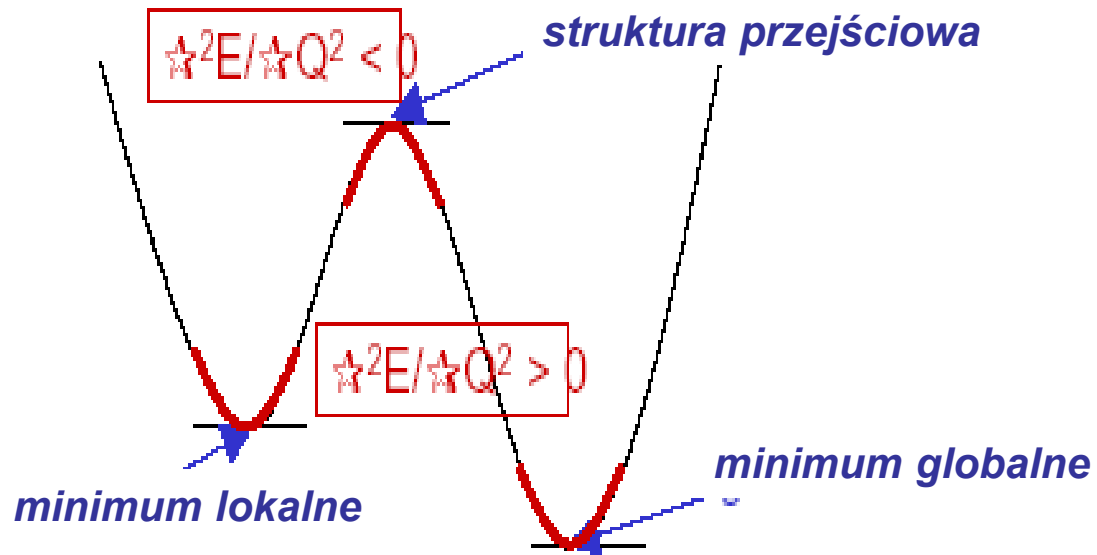
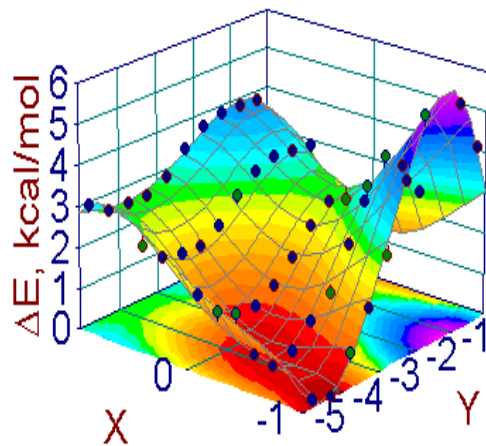


# Hyperpowierzchnia energii potencjalnej





# Wyznaczenie hyperpowierzchni energii potencjalnej układu molekularnego $E = f(Q)$



**Cel: wyznaczenie stabilnej struktury cząsteczki  
lub układu cząsteczek**

# Metody obliczeniowe

- ❖ Teoria orbitali molekularnych

- ❖ Metody ab initio w przybliżeniu Hartree-Focka

- ❖ GAMESS, Gaussian

- ❖ Teoria funkcjonału gęstości (DFT)

- ❖ Metody półempiryczne

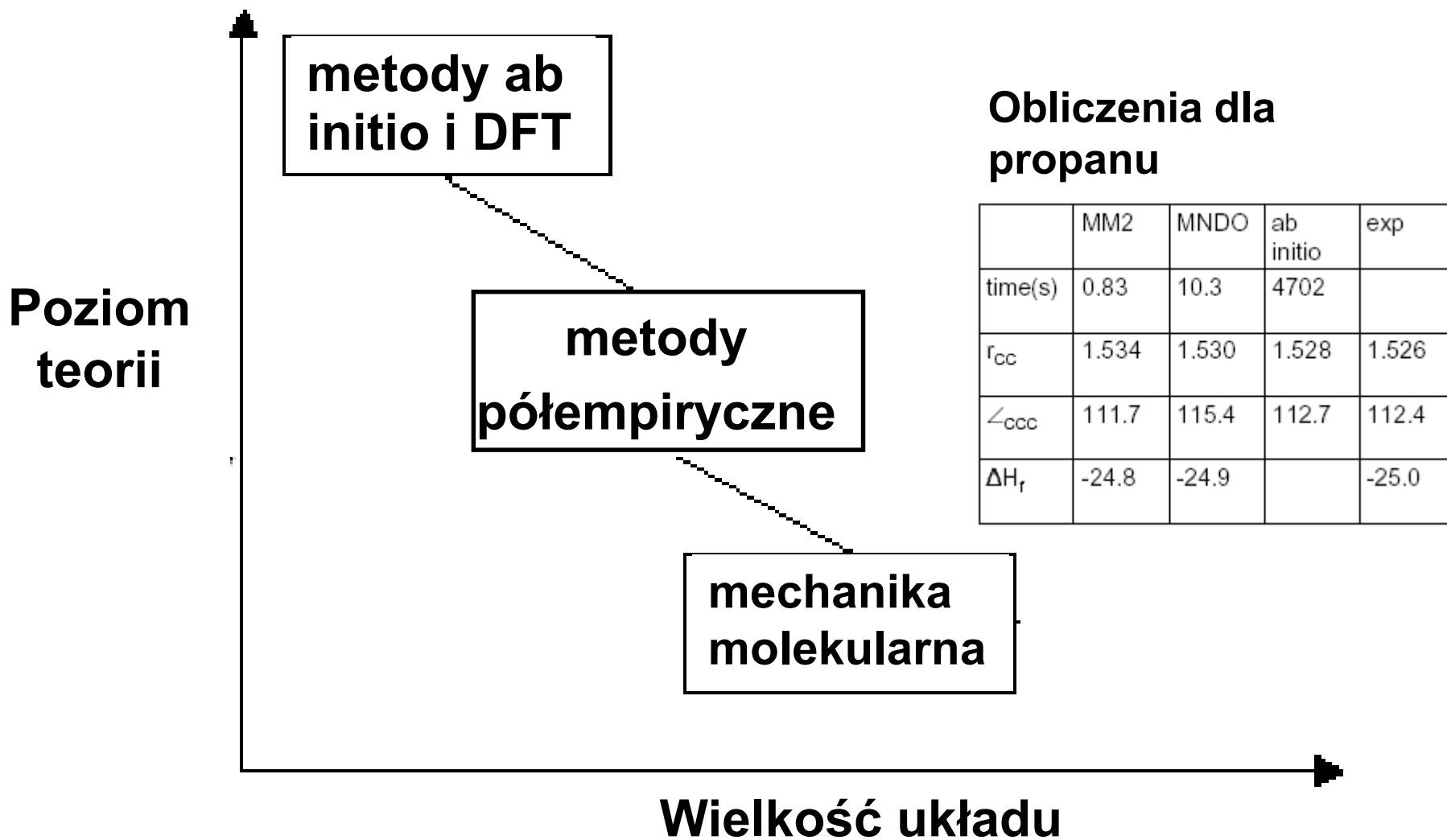
- ❖ CNDO, ZINDO, AM1, PM3, PM5

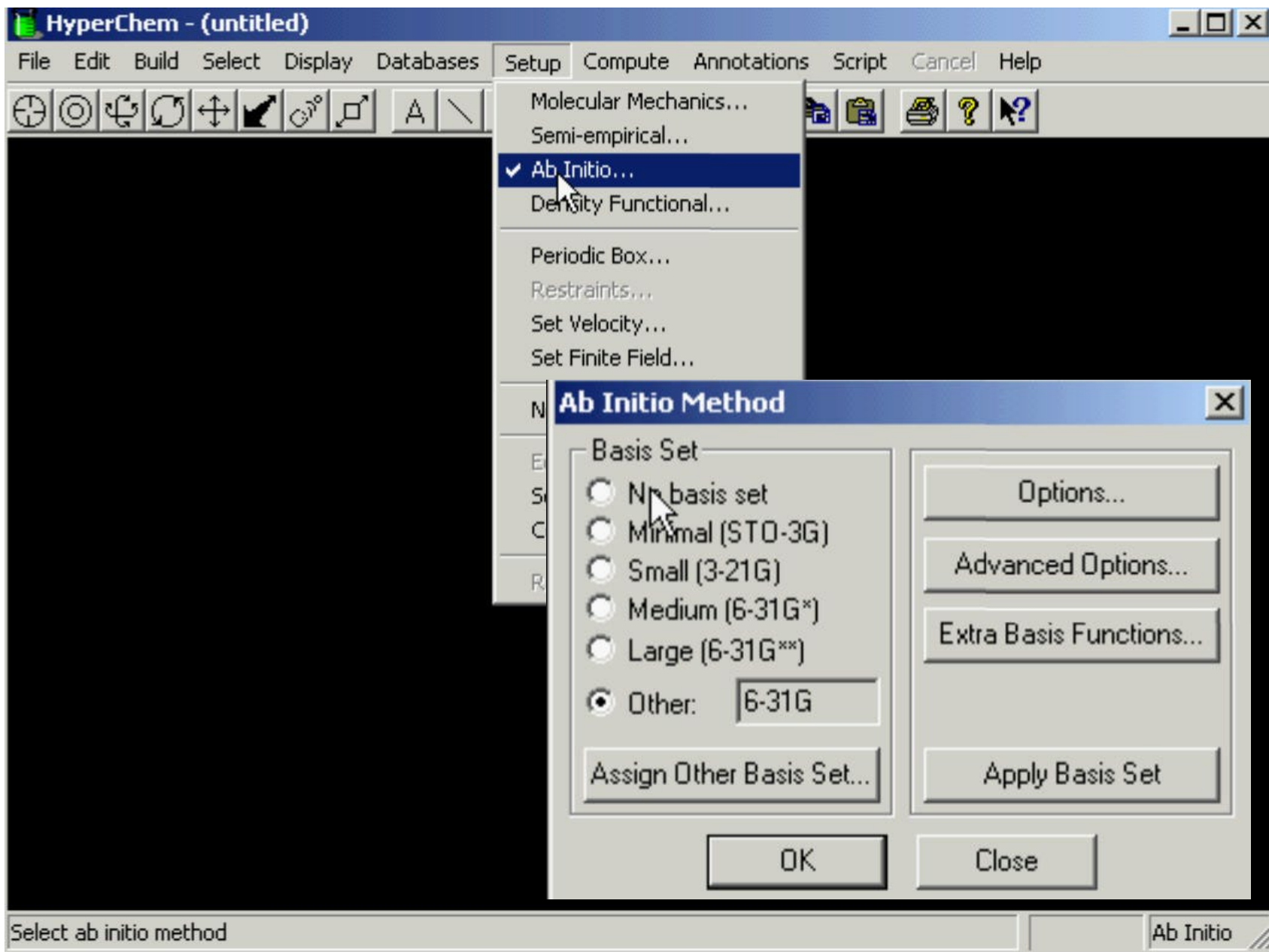
- ❖ Mechanika molekularna

- ❖ MM, AMBER, CHARMM, SYBYL



# Porównanie metod obliczeniowych



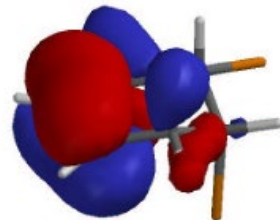




# Metody ab initio w przybliżeniu Hartree-Focka

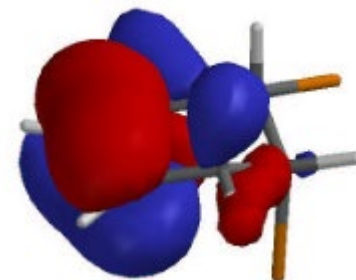
$$H\Psi = E\Psi$$

- ❖ Przybliżenie Borna-Oppenheimera
  - ❖ ruch elektronów jest niezależny od ruchu jądra
- ❖ Przybliżenie Hartree-Focka
  - ❖ elektrony znajdują się w uśrednionym polu innych elektronów
- ❖ Przybliżenie LCAO (orbitale cząsteczkowe są liniową kombinacją orbitali atomowych)



# Metody ab initio w przybliżeniu Hartree-Focka

- ❖ Dokładność wyników zależy od stopnia korelacji elektronowej i wielkości bazy funkcyjnej
- ❖ Koszt obliczeń wzrasta szybko jak wzrasta wielkość bazy funkcyjnej i ilość elektronów
- ❖ Czas CPU jest proporcjonalny do  $n^4$ , gdzie  $n$  odpowiada liczbie funkcji bazowych



# Mechanika kwantowa: wyniki

## **Geometria cząsteczki:**

- **wyniki bardziej dokładne niż w innych metodach**
- **możliwa do wyznaczenia dla szerokiej klasy związku**

## **Właściwości elektronowe:**

- **rozkład gęstości elektronowej**
- **moment dipolowy**
- **potencjał elektrostatyczny**

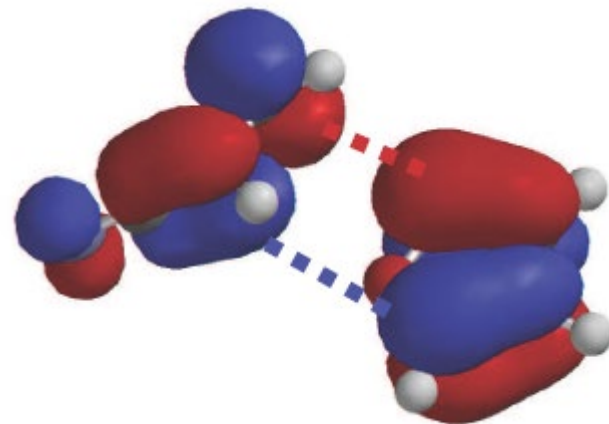
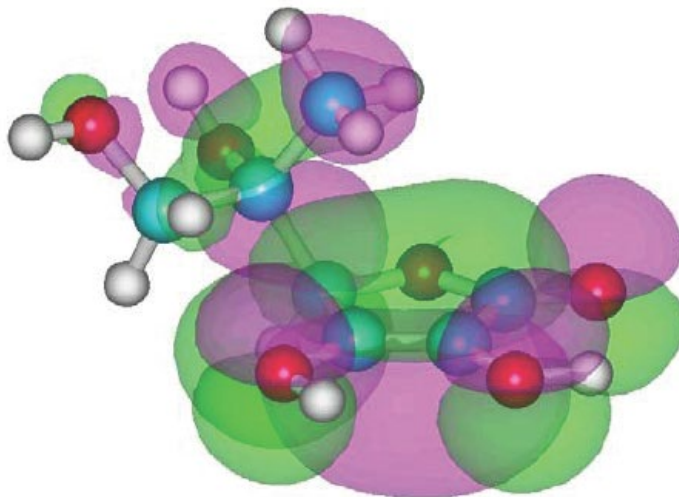
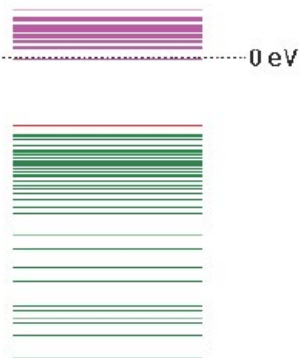
**Właściwości spektroskopowe (widma IR, UV, NMR, ESR)**

**Struktury związków przejściowych i kompleksów aktywnych**

# Mechanika kwantowa: wyniki

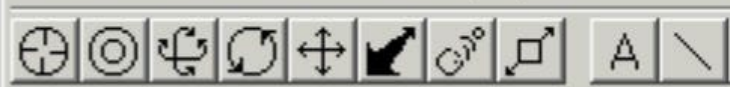
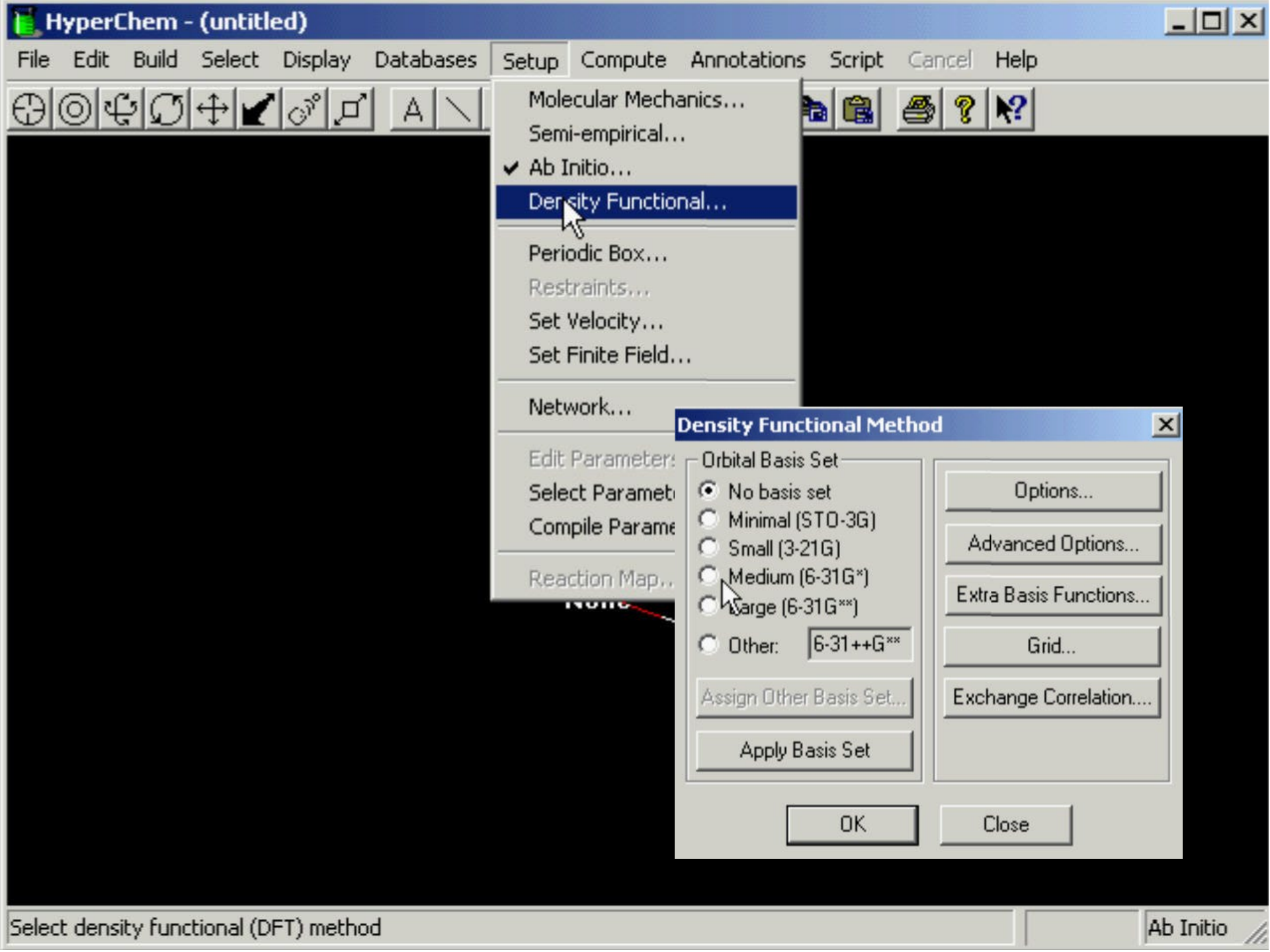
## Orbitale cząsteczkowe:

- Orbitale graniczne – określają gdzie są najbardziej reaktywne elektrony (HOMO) i gdzie podążają (LUMO)
- niezbędne do wyjaśnienia chemicznej reaktywności



Vitamin C LUMO





- Molecular Mechanics...
- Semi-empirical...
- ✓ Ab Initio...
- Density Functional...**
- Periodic Box...
- Restraints...
- Set Velocity...
- Set Finite Field...
- Network...

### Density Functional Method

Orbital Basis Set

- No basis set
- Minimal (STO-3G)
- Small (3-21G)
- Medium (6-31G\*)
- Large (6-31G\*\*)
- Other:

Assign Other Basis Set...

Apply Basis Set

Options...

Advanced Options...

Extra Basis Functions...

Grid...

Exchange Correlation...

OK Close

# Teoria funkcjonału gęstości

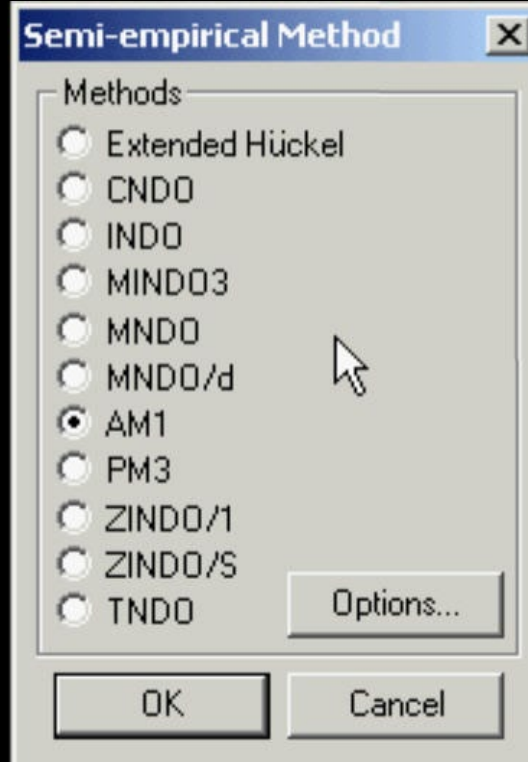
- ❖ Gęstość elektronowa i funkcja falowa mogą być używane alternatywnie to opisu stanu podstawowego układu
- ❖ Istnieje funkcjonał gęstości dla określonej liczby elektronów wyznaczający energię układu
- ❖ Czas CPU jest proporcjonalny do  $n^3$ , gdzie  $n$  odpowiada liczbie funkcji bazowych

Tablica 1. Porównanie obliczonych (przybliżenie harmoniczne) i doświadczalnych wartości częstości drgań ( $\text{cm}^{-1}$ ) dwóch wybranych molekuł

Molekuła	Obliczone metodą			Wartość dośw.
	HF	MP2	LDA	
$\text{C}_2\text{H}_2$	3719	3593	3441	3374
	3607	3516	3343	3289
	2247	2006	2011	1974
	883	783	705	730
	794	444	560	612
$\text{CH}_2\text{NH}$	3719	3463	3355	3297
	3347	3254	3040	3036
	3254	3116	2937	2924
	1901	1724	1690	1640

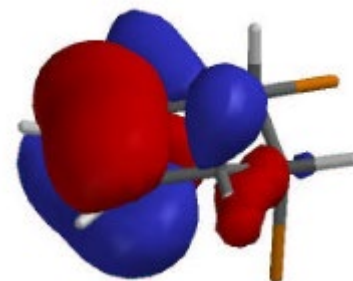
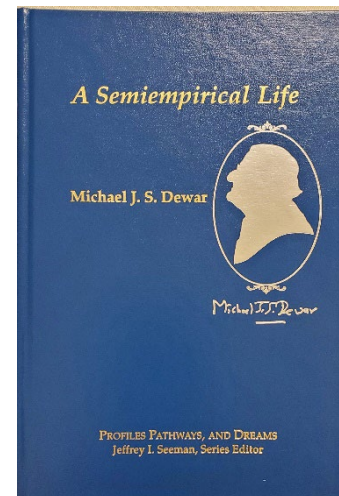


- Molecular Mechanics...
- ✓ Semi-empirical...
- Ab initio...
- Density Functional...
- Periodic Box...
- Restrains...
- Set Velocity...
- Set Finite Field...
- Network...
- Edit Parameters...
- Select Parameter Set...
- Compile Parameter File
- Reaction Map...



# Metody półempiryczne

- ❖ Pominięcie bezpośredniego rozpatrywania elektronów niewalencyjnych
- ❖ całki rezonansowe  $H_{\text{core}}$  są empirycznymi funkcjami parametrycznymi
- ❖ Parametry są dobierane aby najlepiej odtwarzać wartości eksperymentalne (np. ciepło tworzenia)
- ❖ Czas CPU jest proporcjonalny do  $n^2$  , gdzie  $n$  odpowiada liczbie funkcji bazowych

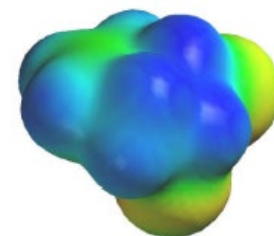


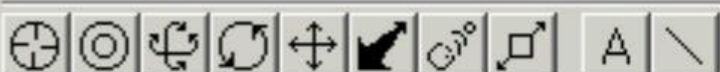


# Porównanie metod półempirycznych

## Średnie wartości błędów dla metod NNDO

Metoda	MNDO	AM1	PM3	PM5
Ciepło tworzenia (kcal/mol)	18.4	11.9	9.9	5.9
Długość wiązań ( Å)	0.066	0.053	0.065	0.051
Kąty (stopnie)	6.3	5.5	5.7	5.4
Moment dipolowy (D)	0.71	0.49	0.57	0.66
Potencjały jonizacji (eV)	0.92	0.72	0.75	0.56





✓ Molecular Mechanics...

Semi-empirical...

Ab Initio...

Density Functional...

Periodic Box...

Restrains...

Set Velocity...

Set Finite Field...

Network...

Edit Parameters...

Sele

Com

Rea

## Molecular Mechanics Force Field

Method

 MM+ AMBER BIO+(CHARMM) OPLS

Options...

Components...

OK

Cancel

# Mechanika molekularna

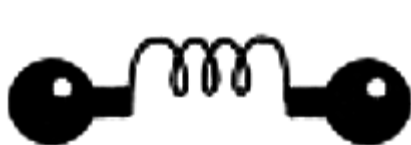
Metoda przewidywania stabilnej konfiguracji poprzez minimalizację elektronowej energii układu cząsteczek

$$\begin{aligned}
 E_{\text{pot}} = & \sum_b K_2(b - b_0)^2 + \sum_{\theta} H_{\theta}(\theta - \theta_0)^2 + \sum_{\phi} \frac{V_n}{2}[1 + \cos(n\phi - \phi_0)] \\
 & + \sum \varepsilon[(r^*/r)^{12} - 2(r^*/r)^6] + \sum q_i q_j / \varepsilon_{ij} r_{ij} + \sum \left[ \frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right]
 \end{aligned}$$

(1)
(2)
(3)

(4)
(5)
(6)

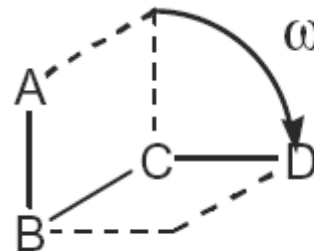
Pole siłowe – to wyrażenie opisujące energię układu jako położenia jej jąder



(1)



(2)

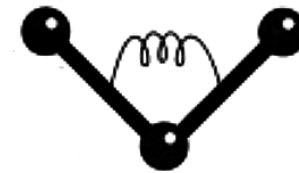


(3)

# Pole siłowe Amber

## Prosta funkcja harmoniczna

$$E_{stretch} = \sum_{\substack{\text{bonds} \\ AB}} \frac{1}{2} k_{AB} (r_{AB} - r_{AB}^0)^2$$

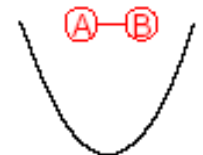


Stała  
siłowa

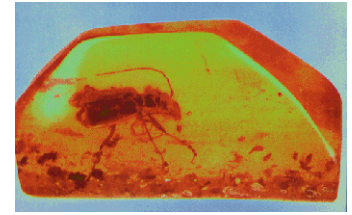
słabnie

Długość  
wiązania

$r_{AB}^0$

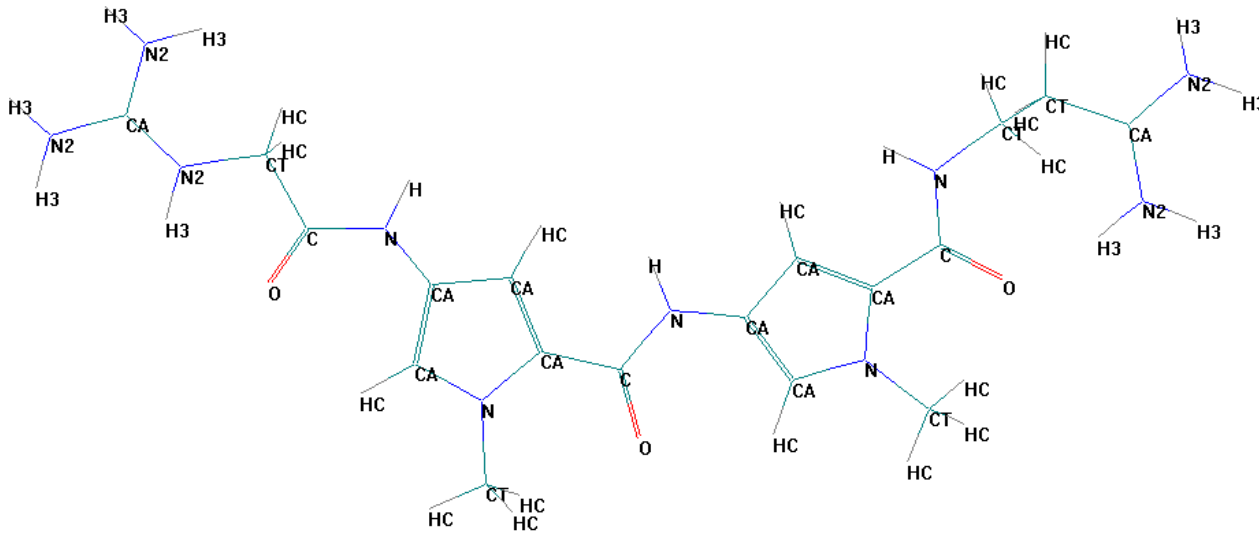


# Pole siłowe Amber



Rozważamy tylko stany podstawowe

- ❖ Wybór „typu atomu” (parametryzacja) jest podstawowy dla wyników obliczeń: np. AMBER ma 5 typów tlenu: karbonyłowy, alkoholowy, kwasowy, estrowy/eterowy i występujący w wodzie



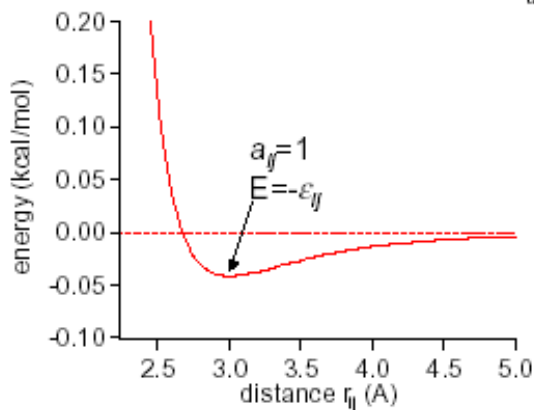
**Netropsyna**



# Pole siłowe Amber: oddziaływania niewiążące

## Oddziaływania Van der Waalsa

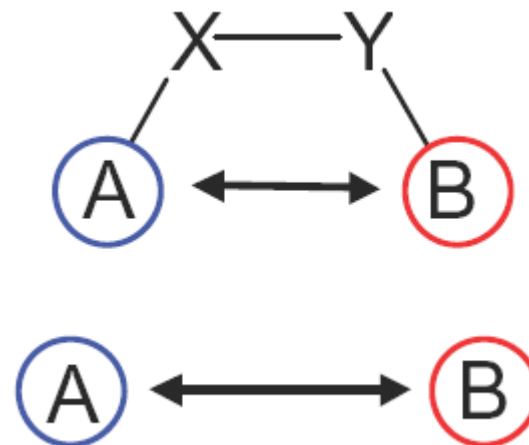
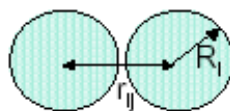
Lennard-Jones



$$E_{vdW} = \sum_{\substack{\text{non-bonded} \\ \text{atom\_pairs}}} \epsilon_{ij} (a_{ij}^{-12} - 2a_{ij}^{-6})$$

$$\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}$$

$$a_{ij} = \frac{r_{ij}}{R_i + R_j}$$

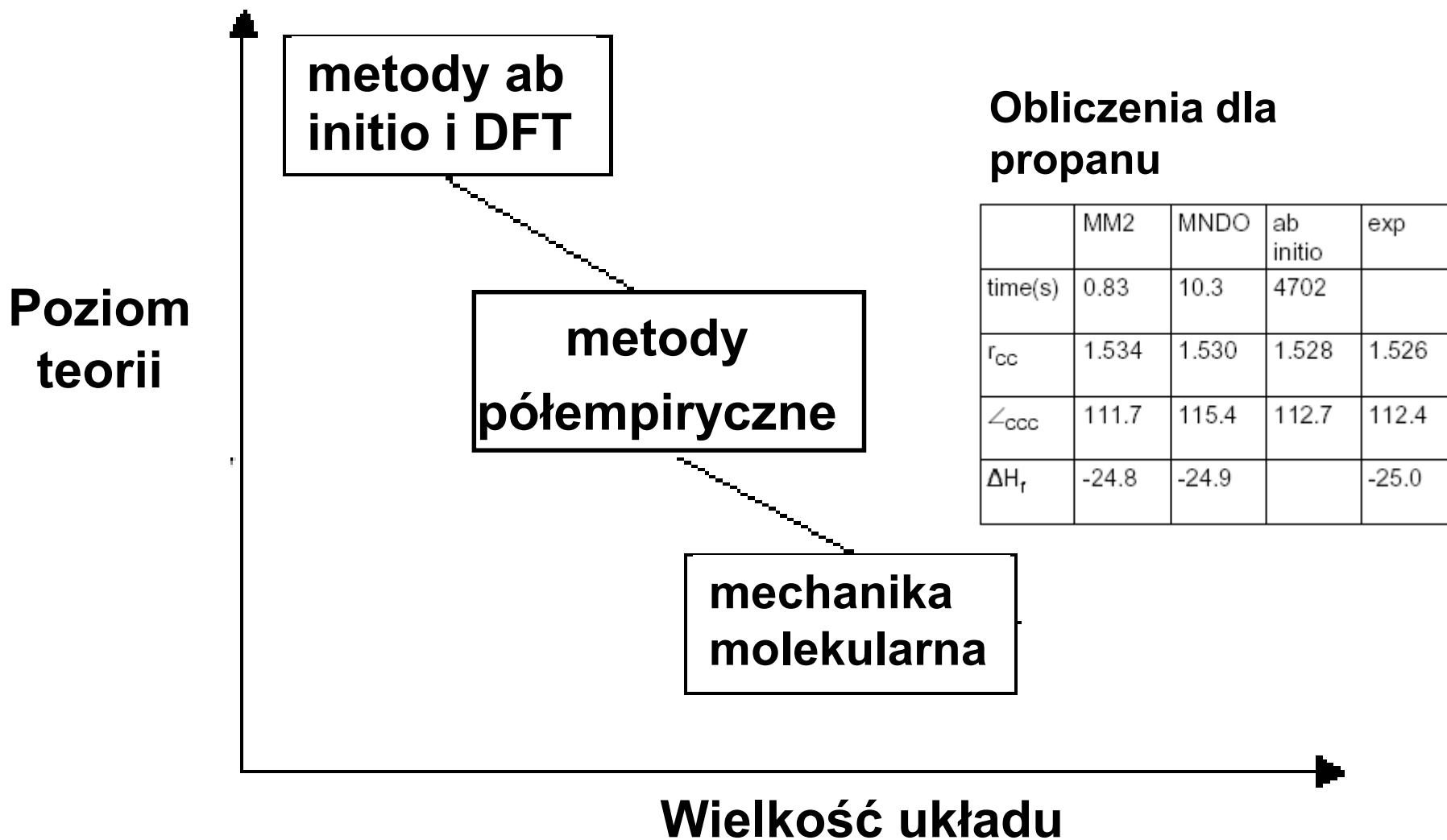


## Oddziaływania elektrostatyczne

$$E_{\text{electrostatic}} = 332.1 \sum_{\substack{\text{atom} \\ \text{pairs} \\ i,j}} \frac{q_i q_j}{\epsilon_{ij} r_{ij}} \quad (\text{kcal/mol})$$

Wartości  $q_i$  i  $q_j$  wyznaczamy metodami ab initio i półempirycznymi

# Porównanie metod obliczeniowych



## Porównanie metod: czas obliczeń i jakość wyników

	MM2	MNDO	ab initio	exp
time(s)	0.83	10.3	4702	
$r_{CC}$	1.534	1.530	1.528	1.526
$\angle_{CCC}$	111.7	115.4	112.7	112.4
$\Delta H_f$	-24.8	-24.9		-25.0

***Właściwości propanu***



Zakład Syntezy i Technologii Środków Leczniczych  
Uniwersytet Medyczny w Białymstoku  
Mickiewicza 2, 15-230 Białystok  
tel. (085) 748-57-01  
fax (48-85) 748-54-16  
E-mail: Krzysztof.Bielawski@umb.edu.pl

