

## Ćwiczenie (2h): Od RCSB do filmu z dynamiki – wizualizacja, optymalizacja i mini-MD

*Modelowanie molekularne – studia podyplomowe*

Struktura: 2NWD (Lizozym) | Wersja: 13 grudnia 2025

### Cel ćwiczenia

W tym ćwiczeniu przejdziesz pełny, minimalny pipeline (potok danych):

1. **RCSB PDB** – zapoznasz się z białkową bazą danych, odczytasz metadane i obejrzysz strukturę w Mol\* (przeglądarka struktur 3D na stronie RCSB),
2. **AlphaFold** – porównasz strukturę białka przewidzianą *de novo* przez model AI (AlphaFold) ze strukturą uzyskaną przy wykorzystaniu metod eksperymentalnych,
3. **GROMACS** – przygotujesz układ symulacyjny, wykonasz jego optymalizację poprzez *minimizację* jego energii potencjalnej i przeprowadzisz krótką symulację dynamiki molekularnej (*MD*),
4. **PyMOL** – zwizualizujesz trajektorię (render, animacja),
5. **Analiza** – sprawdzisz stabilność układu: energia, RMSD (root mean square distance - średnie przesunięcie kwadratowe).

### Wykorzystywane oprogramowanie

- Przeglądarka: Firefox/Chrome/Safari (WebGL).
- PyMOL 2.5+ (zalecane 3.x).
- GROMACS 2022+ (polecenia `gmx`).

## 1 Część 1 (RCSB): jak czytać wpis PDB? (15–20 min)

### PODSTAWY TEORETYCZNE

**Dlaczego zaczynamy od RCSB?** Struktura w PDB nie jest „prawdziwym obiektem 3D”, tylko modelem zbudowanym na podstawie danych eksperymentalnych. Dlatego zawsze pytamy:

- **metoda** (np. X-ray, cryo-EM) i **rozdzielczość** (jakość szczegółów),
- **co jest w modelu** (łańcuchy, ligandy, braki, modyfikacje),
- **jak wygląda walidacja** (czy model ma problemy).

To ogranicza ryzyko, że uruchomisz symulację na strukturze „problematicznej” (braki, ligandy bez parametrów, dziwne atomy).

## 1A. Otwórz wpis 2NWD w RCSB

Wejdź na [rcsb.org](https://www.rcsb.org) i wyszukaj 2NWD.

### KLIKNIJ W INTERFEJSIE

**RCSB:** Search → wpisz 2NWD → kliknij wynik.

### ZADANIE NA ZAJĘCIACH

**Zadanie 1 (metadane w 5 minut).** Znajdź i wypisz:

- metodę eksperymentalną i rozdzielczość,
- liczbę łańcuchów białkowych,
- czy są ligandy/kofaktory (w tym ćwiczeniu celowo wybieramy strukturę bez nich),
- źródło pochodzenia: organizm.

### CHECKLIST (co ma wyjść?)

Po tej części powinienś mieć:

- 4 krótkie punkty tekstu (metoda, rozdzielczość, łańcuchy, ligandy),
- ogólny obraz tego, co w ogóle zawiera wpis PDB.

## 1B. Obejrzyj strukturę w Mol\*

### KLIKNIJ W INTERFEJSIE

Zakładka **Structure** → **3D View (Mol\*)**. W Mol\* znajdź panel **Components** oraz sprawdź predefiniowane reprezentacje białka (Preset/Automatic). Kliknij w dowolny aminokwas na strukturze białka a następnie przejdź do Density, kliknij Enable - w ten sposób wyświetlasz mapę gęstości elektronowej w sąsiedztwie zaznaczonego aminokwasu. Wyłącz wyświetlanie mapy różnicowej (kliknij na symbol oka przy Fo-Fc\*)

### ZADANIE NA ZAJĘCIACH

**Zadanie 2 (Mol\* – szybkie oglądanie).** Zrób zrzut ekranu: białko w reprezentacji cartoon, tło jasne.

### DO KRÓTKIEGO RAPORTU

**Do raportu:** 1 obrazek z Mol\* + 1-2 zdania wyjaśniające dlaczego atomy wodoru są często pomijane w strukturach rozwiązanych metodą dyfrakcji rentgenowskiej (Wskazówka: spójrz na mapę gęstości elektronowej).

### CHECKLIST (co ma wyjść?)

Po tej części masz:

- 1 zrzut ekranu Mol\*,
- świadomość: „jak wygląda białko” i dlaczego często nie ma w nim atomów H.

## 2 Część 2 (AlphaFold): AI vs eksperyment (15–20 min)

### PODSTAWY TEORETYCZNE

**Co porównujemy?** AlphaFold przewiduje strukturę z sekwencji. Eksperyment (tu: X-ray) daje model oparty o dane. Porównanie ma sens, bo:

- regiony elastyczne (pętle, końce) często mają niższą pewność w AF (pLDDT),
- model eksperymentalny może zawierać stabilizacje wynikające z warunków krystalizacji,
- różnice uczą interpretacji: *gdzie ufać AI, a gdzie zachować ostrożność.*

**pLDDT (0–100):** im wyżej, tym model AI pewniejszy lokalnie. W plikach AF pLDDT bywa zapisany w kolumnie B-factor.

### 2A. Na stronie RCSB dla Znajdź powiązanie PDB → UniProt

#### KLIKNIJ W INTERFEJSIE

RCSB wpisz 2NWD → sekcja **Structure Summary / Protein** → znajdź identyfikator **UniProt**. Następnie przejdź na stronę [AlphaFold Protein Structure Database](#) i wyszukaj skopiowany wcześniej identyfikator UniProt).

Następnie ściągnij sktstrukturę (Download files/PDB File) i zmień nazwę pliku na AF.pdb

### 2A-1. Dlaczego w AlphaFold czasem widać „długi ogon”, którego nie ma w PDB?

#### PODSTAWY TEORETYCZNE

**PDB vs AlphaFold:** porównujemy inne „warianty biologiczne” tej samej cząsteczki.

Wpis PDB zwykle przedstawia **dojrzałą, funkcjonalną formę białka** – czyli taką, jaka występuje po obróbce potranslacyjnej (np. po odcięciu sekwencji sygnałowej) i w warunkach eksperymentu. Z kolei AlphaFold zazwyczaj modeluje **pełną sekwencję z UniProt** (produkt genu), która może zawierać dodatkowe fragmenty:

- **signal peptide** (peptyd sygnałowy) – kieruje białko do szlaku sekrecyjnego; jest hydrofobowy i **zwykle ucinany** w komórce,
- **pro-peptyd** – fragment aktywujący/dojrzewający; także może być odcinany,
- fragmenty **nieustrukturyzowane** (IDR) – elastyczne, bez jednej stabilnej konformacji.

Dlatego w AlphaFold możesz zobaczyć długi, cienki fragment (często żółto–czerwony przy kolorowaniu pLDDT), którego **nie ma w strukturze eksperymentalnej** użytej w PDB.

### 2B. Porównanie w PyMOL (dopasowanie + kolor pLDDT)

Otwórz program terminal (kliknij w lewym dolnym rogu ekranu), a następnie uruchom program do wizualizacji makrocząsteczek Pymol wpisując to słowo w terminal.

W PyMOL otwórz pobraną z AlphaFold strukturę (File/Open) a następnie wpisz komendy:

```
fetch 2NWD, type=pdb, async=0

# dopasowanie po atomach CA (wystarczy do ćwiczenia):
align AF and name CA, 2NWD and name CA

# kolorowanie AF po pLDDT (B-factor):
spectrum b, blue_white_red, AF, 50, 100
select low_conf, AF and b < 90
show spheres, low_conf

bg_color white
hide lines
```

### ZADANIE NA ZAJĘCIACH

**Zadanie 3 (3 obserwacje).** Napisz:

- gdzie AF ma niski pLDDT (jakie fragmenty: pętle/końce/rdzeń),
- czy po dopasowaniu widać obszary większego rozjazdu.

### CHECKLIST (co ma wyjść?)

Po tej części masz:

- nałożone struktury w PyMOL,
- zaznaczone regiony niskiej pewności (kulki),
- krótką interpretację AI vs eksperyment.

## 3 Część 3 (GROMACS): przygotowanie układu do symulacji (25–35 min)

### PODSTAWY TEORETYCZNE

**Dlaczego nie da się zrobić MD „na PDB”?** PDB nie zawiera wszystkich informacji potrzebnych do obliczeń energii i sił (ładunki, typy atomów, parametry wiązań itd.). GROMACS musi stworzyć:

- topologię (topol.top) i parametry sił pola,
- układ w pudełku periodycznym,
- wodę i jony,
- stan startowy bez kolizji (minimizacja).

### 3A. Katalog roboczy i pobranie PDB

Otwórz nowe okno terminala i wpisz poniższe komendy (najlepiej po jednej)

```
mkdir -p ModMol_Postgrad_2NWD && cd ModMol_Postgrad_2NWD
wget https://files.rcsb.org/download/2NWD.pdb
```

### 3B. pdb2gmx – topologia białka

#### PODSTAWY TEORETYCZNE

Co wybieramy w **pdb2gmx** i dlaczego?

**Force field (FF)** to zestaw parametrów opisujących energię układu (wiązania, kąty walencyjne, kąty torsyjne, oddziaływania elektrostatyczne i van der Waalsa). **Model wody** (TIP3P, SPC/E itd.) jest częścią tego samego „języka parametrów” i powinien być dobrany spójnie z FF.

W tym ćwiczeniu chcemy wybór:

- powszechny w symulacjach białek,
- stabilny i niezawodny,
- spójny (FF + woda).

Dlatego wybieramy: **AMBER99SB-ILDN** (dla białek) oraz **TIP3P** (woda).

#### KLIKNIJ W INTERFEJSIE

Uruchom polecenie:

```
gmx pdb2gmx -f 2NWD.pdb -o processed.gro -p topol.top -ignh
```

Gdy program poprosi o wybór:

- **Force Field:** wybierz **AMBER99SB-ILDN** (na Twojej liście: 6)
- **Water Model:** wybierz **TIP3P** (na liście modeli wody: 1)

#### CHECKLIST (co ma wyjść?)

Po poprawnym wykonaniu **pdb2gmx** powinieneś mieć:

- plik **processed.gro** (wstępnie przetworzona struktura),
- plik **topol.top** (topologia),
- plik **posre.itp** (więzy pozycyjne, gdybyśmy chcieli unieruchomić białko w trakcie symulacji).

Jeśli **processed.gro** lub **topol.top** nie powstały — nie idź dalej: przewiń terminal do góry i znajdź pierwszy komunikat **Fatal error**.

## PODSTAWY TEORETYCZNE

### Dlaczego potrzebujemy pudełka symulacyjnego?

Symulacje dynamiki molekularnej są wykonywane w tzw. **warunkach periodycznych** (PBC – *periodic boundary conditions*). Oznacza to, że symulowany układ jest matematycznie powielany we wszystkich kierunkach przestrzeni, tak aby:

- uniknąć „efektu ściany” (białko nie odbija się od nich),
- zasymulować środowisko zbliżone do nieskończonego roztworu,
- umożliwić poprawne obliczanie oddziaływań długozasięgowych (elektrostatyka).

**Pudełko symulacyjne** to więc objętość, która:

- zawiera cząsteczkę białka,
- jest wypełniona cząsteczkami wody,
- jest wystarczająco duża, aby białko **nie oddziaływało samo ze sobą** przez granice periodyczne.

W praktyce przyjmuje się bufor **1.0 nm** między białkiem a ścianą pudełka jako bezpieczne minimum dla krótkich symulacji dydaktycznych.

## 3C. Pudełko + solwatacja

### PODSTAWY TEORETYCZNE

#### Dlaczego dodekaedr?

W tym ćwiczeniu używamy pudełka typu **dodecahedron**, ponieważ:

- lepiej „otula” kuliste/owalne białka niż sześcian,
- wymaga **mniej cząsteczek wody** niż klasyczne pudełko prostokątne,
- jest standardowym wyborem w symulacjach białek globularnych.

Mniej wody = mniej atomów = szybsza symulacja, bez utraty poprawności fizycznej.

Tworzymy pudełko symulacyjne zawierające cząsteczkę białka i pustą przestrzeń wypełniamy cząsteczkami wody

```
gmx editconf -f processed.gro -o boxed.gro -c -d 1.0 -bt dodecahedron
gmx solvate -cp boxed.gro -cs spc216.gro -o solvated.gro -p topol.top
```

### UWAGA / TYPowe PROBLEMY

**Uwaga praktyczna (częsty błąd):** Po solwatacji liczba cząsteczek wody jest zapisywana w pliku `topol.top`. Jeśli po tym kroku ręcznie modyfikujesz strukturę i ponownie uruchamiasz solwatację, **musisz zaktualizować topologię**, inaczej GROMACS zgłosi niespójność.

### 3D. Jony (neutralizacja)

#### PODSTAWY TEORETYCZNE

##### Dlaczego neutralizujemy ładunek białka?

Białka niemal zawsze mają **ładunek netto** (wynikający z aminokwasów kwaśnych i zasadowych). W symulacjach MD z periodycznymi warunkami brzegowymi i elektrostatyką typu PME:

- układ o niezerowym ładunku jest **niefizyczny**,
- algorytmy elektrostatyczne zakładają globalną neutralność,
- brak neutralizacji może prowadzić do artefaktów energetycznych.

Dlatego do układu dodaje się **przeciwjony** (np.  $\text{Na}^+$  lub  $\text{Cl}^-$ ), aby:

- całkowity ładunek układu wynosił zero,
- symulacja była stabilna i poprawna fizycznie.

W tym ćwiczeniu:

- **nie modelujemy konkretnego stężenia soli**,
- interesuje nas tylko neutralizacja ładunku.

```
gmx grompp -f em.mdp -c solvated.gro -p topol.top -o ions.tpr -maxwarn 1
gmx genion -s ions.tpr -o solv_ions.gro -p topol.top -pname NA -nname CL -neutral
```

Wybierz grupę do podmiany: **SOL**.

#### UWAGA / TYPOWE PROBLEMY

**Dlaczego używamy `-maxwarn 1`?** Na tym etapie GROMACS może zgłosić drobne ostrzeżenia dotyczące niezerowego ładunku układu. Ignorujemy to ostrzeżenie, ponieważ druga komenda i tak dodała brakujące przeciwjony.

#### CHECKLIST (co ma wyjść?)

W folderze powinny pojawić się (minimum):

- `processed.gro`, `topol.top`
- `boxed.gro`, `solvated.gro`
- `ions.tpr`, `solv_ions.gro`

Jeśli brakuje któregoś pliku, nie idź dalej — wróć o krok i sprawdź komunikat błędu.

## 4 Część 4: minimalizacja energii (8–12 min)

#### PODSTAWY TEORETYCZNE

**Po co minimalizacja?** Po dodaniu wody/jonów mogą powstać złe kontakty (atomy „za blisko”), które w MD powodują niefizyczne siły i niestabilność. Minimalizacja obniża energię potencjalną i usuwa potencjalne kolizje.

## UWAGA / TYPOWE PROBLEMY

Przekopiuj ściągnięte wcześniej z github'a pliki do podkatalogu Mod-Mol\_Postgrad\_2NWD w katalogu domowym

W tym celu najlepiej użyj przeglądarki plików

Wpisz komendy:

```
gmx grompp -f em.mdp -c solv_ions.gro -p topol.top -o em.tpr
gmx mdrun -v -deffnm em
```

Kontrola energii:

```
gmx energy -f em.edr -o em_potential.xvg
# wpisz numer przy: Potential i zatwierdź dwukrotnie klawiszem Enter
```

Obejrzyj wykres zmian energii potencjalnej układu

```
xmgrace em_potential.xvg &
```

## CHECKLIST (co ma wyjść?)

Po minimizacji powinieneś mieć:

- em.gro, em.log, em.edr
- wykres em\_potential.xvg z trendem stabilnym (bez eksplozji)

## 5 Część 5: krótka MD (NVT + MD) (20–25 min)

### PODSTAWY TEORETYCZNE

Co sprawdzamy krótką MD? To nie jest „badanie mechanizmu choroby”, tylko nauka:

- jak uruchomić trajektorię,
- jak poprawnie ją przygotować (PBC, centrowanie),
- jak ocenić podstawową stabilność (energia, RMSD, RMSF).



## 5A. NVT

### PODSTAWY TEORETYCZNE

#### Co oznacza NVT?

Skrót **NVT** oznacza zespół statystyczny (ensemble), w którym:

- **N** – liczba cząstek jest stała,
- **V** – objętość pudełka jest stała,
- **T** – temperatura jest kontrolowana.

Etap NVT służy do **termalizacji układu**, czyli:

- nadania atomom realistycznych prędkości odpowiadających zadanej temperaturze (np. 300 K),
- łagodnego „rozruchu” dynamiki po minimizacji energii,
- uniknięcia gwałtownych zmian struktury na początku symulacji.

**Ważne:** NVT nie służy do analizy ruchów białka — to etap przygotowawczy, a nie produkcyjny.

Wpisz komendy:

```
gmx grompp -f nvt.mdp -c em.gro -r em.gro -p topol.top -o nvt.tpr
gmx mdrun -v -deffnm nvt
```

## 5B. MD (krótka produkcja)

### PODSTAWY TEORETYCZNE

#### Czym jest „produkcja” w MD?

Etap **produkcji** to właściwa symulacja dynamiki molekularnej, w której:

- zapisujemy trajektorię ruchu atomów,
- analizujemy stabilność struktury (RMSD, RMSF),
- obserwujemy naturalne fluktuacje białka.

W klasycznym protokole symulacyjnym często stosuje się:

- **NVT** → **NPT** → **produkcja NPT**.

**Dlaczego w tym ćwiczeniu pomijamy NPT?** Aby:

- uprościć pipeline i zmieścić się w 2 godzinach,
- uniknąć problemów ze stabilizacją ciśnienia,
- skupić się na mechanice MD, a nie na parametryzacji.

Wpisz komendy:

```
gmx grompp -f md.mdp -c nvt.gro -p topol.top -o md.tpr
gmx mdrun -v -deffnm md
```

## 5C. PBC fix do wizualizacji

```
gmx trjconv -s md.tpr -f md.xtc -o md_noPBC.xtc -pbc mol -center
```

Wybierz do centrowania: **Protein**. Do zapisu: **System**.

### CHECKLIST (co ma wyjść?)

Po MD powinieneś mieć:

- md.xtc (trajektoria), md.tpr (parametry), md.edr (energije)
- md\_noPBC.xtc (trajektoria do filmu)

## 6 Część 6 (PyMOL): film i obrazek (10–15 min)

### PODSTAWY TEORETYCZNE

**Dlaczego film jest ważny?** Film uczy intuicji: które fragmenty są ruchliwe i jak wygląda stabilna trajektoria. Nie wyciągamy z tego wniosków biologicznych „na poważnie”, ale uczymy się czytać dynamikę.

W PyMOL’u wpisujemy (upewniając się, że pymol został uruchomiony w katalogu, w którym znajdują się potrzebne pliki):

```
load md.gro, MD
load md_noPBC.xtc, MD
# aby odtworzyć film z trajektorii, kliknij na przycisk Play w prawym górnym rogu
# okna Pymola

# aby ukryć cząsteczki wody
hide lines
# aby ukryć przeciwjony
hide spheres
# aby zmienić tło na białe
bg_color white
# aby przyspieszyć/zwolnić film
set movie_fps, 35

# aby wyrenderować ładny obrazek
ray 1600,1200
# aby go zapisać (uważać aby nie poruszyć struktury)
png md_moja_nazwa.png, dpi=300
```

## 7 Część 7: prosta analiza (energia, RMSD, RMSF) (10–15 min)

### PODSTAWY TEORETYCZNE

#### Co właściwie analizujemy po krótkiej symulacji MD?

Celem analizy nie jest „odkrycie biologii lizozymu”, lecz **sprawdzenie poprawności i stabilności symulacji** oraz nauczenie się, jak czytać podstawowe wykresy diagnostyczne.

W tej części skupiamy się na trzech wielkościach, które odpowiadają na trzy różne pytania:

- **Energia potencjalna (Potential)** – czy układ jest fizycznie stabilny i nie „eksploduje” numerycznie,
- **RMSD** – czy struktura globalnie się stabilizuje, czy nadal znacząco zmienia konformację,
- **RMSF** – które fragmenty białka są bardziej elastyczne, a które sztywne.

Każdy z tych wykresów opisuje *inny aspekt* dynamiki i nie należy ich interpretować zamiennie.

### UWAGA / TYPOWE PROBLEMY

Dane do wykresów znajdują się w plikach z rozszerzeniem .xvg.

Aby wygenerować wykres, w **Terminal** wpisujemy:

```
xmgrace nazwa_pliku.xvg &
```

## 7A. Energia potencjalna

### PODSTAWY TEORETYCZNE

#### Energia potencjalna – test stabilności numerycznej

Energia potencjalna w MD jest sumą:

- energii wiązań, kątów walencyjnych i torsyjnych,
- oddziaływań van der Waalsa,
- oddziaływań elektrostatycznych.

W krótkiej symulacji dydaktycznej:

- **nie oczekujemy stałej energii** (to nie NVE),
- oczekujemy **stabilnych fluktuacji wokół pewnego poziomu**,
- brak długotrwałego dryftu oznacza poprawnie przygotowany układ.

#### Nieprawidłowe zachowania:

- gwałtowny spadek energii – często oznaka złych kontaktów,
- niekontrolowany wzrost energii – możliwa niestabilność symulacji.

W Terminalu wpisz:

```
gmx energy -f md.edr -o md_potential.xvg
# wybierz: Potential
```

## 7B. RMSD ( $C\alpha$ )

### PODSTAWY TEORETYCZNE

**Czym jest RMSD i co nam mówi?**

RMSD (Root Mean Square Deviation) mierzy **średnie przesunięcie atomów** względem struktury referencyjnej po dopasowaniu geometrycznym.

W tym ćwiczeniu:

- analizujemy RMSD atomów  $C\alpha$ ,
- dopasowujemy każdą klatkę do struktury referencyjnej,
- obserwujemy **globalne zmiany konformacyjne**.

**Typowy przebieg RMSD w poprawnej symulacji:**

- szybki wzrost na początku (relaksacja struktury),
- osiągnięcie plateau,
- fluktuacje wokół stałej wartości.

**Ważne:** Wzrost RMSD *nie oznacza* „rozpadu białka” – często oznacza jedynie odejście od struktury krystalograficznej do bardziej naturalnej konformacji w roztworze.

```
gmx rms -s md.tpr -f md_noPBC.xtc -o rmsd_ca.xvg -tu ps
```

Wybierz: **C-alpha** jako referencję i jako grupę dopasowania.

## 7C. RMSF (na resztę)

### PODSTAWY TEORETYCZNE

**RMSF – które fragmenty białka się poruszają?**

RMSF (Root Mean Square Fluctuation) mierzy:

- jak bardzo **każda reszta** odchyła się od swojej średniej pozycji,
- lokalną elastyczność struktury w trakcie symulacji.

Typowe obserwacje:

- pętle i końce łańcucha – wysokie RMSF,
- helisy i rdzeń białka – niskie RMSF,
- mostki disiarczkowe często stabilizują lokalnie RMSF.

**RMSF nie mówi o globalnej stabilności**, lecz o *lokalnej ruchliwości*. Dlatego zawsze interpretujemy RMSF razem z RMSD.

```
gmx rmsf -s md.tpr -f md_noPBC.xtc -o rmsf_res.xvg -res
```

## ZADANIE NA ZAJĘCIACH

### Zadanie końcowe.

1. Czy Potential wygląda stabilnie?
2. Jak zachowuje się RMSD?
3. Co pokazuje RMSF - znajdź najbardziej ruchliwą resztę.

## DO KRÓTKIEGO RAPORTU

### Raport 1 strona (PDF)

- Mol\* screenshot + 2–3 zdania,
- AF vs eksperyment screenshot + 3–5 zdań,
- MD snapshot + (film lub informacja o klatkach),
- Krótka analiza (Potential, RMSD, RMSF).

**Uwaga dydaktyczna:** To celowo uproszczony protokół (bez NPT), żeby zmieścić się w 2h i nauczyć pipeline. W praktyce produkcyjnej dodaje się NPT, dłuższe czasy, walidację stabilności i świadomy dobór parametrów.