

Bioinformatyka: UNRES - symulacje struktury i dynamiki białek przy użyciu podejścia scalonych reszt (UNited RESidue) opartego na fizyce oddziaływań

Krótki opis usługi

Usługa UNRES jest przyjazną dla użytkownika realizacją [pakietu modelowania gruboziarnistego białek, opartego na polu siłowym UNRES](#) opracowanym w wyniku współpracy grupy prof. Adama Liwo z Wydziału Chemii Uniwersytetu Gdańskiego oraz grupy prof. Harolda A. Scheragi z Cornell University, USA. Dzięki znacznemu uproszczeniu reprezentacji łańcucha polipeptydowego przy jednoczesnym ścisłym związku wyprowadzonego pola siłowego z fizyką oddziaływań symulacje są przyspieszone 1000 - 10000 razy w stosunku do obliczeń w reprezentacji pełnoatomowej. Adresatami usługi są biofizycy, biologowie molekularni i bioinformatycy, zainteresowani symulowaniem zmian strukturalnych, dynamiki i termodynamiki dużych układów białkowych w czasie rzeczywistym. W szczególności usługa będzie przydatna dla badaczy, którzy chcą wykonać wspomniane wielkoskalowe symulacje białek, natomiast brak czasu nie pozwala im na zapoznanie się z uruchamianiem zadań bezpośrednio na serwerach przy użyciu systemu kolejkowania.

W usłudze dostępne są dwie wersje pola siłowego. Wersja pierwsza, ogólna, oznaczana dalej jako E0LL2Y, umożliwia obliczenia dla białek o wszystkich typach struktury drugorzędowej, jednak wymaga użycia metod przewidywania struktury drugorzędowej i nałożenia słabych więzów na tę strukturę (He et al., J. Comput. Chem., 2009, 30, 2127-2135). Wersja druga, oznaczana dalej jako GAB, umożliwia obliczenia dla białek o strukturze helikalnej (Liwo et al., J. Phys. Chem. B, 2007, 111, 260-285).

Usługa składa się z następujących elementów:

- UNRES: właściwe symulacje,
- WHAM : przetwarzanie wyników dynamiki molekularnej z wymianą replik metodą analizy ważonych histogramów,
- CLUSTER: analiza skupień wyników symulacji; analiza wyników dynamiki molekularnej z wymianą replik wymaga wstępnego przetworzenia metodą analizy ważonych histogramów.

Usługa UNRES umożliwia realizację następujących typów obliczeń:

- Obliczanie i lokalna minimalizacja energii cząsteczki białka.
- Kanoniczne symulacje gruboziarniste dynamiki białek.
- Symulacje białek przy użyciu metody wymiany replik lub zwielokrotnionej wymiany replik i konstrukcja na tej podstawie profili zależności wielkości geometrycznych, mechanicznych i termodynamicznych od temperatury, badanie krajobrazów energii swobodnej białek, termodynamiki zwijania, itp. W tym przypadku konieczne jest użycie programu WHAM.
- Przewidywanie struktury białek oparte o fizykę oddziaływań. W przypadku przewidywania struktury metodą znajdowania najbardziej prawdopodobnych zespołów statystycznych (patrz rozdział **Zaawansowane użycie**), konieczne jest przetworzenie wyników symulacji dynamiki z wymianą replik programami WHAM i CLUSTER.

Z usługi można korzystać na dwa sposoby: za pośrednictwem interaktywnego portalu UNRES oraz przy pomocy systemu wsadowego [QosCosGrid](#).

Aktywowanie usługi

Aby skorzystać z usługi UNRES należy posiadać [konto](#) w portalu PLGRID oraz posiadać aktywny [grant właściwy PLGrid](#). Należy również złożyć wniosek o usługę [UNRES w Katalogu Aplikacji i Usług](#). Spowoduje to dodanie użytkownika do zespołu *plggunresng* (zespół użytkowników usługi UNRES).

Ograniczenia w korzystaniu

Obecnie aplikacje składające się na usługę są zainstalowane w ramach infrastruktury PLGrid na klastrach:

- Inula w Poznańskim Centrum Superkomputerowo-Sieciowym
- Prometheus w Cyfronecie.

Portal QCG umożliwiający interaktywne uruchamianie zadań kieruje zadania tylko na klaster inula. Za pośrednictwem portalu można uruchomić tylko główny moduł obliczeniowy UNRES, natomiast moduły analizy wyników dynamiki molekularnej z wymianą replik można uruchomić jedynie bezpośrednio w trybie wsadowym. Rozszerzenie funkcjonalności portalu QCG na ten moduły jest w toku.

Ponieważ w chwili obecnej system QosCosGrid nie obejmuje kierowania zadań na klaster Prometheus, w przypadku uruchamiania zadań na tym komputerze należy używać systemu slurm, podobnie jak w przypadku uruchamiania innych zadań na tym komputerze. Szczegółowy opis znajduje się w podręczniku [Prometheus:Podstawy](#).

Pierwsze kroki

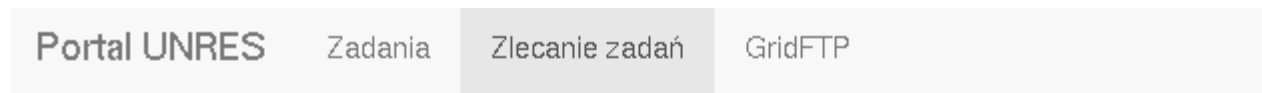
Zalecane jest zapoznanie się z [dokumentacją](#) części składowych pakietu UNRES dostępną na stronie pakietu (dokumentacja w języku angielskim).

Uruchamianie przez portal UNRES

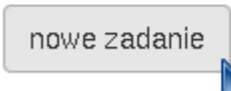
Zlecanie zadań przez portal UNRES możliwe jest tylko na klaster inula w PCSS.

Portal UNRES jest dostępny pod adresem <https://unres-portal.plgrid.pl>. Portal jest zintegrowany z OPENID. Należy zalogować się na portal używając identyfikatora użytkownika w systemie PL-GRID.

W celu zlecenia zadania z poziomu portalu UNRES należy wybrać zakładkę "Zlecanie zadań", a następnie przycisk "nowe zadanie":



Zlecanie zadań

Zleć  lub wybierz z listy zapisanych szablonów poniżej.

Zapisane szablony zadania

 Brak zapisanych szablonów zadań

W celu zlecenia dynamiki molekularnej należy w zakładce "MD" :

- ustawić pole siłowe (GAB lub E0LL2Y)
- ustawić liczbę kroków dynamiki
- wybrać plik PDB będący strukturą referencyjną
- podać sekwencję aminokwasów w pliku PDB w zapisie jednoliterowych uwzględniając aminokwasy terminalne w [zapisie pakietu UNRES](#).

Portal UNRESZadaniaZlecanie zadańGridFTPDavid Jagieła

Zleć zadanieZapisz jako szablon

MDZasobyPlikiŚrodowiskoInne

Pole siłoweE0LLZY

NSTEP500000
Liczba kroków w trajektorii

Plik PDBqcg.man.poznan.pl/home/plgrid/plglighnir/reef/example/1L2Y.pdb xWybierz

SekwencjaXNLVIQWLKDGGPSSGRPPPSX
Sekwencja aminokwasów w zapisie jednoliterowym

OpisPrzykładowa dynamika

Zleć zadanieResetujPokaż zaawansowane

Uwaga 1: Wybranie pliku pdb, który zawiera strukturę odniesienia danego białka (na ogół strukturę eksperymentalną z bazy pdb), jest o w obecnej konfiguracji portalu konieczne. Jeżeli ten plik nie zostanie podany, zadanie zakończy się z błędem.

Uwaga 2: Sekwencja aminokwasowa musi zaczynać/kończyć się resztą "dummy" oznaczaną jako X jeżeli łańcuch nie zaczyna się/kończy pełną grupą peptydową (czyli grupa początkowa jest grupą $-NH_3^+$ lub grupa końcowa jest grupą COO^- ; sekwencja "nieblokowana"). Jeżeli łańcuch zaczyna się/kończy pełną grupą peptydową lub pierwszą/ostatnią resztą jest glicyna, należy wstawić G. Wstawienie innych reszt niż X lub G na początku lub końcu spowoduje błędy w obliczeniach. Wstawienie grup blokujących powoduje pozorne wydłużenie sekwencji, np. jeżeli chcemy wykonać obliczenia nieblokowanego pentapeptydu Ala-Lys-Gly-Pro-Ser to musimy wprowadzić sekwencję XAKGPSX a jeżeli chcemy wykonać obliczenia analogicznego blokowanego pentapeptydu Ac-Ala-Lys-Gly-Pro-Ser-NHCH₃ to wprowadzamy sekwencję GAKGPSG. W przypadku sekwencji np. Gly-Leu-Pro-Leu-Ala-Gly nie trzeba dodawać grup blokujących i wprowadzić sekwencję GLPLAG

Następnie w zakładce "Zasoby" ustawiamy klaster obliczeniowy oraz czas obliczeń. **UWAGA: W chwili obecnej portal uruchamia zadania jedynie na klastrze inula).**

Więcej opcji dynamiki można uzyskać po kliknięciu napisu "Zaawansowane" w zakładce MD. Odpowiednie pola są objaśnione poniżej; wartości domyślne podano w nawiasach:

NTWE (100): częstość zapisu energii w dynamice molekularnej.

NTWX (1000): częstość zapisu współrzędnych cząsteczki białka.

DT: (0.1) długość kroku czasowego; jednostka jest równa 48,9 fs.

DAMAX (1): Maksymalna dopuszczalna zmiana przyspieszenia podczas jednego kroku czasowego

Aby wybrać grant w ramach którego będą wykonywane obliczenia, należy wykonać następujące czynności:

1. wybrać zakładkę MD,
2. kliknąć napis "pokaż zaawansowane",
3. wpisać w pole Grant ID grantu.

Zleć zadanie

[Zapisz jako szablon](#)

MD Zasoby Pliki Środowisko Inne

Host

Kolejka

galera.task.gda.pl

hydra.icm.edu.pl

inula.man.poznan.pl

moss.man.poznan.pl

Wall time

nova.wcss.wroc.pl

reef.man.poznan.pl

zeus.cyfronet.pl

[Zleć zadanie](#)[Resetuj](#)[Pokaż zaawansowane](#)

Zleć zadanie

[Zapisz jako szablon](#)

MD Zasoby Pliki Środowisko Inne

Host

Kolejka

Liczba procesów

Wall time

[Zleć zadanie](#)[Resetuj](#)[Pokaż zaawansowane](#)

Program UNRES wygeneruje tyle niezależnych trajektorii ile wynosi liczba procesorów.

Plikami wynikowymi, które po poprawnym zakończeniu zadania pojawią się w katalogu roboczym i będą mogły być pobrane przez gftp, są:

ZADANIE.out_GB000 - ZADANIE.out_GB00n (gdzie zadanie jest prefiksem nadawanym przez portal a n+1 jest liczbą procesorów): główne pliki wynikowe. Zawierają podstawowe informacje o układzie, informacje o ustawieniach obliczeń dynamiki molekularnej oraz ewentualne informacje o błędach w danych.

ZADANIE_GB000.stat - ZADANIE_GB00n : pliki zawierające podstawowe informacje o przebiegu dynamiki dla każdej trajektorii (nr kroku, czas, energia, rmsd od struktury odniesienia). Proszę zapoznać się z [opisem plików danych i wynikowych programu UNRES](#) (dokumentacja w języku angielskim).

ZADANIE_MD000.pdb - ZADANIE_MD00n.pdb : zrzuty konformacji w formacie pdb dla każdej trajektorii. Pliki te można bezpośrednio oglądać przeglądarkami do molekuł np. [pymol](#).

Uruchamianie bezpośrednio w systemie wsadowym

Uruchamianie zadań na klastrze inula

W celu zdefiniowania ścieżek dostępu oraz ustawienia wartości zmiennych środowiskowych dla usługi należy załadować moduł plgrid/apps/unres-e01l2y/3.2.1 (pole E0LL2Y0) lub plgrid/apps/unres-gab/3.2.1 (pole GAB), w zależności od żadanego pola siłowego. W przypadku uruchamiania zadań przez portal odpowiedni moduł jest ładowany poprzez wybór odpowiedniego pola siłowego z menu.

Przykłady przygotowywania plików wsadowych oraz danych do uruchamiania usługi UNRES w różnych wariantach wraz z opisem zawarte są w archiwum zip. [przykłady.zip](#) można pobrać korzystając z przytoczonego linku.

Poniżej krótki opis wraz z objaśnieniem uruchamiania kanonicznej dynamiki molekularnej Langevina. Używany jest wariant pola siłowego UNRES oznaczony jako pole E0LL2Y (pole "ogólnych zastosowań"). Ten przykład jest zawarty w przytoczonym archiwum w katalogu ff_112y_1le1/LANG.

Skrypt QCG (plik langdyn.qcg; nie ma w nim dodanych tutaj komentarzy):


```
#QCG host=inula
## Obliczenia na inuli (na razie jedyna moliwo)
#QCG walltime=PT15M
## maksymalny czas zadania 15 minut
#QCG nodes=1:1
## uywamy 1 procesor
#QCG note=l2y-langdyn
## identyfikator zadania
#QCG output=${JOB_ID}.output
## specyfikacja nazw pliku z wydrukami z programu i systemu qcg na ekran
#QCG error=${JOB_ID}.error
## specyfikacja nazwy pliku z komunikatami o bdach
#QCG stage-in-dir=.
## pliki danych bd brane z biecego katalogu
#QCG stage-out-dir=.
## pliki wynikowe bd umieszczane w biecyym kagalogu
module load plgrid/apps/unres-e0ll2y/3.2.1
## adowanie moduu zawierajcego komponentu usugi dla pola E0LL2Y
export NPROCS=`cat $PBS_NODEFILE | wc -l`
cd $PBS_O_WORKDIR
mpirun -machinefile $PBS_NODEFILE -np $NPROCS unres.csh
## uruchamianie skryptu unres.csh
```


Plik unres.csh (nie ma w nim dodanych tutaj komentarzy)


```
#!/bin/csh
setenv FGPROCS 1
# obliczenia energii nie s zrównoleglone
setenv POT GB
# specyfikacja potencjau oddziazywa midzy acuchami bocznymi (musi by taka dla pola E0LL2Y)
setenv PREFIX 1L2Y_lang
# specyfikacja nazw plików; plik wynikowy bdzie si nazywa 1L2Y_lang.inp
setenv OUT1FILE YES
# plik wynikowy bdzie produkowany tylko przez procesor gówny (tutaj nie jest to wane bo obliczenia s
jedenprocesorowe)
unresMD_gfortran_MPICH_E0LL2Y.exe $*
# uruchamianie aplikacji.
```


Plik danych 1L2Y_lang.inp (uwaga: formatowanie przy wyświetlaniu na stronie podręcznika nie jest poprawne); plik poprawnie sformatowany można pobrać [stąd](#). Proszę NIE KOPIOWAĆ jego treści podanej dalej bezpośrednio z ekranu ponieważ zawiera ona komentarze, które będą powodowały błędy czytania. Poniżej podany jest tylko przykładowy skróty opis; pełen opis danych znajduje się w [dokumentacji pakietu UNRES](#) (w języku angielskim).


```
1L2Y                                     # tytu (acuch tekstowy)
80-znakowy).

SEED=-3059743 PDBREF MD                 # ogólne dane; startowa
liczba losowa, specyfikacja, e bdzie struktura odniesienia i specyfikacja oblicze MD

nstep=1000000 ntwe=100 ntwx=10000 dt=0.01 damax=10000.0 lang=1 scal_fric=0.01 & # 10 mln kroków
dynamiki, rzrzt energii co 100 kroków, rzrzt współrzdnych co 10000 kroków, krok 0.01 jednostek
"gruboziarnistych" (4.89 fs), obliczenia w trybie Langevina, skalowanie lepkości wody przez 1/100, kontynuacja
tych danych w nastpnej linii ("&" w 80-tej kolumnie).

t_bath=300                               # temperatura 300 K.

WLONG=1.00000 WSCP=1.23315 WELEC=0.84476 WBOND=1.00000 WANG=0.62954           & # wagi wkładów do energii
(nie mog by zmienione dla uywanej wersji pola siowego)

WSCLOC=0.10554 WTOR=1.84316 WTORD=1.26571 WCORRH=0.19212 W CORR5=0.00000      &

W CORR6=0.00000 WEL_LOC=0.37357 WTURN3=1.40323 WTURN4=0.64673 WTURN6=0.00000  &

WVDWPP=0.23173 WHPB=1.00000 WSCCOR=0.0                                         &

CUTOFF=7.00000 W CORR4=0.00000

1L2Y.pdb                                # nazwa pliku ze
struktur eksperymentaln (struktur odniesienia); podlinkowany do pobrania.
```

```

22                                     # 22 reszty aminokwasowe
D  ASN LEU TYR ILE GLN TRP LEU LYS ASP GLY GLY PRO SER SER GLY ARG PRO PRO PRO # sekwencja biała w
kodzie trójliterowym (D oznacza grup blokując, która nie zawiera penej grupy peptydowej); format 20 (1X,A3).
SER D
0                                     # brak reszt cysteiny
mogcych tworzy wizania disulfidowe
0                                     # brak wizów
104.9286  87.6954  81.9040  82.4664  84.4658  83.9446  85.3958  91.4490 # specyfikacja
geometrii wirtualnej: wirtualne kty walencyjne, wirtualne kty torsyjne oraz kty sferyczne okrelajce geometri
acuchów bocznych.
94.6016  101.8618  119.3634  94.3628  96.2641  138.1186  96.2995  129.7022
109.4446  106.3486  106.0416  108.7182
-100.4302  72.4264  66.0265  51.4337  53.5083  60.5686  44.1766  67.6037
-72.8136  -61.9155  -75.8939  67.3579  129.7011  -95.5708  63.9717  -74.5043
-122.3059  -134.6048  92.1133
102.3561  120.0919  152.3636  134.9756  122.3426  152.1780  159.0518  100.5580
139.9612  0.0000  0.0000  117.4525  137.0250  146.2899  0.0000  93.9014
101.0251  113.0432  93.7775  153.8351
-82.3166  -56.6854  85.0901  -88.6662  -140.9453  38.0240  179.4707  -73.0903
-144.7967  0.0000  0.0000  -133.1635  -106.6593  -130.3055  0.0000  -102.7467
-111.6412  -122.0445  -102.3738  -143.3033

```

Uruchamianie

Po zalogowaniu się na serwer, z którego można uruchamiać zadania w systemie [Qcg-Qoscos](http://qcg.man.poznan.pl), np. qcg.man.poznan.pl i przeniesieniu tam wymienionych powyżej plików, należy wydać polecenie:

```
qcg-sub langdyn.qcg
```

Do monitorowania zadań służą [polecenia qcg](#).

Plikami wynikowymi będą:

[1L2Y_lang.out_GB000](#) : plik zawierający podstawowe informacje o danych wejściowych, typie obliczeń, zastosowanych ustawieniach oraz przebiegu obliczeń.

[1L2Y_lang_GB000.stat](#) : plik zawierający skróconą informację o powszechnych krokach czasowych dynamiki; opis (w języku angielskim) znajduje się na [stronie opisu programu UNRES](#).

[1L2Y_lang_GB000.cx](#) : plik zawierający współrzędne poszczególnych konformacji z trajektorii dynamiki Langevina (zgodnie z informacją podaną w pliku wejściowym, są one zapisywane z częstotliwością co 10000).

Trajektoria zapisana w formie skompresowanej w pliku cx nie jest gotowa do oglądania/przetwarzania programami grafiki molekularnej. Należy przekształcić go do formatu [Protein Data Bank \(PDB\)](#) przy użyciu programu xdrf2pdb. Program jest dostępny w ramach usługi UNRES i może być uruchamiany w trybie wsadowym jednak z uwagi na to, że pliki pdb zajmują znacznie więcej miejsca niż pliki cx, lepiej skopiować plik cx na swój lokalny komputer i tam uruchomić xdrf2pdb interaktywnie. Program xdrf2pdb można uzyskać ze [strony "downloads" pakietu UNRES](#). Po wejściu na stronę należy kliknąć na guzik "ACCEPT LICENSE TERMS" i ściągnąć archiwum [xdrfpdb-src-v.3.2.1.tar.gz](#). Program xdrf2pdb wymaga biblioteki xdrf, która jest w pakiecie UNRES. Jeżeli jeszcze jej nie posiadamy to trzeba ściągnąć archiwum [lib.tar.gz](#) ze strony pakietu UNRES. Można też ściągnąć cały pakiet UNRES [unrespack-v.3.2.1.tar.gz](#); biblioteka jest w sources/lib/xdrf lub xdrf_em64 (64 bit) albo xdrf-Argonne (BlueGene Q) gdyby ta nie zadziałała.

Polecenie konwersji dla podanego przykładu jest następujące:

```
xdrf2pdb three seq-3let 1L2Y_lang_MD000.cx 1 1 100 1L2Y_lang.pdb
```

gdzie

three oznacza, że podawana będzie sekwencja aminokwasowa białka w kodzie trójliterowym

[seq-3let](#) jest plikiem z sekwencją; format 20(A3, 1X); zawartość poniżej

```

D  ASN LEU TYR ILE GLN TRP LEU LYS ASP GLY GLY PRO SER SER GLY ARG PRO PRO PRO
SER D

```

1L2Y_lang_MD000.cmx jest nazwą pliku ze współrzędnymi konformacji białka z obliczonej trajektorii w formacie skompresowanym.

1 1 100 : konformacje będą zapisywane co 1, począwszy od konformacji 1 a skończywszy na konformacji 100

[1L2Y_lang.pdb](#) : wynikowy plik z konformacjami w formacie PDB; konformacje te zawierają geometrię na poziomie gruboziarnistym; tylko atomy węgla alfa (CA) i centra łańcuchów bocznych (CB).

Uruchamianie zadań na klastrze Prometheus

Zadania są uruchamiane w systemie kolejki slurm. W zależności od używanego pola należy załadować jeden z modułów:

- plgrid/apps/unres-e0ll2y/3.2.1 (w tym przypadku należy uruchomić aplikację unresMD_ifort_MPI_E0LL2Y.exe lub odpowiedni stowarzyszony z nią program przetwarzający wyniki symulacji)
- plgrid/apps/unres-gab/3.2.1 (w tym przypadku należy uruchomić aplikację unresMD_ifort_MPI_GAB.exe lub odpowiedni stowarzyszony z nią program przetwarzający wyniki symulacji)

Dodatkowo należy załadować moduł plgrid/tools/mpi.

Podobnie jak w przypadku klastra inula, załadowanie modułów powoduje zdefiniowanie plików z właściwymi dla danego pola parametrami; pliki te można jednak podmienić na pliki z parametrami użytkownika. Przykładowy skrypt jest podany poniżej.

Plik energy-pdb.sbatch

#!/bin/bash -l

#SBATCH -J unres

#SBATCH -N 1

#SBATCH --ntasks-per-node=1

#SBATCH -n 1

#SBATCH --time=00:01:00

#SBATCH -A unresprometheustest

#SBATCH -p plgrid

#SBATCH --output="output.out"

#SBATCH --error="error.err"

module load plgrid/apps/unres-e0ll2y/3.2.1

module add plgrid/tools/mpi

export PREFIX=1L2Y_ene-pdb

export OUT1FILE=YES

export FGPROCS=1

mpirun unresMD_ifort_MPI_E0LL2Y.exe

Przykładowe pliki wsadowe dla systemu slurm oraz pliki danych i pliki wynikowe dla różnych trybów obliczeń wraz z krótkim opisem są dostępne w katalogu `/net/software/local/unres/Przyklady` na klastrze Prometheus; archiwum można pobrać stąd w postaci pliku [Przyklady_prometheus.zip](#).

Zaawansowane użycie

Przewidywanie *de novo* struktury białka przy pomocy pola siłowego UNRES

Pole siłowe UNRES umożliwia przeprowadzanie przewidywania *de novo* struktury białka, korzystając jedynie z sekwencji aminokwasowej. W tym celu należy wykonać symulację dynamiki molekularnej ze zwielokrotnioną wymianą replik (MREMD) wprowadzając sekwencję białka o poszukiwanej strukturze do przykładowego (Rysunek 1) pliku wejściowego MREMD.

```

T0810
SEED=-3059743 RE rescale_mode=2 one_letter outfile extconf timlim=1420.0
nstep=50000000 dt=0.10 ntwe=10000 ntwx=10000 lang=0 tbf tau_bath=1 t_bath=300 &
damax=20.0 EDRIFTMAX=20.0 reset_moment=1000 reset_vel=0 restart respa &
ntime_split=1 maxtime_split=512
nrep=32 nstex=20000 tlist mlist traj1file rest1file sync nsyn=20000
250 255 260 265 270 275 280 285 290 295 300 305 310 315 320 325 330 335 340 345&
 350 360 380 390 400 410 420 430 440 460 480 500
  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2&
  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2&
WLONG=1.00000 WSCP=1.23315 WELEC=0.84476 WBOND=1.00000 WANG=0.62954 &
WSCLOC=0.10554 WTOR=1.34316 WTORD=1.26571 WCORRH=0.19212 WCORR5=0.00000 &
WCORR6=0.00000 WEL_LOC=0.37357 WTURN3=1.40323 WTURN4=0.64673 WTURN6=0.00000 &
WVDWPP=0.23173 WHPB=1.00000 WSCCOR=0.25 &
CUTOFF=7.00000 WCORR4=0.00000
385
XMYMKKLIVFGMLIVSAASLVGCGSQIGSSVRQAVSDNQSAQTLVEWENSEANPEALFANWRHEFMVDSSKRESMKTELC
KELQALPAQDLTLFENEIRDENNRAIVSGCKEELLAQVDEHFDEQRESMSVPGHALKAVQSRNSFRFPDNTQKRDMNGY
MAVRGDVARKEVVLTFDDGPHGLYTDAILRALKEVNAKAMFFATGKSVRTNPEALKRVAADGHVIGSHSITHACLGTSVA
CYKQMGNRNLTFDAAAEEVRGGHQAVFDVLGWVDPVFRFPYGETSKDLKAFLKTKSTGEFAWNIESDDWRTQSNEQLLAR
VLAVNESQGRGIVLFHDIQRRTAEIMPQFLRELYNRGYSVVLTLAADPSAKYNSKLVKRKQQLPX
0
385
50.0
 1  0.000 180.000
 2  0.000 180.000
 3  0.000 180.000
 4  0.000 180.000
 5  0.000 180.000
 6  50.000 20.000
 7  50.000 20.000
 8  50.000 20.000

```

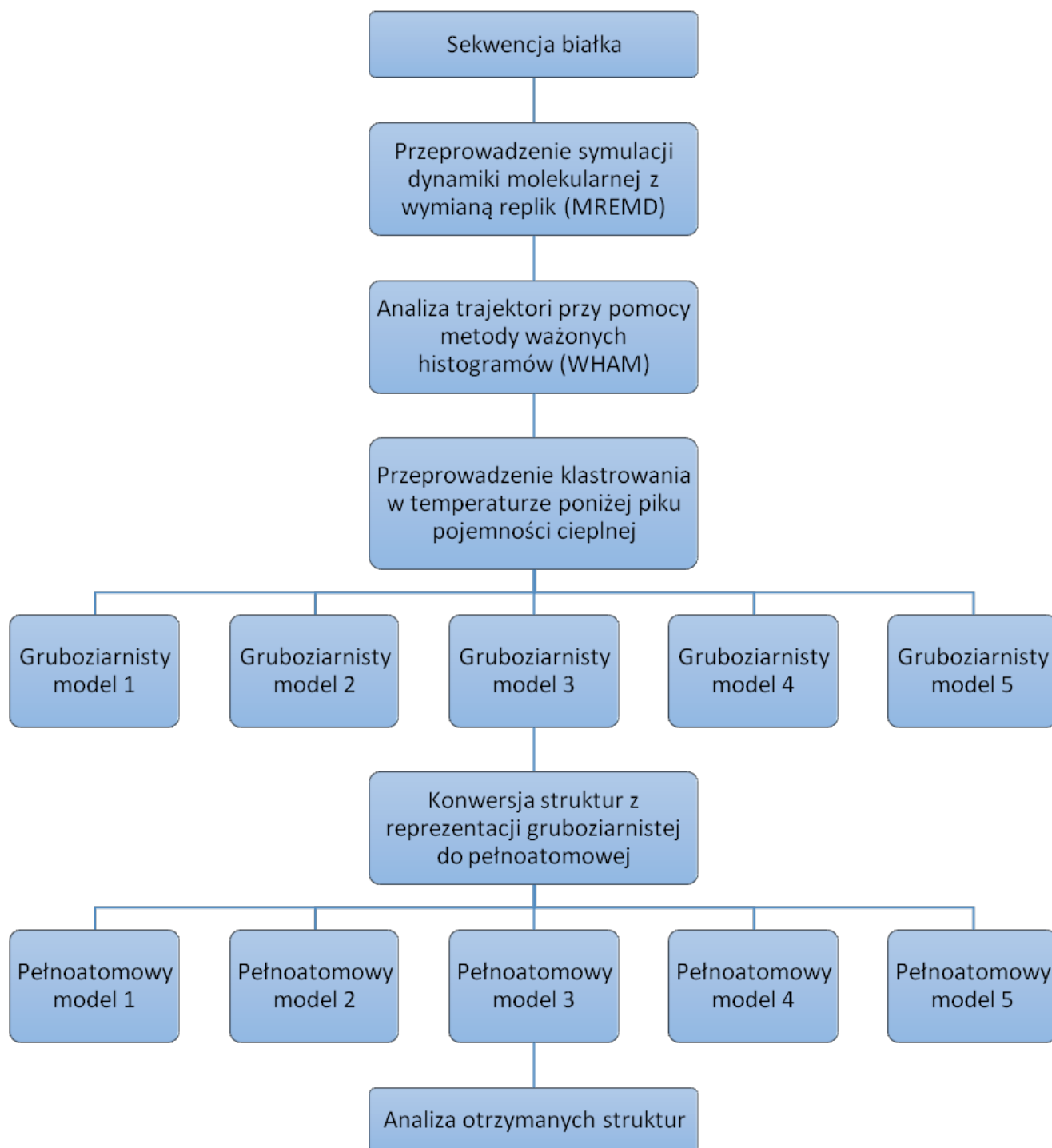
Rysunek 1. Przykładowy plik wejściowy używany w polu siłowym UNRES. Linia pierwsza stanowi komentarz, linie 2-14 zawierają parametry używane w polu siłowym, linia 15 zawiera całkowitą liczbę reszt aminokwasowych symulowanego układu z uwzględnieniem „dummy atom”, natomiast kolejne linie zawierają sekwencję aminokwasową w kodzie jednoliterowym wraz z „dummy atom” (litery X). Ostatnie linie wykorzystywane są do wprowadzenia ograniczeń swobody układu (więzów) np. na strukturę drugorzędową lub odległości między parami atomów.

Przy wprowadzaniu sekwencji należy pamiętać o dodawaniu „dummy atom” na początku i końcu każdego łańcucha białka (litery X w kodzie jednoliterowym).

W celu przyspieszenia symulacji można wprowadzić do pliku wejściowego dane o strukturze drugorzędowej (przewidzianej przy pomocy zewnętrznego narzędzia, takiego jak PSIPRED czy JPred) w postaci listy wartości brzegowych kątów torsyjnych ograniczających swobodę łańcucha (ostatnie linie pliku wejściowego).

W celu przeprowadzenia analizy otrzymanych trajektorii należy użyć podprogramu WHAM, wykorzystującego metodę analizy ważonych histogramów do określenia prawdopodobieństwa wystąpienia struktur w danych temperaturach. Następnie konieczne jest przeprowadzenie analizy skupień (klastrowania) przy użyciu podprogramu Cluster, w temperaturze poniżej temperatury piku pojemności cieplnej, w celu wytypowania najbardziej prawdopodobnych struktur białka. Otrzymane w ten sposób struktury należy następnie przekonwertować z postaci gruboziarnistej do pełnoatomowej przy użyciu dostępnych narzędzi (np. programu Pulchra).

Zalecany schemat postępowania został umieszczony na poniższym diagramie:



Gdzie szukać dalszych informacji?

[Strona dokumentacji pakietu UNRES](#) ze szczegółowym opisem wprowadzanych danych (w języku angielskim).