

Wielkoskalowe symulacje biomolekuł przy użyciu modeli gruboziarnistych

Optymalizacja pakietu UNRES

Agnieszka Lipska¹, Adam Liwo^{1,2}

¹Centrum Informatyczne Trójmiejskiej Akademickiej Sieci Komputerowej

²Wydział Chemii Uniwersytetu Gdańskiego

7 grudnia, 2023



Fundusze
Europejskie



Rzeczpospolita
Polska

Unia Europejska
Europejskie Fundusze
Strukturalne i Inwestycyjne



Współautorzy

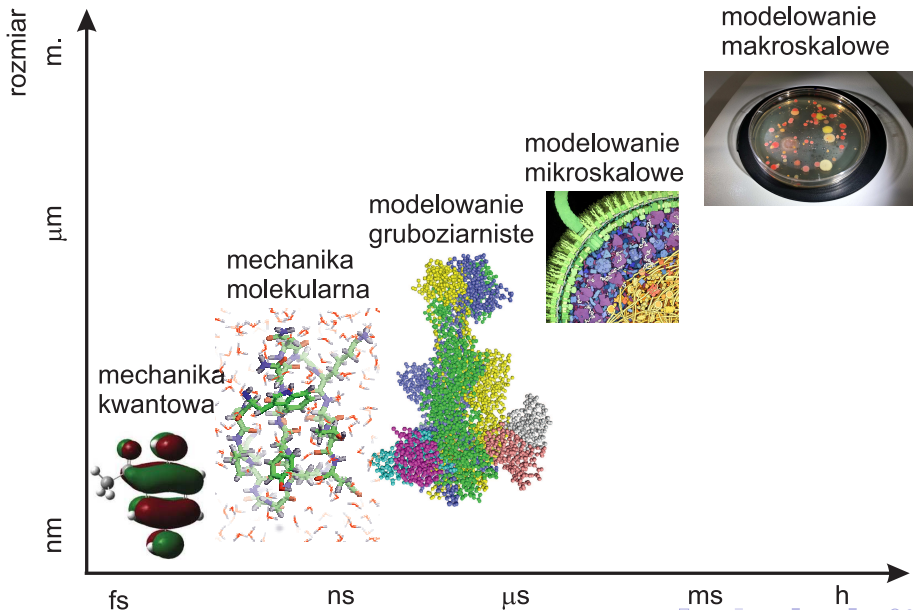
- Cezary Czaplewski, WCH UG, CI TASK PG
- Paweł Czarnul, WETI PG
- Henryk Krawczyk, WETI PG, CI TASK PG
- Emilia Lubecka, WETI PG
- Krzysztof M. Ocetkiewicz, CI TASK PG
- Jerzy Proficz, CI TASK PG
- Adam K. Sieradzan, WCH UG, CI TASK PG

Agenda

- 1 Miejsce modelowania gruboziarnistego w symulacjach biomolekularnych
- 2 Model UNICORN makromolekuł biologicznych
- 3 Dynamika molekularna w modelu UNRES/UNICORN
- 4 Pomijanie centrów zbyt odległych: listy oddziaływań
- 5 Zrównoleglenie kodu: MPI/OpenMP i CUDA
- 6 Skalowalność i możliwości równoległej implementacji modelu UNRES
- 7 Podsumowanie
- 8 Dostępność pakietu UNRES/UNICORN
- 9 Przykład: prezentuje Agnieszka Lipska

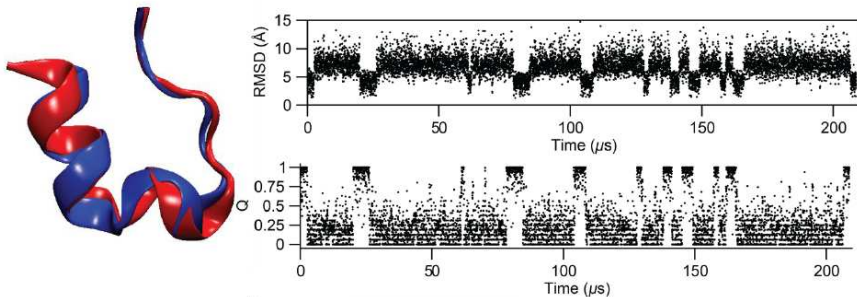
Modelowanie gruboziarniste

Skale obliczeń w modelowaniu biomolekularnym



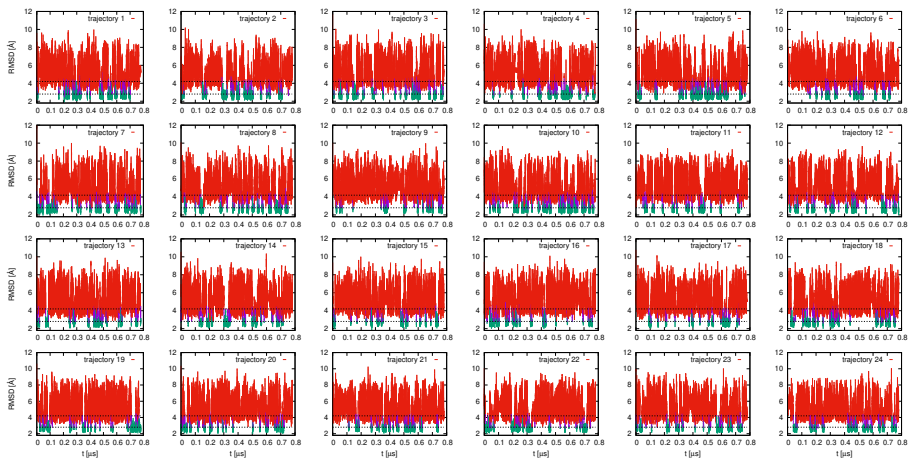
W modelowaniu gruboziarnistym można policzyć więcej i wszystko dzieje się znacznie szybciej

Trp-cage



Lindorff-Larsen et al., *Science*, 2011, 34, 517-520

W modelowaniu gruboziarnistym można policzyć więcej i wszystko dzieje się znacznie szybciej

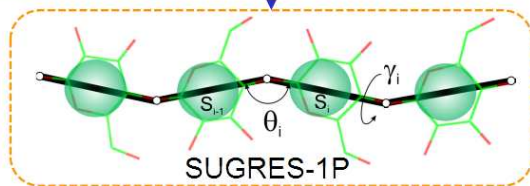
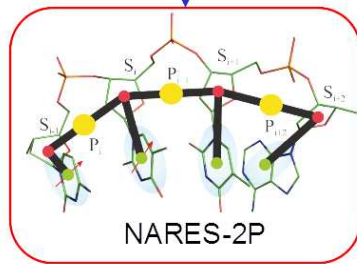
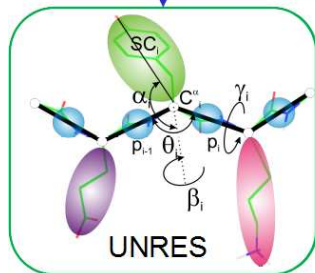


Sieradzan et al., *J. Comput. Chem.*, 2023, 44, 602-625

Model UNICORN

Model UNICORN makromolekuł biologicznych

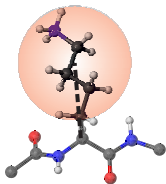
UNICORN



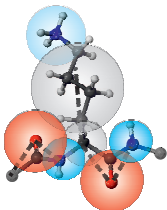
Liwo et al., *J. Molec. Model.*, **2014**, 20, 1-15

Liwo et al., *Prog. Mol. Biol. Trans. Sci.*, **2020**, 170, 73-121

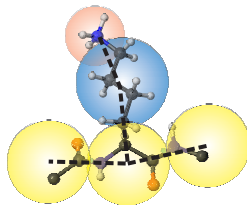
Porównanie modelu UNICORN z innymi modelami gruboziarnistymi



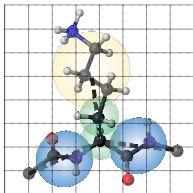
AWSEM/OPEP
ROSETTA



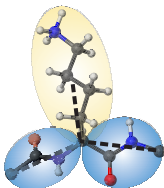
SIRAH



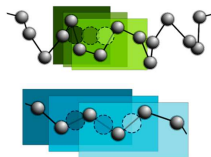
MARTINI



CABS



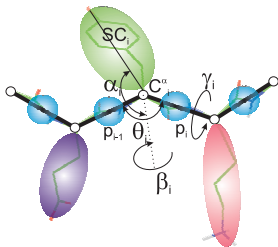
UNRES



SURPASS

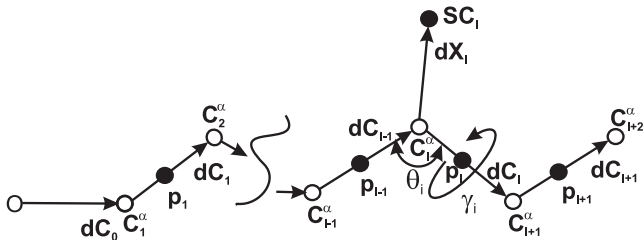
Wyrażenie na energię

$$\begin{aligned} U = & w_{SCSC} \sum_{i < j} U_{SC_i SC_j} + w_{SCp} \sum_{i \neq j} U_{SC_i p_j} \\ & + w_{pp}^{VDW} \sum_{i < j-1} U_{p_i p_j}^{VDW} + f_2(T) w_{pp}^{el} \sum_{i < j-1} U_{p_i p_j}^{el} \\ & + w_{bond} \sum_i U_{bond}(d_i) + w_b \sum_i U_b(\theta_i) \\ & + w_{rot} \sum_i U_{rot}(\theta_i, \alpha_{SC_i}, \beta_{SC_i}) \\ & + f_2(T) w_{tor} \sum_i U_{tor}(\gamma_i, \theta_i, \theta_{i+1}) \\ & + f_3(T) w_{corr}^{(3)} \sum_{i < j-2} U_{corr;i}^{(3)} \\ & + f_3(T) w_{turn}^{(3)} \sum_i U_{turn;p_i p_{i+2}}^{(3)} \end{aligned}$$



Dynamika

Dynamika molekularna w modelu UNRES/UNICORN



$$\mathbf{q} = \begin{pmatrix} \mathbf{q}_x \\ \mathbf{q}_y \\ \mathbf{q}_z \end{pmatrix} \quad \mathbf{q}_\xi = \begin{pmatrix} C_{\xi 1} \\ C_{\xi 2} \\ X_{\xi 2} \\ \vdots \\ C_{\xi n} \end{pmatrix}$$

Równania ruchu (dynamika Langevina)

$$\mathbf{G}\ddot{\mathbf{q}} = -\nabla_{\mathbf{q}}U(\mathbf{q}) - \mathbf{\Gamma}\dot{\mathbf{q}} + \mathbf{f}^{rand}$$

\mathbf{G} : macierz bezwładności,

\mathbf{q} : uogólnione współrzędne gruboziarniste,

U : efektywna funkcja energii w modelu UNRES,

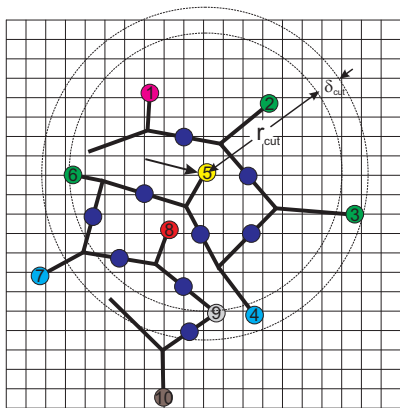
$\mathbf{\Gamma}$: tensor lepkości.

$$\mathbf{f}^{rand} = \left(\frac{k_B T}{\delta t} \mathbf{\Gamma} \right)^{\frac{1}{2}} \mathbf{N}(0, 1)$$

$\mathbf{N}(0, 1)$: wielowymiarowy rozkład Gaussa o zerowej średniej i jednostkowej wariancji.

Listy oddziaływań

Pomijanie centrów zbyt odległych: listy oddziaływań



Lista oddziaływań łańcucha bocznego bocznego 5

centrum	od	do
5	6	6
5	8	9

$$r_{cut} = 25 \text{ \AA}$$
$$\delta_{cut} = 0.5 \text{ \AA}$$

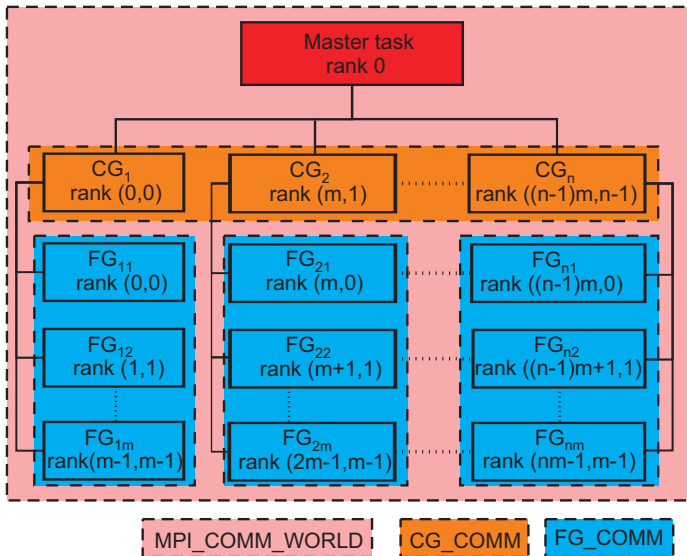
Dzięki pominięciu par centrów odległych więcej niż o r_{cut} czas obliczeń skaluje się jak $N \log N$ a nie jak N^2 z rozmiarem problemu.

Trzy listy oddziaływań SC...SC, SC...p i p...p.

Sieradzan et al. *J. Comput. Chem.*, 2023, 44, 602-625.

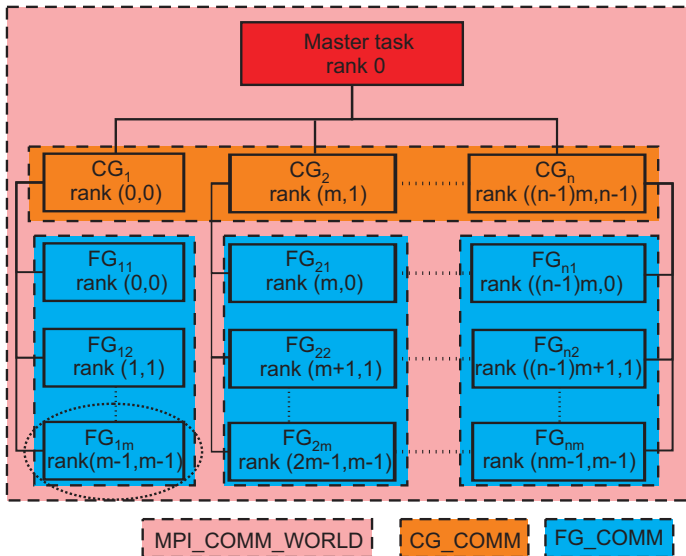
Zrównoleglenie kodu

Schemat zrównoleglenia kodu



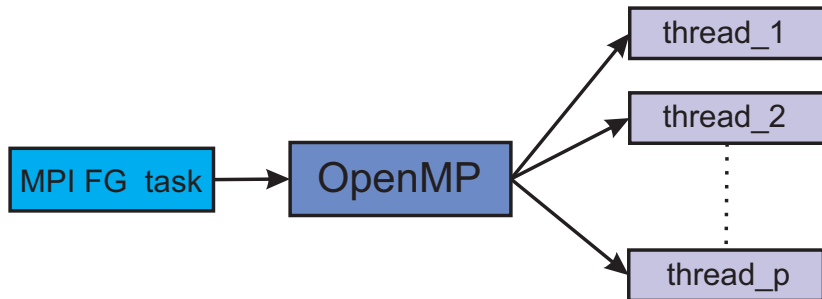
Sieradzan et al. *J. Comput. Chem.*, 2023, 44, 602-625.

Schemat zrównoleglenia kodu



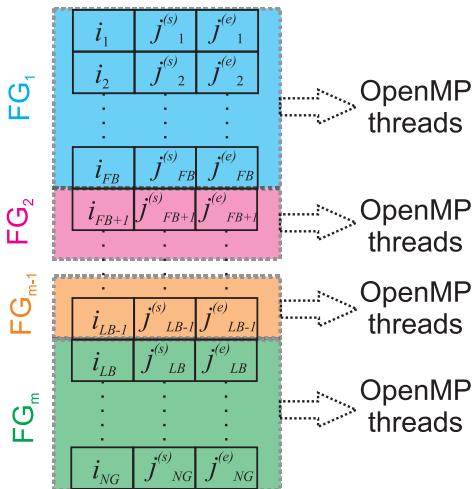
Sieradzan et al. *J. Comput. Chem.*, 2023, 44, 602-625.

Schemat zrównoleglenia kodu



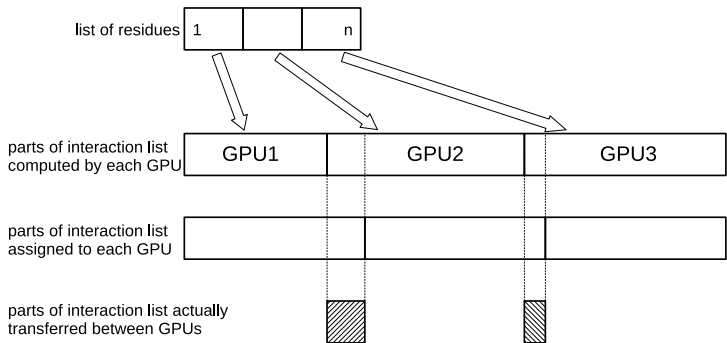
Sieradzan et al. *J. Comput. Chem.*, 2023, 44, 602-625.

Podział listy oddziaływań między procesy/wątki (MPI, OpenMP, 1GPU)



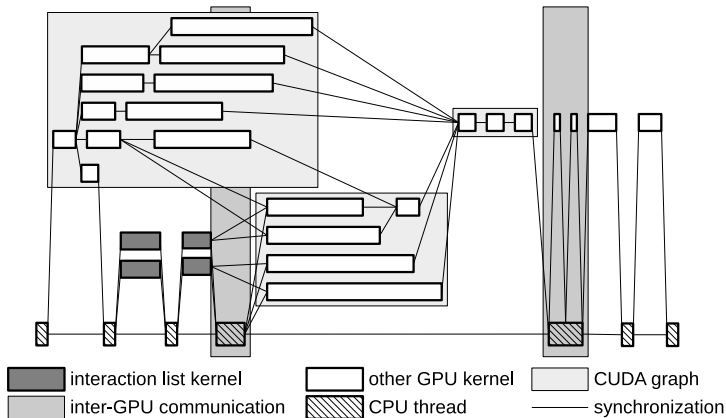
Sieradzan et al. *J. Comput. Chem.*, **2023**, 44, 602-625; Ocetkiewicz et al., *Bioinformatics*, **2023**, 39, btad391.

Podział listy oddziaływań między procesy/wątki (mGPU)



Ocetkiewicz et al. *Comput. Phys. Commun.*, 2023, w recenzji.

Przeptyw danych w implementacji na wiele GPU

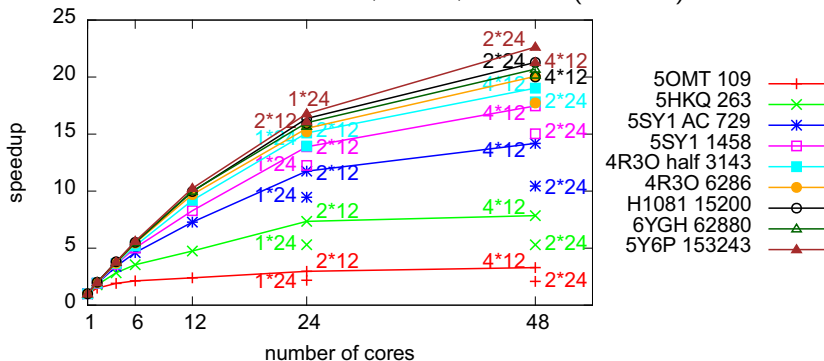


Ocetkiewicz et al. *Comput. Phys. Commun.*, 2023, w recenzji.

Skalowalność

Skalowalność równoległej implementacji modelu UNRES (OpenMP, pojedyncza trajektoria)

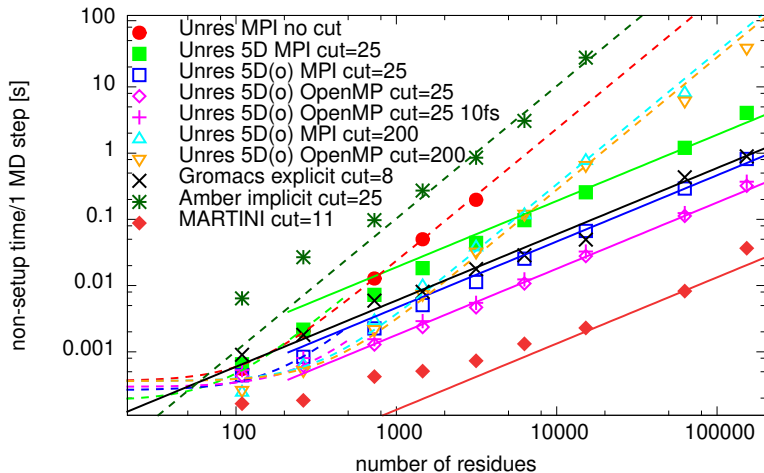
Xeon® E5 v3 @ 2,3 GHz, 12-core (Haswell)



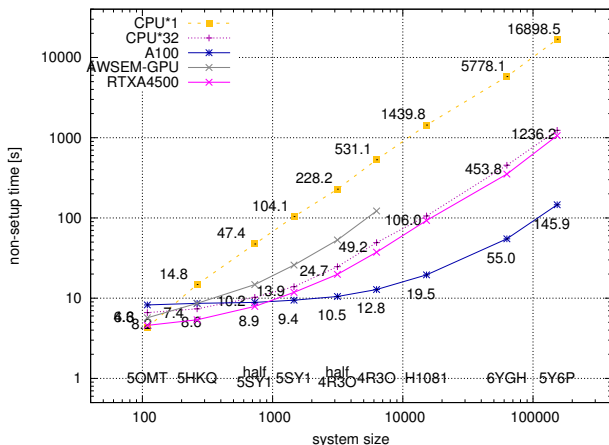
Sieradzan et al. *J. Comput. Chem.*, 2023, 44, 602-625.

Możliwości równoległej implementacji UNRES (MPI i OpenMP).

Xeon[®] E5 v3 @ 2,3 GHz, 12-core (Haswell)

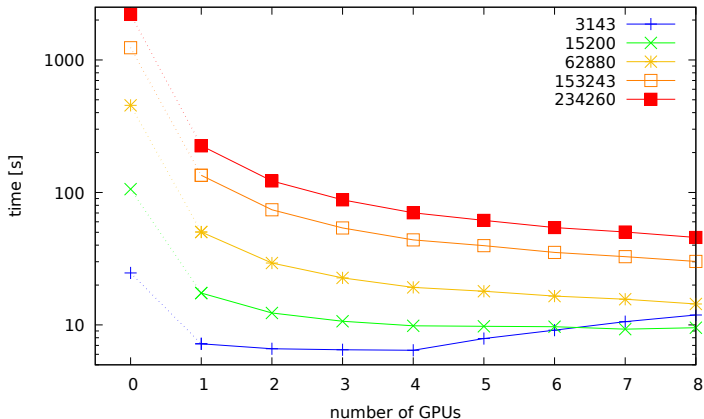


Porównanie długości symulacji MD dla implementacji 1GPU pakietu UNRES z innymi implementacjami oraz pakietami



Ocetkiewicz et al., *Bioinformatics*, 2023, 39, btad391.

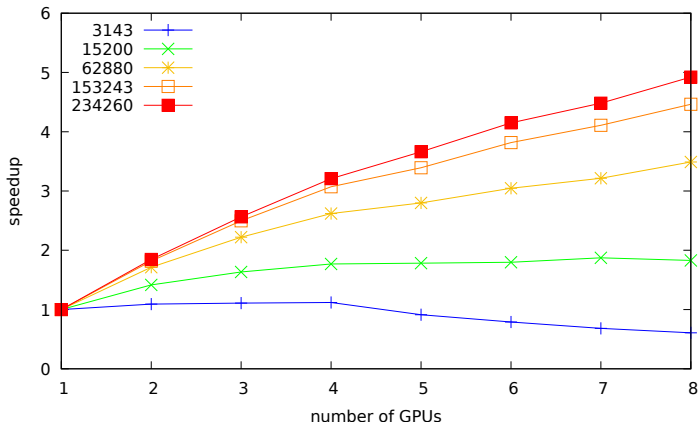
Możliwości implementacji mGPU pakietu UNRES



Czas symulacji w zależności od liczby GPU (A100)

Ocetkiewicz et al., *Comput. Phys. Commun.*, 2023, w recenzji.

Możliwości implementacji mGPU pakietu UNRES



Przyspieszenie w zależności od liczby GPU (A100)

Ocetkiewicz et al., *Comput. Phys. Commun.*, 2023, w recenzji.

Podsumowanie

Podsumowanie

- Zoptymalizowano gruboziarnisty pakiet UNRES pod względem szybkości obliczeń i wymagań pamięci.
- Opracowano i wdrożono efektywny równoległy oraz oszczędny pamięciowo algorytm tworzenia, uaktualniania i używania list oddziałujących centrów modelu UNRES.
- Rozszerzono pierwotną implementację równoległą pakietu z użyciem bibliotek MPI na OpenMP oraz hybrydową OpenMP/MPI.
- Opracowano i wdrożono implementację pakietu do obliczeń przy użyciu akceleratorów graficznych.
- Zoptymalizowany pakiet umożliwia, na 8 kartach A100, uzyskanie 1 ns symulacji ($1 \mu\text{s}$ w skali laboratoryjnej) dziennie dla układów białkowych o rozmiarach powyżej 250 tysięcy reszt aminokwasowych, co odpowiada rozmiarom i skalom czasowym struktur subkomórkowych.

Dostępność

Dostępność pakietu UNRES/UNICORN

- Strona pakietu UNICORN: www.unres.pl
A. Liwo et al., *J. Mol. Model.*, **2014**, 20, 1-15; *J. Chem. Phys.*, **2019**, 150, 155104.
- Wersja zoptymalizowana:
<https://projects.task.gda.pl/eurohpcpl-public/unres>
A.K. Sieradzan et al., *J. Comput. Chem.*, **2023**, 44, 602-625 (MPI, OpenMP); K.M. Ocetkiewicz et al., *Bioinformatics*, **2023**, 39, btad391 (GPU)
- Wersja serwerowa: <http://unres-server.chem.ug.edu.pl>
Czaplewski et al., *Nucleic Acids Res.*, **2018**, 46, W304-W309; Krupa et al., *Bioinformatics*, **2021**, 37, 1613-1615; Ślusarz et al., *Front. Biomol. Sci.*, **2022**, 9, 1071428

Pakiet dostępny bezkosztowo dla użytkowników akademickich (ogólna licencja publiczna GNU, wersja 3.0 i nowsze).

Przykład

Podziękowania

- 1 Program Operacyjny Inteligentny Rozwój 2014-2020, współfinansowany ze środków Europejskiego Funduszu Rozwoju Regionalnego. Projekt o nazwie „Narodowa Infrastruktura Superkomputerowa dla EuroHPC EuroHPC PL”. Nr. projektu POIR.04.02.00-00-D014/20-00.
- 2 Regionalny Program Operacyjny Województwa Pomorskiego na lata 2014-2020. Projekt o nazwie „Utworzenie w Gdańsku Centrum Kompetencji STOS (Smart and Transdisciplinary Knowledge Services) w zakresie Infrastruktury B+R”. Nr. projektu RPPM.01.02.00-22-0001/17-00.
- 3 Centrum Informatyczne Trójmiejskiej Akademickiej Sieci Komputerowej (CI TASK) – serwery obliczeniowe.

Dziękuję za uwagę