

Simulating the Long-timescale Structural Behavior of Bacterial and Influenza Neuraminidases with Different HPC Resources

Yana A. Sharapova¹, Dmitry A. Suplatov², Vytas K. Švedas^{1,2}

© The Authors 2018. This paper is published with open access at SuperFri.org

Understanding the conformational dynamics which affects ligand binding by Neuraminidases is needed to improve the *in silico* selection of novel drug candidates targeting these pathogenicity factors and to adequately estimate the efficacy of potential drugs. Conventional molecular dynamics (MD) is a powerful tool to study conformational sampling, drug-target recognition and binding, but requires significant computational effort to reach timescales relevant for biology. In this work the advances in a computer power and specialized architectures were evaluated at simulating long MD trajectories of the structural behavior of Neuraminidases. We conclude that modern GPU accelerators enable calculations at the timescales that would previously have been intractable, providing routine access to microsecond-long trajectories in a daily laboratory practice. This opens an opportunity to move away from the “static” affinity-driven strategies in drug design towards a deeper understanding of ligand-specific conformational adaptation of target sites in protein structures, leading to a better selection of efficient drug candidates *in silico*. However, the performance of modern GPUs is yet far behind the deeply-specialized supercomputers co-designed for MD. Further development of affordable specialized architectures is needed to move towards the much-desired millisecond timescale to simulate large proteins at a daily routine.

Keywords: neuraminidases, molecular dynamics, long-timescale trajectories, GPU, co-design.

Introduction

Co-infection of the human lower respiratory tract with viruses and pathogenic bacteria can lead to life-threatening complications to influenza including pneumonia. The mechanism that facilitates this synergism is a major focus of scientific research, and thus Neuraminidases, which are key virulence factors of both pathogens, are considered as important drug targets. Despite intensive studies, no inhibitors targeting bacterial Neuraminidases have been approved for clinical use so far. Four drugs targeting the influenza Neuraminidases are available, but their efficacy is challenged by evolution of the pathogen and emerging resistance [1]. It is therefore important to design new drugs before the currently available ones become useless. Probably, the cause for the poor success rate of delivering approved drugs to the market is that drug discovery is usually driven by optimizing the binding affinity and selectivity of the respective drug candidate in a static crystallographic structure of a target protein. Previous studies indicated that structures of bacterial and influenza Neuraminidases contain flexible regions which can be involved in a conformation-specific accommodation of ligands [2, 5, 7]. A deeper understanding of conformational flexibility and dynamics in Neuraminidases that can affect the binding of ligands, can help to improve the selection of novel drug candidates *in silico*.

The conventional molecular dynamics (MD) is a powerful computational method which can take on the challenge of accounting for protein flexibilities at studying drug recognition and binding to a target site [3]. The key problem is that computational cost of a simulation is very high even for a small protein, and much higher for the large multi-domain/chain Neuraminidases, thus the average MD trajectory so far has been limited to tens of nanoseconds. E.g., in one of the most recent studies of the influenza Neuraminidase ten MD trajectories, each of at most 100 ns, were calculated [2]. These computations can help to understand protein flexibility and function,

¹Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia

²Belozersky Institute of Physicochemical Biology, Lomonosov Moscow State University, Moscow, Russia

but overall are far below timescales relevant for the majority of functionally important structural rearrangements. The use of CPU-based computing clusters can provide some speed-up, but the most promising solution so far is the emergence of specialized architectures. The Anton-1 and Anton-2 are special-purpose supercomputers for MD built by a privately held biochemistry company D. E. Shaw Research [6]. Antons run calculations entirely on specialized application-specific integrated circuits and can accommodate only custom-designed MD software, presenting a successful example of a deep co-design of the computer software and hardware to execute it. Although some access to this unique and apparently very expensive hardware is provided to scientific community in the US, its overall availability so far has been very limited. Alternatively, the graphical processing units (GPU) are serially produced specialized architectures which offer an affordable speed-up to MD.

In this work the advances in computer technologies were evaluated at simulating long MD trajectories of Neuraminidases. The article is organized as follows. Section 1 describes the molecular modeling protocol, hardware and software setup. In section 2, we report the results and their analysis. The Conclusion section summarizes the study.

1. Methods

1.1. Hardware and Software

The computations were carried out on the “Lomonosov-2” supercomputer [4] using nodes equipped with (1) one Tesla K40 GPU, one Intel Xeon E5-2697 v3 CPU, and 64 GB RAM; and (2) two Tesla P100 GPUs, one Intel Xeon Gold 6126 CPU, and 96 GB RAM; and connected by Infiniband FDR. The AMBER16 package was compiled with Intel collection v.15.0.3, GNU collection v.4.8.5, MKL v. 11.1.3, CUDA v. 8.0, and OpenMPI v. 1.10.7.

1.2. Structure Preparation and the Molecular Modeling Protocol

The molecular systems of H1N1 influenza virus Neuraminidase (PDB 3B7E) and Neuraminidase A (NanA) from *Streptococcus pneumoniae* bacteria (PDB 2YA8) were prepared and simulated in the FF15IPQ force field with SPC/E_b water model as recently described [5]. The homotetrameric structure of the H1N1 Neuraminidase was reconstructed based on the BIOMT record in PDB. Only the catalytic domain (including the insertion domain) of the NanA Neuraminidase was selected for this study as the key catalytically competent part of the structure [5]. The size of influenza and bacterial Neuraminidases in a water box was ~ 150000 and ~ 65000 atoms, respectively. To compare the speed of MD on contemporary GPUs with the special-purpose Anton family of supercomputers, the molecular system of Human Dihydrofolate Reductase (PDB 2C2T) in water worth of ~ 24000 atoms was constructed, which was equivalent to one previously used to benchmark Antons [6].

2. Results

Speed of MD simulation of H1N1 influenza virus Neuraminidase and the bacterial NanA in a water box was compared between classical Intel Xeon Gold CPUs and two types of GPUs — Tesla K40 based on Kepler architecture, and the recently introduced Tesla P100 based on Pascal architecture. The maximum speed-up reached on CPUs and Tesla K40 GPUs were comparable; however, it took at most 4–6 GPU-nodes to gain the performance of 20+ CPU-nodes (Fig. 1,

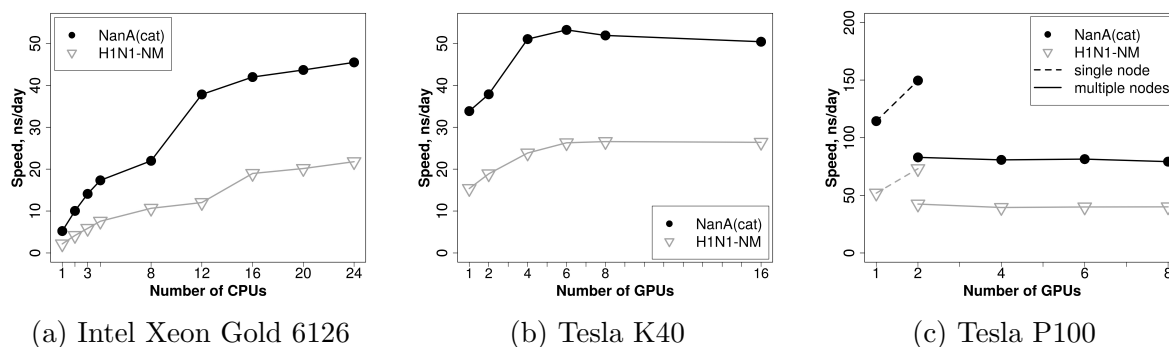


Figure 1. Efficiency and scalability of the MD implemented in AMBER on different hardware

a and b). Therefore, the CPU-based acceleration of MD is possible, but not efficient. The computational yield of Pascal-based cards, although superior in absolute values, was qualitatively different compared to its Kepler-based predecessor. The speed of MD on a single Tesla P100 was 1.95x–2.14x times higher than the top output from 4–6 nodes equipped with Tesla K40 (Fig. 1, c). However, the performance of Tesla P100 did not improve when executed on multiple nodes, instead it degraded by 22–28%. The use of multiple Tesla P100 to run MD in parallel was efficient only within a single node in the peer-to-peer mode, when GPUs communicate directly via the PCI-E bus, providing the speed-up of 1.31x–1.41x with the yield of 73–150 ns/day. Therefore, a microsecond-long trajectory can be reached in at most 1–2 weeks.

The performance of the above-mentioned hardware was compared with special-purpose supercomputers by using a previously described benchmark based on a small protein system (see Section 1.2 of Methods). The peak performances of Anton-1 and Anton-2 machines were ~ 40 x and ~ 228 x times superior to two Tesla P100 GPUs in the peer-to-peer mode (Fig. 2).

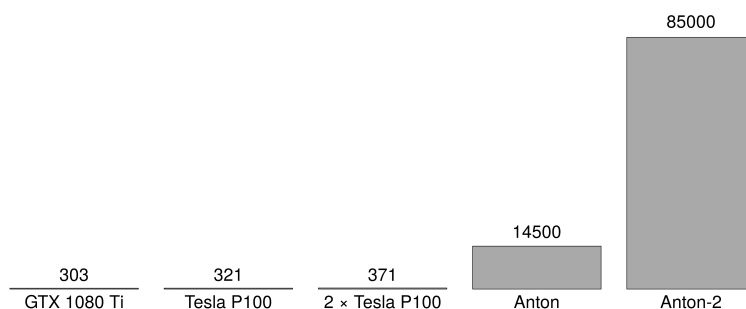


Figure 2. Speed of the MD at different GPU setups and Anton supercomputers (in ns/day)

Discussion and Conclusions

Molecular dynamics has a potential of becoming a particularly important tool in drug discovery to evaluate ligand recognition and binding to a flexible target site, but its use at a daily routine has been limited by a significant computational cost. We have shown that MD simulation of the structural dynamics of large Neuraminidases on modern GPUs can reach the microsecond timescale. This opens an opportunity to move away from the “static” affinity-driven strategies towards a deeper understanding of conformational dynamics in structures of proteins and mechanisms of association/dissociation of a drug candidate from a target site, leading to a better selection of promising compounds *in silico* to be further profiled experimentally. With the de-

velopment of more powerful video cards, the use of more than one GPU-node per simulation is becoming inefficient, leaving GPUs yet far behind the deeply-specialized Anton supercomputers, and thus supporting the recent trend for co-design of software and hardware for a particular purpose, of which Antons are an example, as a strategy to accelerate calculations. Further development of affordable specialized architectures is needed to move towards the much-desired millisecond timescale to simulate structural behavior of large proteins at a daily routine.

Acknowledgments

This work was carried out with the financial support of the Russian Science Foundation grant No. 15-14-00069-P. Supercomputer modeling was performed using equipment of the shared research facilities of HPC computing resources at the Lomonosov Moscow State University supported by the project RFMEFI62117X0011 [4].

This paper is distributed under the terms of the Creative Commons Attribution-Non Commercial 3.0 License which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is properly cited.

References

1. Abed, Y., Boivin, G.: A review of clinical influenza A and B infections with reduced susceptibility to both Oseltamivir and Zanamivir. *Open Forum Infectious Diseases* 4(3), ofx105 (2017), DOI: 10.1093/ofid/ofx105
2. Durrant, J.D., Bush, R.M., Amaro, R.E.: Microsecond molecular dynamics simulations of influenza neuraminidase suggest a mechanism for the increased virulence of stalk-deletion mutants. *The Journal of Physical Chemistry B* 120(33), 8590–8599 (2016), DOI: 10.1021/acs.jpcc.6b02655
3. Ganesan, A., Coote, M.L., Barakat, K.: Molecular dynamics-driven drug discovery: leaping forward with confidence. *Drug Discovery Today* 22(2), 249–269 (2017), DOI: 10.1016/j.drudis.2016.11.001
4. Sadovnichy, V., Tikhonravov, A., Voevodin, V., Opanasenko, V.: Lomonosov: Supercomputing at Moscow State University. In: *Contemporary High Performance Computing: From Petascale Toward Exascale* (Chapman & Hall/CRC Computational Science). pp. 283–307. CRC Press Boca Raton, Fla, USA (2013), DOI: 10.1201/2F9781351104005
5. Sharapova, Y., Suplatov, D., Švedas, V.: Neuraminidase a from streptococcus pneumoniae has a modular organization of catalytic and lectin domains separated by a flexible linker. *The FEBS Journal* 285(13), 2428–2445 (2018), DOI: 10.1111/febs.14486
6. Shaw, D.E., et al.: Anton 2: Raising the Bar for Performance and Programmability in a Special-purpose Molecular Dynamics Supercomputer. In: *International Conference for High Performance Computing, Networking, Storage and Analysis, 2014*, 16–21 Nov. 2014, New Orleans, LA, USA. pp. 41–53. IEEE Press (2014), DOI: 10.1109/SC.2014.9
7. Xu, Z., et al.: Sequence diversity of nana manifests in distinct enzyme kinetics and inhibitor susceptibility. *Scientific Reports* 6, 25169 (2016), DOI: 10.1038/srep25169