JOINT INSTITUTE FOR NUCLEAR RESEARCH

Frank Laboratory of Neutron Physics

FINAL REPORT ON THE

INTEREST PROGRAMME

MOLECULAR DYNAMICS SIMULATION RESEARCH (FROM ATOMIC FRAGMENTS TO MOLECULAR COMPOUND)

Supervisor:

Prof. Kholmirzo Kholmurodov

Student:

Emmanuel Oluwaseyi Atofarati

University of Pretoria, South Africa

Participation period:

February 14 – March 25, Wave 6

South Africa, 2022

Table of Contents

Abstract

As the need for efficient predictive investigation of the molecular trajectory, intermolecular force of attraction or repulsion in molecules relevant for drug designers, biochemist, and material scientist arises, molecular dynamics simulation has played a vital role in solving some of these problems. In this report the fundamental process for molecular dynamics simulation will be examined and my future prospect in molecular dynamics will be presented.

Introduction

For efficient predictive investigation of the molecular trajectory, intermolecular force of attraction or repulsion in molecules relevant in drug design, biochemical samples, and material engineering, the need for structured calculating tools such as molecular dynamic simulating tool arise. Molecular dynamics simulation have been used to solve scientifically related problems which include proteins in different solvents, macro molecular complex compounds (ribosomes or nucleosomes), [1] and micro-/nano-scale heat transfer [2]–[4]. In recent times molecular dynamics simulation has become a multi-disciplinary tool used in chemistry-physics, biochemist, neuroscientist, optogenetics, drug-design, pharmacophore, engineers among others [2]–[5].

Molecular dynamics is a process of numerically simulating the intricate time dependent kinetics, bonding and thermodynamic behavior of atoms and molecule. It involves computer modelling from atomic point of view to molecular level. In molecular dynamic (MD) conformational ensemble can be explored for any given molecule. Many studies have considered the conformational ensemble profile of large molecules like RNAs and proteins, their shapes and the X-ray structures' refinement using molecular dynamics [6], [7]. Usually, MD is carried out by numerically solving the Newton's equations of motion for the system of molecules considered [7]. MD simulation was first carried out in late 1970's by Mc Cammon Andrew, Gelin Bruce and Martin Karplus in their ground breaking and Noble Prize winning studies on folded protein dynamics [1], [5], [8].

Some common software used for molecular dynamics simulation are Amber (Assisted Model Building with Energy Refinement), DL_Poly, Desmond CHARMM (Chemistry at Harvard Macromolecular Mechanics), NAMD (Nano-scale Molecular Dynamics), OpenMM, GROMACS (Groningen Machine for Chemical Simulation), GROMOS (Groningen Molecular Simulation), AIMD (Ab Initio Molecular Dynamics), DFT (Density function Theory), Martini, WEIN2K, VASP (Vienna Ab Initio Simulation) and LAMMPS (Large-scale Atomic/Molecular Massively Parallel Simulator) [6], [9], [10]. Baretto et al. [10] summarized some molecular dynamics software or force field used for predicting G-protein coupled receptors in different literatures. Most frequent software package used for visualizing MD simulation results is VMD (Visual Molecular Dynamics).

According to Hospital et al. [1] the solvent representation is a vital issue while defining systems in molecular dynamics. Implicit solvent model [11]–[13], generalized born model [14], and explicit solvent models [15]–[17] have been considered in previous studies. However, Hospital et al. concluded that the most accurate solvent representation approach is the simplest ones i.e., explicit representation of molecules of the solvent. The explicit solvent representation, accounts for the molecules of solvents in a more accurate but highly computational costly way, and it often involve assumption of periodic boundary condition. However, Implicit solvent approach approximates the average solvent effect using mathematical models, though faster but less accurate

In this report, basic introduction to molecular dynamics simulation will be discussed. The fundamental equations, potentials and simulation techniques, numerical code for simulating of liquid model (Lenard-Jones potential), and the process for using general-purpose code for the simulation of ionic, polymeric and biochemical molecular systems.

Project Goals

The goal of this project is to briefly introduce MD research Simulation and design of physical and biochemical nanostructures, systems and compound. Also, computer molecular design of new structures with given (by experiment) parameters and conditions will be presented.

Scope of Work

The Scope of this study will be as follows.

i. The fundamental equations, potentials fields and simulation techniques.
ii. Numerical code explanation for simulation of liquid model (Lenard-Jones potential).
iii. Introduction to the fundamental steps for simulating molecular dynamics in systems, atomic structures, ionic structures, macromolecules, polymers, and biological molecules such as Protein will be studied

Methodology

In this project basic steps for simulating molecular dynamics in systems, atomic structures, ionic structures, macromolecules, polymers, and biological molecules such as Protein will be studied. Generally, the steps involved in carrying out molecular dynamic simulation are;

i. Selection of an interaction models (either pairs, triplet, or quadruplet of particles).
ii. Selection of boundary conditions (configuration or position, forces, velocities).
iii. Select the conformational ensemble (canonical ensemble (NPT), isothermal-isobaric ensemble or the microcanonical ensemble (NVE), NAMD, AMBER, …).
iv. Select the target temperature, density or pressure.
v. Select the integrator, thermostat, barostat to be utilized.
vi. Perform the simulation.
vii. Analyze the result using post-processing.

Fundamental Equations and Simulation Techniques

In molecular dynamics, the conventional Newton's law of motion is used to study the energy surface of each atoms/molecule by resolving the interatomic motion of the systems.

$$\vec{F} = m\vec{a}$$

Where, m is the mass of the atom, $\vec{F}$ is the atomic force and $\vec{a}$ acceleration of the atom. It is vital to recollect that atom are always in continuous motion and that this motion tends to affect their energy surface. The ordinary differential equation form of the Newton's equation numerically solved in molecular dynamics simulation is as follows.

$$m_i \frac{d^2 r_i\,(t)}{dt^2} = F_i(r), \quad i - 1,2, \dots, n$$

$$\{r_i, m_{i,} F_i\}$$

$$r = \{r_i, r_{2,\dots,} r_n\}; U(r)$$

$$F_i(r) = -\frac{\partial U(r)}{\partial r_i}\}$$

$$m_i \frac{d^2 r_i\,(t)}{dt^2} = F_i(r_i(t)) - \gamma_i m_i \frac{dr_i\,(t)}{dt} + R_i(t)$$

Force Fields Potentials

In molecular dynamics, atomic forces can either be bonded forces or non-bonded forces. Examples of the bonded force are bond angle bending, bond stretching, bond rotation/torsion (dihedral potential), while examples of the non-bonded forces are electrostatic potential, Van Der Waals interaction potential or Leonard-Jones potential. A typical illustration of the potential function and some of the bonding and non-bonding forces are shown in Figure 1.

Generally, force field describes the mathematical models that relates the system energy to its particles' coordinate during motion. The force field is the combination of atomic types, mathematical equations and related parameter. Interatomic potential energy $\boldsymbol{U}$ ($\mathbf{r}_1$, $\mathbf{r}_2$, ..., $\mathbf{r}_N$) and a set of parameters analytically inputted into this form forms the force field. Parameters used for the force field is usually derived from empirical data, such as neutron scattering, neutron spectroscopy, X-ray diffraction, infrared etc. Also,

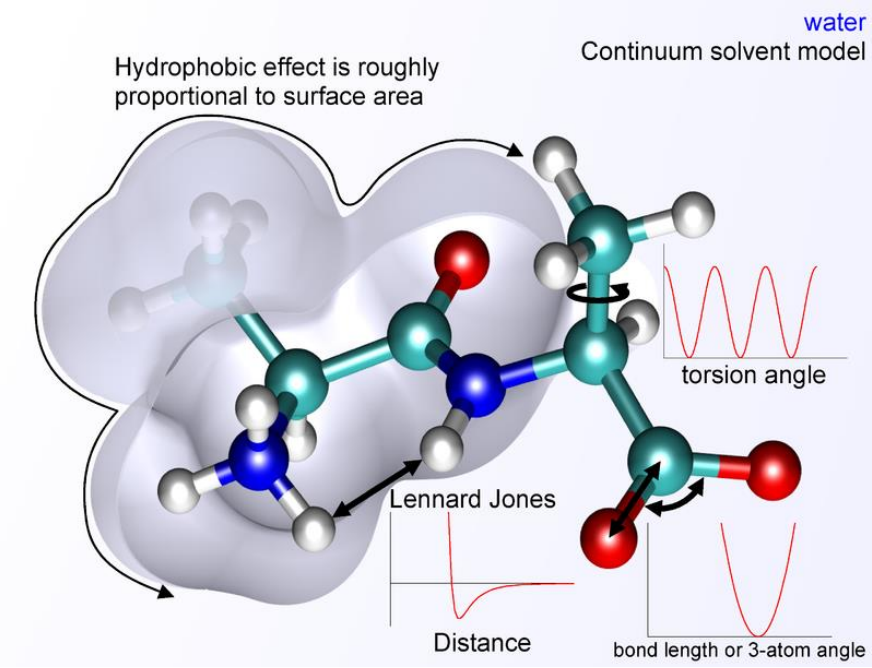


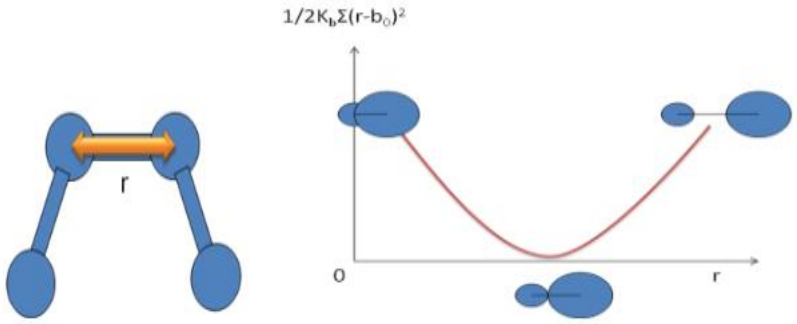
Figure 1. Some Typical Force Field Potentials Energy function in Water (Source from Wikipedia [18]**)**

parameters can be obtained from quantum mechanics correlation or Ab Initio. In recent times the existing and newly developed force field potentials have been optimized and adapted in several

multi-purpose software like DL_POLY, CHARMM, AMBER, NAMD, etc. Some of the common force field potential function are mathematically defined below.

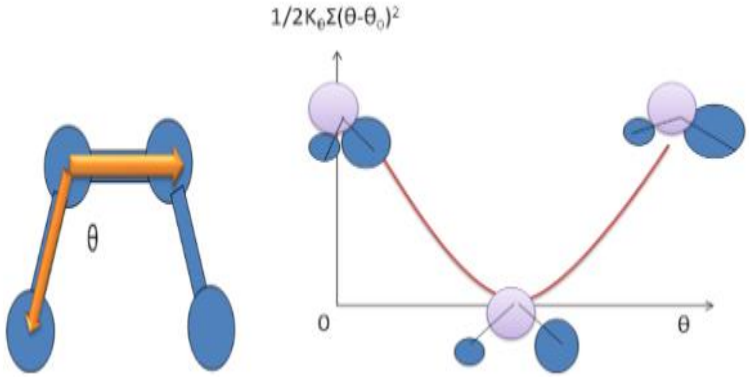
The valency length potential (U_b) is calculated using;

$$U_b = \frac{1}{2}\sum_b K_b(r - b_o)^2$$



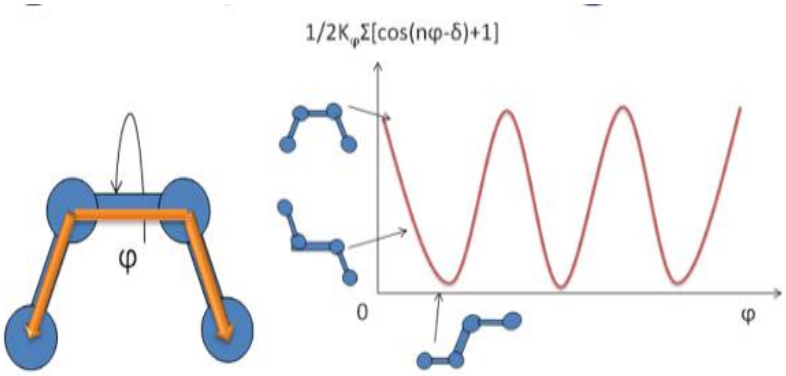
The Valency bond angle potential (U_θ) can be obtained using

$$U_\theta = \frac{1}{2}\sum_\theta K_\theta(\theta - \theta_o)^2$$



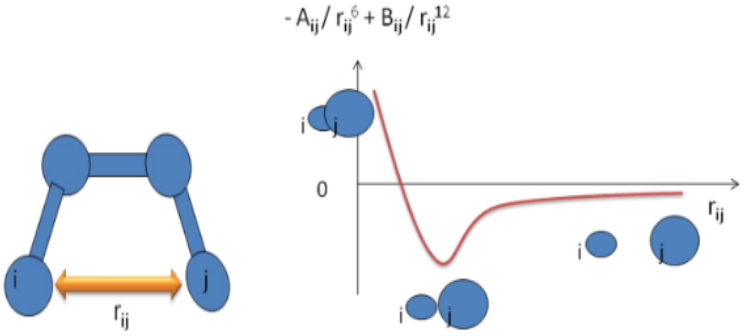
The torsional dihedral potential (U_φ) is calculated from'

$$U_\varphi = \frac{1}{2}\sum_\varphi K_\varphi[cos(n\varphi - \delta) + 1]$$



The Van Der Waals interaction potential or Leonard-Jones potential (U_{LJ}) is obtained using the equation

$$U_{LJ} = \sum_{i,j}\left[\frac{A}{r_{ij}^{12}} - \frac{B}{r_{ij}^{6}}\right]$$



The electrostatic potential (U_{el}) is calculated using

$$U_{el} = \sum_{i,j}\left[\frac{q_i q_j}{\varepsilon r_{lj}}\right]$$

The hydrogen bonding potential (U_{HB}) is calculated by

$$U_{HB} = \sum_{i,j}\left[\frac{A'}{r_{ij}^{12}} - \frac{B'}{r_{ij}^{10}}\right]$$

Hence the total force field potential in the system is obtain from the sum of all bonding forces and non-bonding force as shown in equation below. Bunker et al.[19] summarized the force potential function and their diagram as shown in Figure 2.

$$U\,(r) = U_b + U_\theta + U_\varphi + U_{lj} + U_{el} + U_{HB}$$

Velocity Generation

The actual molecular speed distribution of the atom/molecules is termed the maxwell-Boltzmann distribution. The maxwell-Boltzmann distribution is derived from statistical-mechanics domain. The velocity distribution is denoted by ($f(v_x)$, v_p is the most probable speed, v_m is the mean speed and v_{rms} is the root mean squared speed. The mathematical expression forms the maxwell distribution is presented as follows.

$$T(t) = \frac{1}{3Nk_B}\sum_{i=1}^{n} m_i v_i^2,$$

$$v_i = \frac{dr_i}{dt}$$

$$f(v_x) = 4\pi\left[\frac{M}{2\pi RT}\right]^{\frac{3}{2}} v^2 e^{\left[\frac{-Mv^2}{2RT}\right]}$$

$$v_p = \sqrt{\frac{2RT}{M}}$$

$$v_m = \sqrt{\frac{8RT}{\pi M}}$$

$$v_{rms} = \sqrt{\frac{3RT}{M}}$$

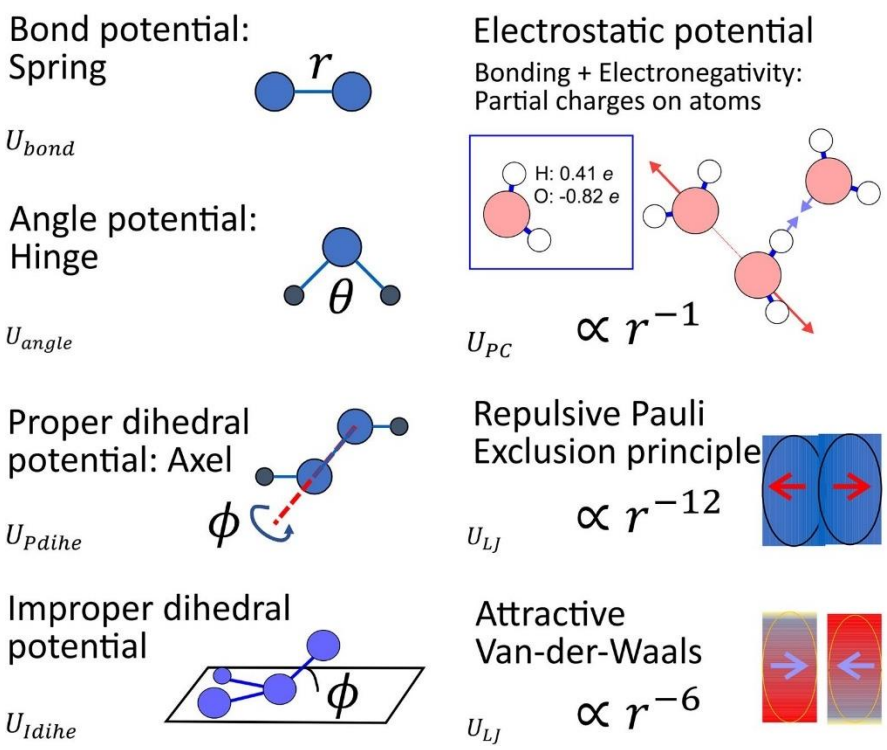


Figure 2. Force Field Potentials Functions and Schematics (sourced from Bunker et al. [19])

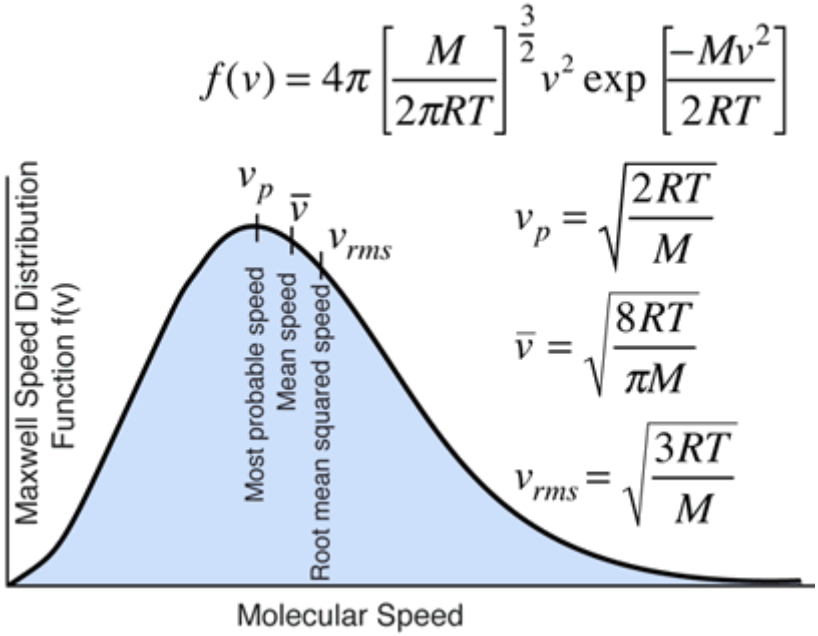


Figure 3. Maxwell-Boltzmann Distribution Function Plot

Selection the Integrator, thermostat, Barostat

In molecular dynamics simulation studies, thermostatting the particles implies controlling their temperature, while barostatting the particles implies controlling the pressure. There are different thermostat algorithms in literatures, namely; Andersen thermostat Algorithm, Nosé-Hoover thermostat Algorithm, Langevin thermostat Algorithm and the Berendsen thermostat Algorithm. While some of the thermostat algorithm follows the statistical approach others follows the deterministic approach. The Andersen thermostat Algorithm involves the simulation of some heat bath using the stochastic impulsive force that acts alternatingly on arbitrarily chosen particles. For the Nosé-Hoover thermostat algorithm, the equation of motion with a heat exchange (dissipation, friction) is solved and the numerical realization is a discrete finite-difference algebraic equation.

A barostat can either change the volume considered isotropically or anisotopically. The Anisotropic Barostat is common Barostat algorithm achieved using the Virial-theorem for non-periodic boundary condition given as

$$P = \rho T + \frac{1}{Vd}\left[\sum_{i=1}^{N} {r_i}^T f_i\right]$$

Ewald summation

According to Wells et al.[20]In order to calculate the electrostatic potential and forces numerically, the Cutoff-based approach have been utilized however it has several drawbacks, The Ewald summation for Coulomb interaction is a highly accurate approach taking into account periodicity. The Ewald summation approach utilizes a convergence function and a Fourier transform, converts the single conditional or slowly converging sum into two rapidly converging sums. Its, rapid convergence sums make the process highly accurate and avoids the computation requirement that occurs dues to cutoff-based method. The ewald sums are usually realized under speacila-purpose architectures like MDGRAPE-2 & 3 as shown in Figure 4. While Figure 5. shows the typical computational cost for some common MD-simulation.

Conclusion

In conclusion the basic steps for simulating molecular dynamics in systems, atomic structures, ionic structures, macromolecules, polymers, and biological molecules such as Protein have been studied.

Future Research Prospect

I would like to study heat transfer performance in nanofluids using molecular dynamic simulation after my PhD. Research is concluded.

Acknowledgement

I want to acknowledge the effort of my trainee Prof. Kholmirzo Kholmurodov, who despite his busy schedule find time to sow to we upcoming researchers. Also, I would like to appreciate my wife who encouraged me to find some other side training while conducting my PhD. research.

(a) (b)

Figure 4. Structure of MDGRAPE-3 Accelerator (a) Memory-Chip and (b) Systems

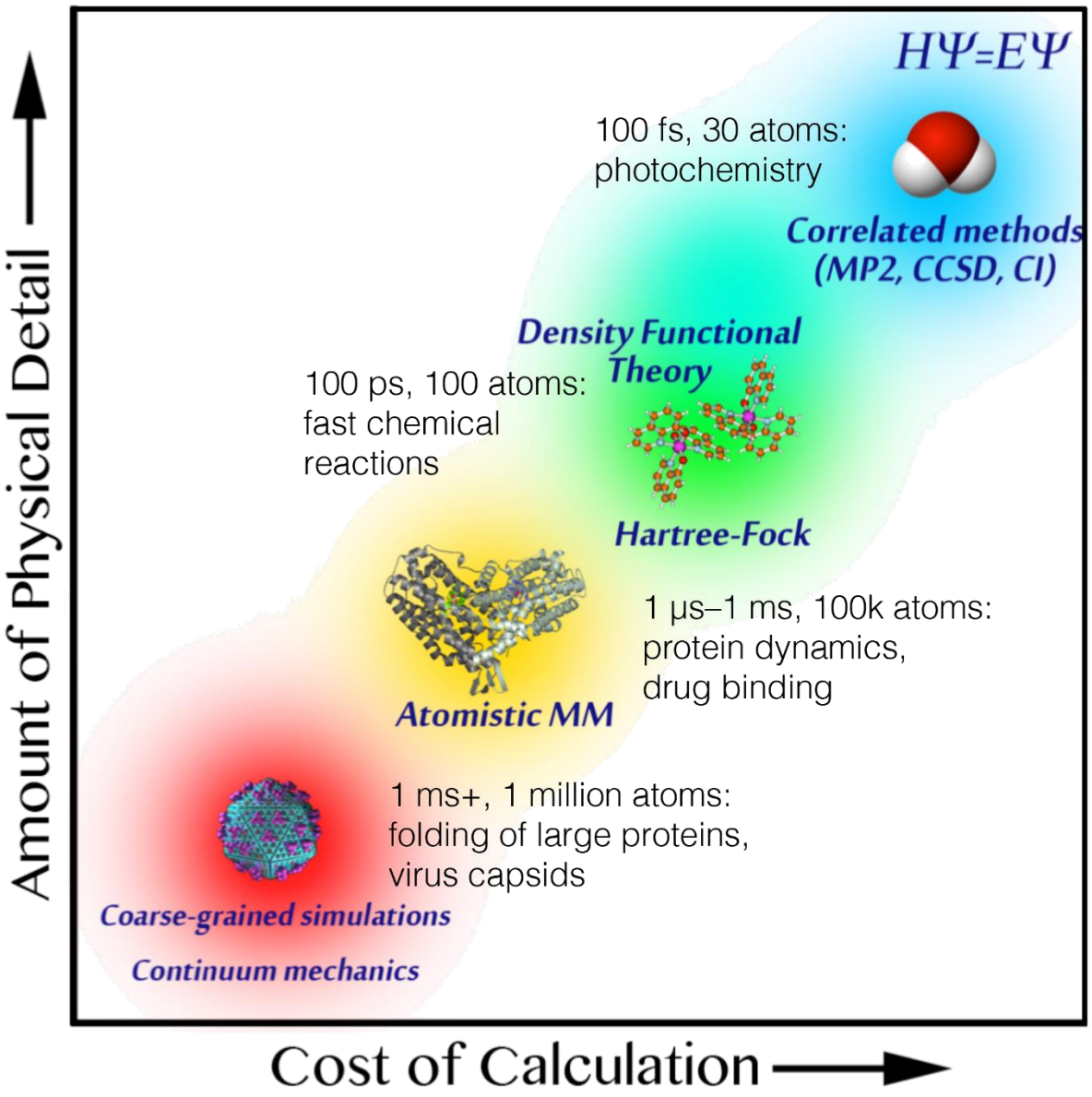


Figure 5. Typical Computational Cost for some common MD-Simulation

Reference

[1] A. Hospital, J. R. Goñi, M. Orozco, and J. L. Gelpí, "Molecular dynamics simulations: advances and applications," *Adv. Appl. Bioinforma. Chem.*, vol. 8, no. 1, pp. 37–47, Nov. 2015, doi: 10.2147/AABC.S70333.

[2] J. Chen, G. Zhang, and B. Li, “Molecular Dynamics Simulations of Heat Conduction in Nanostructures: Effect of Heat Bath,” *http://dx.doi.org/10.1143/JPSJ.79.074604*, vol. 79, no. 7, Jun. 2010, doi: 10.1143/JPSJ.79.074604.

[3] S. Maruyama, “Molecular Dynamics Method For Microscale Heat Transfer,” in *Advances in Numerical Heat Transfer, Volume 2*, 1st Editio., W. . Minkowycz and E. . Sparrow, Eds. CRC Press, 2018, p. 38.

[4] S. Y. Daisy, “Molecular Dynamics Study of Solid-Liquid Heat Transfer and Passive Liquid Flow,” *Diss. - ALL*, Aug. 2016, Accessed: Mar. 24, 2022. [Online]. Available: https://surface.syr.edu/etd/632.

[5] S. A. Hollingsworth and R. O. Dror, “Molecular Dynamics Simulation for All,” *Neuron*, vol. 99, no. 6, pp. 1129–1143, Sep. 2018, doi: 10.1016/J.NEURON.2018.08.011.

[6] J. Polanski, “Chemoinformatics,” in *Comprehensive Chemometrics*, vol. 4, Elsevier, 2009, pp. 459–506.

[7] A. R. Leach, “Ligand-Based Approaches: Core Molecular Modeling,” in *Comprehensive Medicinal Chemistry II*, vol. 4, Elsevier, 2007, pp. 87–118.

[8] J. A. McCammon, B. R. Gelin, and M. Karplus, “Dynamics of folded proteins,” *Nat. 1977 2675612*, vol. 267, no. 5612, pp. 585–590, 1977, doi: 10.1038/267585a0.

[9] X. Han, W. H. Shin, C. W. Christoffer, G. Terashi, L. Monroe, and D. Kihara, “Study of the Variability of the Native Protein Structure,” *Encycl. Bioinforma. Comput. Biol. ABC Bioinforma.*, vol. 1–3, pp. 606–619, Jan. 2019, doi: 10.1016/B978-0-12-809633-8.20148-9.

[10] C. A. V. Barreto *et al.*, “Prediction and targeting of GPCR oligomer interfaces,” in *Progress in Molecular Biology and Translational Science*, vol. 169, Academic Press, 2020, pp. 105–149.

[11] Y. N. Vorobjev, “Advances in Implicit Models of Water Solvent to Compute Conformational Free Energy and Molecular Dynamics of Proteins at Constant pH,” *Adv. Protein Chem. Struct. Biol.*, vol. 85, pp. 281–322, Jan. 2011, doi: 10.1016/B978-0-12-386485-7.00008-9.

[12] J. Kleinjung and F. Fraternali, “Design and application of implicit solvent models in biomolecular simulations,” *Curr. Opin. Struct. Biol.*, vol. 25, pp. 126–134, Apr. 2014, doi: 10.1016/J.SBI.2014.04.003.

[13] B. Roux and T. Simonson, “Implicit solvent models,” *Biophys. Chem.*, vol. 78, no. 1–2, pp. 1–20, Apr. 1999, doi: 10.1016/S0301-4622(98)00226-9.

[14] D. Bashford and D. A. Case, “Generalized Born Models of Macromolecular Solvation Effects,” *Annu. Rev. Phys. Chem.* , vol. 51, pp. 129–152, Nov. 2003, doi: 10.1146/ANNUREV.PHYSCHEM.51.1.129.

[15] P. C. Chen and J. S. Hub, “Interpretation of Solution X-Ray Scattering by Explicit-Solvent Molecular Dynamics,” *Biophys. J.*, vol. 108, no. 10, pp. 2573–2584, May 2015, doi: 10.1016/J.BPJ.2015.03.062.

[16] D. I. Dimitrov, A. Milchev, and K. Binder, “Polymer brushes in solvents of variable quality: Molecular dynamics simulations using explicit solvent,” *J. Chem. Phys.*, vol. 127, no. 8, p. 084905, Aug. 2007, doi: 10.1063/1.2768525.

[17] R. Salomon-Ferrer, A. W. Götz, D. Poole, S. Le Grand, and R. C. Walker, “Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh ewald,” *J. Chem. Theory Comput.*, vol. 9, no. 9, pp. 3878–3888, Sep. 2013, doi: 10.1021/CT400314Y/SUPPL_FILE/CT400314Y_SI_001.PDF.

[18] “Molecular mechanics - Wikipedia.” https://en.wikipedia.org/wiki/Molecular_mechanics#/media/File:MM_PEF.png (accessed Mar. 25, 2022).

[19] A. Bunker and T. Róg, “Mechanistic Understanding From Molecular Dynamics Simulation in Pharmaceutical Research 1: Drug Delivery,” *Front. Mol. Biosci.*, vol. 7, p. 371, Nov. 2020, doi: 10.3389/FMOLB.2020.604770/BIBTEX.

[20] B. A. Wells and A. L. Chaffee, “Ewald Summation for Molecular Simulations,” *J. Chem. Theory Comput.*, vol. 11, no. 8, pp. 3684–3695, Jul. 2015, doi: 10.1021/ACS.JCTC.5B00093.