JOINT INSTITUTE FOR NUCLEAR RESEARCH

FINAL REPORT ON

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

Supervisor:
Prof. Kholmirzo Kholmurodov

Prepared by
Alaa Elsadieque
Research Assistant
Alexandria University Cancer research cluster
Participation period:
Wave 7 (13 June - 22 July)

Egypt, 2022

Contents

Introduction:

Molecular dynamics: models a system of classical particles whose dynamics is described by Newton's equations and its generalizations [1]. Molecular dynamics simulates a classical system of N particles. The core of most simulations is to start with the initial positions and velocities of all particles and to then repeatedly apply a "recipe" to update each particle's position and velocity from time t to time t + Δt (see Fig. 1). The dynamics is governed by Newton's second law.

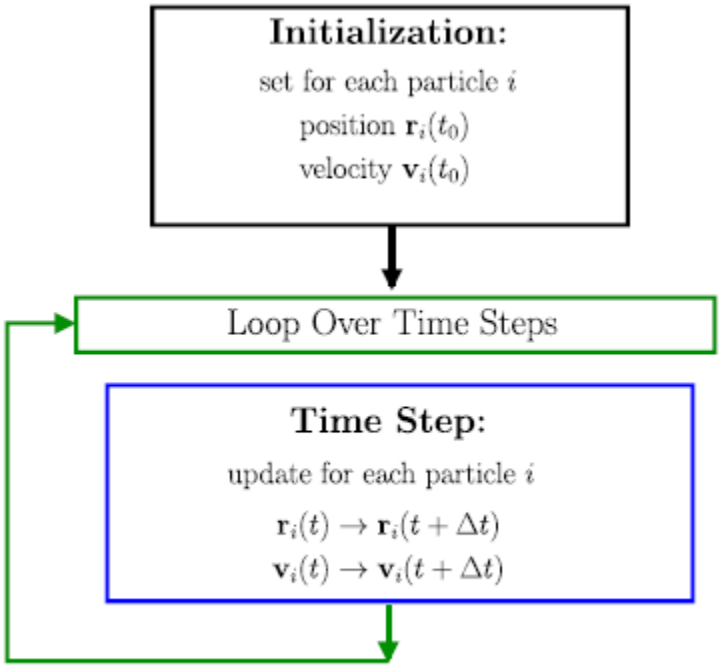

Fig. 1. Flow chart of a molecular dynamics simulation program. [1]

$$F_i = m_i a_i, \qquad (1)$$

where a_i is the acceleration of particle i.

Molecular dynamics (MD) simulations can provide not only plentiful dynamical structural information on biomacromolecules but also a wealth of energetic information about protein and ligand interactions. Such information is very important to understanding the structure-function relationship of the target and the essence of protein–ligand interactions and to guiding the drug discovery and design process [2].

Potential energy function:

The forces acting on the atom i can arise from both internal and external sources. The internal sources are basically the interaction forces acting between bonded and non-bonded atoms. The external sources can be environmental stresses including electric field, heat, and pressure, which are imposed on the system externally according to molecular mechanics. To measure the forces, firstly we should think about how to determine the potential energy of the system that can be expressed as a sum of bonded, non-bonded, and cross-term interactions. The derivatives of the potential energy function are known as force fields.[3]

$$U(r) = U_b + U_\theta + U_\varphi + U_\omega + U_{LJ} + U_{el} + U_{HB} + \cdots \qquad (2)$$

The valency length potential (Ub) 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 ($U\varphi$) 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 (ULJ) 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 (Uel) is calculated using

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

The hydrogen bonding potential (UHB) is calculated by

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

Force Fields Potentials:

Generally, force field describes the mathematical models that relate 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 U (r_1, r_2,… , 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, multi-purpose software like DL_POLY, CHARMM, AMBER, NAMD, etc.

Simulation methods (applications)

Molecular dynamics (MD) simulation, since the late 70s, has advanced from simulating several hundreds of atoms to systems with biological relevance, including entire proteins in solution with explicit solvent representations, membrane embedded proteins, or large macromolecular complexes like nucleosomes or ribosomes. Simulation of systems having ~50,000–100,000 atoms are now routine, and simulations of approximately 500,000 atoms are common when the appropriate computer facilities are available. This remarkable improvement is in large part a consequence of the use of high performance computing (HPC), and the simplicity of the basic MD algorithm.

- ***Molecular Dynamics Simulation of Proteins:***
  Molecular dynamics simulations allow protein motion to be studied, by following their conformational changes

through time. Proteins are typically simulated using an atomic-level representation, where all or most atoms are explicitly present. While the simplest simulation system would entail a single protein molecule solvated in water, other relevant molecules such as ligands or lipid bilayers may also be included, so long as they can all be modeled using the same force field.

The chemical and resultant geometric and mechanical properties of molecules are described in terms of interatomic (see figure 2) "bonded" (bonds, bond angles, torsional angles) and "nonbonded" (van der Waals, Coulombic) potential energy terms—the so-called force field—which sum to give the total potential energy of a system of N atoms.

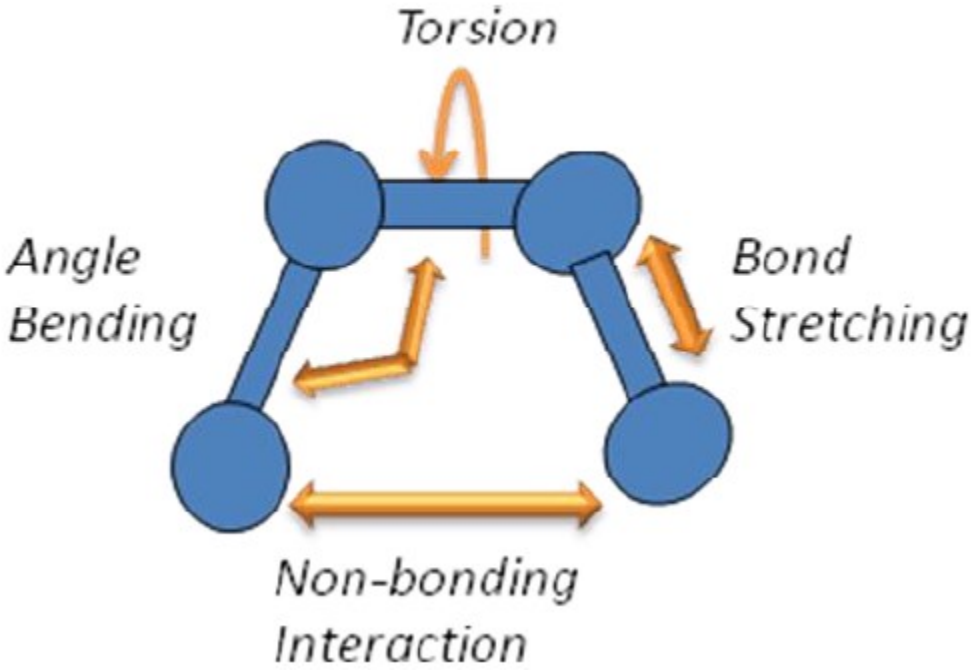

Fig.2 . *Chemical bonds (bond stretching, angle bending, torsion) and non-bonding interaction*

- ***Material :***

  Ensure suitable computing equipment is available with required software installed, choose the force field and solvent model, and obtain all input files before embarking.

- ***Computer:***

  1. Hardware: Minimum specifications are a quad-core machine, e.g., Intel i7, ideally with a compatible CUDA-enabled GPU card and with at least 1 GB RAM per core and 100 GB disk space.

  2. Operating system: Ideally UNIX or Mac. If Windows, run Linux via a virtual machine. Most (all) MD engines are difficult to compile in Windows.

- ***Software:***

  1. MD engine: Commonly used and well-maintained options with active user communities include GROMACS , NAMD, CHARMM, and AMBER, (see table 1) although only the former two options are free and can accommodate a variety of different force fields. As with most modern codes, all four are highly parallelized for CPUs and GPUs.

  2. Preparation and analysis: Most MD engines, including those listed above, provide programs for simulation preparation and analysis. Other options for bespoke analysis include the MDAnalysis package. Basic scripting (e.g., bash) or programming skills (e.g., python) are also helpful for manipulating file and data formats and post-processing data.

3. Visualization: Additional software is required to view simulation coordinate trajectories. VMD is free and compatible with the trajectory file formats produced by most MD engines. PyMOL is another common option.
4. Graphical software: Common options for plotting analysis data include gnuplot, xmgrace, qtgrace, R, and even Excel.

- ***Coordinates:***

  1. Initial (Cartesian) coordinates for proteins can often be obtained from the Protein Data Bank (PDB) in PDB file format. These should be checked for missing or incomplete residues, poorly resolved electron density, mutations introduced to aid crystallization, etc. Missing residues can be built using software such as Modeller or DeepView.
  2. Software such as VMD can generate chemically plausible coordinates for biopolymers. Modeller can build structures based on homology; there are also myriad homology modeling web servers.
  3. Coordinates for small molecules may be available from the Cambridge Structural Database or generated using software such as Avogadro.
  4. Coordinates not derived from experimental data should ideally be optimized using quantum chemical calculations or extensive MD simulations.
  5. Pre-equilibrated boxes of solvent coordinates are available for common solvents such as water.

Biomolecular force field	General-purpose force field	Water model
CHARMM [2–4]	CGenFF [36]	TIP3P [71], cTIP3P[a] [72, 73]
AMBER [5]	GAFF [74]	TIP3P, TIP4P
GROMOS [7, 8, 9]	N/A	SPC [75], SPC/E [76, 77]
OPLS [10, 11]	N/A	TIP3P

Table 1 Common biomolecular force fields, their corresponding general force fields, and the most commonly used compatible water models

- ***Parameters:***

  1. Force field files list parameter values and specify their assignment to atom types and to bonded interactions between particular combinations of atom types. Each MD software package requires these files to be in a particular format and typically provides these files for one or more commonly used force fields.

  2. Each force field may use different terms, and even if these are the same, parameter values are not interchangeable between force fields as they are parameterized using different strategies.

  Thus, all molecules included in the simulation must be described using the same or a compatible force field (Table 1).

  3. Molecule or fragment "building blocks" provide parameters for a particular molecule; for polymers such as

proteins, preparation programs provided with the MD software join together amino acid building blocks to generate the molecular “topology” .
4. If there is not a building block available, the molecule must be parameterized. This may be done manually or using automated software such as the ATB (GROMOS), CherryPicker, CGenFF (CHARMM) or antechamber (AMBER/GAFF), or LigParGen (OPLS).
5. Similarly, each force field is compatible with particular solvent models (Table 1).

- ***Simulation Settings:***

  1. MD run files contain the settings (e.g., algorithm choices) and associated parameter values for carrying out an energy minimization or MD simulation. The file format is specific to each MD package.
  2. Key settings for energy minimization:
  (a) Specification of energy minimization method. Table 1 Common biomolecular force fields, their corresponding general force fields, and the most commonly used compatible water models Biomolecular force field General-purpose force field Water model CHARMM; The CHARMM force field was originally parameterized using a modified version of the TIP3P model, cTIP3P, which adds Lennard-Jones interactions between the hydrogen and oxygen atoms, but the latest version, CHARMM36,was optimized.
  (b) Size and maximum number of minimization steps

(c) Criteria for convergence.
(d) Frequency and properties to write to file. 3. Key settings for initialization and heating (typically run in the NVT.
(e) The use of constraints, constraint algorithm, and associated parameters.
(f) Distances within which to explicitly calculate nonbonded interactions.
(g) Method for treatment of long-range electrostatic interactions and associated parameters. [4]

Conclusion

Molecular physics needs realistic non-probability molecular simulation such as molecular dynamic MD for research purposes as targeted therapy used with all precutions in cancer treatment to save effort and time. There is a lot that we did not get through experiment, but of course we will try through molecular simulations so Medical physicists can benefit significantly from molecular dynamic.

Future Goals

After I have finished writing my master thesis of medical biophysics I want to use computational simulation specially in using irradiation nanoparticles for targeted therapy in the coming research project.

Acknowledgement

All appreciation for Professor Kholmirzo Kholmurodov who has introduce the course with perfectly organized and regular weekly lectures and very caring about what we understand actually

References:

1. Vollmayr-Lee, K. (2020). Introduction to molecular dynamics simulations. American Journal of Physics, 88(5), 401-422.
2. Liu, X., Shi, D., Zhou, S., Liu, H., Liu, H., & Yao, X. (2018). Molecular dynamics simulations and novel drug discovery. Expert opinion on drug discovery, 13(1), 23-37.
3. KHUSENOV, M., NEMATOV, D., BURHONZODA, A., KHOLMURODOV, K., DOROSHKEVYCH, A., DOROSHKEVYCH, N., ... & MAJUMDER, S. MOLECULAR DYNAMICS OF NANOSCALE PHENOMENA: COMPUTER DESIGN FOR NEW DRUGS AND MATERIALS. changes, 9, 14.
4. Collier, T. A., Piggot, T. J., & Allison, J. R. (2020). Molecular dynamics simulation of proteins. In Protein Nanotechnology (pp. 311-327). Humana, New York, NY.