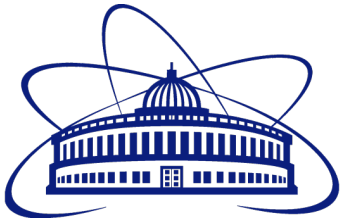

JOINT INSTITUTE FOR NUCLEAR REASEARCH
Frank Laboratory for Neutron Physics

FINAL REPORT ON THE INTEREST PROGRAMME

Introductory course: MD-simulation research (from atomic fragments to molecular compound)

Supervisor:
Prof. Kholmirzo Kholmurodov

Student:
An Binh Le, Vietnam
University of Science, VNU

Participation period:
Oct 30 - Dec 10, Wave 9

Dubna, 2023

Contents

Abstract

This paper is the final report for the INTEREST program at wave 9 and includes a general perspective on molecular dynamics simulation research. The course is to evaluate the empirical Newtonian equations, with potentials depicting interaction phenomena from atomic fragments to molecular compounds and simulation techniques. Also, the fundamental description for simulation of liquid model, ionic, polymeric, and biochemical molecular systems are illustrated by their application in realistic problems.

1 Introduction

Compared with experiments, computer simulations can tackle problems of great complexity, save time and expense, exercise better control over important intermediate processes, and track system responses on a temporal and spatial scale inaccessible by conventional experiments. As a result, a large portion of the key questions that physicists and chemists are devoted to solving can be traced back to the relationship between the bulk properties of matter and the underlying interactions among its constituent atoms or molecules.

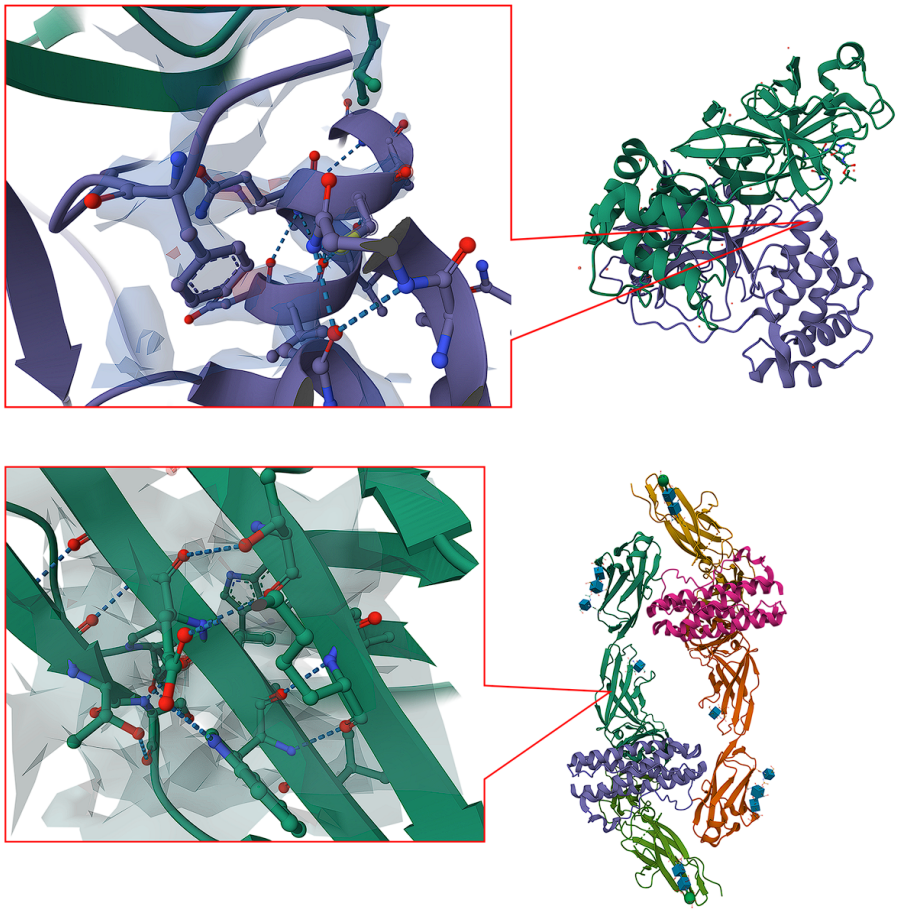

Figure 1.1. Structures of SARS-CoV-2 Mpro (Omicron, P132H) in complex with alpha-ketoamide 13b-K at pH 6.5 and interleukin 11 signalling complex, truncated gp130 with their electron density at a specific residue from top to bottom, respectively.

The MD simulation method applies classical Newtonian mechanics to describe the movement of atoms and obtain the dynamic properties of the modeled system (Fig.

1.1). An important characteristic of MD simulation is that the calculation of atomic movements is based on a solid physical foundation and is highly accurate. Hence, MD simulation can simultaneously provide the statistical and dynamic properties of a system and is applicable to a broad range of material systems. Due to the rapid development of the MD simulation method, different types of force fields have been formulated systematically for various molecular systems, such as biomolecules, polymers, metals, and semiconductors. With the increase in the processing speed of computing resources, MD simulations will be able to model larger and more complicated systems.

2 Simulation methods

2.1 The force field potential

The MD simulation method applies classical Newtonian mechanics to describe the movement of atoms and obtain the dynamic properties of the modeled system. In an MD simulation, the system obeys the classical Newton's laws of motion. Consider a molecular system containing N particles with their position vectors and momentum vectors denoted by $r_i = (x_i, y_i, z_i)$ and $p_i = (p_{i,x}, p_{i,y}, p_{i,z})$, respectively. The Hamiltonian H of the system is written as

$$H\left(\boldsymbol{R}^N, \boldsymbol{P}^N\right) = \sum_{i}^{N} \sum_{\alpha} \frac{p_{i,\alpha}^2}{2m_i} + U\left(\boldsymbol{R}^N\right). \tag{1}$$

Here, $R^N = \{r_1, r_2, \ldots, r_N\}$ and $P^N = \{p_1, p_2, \ldots, p_N\}$, which represent the generalized spatial coordinates and momenta of all the particles in the system, respectively. The first and second terms on the right-hand side represent the kinetic and potential energy of the particles, respectively; α represents each of the three directions x, y, z; m_i denotes the mass of the ith particle.

The force field acting on the system is conservative in nature, which means that the force acting on each particle in the system can be obtained based on the derivative of the potential energy of the system with respect to the position of the particle. The direction of the first-order derivative of the potential energy (i.e., gradient) indicates where the minimum lies, and the magnitude of the gradient indicates the steepness of the local slope. The energy of the system can be lowered by moving each particle

in response to the force acting on it. Hence, the force is given by the negative of the gradient and is written as

$$\boldsymbol{F}_i\left(\boldsymbol{R}^N\right) = -\frac{\partial U\left(\boldsymbol{R}^N\right)}{\partial \boldsymbol{r}_i}. \tag{2}$$

Hence, the movement of a particle can be described by Newton's second law as

$$m_i \frac{\mathrm{d}^2 r_i}{\mathrm{d}t^2} = \boldsymbol{F}_i\left(\boldsymbol{R}^N\right), \tag{3}$$

given the initial positions $\boldsymbol{R}^N$ and momenta $\boldsymbol{P}^N$ of the particles, the integration of the expressions in Eq. (3) with respect to time produces the trajectories of the particles. As a result, MD simulation is a deterministic method whereby with predefined initial conditions, the subsequent time evolution is determined in principle.

In MD simulation, the solution of Eq. 1 is more difficult to obtain because the movement of each atom is under the influence of its interactions with all the other atoms in the system. Hence, MD simulation is, in essence, handling a many-body problem. For a system containing N particles, there are $6N$ variables ($3N$ positions and $3N$ momenta) to solve for, and an analytical solution is virtually impossible to obtain since these variables are closely coupled. Using the finite difference method discussed above, the $6N$ variables can be decoupled, and the key to resolving the Hamiltonian in Eq. 1 is to obtain an expression for $\boldsymbol{F}_i\left(\boldsymbol{R}^N\right)$ in Eq. 3. In MD simulation, a conservative force field with a concomitant potential energy function $U\left(\boldsymbol{R}^N\right)$ is introduced to describe the interatomic interactions, and $\boldsymbol{F}_i\left(\boldsymbol{R}^N\right)$ is calculated using Eq. 2.

According to the complexities of the force fields for modeling molecular systems, they can be classified into three classes. For the first-class force fields, harmonic functions are used to describe the bond stretching, bond angle bending, and bond torsion, and their coupling effects are not counted. Nonbonding interactions are described by the LJ (Lennard-Jones) potential. For the second-class force fields, cubic and biquadratic functions are added to describe the bond stretching, bond angle bending, and bond torsion, with their coupling effects being considered. Exponentialtype functions are added to describe nonbonding interactions. In the third-class force fields, hyperconjugation and polarization effects are further included.

The five terms on the right-hand side of Eq. 4 correspond to the five types of

interactions illustrated in Fig. 2.1. Here, bond stretching and bond angle bending interactions are modeled by harmonic potentials that increase the potential energy as the bond length l_i and the bond angle θ_i deviate from their reference values $l_{i,0}$ and $\theta_{i,0}$, respectively. The parameters k_i^b and k_i^θ are material-dependent constants that determine the interaction strength.

$$U = \sum_{\text{bonds}} \frac{k_i^b}{2}\left(l_i - l_{i,0}\right)^2 + \sum_{\text{angles}} \frac{k_i^\theta}{2}\left(\theta_i - \theta_{i,0}\right)^2 + \sum_{\text{torsion}} \frac{V_n}{2}\left(1 + \cos\left(n\phi - \phi_0\right)\right) + \sum_{i=1}^{N}\sum_{j=i+1}^{N}\left(\frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} + 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]\right). \quad (4)$$

The third term represents the contribution of the bond torsion to the total potential energy of the system. The variable V_n describes the energy barriers for the bond torsion, n denotes the number of minimum points in the potential energy function as the bond is rotated 360°, ϕ denotes the torsion angles, and ϕ_0 denotes the values of ϕ at the points of minimum potential energy. The fourth term represents Coulombic interactions, where q_i and q_j are the charges, r_{ij} is the distance between two charges, and ε_0 is the permittivity of free space or vacuum permittivity. And the last term, modeled by the LJ potential, describes van der Waals interactions. The diameter or distance σ_{ij} at which the interparticle potential becomes zero, ε_{ij} is the wall depth, and r_{ij} is the distance between the particles. The intermolecular distance (r) between the particles plays a very significant role in determining the intermolecular force. The repulsive term r_{ij}^{-12} elucidates the short-range overlapping of electron orbitals, whereas attractive term r_{ij}^{-6} explains the long-range intermolecular forces. In short, it can be said that the particles repel each other at small values of r_{ij} and attract at higher values of r_{ij}.

Examples of potential functions that follow the above functional forms with different levels of modification include the AMBER (assisted model building with energy refinement), CHARMM (Chemistry at Harvard Macromolecular Mechanics) and MM (molecular mechanics) force fields, CVFF (consistent-valence force field), and PCFF (polymer consistent force field). It should be noted that in the subject of chemistry, it is a common practice to directly describe a force acting on a particle without defining its corresponding potential. Hence, interaction potential functions are often termed "force fields". These force fields have been designed for a wide range of material

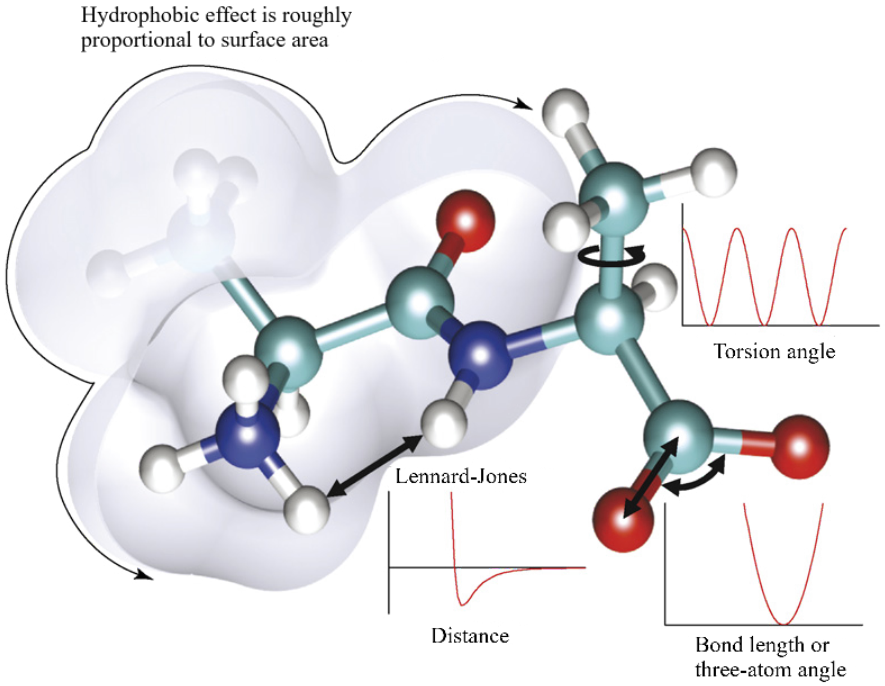

Figure 2.1. Schematics of the possible interactions in material systems, such as semiconductors and complex molecular systems.

systems and are usually considered as generalized force fields.

2.2 System initialization

2.2.1 *Position*

To perform a simulation, an MD box should first be defined, and the positions and velocities of atoms should be initiated. If system initialization is not properly performed, it will usually take a long time for the system to reach thermal equilibrium. There are two types of MD simulation initialization, which are starting a simulation from scratch or continuing from a previous simulation.

If the simulation is started from scratch, the positions of the atoms will usually be initiated based on a lattice structure. The lattice constants are set according to actual reported values from experiments or theoretical calculations.

2.2.2 *Velocity*

In principle, the initial velocities of the atoms can take any arbitrary values, provided that system stability is retained. After the simulation ensues for a certain amount of

time, the atomic velocities will be automatically refined to approach reasonable values in accordance with the assigned simulation conditions and constraints. However, to accelerate the simulation and ensure its smooth performance, the initial atomic velocities should follow certain distributions. Which is usually based on the Maxwell-Boltzmann distribution that can guarantee a quick relaxation of the initial system. The Maxwell-Boltzmann distribution is a Gaussian distribution of a random variable x and is given by

$$P(x)_{\mu,\sigma} = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \tag{5}$$

with μ denoting the mean and σ^2 the variance. According to statistical mechanics, the temperature of a molecular system and the average atomic velocity are related by

$$\langle v_{i,\alpha}^2 \rangle = \frac{k_B T}{m_i}, \tag{6}$$

where $v_{i,\alpha}$ denotes a component of the velocity of the ith atom, with α representing the $x-$, $y-$, $z-$direction, k_B is the Boltzmann constant, T denotes the temperature, and m_i is the atomic mass. A random variable x following a Gaussian distribution with $\mu = 0$ and $\sigma = 1$ can be transformed into another random variable x' that follows a Gaussian distribution with μ' and σ' by the transformation $x' = \mu' + \sigma' x$. Hence, an initial atomic velocity distribution can be generated by setting $\sigma' = \sqrt{k_B T/m_i}$, $\mu' = 0$, and $x' = v_{i,\alpha}$. The resulting velocity distribution is then written as

$$P\left(v_{i,\alpha}\right) = \left(\frac{m_i}{2\pi k_B T}\right)^{1/2} \exp\left(-\frac{1}{2}\frac{m_i v_{i,\alpha}^2}{k_B T}\right). \tag{7}$$

After the initialization of the velocities, they are usually further adjusted to ensure that the net momentum of the atoms is zero, and the system does not drift during the simulation.

2.3 Temperature control

In molecular dynamics (MD) simulations, a thermostat is a tool used to control the temperature of the system being studied. The temperature of the system is one of the most important parameters in MD approach as it affects the behavior of the

molecules being studied. Temperature control is achieved by coupling the simulated system with a so-called "thermostat" that adds or removes energy from the system by either directly altering the motion of atoms or modifying their Newtonian equations of motion.

A variety of thermostat algorithms are currently available to achieve effective temperature control, and they can be categorized in several ways. For instance, the nature of these thermostat algorithms can be either stochastic or deterministic, depending on whether random processes/numbers are adopted to alter the atomic dynamics. They can also be classified as global or local, depending on whether they are applied to the entire system or each atom/molecule individually.

2.3.1 *Berendsen thermostat*

The Berendsen thermostat is also known as the weak coupling thermostat, which introduces a weak coupling of a simulated system with a "thermal bath". In this thermostat algorithm, deviations of the instantaneous system temperature $T(t)$ from the target temperature T are corrected by multiplying the atomic velocities at each timestep by a factor ζ, so that the system dynamics are steered toward one corresponding to T. The choice of the factor ζ should ensure that the rate of change of the instantaneous temperature $T(t)$ at time (t) is proportional to the difference between $T(t)$ and the target temperature T, which is given by

$$\frac{\mathrm{d}T(t)}{\mathrm{d}t} = \frac{1}{\tau_T}\left[T - T(t)\right]. \tag{8}$$

Here, τ_T is the coupling parameter that determines how strongly the system is coupled with the "thermal bath". The solution of Eq. 8 leads to the instantaneous temperature approaching the target temperature in an exponential manner, which is given by

$$T(t) = T - C\exp\left(-\frac{t}{\tau_T}\right), \tag{9}$$

where C is a constant. The finite-difference form of Eq. 8 is written as

$$\Delta T(t) = \frac{\Delta t}{\tau_T}\left[T - T(t)\right], \tag{10}$$

where Δt is the timestep. Eq. 10 implies that energy will be added to the system if $T(t) < T$ and subtracted from the system if $T(t) > T$.

Provided that the velocity rescaling is expressed by $v'_{i,\alpha} = \zeta v_{i,\alpha}$, the energy added to or subtracted from the system can be calculated as

$$\Delta E = \frac{1}{2}\sum_{i,\alpha}\left(m_i v'^2_{i,\alpha} - m_i v^2_{i,\alpha}\right) = \frac{1}{2}\left(\zeta^2 - 1\right)\sum_{i,\alpha} m_i v^2_{i,\alpha} = \left(\zeta^2 - 1\right)\frac{3Nk_BT(t)}{2}, \quad (11)$$

where $v_{i,\alpha}$ is the component of the velocity $\boldsymbol{v}_i$ in the α (x, y, or z) direction, N is the number of atoms in the system, and k_B is the Boltzmann constant $1.380\,649 \times 10^{-23}\,\mathrm{J\,K^{-1}}$. The combination of Eqs. 10 and 11 leads to the expression of ζ, which is given by

$$\zeta^2 = 1 + \frac{\Delta t}{\tau_T}\left(\frac{T}{T(t)} - 1\right). \quad (12)$$

It is apparent that the parameter ζ varies with time and depends on the timestep Δt and the coupling parameter τ_T. The incorporation of velocity rescaling into the Newtonian equations of motion of the atoms/ molecules leads to

$$m_i\frac{\mathrm{d}^2\boldsymbol{r}_i}{\mathrm{d}t^2} = -\frac{\partial U\left(\boldsymbol{R}^N\right)}{\partial \boldsymbol{r}_i} - \frac{1}{2\tau_T}\left(\frac{T}{T(t)} - 1\right)m_i\frac{\mathrm{d}\boldsymbol{r}_i}{\mathrm{d}t} \quad (13)$$

The Berendsen thermostat is generally not recommended for modern MD simulations. However, it has high efficiency in controlling the temperature and can be adopted to produce relatively accurate estimations of statistical properties of an MD simulation system, particularly when the system contains a large number of atoms.

2.3.2 *Nosé-Hoover thermostat*

The basic idea behind this thermostat is that it replaces the "thermal bath" in the previous thermostats with a fictitious degree of freedom s that is added to the energy function. This additional degree of freedom can be regarded as the frictional term in the Langevin equation, and it functions as a "thermal bath" to slow down or accelerate particles until the system temperature reaches the target value. The potential energy associated with s is expressed as $(3N + 1)\,k_BT\ln s$, with $3N + 1$ being the total degrees of freedom. The kinetic energy related to s is given by $Q(\mathrm{d}s/\mathrm{d}t)^2/2$, with Q denoting

the fictitious mass of s. The equations of motion of the atoms/molecules of an MD system coupled with the Nosé-Hoover thermostat can be written as

$$m_i \frac{\mathrm{d}^2 \boldsymbol{r}_i}{\mathrm{d}t^2} = -\frac{\partial U\left(\boldsymbol{R}^N\right)}{\partial \boldsymbol{r}_i} - \gamma_i m_i \frac{\mathrm{d}\boldsymbol{r}_i}{\mathrm{d}t}, \tag{14}$$

$$\frac{\mathrm{d}\gamma_i\left(t\right)}{\mathrm{d}t} = \frac{1}{Q}\left[\sum_{i=1} m_i \frac{\boldsymbol{v}_i \cdot \boldsymbol{v}_i}{2} - \frac{3N+1}{2} k_B T\right]. \tag{15}$$

Here, γ_i is defined to be a positive atomic friction coefficient that dissipates energy from the system, v_i denotes the velocity vector of the ith atom. The strength of coupling between the system and the thermostat, which depends on the temperature relaxation time τ_T (defined in Eq. 10) of the system, is determined by the fictitious mass Q.

2.4 Pressure control

2.4.1 *Berendsen barostat*

The most straightforward method of tuning the pressure of an MD system is to directly rescale the system volume by remapping the atomic coordinates. However, this approach does not result in a proper thermodynamic ensemble, and instead it introduces severe unphysical artifacts to the dynamic behaviors of the system. In addition, the simple volume rescaling method does not allow for pressure fluctuations, and there could be abrupt jumps in the total potential energy of the system.

The Berendsen barostat invokes the weak coupling method to tackle the problem encountered by the simple volume rescaling method, which is analogous to the algorithm adopted by the Berendsen thermostat. For the Berendsen barostat, deviations of the instantaneous system pressure $P\left(t\right)$ from the target pressure P are corrected by multiplying the atomic coordinates at each timestep by a factor of ζ to steer the system state toward the one corresponding to P. The choice of the factor ζ ensures that the rate of change of the instantaneous pressure $P\left(t\right)$ is proportional to the difference between $P\left(t\right)$ and P according to the following formula

$$\frac{\mathrm{d}P(t)}{\mathrm{d}t} = \frac{1}{\tau_P}\left[P - P(t)\right], \tag{16}$$

which is similar to Eq. 8. Here, τ_T is the coupling parameter that determines how closely the system is coupled with the barostat. Thus, the Newtonian equations of motion are modified to be

$$\frac{\mathrm{d}r_{i,\alpha}}{\mathrm{d}t} = v_{i,\alpha} - \frac{\beta\left[P - P(t)\right]}{3\tau_P} r_{i,\alpha}. \tag{17}$$

The above equation indicates that at each timestep, the atomic coordinates $r_{i,\alpha}$ and the box length l (assuming an isotropic system in a cubic box) are multiplied by a scaling coefficient ζ that is written as

$$\zeta = 1 - \frac{\beta\Delta}{3\tau_P}\left[P - P\left(t\right)\right]. \tag{18}$$

In principle, the choice of τ_P can be arbitrary. However, the value of τ_P determines the efficiency of the Berendsen barostat in pressure control, which is analogous to the choice of τ_T for the Berendsen thermostat.

Compared with the simple volume scaling method, coupling a system with the Berendsen barostat allows for pressure fluctuations, and the system pressure approaches the target value more realistically. Hence, the Berendsen barostat may be a good choice for performing an initial pressure equilibration of the system. However, it will lead to inaccuracies in the investigation of atomic dynamics under a constant pressure.

2.4.2 *Andersen barostat*

In the Andersen barostat, an additional degree of freedom is added to the Newtonian equations of motion of the atoms/molecules, which is similar to the strategy used for developing the Nosé-Hoover thermostat Specifically, in the Andersen barostat, the coordinate vector $\boldsymbol{r}_i$ is replaced by the scaled coordinate vector $\boldsymbol{\kappa}_i$ that is defined as $\boldsymbol{\kappa}_i = \boldsymbol{r}_i/V^{1/3}$. By adding the extra degree of freedom Q into the Lagrangian of the MD system, it becomes

$$L\left(\boldsymbol{\kappa}^N, \frac{\mathrm{d}\boldsymbol{\kappa}^N}{\mathrm{d}t}, Q, \frac{\mathrm{d}Q}{\mathrm{d}t}\right) = \frac{1}{2}Q^{2/3}\sum_{i=1}^{N} m_i \frac{\mathrm{d}\boldsymbol{\kappa}}{\mathrm{d}t}\cdot\frac{\mathrm{d}\boldsymbol{\kappa}}{\mathrm{d}t} - \sum_{i<j=1}^{N} U\left(Q^{1/3}\boldsymbol{\kappa}_{ij}\right) + \frac{1}{2}M\frac{\mathrm{d}Q^2}{\mathrm{d}t} - \eta Q, \tag{19}$$

where $\boldsymbol{\kappa}^N = \{\boldsymbol{\kappa}_1, \boldsymbol{\kappa}_2, \boldsymbol{\kappa}_3 \ldots \boldsymbol{\kappa}_N\}$, $\dot{\boldsymbol{\kappa}}^N = \{\dot{\boldsymbol{\kappa}}_1, \dot{\boldsymbol{\kappa}}_2, \dot{\boldsymbol{\kappa}}_3 \ldots\}\dot{\boldsymbol{\kappa}}_N$, and $\boldsymbol{\kappa}_{ij} = \boldsymbol{\kappa}_i - \boldsymbol{\kappa}_j$; $\dot{\boldsymbol{\kappa}}_i$ denotes the derivative of $\boldsymbol{\kappa}_i$ with respect to time; M is the fictitious mass having the extra degree of freedom Q; η is a constant. The first two terms on the right-hand side are simply the Lagrangian of the unscaled system. The third and fourth terms are the kinetic energy and potential energy associated with Q, respectively. Physically, this Lagrangian can be understood by considering a fluid constrained in a container with a variable volume. Hence, Q, whose value is given by the volume of the container, can be interpreted as the position of a piston that exerts a pressure on the fluid.

The expression for the Newtonian equations of motion of the atoms in the initial system, which reads as

$$\frac{\mathrm{d}\boldsymbol{r}_i}{\mathrm{d}t} = \frac{\boldsymbol{p}_i}{m_i} + \frac{1}{3}\boldsymbol{r}_i\frac{\mathrm{d}\ln V}{\mathrm{d}t},$$

$$\frac{\mathrm{d}\boldsymbol{p}_i}{\mathrm{d}t} = -\sum_{j(\neq i)=1}^{N} \hat{\boldsymbol{r}}_{ij}U'\left(\boldsymbol{r}_{ij}\right) - \frac{1}{3}\boldsymbol{p}_i\frac{\mathrm{d}\ln V}{\mathrm{d}t},$$

$$M\frac{\mathrm{d}^2V}{\mathrm{d}t^2} = -\eta + \left(\frac{2}{3}\sum_{i=1}^{N}\frac{\boldsymbol{p}_i\cdot\boldsymbol{p}_i}{2m_i} - \frac{1}{3}\sum_{i<j=1}^{N}\boldsymbol{r}_{ij}\cdot U'\left(\boldsymbol{r}_{ij}\right)\right)/V. \tag{20}$$

Here, $\hat{\boldsymbol{r}}_{ij}$ is a unit vector parallel to the vector $\boldsymbol{r}_{ij}$, and $U'\left(\boldsymbol{r}_{ij}\right)$ denotes the derivative of $U\left(\boldsymbol{r}_{ij}\right)$ with respect to $\boldsymbol{r}_{ij}$. On the right-hand side of Eq. 20, the first term η represents the external pressure exerted by the fictitious piston, and the second term represents the internal pressure due to the simulated particles, which fluctuates as the number of particles is not sufficiently large. An imbalance between these two pressures will result in an acceleration of the piston, whose position will fluctuate with the internal pressure. The timescale for the volume fluctuation is determined by the mass M of the piston.

2.5 Calculation fundamental

2.5.1 *Periodic boundary conditions*

During the initialization of an MD simulation, a simulation box with a finite size is defined first. Due to the surface effect, atoms or molecules located at the boundaries have fewer neighbors than their counterparts inside the box; thus, the

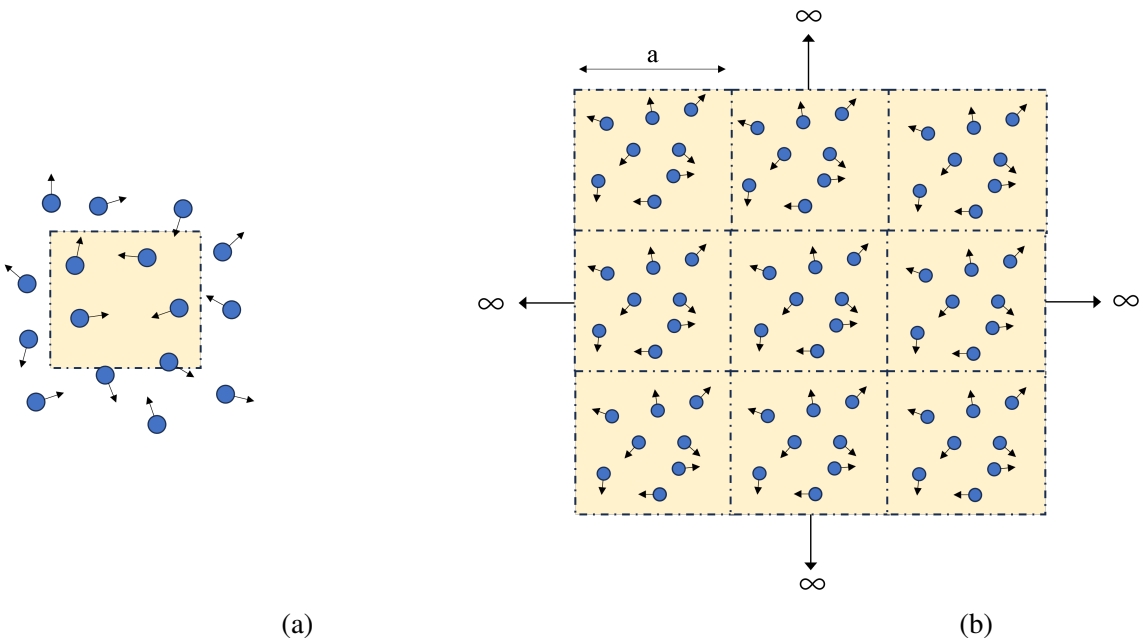

Figure 2.2. (a) Schematic of a realistic atomic system. Black arrows indicate the movement direction of the atoms. Atoms enclosed by the dotted lines are selected to form a simulation cell. (b) Schematic of an artificial atomic system constructed by duplicating the simulation cell in all three dimensions and extending them to infinity.

dynamic behaviors of the former deviate dramatically from those of the latter. These surface effects cause the simulation results to deviate from those obtained on the macroscopic scale, and the extent of such a deviation depends on the ratio of the surface atoms to the total atoms. In fact, no matter how large the simulation system is, its number of atoms will always be negligible compared with the number of atoms contained in a macroscopic chunk of material. As a result, the ratio of the surface atoms to the total atoms will always be much larger in MD simulation than in reality.

A solution to this surface problem is to apply Periodic Boundary Conditions (PBCs), which are some of the most frequently used boundary conditions in MD simulations. To apply the PBCs, a primary simulation cell is extracted from a realistic atomic system, as shown in Fig. 2.2(a). Atoms selected to form the primary cell are far from the surface and are free from boundary effects. By duplicating and packing the primary cell periodically in all three directions, the simulation space is completely filled with atoms to model macroscopic materials, as shown in Fig. 2.2(b). Hence, an atom in the primary cell with the position vector $\boldsymbol{r}_i$ represents an infinite set of image

atoms located at $\boldsymbol{r}_i + l\boldsymbol{a} + m\boldsymbol{b} + n\boldsymbol{c}$, where l, m, and n are integers, and $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{c}$ are the edge vectors of the primary simulation cell.

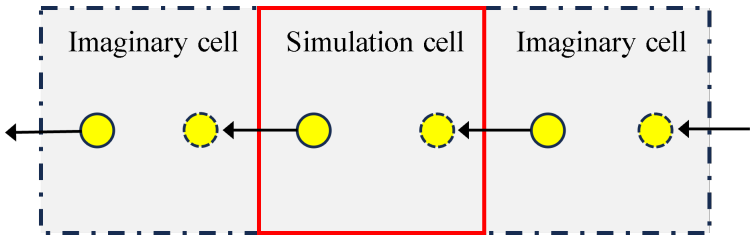

Figure 2.3. Schematic showing the synchronous movement of an atom in the primary simulation cell and that of its images. An atom (denoted by the circle with the solid line) moving out of the simulation cell is accompanied by one image atom (denoted by the circle with the dash line) moving inside.

The application of PBCs brings about two main advantages. First, each atom in the primary simulation cell interacts not only with other atoms inside the cell but also with their images in nearby virtual cells. Secondly, when an atom in the primary simulation cell moves out from one side, an image atom in the neighboring cell enters the simulation cell from the other side, as illustrated in Fig. 2.3. Therefore, the number of atoms inside the simulation cell will remain constant, which is a common requirement for most MD simulations.

A problem resulting from PBCs is that the number of interaction pairs increases to be arbitrarily large as there are infinite image atoms after imposing PBCs. Apparently, the calculation of all these interactions is impossible for practical MD simulations. Fortunately, most atomic interactions diminish rapidly with distance. Beyond a certain critical distance R_c, the interaction between two atoms is sufficiently weak, and it is thus usually truncated in practical simulations.

For the ith atom in the primary simulation cell, both atom j and several of its images are within the distance R_c of it and thus contribute to the net force acting on it. This situation increases the complexity of MD simulations and may lead to unrealistic results. To solve the above problem, the minimum image criterion is applied. This criterion states that the ith atom of the simulation cell interacts with either atom j or one of its images nearest to atom i. To demonstrate this, consider a rectangular simulation cell containing four atoms denoted by A (orange), B (green), C (purple), D (blue), as illustrated in Fig. 2.4. The primary simulation cell is surrounded by its 8 image cells. All of the neighbors of atom A within the critical distance Rc do not

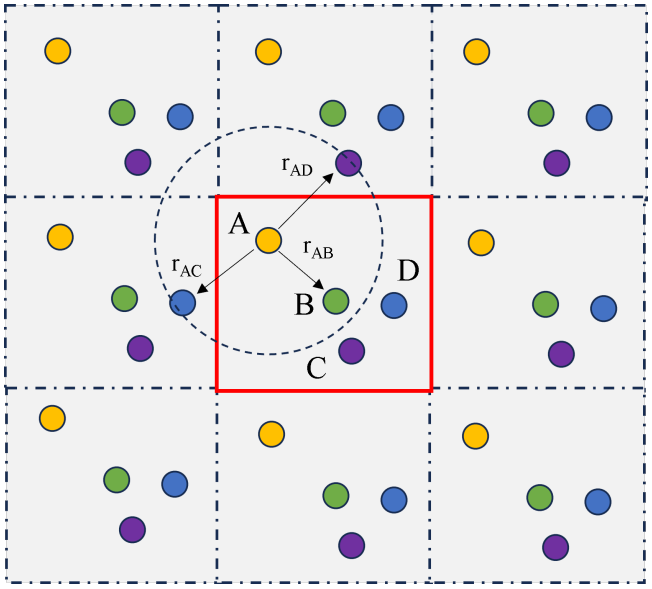

Figure 2.4. Schematic showing the closest neighbors of atom A in the primary simulation cell as determined by the minimum image criterion. The atoms located within a distance of R_c of atom A are listed as its closest neighbors and can interact with it.

remain inside the primary cell. Hence, when calculating the force on atom A, atom B and the images of atoms C and D in the neighboring "virtual" cells are taken into account.

2.5.2 *Force calculation*

Given a potential energy function, the force acting on an atom is calculated by obtaining the first-order derivative of $U\left(\boldsymbol{R}^N\right)$ as indicated by Eq. 2. For such short-range interactions, the total force acting on an atom can be represented by a direct summation over all the forces exerted by a certain number of neighboring atoms. In principle, for long-range interactions, the force calculation should consider the interactions of all the atom pairs. For a system containing N particles, there are $N(N-1)/2$ pair interactions, and hence $N(N-1)/2$ force calculation operations are required. Therefore, the computing task increases dramatically with the system size. As mentioned in Section 2.6.1, most atomic interactions diminish rapidly with increasing interatomic distances. Thus, the interaction between two atoms is usually truncated at a certain critical distance, which reduces the computational cost of the force calculation.

To safely truncate a potential function at a cutoff distance, the potential should decay with distance at a rate no lower than $U(r) \propto r^{-3}$. However, in ionic systems, the Columbic interaction decays slowly with distance according to the relationship $U(r) \propto r^{-1}$, and only approaches zero at very large distances. Hence, the truncation of the Columbic interaction at a small cutoff distance will inevitably lead to inaccuracies in the simulation results. However, a complete consideration of the long-range Columbic interactions will greatly increase the system size and computational cost, which significantly prohibits the simulation of many ionic systems. To circumvent this problem, different approaches have been developed to handle the long-range Columbic interaction, of which the Ewald summation method is the most widely used.

For the Ewald method, an atom in the simulation cell interacts not only with other atoms in the same cell but also with all the atoms in the image cells. Provided that the ions are subjected to PBCs, which are described by three repeat vectors $\boldsymbol{c}_1, \boldsymbol{c}_2, \boldsymbol{c}_3$ (forming a primary cell). This means that whenever there is an ion q_i at location $\boldsymbol{r}_i$, there are also ions with charge q_i at $\boldsymbol{r}_i + n_1\boldsymbol{c}_1 + n_2\boldsymbol{c}_2 + n_3\boldsymbol{c}_3$, where n_1, n_2, n_3 are arbitrary integers. To simplify the notation, an arbitrary repeat vector $n_1\boldsymbol{c}_1 + n_2\boldsymbol{c}_2 + n_3\boldsymbol{c}_3$ is written as $\boldsymbol{n}L$, where L represents the characteristic length of the initial cell. In this case, $L = |\boldsymbol{c}_1| = |\boldsymbol{c}_2| = |\boldsymbol{c}_3|$ and vectors $\boldsymbol{n}$ form a simple cubic lattice n_1, n_2, n_3. The total Coulomb interaction energy for these ions under PBCs has to include the interactions between periodic images

$$U\left(\boldsymbol{R}^N\right) = \frac{1}{2}\sum_{\boldsymbol{n}}\left[\sum_i\sum_{j\neq i}\frac{q_i q_j}{4\pi\varepsilon_0\left|\boldsymbol{r}_{ij}+\boldsymbol{n}L\right|}\right]. \tag{21}$$

The infinite sum in Eq. 21 not only converges very slowly but also is conditionally convergent, meaning that the result depends on the order of the summation. The Ewald method evaluates $U\left(\boldsymbol{R}^N\right)$ by transforming it into summations that converges not only rapidly but also absolutely. The final expression of potential in the Ewald

method is given by

$$
\begin{aligned}
U_{Ewald} &= U^{(S)} + U^{(L)} - U^{(Self)} \\
&= \frac{1}{2}\sum_{\boldsymbol{n}}\left[\sum_{i=1}^{N}\sum_{j(\neq i)=1}^{N}\frac{q_i q_j}{4\pi\varepsilon_0\left|\boldsymbol{r}_i-\boldsymbol{r}_j+\boldsymbol{n}L\right|}\right]\operatorname{erfc}\left(\frac{\left|\boldsymbol{r}_i-\boldsymbol{r}_j+\boldsymbol{n}L\right|}{\sqrt{2}\sigma}\right) \\
&\quad + \frac{1}{2V\varepsilon_0}\sum_{\boldsymbol{k}\neq 0}\frac{e^{-\sigma^2k^2/2}}{k^2}|S(\boldsymbol{k})|^2 - \frac{1}{4\pi\varepsilon_0}\frac{1}{\sqrt{2\pi}\sigma}\sum_{i=1}^{N}q_i^2. \qquad (22)
\end{aligned}
$$

The summation for $U^{(S)}$ is short-ranged in real space (truncated by the erfc function) and the summation for $U^{(L)}$ is short-ranged in reciprocal space (truncated by $e^{-\sigma^2k^2/2}$). As well as $U^{(Self)}$, a correction sums for the intermolecular interaction (chemical bonds). Where σ is the standard deviation of the Gaussian distribution, V denotes the volume of primary cell. And $k = |\boldsymbol{k}|$ with the fact that $\boldsymbol{k}$ is a reciprocal vector, $S(\boldsymbol{k})$ is the structure factor of the charge distribution.

3 Application of molecular dynamics simulation in irradiation processes

Irradiation of materials plays a vital role in determining their atomic structures and properties. Some of the most profound effects of irradiation on materials are observed in the cores of nuclear power reactors. Atoms comprising the structural components of such reactors are displaced consistently due to their collision with high-energy radiation particles, including electrons, protons, neutrons, and ions. A similar situation is encountered by the components of a spacecraft in outer space exposed to a plethora of high-energy particles. Exposure to radiation by high-energy particles may introduce significant microstructural damage to a variety of materials. Such damage is closely related to the formation, distribution, and interaction of point defects (vacancies and interstitials), point defect clusters (e.g., Frenkel pairs, vacancy clusters, and interstitial loops), dislocation segments and networks, inert gas bubbles, and voids.

However, the interaction between radiation particles and the atoms in a material is a complex process that occurs on temporal and spatial scales that are generally

inaccessible to contemporary experimental approaches. From an atomic point of view, the effects of irradiation on materials are rooted in the initial collision event in which an energetic projectile strikes a target atom. With an energy that exceeds the threshold displacement energy of the atom, such an initial collision displaces the target atom in a crystalline material from its equilibrium lattice site, leaving a vacancy site behind. The displaced atom eventually comes to rest at a location between lattice sites and becomes an interstitial atom. The vacancy-interstitial pairs (also known as the Frenkel pairs) and their interactions play a crucial role in determining the physical and mechanical properties of the irradiated crystalline material. Besides the atomic displacement, an energetic charged particle propagating within the material provides energy to electrons via the electron stopping power. To capture the detailed dynamics of the displaced atoms and the associated energy transfer, MD simulation has been increasingly adopted to investigate phenomena pertaining to irradiation processes.

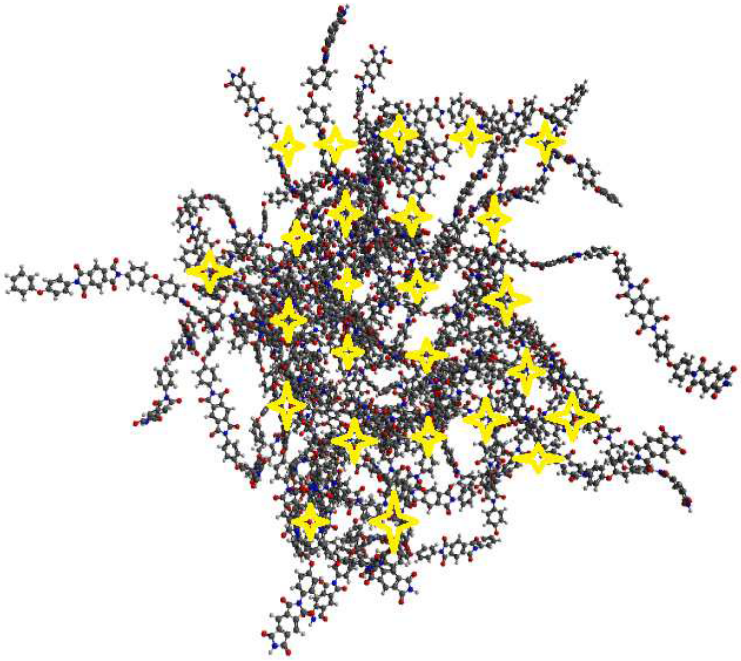

Figure 3.1. Schematic of the MD simulation model for the PI molecular system under electron beam irradiation. Each yellow star represents the location of an electron beam, which is modeled by a string of massless pseudoparticles.

The first example is modeling the damage to a polyimide (PI) system caused by electron beam irradiation. This example is selected as it models the electron beam explicitly and presents a novel approach for modeling the interaction between an

electron beam and a target material. PI is a polymer that is extensively used in the spacecraft industry. Unfortunately, the properties of PI change significantly when it is exposed to high-energy electrons and protons, and heavy ions in outer space, which inevitably induces structural deformations such as chain scission and cross-linking (Fig. 3.1).

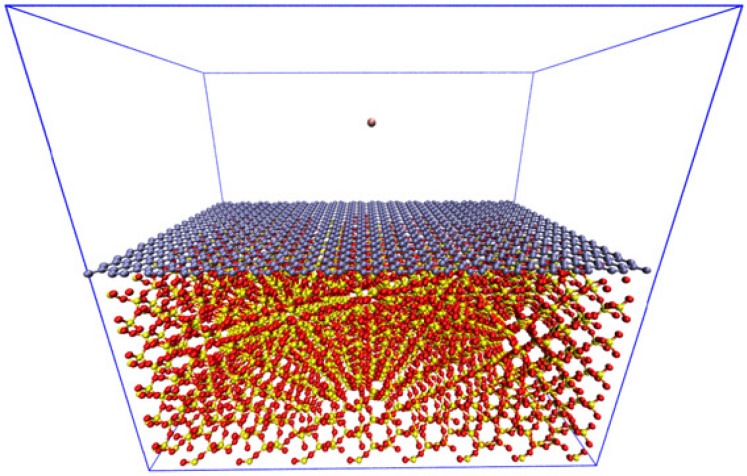

Figure 3.2. Illustration of the simulation setup for the irradiation of a graphene nanosheet supported by a SiO_2 substrate. The indigo, red, and yellow spheres represent the C, O, and Si atoms, respectively. The incident ion is denoted by the pink sphere located above the graphene nanosheet.

The second example (Fig. 3.2) pertains to creating defects in supported graphene nanosheets by exposing them to energetic heavy ion irradiation. Graphene has a broad range of applications due to its unique mechanical and physical properties. Engineering graphene by ion beam irradiation is considered rather promising because the direct doping of foreign atoms can be conveniently realized, and defects can be created in a controllable manner for further functionalization. The substrate that supports the graphene nanosheets has a significant impact on their defect formation process. To understand the underlying defect formation mechanism and the effects of the substrate, experimenters have performed MD simulations on the ion irradiation process of graphene nanosheets supported by a SiO_2 substrate.

Acknowledgement

I would like to express my gratitude to the INTEREST team for creating such an opportunity for me to join the most intriguing scientific project, which brings me in-

sight and advantageous skills for my development as a young researcher. Additionally, the success of this project could not have been possible without the supportive and devoted guidance of Prof. Kholmirzo Kholmurodo during his supervision.

References

[1] Schneider, R., Sharma, A.R., Rai, A. (2008). Introduction to Molecular Dynamics. In: Fehske, H., Schneider, R., Weiße, A. (eds) Computational Many-Particle Physics. Lecture Notes in Physics, vol 739. Springer, Berlin, Heidelberg.

[2] Vollmayr-Lee, K. (2020). Introduction to molecular dynamics simulations. American Journal of Physics, 88(5), 401-422.

[3] Kumar, G., Mishra, R.R., Verma, A. (2022). Introduction to Molecular Dynamics Simulations. In: Verma, A., Mavinkere Rangappa, S., Ogata, S., Siengchin, S. (eds) Forcefields for Atomistic-Scale Simulations: Materials and Applications. Lecture Notes in Applied and Computational Mechanics, vol 99. Springer, Singapore.

[4] Pal, S., & Ray, B.C. (2020). Molecular Dynamics Simulation of Nanostructured Materials: An Understanding of Mechanical Behavior (1st ed.). CRC Press.

[5] Kholmirzo KHOLMURODOV (Editor), Computational Materials and Biological Sciences, Editor: Kholmirzo T. Kholmurodov (Leading Scientist, Frank Laboratory of Neutron Physics, Joint Institute of Nuclear Research, Dubna, Moscow Region, Russia), Nova Science Publishers (N.Y.), ISBN: 978-1-63482-541-2, (2015).

[6] Kholmirzo KHOLMURODOV (Editor), Computer Design for New Drugs and Materials: Molecular Dynamics of Nanoscale Phenomena, Nova Science Publishers (N.Y.), ISBN: 978-1-53612-082-0, (2017).

[7] Zhao, S., Xue, J., Wang, Y., & Yan, S. (2012). Effect of SiO2 substrate on the irradiation-assisted manipulation of supported graphene: a molecular dynamics study. Nanotechnology, 23(28), 285703.

[8] Frenkel, D., & Smit, B. (2023). Molecular Dynamics simulations. Understanding Molecular Simulation, 97–124.

[9] Rahnamoun, A., Engelhart, D. P., Humagain, S., Koerner, H., Plis, E.,

Kennedy, W. J., Cooper, R., Greenbaum, S. G., Hoffmann, R., & van Duin, A. C. T. (2019). Chemical dynamics characteristics of Kapton polyimide damaged by electron beam irradiation. Polymer, 176, 135–145.

[10] Tomobe, K., Yamamoto, E., Kholmurodov, K., & Yasuoka, K. (2017). Water permeation through the internal water pathway in activated GPCR rhodopsin. PloS one, 12(5), e0176876.

[11] Ozboyaci, M., Kokh, D. B., Corni, S., & Wade, R. C. (2016). Modeling and simulation of protein-surface interactions: achievements and challenges. Quarterly Reviews of Biophysics, 49.