Sustaining Orientation of Ubiquitin for Single Particle Imaging Using Electric Fields

Adi Bijedic

Degree Project C in Physics, 15 ECTS
Department of Physics and Astronomy, Uppsala University

Supervisors: Carl Caleman, Anna Sinelnikova & Emiliano De Santis
Subject reader: Oscar Grånäs
Examiner: Matthias Weiszflog

Juli 12, 2021
Abstract

Single-particle imaging, or SPI, is a method used to obtain the three-dimensional structure of particles. Repeatedly aiming X-rays at samples of a particle produces diffraction patterns, which are combined to a best-fit three-dimensional model of the particle. SPI of proteins can be improved by orienting the protein before imaging. Protein dipole orientation makes use of a protein’s dipole moment and an external electric field to generate torque, which can orient the protein. A protein subject to an electric field may however result in damage of the protein’s geometrical structure, or insufficient protein orientation, depending on the magnitude of the electric field.

Sufficient protein orientation without substantial protein damage is possible in an interval of electric field strengths. The results in this report reveal that the method of SPI can be further improved. With a protein being fully oriented in an electric field, it is possible to reduce the electric field strength and yet sustain sufficient orientation, with some constraints. Longer times for imaging and less structural damage to the protein are hence possible. This study implements Molecular Dynamics (MD) and the most extensively used open-source MD software, GROMACS, with ubiquitin as a sample protein.

Sammanfattning


## Contents

1 Introduction ........................................... 3

2 Background ........................................... 5
   2.1 Dipole Orientation ............................... 5
   2.2 Sample Protein ................................. 5
   2.3 GROMACS ........................................ 6
   2.4 Molecular Dynamics Simulations ................. 6
      2.4.1 Bonded Interactions ....................... 7
      2.4.2 Non-bonded Interactions ................... 9
      2.4.3 External Electric Field Potential .......... 10
      2.4.4 Total Potential ............................ 11
   2.5 Computation ..................................... 11
   2.6 Kinetic Energy and Temperature .................. 11
   2.7 Root Mean Square Deviation, RMSD ................ 12

3 Method ............................................... 12
   3.1 Preparation and Pre-orientation .................. 12
   3.2 Electric Field Implementation .................... 13
   3.3 Dipole Moment Ratio and Protein Orientation .... 14

4 Results and Discussion ............................... 15
   4.1 Orientation ...................................... 15
      4.1.1 Dipole Moment Components ................. 15
      4.1.2 Degree of Orientation ...................... 17
   4.2 Energy and Damage ................................ 19
      4.2.1 Kinetic Energy ............................ 19
      4.2.2 Temperature ............................... 21
      4.2.3 Structural Damage ......................... 22
   4.3 Further Research .................................. 23

5 Conclusions ......................................... 23
1 Introduction

An understanding of protein structure and behavior is of high importance in the expansion of atomic biochemistry. Knowledge of proteins’ geometrical three-dimensional structures reveals specific protein properties and behaviors, which helps widen possibilities in medicine and drug discovery. Computational methods greatly facilitate this development [1].

In recent decades, a developed understanding of atomic structures in different proteins has been made possible mainly through the method of crystallization in a solvent, followed by X-ray diffraction and/or nuclear magnetic resonance (NMR) spectroscopy. There are also other methods for determining a protein’s structure, such as cryo-electron microscopy, but crystallography is by far the most widely used method. While these methods have immensely expanded our knowledge for a vast number of proteins [2], they have their limitations.

A preeminent problem is that proteins do not necessarily remain physiologically unaltered during the process of crystallization, thus the obtained image of a protein may not reveal its true, unaltered atomic structure. Moreover, the crystallization of a protein in a solvent can be a complex process. The probability of a successful outcome varies greatly depending on the protein. Also, crystallization of a protein is generally time-consuming and mainly consists of trial and error approaches [3].

Single-particle imaging, or SPI, is a method where a single molecule in gas phase can be imaged through the use of X-ray diffraction, without the need for prior crystallization [4]. SPI could thus be used to determine the atomic structures of non-crystallized proteins with higher accuracy. There are however some technical challenges with SPI, such as radiation damage and sample preparation, and significant efforts have been put into making SPI successful [5]. Recently, the Mass Spectrometry for Single-Particle Imaging of Dipole Oriented Protein Complexes (MS SPIDOC) project launched to further advance the development of the technique of SPI [6].

A crystallized structure of multiple proteins can withstand the intensity of X-ray beams and produce a diffraction pattern. A single protein, however, is highly fragile and is easily damaged by an X-ray beam. It has been found that highly intense and short enough X-ray pulses (in the range of femtoseconds) provides a diffraction pattern of the exposed protein before the sample is destroyed by the high intensity of the beam [4], a phenomenon commonly referred to as ”diffraction before destruction”. SPI can thus be used to successfully image proteins and attain their unaltered atomic structures. A single diffraction image of a protein is however of low significance since its orientation at the moment of diffraction is unknown. There is also the presence random noise in diffraction images. Additionally, the protein is destroyed almost instantly as the X-ray beam passes through it, and can thus only be used to provide a single image.
To obtain a complete understanding of a protein through SPI, and minimize random noise, one needs to image a large number of different samples of the same protein. Algorithms then combine the images, producing a three-dimensional understanding of the protein’s atomic structure.

One of the main objectives of the MS SPIDOC project is to orient proteins prior to SPI for higher preciseness and efficacy. Simulations by Marklund et al. (2017) have shown that it is possible to control the orientation of proteins holding a non-zero dipole moment (which applies to a vast number of proteins) with the use of external electric fields. SPI of proteins using this method can give promising outcomes, as multiple imaging samples with known orientations can give a more accurate three-dimensional image with higher efficacy. However, a single protein subjected to an external electric field is easily damaged and can lose its original structure if the external electric field is too strong. On the other hand, an electric field being too weak might not manage to orient the protein. Luckily, previous studies have found that there exists an interval of electric field strengths in which proteins can be successfully oriented without significant loss of structure [7]. It has also been found that a time-dependent, increasing electric field yields an even better outcome in orienting ubiquitin (a relatively non-complex protein) while its structure remains intact [8]. Furthermore, recent studies show that in orienting a protein with the use of an external electric field, its structure at the time of full orientation is preserved [9].

This report investigates the possibilities of sustaining the orientation of an already oriented protein using lower strength electric fields in the range of 0.06-0.5 V/nm. If possible, one could apply a strong electric field to orient a protein successfully without causing damage, and then lower the strength of the electric field yet keep orientation. This would be useful for sustaining protein orientation for a prolonged period of time for SPI without unfolding or changing its atomic structure.

In this study, Molecular Dynamics (MD) simulations will be used to simulate the behavior of a model protein in the process described above. By investigating field strength breakpoints for orientation and their respective effects on the properties of the protein, the results of this experimental report can hopefully serve as a contribution towards the realization of an apparatus for improved imaging of proteins using SPI.

In the following section, the main components in this study are presented. The foundations of MD are outlined, along with important protein and electric field properties. Section 3 depicts the preparations needed and the methods behind specific implementations, relevant for the results of this project. The results and discussions are presented in Section 4 along with recommendations for future projects. A summary of the entire project is found in Section 5.
2 Background

2.1 Dipole Orientation

The existence of a dipole moment is the result of an uneven charge distribution within a body, or a protein as in the case of this study. The dipole moment in proteins can be made use of to orient the protein. Two points charged with \( +q \) and \( -q \) displaced by a distance \( \vec{d} \) give rise to a dipole moment

\[
\vec{p} = q\vec{d},
\]

with the convention that \( \vec{d} \), and thus \( \vec{p} \) are directed from the negative to the positive charge as presented in Figure 1.

![Figure 1: Two oppositely charged particles giving rise to dipole moment.](image)

The total dipole moment for a system is hence

\[
\vec{p} = \sum_i q_i \vec{d}_i,
\]

where \( q_i \) and \( d_i \) are the charges and displacement vectors respectively for each atom \( i \) in the system. When subject to an electric field of magnitude \( \vec{E} \), each atom \( i \) will be acted upon with a force

\[
\vec{F}_i = q_i \vec{E}.
\]

With forces being present, and a lever connecting the particles, torque is generated. The torque \( \vec{\tau} \) is dependent on the dipole moment and the electrical field strength as

\[
\vec{\tau} = \vec{p} \times \vec{E}.
\]

The sample proteins in this study are initially already oriented. When subject to an electric field of some magnitude, there should be minimal to no torque acting on the protein. As the protein is transitioned into a lower strength electric field, the protein will tend to revert to oscillating randomly, but as soon as any deviation occurs, torque is induced. Depending on the lower electric field strength, it may be sufficient to sustain the protein’s orientation.

2.2 Sample Protein

The model protein of choice for this study is ubiquitin (Figure 2). Ubiquitin holds a non-zero dipole moment and contains a relatively small number of roughly
1300 atoms, making simulations with ubiquitin efficient yet fruitful for analysis. Furthermore, it is a well-studied protein that has been used in dipole orientation simulations in previous studies [7][8], which this report is based on.

Figure 2: Cartoon representation of a ubiquitin sample.

2.3 GROMACS

The software used for all simulations in this study is GROMACS, version 4.6.5. GROMACS is open-source and free, designed for simulations on systems containing up to millions of particles [10]. The simulations in GROMACS utilize Newtonian equations of motion, which is commonly known as Molecular Dynamics, or MD.

2.4 Molecular Dynamics Simulations

Molecular Dynamics (MD) simulations make use of classical mechanics to simulate interactions between atoms in a system. Despite being built on classical interactions, some quantum phenomena are parametrized with the help of classical force fields. With this, one can successfully and accurately predict real-life behavior and dynamics of proteins, as well as systems of much larger sizes, given that the atoms in the system are at reasonable temperatures. There are exceptions where MD simulations fail to do this however [11].

In MD, a system comprised of N spherical atoms with electrons in their ground state is assumed. The force acting on an atom is defined by Newton’s second
law of motion as
\[
\vec{F}_i = m_i \frac{\partial^2 \vec{r}_i}{\partial t^2}, \quad i = 1, 2...N, \quad (2.5)
\]
where \(m_i\) is the mass for each atom \(i\) and \(\vec{r}_i\) is the displacement over some time \(t\) for each atom \(i\). The forces between atoms in the entire system are in turn given by
\[
F_{\text{tot}} = \sum_{j=1}^{N} F_{i,j}, \quad i, j = 1, 2...N. \quad (2.6)
\]
All atoms interact with each other in multiple ways. Interacting forces between atoms can be defined as the negative gradient of the potential \(V(\vec{r}_1, \vec{r}_2...\vec{r}_N)\), being crucial to define in order to successfully represent the evolution the system.
\[
\vec{F}_i = -\frac{\partial V}{\partial \vec{r}_i}. \quad (2.7)
\]
The equations are solved simultaneously and continuously [11]. To optimize simulations and enhance efficiency, different types of interactions between atoms in the system are divided into two categories: bonded interactions and non-bonded interactions. The chemical properties and structure between atoms determines the type of interaction [11].

2.4.1 Bonded Interactions

Since no atom is free in a protein, each atom constituting the protein is covalently bound to at least one neighboring atom, with which it is interacting. Bonded interactions between atoms, or bodies, include bond stretching (two-body), bond angle (three-body), and dihedral angle (four-body).

Bond Stretching Potential
The bond stretching potential \(V_b\) in a two-body interaction can be approximated to a harmonic spring potential, where two covalently bound atoms \(i\) and \(j\) are vibrating along their bound axis as in Figure 3. This potential, derived from Hooke’s law, is given by
\[
V_b(r_{ij}) = \frac{1}{2} k_{b_{ij}} (r_{ij} - r_{0_{ij}})^2, \quad (2.8)
\]
where \(k_{b_{ij}}\) are the pairing strengths, \(r_{ij}\) is the distance and \(r_{0_{ij}}\) is the separation at equilibrium between atoms \(i\) and \(j\).

![Figure 3: Bond stretching between two covalently bound atoms.](image)
Bond Angle Potential

The bond angle potential \( V_\theta \) in a three-body interaction between atoms \( i, j, \) and \( k \) is in similarity to a bond stretching potential also approximated to a harmonic spring potential. The potential between two vibrating springs separated by an angle \( \theta_{ijk} \) is given by

\[
V_\theta(\theta_{ijk}) = \frac{1}{2} k_{\theta_{ijk}} (\theta_{ijk} - \theta_{0_{ijk}})^2,
\]

where \( k_{\theta_{ijk}} \) is the triplet strength and \( \theta_{0_{ijk}} \) is the equilibrium angle between the three atoms. This is illustrated in Figure 4.

\[
\theta_{ijk} \quad \theta_{0_{ijk}}
\]

\[i\quad j\quad k\]

Figure 4: Bond angle in an interaction between three atoms.

Dihedral Angle Potential

In a four-body interaction between atoms \( i, j, k, \) and \( l \), the rotational potential along the dihedral angle (Figure 5) has to be additionally considered when determining the total potential. It is given by

\[
V_d(\phi_{ijkl}) = k_\phi (1 + \cos(n\phi - \phi_s)),
\]

where \( k_\phi \) is the bonded quadruple-strength between the atoms, and \( \phi \) is the angle between the \( ijk \)- and \( jkl \)-planes.
2.4.2 Non-bonded Interactions

Atoms separated by three or more covalent bonds are considered non-bonded to each other and are always pair-additive in GROMACS [11]. Non-bonded interactions between atoms are divided into two categories: Coulomb interaction and Lennard-Jones interaction.

Coulomb Potential

Charged atoms interact with other charged atoms through electrostatic forces, giving rise to the Coulomb potential. For bonded interactions, this is already partially accounted for, since electrostatic forces between bonded atoms directly affect the coupling strength. For a non-bonded interaction, the Coulomb interaction between two charged particles $i$ and $j$ is given by

$$V_C(r_{ij}) = \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon r_{ij}}.$$  (2.11)

where $q_i$ and $q_j$ are the respective charges for atoms $i$ and $j$. $\varepsilon_0$ is the permittivity in vacuum, $r_{ij}$ is the distance between the atoms, and $\varepsilon_r$ is a constant accounting for the reaction field produced. However, in the simulation setup of this project, the reaction field is not accounted for [11], and $\varepsilon_r = 1$ is hence set.

Lennard-Jones Potential

The second type of non-bonded potential takes into account both a repulsive term and a dispersion term. The Lennard-Jones potential between bonded atoms is also partially already accounted for. Using the Pauli exclusion principle for repulsive interactions, and the van der Waals force for dispersive interactions, the Lennard-Jones potential between two atoms $i$ and $j$ is

$$V_{LJ}(r_{ij}) = \frac{C_{ij}^{(12)}}{r_{ij}^{12}} - \frac{C_{ij}^{(6)}}{r_{ij}^6},$$  (2.12)
where $C^{(12)}_{ij}$ and $C^{(6)}_{ij}$ are repulsion and attraction terms respectively, being dependent on pairs of atom types. The distance between the atoms is $r_{ij}$. The simulations follow a matrix with set values for the $C_{ij}$-terms for specific atom pairs [11].

![Figure 6: Coulomb potential (blue) and Lennard-Jones potential (orange) as functions of distance in non-bonded interactions between atoms. The potentials are not depicted to scale.](image)

As the Coulomb potential is dependent on the quadratic inverse of the distance $r_{ij}$, its magnitude is decreasing at a considerably slower rate compared to the rapidly decreasing Lennard-Jones potential. At some distance between the atoms, the Lennard-Jones potential becomes insignificantly small, as seen in Figure 6. To keep the simulations optimized, GROMACS implements a cut-off scheme, where the non-bonded Lennard-Jones potential is neglected at some cut-off distance between the atoms [11]. The cut-off scheme follows a default buffer list for atom pairs and a non-static buffer list for cluster pairs [11]. There also exists a cut-off scheme for Coulomb interactions. However, since the protein is in a vacuum in the simulations of this project, this is not applied.

### 2.4.3 External Electric Field Potential

If the system has an uneven charge distribution, there will be the existence of a total dipole moment (Equation 2.1 and Figure 1). An applied external electric field $\vec{E}$ over this system will exert forces $\vec{F}_i$ on each atom $i$ in the system, defined in Equation 2.3. With the uneven charge distribution in the system, and the forces exerted on each atom, torque will be generated (Equation 2.4) and the
system will gain potential energy given by
\[ V_{\text{ef}}(\theta) = -\vec{p} \vec{E} \cos \theta, \quad (2.13) \]
where \( \theta \) is the angle between the total dipole moment \( \vec{p} \) and the electric field \( \vec{E} \).

### 2.4.4 Total Potential

There are several additional potential terms as well as rather complex and modified potentials, along with restraints implemented in MD simulations. However, only the most relevant interactions have been covered Section 2.

The total potential for the entire system, in this case, is hence the sum of all the above-listed potential terms (Equation 2.8-2.13), for the sum of all atoms, i.e.

\[
V_{\text{tot}} = \sum_{i,j} V_b(r_{ij}) + \sum_{i,j,k} V_a(\theta_{ijk}) + \sum_{i,j,k,l} V_d(\phi_{ijkl}) + \sum_{i,j} V_C(r_{ij}) + \sum_{i,j} V_{\text{LJ}}(r_{ij}) + V_{\text{ef}}(\theta). \quad (2.14)
\]

### 2.5 Computation

To compute the trajectories for each atom with time, an integrator has to be used. There are a number of different integrator algorithms in MD. GROMACS can be set to use the so called leap-frog integrator, which is also what the simulations conducted in this study will implement. While there are other integrators with higher precision, the leap-frog integrator is efficient yet precise enough for the scope of this project. For each step of some length \( \Delta t \), the leap-frog algorithm uses positions \( \vec{r} \) at time \( t \) while using velocities \( \vec{v} \) at time \( t - \frac{1}{2}\Delta t \) as

\[
\vec{v}(t + \frac{1}{2}\Delta t) = \vec{v}(t - \frac{1}{2}\Delta t) + \frac{\vec{F}(t)}{m}\Delta t, \quad (2.15)
\]

\[
\vec{r}(t + \Delta t) = \vec{r}(t) + \vec{v}(t + \frac{1}{2}\Delta t)\Delta t. \quad (2.16)
\]

With the sum of potentials in the defined Equations 2.8-2.13, the resulting force \( \vec{F} \) acting on an atom in the system is calculated using the relationship between force and potential in Equation 2.7 [11].

### 2.6 Kinetic Energy and Temperature

The kinetic energy of the system defines its temperature
\[
T = \frac{2E_k}{kN_d}, \quad (2.17)
\]
where \(k\) is the Boltzmann constant, and \(N_{df}\) is the number of degrees of freedom. \(E_k\) is the kinetic energy, defined using classical mechanics as

\[
E_k = \frac{1}{2} \sum_{i=1}^{N} m_i v_i^2, \tag{2.18}
\]

where \(m_i\) is the mass, and \(v_i\) is the velocity for each atom \(i\) in a system constituted by \(N\) atoms.

### 2.7 Root Mean Square Deviation, RMSD

The root mean square deviation (RMSD) is a measurement of positional disturbance for specified atom types in a system, defined by

\[
RMSD(t, t_0) = \left( \frac{1}{M} \sum_{i=1}^{N} m_i ||\vec{r}_i(t) - \vec{r}_i(t_0)||^2 \right)^{\frac{1}{2}} \tag{2.19}
\]

where \(M\) is the total mass of all atoms, \(m_i\) is the mass for each atom \(i\), \(\vec{r}_i(t)\) is the position of atom \(i\) at time \(t\), and \(\vec{r}_i(t_0)\) is the atom’s initial position. For the purposes of this study, the RMSD is calculated with the protein structure at \(t=0\) as reference.

The RMSD is used to monitor how well the protein preserves its original structure. However, as the main interest for SPI is the backbone structure of the protein, the RMSD is calculated only for alpha-carbon atoms, which are the first carbon atoms directly bound to a functional group. These play a crucial role in the overall structure (secondary and tertiary) of the protein. An alpha-carbon RMSD value closer to 0 indicates smaller deviations and better structural preservation.

### 3 Method

#### 3.1 Preparation and Pre-orientation

Prior to all main simulations in this project, all protein samples have been subject to a ramp up of electric field strength for a duration of 5 ns up to different constant electric field strengths, applied for an additional 5 ns. These structures have been provided by Sinelnikova et al. [9].

The initial values of \(E_i = \{0.5, 0.8, 1.0\} \text{ V/nm}\) have been chosen based on previous findings that electric field strengths in this range have promising outcomes for successfully orienting simple proteins without significantly impacting their structures [8][9][12]. It is of additional interest to investigate to what extent the dipole orientation is sustained with insufficient pre-orientation, whereby \(E_i = 0.2 \text{ V/nm}\) is also investigated.

An example of dipole orientation of ubiquitin is presented in Figure 7, where an
electric field of strength $E_i = 0.8 \text{ V/nm}$ has been applied to a ubiquitin sample for 1 ns.

For statistical reasons, 10 different initial configurations of ubiquitin will undergo every simulation setup. The obtained simulation data of the protein and its properties are averaged over all 10 samples. This generates a more realistic and randomized distribution of the protein’s energy and inner atom velocities, providing more reliable results.

### 3.2 Electric Field Implementation

Following the pre-orientation, the final electric field strengths are instantly initiated in the main simulations. The electric field is applied along the x-axis of the system in the simulations. The ramping down time is set to 0 for all simulations, mimicking an instant ramping down from some initial electric field strength $E_i$ to some final electric field strength $E_f$.

The provided protein structures used in this report have been subject to electric fields ramping up to constant electric fields of strength $E_i = \{0.2, 0.5, 0.8, 1.0\} \text{ V/nm}$
in the pre-orientation phase. Constant electric fields are implemented in the main simulations of strengths $E_f$. The final electric field strengths implemented are $E_f = \{0.06, 0.08, 0.10, 0.12, 0.14, 0.16, 0.18, 0.20, 0.22, 0.24, 0.26, 0.32, 0.38, 0.44, 0.50\}$ V/nm and are kept fixed until completion of the simulations.

The electric field strength experienced by all 10 ubiquitin samples will follow all of the paths once from left to right, displayed in Figure 8, where all combinations of $E_i$ and $E_f$ are illustrated.

![Figure 8: All simulation setups of protein samples transitioning from the pre-orientation phase (left half), with some initial electric strength $E_i$, into the main simulation (right half) with some final electric field strength $E_f$.](image)

A maximum simulation time of 1 ns is used for all main simulations. This is a relatively short period of time, which is plausible for this study however, as any potential loss of orientation should be instantaneous.

### 3.3 Dipole Moment Ratio and Protein Orientation

As the electric field is applied along the x-axis of the system, the ratio between the protein’s dipole moment along the x-axis of the system ($M_x$) and the total dipole moment of the protein ($M_{tot}$) are extracted for each step in the simulations. This ratio represents the cosine value of the angle of deviation between the desired orientation of the protein (parallel to the x-axis) and its actual orientation. A
consistent ratio of 1 hence indicates full orientation while a ratio diverging from 1 indicates increasing loss of orientation. For better conceptualization, the degree of orientation is defined as
\[ \mathcal{D} := 1 - \cos(\theta), \] (3.1)
where \( \cos(\theta) \) is the ratio between \( M_x \) and \( M_{tot} \). Full orientation of the protein is hence indicated by \( \mathcal{D} = 0 \), whereas \( \mathcal{D} = 1 \) and \( \mathcal{D} = 2 \) indicate perpendicular and anti-parallel orientations respectively. As 10 different samples are used for every setup, the average degree of orientation is calculated among all simulations in each setup.

When the protein is subject to an electric field, its charge distribution will be polarized and the total dipole moment of the protein might not remain unchanged. GROMACS does not take this into consideration, as the charge distribution in classical MD is dependent only on the charges and positions of atoms in the system. However, the changes in the total dipole moment of the protein due to polarization are probably insignificant in the case of this study, considering the magnitude of electric field strengths implemented [13].

4 Results and Discussion

A transition of the oriented protein from some electric field strength to a higher one \( (E_f > E_i) \), holds virtually no significance for the results since the protein’s orientation after a decrease of electric field strength is to be investigated. This is the case for simulations in which \( E_i = 0.2 \) V/nm and \( E_f > 0.2 \) V/nm, and these numbers are hence not presented.

In reality, the total dipole moment of the protein is expected to decrease slightly with a decrement of electric field strength. This is not accounted for in the simulations. This effect can however be ignored, as discussed in Section 3.3.

4.1 Orientation

Stronger initial and final electric field strengths, \( E_i \) and \( E_f \), are expected to result in better protein orientation. The orientation can be analyzed at the level of each sample as dipole moment component development or as degree of orientation, as well as an average of all sample simulations for each setup.

4.1.1 Dipole Moment Components

The dipole moment components are obtained for all individual samples and simulation setups. In Figure 9, four simulations with different setups with a specific ubiquitin sample are displayed.
Figure 9: Dipole moment components for a protein sample in four different simulation setups for every step in the simulation. Initial and final electric field strengths are indicated in the plot captions. Alignment of the dipole moment along the x-axis (red) and total dipole moment (blue) of the protein indicates full orientation. Dipole moment along the y-axis and z-axis are represented by the yellow and green curves respectively.
4.1.2 Degree of Orientation

Figure 10: Degree of orientation of one protein sample for every step in four different simulations (blue) and mean over 0.5 ps (red) throughout the simulations (1 ns). Initial and final electric field strengths are indicated in the plot captions. A degree of orientation closer to 0 indicates better orientation.

As discerned from Figure 10, the initial orientation of the protein is lost almost instantaneously as the protein is transitioned from a higher intensity electric field (prior to simulation) to a lower one (course of the simulation). The preservation of the degree of orientation is notably improved with increments of both initial and final electric field strengths (compare Figures 10a and 10c) and seems to remain stable within some range. The amplitude of this range is seemingly decreasing with greater intensity of electric field strengths.
As the preservation of the protein’s orientation is to be investigated, the average degree of orientation over only the second half of the simulation is represented in Figure 11.

We define poor degree of orientation when $\varphi \geq 0.05$. $\varphi = 0.05$ corresponds to approximately 18 degrees of deviation between the protein’s average orientation and the desired orientation. The results shown in Figure 11 reveal that $E_f \leq 0.16$ V/nm results in poor preservation of the orientation, independently of the value of $E_i$. For $0.18$ V/nm $\leq E_f \leq 0.26$ V/nm, a higher initial electric field strength provides better final orientation, although $E_i = 0.5$ V/nm and $E_i = 0.8$ V/nm only provide marginally differing results in this range. For $E_f \geq 0.32$ V/nm, the initial electric field strength is of low significance, and the mean degree of orientation proves to be well preserved in all setups in this range.

Generally, higher initial and final electric field strengths result in better preservation of protein orientation. However, the degree of orientation for most simulations with $E_i = 0.5$ V/nm deviate from this, repeatedly exhibiting better orientation than expected, especially in the lower region of final electric field strengths.
strength as seen in Figure 11. This deviation is traced back to the pre-orientation phase, where the protein samples oriented in $E_i = 0.5 \text{ V/nm}$ on average gained considerably less kinetic energy compared to the other setups, along with better orientation, which is also an explanation for the unexpected results in Figure 13.

4.2 Energy and Damage

4.2.1 Kinetic Energy

![Diagram showing kinetic energy for four different simulation setups.](image)

(a) $E_i = 0.2 \frac{\text{V}}{\text{nm}}, E_f = 0.06 \frac{\text{V}}{\text{nm}}$

(b) $E_i = 0.5 \frac{\text{V}}{\text{nm}}, E_f = 0.24 \frac{\text{V}}{\text{nm}}$

(c) $E_i = 0.8 \frac{\text{V}}{\text{nm}}, E_f = 0.32 \frac{\text{V}}{\text{nm}}$

(d) $E_i = 1.0 \frac{\text{V}}{\text{nm}}, E_f = 0.50 \frac{\text{V}}{\text{nm}}$

Figure 12: Kinetic energy in one specific protein sample for four different simulation setups. Every step in the simulations (purple) and the mean over 0.5 ps (red) for the duration of the simulations (1 ns) are displayed. Initial and final electric field strengths are indicated in the plot captions.

An increase in energy in the protein is expected when the protein is subject to an external electric field, as energy in the protein is absorbed from the electric
field in the process of orientation. The kinetic energy in the protein sample shows a slight increment throughout the simulations, as observed in Figure 12.

![Figure 13: Average kinetic energy in the protein over the last 0.5 ns of the simulations for all combinations of $E_i$ and $E_f$.](image)

From Figure 13, it is clear that the average kinetic energy among the protein samples shows dependence only on the initial electric field strength, with no to slight dependence on final electric field strength. There is, however, a larger spread with lower $E_f$, indicating an existence of dependence on the different protein samples used.
4.2.2 Temperature

Figure 14 shows temperature dependence in the protein being dominated mainly by the initial electric field strength, $E_i$, with no or slight dependence on the final electric field strength, $E_f$. A larger spread with lower $E_f$ is observed.

A higher initial electric field strength should result in higher kinetic energy and temperature in the protein. The deviating behavior of $E_i = 0.5$ V/nm in Figures 13 and 14 are associated with the pre-orientation phase, where less kinetic energy was gained in the protein samples during orientation compared to other initial electric field strength setups.
4.2.3 Structural Damage

Figure 15: Average RMSD in the protein over the last 0.5 ns of the simulations for all combinations of parameters with protein structure at t=0 as reference.

Figure 15 shows that the alpha-carbon RMSD value of the protein mainly depends on initial electric field strength, $E_i$. For low final electric field strengths ($E_f < 0.18 \text{ V/nm}$), higher fluctuations and uncertainties of RMSD are observed. For $E_f \geq 0.18 \text{ V/nm}$, the RMSD remains relatively constant when comparing each initial electric field strength setup. Regardless, the values of the RMSD in all simulations are low, indicating minimal protein damage in all simulations.

Previous simulations and studies have repeatedly disclosed that dipole orientation of a protein, without structural damage to the protein, is achievable. The results in this report reveal promising outcomes, with simulations indicating that the dipole orientation of such a protein can indeed be sustained when the applied electric field strength is lowered. The dependence on the initial electric field strength (in which the protein was originally oriented) is strong for final electric field strengths $E_i \leq 0.26 \text{ V/nm}$, with a higher initial electric field strength resulting in better orientation. Above this limit, the initial electric field strength is of low significance, since orientation is adequately sustained regardless of $E_i$. 

22
As neither the protein’s RMSD nor kinetic energy and temperature are notably affected by the final electric field strength, there are some margins for errors.

4.3 Further Research

This report investigated the possibilities of sustaining protein dipole orientation in lower electric field strengths than that in which the protein was originally oriented. This transition of the protein, or change in electric field strength, was initiated in an instant, meaning that it was not time-dependent. To mimic a more realistic situation, a time-dependent ramping down of the electric field strength can be implemented in future projects.

The results reveal some inconsistencies in the simulations, such as unexpected and deviating average values of the degree of orientation and kinetic energy in the protein in different setups. These inconsistencies are rooted in the pre-orientation phase, where a limited number of 10 different protein structures were used for every setup. For even better accuracy and consistency, a higher number of initial structures have to be included. For a more thorough investigation of protein damage, the total RMSD for both the orientation phase and sustaining of orientation could be calculated, although this should not differ much.

Although ubiquitin consists of a large number of atoms, it is a relatively small protein. The ability to orient larger and more complex proteins using electric fields, as well as sustaining the orientation in electric fields of lower strength, should be investigated. This could provide more of a complete picture of dipole orientation for different proteins and electric field strength requirements for successful SPI using dipole orientation.

5 Conclusions

The protein’s final degree of orientation after a reduction in electric field strength was dependent on both initial and final electric field strengths. Even with the protein being considerably pre-oriented in the weaker electric field of strength $E_i = 0.2 \, \text{V/nm}$, the orientation was lost when the protein transitioned into an electric field of lower intensity. A final electric field strength of $E_f \leq 0.16 \, \text{V/nm}$ resulted in a loss of protein orientation where the angle of deviation averaged at $\theta \approx 18^\circ$ or more regardless of initial field strength.

For $0.18 \, \text{V/nm} \leq E_f \leq 0.26 \, \text{V/nm}$, the higher initial electric field strength of $E_i = 1.0 \, \text{V/nm}$ resulted in better final orientation. For a protein transitioning into an electric field in the range of $0.32 \, \text{V/nm} \leq E_f \leq 0.5 \, \text{V/nm}$, orientation was successfully sustained for all initial field strengths $E_i = \{0.5, 0.8, 1.0\} \, \text{V/nm}$ with minor or insignificant differences in the final degree of orientation.
Differences in kinetic energies and temperatures were mainly accumulated in the pre-orientation phase, meaning that these quantities were highly dependent on the initial electric field strength $E_i$. A higher initial electric field strength resulted in higher kinetic energies and temperatures, with slight to no dependence on the final electric field strength, $E_f$.

The RMSD in the protein depended largely on the initial field strength, $E_i$. For lower final field strengths, $E_f \leq 0.12 \text{ V/nm}$, the fluctuations and uncertainties of the RMSD were notably larger. Above this, the RMSD generally remained constant or slightly decreased. The RMSD remained insignificantly low for all combinations of electric field strengths and that the protein’s structure was well preserved in all cases, meaning that the lower strength electric fields affected the protein’s initial structure insignificantly.

The ability to lower the electric field strength yet maintain dipole orientation of an already oriented protein is possible. This suggests that longer times for SPI and/or less structural damage to the protein can be achieved, producing more accurate and effective protein imaging.

References

