Electronic Structure Calculations of Amino Acids Under the Influence of Electric Fields

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

Christopher Prytz
Supervisors: Carl Caleman and Oscar Grånäs
Subject Reader: Erik Marklund

June 7, 2018

UPPSALA
UNIVERSITET

Abstract

By the use of highly intense X-ray pulses, with durations in the scale of tenths of fs, high resolution diffraction images of the structures of individual macro-molecules can be obtained. To generate a self-consistent 3D-image of the structures the diffraction images of thousands of copies of the same molecules have to be compounded. The composition relies on advanced algorithms with long computational times and is not always successful or even possible. Simulations of proteins in electric fields have proven that pre-orientation of molecules could yield results otherwise unobtainable. When a molecule enters the electric field it will experience an increasing field strength, i.e. a time dependent electric field, and the change in the field strength will be a function of the molecules translational velocity. Proteins are compounded by amino-acids. If amino-acids structures are altered it could indicate that the structures of proteins would be altered as well. The initial object of this project was to determine if there exists any critical molecular velocities at which the structures of amino-acids significantly are altered. If the structures are altered the diffraction images will be of other structures than of those intended. The project was based on computer simulations. When approximating the plates of a capacitor as two point charges it was found that the time-lapse would be too long for the simulations intended. Instead it was asked at which field strengths the amino-acids would start to lose their essential structures. The amino-acids were simulated in constant electric fields on a computer at Uppsala University. It was found that the amino-acids could start to lose their structures in fields intended for pre-orientation of proteins. It was also found that the field strengths required for a change in the intramolecular forces depended on the size of the amino-acid as well as the direction of the field in relation to the amino-acids orientation.

Sammanfattning

Genom att exponera enskilda makromolekyler för högintensiva röntgenstrålepulser kan diffraktionsbilder av dessa molekylers strukturer erhållas. För att framställa en självkonsistent 3D-bild av strukturerna måste diffraktionsbilderna från tusentals kopior sammanställas. Sammanställningen är avhängig på avancerade algoritmer som inte alltid lyckas. Simuleringar av proteiner i elektriska fält har påvisat att en förorientering av molekylerna, innan exponering för röntgenpulserna, kan öka antalet fall där sammanställningen till en 3D-bild lyckas. När en molekyl levereras in i ett elektriskt fält kommer den uppleva en tilltagande elektrisk fältstyrka, alltså ett tidsberoende elektriskt fält. Förändringen i fältstyrka kommer vara beroende på molekylens fart med vilken den far in i det elektriska fältet. Proteiner består utav aminosyror. Om en aminosyra skulle komma att förlora sin struktur, till följd av en yttre påverkan från ett externt elektriskt fält, skulle det kunna implicera att även ett protein skulle kunna komma att förlora sin struktur. Om ett protein förlorar sin struktur ger metoden diffraktionsbilder av en struktur som skiljer sig från den man egentligen är intresserad av. Den initiala problemformuleringen var att undersöka om det finns några kritiska hastigheter för aminosyror för vilka deras strukturer skulle komma att förändras. Studien grundades på datorsimuleringar. Då kondensatorplattorna approximerades till två punktladdningar insågs det att tidsförloppet skulle bli för långt för de ämnade simuleringarna. Problemet omformulerades istället till i vilka fältstyrkor som aminosyror börjar förlora sin struktur. Aminosyrorna simulerades i statiska elektriska fält på en dator vid Uppsala universitet. Baserat på denna studie finns det viss grund till att anta att proteiner skulle kunna komma att förlora sin struktur i fält ämnade för orientering till följd av en förändring i de intramolekylära krafterna hos de simulerade aminosyrorna. Förändringen i de intramolekylära krafterna hos aminosyrorna verkar bero på aminosyrans storlek samt riktningen på det elektriska fältet i förhållande till aminosyrans orientering i rummet.

Contents

1 Introduction

Proteins are key-molecules in many bio-chemical reactions and are therefore of great interest in areas such as physiology and medicine. A protein's structure and composition determine its functions. Therefore, accurate depictions of different proteins' structures are of great interest in numerous fields of science. New techniques for the imaging of the structures of macro-molecules are on the rise. By the use of free electron lasers (FEL), emitting X-ray pulses, high resolution diffraction images of the structures of individual macromolecules might be obtained. New methods, utilising electric fields for a pre-orientation of proteins, are in development. Proteins, carrying a net dipole, will align their dipole moments with the electric field (see figure 1). By orienting the molecules prior to the exposure to the X-ray pulses the success rate of the structure determination could be increased and yield results otherwise unobtainable. A molecule traveling from an area without an external electric field into an area with an external electric field will experience an increase in electric field strength. Therefore, the electric potentials of the particles of the molecule will be time-dependent. The time-dependency implies that the translational speed of the molecule will determine the electric potentials of its particles. The faster the molecules travel into the electric field the greater the change in electric potentials will be. If the potentials are significantly altered the molecule's electronic and molecular structure might be affected and the imaging will be of another structure than of the one intended. Since amino-acids are key-components of proteins their potential structural alteration, due to the influence of electric-fields, are of interest. If the amino-acids structures are altered it implies that the structures of proteins could be altered as well. The initial objective of this project was to determine if there exist any critical speeds at which the structures of the three amino-acids alanine, valine and glycine significantly are altered when entering an external electric field. The speeds considered were to be speeds of practical and theoretical relevance. The objective was to be accomplished by analysing computer simulations conducted with the software SIESTA [1]. More about the background of the project can be found in section 2 Review. The amino-acids were to be simulated in time dependent electric fields. In the study it turned out that for physically relevant speeds the change in electric potential was not fast enough for the original method to be applied. The initial problem formulation was no longer of importance. Instead it was asked at which field strengths the amino acids would lose their structure. The amino-acids were simulated in static electric fields with step-wise increasing strengths. More of the methods used can be found in section 3 Method and the results of the study can be found in section 4 Results. The thesis will end with some remarks regarding the project and its results in section 5 Conclusions and Recommendations.

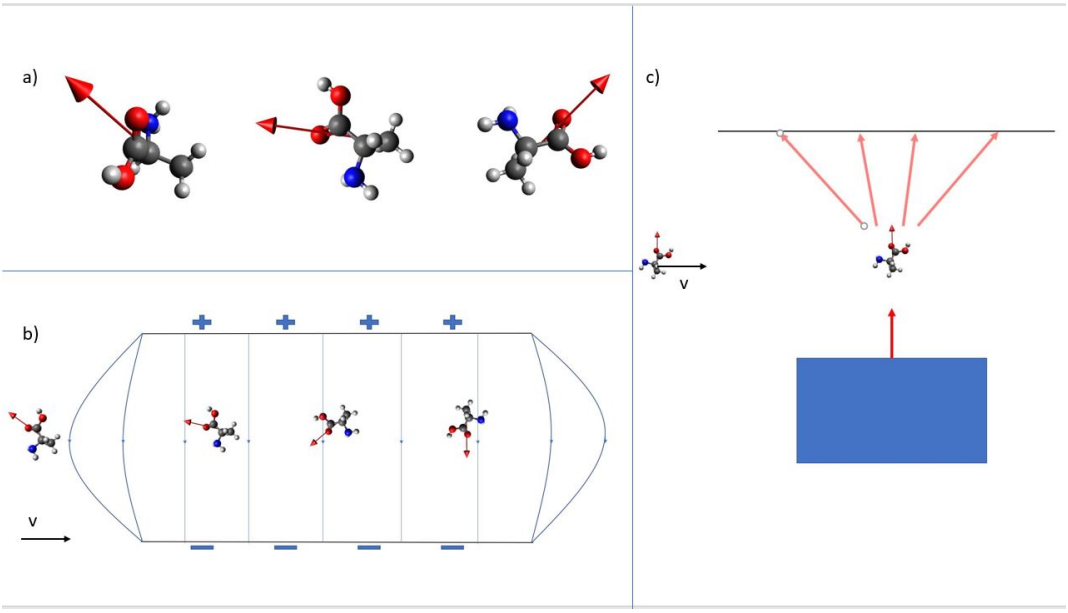

Figure 1: *The process is depicted with the amino-acid alanine and was partly generated with the software Avogadro [2]. The dipole-moment of alanine is depicted with a red arrow. a) Every sample of alanine has its own orientation. b) The alanine-molecule enters an electric field from left to right and becomes oriented. It should be noted that after the orientation the molecule still has unknown rotation around the axis of its dipole-moment. c) The oriented molecules enter the X-ray beam and a diffraction image for every sample is acquired.*

2 Review

In X-ray crystallography a crystalline sample is shot from different angles with an X-ray pulse. Every shot results in a unique diffraction image. The set of all diffraction images is then compounded into a self-consistent 3D-image of the samples structure. Some proteins are hard or impossible to crystallise [3]. Many of these proteins are membrane proteins and could, due to their poteintials as drug targets, be seen as some of the most interesting of proteins [4]. A relatively new set of techniques, single particle imaging (SPI), has the advantage over the more conventional crystallography in that diffraction images can be obtained from single particles in vacuo [5]. When a protein is exposed to an X-ray pulse a diffraction image can be obtained, if the pulse is sufficiently short and intense. Due to the heavy X-ray ionisation the protein is destroyed in the process. The duration of such a pulse is in the scale of tenths of fs. To assemble a sufficient set of diffraction images, for a reconstruction of the structure to be enabled, thousands of samples, each giving a unique diffraction image, have to be exposed to the X-ray pulses. A requisite for the generation of a 3D-image in SPI is therefore that identical copies of the sample can be reproduced and delivered into the X-ray beam.

The next step in the process is the so called orientation recovery. Advanced algorithms are used to determine the orientation of the molecule at the time of exposure. The set of all diffraction images are then, like in crystallography, compounded into a self-consistent 3D-image. Expand Maximize and Compress (EMC) is the preeminent algorithm for the orientation recovery and its first application on experimental data was on the reconstruction of the mimivirus in 2016 [6]. The orientation recovery is a computational heavy process and is in some cases impossible. One way to enhance the accuracy of the orientation recovery is adding information about the molecules orientation prior to the exposure to the X-ray pulses. Molecular dynamics simulations have proven that a pre-orientation could cut the computational time of the EMC by 25-50% and, more importantly, increase the number of cases where the EMC converges [5]. The four proteins, tryptophan cage (Trp-cage), the C-terminal fragment of the ribosomal protein L7/L12 (Ctf), ubiquitin and lysozyme, all carrying a net dipole, were in the simulations successfully oriented in static electric fields with strengths ranging from 100 kV/cm to 30 000 kV/cm. For the shorter simulations, where the fields were of strengths 2000-30000 kV/cm, the time of orientation was in the scale of ps to ns. Ubiquintin kept its structure in fields weaker than 20 000 kV/cm and the other proteins kept theirs in fields weaker than 25 000 kV/cm. For the longer simulations, lasting 50 ns, all proteins, except from Trp-cage which was requiring stronger fields, were oriented to a high degree at a field strength of 1000 kV/cm. For the simulations lasting 50 ns the proteins all kept their structure in fields weaker than 4000 kV/cm. Other studies have found that force induced dissociation and unfolding occur with forces in the scales of 1-100 pN [7]. This range of strengths of forces was in this study taken as a reference for when the changes in the intramolecular forces of the amino-acids studied could result in alterations of the structures of the amino-acids and those of proteins.

Of interest is if and how the electric fields alter the structures of the proteins. Due to the electric field just outside of a parallel plate capacitor not being zero but increasing in strength the closer one gets to the capacitor a molecule traveling towards and eventually enters the capacitor will before the entrance experience an increase in electric field strength. The electric potentials, due to an external electric field E and dependent on the strength of that field, of the electrons of the

molecule will be time-dependent and related to its velocity, that is

$$\frac{dE(r)}{dt} = \frac{dE}{dr}\frac{dr}{dt} \tag{1}$$

where r is the position and t the time It is hypothesised that a fast change in electric potential might alter the electronic, and in turn the molecular, structure of the protein. A wavefunction $\Psi(t)$ of a time dependent many-body-system of N electrons is a function of 3N variables which must satisfy the time dependent Schrödinger equation.

$$i\hbar\frac{\partial}{\partial t}\left|\Psi(t)\right\rangle = \hat{H}\left|\Psi(t)\right\rangle \tag{2}$$

The Hamiltonian will take the following form for the systems considered in this project

$$\hat{H} = [\hat{T} + \hat{W} + \hat{V} + \hat{U}] \tag{3}$$

were $\hat{T}$ is the kinetic energy operator, $\hat{W}$ the operator of the electron-electron interactions, $\hat{V}$ the operator of the external potential due to the electron's interaction with the nuclei and $\hat{U}$ the operator of the external electric field. $\hat{U}$ would be time-dependent. According to the Runge and Gross theorem, for a given initial wavefunction satisfying

$$\left|\Psi(t_0)\right\rangle = \left|\Psi_0\right\rangle \tag{4}$$

and a finite time interval $t_0 < t < t_1$, time dependent external potentials differing only by a trivial gauge transformation corresponds to a unique time dependent electron density [8]. This one-to-one correspondence is the foundation of the time dependent density functional theory (TDDFT). TDDFT reduces many-body-problems of 3N variables into N single-body problems of three variables by approximating the interactions of N electrons with the corresponding time dependent electron density function.

3 Method

The project was limited to three amino-acids. The choice of alanine (fig. 2a), valine (fig. 2b) and glycine (fig. 2c) was based on two reasons. The first reason was that they are three of the smallest amino-acids in existence. A smaller molecule generally entails shorter simulation times. The second reason was that there already existed structural files of them from a previous project at Uppsala University [9]. The simulations were carried out on a computer located at Uppsala

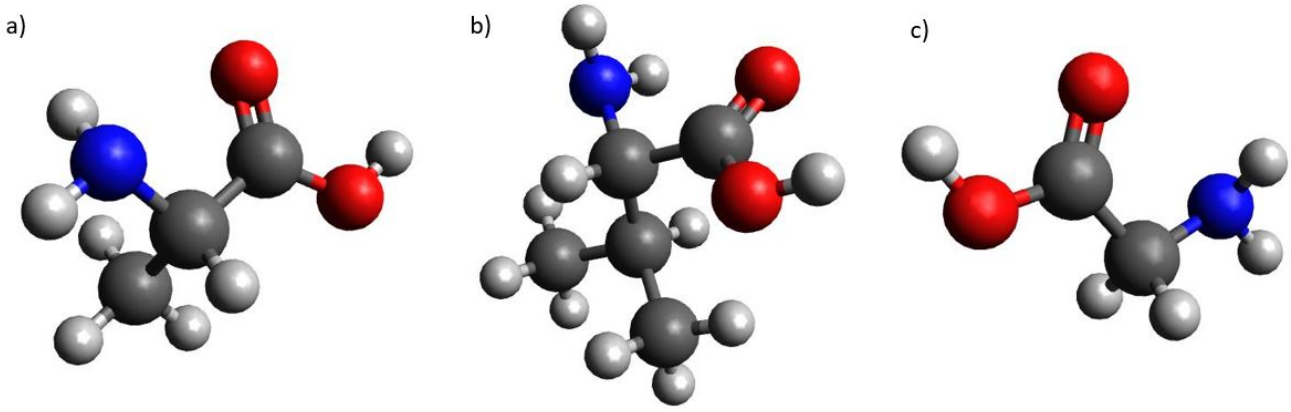

Figure 2: *The three amino-acids of the project are here depicted by the use of Avogadro. The red spheres represent oxygen, the dark grey represent carbon, the lighter grey represent hydrogen and the blue represent nitrogen. The chemical formulas are respectively* $C_3H_7NO_2$, $C_5H_{11}NO_2$ *and* $C_2H_5NO_2$ *for alanine (a), valine (b) and glycine (c).*

University with the software SIESTA. The execution of a simulation with SIESTA requires, apart from the software itself, files containing pseudo-potentials for the different types of atoms found in the amino-acids. The pseudo-potentials function as the ion-electron interactions. The pseudo-potentials were acquired from previous simulations at Uppsala University and could therefore be considered to be thoroughly tested. SIESTA is executed by being fed an input-file containing information about the desired run. The information contains structural information of the considered amino-acid, characteristics of the external electric field, values for different parameters of the run and information about which data that should be out-written to different output-files. The structural information of the amino-acids was obtained from previous simulations at Uppsala University [9]. Essentially it is coordinates for the atoms of the amino-acid as well as the atomic species. A time-dependent external electric field has to be described, in the input-file, as a sum of a gaussian function, a sinus function and a heaviside step function although each of these terms can be set to zero. A time-independent external electric field has its field strength in the x-, y- and z-direction written out. Essential parameters of the run are the number of time-steps, the length of the time-steps and the MeshCutoff. The MeshCutoff defines the plane-wave cutoff energy which affects the convergence and accuracy of the run.

Because of the form of the time-dependent electric field in the input-file the field had to be approximated as a gaussian function of the following form.

$$f(x) = ae^{-(\frac{x-b}{c})^2} \tag{5}$$

The two plates of the capacitor were approximated by two point charges. The field strength, E, was described as a function of the translational speed, v, of the molecule. A more elaborate derivation, than what will be presented in this subsection, can be found in the appendix (Derivation). The following function was derived

$$E(r) = \frac{2q}{4\pi\epsilon_0 r^3} \tag{6}$$

where r is a function of v and the time t, q the charge of the point charges, ϵ_0 the permittivity of free space, d the distance between the two point charges and t_1 the time at which the molecule is at its closest to the two point charges.

$$r = \sqrt{(v(t-t_1))^2 + (\frac{d}{2})^2} \tag{7}$$

For physically relevant speeds the change in electric potential was not fast enough to induce electronic excitation. Even when the speeds approached one third of the speed of light the time lapse would be in about 0.5 ns which would, for this project, take too much time to simulate. Instead it was asked at which field strengths the amino-acids would start to lose their structure.

The three amino-acids were simulated in static fields that increased in strength for each simulation. The intramolecular forces of the molecule in an external electric field were compared to the intramolecular forces of the molecule without the influence of an external field. Stronger fields were applied until the change in intramolecular forces reached an absolute value of 0.1 eV/Å which approximately is 160 pN. In order to simulate the molecules entering the field with different orientation five different directions of the electric field was simulated for each amino-acid. The fields were in the $\hat{x}$-, the $\hat{y}$-, the $\hat{z}$-, the $(\hat{x}+\hat{y}+\hat{z})$- and the $-(\hat{x}+\hat{y}+\hat{z})$-direction. The fields in the $\hat{x}$-, the $\hat{y}$- and the $\hat{z}$-direction were increased in strength, compared to the previous field, with 0.01 V/Å until a change in the intramolecular forces of 0.1 eV/Å was reached. The fields in the $(\hat{x}+\hat{y}+\hat{z})$- and the $-(\hat{x}+\hat{y}+\hat{z})$-direction had, respectively, equal components in the $\hat{x}$-, the $\hat{y}$- and the $\hat{z}$-direction and for these fields the strength of each component was changed by 0.01 V/Å until a change of 0.1 eV/Å in the intramolecular forces was reached. When 0.1 eV/Å was reached additional simulations were conducted with fields varying by 0.001 V/Å in the interval of 0.1 V/Å were 0.1 eV/Å had been reached. For each amino-acid one representative geometry was used. The MeshCutoff was set to 200 Ry.

4 Results

Glycine was, overall, the amino-acid most susceptible to a change in its intramolecular forces due to the influence of an external electric field (table 1 and fig. 3). It had the second smallest absolute value of its inital dipole-moment and the second smallest difference in the energy of the highest occupied electron orbital and the lowest occupied electron orbital. Valine was the least susceptible to a change in its intramolecular forces. It had the smallest absolute value of its initial dipole-moment and the smallest difference in the energy of the highest and lowest occupied electron orbitals. Alanine had the biggest absolute value of its dipole moment and the biggest difference in energy of the highest and lowest occupied electron orbitals.

Amino-acid	E [V/Å]	k [e]	p_0 [D]	p_1 [D]	$\Delta p = p_1 - p_0$ [D]	Energy Difference [eV]
Glycine	0.068 ($\hat{x}$)	1.48	4.80	4.70	-0.10	5.02
Alanine	0.090 ($\hat{z}$)	1.12	5.00	5.24	0.24	5.20
Valine	0.129 ($\hat{z}$)	0.76	4.67	5.00	0.33	4.75

Table 1: *The table contains data from the simulations of the weakest fields required for a change in the intramolecular forces of 0.1 eV/Å for each amino-acid respectively. E stands for the electric field strength at which 0.1 eV/Å was attained and the direction of the field is given in the parenthesis. See fig. 5, fig. 7 and fig. 9 for directional references and tables 4-6 in the appendix for references of the electric field strengths. k is the k-value of a linear curve-fit of the form $y = kx$ to the data-points of the fields in the directions requiring the weakest field strengths for a change of 0.1 eV/Å in the intramolecular forces for each amino-acid respectively (fig. 3). p_0 is the absolute value of the dipole moment in a state without any external electric field. p_1 is the absolute value of the dipole moment in a state with the external field strength E. The Energy Difference is the difference between the highest occupied orbital and the lowest occupied orbital at the state without any external electric field.*

Amino-acid	$\hat{x}$	ΔIF	$\hat{y}$	ΔIF	$\hat{z}$	ΔIF	$\hat{x}+\hat{y}+\hat{z}$	ΔIF	$-\hat{x}-\hat{y}-\hat{z}$	ΔIF
Glycine	0.05	0.0740	0.13	0.0626	0.09	0.0606	0.0520	0.0629	0.0520	0.0626
Alanine	0.18	0.0615	0.08	0.0667	0.06	0.0661	0.1732	0.0634	0.1732	0.0617
Valine	0.14	0.0605	0.29	0.0601	0.08	0.0621	0.2425	0.0646	0.2252	0.0637

Table 2: *Represented in this table are the strengths of the electric fields, given in V/Å, which resulted the change in intramolecular forces to pass 0.06 eV/Å for each amino-acid in each field direction. ΔIF stands for the change in intramolecular forces and is given in eV/Å.*

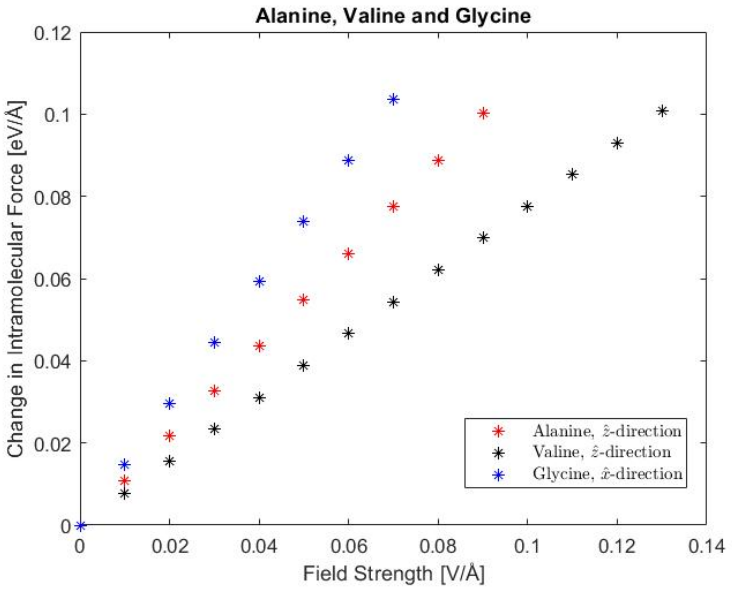

Figure 3: *The absolute value of the largest change in the intramolecular forces of alanine, valine and glycine are here plotted against the corresponding field strengths. The three sets of data corresponds to the weakest field strengths required for a change of 0.1 eV/Å in the intramolecular forces for each amino-acid respectively.*

Amino-acid	$\hat{x}$	ΔIF	$\hat{y}$	ΔIF	$\hat{z}$	ΔIF	$\hat{x}+\hat{y}+\hat{z}$	ΔIF	$-\hat{x}-\hat{y}-\hat{z}$	ΔIF
Glycine	0.01	0.0148	0.01	0.0050	0.01	0.0067	0.0173	0.0210	0.0173	0.0210
Alanine	0.01	0.00341	0.01	0.0086	0.01	0.0108	0.0173	0.0063	0.0173	0.0063
Valine	0.01	0.0043	0.01	0.0020	0.01	0.0078	0.0173	0.0045	0.0173	0.0045

Table 3: *Represented in this table are the simulations at field strengths 0.01 V/Å in the $\hat{x}-$, the $\hat{y}-$ and the $\hat{z}$-direction and 0.0173 V/Å in the $(\hat{x}+\hat{y}+\hat{z})$- and $(-\hat{x}-\hat{y}-\hat{z})$-direction. ΔIF, as in table 2, stands for the change in intramolecular forces and is given in eV/Å.*

4.1 Alanine

The weakest field required for a change of 0.1 eV/Å in intermolecular forces for alanine was in the $\hat{z}$-direction with a value of 0.090 V/Å. The strongest field required was in the $\hat{x}$-direction and its value was 0.293 V/Å. In the state without an external electric field alanine had the dipole moment $\mathbf{p}_0 = (-1.48, 1.67, 4.48)$D and in the state with an external electric field of 0.090 V/Å in the $\hat{z}$-direction the dipole moment was $\mathbf{p}_1 = (-1.49, 1.67, 4.73)$D. The difference is $\mathbf{p}_1 - \mathbf{p}_0 = (-0.01, 0, 0.25)$D.

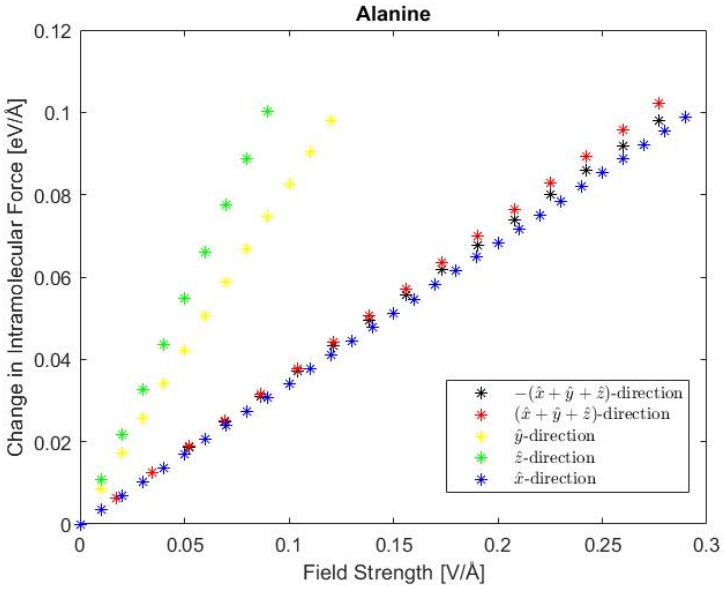

Figure 4: *The absolute value of the largest change in the intramolecular forces of alanine are here plotted against the corresponding field strengths.*

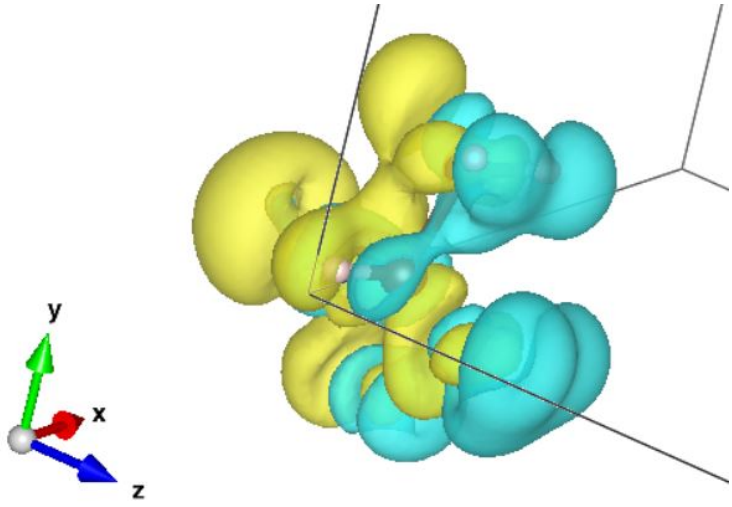

Figure 5: *This figure is a visualisation, generated in VESTA [10], of the change in electron density between the state with no external electric field and the state with a field of 0.090 V/Å in the $\hat{z}$-direction for alanine. The electrons have moved from the blue areas to the yellow areas. The iso-surface level is here 7.9×10^{-5} e/a_0^3, where a_0 is the Bohr radius.*

4.2 Valine

The weakest field required for a change of 0.1 eV/Å in intermolecular forces for valine was in the $\hat{z}$-direction with a value of 0.129 V/Å. The strongest field, 0.471 V/Å, was required in the $\hat{y}$-direction. This field was the overall strongest for all simulations. In the state without an external electric field valine had the dipole moment $\mathbf{p}_0 = (-3.75, -0.25, 2.78)$D and in the state with an external electric field of 0.129 V/Å in the $\hat{z}$-direction the dipole moment was $\mathbf{p}_1 = (-3.78, -0.25, 3.26)$D. The difference is $\mathbf{p}_1 - \mathbf{p}_0 = (-0.03, 0, 0.48)$D.

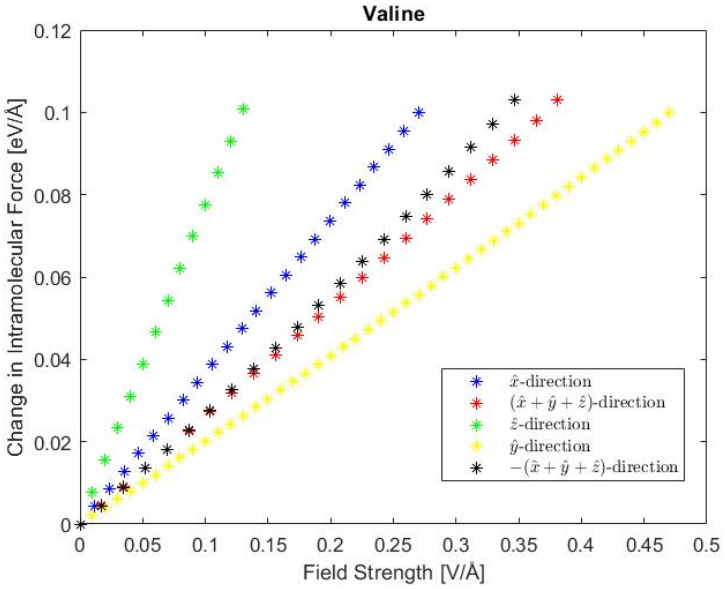

Figure 6: *The absolute value of the largest change in the intramolecular forces of valine are here plotted against the corresponding field strengths.*

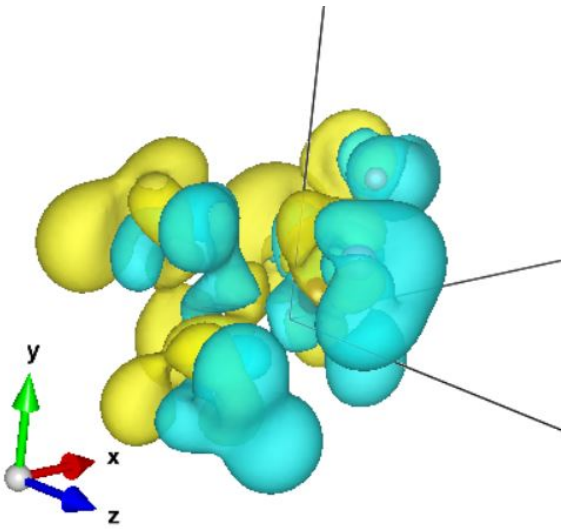

Figure 7: *This figure is a visualisation, generated in VESTA, of the change in electron density between the state with no external electric field and the state with a field of 0.129 V/Å in the $\hat{z}$-direction for valine. The electrons have moved from the blue areas to the yellow areas. The iso-surface level is here $1.1 \times 10^{-4} e/a_0^3$, where a_0 is the Bohr radius.*

4.3 Glycine

Generally glycine was the amino-acid, of the three tested, requiring the weakest fields for the change in intramolecular forces to reach the value of 0.1 eV/Å. The weakest field required was in the $\hat{x}$-direction with a value of 0.068 V/Å. This field was the overall weakest field required for all amino-acids and directions of the fields. The strongest field required was in the $\hat{y}$-direction with a strength of 0.214 V/Å. In the state without an external electric field the glycine molecule had the dipole-moment $\mathbf{p}_0 = (-2.63, -3.41, -2.11)$D and in the state with an external electric field of 0.068 V/Å in the $\hat{x}$-direction the dipole moment was $\mathbf{p}_1 = (-2.48, -3.40, -2.09)$D. The difference is $\mathbf{p}_1 - \mathbf{p}_0 = (0.15, 0.01, 0.02)$D.

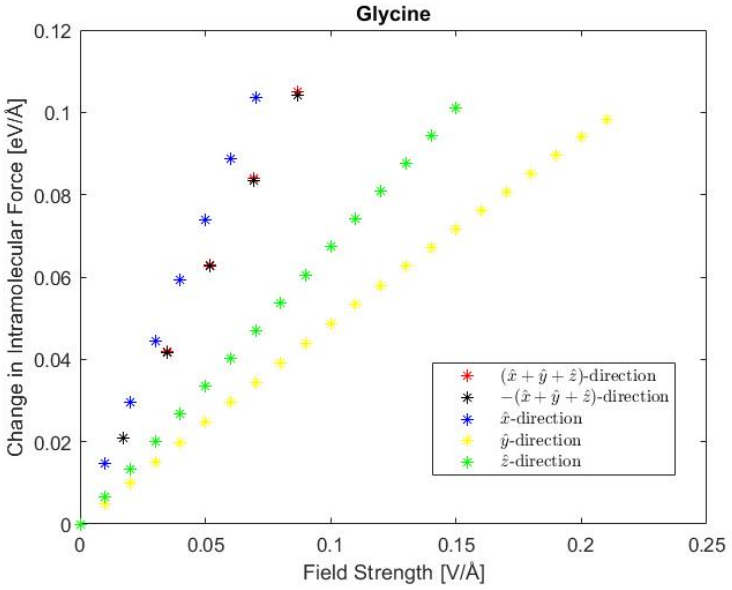

Figure 8: *The absolute value of the largest change in the intramolecular forces of glycine are here plotted against the corresponding field strengths.*

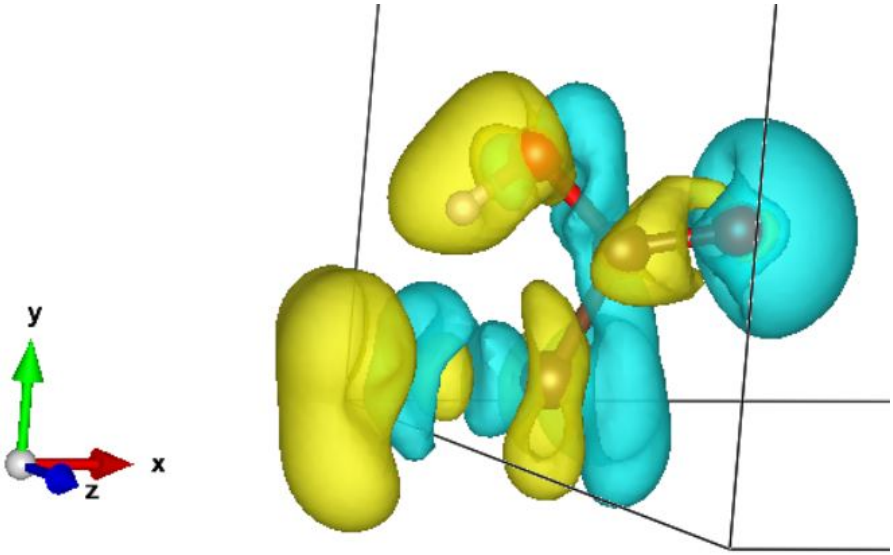

Figure 9: *This figure is a visualisation, generated in VESTA, of the change in electron density between the state with no external electric field and the state with a field of 0.068 V/Å in the $\hat{x}$-direction for glycine. The electrons have moved from the blue areas to the yellow areas. The iso-surface level is here $6.0 \times 10^{-5} e/a_0^3$, where a_0 is the Bohr radius.*

5 Conclusions and Recommendations

The main chain of a protein relies on the three strong bonds the peptide bond, C_α-CO bond and the C_α-N bond [11]. The peptide bond is the covalent bonding between the different amino-acids constituting the protein. The protein folding, on the other hand, relies heavily on weak interactions, such as hydrogen bonding, between the side chains of the protein. It ought to be expected that proteins in electric fields would unfold before it loses its main chain structure. Based on this study, if forces in the scales of 1-100 pN causes protein unfolding, there could be reasons to believe that electric fields might alter the structures of proteins. The field strengths required for a change of 0.1 eV/Å $\approx$ 160 pN in the intramolecular forces of the three amino-acids in this study inquired ranged between 0.068 V/Å = 6800 kV/cm to 0.471 V/Å = 47100 kV/cm. These field strengths are probably much stronger than those required for orientation of proteins, but for the fields of 0.01 V/Å = 1000 kV/cm and 0.0173 V/Å = 1730 kV/cm (table 3) the change in the intramolecular forces ranged between 0.0020 eV/Å $\approx$ 3.204 pN to 0.0210 $\approx$ 33.45 pN which are not insignificantly small.

One should have in mind the limitations of this project. Four of the major limitations are firstly the restriction to only one representative geometry for each amino-acid. A more rigorous study would utilise a greater set of geometries differentiated by vibrations and rotations in the bindings of the molecule. Secondly, the study was limited to five directions of the electric fields. It is clear that the direction of the electric field has an impact on the intramolecular forces. For example, for valine a weaker field in the $\hat{z}$-direction than in the $\hat{y}$-direction was required for the same change in the intramolecular forces. It is possible other results would have followed had more directions of the fields been studied. Thirdly, only three amino-acids were studied. Being three of the smallest amino-acids they are in some sense representative for all amino-acids but to strictly, and without caution, generalise the results to other amino-acids might be ill advised. Lastly is the fact that it was amino-acids, and not proteins, being the subjects of this study.

By comparing the data in table 1 it seems that the bigger the amino-acid is, the stronger the electric fields need to be for the intramolecular forces to increase. It could be of interest to explore if this trend holds for other amino-acids and in extension for proteins as well. The trend holding could be a benefit for protein orientation. When comparing the change in dipole moment, at least for the fields presented in table 1, valine seems to have had its polarisation the most affected of the three amino-acids before reaching a change of 0.1 eV/Å in the intramolecular forces. Since the changes of the dipole moments were, for all three amino-acids, close to zero for the components that were not in the direction of the electric field a change in direction, but not in magnitude, was not overlooked. The result is not surprising considering that the induced dipole moment for a molecule in an external electric field is given by $\mu = \alpha E$ where $\alpha \propto \frac{R^2}{\epsilon}$, E the electric field strength, R the radius of the molecule and ϵ an approximation by a mean value of the excitation energies [12].

Appendix

Tables

$\hat{x}$	ΔIF	$\hat{y}$	ΔIF	$\hat{z}$	ΔIF	$\hat{x}+\hat{y}+\hat{z}$	ΔIF	$-\hat{x}-\hat{y}-\hat{z}$	ΔIF
0.291	0.09921	0.121	0.09884	0.081	0.09000	0.088	0.09636	0.106	0.09853
0.292	0.09955	0.122	0.09961	0.082	0.09115	0.090	0.09701	0.107	0.09912
0.293	0.09988	0.123	0.10037	0.083	0.09229	0.092	0.09766	0.109	0.09972
0.294	0.10022	0.124	0.10113	0.084	0.09345	0.094	0.09832	0.111	0.10032
0.295	0.10056	0.125	0.10189	0.085	0.09460	0.095	0.09897	0.113	0.10092
0.296	0.10090	0.126	0.10265	0.086	0.09575	0.097	0.09962	0.114	0.10151
0.297	0.10124	0.127	0.10341	0.087	0.09690	0.099	0.10027	0.116	0.10211
0.298	0.10158	0.128	0.10417	0.088	0.09806	0.100	0.10092	0.118	0.10271
0.299	0.10192	0.129	0.10492	0.089	0.09921	0.102	0.10157	0.120	0.10330
-	-	-	-	0.090	0.10037	-	-	-	-

Table 4: Presented in this table are the values of the changes in the intramolecular forces of alanine for the electric fields varied by 0.001 V/Å. For the fields in the $(\hat{x}+\hat{y}+\hat{z})$- and the $-(\hat{x}+\hat{y}+\hat{z})$-direction each component in the $\hat{x}$-, the $\hat{y}$- and in the $\hat{z}$-direction was varied by 0.001 V/Å.

$\hat{x}$	ΔIF	$\hat{y}$	ΔIF	$\hat{z}$	ΔIF	$\hat{x}+\hat{y}+\hat{z}$	ΔIF	$-\hat{x}-\hat{y}-\hat{z}$	ΔIF
0.230	0.09990	0.470	0.09982	0.121	0.09381	0.366	0.09854	0.331	0.09779
0.231	0.10034	0.471	0.10004	0.122	0.09458	0.372	0.09902	0.333	0.09837
0.230	0.10078	0.472	0.10027	0.123	0.09535	0.369	0.09951	0.334	0.09895
0.233	0.10122	0.473	0.10050	0.124	0.09613	0.371	0.09999	0.336	0.09953
0.234	0.10166	0.474	0.10072	0.125	0.09690	0.372	0.10048	0.338	0.10011
0.235	0.10210	0.475	0.10095	0.126	0.09767	0.374	0.10096	0.340	0.10069
0.236	0.10254	0.476	0.10117	0.127	0.09844	0.376	0.10145	0.341	0.10127
0.237	0.10298	0.477	0.10140	0.128	0.09921	0.378	0.10193	0.343	0.10185
0.238	0.10341	0.478	0.10163	0.129	0.09998	0.379	0.10242	0.345	0.10243
0.239	0.10386	0.479	0.10185	0.130	0.10075	-	-	-	-

Table 5: Presented in this table are the values of the changes in the intramolecular forces of valine for the electric fields varied by 0.001 V/Å. For the fields in the $(\hat{x}+\hat{y}+\hat{z})$- and the $-(\hat{x}+\hat{y}+\hat{z})$-direction each component in the $\hat{x}$-, the $\hat{y}$- and in the $\hat{z}$-direction was varied by 0.001 V/Å.

$\hat{x}$	ΔIF	$\hat{y}$	ΔIF	$\hat{z}$	ΔIF	$\hat{x}+\hat{y}+\hat{z}$	ΔIF	$-\hat{x}-\hat{y}-\hat{z}$	ΔIF
0.061	0.09021	0.211	0.09872	0.141	0.09494	0.071	0.08608	0.071	0.08541
0.062	0.09168	0.212	0.09915	0.142	0.09562	0.073	0.08819	0.073	0.08749
0.063	0.09316	0.213	0.09958	0.143	0.09629	0.074	0.09030	0.074	0.08956
0.064	0.09464	0.214	0.10002	0.144	0.09697	0.076	0.09240	0.076	0.09163
0.065	0.09611	0.215	0.10045	0.145	0.09764	0.078	0.09451	0.078	0.09370
0.066	0.09759	0.216	0.10088	0.146	0.09832	0.080	0.09662	0.080	0.09577
0.067	0.09907	0.217	0.10131	0.147	0.09899	0.081	0.09873	0.081	0.09784
0.068	0.10054	0.218	0.10174	0.148	0.09966	0.083	0.10084	0.083	0.09991
0.069	0.10202	0.219	0.10217	0.149	0.10034	0.085	0.10294	0.085	0.10198

Table 6: Presented in this table are the values of the changes in the intramolecular forces of glycine for the electric fields varied by 0.001 V/Å. For the fields in the $(\hat{x}+\hat{y}+\hat{z})$- and the $-(\hat{x}+\hat{y}+\hat{z})$-direction each component in the $\hat{x}$-, the $\hat{y}$- and in the $\hat{z}$-direction was varied by 0.001 V/Å.

Derivations

The constants of the gaussian of the form

$$f(x) = ae^{-(\frac{x-b}{c})^2} \tag{8}$$

had to be determined from an approximation of the experienced change in electric field strength for a molecule entering an electric field. Since the electric field strength inside a capacitor can be considered to be constant, and since the project only is to consider the effects until a complete entry of the electric field, only the time until the entry was considered. The endpoints of the capacitor plates, being at a distance d from each other, were approximated as two point charges $q_1 = q$ and $q_2 = -q$ (see fig. 10). At the point p of a distance $d/2$ to q_1 and to q_2, respectively, the field strength was assumed to be E and therefore $a = E$. By superposition the sum of the electric fields E_1 and E_2 of q_1 and q_2, respectively, is E at the point p. Take r_1 and r_2 to be the distance from the point charges q_1 and q_2 respectively. With

$$\mathbf{E}_1 = \frac{q_1}{4\pi\epsilon_0 r_1^3}\hat{r} \tag{9}$$

and

$$\mathbf{E}_2 = -\frac{q_2}{4\pi\epsilon_0 r_2^3}\hat{r} \tag{10}$$

we have

$$\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2 = \frac{q_1}{4\pi\epsilon_0 r_1^3}\hat{r} - \frac{q_2}{4\pi\epsilon_0 r_2^3}\hat{r} \tag{11}$$

Since $r_1 = r_2 = d/2$ field strength from (11) can be written as

$$E = \frac{2q}{4\pi\epsilon_0(\frac{d}{2})^3} \tag{12}$$

From (12) an expression for q can be acquired

$$q = \frac{E\pi\epsilon_0 d^3}{4} \tag{13}$$

Having found q the electric field can generally, outside of the capacitor plates, be described as

$$\mathbf{E} = \frac{q}{4\pi\epsilon_0 r_1^3}\hat{r} - \frac{q}{4\pi\epsilon_0 r_2^3}\hat{r} \tag{14}$$

where, if the x-coordinate of the starting point be taken as zero, v is the velocity of the molecule, t being the time the molecule has traveled from its starting position and $t_1 = b$ is the time just before the molecule has entered the capacitor (that is when the molecule has reached the point p)

$$\mathbf{r}_1 = v(t - t_1)\hat{x} + \frac{d}{2}\hat{y} \tag{15}$$

and

$$\mathbf{r}_2 = v(t - t_1)\hat{x} - \frac{d}{2}\hat{y} \tag{16}$$

The absolute values of $\mathbf{r}_1$ and $\mathbf{r}_2$ will be the same

$$r = \sqrt{(v(t - t_1))^2 + (\frac{d}{2})^2} \tag{17}$$

By antisymmetry the $\hat{x}$-components of $\mathbf{E}$ will negate one another and by symmetry

$$\mathbf{E} = \frac{2q}{4\pi\epsilon_0 r^3}\hat{y} \tag{18}$$

Different values for c were to be found by curve-fitting a gaussian function to the function $\mathbf{E}$ for different values of v.

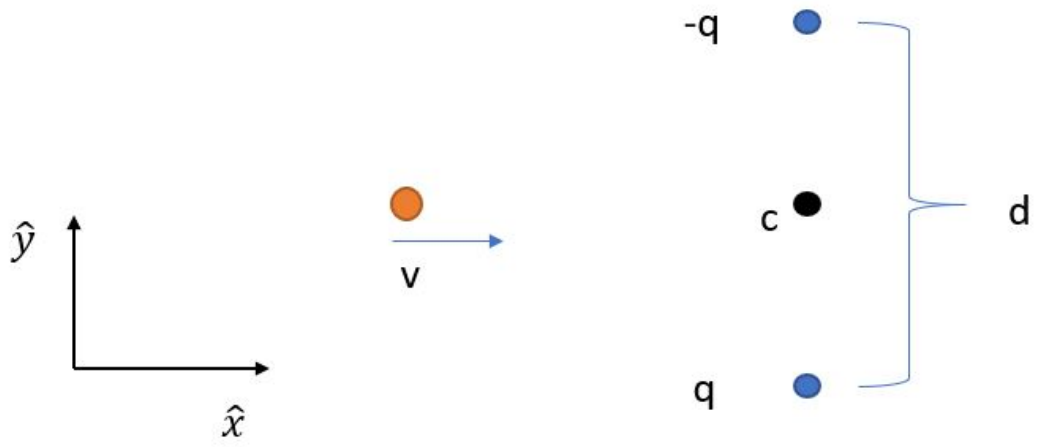

Figure 10: *This is a sketch of the situation. The molecule is represented as an orange circle and the two point charges, with charges q and -q respectively, are represented as blue circles.*

References

[1] MaX Center of Excellence, *SIESTA*. [cited 2018-06-06] Available from: https://departments.icmab.es/leem/siesta/

[2] Avogadro Chemistry, *Avogadro*. Version 1.2.0. Avogadro Chemistry; 2017. [cited 2018-06-06] Available from: https://avogadro.cc/

[3] Dale GE, Oefner C, D'Arcy A. The Protein as a Variable in Protein Crystallization. *J. of Struct. Biol.* 2013 Feb; 1(142), 88-97. DOI:10.1016/S1047-8477(03)00041-8

[4] Chapman HN, Fromme P, Barty A, White TA, Kirian RA, Aquila A, et al. Femtosecond X-Ray Protein Nanocrystallography. *Nature*. 2011 Feb; 470(7332), 73-77. DOI:10.1038/nature09750

[5] Marklund EG, Ekeberg T, Moog M, Benesch JLP, Caleman C.Controlling Protein Orientation in Vacuum Using Electric Fields. *J. Phys. Chem. Lett.* 2017 Sep; 8(18), 4540-4544. DOI:10.1021/acs.jpclett.7b02005

[6] Ekeberg T, Svenda M, Abergel C, Maia FRNC, Seltzer V, C JM. Three-Dimensional Reconstruction of the Giant Mimivirus Particle with an X-ray Free-Electron Laser. *Phys. Rev. Lett.* 2015; 114(9), 1-6. DOI: 10.1103/PhysRevLett.114.098102

[7] Strigl M, Simson DA, Kacher CM, Merkel R. Force-Induced Dissociation of Single Protein A-IgG Bonds. *Langmuir*. 1999 Okt; 15(21), 7316-7324. DOI:10.1021/la990259z

[8] Engel E, Dreizler RM. *Density Function Theory: An Advanced Course*. Berlin Heidelberg: Springer; 2011. DOI:10.1007/978-3-642-14090-7

[9] Trygg, S. A Computational Study of Dissociation Pathways in Highly Ionized Biological Molecules.

[10] Momma K, Izumi F. VESTA 3 for three-dimensional visualization of crystal, volumetric and morphology data. *J. Appl. Crystallogr.* 2011; 44, 1272-1276

[11] Fromm HJ, Hargrove MS. *Essentials of Biochemistry*. Berlin Heidelberg: Springer; 2012. DOI:10.1007/978-3-642-19624-9

[12] Atkins PW. *Physical Chemistry. Fifth Edition.* Oxford: Oxford University Press; 1994.