BPTI Report

Alicja Nowakowska, 234857

1. Introduction

BPTI – Bovine Pancreatic Trypsin Inhibitor (Organism: Bos Taurus [1]) is a small protein responsible for the suppression of protein digestion (e.g. the partition into peptide building blocks). It belongs to the serine protease family of inhibitors and is one of the most frequently studied globular proteins due to its small size and simplicity. Its other important characteristic is extreme conformational stability that influences significantly on its particular folding pathway frequently used as a model for many disulfide-rich proteins [2].

It consists of 58 residues forming a single-chain polypeptide: RPDFC LEPPY TGPCK ARIIR YFYNA KAGLC QTFVY GGCRA KRNNF KSAED CMRTC GGA of the molecular mass 6512. It has three disulfide bonds between: Cys5-Cys55, Cys14-Cys38 and Cys30-Cys51 that guarantee, along with the relatively small hydrophobic core, the high protein stability. Below 100℃ it's rather inert to urea but once one of the disulfide bonds is reduced it gets denatured easily. Furthermore, it was shown that the high pressure has slight effect on the BPTI structure and only minor reversible rearrangements appear if it is imposed (does not fully unfold even at 100 MPa). It is also worth to mention that experiments demonstrated that if all disulfide bonds are eliminated it unfolds even at room temperature [2] [1].

BPTI is known to possess two α-helix regions (residues 3-7 and 47-56) and two β-strands (residues 18-24 and 29-35) that form an anti-parallel β-sheet. Between Lys15 and Ala16 (near the second disfulide bond) the reactive site that binds trypsin or other proteolytic enzymes is located [3]. The matter of salt bridges in the case of BPTI is more complicated. Two methods of determination of the atomic structure are used: NMR and X-Ray and they give different results. Mainly, the NMR research demonstrated existence of the salt bridge between N-terminus and C-terminus while X-Ray studies performed at higher pH caused deprotonation of N-terminus and no salt bridge was observed [4]. In terms of hydrogen bonds, it contains an interesting aromatic hydrogen bonding interaction between residues Gly37, Tyr35, Asn44 . The backbone NH of Gly37 and the side-chain NH2 of Asn44 are hydrogen bond donors on opposite sides of the aromatic Tyr35 ring that are also responsible for stabilization [5].

The root mean square positional fluctuations were also determined on the basis of X-Ray studies and have a value of 0.68 for backbone atoms and 0.78 for sidechain. The radius of gyration tends to oscillate around value of 1.1 nm [5].

1.1. Molecular dynamics studies of BPTI

Given that the experimental studies do not fully reveal the timescale of the atomic motions the molecular dynamics method seems to be an attractive technique for investigation of the proteins. Two important issues are needed to be taken into account when performing the MD simulation: first how well the used force field reproduces movements and second what time scale is available. Such a computational experiment serves analysis of properties that are not measured in a laboratory as well as comparison of the results.

The first simulation of BPTI in vacuum was done already in 1977 by a research group of McCammon. It had a length of 8.8 ps which was a great success that time [6]. Since then various studies have been applied. The first criteria that distinguish them is the usage of the initial structure that can come either from NMR or X-Ray. In [7] both approaches were compared and the

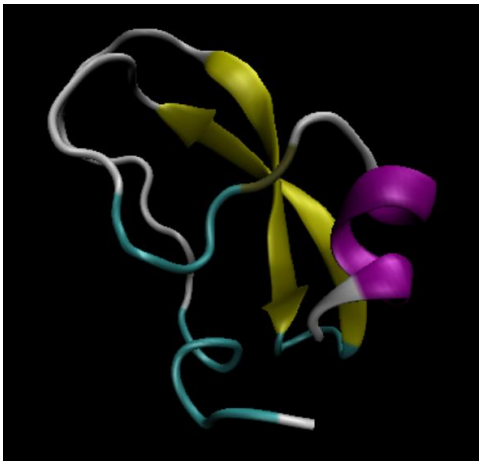
main difference was observed in the behaviour of N-terminus and C-terminus which created the salt bridge if the structure that came from NMR and did not for X-Ray. The question of the proper choice of the force field was discussed in [8] when the close agreement of MD results and the experimental ones for BPTI's peptides were obtained for a combination of SPC/E and GROMOS87 force field while usage of SPC or TIP3P instead of SPC/E gave poor approximation. Also the simulation using OPLS and TIP3P model was done and did not reproduce the desirable data. It was shown that it is essential to have a simulation of length of few hundreds picoseconds to properly sample a conformational equilibrium.

In [5] the impact of hydrostatic pressure as well as solvent density was investigated. It was concluded using AMBER and TIP3P model that the low density (32%) and high density (42%) give larger deviations from the experimental results while those from the range 36%-40% provide good approximation. Nevertheless, the obtained values for the radius of gyration diverged slightly since they oscillated around 0.5 nm and not all the hydrogen bonds were present during whole simulation. In [9] BPTI was simulated at high pressure and low one. The corresponding compressibility coefficients were calculated, radius of gyration were compared and the MD simulation was proven to correctly predict no pressure induced protein unfolding. RMSD oscillated around 1.5 nm during first 5 ns and then around 2.0 nm during next 15 ns.

2. Simulation of BPTI

This report aims to investigate how different parameters such as time step, cutoff, force field and the intital BPTI configuration influence on computational time and obtained results. The provided examination has a goal of analysing what choice guarantees balance between the accuracy of the results and the simulation duration. The MD of 0,5 ns is done with pressure of 1 bar and temperature 300 K and the starting system density equals 698.798 g/l. TIP3P model for water molecules is used and AMBER for protein (excluding the last section). The initial structure of BPTI was extracted from Protein Database (Figure 1.).

Figure 1. Initial structure of BPTI



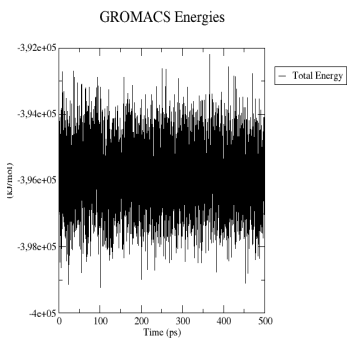
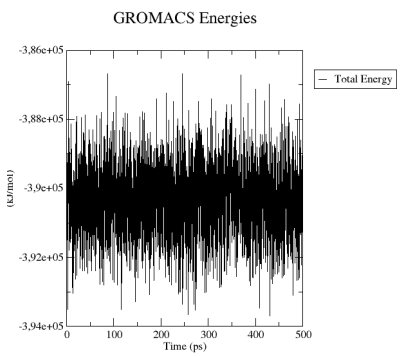
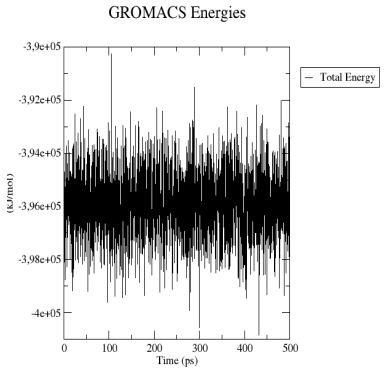
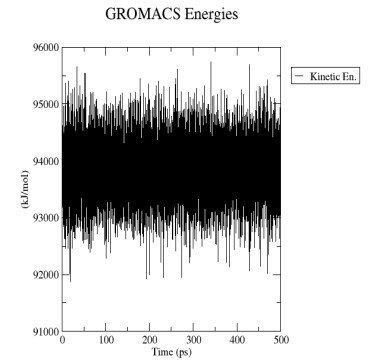
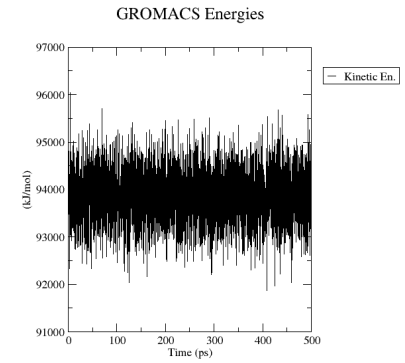
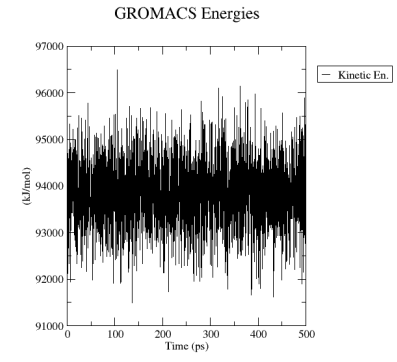
2.1. Structure minimization

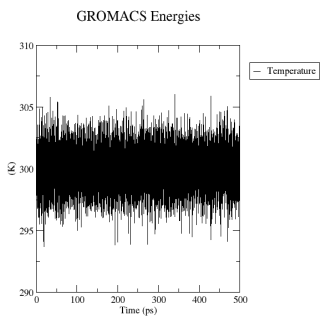
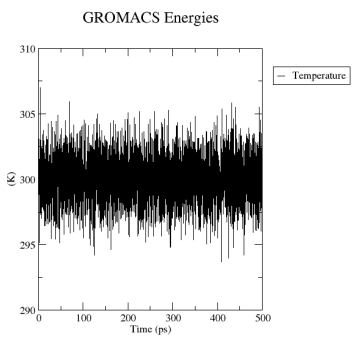
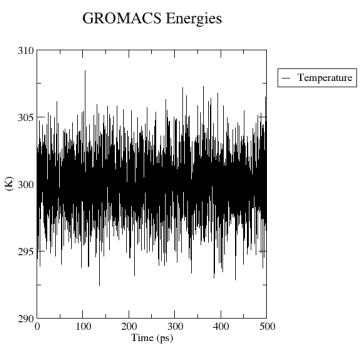
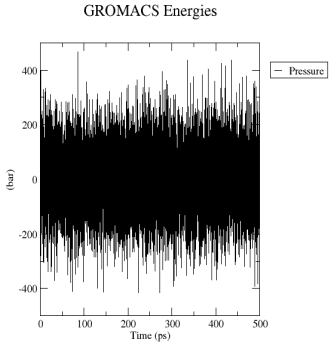
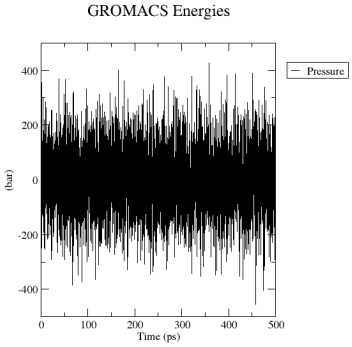
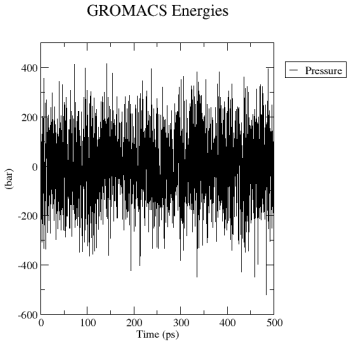
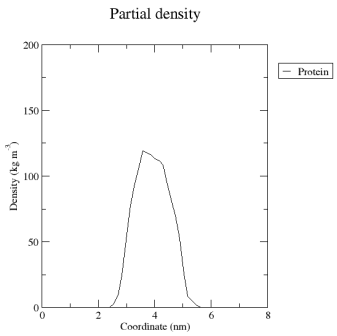
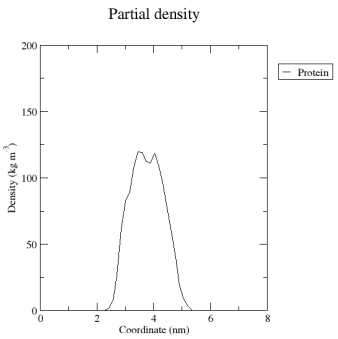
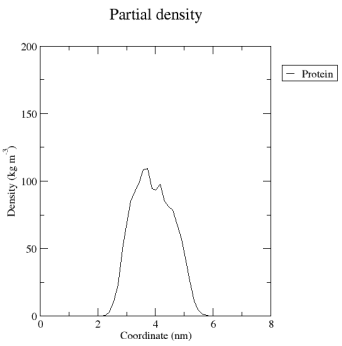
Among known methods for the minimization of the structure the steepest descent and the conjugate gradient are presented. Both are used for solving the system of equations but have different approach and productivity. Mainly, the steepest descent is a fast method since it enables solving equations quickly though requires more iterations to converge comparing to the slower but more productive conjugate gradient. The number of steps required for performing the steepest descent was 2292 and another 2 steps of conjugate gradient were needed.

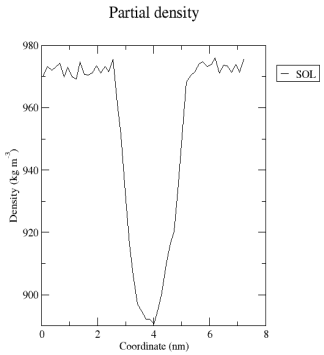
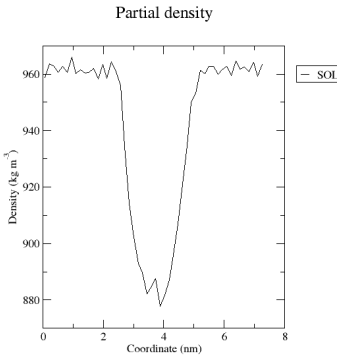
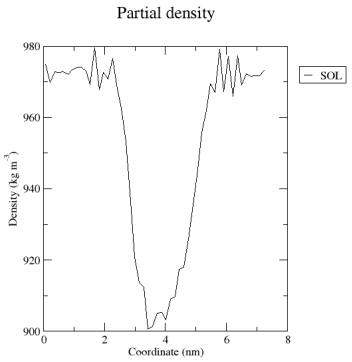
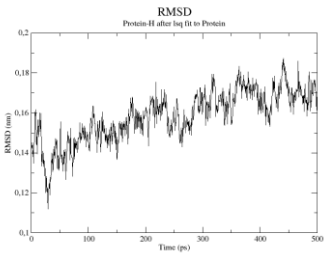
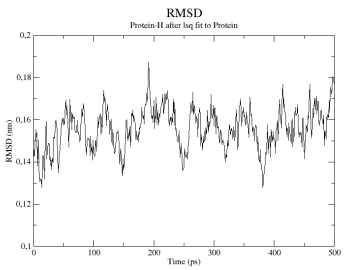
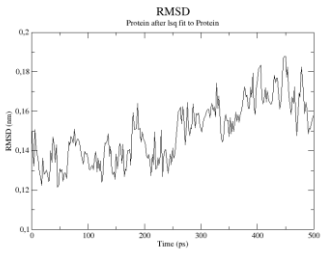
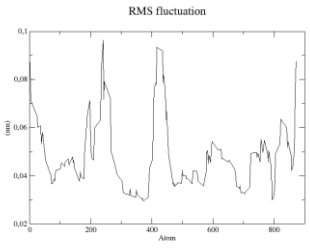
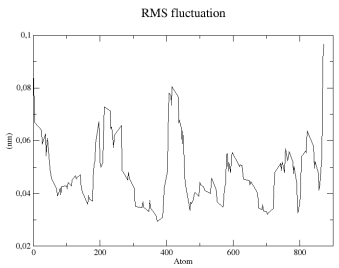
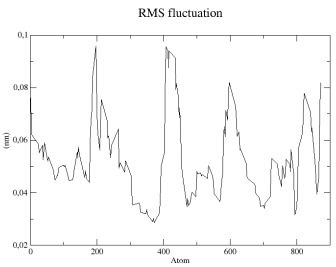
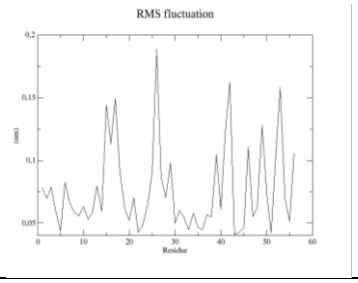
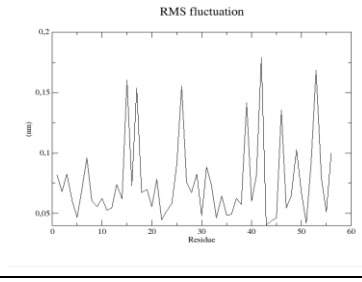
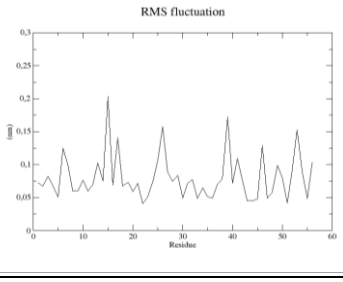
2.2. Choice of the time step

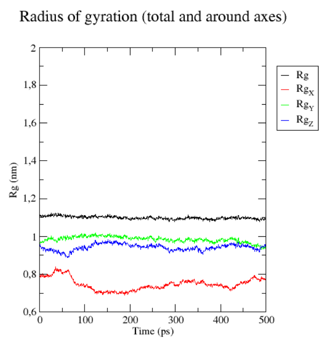
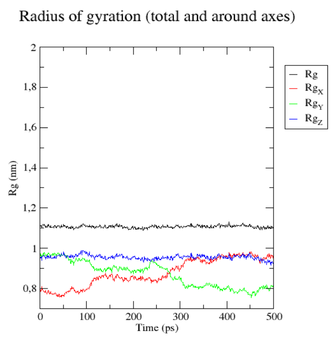
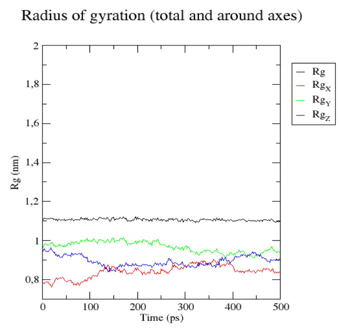
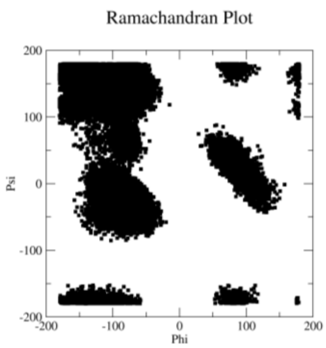
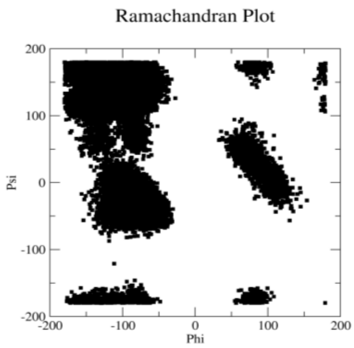
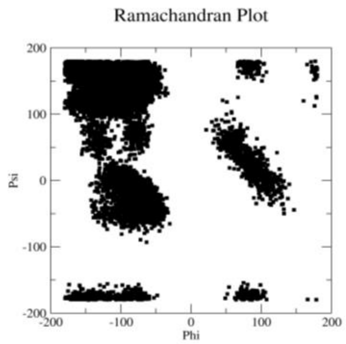
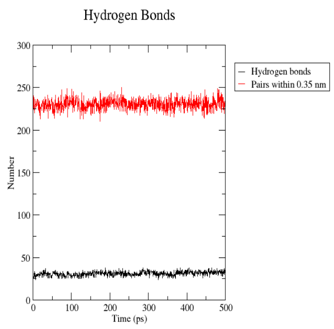
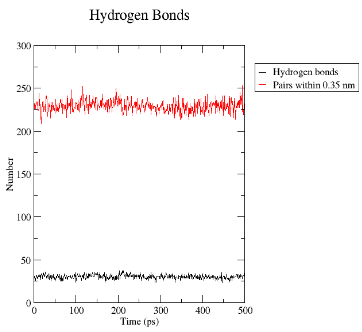
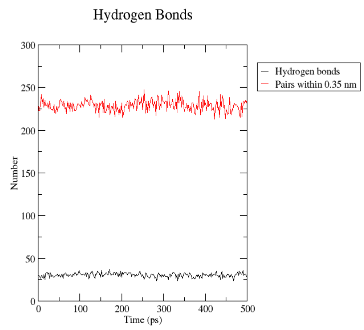
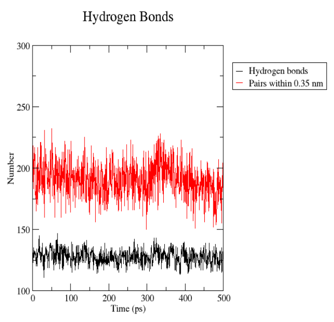
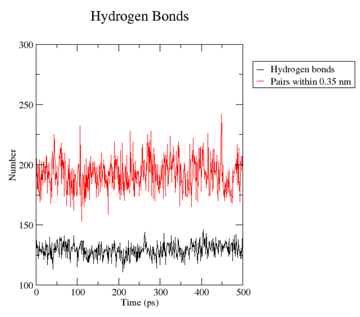
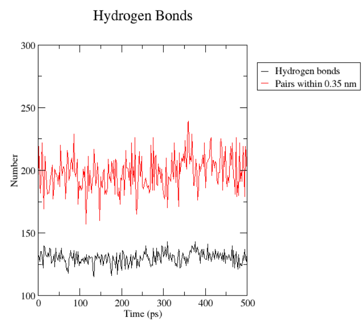
In the first step of the investigation of the final parameters choice the impact of different time step is checked. To obtain a fast MD simulation a bigger time step choice seems to be an attractive option since it enables longer simulation time and thus better sampling. Nevertheless, it comes for the price of poorer numerical stability and more significant truncation error when integrating. Therefore, the aim of the time step research is finding the largest time step that will maintain the conservation of energy. The realization of this goal is usually done through trial and error on the basis of NVE simulations [10]. The analysis for BPTI is presented in the Table 1.

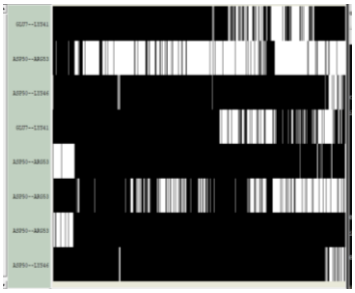
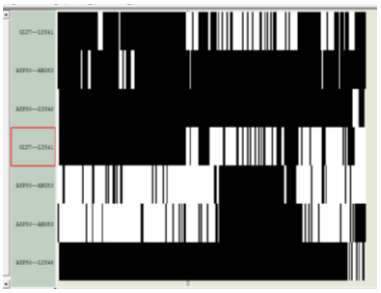
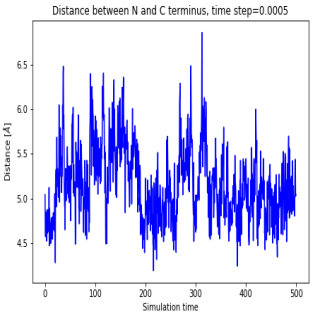
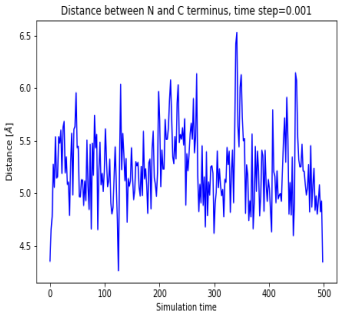
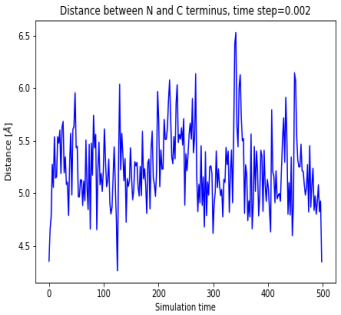
Table. 1. Comparison of the results for different time step values

		Time step = 0.0005	**Time step = 0.001**	**Time step = 0.002**
		Number of steps = 1 000 000	**Number of steps = 500 000**	**Number of steps = 250 000**
		Computation time = 7h 20	**Computation time = 5h 26**	**Computation time = 2h 43**
Total energy	**Plot**			
	Average	-395801.1	-390340.1	-395898.5
	Min	-399226.9	-393698.8	-400846.0
	Max	-392188.2	-386674.7	-390272.9
	RMSD	915.4405	984.5149	1262.7197
Kinetic energy	**Plot**			
	Average	93857.9	93859.44	93841.70
	Min	91875.48	91877.93	91491.25
	Max	95739.95	96038.36	96493.03
	RMSD	510.0995	544.6987	718.5705

Temperature	**Plot**			
	Average	300.0109	300.0159	299.9591
	Min	293.6743	293.6821	292.4461
	Max	306.0268	306.9806	308.4340
	RMSD	1.630501	1.741095	2.296866
Pressure	**Plot**			
	Average	0.4220186	0.936977	1.47605
	Min	-416.5711	-454.220459	-520.07953
	Max	466.9795	428.072906	414.23242
	RMSD	119.1040	119.9219	138.6566
Density of the protein	**Plot**			
	Average	27.76625	27.45692	27.76909
	Min	0	0	0
	Max	119.254	119.70200	109.23800
	RMSD	43.50179	43.68535	39.32106

Density of the solvent	**Plot**			
	Average	953.113	942.4945	953.2095
	Min	890.697	877.7590	900.6410
	Max	975.944	965.8930	979.6700
	RMSD	30.30799	30.23333	27.14432
RMSD of the protein with respect to the X-ray initial structure	**Plot**			
	Average	0.1581963	0.1556792	0.1495056
	Min	0.1119144	0.1278233	0.1215946
	Max	0.1873034	0.1870404	0.1878924
	RMSD	0.0126985	0.0097779	0.0154890
RMSD with respect to initial X-ray structure	**Value**	0.176597 nm	0.182058 nm	0.158174 nm
RMSF of the backbone atoms	**Plot**			
	Average	0.04881726	0.0486756	0.05321845
	Min	0.0295	0.0294000	0.02870000
	Max	0.0965	0.0966000	0.09580000
	RMSD	0.01419451	0.01244497	0.01492791
RMSF for each amino acid residue	**Plot**			

	Average	0.07655357	0.07686786	0.08022857
	Min	0.0407	0.04040000	0.04110000
	Max	0.1885	0.17900000	0.20340000
	RMSD	0.03450388	0.03417331	0.03418021
Radius of gyration	Plot			
	Average	1.09895	1.105342	1.106852
	Min	1.07807	1.088850	1.092000
	Max	1.12305	1.131270	1.122210
	RMSD	0.007782462	0.006388267	0.005786497
Ramachandran Plot	Plot			
Hydrogen bonds in protein-protein pairs	Plot			
	Average	31.2008	30.34331	30.61753
	Min	24	23	24
	Max	38	38	37
	RMSD	2.543352	2.487951	2.401069
Hydrogen bonds in protein-solvent pairs	Plot			
	Average	127.7672	129.2395	130.1594

	Min	111	111	115
	Max	147	146	143
	RMSD	5.653916	5.797803	5.070552
Salt bridges	**Analysis**	 No salt bridge between N and C terminus	 No salt bridge between N and C terminus	 No salt bridge between N and C terminus
	Distance between N terminus and C terminus	 **Mean:** 5.168999	 **Mean:** 5.178435	 **Mean:** 5.239254

On the basis of the presented table a first difference that is evident is the size of the oscillations for kinetic energy, total energy, temperature and pressure that increases along with the time step value. To highlight this observation the plots of the RMSD (Root Mean Square Deviation) of the mentioned properties were created (Figure 2.). Now it may be observed that the growth in the fluctuations magnitude is quite sharp from 0.001 to 0.002 though the rather similar mean values and ranges from the Table 1. did not suggest such a discrepancy. The opposite pattern is noticeable for the number of hydrogen bonds that varies more with the decrease of the time step indicating the bigger number of interactions between the protein and solvent and more breaking events. It is also worth to underline that detectable changes in the root mean square deviation of RMSF of both the backbone atoms and of amino acids residues as well as in the RMSD value are not so remarkable.

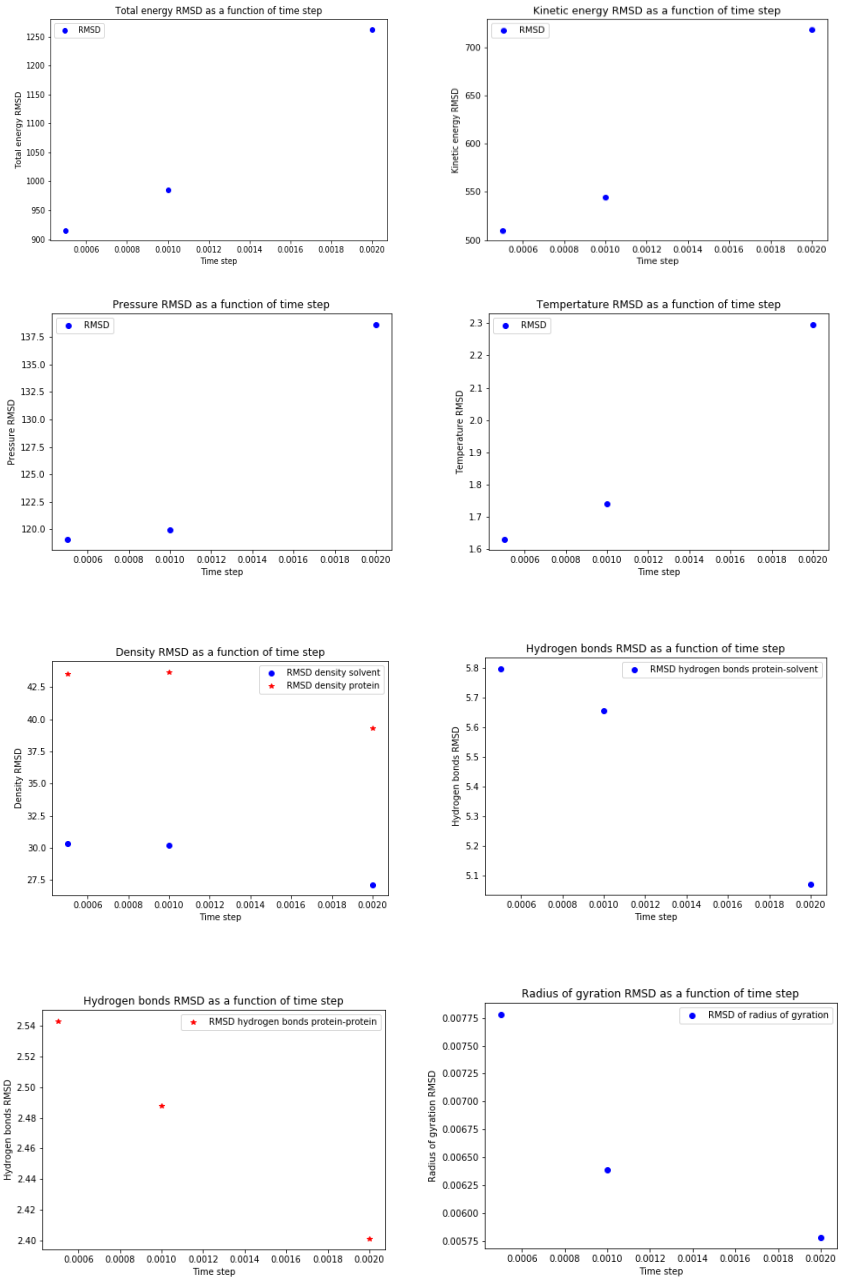


Figure 2. Properties RMSD as a function of time step

Another plot that reveals us the decrease in the simulation accuracy with the raise of the time step is the Ramachandran graph which gets more scattered and is the most condensed for 0.0005. Thus, the larger the interval between subsequent iterations the worse the approximation of the structure is obtained. Moreover, protein creates and breaks more hydrogen bonds for the lowest interval value what demonstrates that important information is lost for bigger simulation intervals. On the other hand, no significant difference is observed in the determination of existence of salt bridges and as expected for no simulation the one between C-terminus and N-terminus is identified. In spite of that, it is noted that the distance between the mentioned protein parts oscillates around the smaller value for 0.0005 while for 0.002 the residues are much further.

Nevertheless, the time needed to perform the full simulation of 0.0005 is extremely long since much more iterations are performed in such a case and at the same time the discrepancies between 0.0005 and 0.001 are not so outstanding as those that appear for 0.002. Therefore at this level of examination the choice of 0.001 seems to be a reasonable one since it lowers down a bit the MD duration but rather preserves the overall system behaviour. However, if one looks for the general

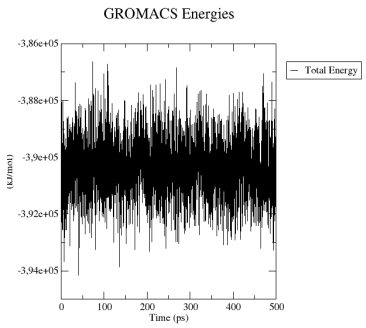
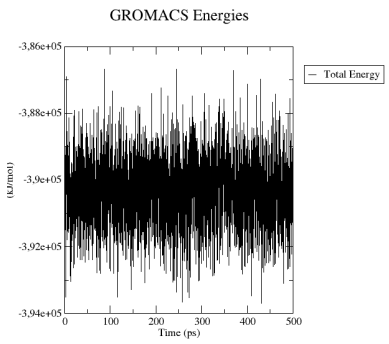
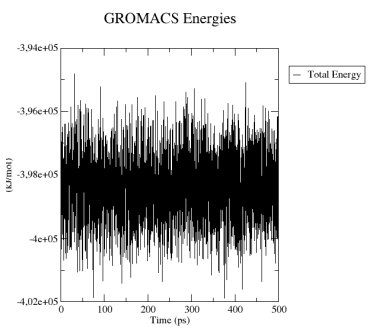
information about the simulation results such as the mean values of the properties the less time-consuming simulation may be done.

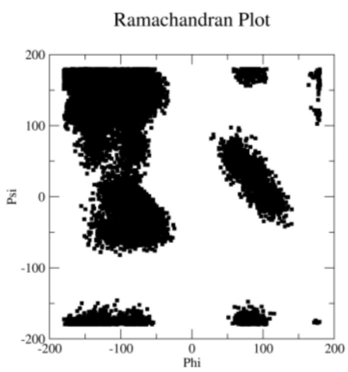
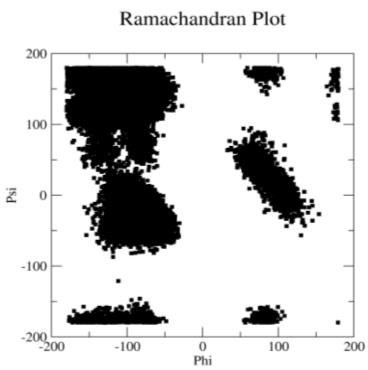
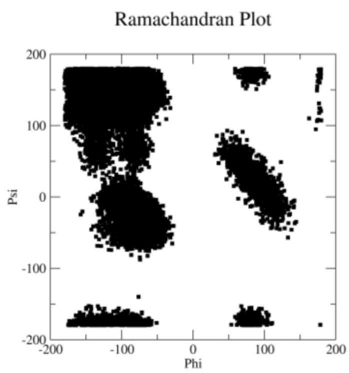
In spite of the above conclusions, the further investigation of the time step values between 0.0005 and 0.001 would be helpful for the diagnosis of the best trade-off point.

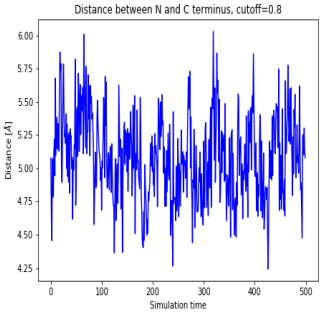
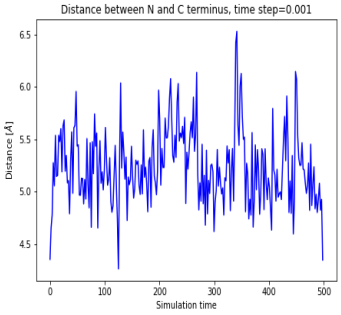
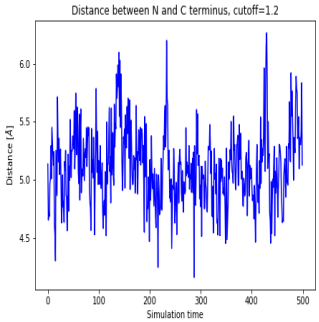
2.3. Choice of the cutoff value

It was widely confirmed that the correct treatment of Van der Walls interactions and electrostatic non-bonded interactions is the key for the realistic MD simulation at the same time it is the most computationally expensive part. It's worth to recall that the higher the cutoff value the more accurate the simulation is though more time consuming. Nowadays, the typical distance used is from the range [0.8 nm, 1.2 nm]. In [11] it was shown that the significant truncation errors appear once the value of the choice is below 0.9 nm. In [12] was shown that the stability of α-helix is a function of cutoff value and that the commonly chosen 1.0 nm is inappropriate since does not ensure the convergence of Coulomb interaction. In this report the results for 0.8 nm, 1.0 nm and 1.2 nm are compared what is shown in Table 2. The times of the simulations obtained are not consistent with the expectation since the one for 1.0 is larger than for 1.2. Therefore, in order to fully approximate the MD duration for different cutoffs, simulations should be repeated and performed for a bigger number of optimized structures.

Table. 2. Comparison of the results for different time cutoff values

		Cutoff = 0.8	Cutoff = 1.0	Cutoff = 1.2
		Computation time = 2h 28	Computation time = 5h 26	Computation time = 4h 24
Total energy	**Plot**			
	Average	-390343.8	-390340.1	-398387.0
	Min	-394147.1	-393698.8	-401877.6
	Max	-386637.0	-386674.7	-394807.5
	RMSD	965.1930	984.5149	978.2929
Kinetic energy	**Average**	93858.85	93859.44	93857.29
	Min	91901.62	91877.93	91849.75
	Max	95755.69	96038.36	95709.52
	RMSD	541.6882	544.6987	542.5701
Temperature	**Average**	300.0140	300.0159	300.0090
	Min	293.7578	293.6821	293.5920
	Max	306.0771	306.9806	305.9295

	RMSD	1.731473	1.741095	1.734291
Pressure	Average	-0.1662923	0.936977	0.7154135
	Min	-410.1463930	-454.220459	-483.7586360
	Max	409.7832950	428.072906	484.9631350
	RMSD	120.5766	119.9219	122.2612
Density of the protein	Average	27.45673	27.45692	27.89925
	Min	0	0	0
	Max	121.40800	119.70200	111.65400
	RMSD	42.77298	43.68535	40.68307
Density of the solvent	Average	942.4879	942.4945	957.678
	Min	881.945	877.7590	896.830
	Max	965.5720	965.8930	980.798
	RMSD	29.95787	30.23333	28.96568
RMSD of the protein with respect to the X-ray initial structure	Average	0.1622488	0.1556792	0.1452263
	Min	0.1261773	0.1278233	0.1107882
	Max	0.2000886	0.1870404	0.1869248
	RMSD	0.013041993	0.009777986	0.014453707
RMSD with respect to initial X-ray structure	Value	0.162887 nm	0.182058 nm	0.145067 nm
RMSF of the backbone atoms	Average	0.04736488	0.0486756	0.04894643
	Min	0.02920000	0.0294000	0.02760000
	Max	0.11420000	0.0966000	0.08520000
	RMSD	0.01232311	0.01244497	0.01169979
RMSF for each amino acid residue	Average	0.07508929	0.07686786	0.0757875
	Min	0.04120000	0.04040000	0.0401000
	Max	0.16060000	0.17900000	0.1610000
	RMSD	0.03268883	0.03417331	0.03090794
Radius of gyrat ion	Average	1.104685	1.105342	1.108754
	Min	1.086420	1.088850	1.090590
	Max	1.123460	1.131270	1.128840
	RMSD	0.006652990	0.006388267	0.006389145
Ramachandran plot	Plot			

Hydrogen bonds in protein-protein pairs	**Average**	30.18762	30.34331	31.01198
	Min	23	23	23
	Max	39	38	38
	RMSD	2.847583	2.487951	2.456798
Hydrogen bonds in protein-solvent pairs	**Average**	128.8483	129.2395	129.3293
	Min	107	111	113
	Max	146	146	144
	RMSD	5.604368	5.797803	5.600475
Salt bridge	**Distance between N and C terminus**	 **Mean:** 5.08403	 **Mean:** 5.178435	 **Mean:** 5.112834

At the first sight the results seem to be rather similar. Property that seems to be affected the most by the change of the cutoff value is the total energy (Figure 3.) that significantly lowers down for 1.2 due to the larger number of interactions considered comparing to the simulations for 0.8 and 1. 0. However, now the changes in the oscillations are not so evident as previously though exist.

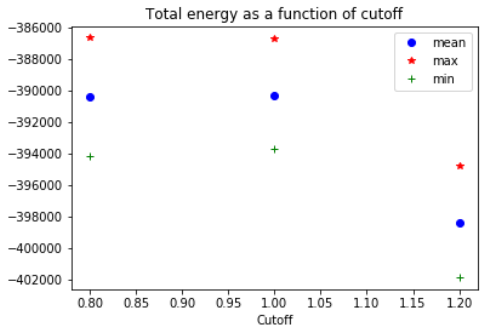


Figure 3. Total energy as a function of cutoff

The fluctuations are described using mean square deviation. When the cutoff is too small the syste m experiences a low number of changes. When it raises to 1.0 the number of considered interaction s raises rapidly but it is not big enough since for 1.2 the oscillations size again drops indicating that the higher the value the smoother and more realistic changes take place. This pattern is observed f or total energies, densities, temperature, RMSF and hydrogen bonds in protein-solvent pairs.

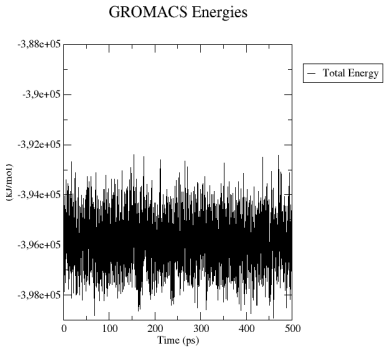
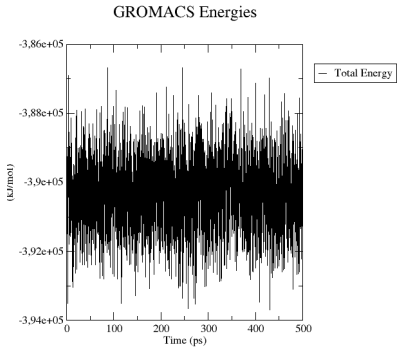
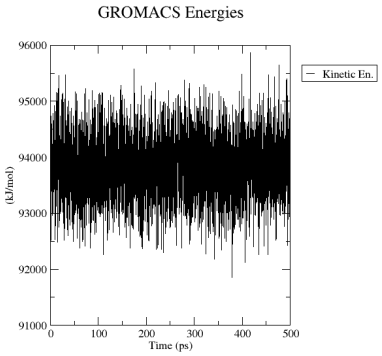
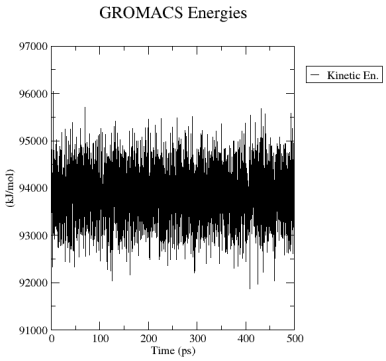
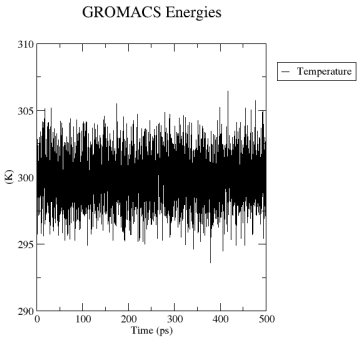
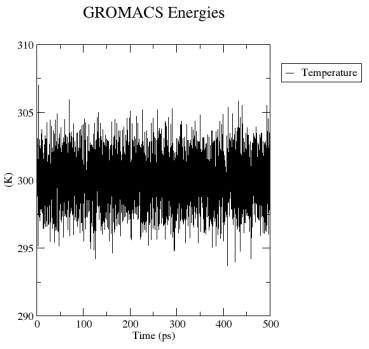
The density and the number of hydrogen bonds raise along the cutoff value though the changes seem to me rather minor. The Ramachandran plot does not distinguish clearly in contrary to the comparison of time steps and values of RMSD and RMSF represent slight increase or decrease pattern though the differences are too small to state it with confidence. The radius of gyration gets closer to the real value with the raise of the cutoff value but changes are slight. Moreover, no evident diversity is observed on the basis of mean and range values. Some dissimilarities are observed when analysing the distance between N-terminus and C-terminus and surprisingly the lowest one is noted for 0.8 probably because the smaller cutoff disables long range interactions and thus residues experience smaller positions changes. For 1.0 the rapid raise of the value is detected since more relations are taken into account and for 1.2 it decreases.

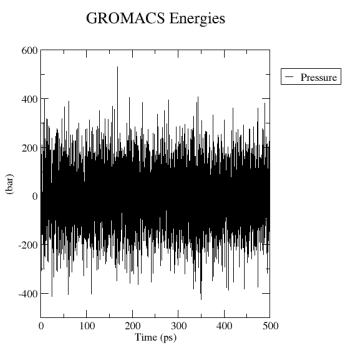
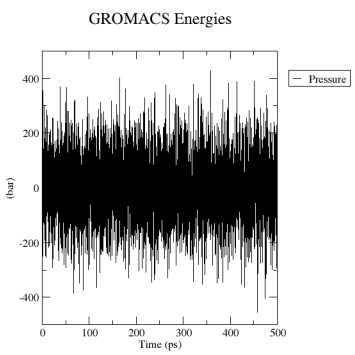
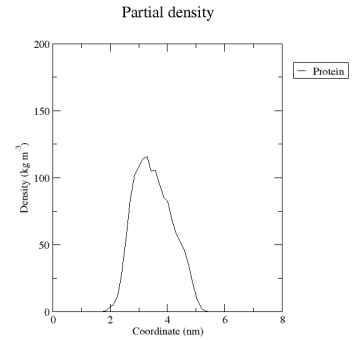
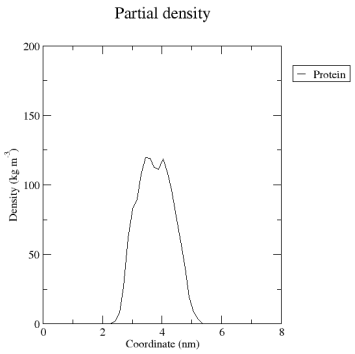
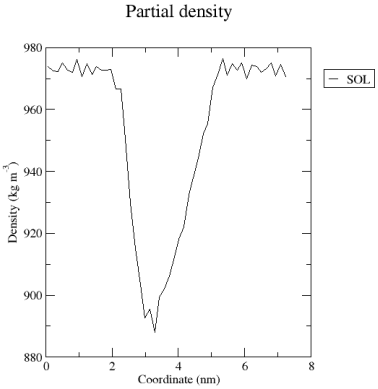
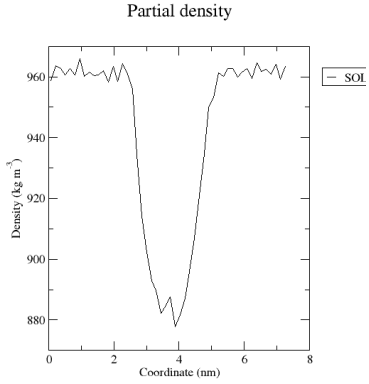
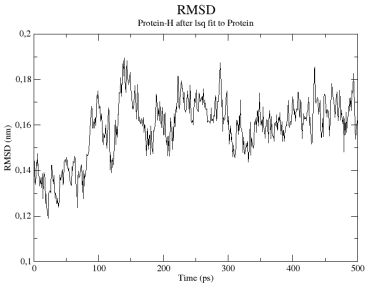
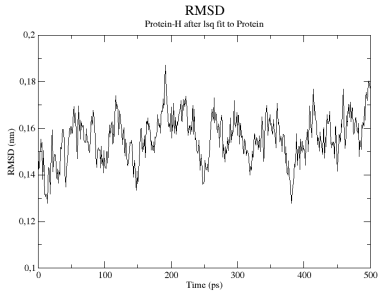
Finally, it may be concluded that the mentioned discrepancies are rather minor excluding the total energy and the distance between N-terminus and C-terminus. Nevertheless, having in mind the results from other papers and the rather similar computation time the best choice seems to be the biggest value of the cutoff, that is 1.2 nm that shall give the lowest number of artifacts since the biggest number of interactions is considered and thus it is the closest to the reality.

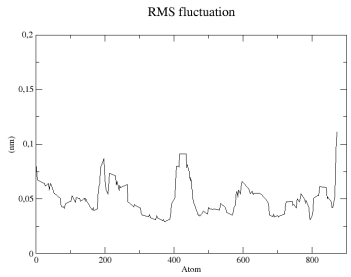
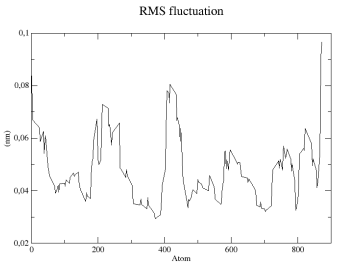
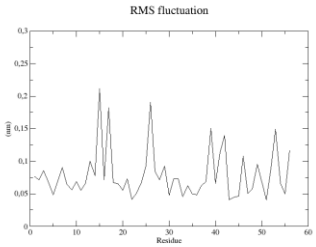
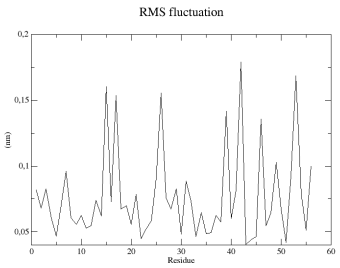
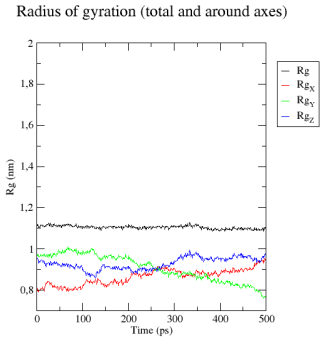
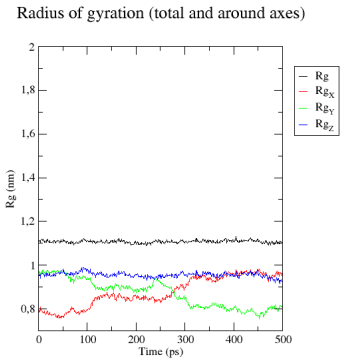
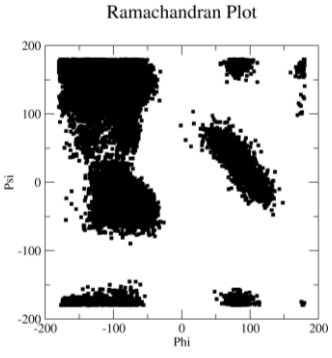
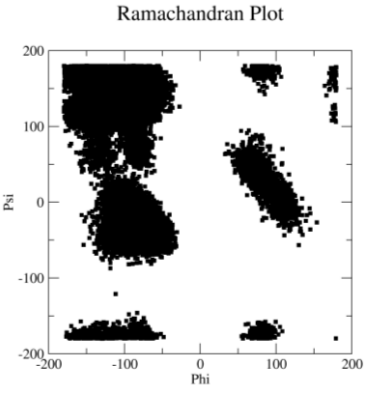
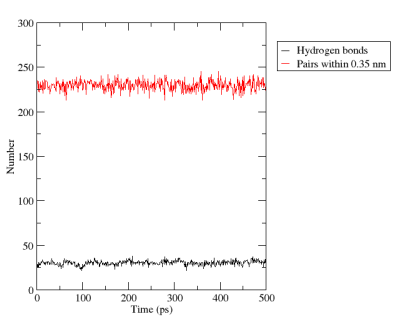
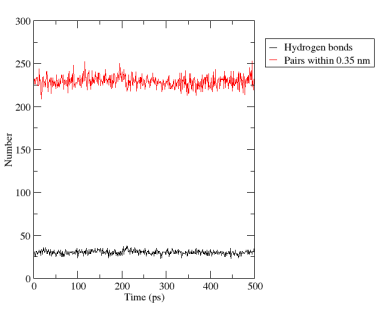
2.4. Molecular coordinates impact

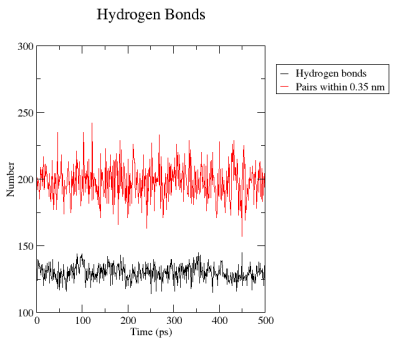
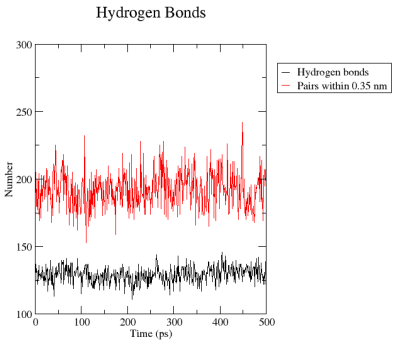
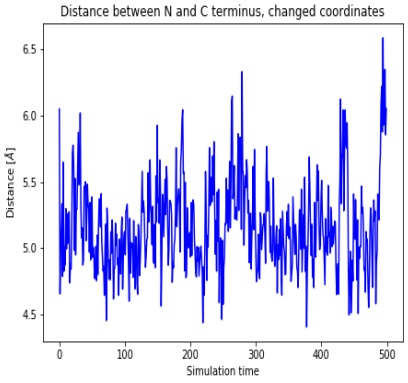
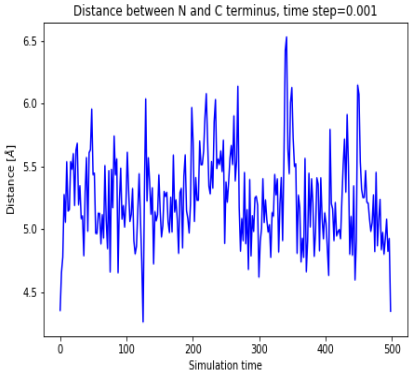
Now the impact of the initial structure on the MD simulation was examined. Specifically, few atoms coordinates were changed and the plots of the properties were done what is shown in the Table 3.

Table 3. Comparison of the results for the run with changed coordinates

		Run with changed coordinates	First run
Total energy	**Plot**		
Kinetic energy	**Plot**		
Temperature	**Plot**		
Pressure	**Plot**		

			
Density of the protein	**Plot**		
Density of the solvent	**Plot**		
RMSD of the protein with respect to the X-ray initial structure	**Plot**		
RMSD with respect to initial X-ray structure	**Value**	0.164281 nm	0.182058 nm
RMSF of the backbone atoms	**Plot**		

			
RMSF for each amino acid residue	**Plot**		
Radius of gyration	**Plot**		
Ramachandran Plot	**Plot**		
Hydrogen bonds in protein-protein pairs	**Plot**		

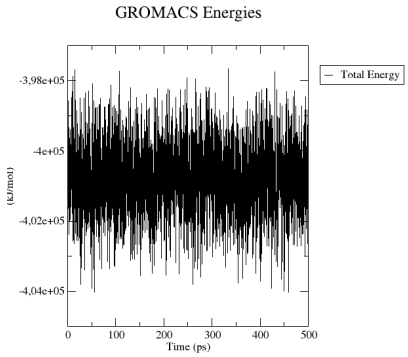
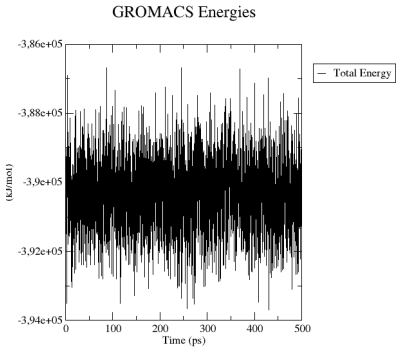
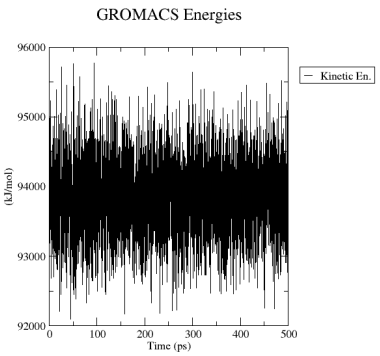
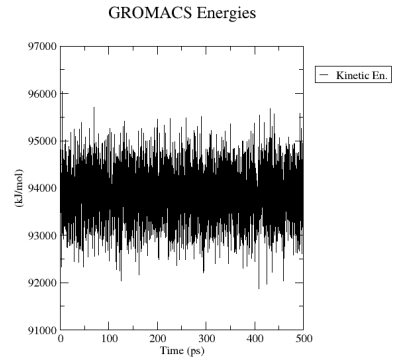
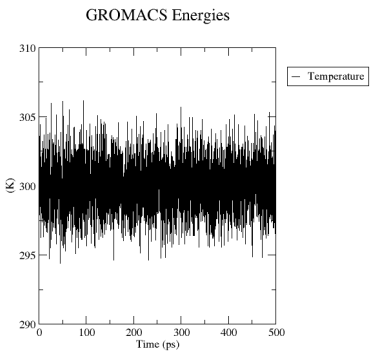
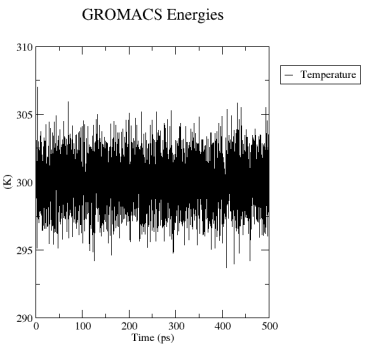
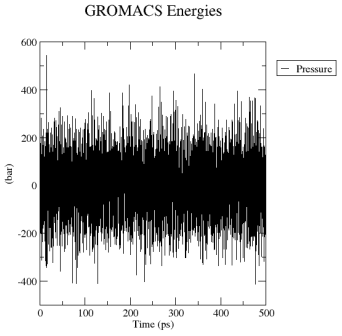
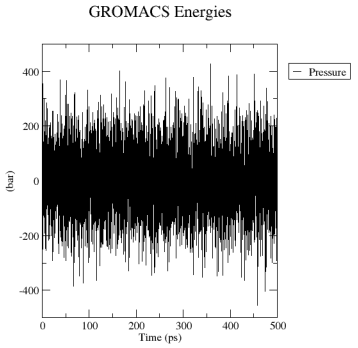
Hydrogen bonds in protein-solvent pairs	**Plot**		
Salt bridges	**Distance between N and C terminus**	 **Mean:** 5.170638	 **Mean:** 5.178435

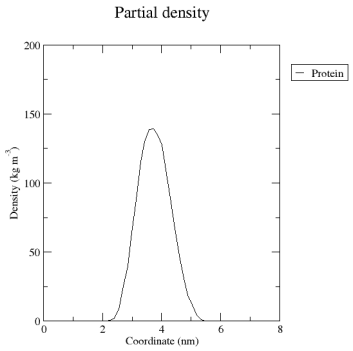
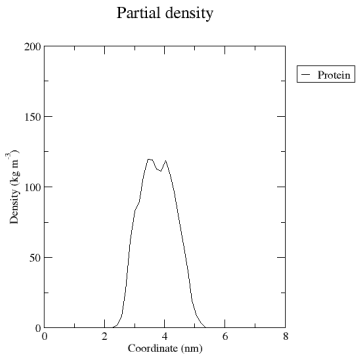
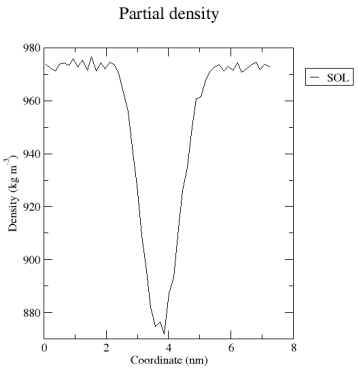
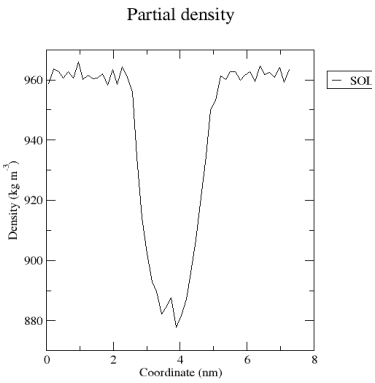
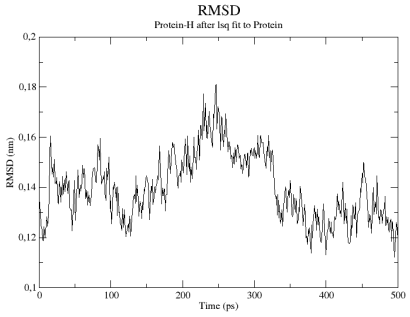
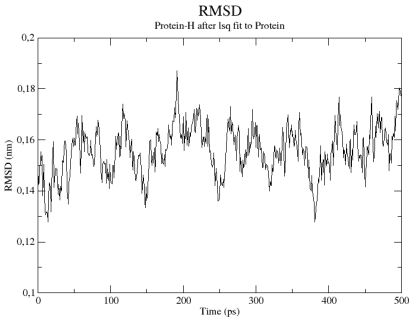
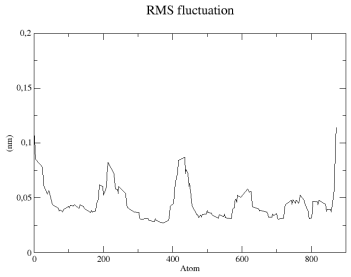
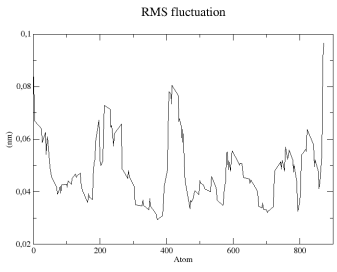
When analysing the above table it may be easily noted that all the properties excluding hydrogen bonds, radius of gyration and distance between N-terminus and C-terminus are significantly influenced by the coordinates changes. It especially concerns energies, densities, RMSF, RMSD and Ramachandran plot indicating that even the slight movements of the atom positions have tremendous impact on the quality of the structure and basic properties approximation. Such disarrangements affect the prediction of helices and strands what is noticed on the Ramachandran plot, they increase the time needed to reach the equilibrium what is evident on the RMSD plot and change the size of the oscillations and obtained values of the energies. All this suggests that MD simulations are very sensitive to the initial PDB structure and the obtained results are strongly correlated with it.

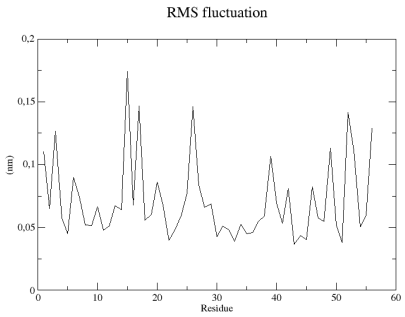
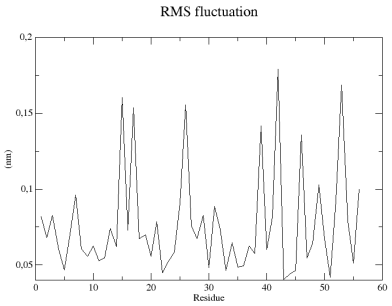
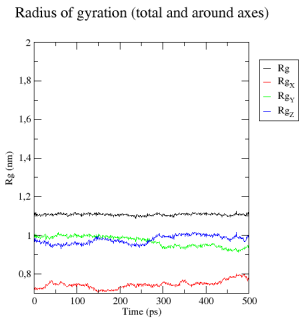
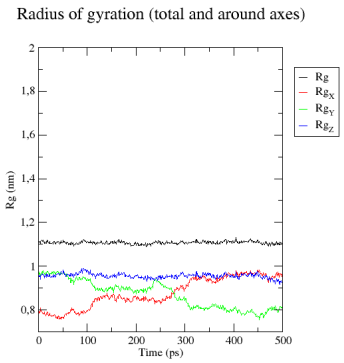
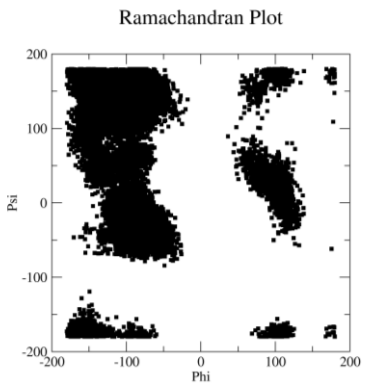
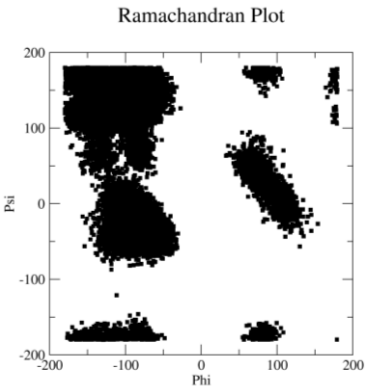
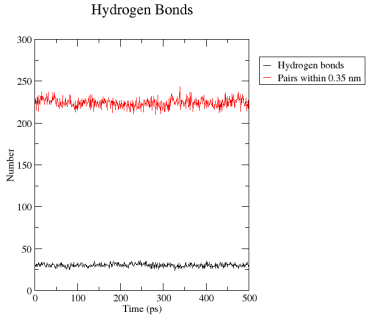
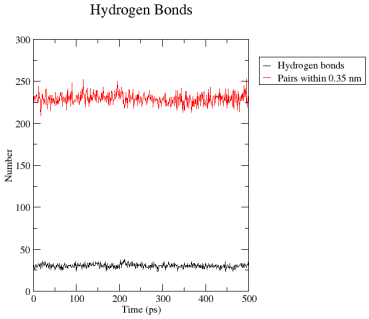
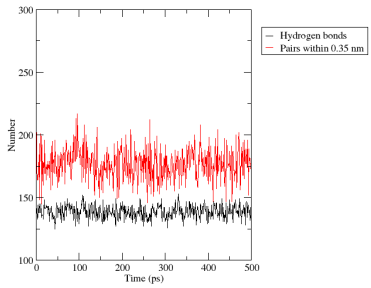
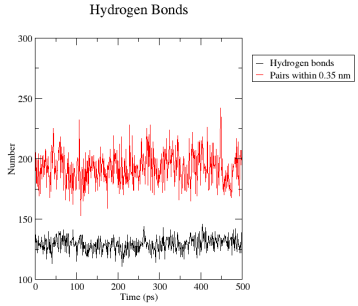
2.5. Force field comparison

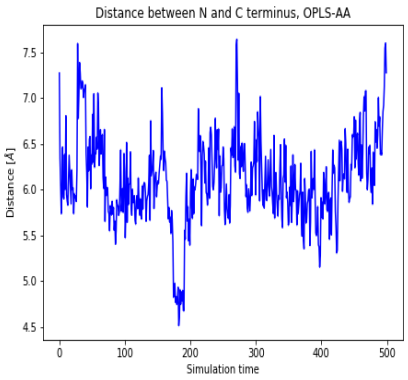
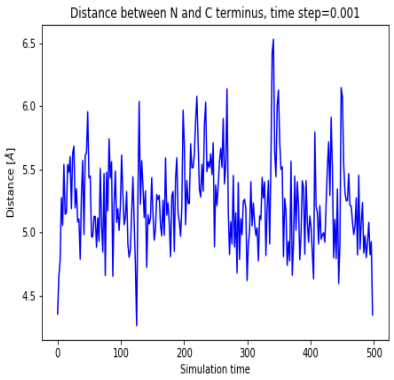
Both OPLS-AA (Optimized Potential for Liquid Simulations – All Atom) and AMBER force field are one of the most frequently used in the simulation of biomolecules. AMBER is designed for simulation of protein and nucleic acids and was originally developed as united-atom force field. Nowadays, the extended version with parameters optimized to fit organic molecules is widely available. The hydrogen bonding values and those for Van der Walls interactions are extracted from crystal structures and lattice energies and the atomic partial charges are calculated using quantum mechanics methods. On the other hand, the OPLS has the aim of simulating organic molecules in liquids. The Van der Walls constants are optimized using experimental liquid properties, mainly enthalpy of vaporizations and densities and the charges are derived not only from the quantum approach but also from experimental condensed-phase properties [13]. In the Table 4. both approaches are compared.

Table 4. Comparison of force fields

		OPLS-AA Computation time = 3h 05	AMBER Computation time = 5h 26
Total energy	**Plot**		
Kinetic energy	**Plot**		
Temperature	**Plot**		
Pressure	**Plot**		

Density of the protein	**Plot**		
Density of the solvent	**Plot**		
RMSD of the protein with respect to the X-ray initial structure	**Plot**		
RMSD with respect to initial X-ray structure	**Value**	0.12325617 nm	0.182058 nm
RMSF of the backbone atoms	**Plot**		

RMSF for each amino acid residue	**Plot**		
Radius of gyration	**Plot**		
Ramachandran Plot	**Plot**		
Hydrogen bonds in protein-protein pairs	**Plot**		
Hydrogen bonds in protein-solvent pairs	**Plot**		

Salt bridges	**Distance between N and C terminus**	 **Mean:** 6.108800	 **Mean:** 5.178435

When comparing the obtained results, at the first sight it is noted that the Partial Densities have smoother shape indicating more desirable distribution of atoms in the box what is consistent with the OPLS-AA advantage of better resemblance of the experimental data. At the same time on the basis of RMSD plot it may be concluded that system is less equilibrated than the one created from AMBER and experiences significantly higher oscillations of the kinetic and total energy, pressure and temperature. Moreover, the structure prediction is worse what is demonstrated on the Ramachandran plot. Similarly, the distance between N-terminus and C-terminus remarkably raises.

Taking all this into account, the choice of the force field should depend on the aim of performed studies. If one applies a more general approach the AMBER shall be used since it provides a more equilibrated system but for comparison of MD with data the faster OPLS-AA is better.

2.6. Conclusions

In this report different MD simulation parameters and their impact on the obtained results was analysed. It was shown that the usage of the proper time step is one of the most problematic issues since the decrease in the interval size dramatically increases the computation time and the raise of its value influences highly on the quality of the information provided by the simulation. Nevertheless, the modern ordinary computers do not supply an efficient enough software for the accurate simulations thus the time step of 0.001 is the best compromise. Analogues study was performed for the cutoff values though here the discrepancies were not so large. At this point it should be emphasized that this conclusion should be interpreted in terms of the calculated properties and more extensive studies revealed that the bigger the cutoff the better the results. Nevertheless, the provided analysis did not show a clear relation between the cutoff and the time needed. Therefore, according to the mentioned papers the value of 1.2 shall be used though the bigger it is the less number of artefacts is present. Also some changes in the atomic positions in PDB file were applied and the simulation comparison was done. This approach demonstrated that event slight discrepancies like the ones done affect a lot on the overall system behaviour indicating that the correct PDB is the basis of the successful MD. Finally, AMBER and OPLS-AA results were contrasted and the advantage of OPLS-AA on the calculation of experimental properties was noticed though the quality of MD was proven to be better for AMBER force field.

Finally, one should always have in mind that the choice of the simulation parameters highly depends on the analysis that is to be conducted and its aims. If one wants to have a general information about the system behaviour or quickly check if some hypothesis have a chance to be true and at the same time does not focus too much on the accuracy of the results then the selection of OPLS-AA, time step of 0.002 and the cutoff 0.8 will guarantee a fast simulation with the desirable though rough approximation. However, such an approach should not be done for even a

bit more demanding examination and in such a case AMBER, time step of 0.001 and cutoff of 1.2 will be the best choice for the majority of mainstream purposes. On the other hand, if there is a need of carrying out a profound, rigorous and professional investigation then the lower the time step and the higher the cutoff (maybe 0.0001 and 1.5 respectively would be appropriate) shall be chosen. Different combination of force fields shall be used in order to verify which is the best for the considered problem. All this confirms that MD simulation ought to be study-specific.

2.7. References

[1] Protein data bank. http://www.rcsb.org/pdb/

[2] Bocedi A. Bolognesi M. Spallarossa A. Coletta M. De Cristofaro R. Menegatti E. Ascenzi, P. The bovine basic pancreatic trypsin inhibitor (kunitz inhibitor): a milestone protein. Curr. Protein Peptide Sci., 4:231–251, 2003.

[3] J. Deisenhofer W. Bode M. Marquart, J. Walter and R. Huber. The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors. 389:480–490, 1983.

[4] Lawrence J. Berliner Jacques Reuben, Biological Magnetic Resonance: Volume 4, Chapter 3, Springer Science & Business Media 2013

[5] Pullman, A., Jortner, Joshua. Modelling of Biomolecular Structures and Mechanisms, pages 200-220 Kluwer Academic Publisher (1994)

http://theory.rutgers.edu/resources/pdfs/The%20Effect%20of%20Hydrostatic%20Pressure%20on%20Protein%20Crystals.pdf

[6] R.M Brunne K.D Berndt P.Guntert K.Wuthrich and W.F von Gunsteren Structure and

Internal Dynamics of Bovine Pancreatic Trypsin Inhibitor in Aqueous Solution from Long-Time Molecular dynamics Simulation, PROTEINS: Structure, Function and Genetics, 23:49-62 (1995)

[7] S.T. Wlodek,' J. Antosiewicz and J.A. McCammon Prediction of titration properties of structures of a protein derived from molecular dynamics trajectories, Protein Science 6:373-382 (1997)

[8] David van der Spoel, Aldert R. van Buuren, D. Peter Tieleman and Herman J.C. Berendsen, Molecular dynamics simulations of peptides from BPTI: A closer look at amide-aromatic interactions, Journal of Biomolecular NMR, 8:229-238 (1996)

[9] Douglas B. Kitchen,*Lynne H. Reed, and Ronald M. Levy' Molecular Dynamics Simulation of Solvated Protein at High Pressure, Biochemistry 31:10083-10093 (1992)
https://ronlevygroup.cst.temple.edu/media/pdf/publications/1992.Biochem.31.10083-93.pdf

[10] Sangrak Kim¸Issues on the Choice of a Proper Time Step in Molecular Dynamics, Physics Procedia 53: 60 – 62 (2014)

[11] Stefano Piana, Kresten Lindorff-Larsen, Robert M. Dirks, John K. Salmon, 1 Ron O. Dror, 1 and David E. Shaw Evaluating the Effects of Cutoffs and Treatment of Long-range Electrostatics in Protein Folding Simulations, PLoS One. 2012; 7(6)

[12] H.Schreiber, O.Steinhauser Cutoff Size Does Strongly Influenece Molecular Dynamics Results on Solvated Polypeptides, Biochemistry 31:5856-5860 (1992)

[13] Augustinus J. M. Sweere and Johannes G. E. M. Fraaije, Accuracy Test of the OPLS-AA Force Field for Calculating Free Energies of Mixing and Comparison with PAC-MAC, J. Chem. Theory Comput.13:5 (2017)