

# Analysis of molecular dynamics simulations of PQQ-GDH in interaction with carbon nanotube

---

**Buljac, Lucia**

**Master's thesis / Diplomski rad**

**2024**

*Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj:* **University of Split, Faculty of Science / Sveučilište u Splitu, Prirodoslovno-matematički fakultet**

*Permanent link / Trajna poveznica:* <https://urn.nsk.hr/urn:nbn:hr:166:044828>

*Rights / Prava:* [In copyright](#) / [Zaštićeno autorskim pravom.](#)

*Download date / Datum preuzimanja:* **2024-11-29**

*Repository / Repozitorij:*

[Repository of Faculty of Science](#)



UNIVERSITY OF SPLIT



DIGITALNI AKADEMSKI ARHIVI I REPOZITORIJ

University of Split  
Faculty of Science

**Analysis of molecular dynamics simulations of  
PQQ-GDH in interaction with carbon nanotube**

Master thesis

Lucia Buljac

Split, July 2024.

## Temeljna dokumentacijska kartica

Sveučilište u Splitu  
Prirodoslovno–matematički fakultet  
Odjel za fiziku  
Ruđera Boškovića 33, 21000 Split, Hrvatska

Diplomski rad

### Analiza molekularne dinamike PQQ-GDH proteina u interakciji s ugljikovom nanocjevčicom

Lucia Buljac

Sveučilišni diplomski studij Fizika; smjer: Biofizika

#### Sažetak:

Interes za enzimatske biogorivne ćelije raste proteklih godina, zbog njihovih bioloških komponenti koje se suprotstavljaju gorivnim ćelijama na bazi plemenitih metala i legura. Topljiva glukoza dehidrogenaza, ili enzim PQQ-GDH, pokazao je obećavajuće pokušaje primjene u biosenzorima i bioenergetici, zbog svoje visoke katalitičke aktivnosti neovisne o kisiku. Koristeći metode i alate programa VMD, provedena je analiza simulacija molekularne dinamike monomera proteina PQQ-GDH u kompleksu s kratkom i beskonačno dugom ugljikovom nanocjevčicom (CNT). U svrhu ispitivanja stabilnosti i snage interakcija između monomera i nanocjevčice, izračunate su udaljenosti između određenih aminokiselinskih ostataka i stijenke CNT-a. U svrhu ispitivanja stabilnosti i očuvanja funkcije proteina, izračunate su standardne devijacije i radijusi giracije monomera sa i bez nanocjevčice. Na kraju, uzimajući u obzir rezultate analize simulacija, konstruiran je teorijski prototip enzimatskog biosenzora koji sadrži PQQ-GDH enzim u kompleksu s kratkom CNT.

**Ključne riječi:** EBFC, monomer, ugljikova nanocjevčica, simulacije molekularne dinamike, VMD

**Rad sadrži:** 34 stranice, 22 slike, 2 tablice, 30 literaturnih navoda. Izvornik je na engleskom jeziku.

**Mentor:** doc. dr. sc. Željka Sanader Maršić

**Neposredni voditelj:** dr. sc. Antonija Mravak

**Ocjenjivači:** doc. dr. sc. Željka Sanader Maršić,  
dr. sc. Antonija Mravak,  
doc. dr. sc. Martina Požar

**Rad prihvaćen:** 30. 9. 2024.

Rad je pohranjen u Knjižnici Prirodoslovno–matematičkog fakulteta, Sveučilišta u Splitu.

<b>Basic documentation card</b>
---------------------------------

University of Split  
Faculty of Science  
Department of Physics  
Ruđera Boškovića 33, 21000 Split, Croatia

Master thesis

**Analysis of molecular dynamics simulations of PQQ-GDH in interaction with carbon nanotube**

Lucia Buljac

University graduate study Physics, specialization in Biophysics

**Abstract:**

Interest for enzymatic biofuel cells grows in the past years, due to their biological components which counteract noble-metal and alloy-based fuel cells. Soluble glucose dehydrogenase, or PQQ-GDH enzyme, showed exciting possibility for application in biosensing and bioenergetics, due to their high catalytic activity which is oxygen independent. Analysis of molecular dynamics simulations of the monomer of protein PQQ-GDH in complex with short and infinitely long carbon nanotube (CNT) has been carried out, using methods and tools of VMD program. To examine the stability and the strength of protein nanotube interactions, distances between certain residues and CNT wall were calculated. To examine protein stability and function preservation, root-mean-square deviation and radius of gyration of monomers with and without nanotube were computed. Finally, considering results from simulation analysis, a theoretical prototype of the enzymatic biosensor comprising PQQ-GDH enzyme in complex with a short CNT was constructed.

**Keywords:** EBFC, monomer, carbon nanotube, molecular dynamics simulations, VMD

**Thesis consists of:** 34 pages, 22 figures, 2 tables, 30 references. Original language: English.

**Supervisor:** Assist. Prof. Dr. Željka Sanader Maršić

**Leader:** Dr. Antonija Mravak

**Reviewers:** Assist. Prof. Dr. Željka Sanader Maršić,  
Dr. Antonija Mravak,  
Assist. Prof. Dr. Martina Požar

**Thesis accepted:** September 30th, 2024

Thesis is deposited in the library of the Faculty of Science, University of Split.



I would like to thank my mentor, professor Željka Sanader Maršić, for her patience, advice and guidance during the writing of this thesis, and for the support shown throughout my studies. Her support and trust encouraged me to continue showing my best version. I also want to thank professor Antonija Mravak for her support and guidance. I would also like to thank professor Emmanuelle Bignon for her hard work on NAMD protein simulations and for sending them.

And finally, I would like to thank my loved ones, my parents, my sister and my boyfriend, for their undeniable patience and support in the moments when I needed it the most.

# Contents

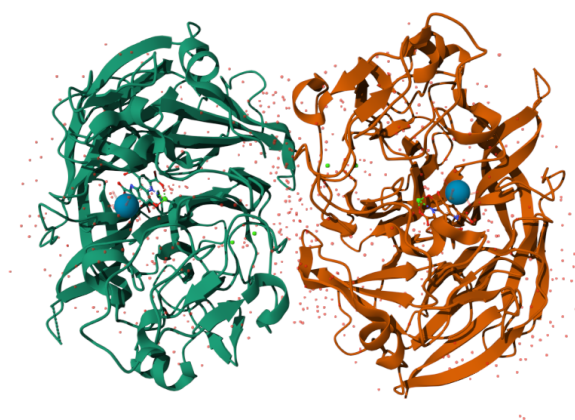
<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Enzyme PQQ-GDH in complex with carbon nanotube	1
1.2	Application in bioelectrochemistry	3
<b>2</b>	<b>Methods and Tools</b>	<b>6</b>
<b>3</b>	<b>Results and discussion</b>	<b>9</b>
3.1	Simulation systems	9
3.2	Protein interaction with carbon nanotube	10
3.3	Carbon nanotube residues distance analysis	15
3.4	RMSD analysis	20
3.5	Radius of gyration analysis	26
3.6	Electron hopping process in protein	27
<b>4</b>	<b>Conclusion</b>	<b>29</b>
<b>A</b>	<b>Python coding scripts</b>	<b>34</b>

# 1 Introduction

## 1.1 Enzyme PQQ-GDH in complex with carbon nanotube

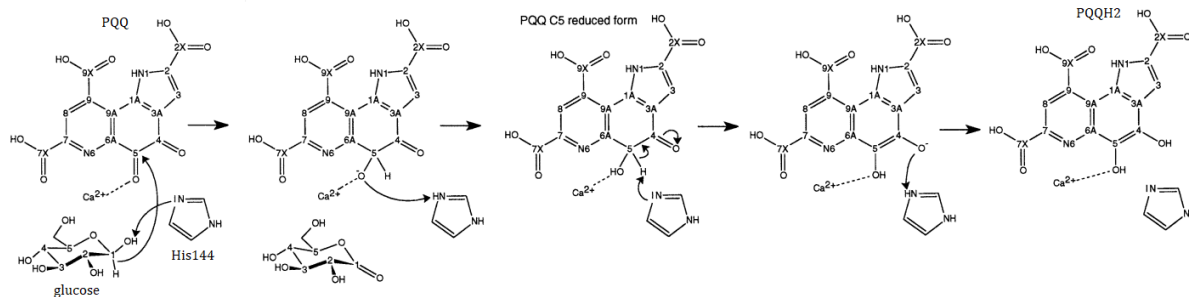
Soluble glucose dehydrogenase (s-GDH) is a classical quinoprotein which requires the cofactor pyrroloquinoline quinone (PQQ) to oxidize glucose (beta-D-glucopyranose, BGC) to gluconolactone. Former enzyme has been identified in *Acinetobacter calcoaceticus*, Gram-negative bacteria. X-ray structures of s-GDH reveal its active site and explain how a functionally bound substrate interacts with the cofactor in an enzyme. s-GDH is a dimeric enzyme of identical subunits, with one monomer (50 kDa, 454 residues) binding one PQQ molecule and three calcium ions. One of the calcium ions is required for activation of the cofactor, the other two are needed for functional dimerization of the protein. Each monomer has a  $\beta$ -propeller (superbarrel) fold composed of six four-stranded anti-parallel  $\beta$ -sheets. These 24  $\beta$ -strands form the structural core of the enzyme. [1] Three-dimensional conformation of enzyme in aqueous solution is shown in Figure 1.

There are three important molecules which participate in a chemical reaction of s-GDH enzyme; cofactor PQQ, molecule of glucose and histidine, His144, active site residue which catalyzes the reaction. Oubrie et al. suggest reaction mechanism comprising general base-catalyzed proton abstraction in concert with direct hydride transfer from substrate to PQQ, and ending with tautomerization to PQQH<sub>2</sub>. The first step of reaction is His144-catalyzed proton abstraction from O1 glucose atom and reduction process of the cofactor. The last part of the reaction ends with reduced cofactor, PQQH<sub>2</sub>, and protonated histidine on one nitrogen atom, as in the beginning of the reaction. Scheme of reaction mechanism is depicted in Figure 2.



**Figure 1:** Enzyme PQQ-GDH in complex with cofactor PQQH<sub>2</sub> and glucose (blue ball), surrounded with water molecules (red balls) and calcium ions (green balls). [3].

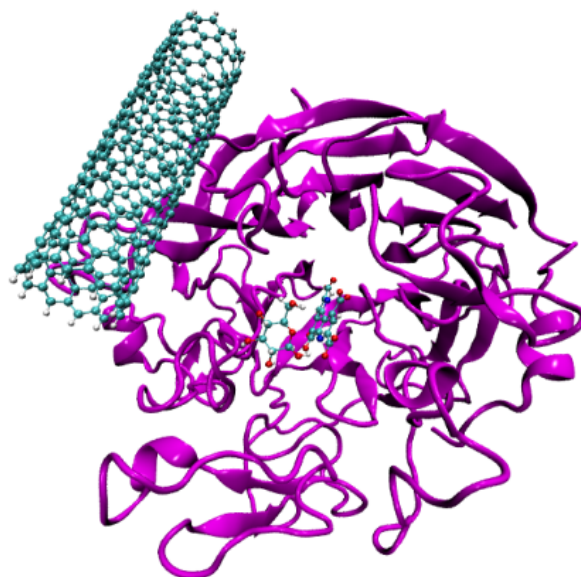
Enzymes from the oxidoreductase class catalyze reactions that involve electron transfer from



**Figure 2:** Reaction mechanism of catalyzed hydride transfer from glucose to cofactor and tautomerization of cofactor to PQQH<sub>2</sub>. [1].

one molecule, the electron donor, to another molecule, the electron acceptor. The source of electrons are small biomolecules such as glucose. With the goal of replacing metal electrodes with environmentally friendly products, these enzymes offer a solution and are a key ingredient in enzymatic biosensors and biofuel cells. Enzymatic biofuel cells (EBFCs) are enzyme based fuel cells that generate electrical current through electrochemical process. Diverse nanomaterials are used as electrodes. Nanomaterials based on carbon gave promising results, especially carbon nanotubes (CNTs). Their unique structure and nanoscale dimensions mean that CNTs possess many properties that are of great benefit for chemical and biological sensing applications. These properties include a very large surface area to volume ratios and large aspect ratios, excellent electrical conductivities and a high chemical stability. [2]

CNTs have a one-dimensional cylindrical shape with nanometre scale diameters and micrometer scale length. Structurally, CNTs can be thought of as a rolled-up sheet of graphene. CNTs have a very high melting point and a high tensile strength as each carbon atom is joined to three other carbon atoms by strong covalent bonds. This also leaves each carbon atom with a spare electron, which forms a sea of delocalised electrons within the tube, allowing for good electrical properties, and redox reactions with fast electron-transfer rates. Importantly, CNTs exhibit good biocompatibility, allowing for integration with biological components, such as enzymes. [2] The high surface area offered by CNTs comes useful for enzyme support, advantages include increased enzyme stability and possibility to recycle the enzymes. If the enzyme catalytic site is involved in CNT interaction, there will be a dramatic or even total loss of substrate conversion into product. [4] The type of CNT used in this thesis is single-walled carbon nanotube (SWCNT). The Figure 3 depicts one of the simulation systems, which was analyzed further in this work, presenting a monomer subunit of PQQ-GDH enzyme in complex with SWCNT. CNT is not covalently bounded to enzyme, and moves freely around the enzyme during simulation time.

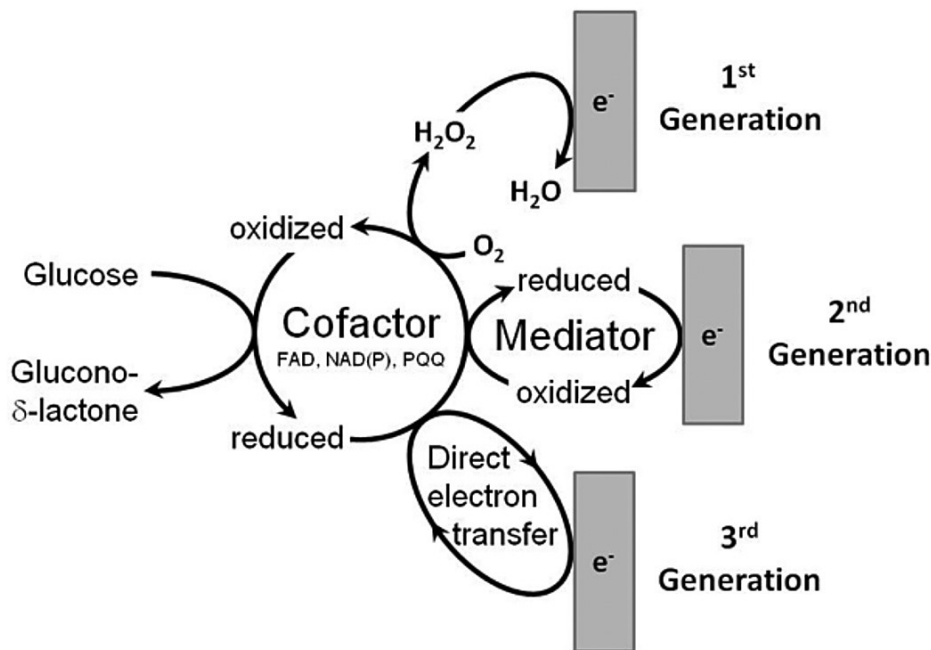


**Figure 3:** Simulation replica of the monomer of PQQ-GDH enzyme in complex with single-walled carbon nanotube. Enzyme is presented with magenta ribbon, a short cyan cylinder represents carbon nanotube and inside the enzyme active pocket resides the substrate molecule glucose and cofactor molecule PQQ.

## 1.2 Application in bioelectrochemistry

Because of the high catalytic activity and the oxygen independent function of PQQ-GDH, the soluble enzyme has found considerable application in biosensing and bioenergetics. [5] The first example of an enzymatic biofuel cell using enzymes as catalysts was proposed in 1964. The principle of EBFCs is similar to that of classic fuel cells, which is based on a catalytic fuel-oxidizing anode and a catalytic oxidizer-reducing cathode. The difference lies in the nature of the catalysts used, which are of biological origin in the case of biofuel cells, contrary to abiotic fuel cells where principally noble-metal-based catalysts or alloys are used. Compared with conventional fuel cells, EBFCs are safe due to the enzyme reactions that can operate under mild conditions such as room temperature, atmospheric pressure, and neutral pH. Additionally, EBFCs can be used for several biologically related reductants (fuels) as electron donors such as sugars, alcohols, amines, organic acids, and hydrogen at the anode side. Extensive studies have been carried out to develop improved enzyme-based systems for monitoring glycemic levels. First-generation blood glucose monitoring systems employed oxygen as the electron acceptor, determining glucose concentration by following either the consumption of oxygen or the liberation of hydrogen peroxide. In second generation sensors, enzymes transfer electrons to artificial electron acceptors (also referred to as electron mediators or redox dyes) instead of oxygen to avoid interference from other redox species. Third-generation sensors employ direct electron transfer (DET) to the electrode, thus eliminating toxic artificial electron mediators and avoiding errors due to variations in the concentration of oxygen in blood samples. [7] Schematic representation of the principles of

first-, second-, and third-generation glucose sensors is illustrated in Figure 4. DET reactions between proteins (enzymes) and electrodes have been extensively studied from the viewpoints of both understanding the fundamental features and for applications as biosensors and biofuel cells. Various approaches were developed for the DET reactions of redox proteins at early stages, but remained at the fundamental level. [6]



**Figure 4:** Schematic representation of the principles of first-, second-, and third-generation glucose sensors. Electrons from the glucose oxidation reaction are first taken up by the enzyme's cofactor (primary electron acceptor) and transferred to either oxygen (first generation), an electron mediator (second generation), or directly to the electrode (third generation). [7]

Electron transfer (ET) is a fundamental process which plays a central role in physics, chemistry, and biology. Biological electron transfer reactions are required for respiration, photosynthesis, and redox reactions of intermediary metabolism. Unlike typical chemical reactions, electron transfer reactions do not involve the making or breaking of bonds. [8] In 1992, Rudolph A. Marcus was awarded with the Nobel Prize in Chemistry for his contribution to the development of the ET theory in chemical systems. In an ET reaction, we first define the electron donor species (D) and the electron accepting ones (A). To enhance the coupling probability of their electronic orbitals, D and A (reactants) should be as close as possible. In biological systems, the dependence of the ET rate on the distance between D and A has been widely elucidated both theoretically and experimentally. In this regard, the ET theory for biological systems, named afterwards “Marcus Theory”, is able to predict the ET rate constant value as given by the equation:

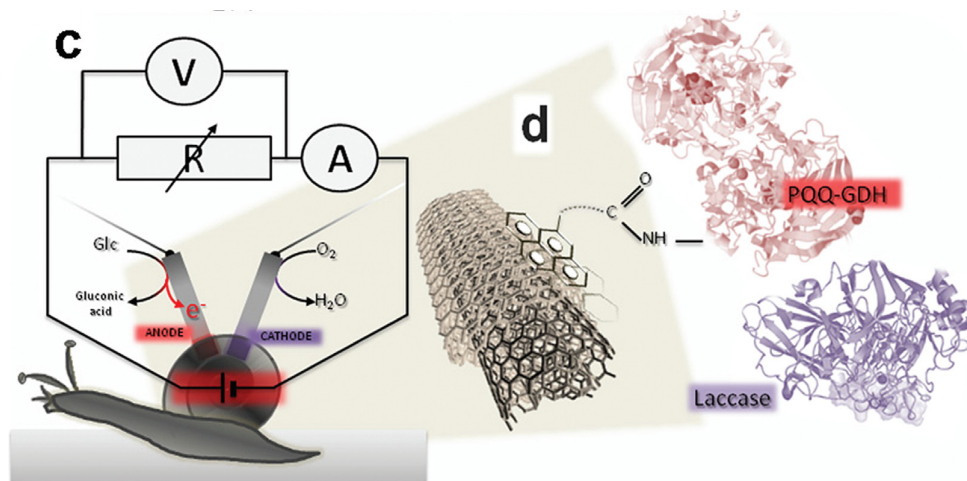
$$k_{ET} \propto e^{-\beta(d-d_0)} \cdot e^{-\frac{(\Delta G^0 + \lambda)^2}{4RT\lambda}}, \quad (1.1)$$

where  $\beta$  corresponds to the decay or attenuation factor (about  $10 \text{ nm}^{-1}$  for proteins),  $\Delta G^0$  and

$\lambda$  correspond to the free Gibbs energy and reorganization energy accompanying the ET process.  $d_0$  and  $d$  are the Van der Waals distance and actual distance between redox active site, while  $R$  and  $T$  are the gas constant and the absolute temperature. [9]

The electron transfer rate is exponentially dependent on the distance between the redox-active centers, as predicted by the Marcus theory, presented and explained by the above equation (1.1). Therefore, in order to shorten the electron transfer distance between the active center of the enzyme and the electrode, surface functionalization of an electrode might be an essential approach. When no DET can be achieved, redox molecules with appropriate redox kinetics and diffusion capacities are necessary to shuttle the electrons via mediated electron transfer (MET). A mediator can be organic molecules or metal complexes, which may have access to a redox- active center located deeply inside the protein molecule. In general, MET-based EBFCs show higher electric power density in comparison with DET-based EBFCs. Highly porous conductive nanostructures became the material of choice for efficient wiring of enzymes. Carbon nanotubes (CNTs) have very favored properties as electrode material for electron transfer with enzymes due to their nanowire structures enabling naturally close contact to the redox centers of some enzymes. Quinones are becoming more popular as mediators in bioanodes. Therefore, quinones are excellent mediators for other glucose-oxidizing enzymes such as glucose dehydrogenases (GDH). The advantage of GDHs over glucose oxidases is that no oxygen is reduced to hydrogen peroxide, which is generally considered a negative effect. The regeneration of these enzymes is assured by mediators, which can especially be designed for biofuel cell applications. Various experiment incorporating functionalized CNT layers with PQQ-GDH enzyme showed successful DET and MET reactions. [6]

For the DET approach, it has been found that among different CNTs, particularly, SWCNTs provide better conditions for the enzyme interaction. It has also been demonstrated that GDH-based biofuel cells operate in living organisms into which a complete biofuel cell has been inserted. Figure 5 illustrates the structure of the biofuel cell used for insertion of the EBFC inside a snail. Biofuel cell based on laccase enzyme and s-GDH with CNTs as electrodes is inserted into a snail. Rather early PQQ-GDH has been considered as an ideal enzyme for glucose detection mainly because of its oxygen independent catalytic reaction and the high turnover rate. Early blood sugar sensors have been developed which used the enzyme in conjunction with a mediator. PQQ-GDH has turned out to be a favourable enzyme for biofuel cell application, since here, a high turnover is required to convert efficiently chemical energy into electrical energy. [5]



**Figure 5:** Biofuel cell based on laccase enzyme and s-GDH, and CNTs as electrode material, inserted into a snail. On the anode side, glucose is converted to gluconic acid and on the cathode side, oxygen molecule is used as an electron acceptor. [5]

## 2 Methods and Tools

The visualization program VMD ("Visual Molecular Dynamics"), a program for displaying and analyzing biomolecular systems, was used for molecular dynamics analysis in this thesis. The program uses 3-D graphics and built-in scripts. In addition to the visualization of molecules using different methods, the program enables animation and analysis of molecular dynamics simulation trajectories. [10]

The root-mean-square deviation, abbreviated RMSD, is a useful measure to monitor enzyme stability during simulation. The RMSD built-in *RMSD Trajectory Tool* calculates the mean deviation of the selected structure, for each time frame of the simulation, from some reference configuration, that is, from the structure at the initial time frame. It is accessed via the following selection in the VMD main window: *Extensions* → *Analysis*. The RMSD value is calculated according to the following formula:

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N (\vec{r}_i(t_1) - \vec{r}_i(t_2))^2}, \quad (2.1)$$

where  $N$  is the number of specified atoms and  $\vec{r}_i$  denotes their positions. RMSD is the mean distance for each simulation time frame,  $t_2$ , relative to the initial time frame,  $t_1$ . In molecular dynamics simulation, the system is relaxed from the initial configuration, and the fluctuations of the RMSD values show the relaxation path to the most stable configuration. Each RMSD-graph describes the behavior of the system and the equilibration process of the protein in the simulation system. [11] RMSD values provide information on the stability of the structure. Smaller values indicate greater stability and smaller deviations from the reference



structure. [13]

In the *RMSD Trajectory Tool* text entry, you can specify a group of atoms on which you want to apply the RMSD calculation, for example, the keyword *"protein"* selects the RMSD calculation of all protein atoms, while the selection *"resid 400 to 500"* selects to calculate the RMSD of only the residues with ordinals within the specified range. The *Top* option indicates the calculation of the RMSD of the molecule that was last loaded into the main window of the program. The *Align* option applies rigid-body translations and rotations to the protein at each time frame, aligning it with the configuration of the reference, zero time frame. [11] Clicking on the *RMSD* button calculates the RMSD value in Angstroms for each time frame of the simulation. The *Plot* option enables the plot display of the RMSD values, in each time frame. The following selection: *File* → *Save Data* allows saving the data for generating the graph afterwards in the Gnuplot program or similar program.

In order to calculate the RMSD values between two protein structures in each time frame of the simulation, the script *rmsd.tcl*, shown in Figure 18 in Appendix A, was used. The script is loaded into the VMD-program by typing the command *"source rmsd.tcl"* in the Tk console. The root-mean-square deviation calculation process initiates with the command *"print\_rmsd \$seltext1"*. The script prints an output file, where, all the frames of the simulation are printed in the first column, and the RMSD values in the second column. [12]

Measuring the gyration radius of proteins during simulations is a useful analysis to determine the compactness and stability of the protein structure. *R<sub>g</sub>* measures the distance between the centre of mass and the rotational axis of the ligand-bound protein assembly. Lower values of the radius of gyration indicate higher stability and compactness of the complex, while higher values suggest lower stability and compactness. [13] Radius of gyration of a structure, that is, its square value, is calculated according to the following formula:

$$r_G^2 = \frac{\sum_{i=1}^N m_i (\vec{r}_i - r_{CM})^2}{\sum_{i=1}^N m_i}, \quad (2.2)$$

where  $m_i$  is the mass of each of the total  $N$  atoms, and  $r_{CM}$  is the position of its center of mass, which is defined by the expression [14]:

$$r_{CM} = \frac{\sum_{i=1}^N m_i \vec{r}_i}{\sum_{i=1}^N m_i}. \quad (2.3)$$

To calculate the radius of gyration on the loaded protein structures in the VMD program, three scripts were used; *center\_of\_mass.tcl*, *gyr\_radius.tcl* and *rog\_loop\_dcd.tcl*. These scripts containing Python code are shown in Figures 19, 20 and 21 in the Appendix A. The scripts *center\_of\_mass.tcl* and *gyr\_radius.tcl* calculate the center of mass and radius of gyration, according to the expressions (2.3) and (2.2), respectively, of all specified atoms in the

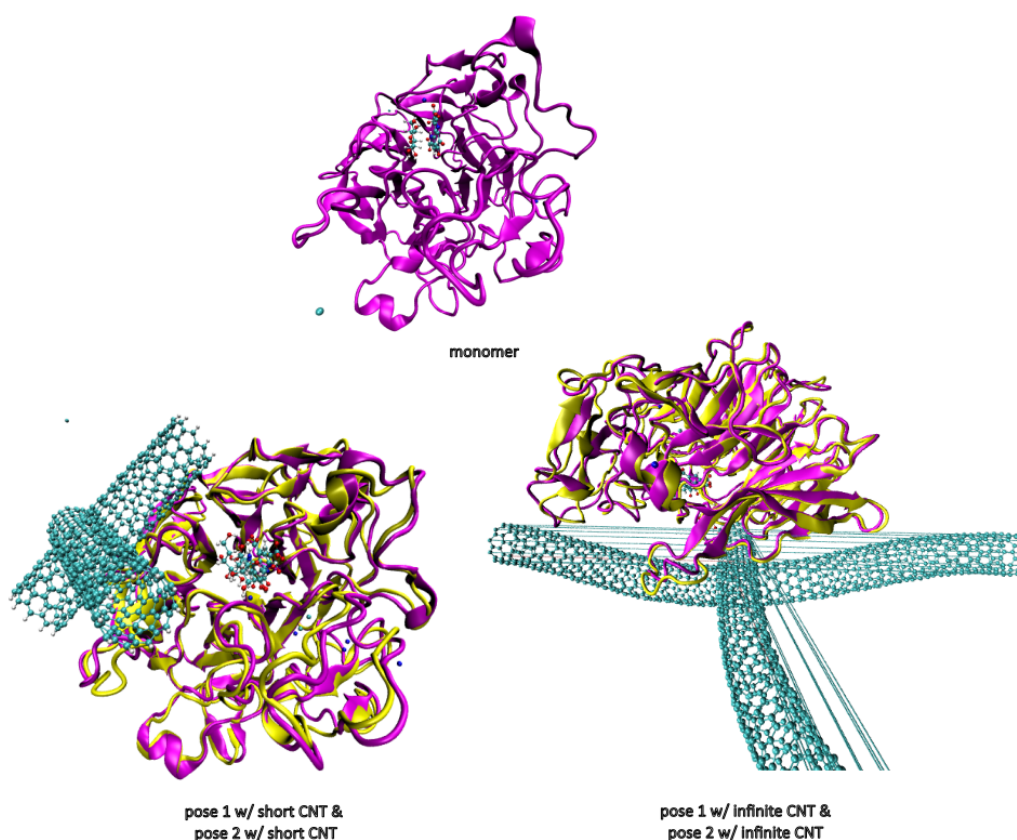
simulation system. The script *rog\_loop\_dcd.tcl* loads the other two scripts and calculates the radius of gyration for each frame of the simulation for the specified group of atoms, in this case, for the group of residues "*resid 1 to 451*", which denote all amino acids of the observed of the PQQ-GDH enzyme. The script is loaded into the VMD-program by typing the command "*source rog\_loop\_dcd.tcl*" in the Tk console. The script prints an output file, where, all the frames of the simulation are printed in the first column, and the values of the radius of gyration in the second column. [14]

To calculate the distance between certain groups of atoms, the script *distance.tcl*, shown in the Figure 22 in the Appendix A, was used. The first part of the script calculates the distance in time, and the second one its distribution. The script is loaded into the VMD-program by typing the command *source distance.tcl* in the Tk console. The distance calculation process is initiated with the command "*distance \$seltext1 \$seltext2 N\_d f\_r\_out f\_d\_out*". Commands "*seltext1*" and "*seltext2*" indicate groups of atoms between which the distance is calculated. *N\_d* indicates the number of bins (class; interval of values from which the distribution is calculated) for generating the distribution, and *f\_r\_out*, *f\_d\_out* indicate the output files in which data are saved, distances in each frame of the simulation and distribution data, respectively.

### 3 Results and discussion

#### 3.1 Simulation systems

In order to describe the behavior of the PQQ-GDH protein on the CNT surface, MD simulations of 5 different systems were performed: monomer, monomer with short CNT (pose 1 and pose 2) and monomer with infinite CNT (pose 1 and pose 2). Simulations were carried out in the program NAMD [16] by collaborator associate professor Emmanuelle Bignon, at Université de Lorraine and CNRS, in Nancy. As MD simulations have different duration times, all graphs are generated for simulation time up to 1200 ns. To obtain statistically independent results, three replicas of MD simulations (MD1, MD2, MD3), with different initial velocity parameters, were made for each initial position. Simulations were made with different positions of CNTs, in order to reject the influence of nanotube's orientation on protein activity. A total of fifteen MD simulations of previously mentioned 5 different systems were analyzed in this thesis (3 simulations for each system). All structures of simulation systems are shown in Figure 6.

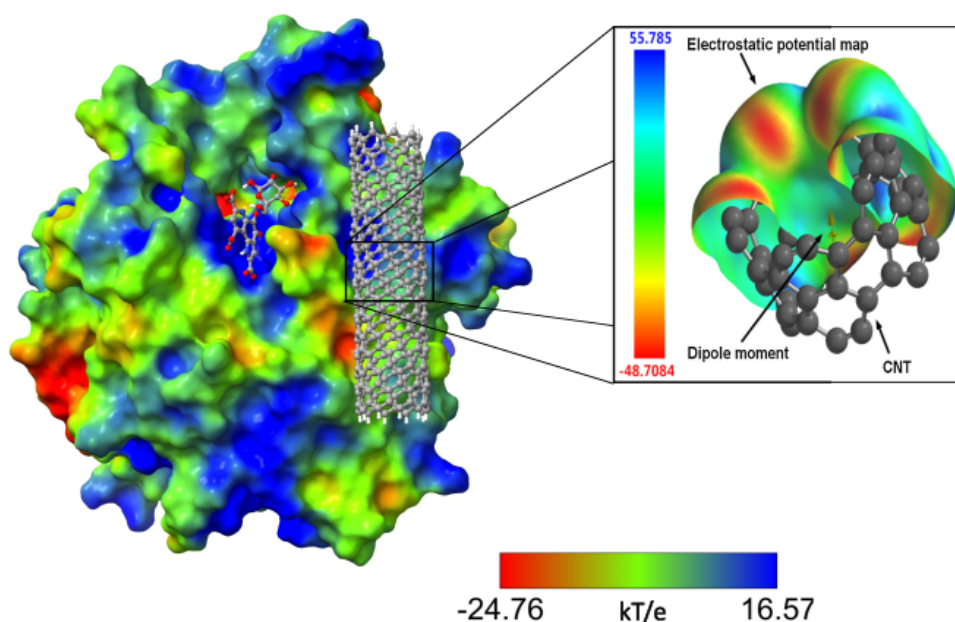


**Figure 6:** Three-dimensional structures of the simulation systems, in order: PQQ-GDH monomer protein with its substrate, molecule of glucose, and its cofactor molecule PQQ, aligned structures of the same monomers in two poses, each with a short carbon nanotube and with an infinite one. Structures were created in the visualization program VMD, using NewCartoon drawing method for proteins and CPK drawing method for BGC, PQQ and CNT structures.

### 3.2 Protein interaction with carbon nanotube

For the analysis of the protein interaction with carbon nanotube, numerical data of certain residues binding free energies with a short CNT, in poses 1 and 2, were used. Of all the residues for which the calculation of the binding free energy was carried out, those with the highest energies were selected ( $\leq -2.5$  kcal/mol for both poses). These calculations gave a qualitative analysis of the interactions of certain residues with the carbon nanotube. They also give rise to future plans for eventual protein mutations at certain residue positions in order to achieve stronger and more stable interactions with the carbon nanotube. For each residue in both poses, side chain atoms contribution to the binding free energy dominates. For the protein monomer PQQ-GDH in pose 1, those are the following residues: Thr47, Val70, Asp72, Pro103, Lys108 and Pro111. For the same protein in pose 2, the selected residues are: Thr47, Val70 and Ala73.

Research done by C. Wang *et al.* [17] confirmed that the H/ $\pi$  interactions between amino acids and CNTs play more important roles than the hydrogen bonds between amino acids in stabilizing the complex structures. In this type of interaction, CNT is a typical H/ $\pi$  acceptor and amino acids are H/ $\pi$  donors. [17] H/ $\pi$  interaction is a typical interaction formed between hydrogen compounds and any  $\pi$ -system (for example, benzene). The nature of this interaction is that of a hydrogen bond between the hydrogen atom and the  $\pi$ -cloud. [18] Research done by Z. He and J. Zhou [22] showed that among the 20 standard amino acids, aromatic amino acids Phe, Tyr and Trp and positively charged amino acid Arg exhibit the strongest affinity for CNTs. Aromatic amino acids interact with the tube via their aromatic rings to form strong  $\pi$ - $\pi$  stacking interactions. Aliphatic (as Ala, Val, Pro and others) interact with CNT by the hydrophobic interaction of their aliphatic side chains, which is directly related to their hydrophobicities. Furthermore, it was found that proline makes C-H/ $\pi$  interactions with  $\pi$  orbitals of aromatic compounds, in this case, the benzene rings forming the shell of the nanotube. Regarding polar amino acids, it was found that Thr47 makes a specific O-H/ $\pi$  interaction with the aromatic rings of the benzene components of the nanotube. O-H/ $\pi$  interactions, together with N-H/ $\pi$  interactions, are exactly 4.2 times weaker than the classic hydrogen bond, while C-H/ $\pi$  interactions are even weaker, even 75 times weaker than the hydrogen bond. [19, 20] Charged amino acids, such as Lys, make so called cation- $\pi$  interactions with aromatic  $\pi$ -cloud. Such interactions are dominated by the electrostatic attraction between an electron-rich aromatic compound and electron-deficient cation. [21] Compared with aromatic amino acids, interactions between aliphatic or polar amino acids and CNT are much weaker. [22]



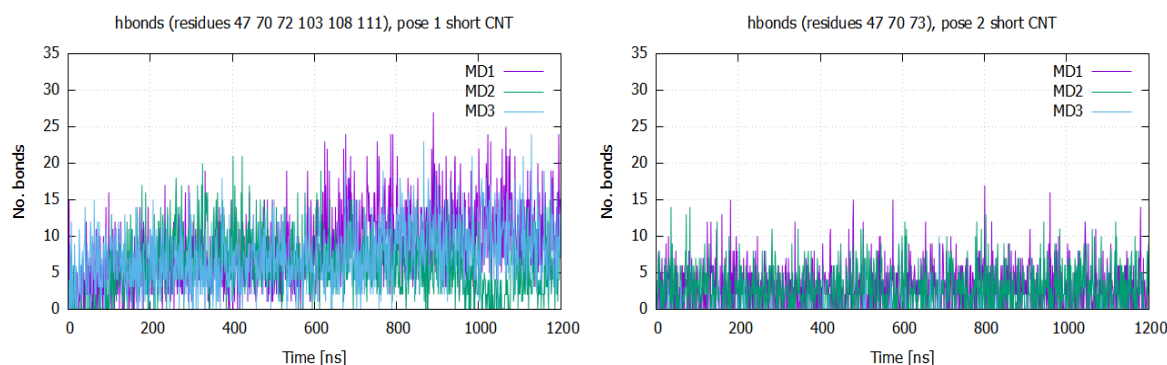
**Figure 7:** Electrostatic potential map of the PQQ-GDH protein monomer in complex with a carbon nanotube. Different regions of the protein are colored according to the displayed potential value scale in the  $kT/e$  unit, where  $k$  is the Boltzmann constant,  $T$  is the temperature of the system and  $e \approx 1.6 \cdot 10^{-19} C$  is the electron charge. The figure next to the surface of the protein shows the potential map for a segment of the carbon nanotube, with an appropriate potential scale, taken from [26].

Figure 7 shows the electrostatic potential map of the PQQ-GDH protein monomer in complex with a carbon nanotube. Red regions represent lower electrostatic potential, thus higher electronic charge density. The map of the protein was created using the next-generation molecular visualization program UCSF ChimeraX [23], using calculated potential map data from the APBS ("Adaptive Poisson-Boltzmann Solver") web server [24]. APBS is a software package for the numerical solution of the Poisson-Boltzmann equation (PBE), one of the most popular continuum models for describing electrostatic interactions between molecular solutes in salty, aqueous media. APBS was designed to efficiently evaluate electrostatic properties for a wide range of molecules. [25] The image next to the protein shows the potential map for a segment of a carbon nanotube, obtained by the density functional theory (DFT) calculations. [26]

To begin with, the calculation of number of hydrogen bonds formed between certain residues and the short carbon nanotube, throughout simulation, is carried out. The *Hydrogen bonds* plugin, accessed by the selection in the main window: *Extensions*  $\rightarrow$  *Analysis*  $\rightarrow$  *Hydrogen Bonds*, counts the number of hydrogen bonds formed throughout a trajectory. The search can be restricted to a single selection or between two distinct selections, as well as a frame range given by the user. The plugin provides options for up to two selections (which should not overlap) as well as the frames over which to calculate. The number of hydrogen bonds vs. time can be plotted immediately within VMD, and/or saved to a file (default hbonds.dat). [10] For the calculation of hydrogen bonds in pose 1, selection "*resid 47 70 72 103 108 111*" is used,

while for calculating bonds in pose 2, selection "*resid 47 70 73*" is used. Donor-acceptor distance cutoff is set to 3.5 Å for pose 1 residues, and 3.7 Å for pose 2 residues. Angle cutoff is set to 120 degrees. It is checked that these cutoff values give optimal results for each residue, regarding the distances of residue hydrogen atoms interacting with CNT carbon atoms in two poses. Figure 8 presents results of this calculation, therefore, change in the number of hydrogen bonds formed between specified residues and the short CNT, during simulation time in each of the two poses. It is evident from the graphs that pose 1 with short CNT has more hydrogen bonds, which is expected due to the higher number of residues having interaction with the nanotube. At the end of the simulation time in pose 1, simulation MD1 has the greatest number of bonds, more than 25 bonds. The same simulation achieves the greatest value in pose 2, more than 15 bonds. In contrast to the graph in pose 1, the graph in pose 2 shows more regular pattern.

Figure 9 shows the corresponding residues interacting with carbon nanotube. Since there is

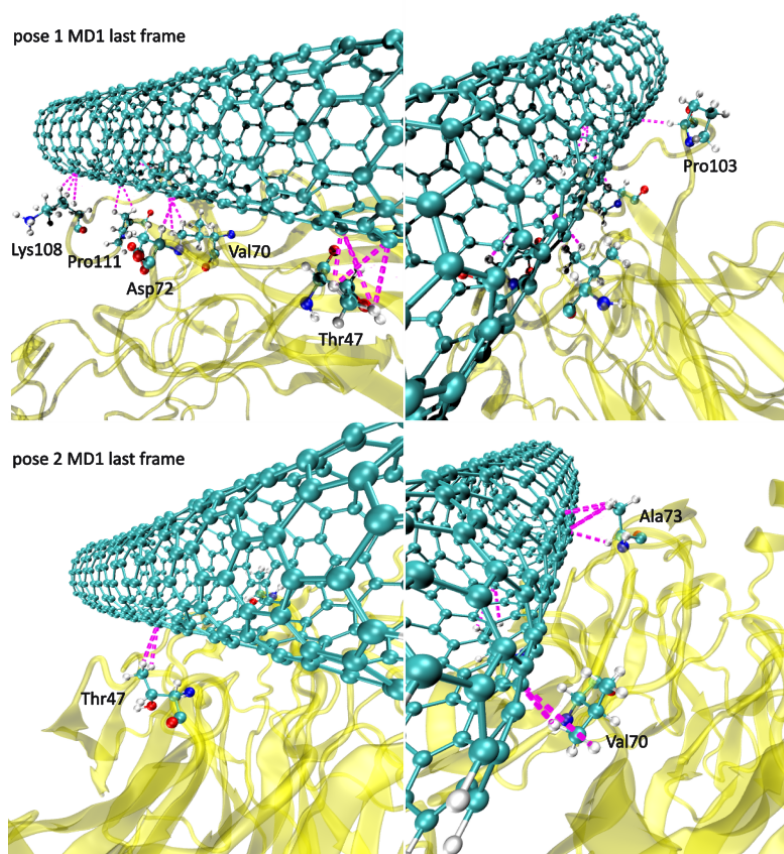


**Figure 8:** Time plots of number of hydrogen bonds, throughout the simulations, between certain residues and the short carbon nanotube in pose 1 and 2. Bonds were calculated using the Hydrogen bonds plugin. Calculations were carried out on residues Thr47, Val70, Asp72, Pro103, Lys108 and Pro111 in pose 1, and on residues Thr47, Val70 and Ala73 in pose 2.

only a VMD tool for finding hydrogen bonds between appropriate group of atoms, it is possible to visualize their interaction. As seen in Figure 9, interaction in the form of hydrogen bonds between hydrogen atoms (hydrogen-donor) of residues and carbon atoms (hydrogen-acceptor) of the nanotube is presented. Bonds are created using the VMD visualization tool *Hbonds*, accessed by the following selection: *Graphics* → *Representation* → *Draw style* → *Drawing method* → *Hbonds*. Bonds between specified atoms are created by typing *resid no. or resname CNT* in the *Selected Atoms* text entry. The cutoff distance for hydrogen bond is set to 3.6 Å for Pro103 and to 3.5 Å for other residues in pose 1. In pose 2, different cutoff distance is set for each residue; 3.7 Å for Thr47, 3.6 Å for Val70 and 3.5 Å for Ala73. The cutoff angle for hydrogen bond is set to 120 degrees for each residue in both poses.

According to Jeffrey, hydrogen bonds can be divided into three categories, depending their strengths. Hydrogen bonds with donor-acceptor distances 2.2-2.5 Å are considered as strong, with mostly covalent character. If donor-acceptor distance ranges between 2.5-3.2 Å, then the

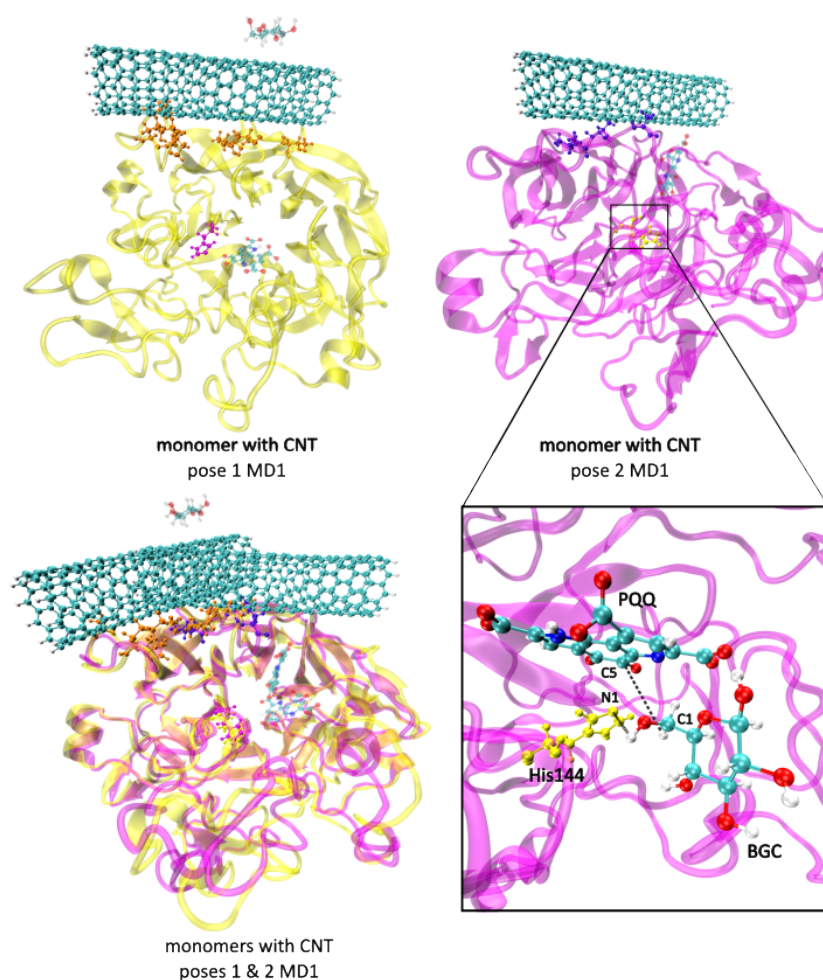




**Figure 9:** Interaction between residues and a short carbon nanotube in the form of hydrogen bonds (magenta dashed lines) in poses 1 and 2 is visualized. For the protein PQQ-GDH with CNT in pose 1, six residues are presented and labeled in this figure: Thr47, Val70, Asp72, Pro103, Lys108 and Pro111. For the pose 2, three residues are presented and labeled: Thr47, Val70 and Ala73.

bond is considered moderate, mostly electrostatic. However, if the distance ranges between 3.2-4.0 Å, bond is considered weak and electrostatic. [27] According to the cutoff distances between 3.5-3.7 Å used for the generation of graphs in Figure 8 and for illustrating the interactions in Figure 9, it is evident that interactions between the residues and the short CNT in both poses are weak electrostatic forces.

The difference between the protein with a short CNT in pose 1 and pose 2 is displayed in Figure 10. It is evident that CNT in pose 1 interacts with more residues than in pose 2. Figure 10 also shows His144, the active site residue, which is highlighted in magenta and yellow in each pose. The fourth image in Figure 10 shows magnified display of the active site, consisting of residue His144, presented with yellow balls, and molecules of glucose BGC and the cofactor PQQ. As shown in the figure, the residues that interact with CNT are far enough from the residues of the active site of the protein, therefore, different positions of the nanotube do not affect the residues essential for the catalytic activity of the protein. A more detailed analysis of the difference between pose 1 and pose 2, regarding protein and CNT interaction, will be carried out in the next subsection, by analyzing the distance of individual residues from the short CNT in each pose.



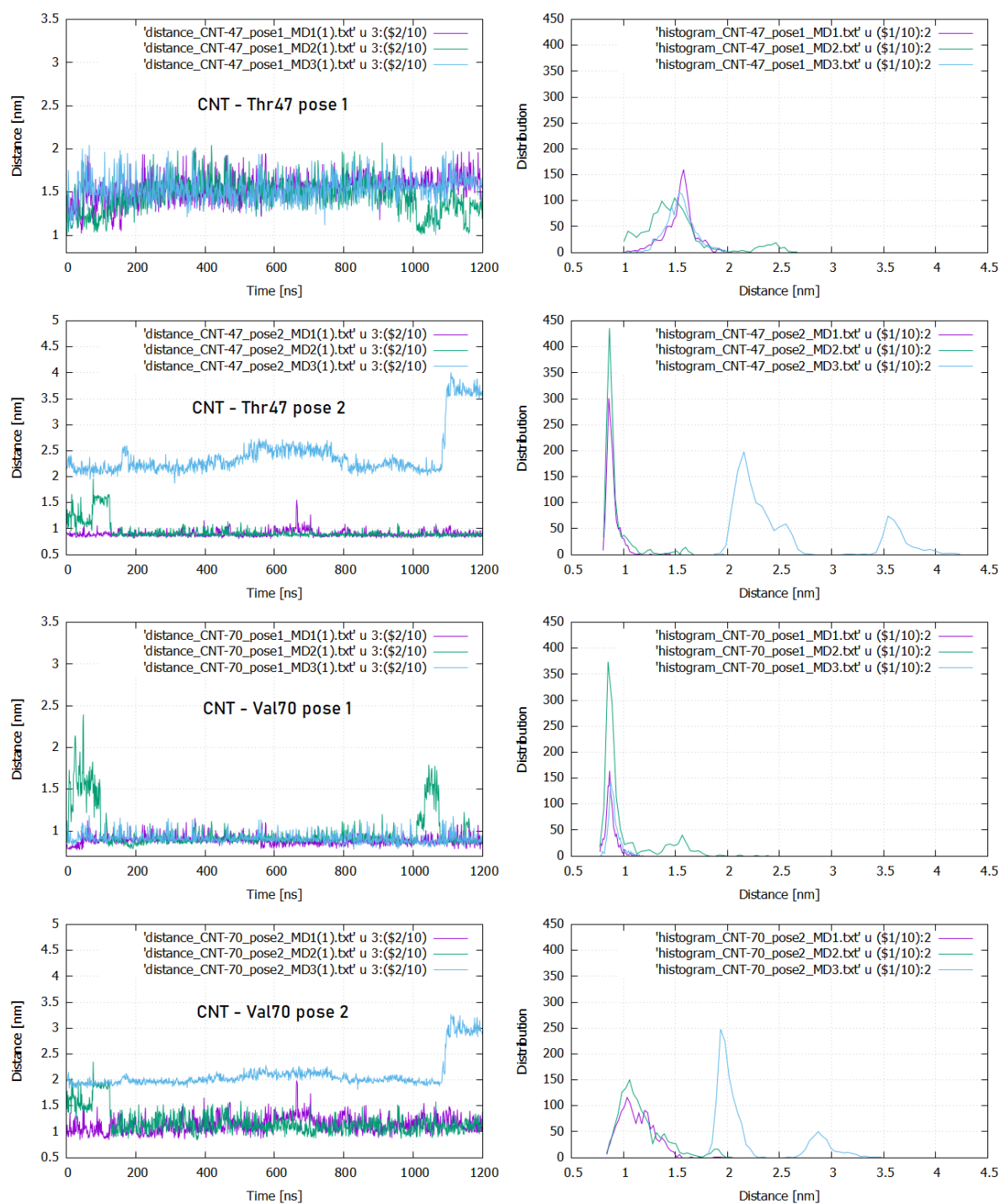
**Figure 10:** Image shows the PQQ-GDH protein monomer with a short carbon nanotube in pose 1 and pose 2. The residues interacting with the CNT in pose 1 are shown in orange, and in the adjacent picture, the interacting residues in pose 2 are shown in blue. The image below shows aligned structures of both CNT positions. Also, the active site residue His144 is highlighted in magenta and in yellow for each pose. The last image represents magnified display of the active site molecules and their atoms which have a crucial role in catalytic activity of the protein.



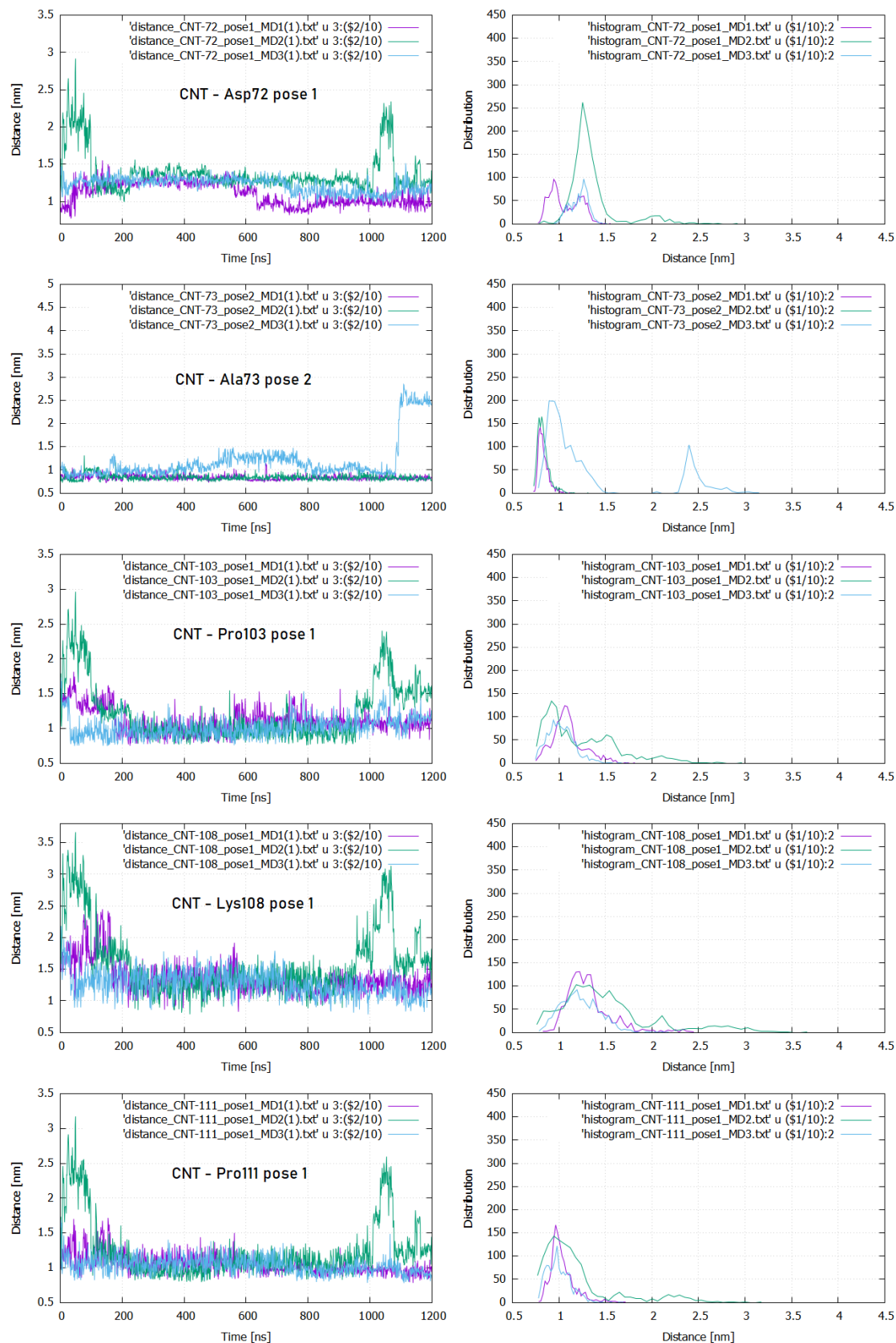
### 3.3 Carbon nanotube residues distance analysis

In order to compare interactions between the protein and the carbon nanotube, in different poses, analysis of the distance of each selected residue from the CNT in two poses was performed. The analysis was done using the Python script *distance.tcl*, described in more detail in Methods and Tools chapter and presented in Appendix A. Obtained distances in nanometers were plotted in time (nanoseconds) and visualized using the Gnuplot program. The script calculates time-dependent distances, that is, in each frame and a histogram of distances for each selected group of atoms. The command *distance "resid no." "resname CNT"* in Tk console specifies the sequence number of residues and all CNT atoms between which the distance is calculated. To calculate the histogram, in each simulation, the number of classes is set to 42 ( $N_d = 42$ ).

Figure 11 shows the time-dependent distances of residues Thr47 and Val70 from the short nanotube in poses 1 and 2. Next to each graph of time dependence, a histogram is shown, that is, the distance distribution of the same residue. In simulation MD2, pose 2, presented by the blue graph, both residues, Thr47 and Val70, are at greater distances from CNT, and their graphs have the same pattern. Indeed, the distance varies between 2 and 2.5 nm for Thr47 and about 2 nm for Val70, and after 1000 ns the distance increases strongly; between 3.5 and 4 nm for Thr47 and about 3 nm for Val70. This is demonstrated in the first two images of Figure 13 by contrasting two aligned structures at frames 1008 and 1108. The first image displays Thr47, presented with yellow and magenta spherical clusters, in relation to the nanotube at two frames. The second image displays same two frames for Val70 in the same manner. Images demonstrate a significant difference in the distances of the indicated residues from the CNT between the displayed frames. On the other hand, the graphs of the individual residues in pose 1 show same patterns for each simulation, with small deviations in the case of Val70 in simulation MD3 (green graph), which reaches slightly higher values between 2 and 2.5 nm right at the beginning of the simulation. The distance of Thr47 in pose 1 oscillates the most, i.e. it has the largest oscillations of all the residues shown, which is reflected in a larger distribution width and larger distance values, with mean values around 1.4 and 1.6 nm. This suggests less stable interactions with the nanotube. Larger distance values in the MD2 simulation in pose 2 and changes in oscillation level are best seen in the histograms, where they appear as two blue peaks at larger distances. With the exception of simulation MD2 in pose 2, residues Thr47 in pose 2 and Val70 in pose 1 reach the smallest distance from CNT. From this, it can be inferred that these residues, in their respective poses, have stronger interactions with the nanotube during the first 1200 ns of the simulation.

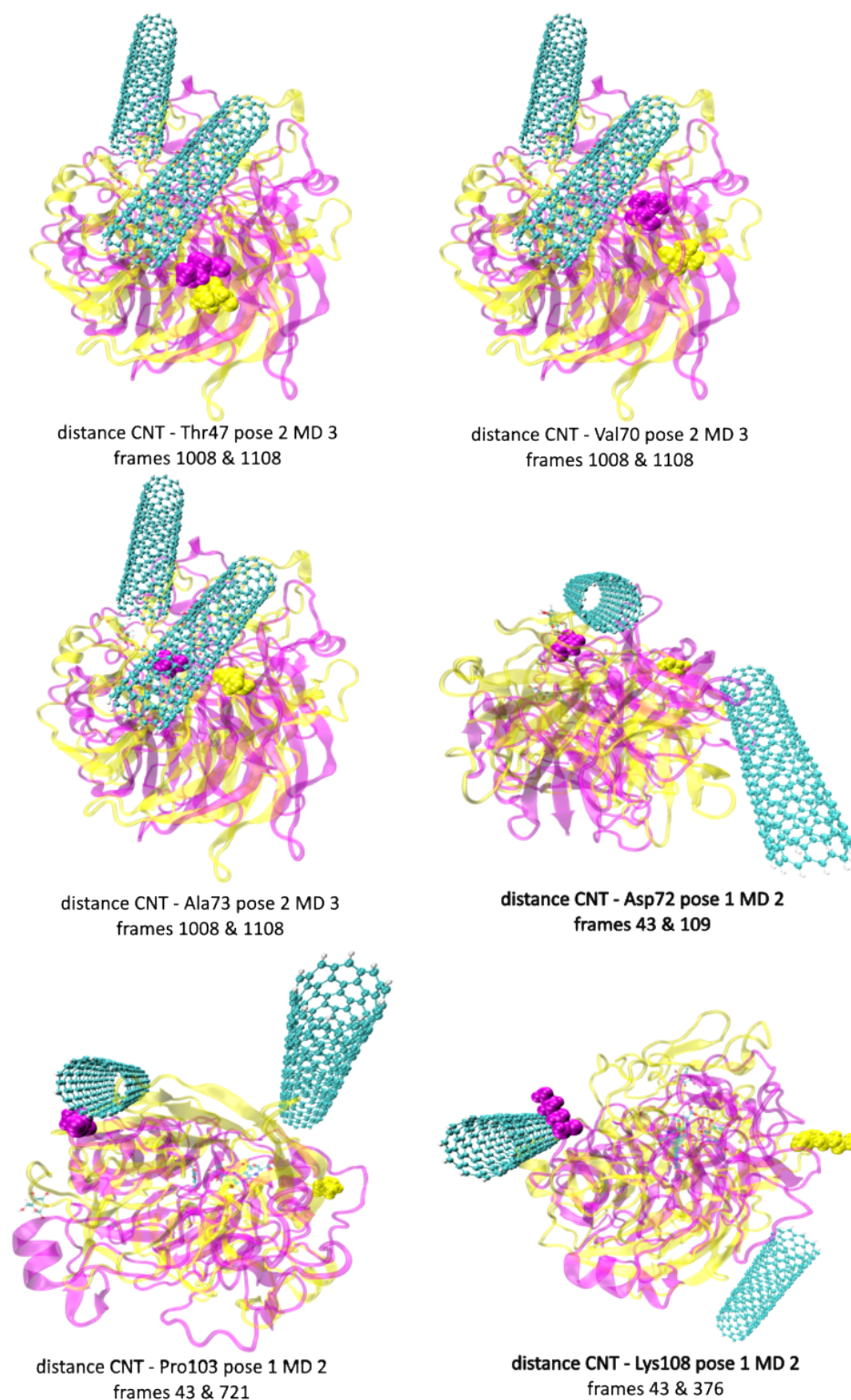


**Figure 11:** The time-dependent distances of residues Thr47 and Val70 from the short nanotube in poses 1 and 2. Next to each graph of time dependence, a distance distribution of the same residue is shown. Distance is calculated for each simulation system, presented with different coloured graph.



**Figure 12:** The time-dependent distances of residues Asp72, Pro103, Lys108 and Pro111 from the short nanotube in pose 1 and distances of Ala73 in pose 2. Next to each graph of time dependence, a distance distribution of the same residue is shown. Distance is calculated for each simulation system, presented with different coloured graph.

Figure 12 shows the time-dependent distances and corresponding distance histograms of residues Asp72, Pro103, Lys108 and Pro111 in pose 1 and residue Ala73 in pose 2 from the short nanotube. The identical graph patterns for residues Asp72, Pro103, Lys108, and Pro111 in posture 1 are instantly visible, especially in the MD2 simulation, presented by the green graph. The graph shows how the protein behaves almost equivalently at the positions of the all the mentioned residues. In the mentioned MD2 simulation, protein records an increase in the value at the beginning of the simulation and after 1000 ns. This increase in value is illustrated in the fourth, fifth and sixth image of the Figure 13, which show different distances between the nanotube and residues Asp72, Pro103 and Lys108 at two frames. At the frame 43 and frame 109, 721 and 376, for Asp72, Pro103 and Lys108, respectively. Mean distance values for mentioned residues range from 1 to 1.25 nm, while their maximum distances are up to 3 and slightly more than 3.5 nm for Lys108. In the MD2 simulation, Val70 in pose 1 shows similar behavior, although with noticeably less oscillations (see Figure 11, green graph CNT - Val70 pose 1). In general, Pro103, Lys108, and Pro111 have larger and more distance oscillations throughout simulations, resulting in broader distributions and indicating less stable interactions with the nanotube. On the other hand, with oscillations smaller than 1 nm in simulations MD1 and MD2, Ala73 achieves the smallest distances from CNT throughout simulations, excluding the MD3 simulation, presented by the blue graph. Its sudden increase in value after 1000 ns is illustrated in the third image of the Figure 13, through different Ala73 distances from CNT in frames 1008 and 1108. This behaviour of Ala73 indicates stable and stronger interactions with the carbon nanotube.



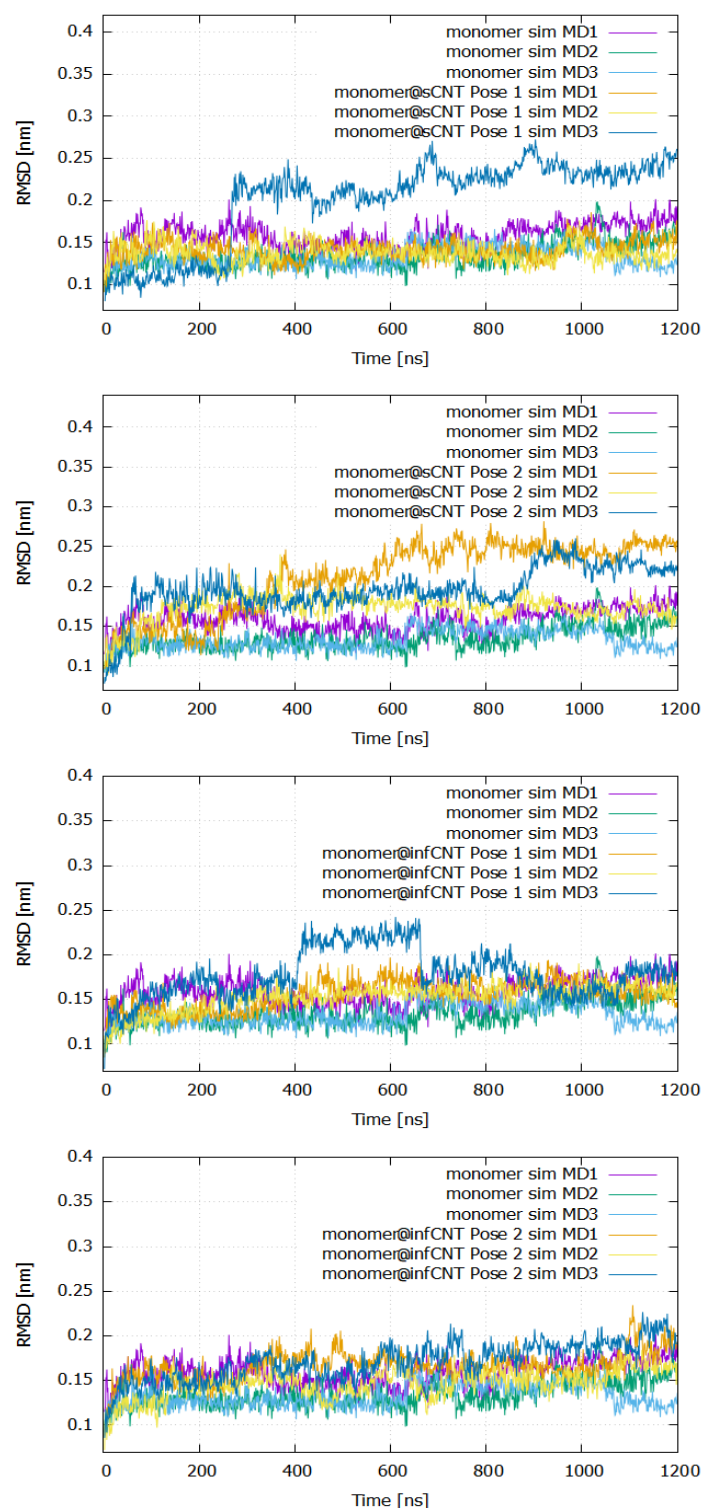
**Figure 13:** Differences in the distances between individual residues and the short carbon nanotube, as seen in Figures 11 and 12, are visualized by comparing two structures in two different frames. Structures from different frames are depicted in yellow and magenta. Each image depicts a distinct residue, represented by yellow and magenta spherical clusters, as well as the positions of a short nanotube in two frames.

### 3.4 RMSD analysis

The RMSD values of the PQQ-GDH protein monomer without the carbon nanotube and in the system with short and infinite nanotube were computed and analyzed, in order to assess the conservation of the protein's monomeric structure and, consequently, its biological activity. This analysis compares the monomeric structures of proteins by calculating their RMSD values in two frames, in the first (labeled as 0 frame in VMD) and every other simulation frame. Therefore, each one of the resulting graphs presents the equilibration process of the protein system, from its initial to some final configuration state. By typing a keyword "*resid 1 to 451*" in the RMSD window tool, the RMSD-value for every protein monomer atom was calculated. When the extra *Backbone* option is chosen, the RMSD values of N - C $_{\alpha}$  - C atoms, i.e. atoms that form the protein skeleton, are computed. By calculating the RMSD-value of the atoms on the protein skeleton, all other side chain atoms, which contribute to higher RMSD values, are eliminated. The Methods and Tools chapter provides a full explanation of how to use the *VMD RMSD Trajectory Tool* to calculate the RMSD values of a group of atoms.

Figure 14 shows the RMSD plots, as a function of time, of the pure monomer system and, respectively, the monomer with short CNT in pose 1, the monomer with short CNT in pose 2 and the monomer with infinite CNT in pose 1 and the same monomer in pose 2. RMSD values were calculated and plotted for all three simulations in each protein system. For all protein systems, during most of the simulation time, mostly the monomer in the nanotube system achieves higher RMSD values than the corresponding pure monomer. An exception is the monomer in the system with a short nanotube in pose 1, in simulation MD1, where the pure monomer (purple graph) achieves higher RMSD values than the corresponding monomer with a short CNT (orange graph). Also, in the same simulation, the pure monomer records higher amplitudes than that in the system with CNTs, in both poses; in pose 1 in the system with infinite CNT, the first 300 ns of the simulation, and in pose 2, also with infinite CNT, the first 300 ns and at certain moments until the end of the simulation. The monomer with short CNT in the MD3 simulation (blue graph) achieves much greater amplitudes than the analogous pure monomer (cyan graph), as seen in the first image of Figure 14. On the other hand, monomers with CNTs in simulations MD1 (orange graph) and MD2 (yellow graph) have significantly lower amplitudes, which indicates a faster equilibration of the system and, consequently, a more stable protein structure. The RMSD values of monomers with short CNT in pose 2 are shown in the second image of Figure 14. Higher values of monomers with CNT are shown for nearly all 1200 ns of simulation time, however, there are significant variations in amplitudes between individual simulations. The simulation MD1 (orange graph) achieves the largest RMSD amplitudes. On the other hand, the monomer with CNT achieves the smallest amplitudes in the MD2 simulation (yellow graph), so it is also the most stable monomer in the system with a short nanotube in pose 2.





**Figure 14:** Time-dependent RMSD values, of the PQQ-GDH protein monomer and the same monomer in short and infinite CNT systems, for each simulation system are shown. The values of the pure monomer are displayed in violet, green, and cyan graphs for simulations MD1, MD2, and MD3, respectively. In the system of monomers with nanotubes, simulations in the same order, are displayed, respectively, in orange, yellow, and blue graphs.

The RMSD graphs of monomers with an infinite CNT in pose 1 are shown in the third image of Figure 14. In simulations MD1 and MD2 (orange and yellow graph), the monomer shows stability; however, in simulation MD3 (blue graph), the monomer with CNT experiences a sudden increase in the RMSD value, which it maintains between 0.2 and 0.25 nm, in the approximate 420 to 660 ns time frame. With the exception of the final 100 ns, when the RMSD amplitudes of the monomers in simulations MD1 and MD3 rise above 0.2 nm, the final image of the Figure 14, achieving the lowest amplitudes during simulations, shows the greatest stability out of all presented monomers.

**Table 1:** Table shows the mean RMSD values, in nanometers, of pure PQQ-GDH protein monomer and monomer with short and infinitely long carbon nanotube, in all simulation systems. Values, rounded to two decimal places, were calculated using RMSD Trajectory Tool.

	monomer w/ short CNT		monomer w/ infinite CNT		pure monomer
	pose 1	pose 2	pose 1	pose 2	
MD1	0.14	0.22	0.16	0.17	0.16
MD2	0.14	0.18	0.15	0.15	0.14
MD3	0.21	0.2	0.18	0.17	0.13

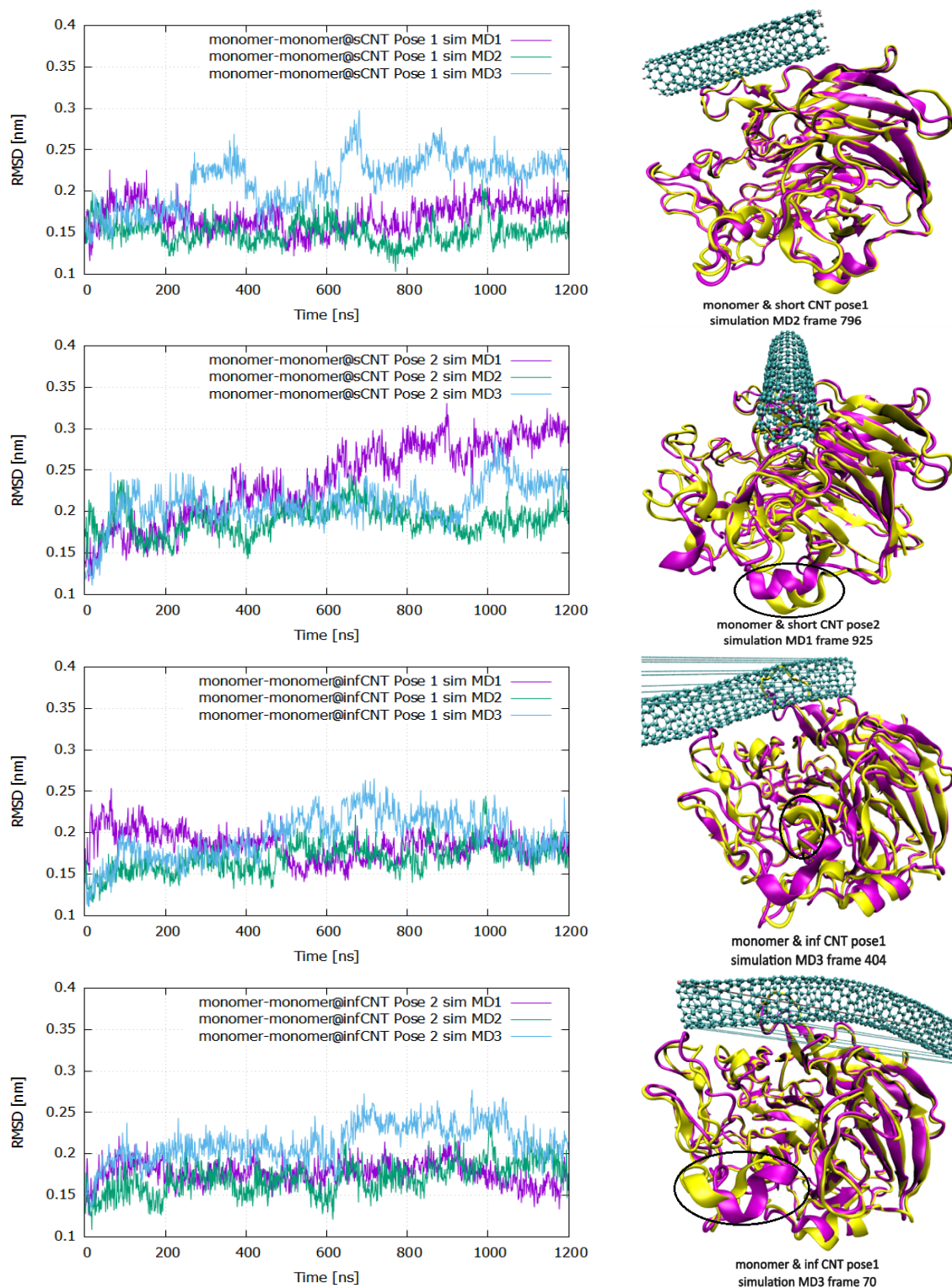
Mean RMSD values of all three systems; pure monomer, monomer in the system with a short and monomer in the system with an infinite carbon nanotube, are shown in Table 1. The values were calculated using a VMD tool, *RMSD Trajectory Tool*, converted to nanometers and rounded to two decimal places. The monomer in the system with short CNT in pose 2 has the greatest values, and the monomer in the system with infinite CNT, also in pose 2, has all values higher than the pure monomer in the corresponding simulations. It is evident from the presented table values, that all the monomers with CNTs, in the MD3 simulation, differ the most from the pure monomer values, whose mean RMSD value is only 0.13 nm.

Therefore, it can be concluded that the monomer in pose 2, in the system with a short CNT, generally shows a different equilibration path, that is, variable patterns of behavior during the simulations, relative to the pure monomer, based on the mean values presented in the Table 1 and the presented RMSD-time plots. Furthermore, monomers with nanotubes in both poses, particularly in the MD3 simulation, show deviations from the pure monomer in terms of equilibration trajectories and, therefore, structural stability.



To gain better insight into the structural differences between the pure monomer and the monomer in system with a carbon nanotube, RMSD values between two protein structures were computed using the script *rmsd.tcl*, which is described in more detail in the Methods and Tools chapter, and presented in Appendix A. The computation was done on each residue including only backbone residues, in each simulation time frame, by typing a following command in the Tk console: `"print_rmsd "resid 1 to 451 and backbone""`. Unlike the previous RMSD analysis of the monomers equilibration trajectories, this analysis gives a better understanding of structural differences of the two monomers in each simulation time frame. The RMSD values of the PQQ-GDH protein monomer without the carbon nanotube and in the system with short and infinite nanotube were computed in this analysis.

Figure 15 shows the RMSD time plots between the pure monomer system and the monomers with carbon nanotubes, respectively, with a short CNT in poses 1 and 2, and the monomer with infinite CNT in poses 1 and 2. RMSD values were calculated and plotted for all three simulation systems. The three-dimensional structures of two monomer systems are displayed in the frame next to each plot; a magenta ribbon representing a pure monomer structure and a yellow ribbon representing a monomer with one type of CNT. It is evident that the monomers with short nanotubes show larger structural deviations from the pure monomer, which is reflected in higher RMSD amplitudes in graphs. The most pronounced difference between the two structures presents in the MD3 simulation (cyan graph) in pose 1. Even larger difference of the two structures presents in the MD1 simulation (purple graph) in pose 2, with a maximum of roughly 0.33 nm, which is the highest RMSD value out of all presented simulation systems. This is illustrated in the figure next to the plot. The structures of the pure monomer and the monomer with a short CNT in pose 2 are displayed, in the MD1 simulation at the time frame 925, when the biggest RMS deviation between the structures occurs. The differences in the structures are most visible in the different positions and shapes of alpha coils and the mismatches in various loops. The monomer with a short CNT in pose 1 in the MD2 simulation (green graph) achieves the smallest RMSD amplitude, indicating the best structural overlaps with the pure monomer. This is shown in the figure next to the plot, which depicts two almost completely overlapped structures at the time frame 796 and represents an RMSD value of approximately 0.1 nm. Furthermore, the MD2 simulation (green graphs) in almost every simulation system achieves the least RMSD values, which indicates two more similar structures, and in addition, reduces the influence of the simulation parameters themselves on the same RMSD values. Monomers with infinite CNTs achieve noticeably lower RMSD values, mostly below 0.25 nm (simulations MD1 and MD2; purple and green graphs) and, in some moments during simulations, values higher than 0.25 nm (MD3 simulation; cyan graph). Figures next to the graphs show monomer structures with infinite CNT from the MD3 simulation system in pose 1, in different time frames.



**Figure 15:** Time plots of RMSD between the PQQ-GDH protein monomer and the same monomer in short and infinite CNT systems, for each simulation system are shown. The values are displayed in purple, green and cyan graphs for simulations MD1, MD2 and MD3, respectively. Images of aligned three-dimensional structures of two monomer systems, one of the pure monomer and the other with the nanotube, are displayed next to the plots. These images were taken at specific simulation frames to better illustrate the similarities and the differences of the two protein structures.

In frame 404, the RMSD value between the two structures is approximately 0.2 nm, and in frame 70, i.e. at the beginning of the simulation time, the RMSD is approximately 0.15 nm. The difference in the similarity of the two structures is mostly noticeable in the positions of the two alpha coils, circled in the pictures. In frame 404, the yellow coil is not completely visible, because it is in a different position compared to the purple one, while in frame 70, both alpha coils are completely visible.

**Table 2:** Table shows the mean RMSD values, in nanometers, between the two structures, the structure of the pure PQQ-GDH protein monomer and the same monomer with the short and infinite carbon nanotube. Values, rounded to two decimal places, were calculated for each simulation system. Values were obtained by taking the average RMSD value by frames, calculated using the script *rmsd.tcl*.

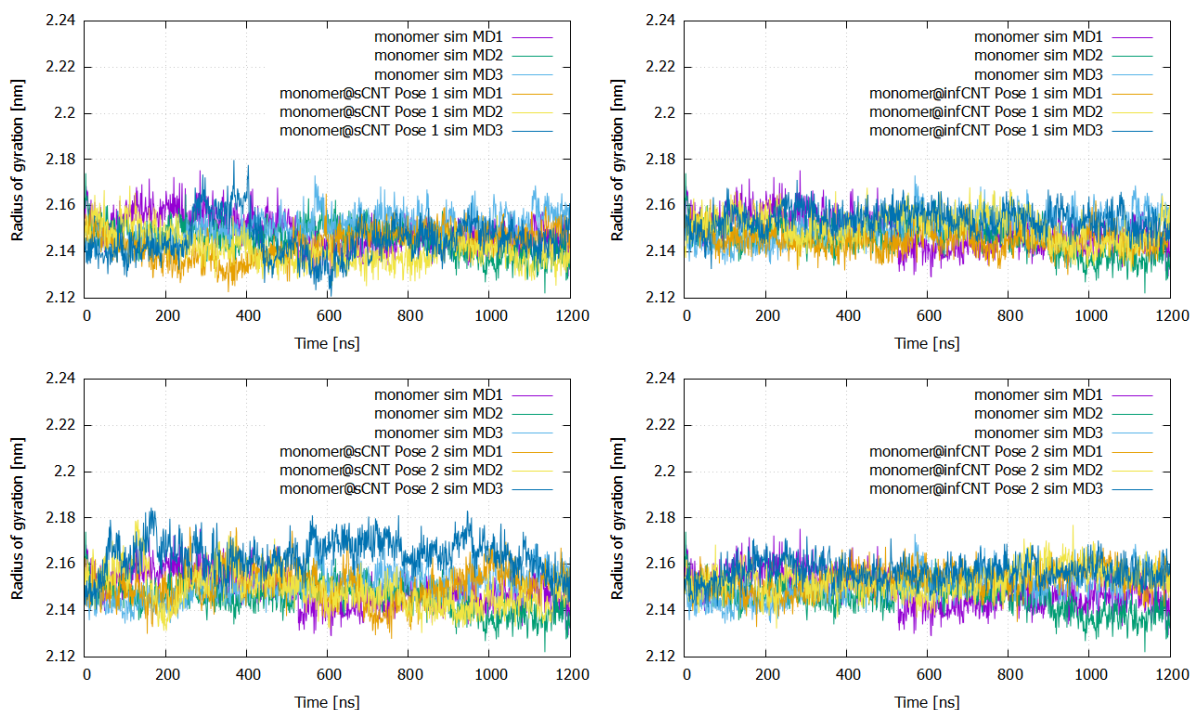
	mon-pose1 sCNT	mon-pose2 sCNT	mon-pose1 infCNT	mon-pose2 infCNT
MD1	0.17	0.24	0.18	0.18
MD2	0.15	0.19	0.17	0.17
MD3	0.21	0.21	0.19	0.21

Table 2 shows mean RMSD values between the two protein structures for each simulation systems, calculated using the script *rmsd.tcl*. As mentioned before describing the graphs, the MD2 simulations in each protein system have the least RMSD values, thus indicating more structural similarity between the pure monomer and the one with a CNT in its system. The monomer in pose 2 with a short CNT shows the highest RMSD values, which indicates large differences in the mutual orientations and positions of the two structures, especially in simulations MD1 and MD3.

Carugo [29] suggests that RMSD values between two protein structures greater than 2 Å can be interpreted as statistically significant differences between the structures, but emphasizes that the size of the proteins and the alignment of their structures should be taken into account, before making similar conclusions. According to the presented graphs and the values in the Table 2, both monomers with short CNT in pose 1 (simulations MD1 and MD2) and monomers with an infinite CNT (all simulations in both poses except simulations MD3) show great similarity with pure monomers. This points to the non-negligible influence of the short and infinite carbon nanotube on the molecular motions of protein monomers. Also, considering the low RMSD values, of the protein systems mentioned above and shown in the previous graphs in Figure 15, in a way, it can be said that the carbon nanotube in these simulation systems stabilizes the entire protein, limiting the fluctuations of its structural parts. The reduced fluctuations lead to smaller differences in the structural overlaps of the two monomers, caused by different mutual positions and orientations during the simulations, which ultimately contributes to a smaller RMSD value.

### 3.5 Radius of gyration analysis

In order to gain a better insight into the compactness and stability of the protein structures analyzed in this work, the calculation of the radius of gyration was performed on the PQQ-GDH protein monomer systems, with and without carbon nanotube. The radius of gyration for each structure was calculated using the scripts *center\_of\_mass*, *gyr\_radius* and *rog\_loop\_dcd*, which are described in more detail in the Methods and Tools chapter and presented in Appendix A.



**Figure 16:** Time plots of the radius of gyration for all monomeric systems without and with a short and infinite carbon nanotube, in all simulation systems, are shown. Simulations MD1, MD2 and MD3 of the pure monomer are shown, respectively, in purple, green and cyan graphs. Simulations, in the same order, of monomers in the system with nanotubes are shown, respectively, in orange, yellow and blue graphs. From the attached graphs, it is evident that all monomer structures are compact and stable.

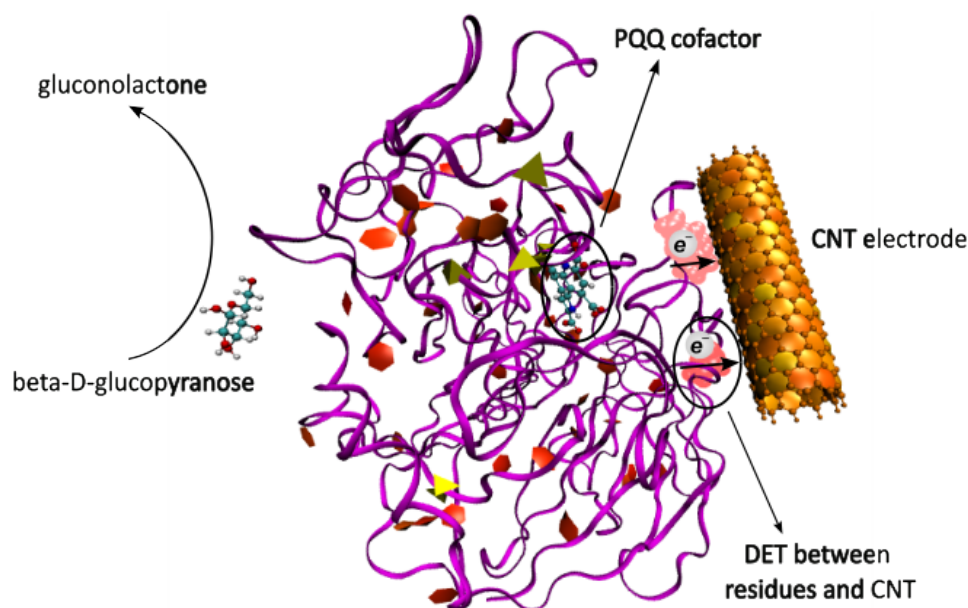
From the graphs shown in the Figure 16, it is evident that these are compact and stable protein structures. The amplitudes of the radius of gyration of all monomer structures almost overlap during the entire simulation time. The overlap is more pronounced in the systems with an infinite nanotube, while slightly higher values of the radius of gyration are achieved by the systems with a short nanotube. In particular, the monomer with a short nanotube in pose 2, in the MD3 simulation (blue graph), achieves the highest values of the radius of gyration, above 2.18 nm. The monomer with a short nanotube in pose 1, in the MD3 simulation, also achieves the largest amplitudes (up to 2.18 nm), but also the smallest amplitude (around 2.12 nm) around 600 ns. Therefore, the mentioned monomer has the largest value differences of the radius of gyration. However, when considering the order of magnitude of the difference, about

$2.18 - 2.12 = 0.06$  nm, the difference is insignificant when compared to the monomer's overall dimensions. The drop in oscillations of radius is also visible in other monomers in pose 1 with a short CNT (orange and yellow graph) and, as well as, in the pure monomer, in the MD2 simulation.

### 3.6 Electron hopping process in protein

Electron hopping in biomolecules is a process where electrons migrate through molecules in a multi-step reaction. This enables long-distance electron transfer in biomolecules. Electrons can migrate in biopolymers like DNA and proteins over long distances (up to 30 nm). According to the electron transfer theory (see equation (1.1)), a single-step electron transfer (ET) over these distances is too slow to occur because the rate  $k_{ET}$  decreases exponentially with the distance  $(d - d_0)$  between electron donor and electron acceptor. Long-distance ET becomes possible if biopolymers contain groups that can be oxidized or reduced during electron transfer. These groups act as relay stations (stepping stones) where the charge has a definite life time. The electron hops through the molecule in a multi-step reaction, while the rate of each hopping step follows the Marcus equation. The overall process can be mathematically described as a diffusion of the charge through the biomolecule. Long-range electron transfer through proteins and enzymes is a fundamental reaction for all living organisms. In energy conversion processes like photosynthesis and respiration, multiprotein assemblies control ET through the cell membrane over more than 4 nm. Relay stations for these processes are either electron transfer cofactors or certain amino acids, whose side chains have appropriate redox potentials for a hopping process. This condition is fulfilled for tyrosine, tryptophan, and histidine, which carry aromatic side chains, but also for the sulfur-containing aliphatic amino acids, cysteine and methionine. Electron hopping process is possible through strong  $\pi$ - $\pi$  interactions between aforementioned aromatic amino acid side chains. For example, in ribonucleotide reductase both tyrosine and tryptophan are the stepping stones of the long distance ET. [8]

Schematic representation of the biosensor based on the studied PQQ-GDH protein simulation system in complex with a short carbon nanotube is shown in the Figure 17. Figure represents a snapshot of the protein in the MD1 simulation with the nanotube in pose 1. Bearing in mind the idea of electron hopping between the aromatic amino acids His, Tyr and Trp, and those containing sulphur, Met and Cys, Figure 17 shows these amino acids, dispersed in the protein's backbone structure, which is presented with a magenta ribbon. Residues aromatic side chains along with aromatic benzene components of the CNT, are presented with the special drawing method *PaperChain*. This method displays ring structures as polygons, colored by ring pucker.



**Figure 17:** Schematic representation of the working principle of the biosensor, constructed with the PQQ-dependent glucose dehydrogenase in complex with short carbon nanotube, serving as electrode. An electron, which is generated during the catalytic reaction in the enzyme active site, transfers to the nanotube by the process of DET between the edge residues of protein and the benzene components of nanotube. Such system has a potential of measuring the current of generated electrons, depending on the concentration of glucose in the system.

The sulfur-containing amino acids are presented with yellow tetrahedrons, created with the drawing method *Polyhedra*, which draws a collection of triangles that connect all triplets of groups of atoms within a user-defined radius. [10] A brief description of the working principle of the biosensor illustrated in Figure 17 would be as follows. In the presence of glucose, that is, beta-D-glucopyranose molecule, protein PQQ-GDH oxidizes glucose to gluconolactone, while the cofactor PQQ is being reduced. Then, the PQQ electron transfers to the active site residue His144, from where it hops between aromatic and sulfur-containing aliphatic residues. When it reaches the last residues closest to the nanotube, it transfers by the direct electron transfer from the edge residues to the nanotube. In pose 1, these edge residues are Pro67, Pro103 and Pro111, which make C-H/ $\pi$  interactions with the nanotube, as mentioned previously. In pose 2, the residue closest to CNT is Thr47, which makes an O-H/ $\pi$  interaction with the nanotube. Zondlo in his research article discussed about aromatic-proline interactions and did comparison with aromatic-aromatic interactions. He found that the aromatic-proline interactions are enthalpically favorable, consistent with a C-H/ $\pi$  interaction and inconsistent with a classical hydrophobic effect, meaning that aromatic-aromatic interactions are similar in enthalpy to aromatic-proline interactions and, consequently, can counteract them. Additionally, what is certainly the most important information from that article, regarding the protein-CNT system, is that C-H/ $\pi$  interactions are electronically tunable, which means that there is a possibility of direct electron transfer happening between the two interacting compounds. [30]



This suggests that there is a possibility of electron transfer between the proline residues and the carbon nanotube. Due to its great conductivity, carbon nanotube serves as electrode which can easily conduct the current of generated electrons. In such a system, it is possible to measure the current of generated electrons, depending on the concentration of glucose in the system, with a constant influx of new cofactor molecules, which restore enzyme function.

## 4 Conclusion

Popularity of enzymatic biofuel cells has grown over the past few years, due to their biological origin which counteracts abiotic, noble-metal or alloys-based fuel cells. Soluble glucose dehydrogenase (s-GDH) is a classical quinoprotein which requires the cofactor pyrroloquinoline quinone (PQQ) to oxidize one molecule of glucose. Enzymes from the oxidoreductase class catalyze reactions that involve electron transfer from glucose molecule, the electron donor, to another molecule, the electron acceptor. Because of the high catalytic activity and the oxygen independent function of PQQ-GDH, the soluble enzyme has found considerable application in biosensing and bioenergetics. Carbon nanotubes are used as electrodes in such systems, because of its excellent electrical conductivity and high chemical stability. Third-generation sensors employ direct electron transfer from the cofactor, in this case PQQ, to the carbon nanotube electrode, thus eliminating toxic artificial electron mediators. This thesis presents molecular dynamics analysis of the protein structures in complex with a short and infinite carbon nanotubes, using different mathematical tools.

Calculations of distances between certain residues and CNT showed that residues Thr47 and Ala73 in pose 2 and Val70 in pose 1 make more stable and stronger interactions with CNT. Interactions with CNT are generally weak electrostatic forces, by one order of magnitude weaker than classical hydrogen bonds. Analysis of the number of hydrogen bonds formed between residues and CNT showed higher numbers for pose 1, which was expected, considering the pose 1 having more interacting residues, which is extracted earlier from the binding free energy calculations. Calculating root-mean-square deviations of proteins structures of pure monomer and with a CNT during simulation times, and measuring RMSD between mentioned structures, gave us a better insight in the protein-CNT complex stability and preservation of its function. Results of RMSD analysis showed that the pure monomer systems are slightly more stable than those with a CNT in their system. Furthermore, calculating the direct RMSD between the pure monomers and the monomers with carbon nanotubes, we concluded, regarding their lower RMSD values, that some simulation systems, those with short CNT (pose 1, simulations MD1 and MD2) and with an infinite CNT, show great structural overlap with pure monomers. Therefore, it can be said that the carbon nanotube in some simulation systems stabilizes the entire protein, limiting the fluctuations of

its structural parts. Furthermore, analysis of the radius of gyration showed overlapping radius amplitudes of all monomer systems during simulation times.

By introducing working principle of the third-generation biosensors in the *Application in bioelectrochemistry* chapter, and considering information gathered from our molecular simulations analysis and literary sources, we have finally been able to construct, at least theoretically, a model for biosensor comprising the PQQ-GDH enzyme in complex with a carbon nanotube. Such system, due to CNT's great electrical conductivity and high enzyme catalytic activity, can measure the current of chemically generated electrons depending on the glucose concentration in the system. Therefore, by explaining functioning of such systems on atomic level, this model, comprising PQQ-GDH enzyme and a short CNT, serves as an adequate theoretical prototype of the enzymatic biosensors, some of which have the opportunity of implementation in the system of pacemakers and glucose detectors.



## 5 Literature

- [1] A. Oubrie, H. J. Rozeboom, K.H. Kalk, A. J. J. Olsthoorn, J. A. Duine and B. W. Dijkstra: *Structure and mechanism of soluble quinoprotein glucose dehydrogenase*, The EMBO Journal, vol. 18., 1999.  
DOI: <https://doi.org/10.1093/emboj/18.19.5187>
- [2] David C. Ferrier and Kevin C. Honeychurch: *Carbon nanotube (CNT)-based biosensors*, Biosensors, Multidisciplinary digital publishing institute, 2021.  
DOI: <https://doi.org/10.3390/bios11120486>
- [3] *RCSB PDB*  
URL: <https://www.rcsb.org/structure/1CQ1> (12. 6. 2024.)
- [4] S. Marchesan and M. Prato: *Under the lens: carbon nanotube and protein interaction at the nanoscale*, Royal society of chemistry, January 2012.  
DOI: <https://doi.org/10.1039/C4CC09173F>
- [5] F. Lisdat: *PQQ-GDH – Structure, function and application in bioelectrochemistry*, Bioelectrochemistry, 2020.  
DOI: <https://doi.org/10.1016/j.bioelechem.2020.107496>
- [6] M. Holzinger, Y. Nishina, A. Goffl, M. Tominaga, S. Cosnier and S. Tsujimura: *Molecular design of glucose biofuel cell electrodes*, Molecular Technology: Energy Innovation, Volume 1, Wiley-VCH Verlag GmbH & Co, 2019.  
DOI: [https://doi.org/10.1002/9783527823987.vol1\\_c11](https://doi.org/10.1002/9783527823987.vol1_c11)
- [7] S. Ferri, K. Kojima and K. Sode: *Review of glucose oxidases and glucose dehydrogenases: A bird's eye view of glucose sensing enzymes*, Journal of diabetes science and technology Volume 5, Issue 5, September 2011.  
DOI: <https://doi.org/10.1177/193229681100500507>
- [8] David N. Beratan and Spiros S. Skourtis: *Electron transfer through proteins*, Encyclopedia of Biophysics, Springer, 2013.  
DOI: [https://doi.org/10.1007/978-3-642-16712-6\\_13](https://doi.org/10.1007/978-3-642-16712-6_13)
- [9] P. Bollella and E. Katz: *Enzyme-based biosensors: Tackling electron transfer issues*, Sensors, Multidisciplinary digital publishing institute, 2020.  
DOI: <https://doi.org/10.3390/s20123517>
- [10] *Theoretical and computational biophysics group*, VMD  
URL: <https://www.ks.uiuc.edu/Research/vmd/> (12. 6. 2024.)

- [11] *Using VMD*  
URL:  
<https://www.ks.uiuc.edu/Training/Tutorials/vmd/tutorial-html/vmd-tutorial-main.html>  
(12. 6. 2024.)
- [12] NIH Biomedical Research Center for Macromolecular Modeling and Bioinformatics:  
*VMD User's guide*, November 27, 2016.
- [13] Tagyedeem H. Shoaib, W. Ibraheem<sup>1</sup>, M. Abdelrahman *et al.*: *Exploring the potential of approved drugs for triple-negative breast cancer treatment by targeting casein kinase 2: Insights from computational studies*, PLoS ONE, August, 2023.  
DOI: <https://doi.org/10.1371/journal.pone.0289887>
- [14] *How to calculate the radius of gyration for a protein using VMD?*, M. Shehata, YouTube video  
URL: <https://www.youtube.com/watch?v=U15FVwQ8ynQ> (16. 6. 2024.)
- [15] *VMD Distance measurements of a simulation in NAMD*, T. Messina, YouTube video  
URL: <https://youtu.be/KE268sYqrmw?si=M55ZDETKZPp3bbRs> (16. 6. 2024.)
- [16] J. C. Phillips, D. J. Hardy, J. D. C. Maia *et al.*: *Scalable molecular dynamics on CPU and GPU architectures with NAMD*, The Journal of chemical physics, Volume 153, Issue 4, 28 July 2020.  
DOI: <https://doi.org/10.1063/5.0014475>
- [17] C. Wang, G. Yang and Y. Jiang: *Structure and property of multiple amino acids assembled on the surface of a CNT*, Physica E: Low-dimensional systems and nanostructures, 2016.  
DOI: <http://dx.doi.org/10.1016/j.physe.2016.08.018>
- [18] I. Rozas, I. Alkorta and J. Elguero: *Unusual Hydrogen Bonds: H... $\pi$  Interactions*, The Journal of physical chemistry A, 1997.  
DOI: <https://doi.org/10.1021/jp971893t>
- [19] M. Brandl, M. S. Weiss, A. Jabs, J. Suehnel and R. Hilgenfeld: *C-H... $\pi$ -interactions in proteins*, Journal of molecular biology, 2001.  
DOI: 10.1006/jmbi.2001.4473
- [20] M. J. Plevin, I. Hayashi and M. Ikura: *Characterization of a conserved "threonine clasp" in CAP-Gly domains: Role of a functionally critical OH/ $\pi$  interaction in protein recognition*, Journal of the American chemical society, 2008.  
DOI: 10.1021/ja805576n

- [21] K. Kumar, Shin M. Woo, T. Siu, W. A. Cortopassi, F. Duarte and R. S. Paton: *Cation– $\pi$  interactions in protein–ligand binding: theory and data-mining reveal different roles for lysine and arginine*, Chemical science, 2018.  
DOI: <https://doi.org/10.1039%2Fc7sc04905f>
- [22] Z. He and J. Zhou: *Probing carbon nanotube–amino acid interactions in aqueous solution with molecular dynamics simulations*, Carbon 78, July 2014.  
DOI: <http://dx.doi.org/10.1016/j.carbon.2014.07.031>
- [23] UCSF ChimeraX  
URL: <https://www.cgl.ucsf.edu/chimerax/> (11. 9. 2024.)
- [24] APBS  
URL: <https://server.poissonboltzmann.org/> (11. 9. 2024.)
- [25] Adaptive Poisson-Boltzmann Solver (APBS), CHARMM - GUI  
URL: <https://www.charmm-gui.org/charmmdoc/apbs.html#> (11. 9. 2024.)
- [26] M. Azeem and M. A. Saleem: *Role of electrostatic potential energy in carbon nanotube augmented cement paste matrix*, Construction and Building Materials, Elsevier, 2019.  
DOI: <https://doi.org/10.1016/j.conbuildmat.2019.117875>
- [27] George A. Jeffrey: *An introduction to hydrogen bonding*, Oxford University Press, 1997.
- [28] O. Carugo and S. Pongor: *A normalized root-mean-square distance for comparing protein three-dimensional structures*, Protein Science, July 2001.  
DOI: <https://doi.org/10.1110/ps.690101>
- [29] O. Carugo: *Statistical validation of the root-mean-square-distance, a measure of protein structural proximity*, Protein Engineering, Design & Selection, vol. 20, January 11th 2007.  
DOI: <https://doi.org/10.1093/protein/gzl051>
- [30] Neal J. Zondlo: *Aromatic-proline interactions: Electronically tunable CH/ interactions*, Accounts of chemical research, 2013.  
DOI: <https://doi.org/10.1021%2Far300087y>

## A Python coding scripts

Python script used for the calculation of root-mean-square deviation of the two protein structures in each simulation frame is presented in Figure 18. Python scripts used for the calculation of the radius of gyration of the protein structure are presented in Figures 19, 20 and 21. Python script used for the calculation of the distance between specified groups of atoms of the protein structures is presented in Figure 22.

```
1 proc print_rmsd {seltext1} {
2
3 #set protein selections
4 set monomer [atomselect 22 "$seltext1"]
5 set cnt [atomselect 23 "$seltext1"]
6
7 set outfile [open rmsd_monomer_infCNT_pose2_MD3.dat w]
8 puts $outfile "frame rmsd"
9 set num_steps [molinfo 22 get numframes]
10 for {set frame 0} {$frame < $num_steps} {incr frame} {
11     $monomer frame $frame
12     $cnt frame $frame
13
14     #compute the transformation
15     set trans_mat [measure fit $monomer $cnt]
16     $monomer move $trans_mat
17
18     #compute the RMSD
19     set rmsd [measure rmsd $monomer $cnt]
20
21     #print the RMSD
22     puts $outfile "$frame $rmsd"
23 }
24 close $outfile
25 }
```

**Figure 18:** The script *rmsd.tcl* calculates the root-mean-square deviation of the two loaded protein structures. [12]

```

1 proc center_of_mass {selection} {
2     # some error checking
3     if {[$selection num] <= 0} {
4         error "center_of_mass: needs a selection with atoms"
5     }
6     # set the center of mass to 0
7     set com [veczero]
8     # set the total mass to 0
9     set mass 0
10    # [$selection get {x y z}] returns the coordinates {x y z}
11    # [$selection get {mass}] returns the masses
12    # so the following says "for each pair of {coordinates} and masses,
13    # do the computation ..."
14    foreach coord [$selection get {x y z}] m [$selection get mass] {
15        # sum of the masses
16        set mass [expr $mass + $m]
17        # sum up the product of mass and coordinate
18        set com [vecadd $com [vecsacle $m $coord]]
19    }
20    # and scale by the inverse of the number of atoms
21    if {$mass == 0} {
22        error "center_of_mass: total mass is zero"
23    }
24    # The "1.0" can't be "1", since otherwise integer division is done
25    return [vecsacle [expr 1.0/$mass] $com]
26 }

```

**Figure 19:** The script *center\_of\_mass.tcl* calculates the center of mass of the loaded protein structure. [14]

```

1 proc gyr_radius {sel} {
2     # make sure this is a proper selection and has atoms
3     if {[$sel num] <= 0} {
4         error "gyr_radius: must have at least one atom in selection"
5     }
6     # gyration is sqrt( sum((r(i) - r(center_of_mass))^2) / N)
7     set com [center_of_mass $sel]
8     set sum 0
9     foreach coord [$sel get {x y z}] {
10        set sum [vecadd $sum [veclength2 [vecsacle $coord $com]]]
11    }
12    return [expr sqrt($sum / ([$sel num] + 0.0))]
13 }

```

**Figure 20:** The script *gyr\_radius.tcl* calculates the radius of gyration of the loaded protein structure. [14]

```
1 # load necessary tcl functions
2 source gyr_radius.tcl
3 source center_of_mass.tcl
4
5 set outfile [open rg_monomer_MD3.dat w]
6 puts $outfile "i rad_of_gyr"
7 set nf [molinfo top get numframes]
8 set i 0
9
10 set prot [atomselect top "resid 1 to 451"]
11 while {$i < $nf} {
12
13     $prot frame $i
14     $prot update
15
16     set i [expr {$i + 1}]
17     set rog [gyr_radius $prot]
18
19     puts $outfile "$i $rog"
20
21 }
22
23 close $outfile
```

**Figure 21:** *The script rog\_loop\_dcd.tcl calls the previous two scripts, and calculates the radius of gyration of the loaded protein structure for each simulation frame. [14]*

```

1 proc distance {seltext1 seltext2 N_d f_r_out f_d_out} {
2
3 set sel1 [atomselect top "$seltext1"]
4 set sel2 [atomselect top "$seltext2"]
5
6 set nf [molinfo top get numframes]
7
8 set outfile [open $f_r_out w]
9
10 for {set i 0} {$i<$nf} {incr i} {
11     puts "frame $i of $nf"
12     $sel1 frame $i
13     $sel2 frame $i
14     set com1 [measure center $sel1 weight none]
15     set com2 [measure center $sel2 weight none]
16
17     set simdata($i.r) [veclength [vecsub $com1 $com2]]
18
19     puts $outfile "$i , $simdata($i.r)"
20 }
21
22 close $outfile
23
24 set r_min $simdata(0.r)
25 set r_max $simdata(0.r)
26 for {set i 0} {$i<$nf} {incr i} {
27     set r_tmp $simdata($i.r)
28     if {$r_tmp<$r_min} {set r_min $r_tmp}
29     if {$r_tmp>$r_max} {set r_max $r_tmp}
30 }
31
32 set dr [expr ($r_max - $r_min) / ($N_d - 1)]
33 for {set k 0} {$k<$N_d} {incr k} {
34     set distribution($k) 0
35 }
36
37 for {set i 0} {$i<$nf} {incr i} {
38     set k [expr int(($simdata($i.r) - $r_min) / $dr)]
39     incr distribution($k)
40 }
41
42 set outfile [open $f_d_out w]
43 for {set k 0} {$k<$N_d} {incr k} {
44     puts $outfile "[expr $r_min + $k*$dr] , $distribution($k)"
45 }
46
47 close $outfile

```

**Figure 22:** The script *distance.tcl* calculates the distance and its distribution of specified groups of atoms in each simulation frame. [15]