Proceedings of The National Conference
On Undergraduate Research (NCUR) 2008
Salisbury University
Salisbury, Maryland
April 10 - 12, 2008

Study of Stable Conformations of CNT-DNA Hybrids by Means of Principal Component Analysis

Myrna I. Merced Serrano
Department of Mathematics
University of Puerto Rico at Humacao
100 Tejas Avenue
Humacao Puerto Rico 00791

Faculty Adviser: José Sotero Esteva

Abstract

Single walled carbon nanotubes (swCNT) functionalized with single strand DNA (ss-DNA) in a non-covalent interaction have been successfully used as gas sensors. Molecular Dynamics (MD) simulations have been used before to understand the interactions between swCNT and ss-DNA that influence electron transport such as π-π stacking. But studying how gas analytes interact with these hybrids requires detailed observations of the conformation of ss-DNAs. Principal Components Analysis (PCA) is often used to analyze conformations of molecules, mainly proteins, produced by MD simulations. A set of *N* atoms is selected within the molecule. The conformations of the molecule at different moments can be traced by using PCA to map the sequence of points in a 3N-dimensional space into a 2D or 3D space. Clusters of points indicate similar conformations. A PCA analysis was performed on trajectories generated by MD simulations of swCNT-ss-DNA hybrids at room temperature in aqueous solution. Ss-DNAs corresponding to three specific sequences of 21 oglionucleotides used in actual gas sensor experiments as well as four Poly-Cytosine of various lengths were explored. The selection of atoms sets for the PCA analysis was also varied ensuring uniform selections along the DNA. The PCA analysis shows that the paths of the 3D points progress to eventually form clusters indicating that the shapes of the DNAs evolve into stable configurations. No returns in the paths nor jumps between clusters indicate that the DNA reaches a final conformation. These results are consistent with observations made in the laboratory proving that the PCA method is suitable for the study of the conformation of CNT-DNA hybrids.
Keywords: molecule conformations, DNA-CNT hybrids, Principal Components Analysis (PCA)

1. Introduction

The combination of single stranded DNA (ss-DNA) with a single walled carbon nanotube (swCNT) is a particularly elegant example of a self assembled nanodevice that has been experimentally proven to be sensitive to a set of gases. The robustness of the functionalization of the swCNT is crucial for the usefulness of the device[1]. Molecular Dynamics (MD) simulations are well-suited to provide insights into the fundamental properties of DNA-CNT hybrids because they enable calculations of structural properties with an atomic resolution. MD simulations have been used by A.T. Johnson *et.al.* to understand the interactions between swCNT and ss-DNA[2] that influence electron transport such as π-π stacking[3].

Since its first use by A. García[4] for the analysis of MD simulations results, Principal Component Analysis (PCA) is often used to trace the conformations of molecules especially proteins (see R.L. Jernigan *et.al.*[5] for a recent example). Often the *x, y, z* coordinates of the positions of atoms of the molecule serve as the input for the PCA method although the use of the dihedral angles in the backbone has been explored recently[6]. The problem of determining an appropriate sample size has been studied from the point of view of the length of the simulation[7]. Results on the influence of the selection criteria of atoms or angles on the PCA are difficult to find.

This paper presents the application of PCA to the novel studies of conformation of ss-DNA onto CNTs. In order to validate the selection of atoms for the PCA in further studies a comparison is made between the results obtained from several selections of atoms.

2. Background

2.1 **carbon nanotubes**

Carbon nanotubes (CNT) are cylindrical sheets of carbon. CNT have diameters of ~1nm and lengths up to a few centimeters[8]. They have many interesting properties. They can have tensile strength as high as sixty times larger than steel. They also show electronic stability. Nanotubes can accommodate current densities 1000 times higher than copper and silver.

2.2 **ss-DNA**

Single stranded DNA (ss-DNA) is a DNA molecule consisting of only one chain of alternating sugars and phosphates. It can assume different structures depending on the solvent and ionic environment. For this work some ss-DNA composed of a repeating sequence of cytosine (Poly-Cytosine) and specific sequences of ss-DNA were used. Figure 1 shows a ss-DNA onto a CNT.

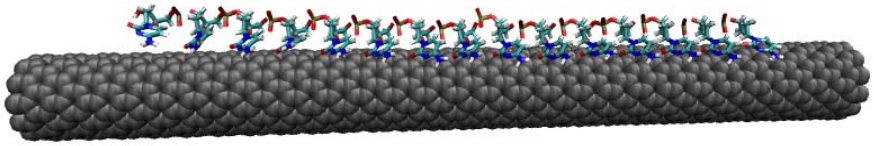


Figure 1. ss-DNA onto a CNT.

2.3 **MD simulations**

MD simulations calculate the trajectories of *N* interacting atoms by numerically solving Newton's equations of motion for each atom. Since atomic forces are conservative, they can be described by a potential function. There are a variety of atom's interactions some of these exists between atoms that do not share a chemical bond (electrostatic and the Van der Waals forces). They are other forces that act on atoms that share chemical bonds. These forces are bond stretching (bond force), bond bending (angle force) and bond twisting (torsion force).

2.4 **PCA**

Principal Component Analysis (PCA) is a technique used to reduce data that is represented in a high number of dimensions to 2D or 3D. PCA allows to visualize the similarities and/or differences in a set of data. Among the wide range of applications of the PCA, it can be used to study the shapes of molecules studied by means of MD simulations[4]. The technique has been used successfully in the analysis of proteins[9].

3. **Software and methods**

The GROMACS MD package[10] was used to perform MD simulations. A detailed script was written in order to precisely the set up of the system[11]. In the following sections is the outline of the procedure to set up and to run the MD simulations. Also, there is a detailed description of the organization of the data and the PCA performed into the data.

3.1. **MoSDAS**

The Model building, Simulation and Data Analysis Script (MoSDAS[12]) was developed to automate the production of the MD system. The development of MoSDAS simplifies and avoids the most of the errors in the simulation process. All the commands of MoSDAS are in *bash* programming language. The main purpose of MoSDAS is to call and run other programs.

The first step is to generate the ss-DNA with *nucleic* program, which is part of the Tinker molecular modeling package[13], a CNT 30Å longer than the ss-DNA is also generated. The ss-DNA and the CNT were joined with *tleap*, which is a subprogram of the AMBER7 MD package[14]. The system was placed inside of a box and hydrated with water. The water inside of the tube was removed with the *tcl* script *rem-wat-interior.tcl* developed by Robert Johnson[3]. The *index* file is the file that classified the atoms by groups, for example WATER, DNA_20L, etc. This file was made with *make_ndx* program, a program that is part of GROMACS. The *topology* file of the system was generated with an *awk* script. The *master input* file was edited from a template. The

editions of the *master input* file were made with a script. Periodic boundary conditions in all directions were used.

3.2 **simulations**

Before each MD simulation a minimization of the system is required. Some systems also need a relaxation after the minimization.

MD simulations of Poly-Cytosines of 5, 15, 25, and 30 monomers as well as the sequences:

Sequence 1: 5' AAA ACC CCC GGG GTT TTT TTT TTG 3'
Sequence 2: 5' CTT CTG TCT TGA TGT TTG TCA AAC 3'
Sequence 3: 5' GAG TCT GTG GAG GAG GTA GTC 3'

These specific sequences are the same used in actual experiments of gas sensors by A.T. Johnson *et.al.*[3]. Each of these simulations was run for 10ns and not all of them have been equilibrated yet.

After the simulation the trajectory data was visualized with Visual Molecular Dynamics (VMD)[15]. The program VMD is a molecular visualization program for displaying, animating and analyzing large biomolecular systems using 3D graphics and built-in scripting[16]. The visualization of the data was made because we wanted to be sure that the system was stable.

3.3 **organization of the trajectory data**

The trajectory file produced by GROMACS that contains all the coordinates, velocities, forces and energies as in the GROMACS *master input* file. This file is in portable binary format and can be read with *gmxdump*, program given by GROMACS.

3.3.1 *atoms sets*

This study is focused on the shape of the ss-DNA only. Different subsets of the ss-DNA atoms were chosen ensuring a uniform distribution across the length of the molecule.

For Poly-Cytosine we have five sets:

1. All ss-DNA atoms except for Hydrogen atoms
2. All atoms in the backbone of the ss-DNA
3. All atoms in all the rings of the ss-DNA
4. The N4 atoms (it is on the end of each ring)
5. The P atoms (they are at backbone)

For the specific sequences we have three sets:

1. All ss-DNA atoms except for Hydrogen atoms
2. All atoms in the backbone of the ss-DNA
3. All atoms in all the rings of the ss-DNA

3.3.2 *script to generate the data matrix*

To perform the PCA of the trajectory data of the different atoms sets it is required to create a data matrix. This matrix needs to have a row for each frame of data, and the columns are the *x, y, z* coordinates of the atoms. For example, the coordinates of atom one are the first three columns of the matrix and all the rows in these three columns are the trajectory of this atom. To generate this matrix a *bash* script was developed. A matrix for each atoms set was generated.

3.4 **PCA in the trajectory data**

After the matrices for each sequence and each atoms set were built, a PCA for all matrices was produced. To make the PCAs, MatLab was used. MatLab is a high-performance language for technical computing[17]. To obtain the PCA the `princomp(x)` function of MatLab was used. After the PCA, the standard scores were analyzed. The first scores were used to make plots of all atoms sets of the ss-DNA. A qualitative comparison of the data was made.

4. **Results and Discussion**

A visual inspection of Poly-Cytosine ss-DNA reveals that the molecule conforms onto the CNT in a counterclockwise helical way[2] (Figure 2).

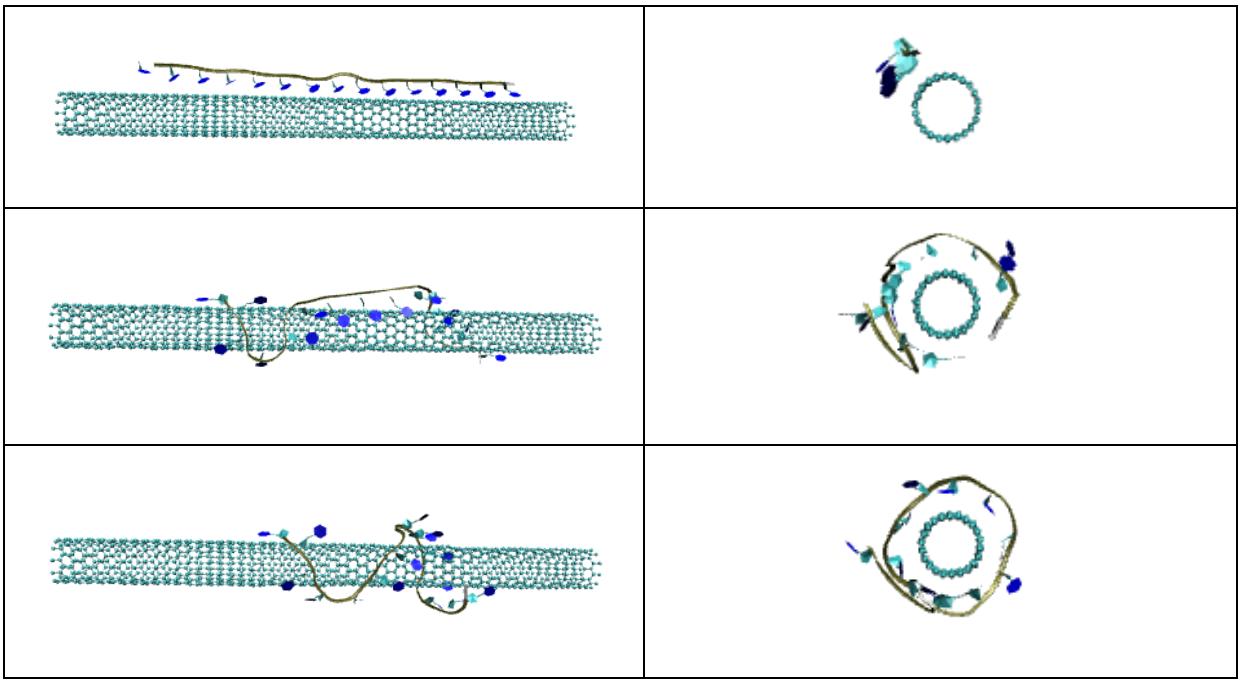
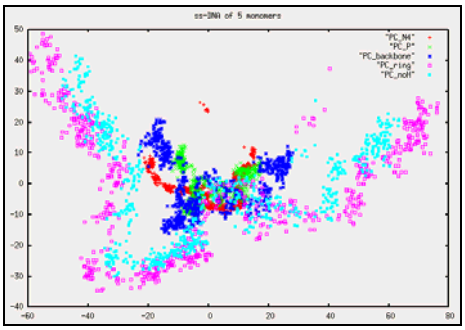
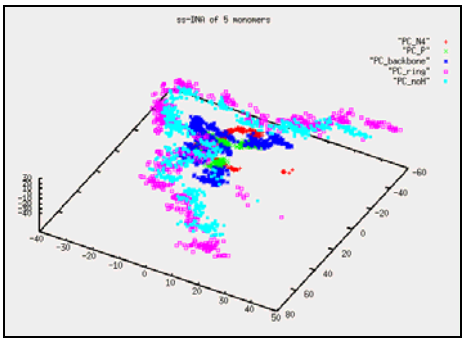
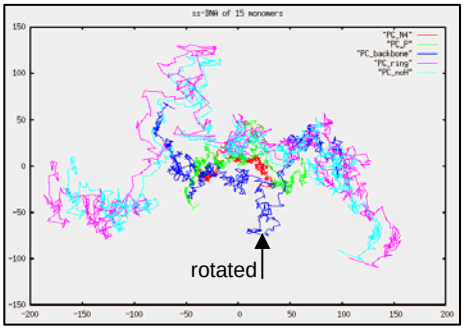
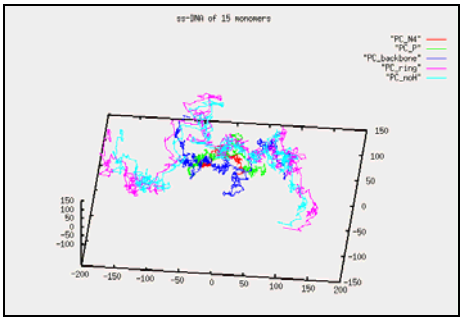


Figure 2. Snap shots of Poly-Cytosine ss-DNA of 15 monomers during the MD simulation.

Figure 3 shows the 2D and 3D PCA reductions of all Poly-Cytosine ss-DNA. In this figure the red line corresponds to N4 atoms set, green corresponds to P atoms set, blue corresponds to backbone atoms set, pink corresponds to ring atoms set, and light blue corresponds to the set of all atoms except for Hydrogen.

Monomers	2D	3D
5		
15		

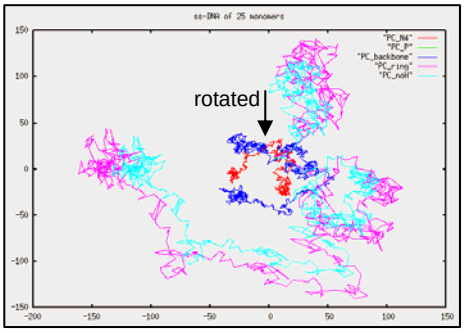
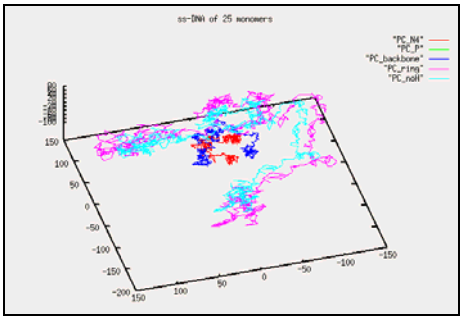
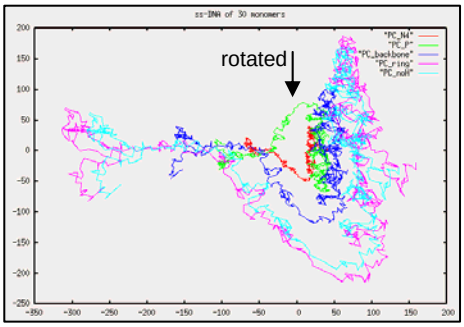
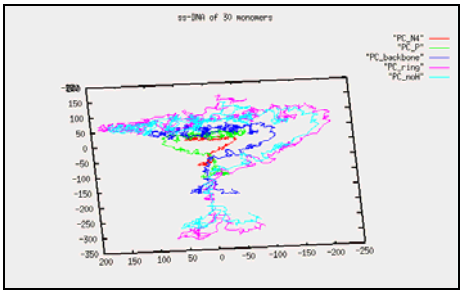
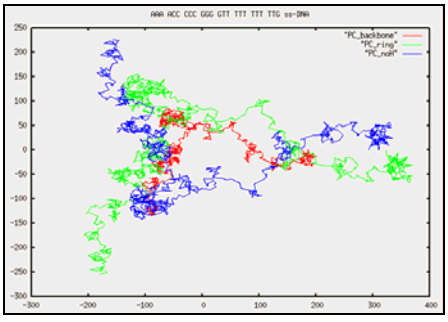
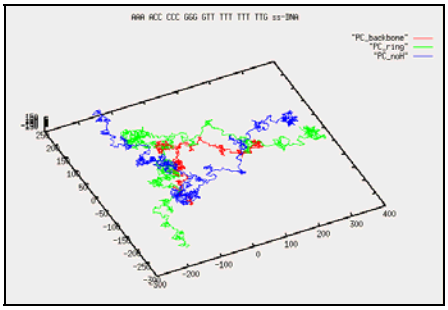
25		
30		

Figure 3. Plots of scores of PCA of all Poly-Cytosines.

In these plots (Figure 3) clusters of points indicate stable configurations of the system. In all cases the clusters around the point that marks the final frame of the simulation are notably densest of all. Moreover, the absence of multiple paths between clusters and of any return path to previous clusters demonstrates that those last conformations are stable and the process is irreversible within the time frame of the simulation. This phenomenon is independent of the length of the Poly-Cytosine.

In Figure 3 is easy to see that the shapes of the lines are similar for all atoms sets. For example, the 2D and 3D plots of Poly-Cytosine of 30 monomers show that the shapes of the lines are almost the same. Also, the set of *P* atoms is rotated compared with the other four sets. The orientation of lines can change occasionally but that does not affect the interpretation of the results. Another important result is that the sets that contain more atoms show that the line in the plot is more extended in comparison with the smallest sets. This is related with the way in which the PCA is computed. In Poly-Cytosine of 30 monomers, the set of all atoms except for Hydrogen and the set of the ring atoms are the biggest ones. This is consistent with all Poly-Cytosine ss-DNA. The smallest set for all ss-DNA is the N4 atoms set. The red line in the plot confirms that because is the most compressed line in each plot. These observations are consistent for all Poly-Cytosine ss-DNA. It is essential to see that although they are some of the lines rotated, compressed or extended that does not affect our results because this study is based on searching for clusters to determinate that the molecule has stable conformations.

Visual inspection of the system of the specific sequences of ss-DNA and CNT shows the ss-DNA conforming onto the CNT in a way similar to Poly-Cytosine. In Figure 4 the red line corresponds to the set of backbone atoms, green corresponds to ring atoms set, and blue corresponds to the set of all atoms except for Hydrogen.

Sequence	2D	3D
1		

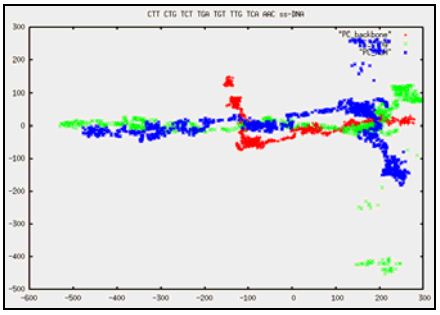
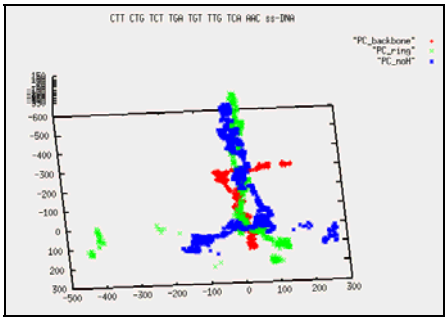
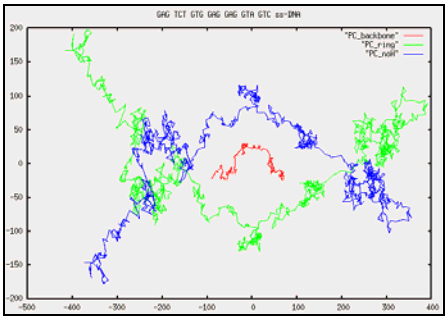
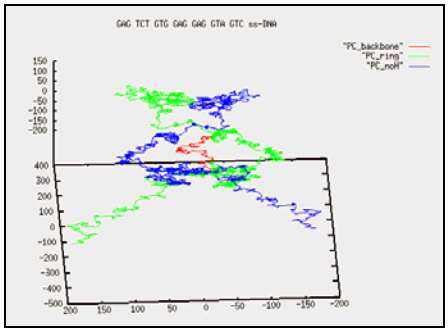
2		
3		

Figure 4. Plots of scores of PCA of the specific sequences.

The PCA depicted in Figure 4 shows patterns that demonstrate that the last conformations of the molecule are stable, and irreversible within the time frame of the simulation for the specific sequences of ss-DNA. The shape in each sequence is conserved no matter the atoms sets. As was in the Poly-Cytosine for a big set of atoms the lines were extended, as well for small sets the lines were compressed. They are lines rotated but this does not affect the results because the shape of the lines is conserved. Some of these plots show the trajectory of the molecule since the simulation was start. Examples for this observation are the sequence 1 and 2, here in the 2D plots the start is at the left side of the plots and the clusters at the right side represent a stable configuration for this ss-DNA. Also, there are no jumps in none of the plots.

5. **Conclusions**

The PCA analysis shows that the paths of the 3D points progress to eventually form clusters indicating that the shapes of the DNAs evolve into stable configurations. This is consistent no matter which atoms set its plotted. No returns in the paths nor jumps between clusters indicate that the DNA reaches a final conformation. These results are consistent with observations made in the laboratory proving that the PCA method is suitable for the study of the conformation of CNT-DNA hybrids.

6. **Future work**

Currently we are working on a Graphical User Interface (GUI) to setup, run and analyze systems of Polymer-CNT hybrids. This study justifies the inclusion of the PCA into this application.

7. **Acknowledgments**

This work was supported by Penn-UPR Partnership for Research and Education in Materials project (NSF-DMR-353730), Humacao Undergraduate Research in Mathematics to Promote Academic Achievement program (NSA-H98230-04-C-0486) and Minority Access to Research Careers program (NIH- 5T34GM008156-18). The author of this paper wants to thank to Robert Johnson and A.T. Johnson from the University of Pennsylvania for the introduction to the problem of MD Simulations, and to Robert Jernigan from Iowa State University for showing some of the analysis techniques used in this paper.

8. **References**

[1]Staii C., Johnson A.T., Chen, M., Gelperin, A., "*DNA-Decorated Carbon Nanotubes for Chemical Sensing*", Nano Letters, 2005, 5(9), 1774-1778.
[2]Johnson, R., Johnson, A.T., Klein, M., "*Probing the Structure of DNA-Carbon Nanotube Hybrids with Molecular Dynamics Simulations*", American Physical Society, APS March Meeting, March 5-9, 2007.
[3]Johnson, A.T., Staii, C., Chen, M., Khamis, S., Johnson, R., Klein, M., Gelperin A., "*DNA-decorated carbon nanotubes for chemical sensing*", 2006 Semicond. Sci. Technol. 21 S17-S21.
[4]Garcia, A.E., "*Large-amplitude nonlinear motions in proteins*", Physical Review Letters, 68 (17). 2696-2699, 1992.
[5]Yang, L., Song, G., Carriquiry, A., Jernigan, R.L., "*Close Correspondence between the Motions from Principal Component Analysis of Multiple HIV-1 Protease Structures and Elastic Network Modes*", Structure, Volume 16, Issue 2, 321 – 330.
[6]Altis, A., Nguyen, P., Hegger, R., Stock, G., "*Dihedral angle principal component analysis of molecular dynamics simulations*", The Journal of Chemical Physics, Vol. 126, No. 24. (2007).
[7]Grossfield, A., Feller, S., Pitman, M., "*Convergence of Molecular Dynamics Simulations of Membrane Proteins*", PROTEINS: Structure, Function, and Bioinformatics 67:31–40 (2007)
[8]Zheng, M., *et.al*, "*Structure-Based Carbon Nanotube Sorting by Sequence-Dependent DNA Assembly*" Science 28 November 2003: Vol. 302. no. 5650, 1545 – 1548.
[9]Teodoro, M., Phillips, G.N. Jr., Kavraki, L.E, "*A dimensionality reduction approach to modeling protein flexibility*", In Proc. ACM Int. Conf. on Computational Biology, 2002, 299-308.
[10]GROMACS, http://www.gromacs.org/gromacs/features/feature-summary.html.
[11]Merced M., "*Metrics for the study of DNA-CNT hybrids*", Proceedings of the National Conference on Undergraduate Research (NCUR 2007).
[12]Merced M., "Automatization of a Molecular Dynamics Simulation and the Evaluation of Metrics for the study of DNA-CNT Hybrids", PREM Technical Report, December 2006.
[13]Tinker Molecular Modeling Package, http://dasher.wustl.edu/tinker/.
[14]The Amber Molecular Dynamics Package, http://amber.scripps.edu/.
[15]Humphrey, W., Dalke, A., Schulten, K., "*VMD- Visual Molecular Dynamics*", J. Molec. Graphics 1996, 14.1, 33-38.
[16]Visual Molecular Dynamics, http://www.ks.uiuc.edu/Research/vmd/.
[17]The Math World- MATLAB and Simulink for Technical Computing, http://www.mathworks.com/.