Metrics for the Study of DNA-CNT Hybrids and a Prototype of MoSDAS Graphics User Interface

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

DNA-carbon nanotube (DNA-CNT) hybrids have many useful properties important in the field of nanoscience. Molecular dynamics (MD) simulations are well-suited to provide insights into the fundamental properties of DNA-CNT hybrids because they enable calculation of structural properties with atomic resolution. Radial distribution functions are often used to detect structural properties of liquids and crystalline materials in MD simulations. In this work other metrics for measuring how well ss-DNA conforms onto CNT were defined, tested, and compared to radial distribution functions. Three measures were defined as the maximum, minimum and average of the molecule separation distances intended to measure how close and tight the ss-DNA wraps around the CNT. An axial distribution function with respect to the z-axis was also defined. It is based on an expanding cylinder centered in the symmetry axis of the CNT instead. A simple ss-DNA (Poly-C) was used for testing purposes. A complex process of model building, simulation and data analysis was completely automated as a script (MoSDAS). Simulations were performed for Poly-C's with varying number of bases. Both of the metrics defined above as well as two radial distribution functions provided by the GROMACS MD package were compared. It was found that MoSDAS greatly simplifies the simulation process and avoids errors. Also, the molecule separation based metrics provides better information than the other measures on when the ss-DNA wraps around the CNT. The axial distribution function was best suited as an indicator of how the DNA atoms conform around the CNT. Analysis of this function shows, for example, a shift to larger CNT-DNA distances as the ss-DNA length increased. A two-peak distribution suggests that different parts of the Poly-C agglomerate at different levels. This information is not evident in the usual radial distribution functions.
Keywords: Molecular Dynamics, DNA-CNT hybrids, Distribution Function

1. Introduction

Carbon nanotubes (CNT) and single stranded DNA (ss-DNA) are both interesting and important systems in nanoscience. For example, ss-DNA-CNT hybrids have been recently used to construct nanoscale chemical sensors.[1] ss-DNA has also been used to separate and sort CNTs based on metallicity.[2] They are completely compatible with one another. Molecular dynamics (MD) simulations are well-suited to provide insights into the fundamental properties of DNA-CNT hybrids because they enable calculation of structural properties with atomic resolution.

It is understood that ss-DNA attaches readily to the CNT by the $\pi - \pi$ stacking interaction. However, these have not yet been studied in detail to determine the specific conformation of ss-DNA about a CNT. Several MD simulations of Poly C ss-DNA adsorbing to a CNT have been done both at University of Pennsylvania and at the University of Puerto Rico at Humacao. The analysis of the results has been mostly quantitative. This work presents the results of a comparison of different quantitative metrics that can be used to study them.

2. Background

2.1. carbon nanotubes

Carbon nanotubes are cylindrical sheets of carbon with diameters of ~1nm and lengths up to a few centimeters.[3]

They are unique for their size and special properties. CNTs have electronic and structural properties that vary depending on the diameter, chirality and length. Nanotubes can form both single-wall (SWNT) and multiple-wall (MWNT) structures.[3] They have many interesting properties such as high mechanical strength (nanotubes can have tensile strengths 60 times larger than steel) and electronic stability (nanotubes can accommodate current densities 1000 times higher than copper and silver). These features make them possible candidates for practical applications.

2.2. single stranded DNA

Single stranded DNA is a DNA molecule consisting of only one chain of alternating sugars and phosphates. They are often represented by sequences of the letters C, A, T, and G that correspond to the different bases unit. For this work, ss-DNA composed of a repeating sequence of one of the bases, Cytosine (Poly-C ss-DNA) was simulated.

2.3. measuring molecule separation

Three distances between the Poly-C and the CNT were defined. These distances are based on the point-to-set distances in metric spaces.

Definition. Let $a,b \in \Re^3$ and $D,C \subset \Re^3$, D and C finite sets. The *Euclidean distance* between $a = (x_a, y_a, z_a)$ and $b = (x_b, y_b, z_b)$ is $d(a,b) = \sqrt{(x_a - x_b)^2 + (y_a - y_b)^2 + (z_a - z_b)^2}$. The *distance* between a and C is $r_{a,C} = \min\{d(a,b) \mid b \in C\}$[5]. The *maximum*, *minimum* and *average separation* between D and C is $r_{D,C}^{\min} = \min\{r_{a,C} \mid a \in D\}$, $r_{a,C}^{\max} = \max\{r_{a,C} \mid a \in D\}$ and $r_{D,C}^{ave} = \frac{1}{|D|}\sum_{a \in D} r_{a,C}$ respectively.

In our application C and D correspond to the sets of atoms of the CNT and Poly-C.[4]

2.4. distribution functions

Three kinds of measures of densities of atoms are compared in this work. They fall within the two following categories.

2.4.1. *radial distribution functions in MD*

In Molecular Dynamics the term distribution function refers to the densities of points as the separation to a reference set is varied. They are based in how many atoms of a given type can be found at the distance r away from the points in the reference sets. Distribution functions are often used to detect structural properties of liquids and crystalline materials in MD simulations. Two kinds of radial distribution functions are for example with respect to a set of atoms and with respect to the center of mass of a set of atoms. They are useful as validation of simulations because their results resemble x-ray diffraction experiments. The following definition uses a formulation of the notion of radial distribution that is useful for computational purposes.

Definition. Let $C,T \subset \Re^3$, C and T finite sets. The *radial distribution function* with bin size Δr is defined by

$$R_{CT}(r) = \frac{V}{4\pi r^2 \Delta r M N_C N_T} \sum_{m=1}^{M} \sum_{i=1}^{N_C} \sum_{j=1}^{N_T} Q_m(r; r_{C_i T_j}) \qquad (1)$$

where V is the volume of the smallest sphere that contains all atoms in the sets C and T, M is the total number of frames; N is the total of atoms of a given type within the volume elements, and Q_m the classification function

$$Q(r;r_{C_iT_j}) = \begin{cases} 1 & \text{if } r - \frac{\Delta r}{2} \leq \quad r_{C_iT_j} < r + \frac{\Delta r}{2} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where r is the radial distance and $r_{C_iT_j}$ is the distance between atoms C_i and T_j.[6]

Figure 4 shows the geometric distribution of the atoms. Where R_g is the reference group and the small disks represents atoms. A ring centered on the reference is drawn with radius r and thickness Δr.[7]

In Formula 1 the term Q_m classifies the atoms in two groups. The atoms that are in the ring at distance r of the reference group and the other ones. The sums are for counting the atoms that are in the ring at distance r of the reference group. The term $\frac{1}{4\pi\ r\Delta r}$ normalizes the geometric density of all the atoms in the system. Finally, the term $\frac{V}{r\,M\,N_C\,N_T}$ normalizes the measurements, so the final measure of all the systems is 1. That is useful for comparison different systems.

When the radial distribution function is computed with respect to the center of mass of the system the set C can be taken to be this point in space and (1) becomes

$$g(r) = \frac{V}{4\pi\ r^2\Delta r MN}\sum_{m=1}^{M}\sum_{i=1}^{N} Q_m(r;r_i).$$

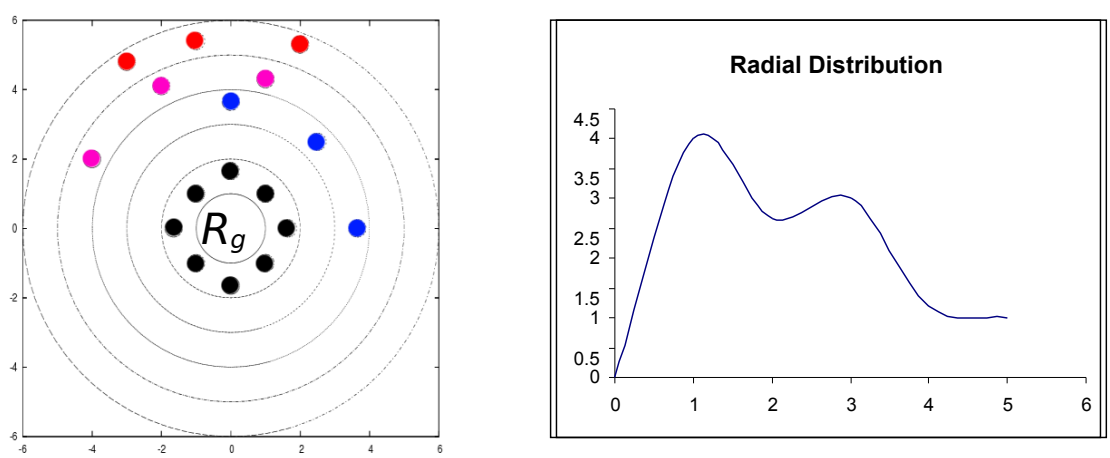


Figure 4. Graphical representation of geometric distribution.

2.4.2. *axial distribution function*

The axial distribution function is the distribution of the atoms of a given type based on an expanding cylinder centered on the reference axis. It is analogous to the radial distribution function described above with an infinite reference set (the axis of reference). It is a probability distribution measure.

Definition[5]**.** A function $G:\Re \rightarrow \Re$ is a *probability distribution function* if

i. G is non-decreasing ,

ii. G is left continuous, $\lim_{\substack{y \to x \\ y < x}} G(y) = G(x), \text{ all } x \in \Re$,

iii. $G(-\infty) = \lim_{x \to -\infty} G(x) = 0, \quad G(x) = \lim_{x \to \infty} G(x) = 1.$

If there is a function $g(x)$ such that $G(x) = \int_{-\infty}^{x} g(t)dt$ then $g(x)$ is called the *density function* corresponding to G.

If the function $G(x)$ satisfies $G(x) = \sum_{j:x_j < x}^{N} p_j$ where $p_j > 0$, $\sum_{j=1}^{N} p_j = 1$ and $S = \{x_j : 1 \le j < n+1 \le \infty\}$ is a subset of $(-\infty, \infty)$ then it is called *discrete*. The associated probability density function is

$$f(x) = \begin{cases} p_j & \text{for } x = x_j \\ 0 & \text{for } x \neq x_j \end{cases}.$$

For the particular application presented here the function $G(x)$ is defined by

$$G(r) = \left(\frac{1}{N} \sum_{i=1}^{N} \delta[r - r_i] \right) \tag{3}$$

where N is the total number of atoms of a given type within the volume elements, r_i is the distance between the atom i and the axis of reference and $\delta[x] = \begin{cases} 1 & \text{if } x \ge 0 \\ 0 & \text{if } x < 0 \end{cases}$. This $G(r)$ is a discrete distribution function. But instead of using the associated density function for this discrete distribution an *axial density function* is defined at the values $\{0 \le r_1 < r_2 < r_3 < \ldots < r_n\}$ by setting $\Delta r_i = r_{i+1} - r_i$ and

$$g(r_i) = \frac{G(r_i + \Delta r_i) - G(r_i)}{\Delta r_i} = \frac{G(r_i + (r_{i+1} - r_i)) - G(r_i)}{r_{i+1} - r_i} = \frac{G(r_{i+1}) - G(r_i)}{r_{i+1} - r_i}. \tag{4}$$

In this work the axial density function is referred by *axial distribution function* to agree with the way it is called in the Molecular Dynamics as pointed out in section 2.4.1.

3. Software and methods

In this work, the performance of molecular dynamics simulations using the **GROMACS** MD package was done.[8] The ss-DNA structure was obtained from the *nucleic* program, which is part of the **Tinker** molecular modeling package.[9] Also, *tleap* program was used. *Tleap* is a sub-program of the **AMBER7** molecular dynamics package.[10] The results of the simulations were visualized with **VMD**.[11] In the following sections, is the outline of the procedure for setting up and running the simulations. Also, is the comparison of the results between two programs developed as part of this work (molecule separation and axial distribution function) and *g_rdf* program that is part of **GROMACS**

MD package.

3.1. MoSDAS

The Model building, Simulation and Data Analysis Script (**MoSDAS**) was developed to automate the production of the complete system. **MoSDAS** was made in bash programming language. The main purpose of **MoSDAS** is to call and run other programs.

3.2 MoSDAS GUI

MoSDAS GUI is a Graphics User Interface made with the Tkinter application of the Python programming language. It has a menu, buttons, radio buttons, text fields, and labels. Figure 5 shows a picture of the prototype of MoSDAS GUI.

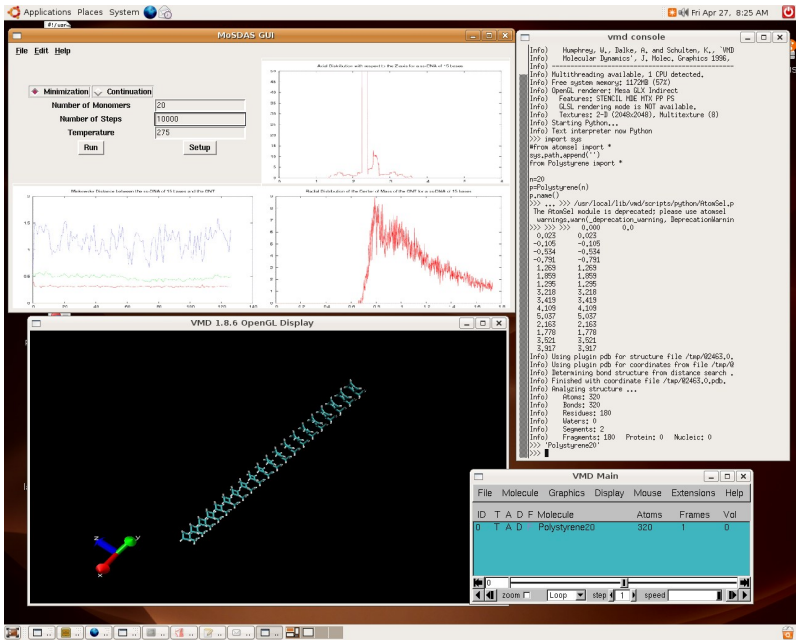


Figure 5. Prototype of MoSDAS GUI.

MoSDAS GUI run as a VMD plugin. Figure 5 shows the windows of MoSDAS GUI, VMD console, VMD main, and VMD display.

3.3 simulations

The molecular dynamics simulation was done with the **GROMACS** MD package. About eight simulations were run. The only variation in the simulations was the number of monomers of the ss-DNA. The obtain simulation data was of Poly-C ss-DNA of 5, 10, 15, 19, 20, 25, and 30 monomers.

3.4. metrics

Here are three different metrics to analyze the data. Those metrics are: the molecule separation, the radial distribution and the axial distribution.

3.3.1. *molecule separation*

For the molecule separation a computer program was developed. This program was written in C++ programming language. The program use the complete trajectory of the atoms to calculated the separation of ss-DNA from the

CNT.

The distance between an atom (name *N4* in the data files) in the ring of the ss-DNA and all the CNT atoms was calculated. The minimum distance of each monomer was stored. Then, the minimum and maximum distances were selected and the average was calculated.

3.3.2. *radial distribution functions*

For the calculation of the radial distribution function the program *g_rdf* that is part of the **GROMACS** package was used. The default measure of this program is with respect to a set of atoms. To use the measure with respect to the center of mass, the `-com` option was used. In both measures the CNT atoms was used as the reference group and the ss-DNA as the `group1`. Also, they are calculated with the complete trajectory of the atoms.

3.3.3. *axial distribution function*

For the calculation of the axial distribution function a computer program was developed. This program is in C++ programming language. The z-axis was used as the reference axis because the CNT is inside of it. The axial distribution of Poly-C atoms was measured with the last frame of each simulation.

The distance between the coordinates (x, y) of Poly-C was measured and stored in a vector. Then the vector was sorted in an ascending position (calculation of $G(r)$). Next the program prints out the distance and the portion of the system that have the atoms at a respective distance of the axis of symmetry.

The following pseudo-C++ code stores a table of values of the probability distribution function $G(r_i)$ in the vector G[i]:

```
valarray <double> r = readAtomCoordinatesAndStoreDistances();
int N = r.size();
sort(r.begin(), r.end());
for(int i=0; i < N; i++)
      G[i] = (double)i / N;
```

Note that, since the distances r_i in `r` have been sorted, the formula `i/N` computes the portion of the atoms that have distances less than `r[i]`.

Appending the following code to the previous one completes the computation of $g(r_i)$ defined in (2):

```
for(int i=0; i < N-1; i++)
      G[i] = (G[i+1] - G[i]) / (r[i+1] - r[i]);
```

Since this approximation of $g(r)$ produces exaggerated fluctuations, the final plot was done with averages of the values of $g(r_i)$ around $g(r_i)$

3.4 post processing

When the results of all the metrics are complete, a qualitative comparison of the developed metrics and the **GROMACS** metrics (radial distribution functions) was done. A comparison with the visual representation of the system revealed which metrics give more information about the specific problem.

4. Results and Discussion

4.1 molecule separation distances

Molecule separation distances between Poly-C and CNT shows how close and tight the Poly-C wraps completely around the CNT. Also, they show is the system is equilibrated or not. Figure 6 depicts the molecule separation measures of Poly-C (30).

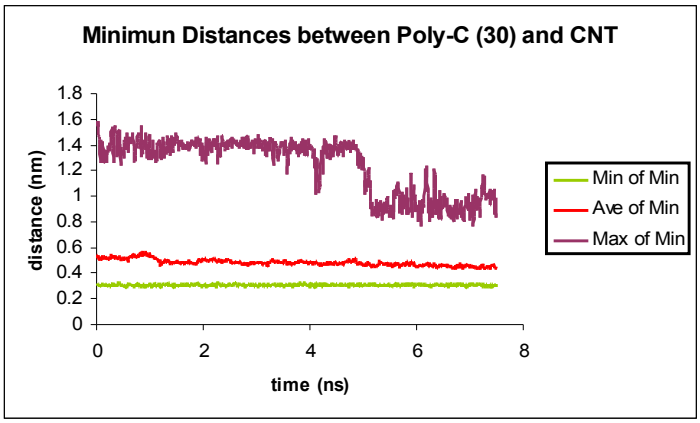


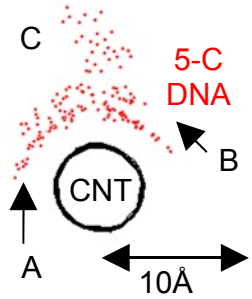
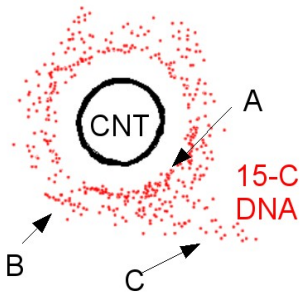
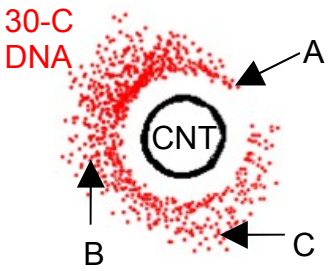
Figure 6. Molecule separation of Poly-C (30)

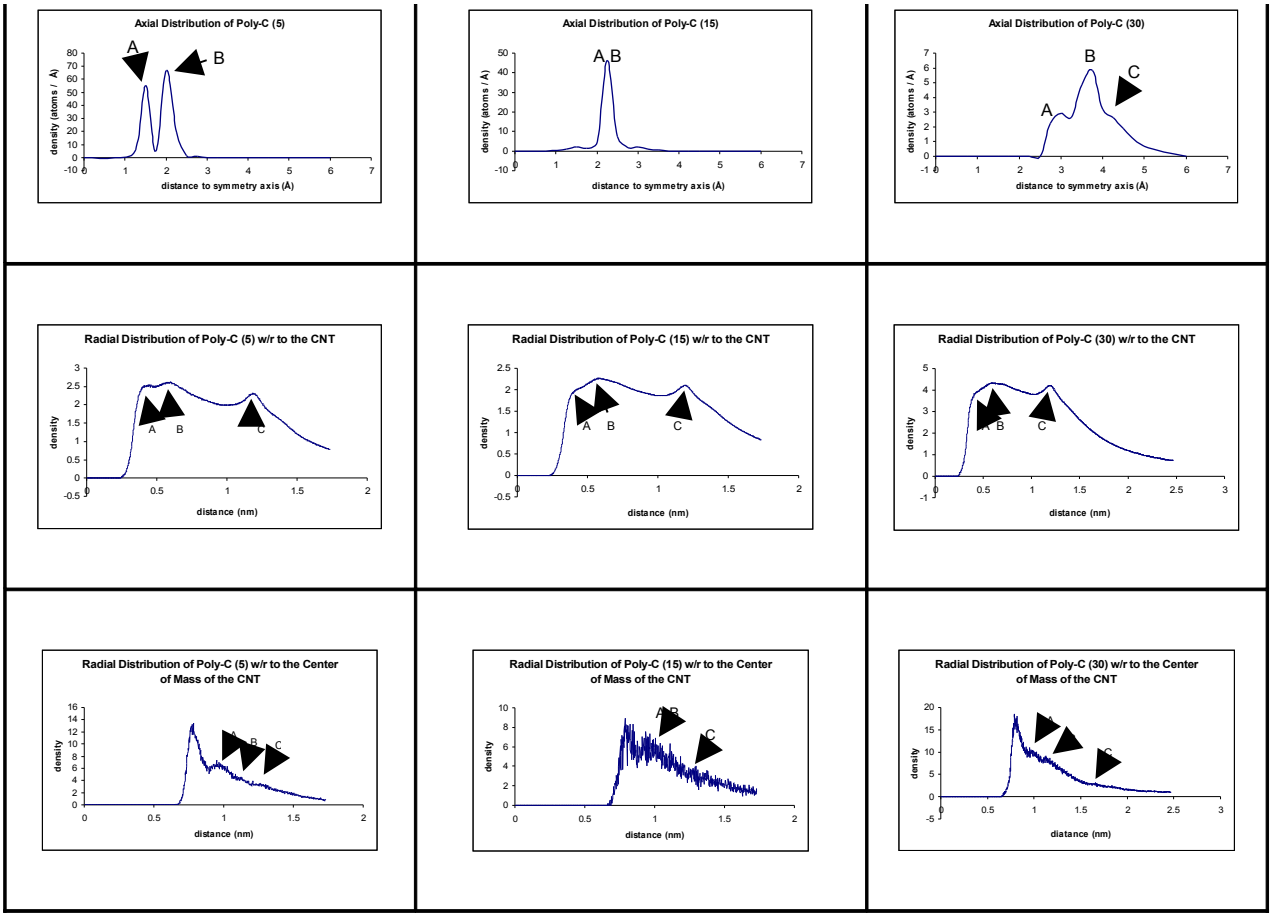
Figure 6 shows a descended step at approximately 5 ns of simulated time. This descent of the maximum distance is an indicator that the system is coming to the equilibrate status. Before this moment the user should infer that the simulation needs to be executed for a longer time. High steps indicate that there are stacks of Cytosine bases due to the $\pi - \pi$ attraction.

4.2 distribution functions

Table 1 depicts the results for each of the distributions that were calculated. They are examples of Poly-C consisting of 5, 15, and 30 monomers.

Table 1. Results for Poly-C of 5, 15, and 30 monomers.

Poly-C (5)	Poly-C (15)	Poly-C (30)
		



In the first row of Table 1 are pictures of the last position of atoms in Poly-C and CNT in the symmetry axis view (Z-axis). These pictures show atoms agglomeration in different groups or levels (A, B, and C) with different densities.

Table 1-second row, shows the density of the atoms of Poly-C ss-DNA with respect to the Z-axis. This measurement was calculated with respect to the Z-axis because Z-axis is the axis of symmetry for the CNT. A two-peak distribution suggests that different parts of the Poly-C agglomerate at different levels. They also show a shift to larger CNT-DNA distances as the Poly-C ss-DNA length increased. This information is not evident in the usual radial distribution functions.

Next in Table 1 is the radial distribution of Poly-C with respect to the CNT. This measure calculates the density of the atoms with respect to the CNT (a group of atoms) in terms of the distance between them. In last row are the radial distributions of Poly-C with respect to the center of mass of the CNT. The only difference between these two radial distributions is the reference group. In the first one the reference group is a group of atoms and in the second one is a point of reference (this point is calculated by the program *g_rdf*).

5. Conclusions

It was found that the development of **MoSDAS** simplifies and avoids the most of the errors in the simulation process. The molecule separation distances provides better information than the other measures on when the ss-DNA wraps around the CNT, and give an idea of how close and tight is the ss-DNA from the CNT. The molecule separation distances can also be used to determine when the system is in the equilibrated stage. The axial distribution function was best suited as an indicator of how the DNA atoms conform around the CNT.

6. Future work

Development of a Graphical User Interface (actually in construction) that will help the user execute the automatization script, the molecular dynamics simulation and the post processing phase in which the analysis of the results will be done.

7. Acknowledgments

The author wish to express their appreciation to 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), UPRH RISE and National Science Foundation's Computer Science, Engineering and Mathematics Scholarships program (NSF-0123169).

8. Reference

1. Staii C, et al, "DNA-Decorated Carbon Nanotubes for Chemical Sensing," *Nano Letters*, (2005): 1774-1778.
2. Zheng, M, et al, "Structure-Based Carbon Nanotube Sorting by Sequence-Dependent DNA Assembly," *Science* (2003): 1545-1548.
3. Johnson Group: Experimental Nanoscale Physics-Resources, "Introduction to Carbon Nanotubes", University of Pennsylvania, http://www.lrsm.upenn.edu/~nanophys/nanotube.html.
4. Cramer, C, *Essentials of Computational Chemistry*: *Theories and Models*, 2nd Ed. (England: John Wiley & Sons, 2004), 86.
5. Shih, Y and Teicher H, *Probability Theory: Independence, Interchangeability, Martingales*, 2nd Ed. (New York: Springer-Verlag, 1978), 25-29.
6. Rapaport, D, *The Art of Molecular Dynamics Simulations*, 2nd Ed. (Cambridge: University Press, 1995), 90.
7. Radial distribution function, "Ljmd: Background," http://www.ccr.buffalo.edu/etomica/app/modules/sites/Ljmd/Background1.html.
8. GROMACS, "Feature Summary," 2005, http://www.gromacs.org/gromacs/features/feature-summary.html.
9. Tinker Molecular Modeling Package, "Tinker Home Page: Introduction," http://dasher.wustl.edu/tinker.
10. The Amber Molecular Dynamics Package, "Assisted Model Building with Energy Refinement," http://amber.scripps.edu.
11. Visual Molecular Dynamics, "Theoretical and Computational Biophysics Group," http://www.ks.uiuc.edu/Research/vmd.