AUTOMATIZATION OF A MOLECULAR DYNAMICS SIMULATION AND THE EVALUATION OF METRICS FOR THE STUDY OF DNA–CNT HYBRIDS

PREM Technical Report, December, 2006

Myrna I. Merced Serrano
Department of Mathematics
University of Puerto Rico at Humacao
100 Tejas Avenue
Humacao Puerto Rico 00791
mymese@mate.uprh.edu

Faculty Adviser: José Sotero Esteva, A.T. Charlie Johnson
Student Adviser: Robert Johnson

Keywords: Molecular Dynamics, DNA-CNT hybrids

Abstract

DNA-carbon nanotube (DNA-CNT) hybrids have many useful properties important in the field of nanoscience. Molecular dynamics 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.

1. Introduction

Carbon nanotubes (CNT) and single stranded DNA (ss-DNA) are both interesting and important systems in nanoscience that are completely compatible with one another. Molecular dynamics 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. 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 metallicity[2].

It is understood that ss-DNA attaches readily to the CNT by the π-π stacking interaction. However, these has not yet been a detailed study 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 UPenn and UPRH. The analysis of results have been mostly quantitative. Here, we present the results of a comparison of different quantitative metrics that can be used to study them.

2. **Background**

2.1. **carbon nanotubes**

CNT were discovered in 1991. They are cylindrical sheets of carbon. CNT have diameters of ~1nm and lengths up to a few centimeters[2]. They are unique for their size and special properties. CNT 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 accomodate current densities 1000 times higher than copper and silver). These features make them possible candidates for practical applications. The figure 1 shows an example of the different forms of CNTs.

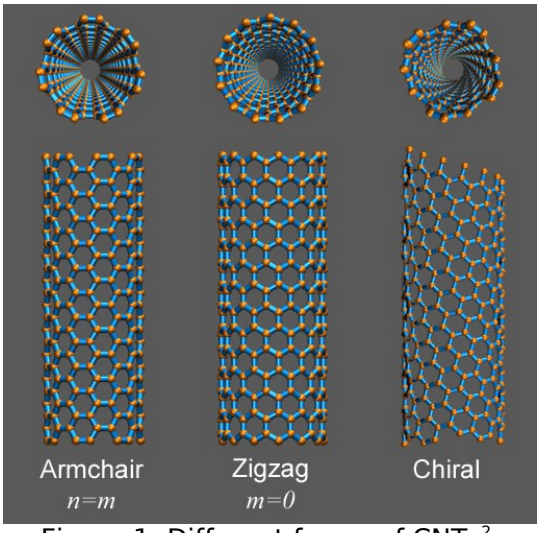

Figure 1. Different forms of CNTs[2].

2.2. **single stranded DNA**

Single stranded DNA (ss-DNA) is a DNA molecule consisting of only one chain of alternating sugars and phosphates[4]. They can assume different structures depending on the solvent and ionic environment. Figure 2 depicts the A structure and Figure 3 depicts the B structure.

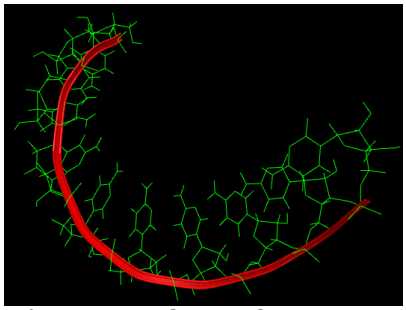

Figure 2. A form of a ss-DNA

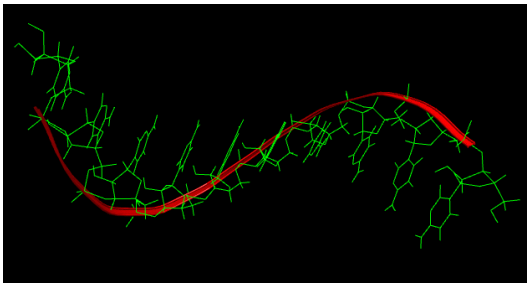

Figure 3. B form of ss-DNA

For this work, we used a ss-DNA composed of a repeating sequence of cytosines (Poly-C ss-DNA).

2.3. **molecule separation measures**

We define three distances between the Poly-C and the CNT. They are based on the molecule separation distances in metric spaces. As usual, atoms are considered as points in a 3D euclidean space. Since, we are considering finite sets of points the molecule separation distances between a point *a* and a finite set *T* is

$$r_{a,T} = min\{r_{a,b} \mid b \in T\}$$ [5].

Let *C* be another finite set. Then we define the maximum, minimum and average separation between *C* and *T* as

$$r_{C,T}^{min} = min\{r_{a,T} \mid a \in C\}\ ,\ r_{C,T}^{max} = max\{r_{a,T} \mid a \in C\}$$ and $$r_{C,T}^{ave} = \frac{1}{|C|}\sum_{a\in C} r_{a,T}\ .$$

2.4. **probability distribution functions**

Distribution functions measures the densities of points as the separation to a reference set is varied. This work used the radial distribution and the axial distribution functions. They are based in how many atoms of a given type I can find at the distance *r* away from it.

Figure 4 shows the probability of the distribution of the atoms. R_g is the is reference group and the color circles are the atoms of a given type. A ring centered on the reference is drawn with radius *r* and thickness *dr*[6].

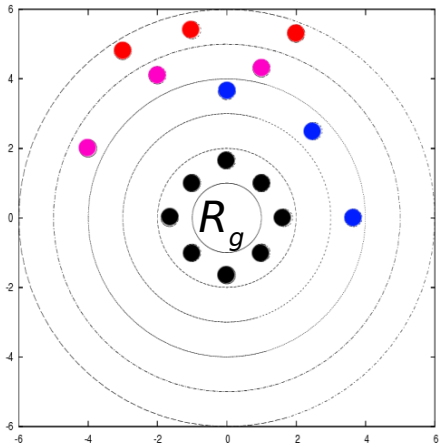

Figure 4. Graphical representation of the probability distribution functions.

2.4.1. *radial distribution function*

Radial distribution functions are often used to detect structural properties of liquids and crystalline materials in MD simulations. Usually two kinds of radial distribution function, 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 for the validation of simulations because their results resemble x-ray diffraction experiments.

A mathematical formulation for the radial distribution function $g_{AB}(r)$ with respect to a set of atoms is

$$\frac{1}{V}\ g_{AB}\ (r) = \frac{1}{N_A N_B}\left(\sum_{i=1}^{N_A}\ \sum_{j=1}^{N_B} \delta[r - r_{A_i B_j}]\right) \qquad (1)$$

where *V* is the spherical volume that contains all atoms in the sets *A* and *B*, *N* is the total number of atoms of a given type within the volume elements, δ is the Dirac delta function, *r* is the radial

distance and $r_{A_iB_j}$ is the distance between atoms A_i and B_j [7].

The Dirac delta function has the value of infinity for x = 0 and the value zero elsewhere. A formula for the Dirac delta function is

$$\delta(x)=\begin{cases}\infty, & x=0\\ 0, & x\neq 0\end{cases}$$

and which is also constrained to satisfy the identity

$$\int_{-\infty}^{\infty}\delta(x)\,dx \;=\; 1 \quad .$$

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

$$\frac{1}{V}\;g(r) \;=\; \frac{1}{N}\left(\sum_{i=1}^{N}\delta[r \;-\; r_i]\right) \;.$$

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).

A mathematical formulation for the axial distribution function $g_{AB}^{axial}(r)$ with respect to a set of atoms is

$$\frac{1}{V}\;g^{axial}(r) \;=\; \frac{1}{N}\left(\sum_{i=1}^{N}\delta[r \;-\; r_i]\right)$$

where *V* is the cylindrical volume that contains all atoms, *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.

If we adapt the method for computing the radial distribution function presented by Rapaport (pag 91) to this axial distribution we obtain the following. Let c be the length of the nanotube. If h_n is the number of atom pairs (i, j) for which $(n-1)\Delta r \le r_{i,j} < n\Delta r$, N_m is the number of atoms we obtain the approximation

$$g(r_n)=\frac{Vh_n}{\pi c r_n^2 N_m^2 \Delta r}$$

where $r_n=\left(n-\frac{1}{2}\right)\Delta r$ [8].

3. **Software and methods**

In this work, we perform molecular dynamics simulations using the **GROMACS** MD package[9]. We obtain ss-DNA structure from *nucleic* program, which is part of the **Tinker** molecular modeling

package[10] . This program generates the coordinates of ss-DNA of a given sequence and structure. Three pre-defined structures are available (A, B, Z) in this program. We also used *tleap,* which is a sub-program of the **AMBER7** molecular dynamics package[11]. This program enables the user to edit configurations of existing molecules, create structures for new ones and generate *topology* and *coordinate* files. The results of our simulations were visualized with **VMD.** The program **VMD** is a molecular visualization program for displaying, animating and analyzing large biomolecular systems using 3-D graphics and built-in scripting[12]. In the following sections, we outline the procedure for setting up and running our simulations. For analyzing the results we used two programs developed by us (molecule separation and axial distribution function) and *g_rdf* a program of **GROMACS** MD package.

3.1. **MoSDAS**

The Model building, Simulation and Data Analysis Script (**MoSDAS**) were developed to automate the production of the complete 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. **MoSDAS** first creates and edited the ss-DNA files, then the CNT files and finally the complete system files for the setup of the simulations.

3.1.1 *production of the Poly-C ss-DNA*

The script first generates the Poly-C ss-DNA with a straight structure (created by us). For the generation of the Poly-C ss-DNA the *nucleic* program, a sub-program of **Tinker** was used. A script on Python programming language was developed to send the input information to *nucleic* program. Then the ss-DNA *pdb* file (coordinate file) was edited with *sed* program.

The coordinates of the ss-DNA were changed because later they has to be on top to CNT. To change the coordinates we used the *Tkconsole* of the **VMD** program. We have to calculated the translation and rotation matrices

$$\begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 & 0 \\ \sin(\theta) & \cos(\theta) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$ (z-axis) and $$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & \cos(\alpha) & -\sin(\alpha) & 0 \\ 0 & \sin(\alpha) & \cos(\alpha) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$ (x-axis)[13]

to move the coordinates. To get the angles of θ and α we selected the first and the last coordinates of the CNT and do some trigonometry.

Because the ss-DNA has negative charge the addition of Na+ ions with *tleap* was done to neutralize them. Also, the *topology* file of the ss-DNA was done with *tleap*, convert to **GROMACS** format with *Conv_7.x* program and edited with *sed*. The system need a CNT so, the length of the ss-DNA was measure for approximate the CNT length.

3.1.2. *production of the CNT*

The CNT was generated with the program *nanotubegen* by Robert Johnson. The CNT length was measure with a commands that were send to **VMD**. The CNT was centered in a box with a coordinates of 10 x 10 x *tube_length.* Then, the *topology* file of the CNT was generated with the *x2top* program, which is a sub-program of **GROMACS** and edited with *sed* program. The *position restraints* file of the CNT was done with an *awk* script.

3.1.3 *production of the complete system*

The Poly-C ss-DNA and the CNT were join with *tleap*. The system was placed inside of a box and hydrated with water. The water that was inside of the tube was removed with the *tcl* script *rem-wat-interior.tcl* by Robert Johnson. The *index* file is the file that classified the atoms by groups, for example WATER, DNA_20L, etc. These file was done with *make_ndx* program, a sub-program of **GROMACS**. Also, the *topology* file of the complete system was generated with an *awk* script. The *master input* file was edited from a template. The editions of these file are done with *shellPruebaInputI* program. The arguments of these program are the quantity of bases of the ss-DNA, if is a minimization, the command *min*; if not the command *cont* and the number of steps. Finally, **MoSDAS** run the simulation with *run_gromacs.sh* bash script.

3.2 **simulations**

The performance of molecular dynamics simulation were done with the **GROMACS** MD package. We ran about eight simulations. The only variation in he simulations is the length of the ss-DNA and we control the length with the quantity of bases. The other variables are in control. We have simulations of the Poly-C ss-DNA of 5, 10, 15, 19, 20, 25, 30 and 35 bases. Not all of then have been equilibrated.

3.3. **metrics**

We used 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 we developed a computer program. These program is in C++ programming language. The program use the complete trajectory of the atoms to calculated the distances and select the need ones.

We calculated the distance between the atom *N4* in the ring of the ss-DNA and the CNT atoms. For example, if the ss-DNA has 10 bases, we have 10 molecule separation distances. With those distances, the minimum and maximum distance were selected and the average of all of the distances was calculated. With those measures, we can have an idea how close and tight the ss-DNA wraps around the CNT. Also, we can know when the system is going to be equilibrated.

3.3.2. *radial distribution function*

For the calculation of the radial distribution function we used the program *g_rdf* that is part of the **GROMACS** package. 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, we have to use the *-com* option.

In both measures we used the CNT atoms 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 we developed a computer program. These program is in C++ programming language.

We used the z-axis as the reference axis because the CNT is inside of it. We measure the distribution of the ss-DNA atoms with respect to the z-axis on the last frame of each simulation.

3.4 **post processing**

When the results of all the metrics are complete, we compare the developed metrics with the **GROMACS** metrics (radial distribution functions). And we selected which ones give us more

information about our specific problem.

4. **Discussion of Results**

For the results of the simulations, we make graphs of the radial distribution function with respect to the CNT and with respect to the center of mass of the CNT. Also, we make graphs of the molecule separation and axial distribution.

Figure 6 shows the radial distribution of ss-DNA atoms with respect to the CNT atoms. Figure 7 shows the radial distribution function with respect to the center of mass of the CNT. Figure 6 and 7 are from a Poly-C ss-DNA of 15 bases. They are similar to the radial distribution function of Poly-Cs of length 5, 10, 19, 25 and 30 (not shown).

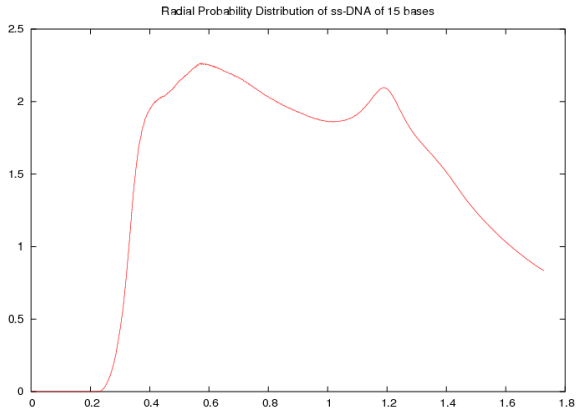

Figure 6. Radial distribution function with respect to the CNT.

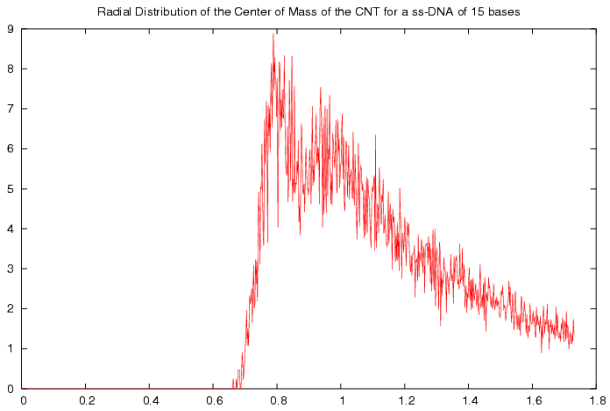

Figure 7. Radial distribution function with respect to the center of mass of the CNT.

The molecule separation measures shows when the Poly-C ss-DNA wraps completely around the CNT. Figure 8 shows the molecule separation measures of a Poly-C ss-DNA of 30 bases.

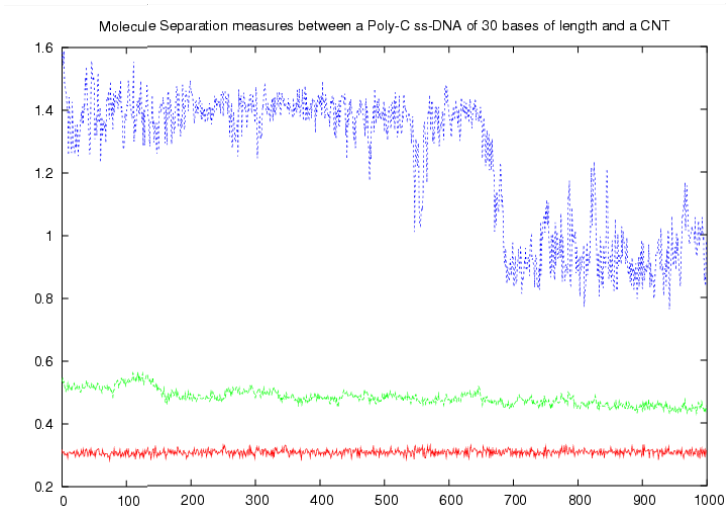

Figure 8. Molecule separation measures between the Poly-C ss-DNA and the CNT.

The axial distribution function graphs shows the distribution of the atoms in the Poly-C ss-DNA with respect to the the z-axis for Poly-C ss-DNA of 30, 25, 19, 15, 10 and 5 monomers. Agglomerations of the atoms at different levels are observed for some molecules. They also show a shift to larger CNT-DNA distances as the Poly-C ss-DNA length increased. Figure 9 shows the axial distribution of the atoms in the Poly-C ss-DNA.

Poly-C of 30 bases

Poly-C of 25 bases

Poly-C of 19 bases

Poly-C of 15 bases

Poly-C of 10 bases

Poly-C of 5 bases

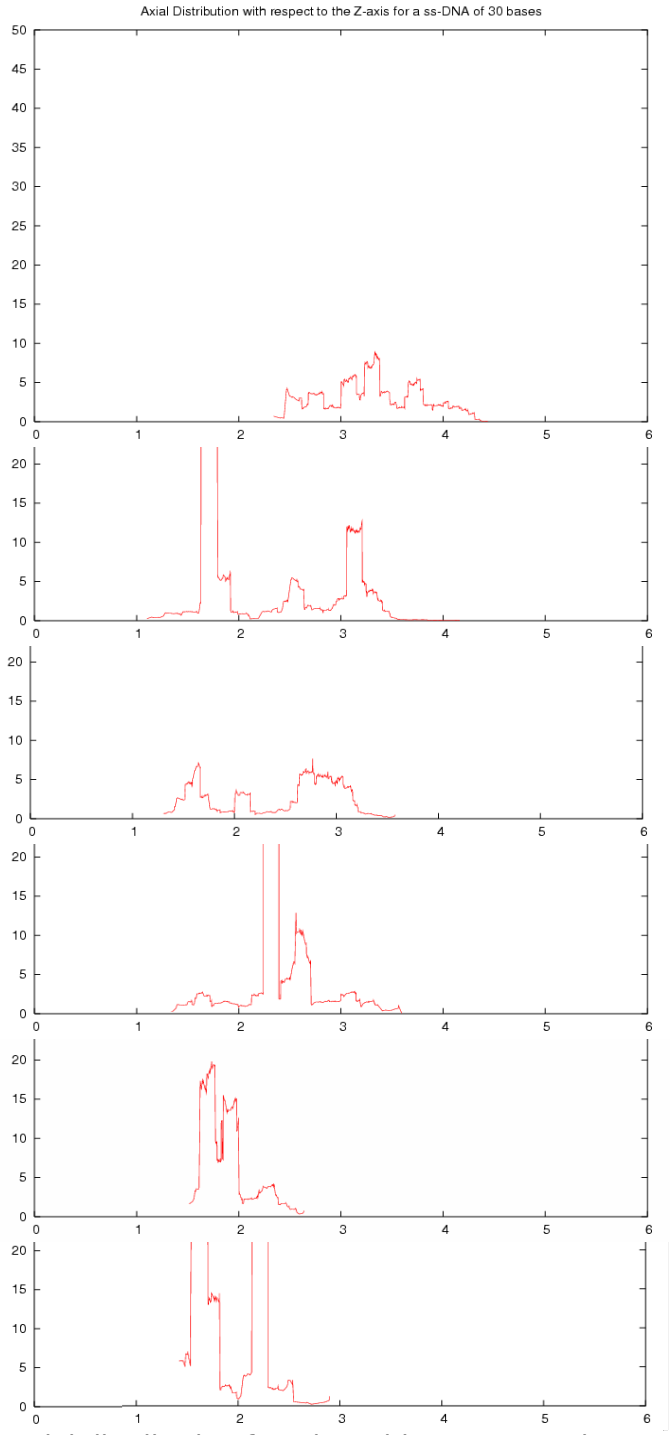

Figure 9. Axial distribution function with respect to the z-axis.

5. **Conclusions**

It was found that MoSDAS greatly simplifies the simulation process and avoids errors. Also, the molecule separation 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.

6. **Future work**

We will develop Graphical User Interface 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 thank the Penn-UPR Partnership for Research and Education in Materials project (NSF-DMR-353730), the 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. "DNA-Decorated Carbon Nanotubes for Chemical Sensing" Staii, C.; Johnson, A. T., Jr.; Chen, M.; Gelperin, A. Nano Lett.; 2005; 5(9); 1774-1778.

2. “Structure-Based Carbon Nanotube Sorting by Sequence-Dependent DNA Assembly”

3. Zheng, M.; Jagota, A.; Strano, M.; Santos, A.; Barone, P.; Chou, S.; Diner, B.; Dresselhaus, M.; Mclean, R.; Onoa, G.; Samsonidze, G.; Semke, E.; Usrey, M. and Walls, D. *Science* 28 November 2003 302: 1545-1548.

4. Johnson Group: Experimental Nanoscale Physics-Resources, http://www.lrsm.upenn.edu/~nanophys/nanotube.html

5. Gallier, J.; Geometric Methods and Applications; *Springer*; p. 245

6. Radial distribution function, http://www.ccr.buffalo.edu/etomica/app/modules/sites/Ljmd/Background1.html

7. Cramer, C; Essentials of Computational Chemistry, Second Edition, *Wiley,* p.84

8. Rapaport, D.; The Art of Molecular Dynamics Simulations, Second Edition, *Cambridge,* p.90

9. GROMACS, http://www.gromacs.org/gromacs/features/feature-summary.html

10. Tinker Molecular Modeling Package, http://dasher.wustl.edu/tinker/

11. The Amber Molecular Dynamics Package, http://amber.scripps.edu/

12. Visual Molecular Dynamics, http://www.ks.uiuc.edu/Research/vmd/

13. Hern D.; Computer Graphics, Second Edition, *Prentice Hall,* p.409