

Segmentation and Additive Approach (SAA) – A Reliable Technique to Study Noncovalent Interactions of Large Molecules at the Surface of Single-Wall Carbon Nanotubes (SWNTs)

Ana M. Torres^{1,4}, Steve Scheiner², Ajit K. Roy³, Andrés Garay-Tapia⁴, John J. Bustamante¹ and Tapas Kar^{2*}

¹ Centro de Bioingeniería, Universidad Pontificia Bolivariana, Bloque 22C, Planta 2^a Circular 1^a No 73–76, Medellín, Colombia

² Department of Chemistry and Biochemistry, Utah State University, Logan, UT 84322-0300, United States

³ Materials and Manufacturing Directorate, Air Force Research Laboratory, Dayton, Ohio 45433, United States

⁴ Centro de Investigación en Materiales Avanzados (CIMAV), Unidad Monterrey, Alianza Norte 202. Parque PIIT. Apodaca, Nuevo León, México.

ABSTRACT: This investigation explores a new protocol, named Segmentation and Additive approach (SAA), to study exohedral noncovalent functionalization of single-walled carbon nanotubes (SWNTs) with large molecules, such as polymers, bio-molecules etc, by segmenting the entire system into smaller units to reduce computational cost. A key criterion of the segmentation process is the preservation of the molecular structure responsible for stabilization of the entire system in smaller segments. Noncovalent interaction of linoleic acid (LA, C₁₈H₃₂O₂), a fatty acid, at the surface of a (10,0) zigzag nanotube is considered for test purposes. Three smaller segmented models have been created from the full (10,0)-LA system and interaction energies were calculated for these models and compared with the full system at different levels of theory, namely ωB97XD, LDA. The success of this SAA is confirmed as the sum of the interaction energies is in very good agreement with the total interaction energy. Besides reducing computational cost, another merit of SAA is an estimation of the contributions from different sections of the large system to the total interaction energy which can be studied in-depth using a higher level of theory to estimate several properties of each segment. On the negative side, bulk properties, such as HOMO-LUMO gap, of the entire system cannot be estimated by adding results from segment models.

*Corresponding author: Tapas Kar, Email: tapas.kar@usu.edu, Fax: 1-435-797-3390

I. INTRODUCTION

Modeling and simulation have been well established as instrumental in understanding molecular structures, structure-property relations, complicated chemical processes, entangling complicated experimental results and the design of new molecular systems and chemical processes to develop application-oriented products. Although the ongoing development of both software and hardware enables researchers to study large molecules, computational investigations of the interactions between two large molecules is still computationally highly demanding and in several instances not even feasible with high performance computer facilities. Noncovalent exohedral interactions [Hirsch, 2002 #1868] between single-wall carbon nanotubes (SWNTs) either with polymers or large surfactant are good examples. Because of such computational limitations, theoretical investigations of polymer-wrapped SWNTs or noncovalent functionalization with large surfactants at the surface of SWNTs at a moderate level of theory are missing from the literature. However significant progress [Star, 2003 #1952;Star, 2003 #1952;Chen, 2001 #2029;Zhao, 2003 #2090;Britz, 2006 #2222;Zhao, 2003 #3158;Zhao, 2009 #3616;Herranz, 2010 #3617;Ma, 2010 #3618] has been made from the experimental side on such materials, especially on nano-composites, aromatic small molecules, biomolecules and polymer based noncovalent functionalizations. Because of the importance of noncovalent functionalization of SWNT, development of a computationally efficient but reliable approach to investigate such interactions would be beneficial in advancing research on these fascinating materials.

Since for any weak interaction, such as noncovalent functionalization of SWNTs, electron correlation is necessary to accurately estimate interaction or adsorption energies, QM/MM methods would not be reliable. Recent development of the Same Level Different Basis (SLDB) protocol [Kar, 2004 #2801;Kar, 2006 #2990;Kar, 2008 #3183;Kar, 2010 #3221;Kar, 2012 #3381;Kar, 2012 #3382] introduced to study covalent interactions of a wide range of molecules at the side-wall and at the end of tubes was found highly reliable and efficient. Significant reduction of computational cost by this method is due to consideration of larger basis functions only for a small number of carbon atoms of SWNT at the active site and adsorbates, and smaller basis functions for the remainder of carbon atoms of the tube. Although this approach may be comfortably applied to investigate [Kar,

2008 #3183] noncovalent interactions between small aromatic molecules at the surface of SWNTs, computational cost increases exponentially as the size of adsorbate increases, felt particularly in the cases of polymer, biomolecules and large surfactant. As a consequence, a large number of carbon atoms of the tube, part of the active site, must be treated with larger basis sets, which causes a sharp increase in computational time. Application of the periodic boundary condition (PBC) will consume more computational time than the molecular model calculation of the same system, as a larger tube model in the unit cell of PBC has to be considered to avoid interaction between adsorbate molecules.

Over the years, several fragmentation approaches have been developed to study large systems, such as large biomolecules and molecular clusters, aiming to achieve computational efficiency and recently Gordon et. al. [Gordon, 2012 #3641] critically reviewed those approaches. In most cases, the total system is fragmented into smaller units and each unit are calculated separately. Such approximations are always associated with error in estimating total energy of the full system. Inclusion/exclusion of interfragment interactions (overlap between smaller units) is a critical issue on the accuracy. One such divide-and-conquer method is Molecular Tailoring Approach (MTA) developed by Gadre and co-researchers [Sahu, 2014 #3642], where maximum fragment size in terms of atoms per fragment and a distance factor, termed as R-goodness, are central to the approach. MTA is found reproduces total energy of the full systems while reduces computational cost. Such approach has been tested on a number of molecular clusters (such as $(C_2H_2)_{10}$, $(CO_2)_{13}$, $(H_2O)_{25}$, $(C_6H_6)_{10}$ etc), [Sahu, 2013 #3647] and also π -conjugated systems, such as graphene sheets ($C_{56}H_{20}$ and $C_{96}H_{24}$) [Yeole, 2010 #3646]. They further developed the MTA approach to optimize fragments models and also introduced a grafting procedure [Sahu, 2014 #3642] to include electron correction methods, such as MP2, CCSD(T) at MTA, which require four-fold calculations on fragments and full system at HF and correlation methods.

These fragmentation methods were applied to a wide range of large molecular systems, but not the interaction between two large systems, as in the case of noncovalent functionalization of SWNTs. Such systems are more complicated as fragmentation of one will have influence on the other and many additional calculations have to be considered for overlapping fragments. Moreover, most of the prescribed fragmentation methods are

associated with developing programming code and locally modified QM softwares which cause hindrance to carry on the black-box way of doing calculations. Such limitations of currently available methods demand a new approach to handle noncovalent functionalizations of SWNTs.

In this investigation we explore the possibility of a new approach termed Segmentation and Additive approach (SAA) where segments are mutually exclusive. To assess the reliability of this protocol, we considered linoleic acid at the surface of a (10,0) SWNT. Linoleic acid (LA, $C_{18}H_{32}O_2$), a polyunsaturated omega-6 fatty acid, is a carboxylic acid with an 18-carbon chain with two C=C double bonds as shown in Figure 1. End-functionalization of LA with SWNT has been recently reported [Azevedo, 2010 #3615]. We considered two possible orientations of LA as shown in the figure.

II. Segmentation and additive approach

The basic concept of the segmentation and additive approach (SAA) is to partition the entire system into smaller segments and calculate interaction energy (E_{Int}) of each of these smaller individual units. A key criterion is that the sum of interaction energies of segments should be close to the total interaction energy of the system. If this criterion is maintained then the SAA will lower the computational cost significantly, and will thus allow theoretical investigators to contribute valuable insights to this arena of noncovalent functionalizations. The advantage of this approach is that each segment model (SM), depending on its importance and relevance to structure and properties, can be studied in depth at a higher level of theory, and would provide information on further functionalization or changes required in functional group(s) of adsorbates to achieve some application-oriented goals. Besides, such models will show which segment of the large molecule is responsible for certain properties that would not be possible to estimate when the entire molecule is calculated. However, some properties, such as band gap (HOMO-LUMO gap), of the entire system cannot be estimated from the SAA approach.

To assess the reliability of this strategy, one needs to calculate the entire system, followed by computations on each segment to compare results. To do so, a full geometry optimization of the entire system is an essential part of the approach; secondly, how to partition the full system into smaller components. In the following section we describe a

step-by-step approach as applied to the SWNT-LA model system. If successful, then SAA is most straightforward approach without modifying or developing any software and convenient of using black-box approach using standard QM softwares.

III. Method of Calculations

Geometry optimization by relaxing all atoms is one of the most computationally demanding steps, and computational cost escalates significantly with increasing number of atoms, basis functions, and number of electrons. Thus, a moderate level of calculation to optimize geometries with a reasonable set of basis functions, such as double-zeta quality with polarization functions (6-31G*), for more than 200-atom systems are the most time-consuming part. The following chart shows the number of atoms, the number of basis functions and number of electrons in the (10,0)-LA testbed. The use of the 6-31G (5d) basis set for all atoms of the entire (10,0)-LA models (C₂₉₈H₅₂O₂) results in 4304 basis functions with 928 occupied molecular orbitals (MOs). In such cases, utilization of DFT with dispersion corrections or MP2 methods seem not feasible for routine geometry optimizations. Moreover, the number of geometric parameters in (10,0)-LA (C₂₉₈H₅₂O₂, 486 distances + 952 angles + 1820 dihedral angles = total 3258 variables in Z-matrix; or 1056 Cartesian coordinates) are another factor of concern in full optimization.

Coupled with these issues, oscillation of the SCF energy (both early and late stages of SCF cycles) is a technical problem that encountered at the intermediate steps of full geometry optimization, as well as single point calculations, of (10,0)-LA and (10,0) structures. To circumvent oscillation of the SCF energy, several non-standard or default options implemented in G09 program [Frisch, 2009 #3222] were used after several trial runs and those are: (1) increasing 2-electron integral accuracy to 12 (i.e., 10¹² instead of default 10¹⁰) and grid = ultrafine (integration grid used for numerical integration) and reducing SCF convergence criterion to 10⁻⁷ instead of default value of 10⁻⁸. Such integral options and criterion were used for all calculations. Besides those, in some single point calculations at different levels energy-based mixing in DIIS extrapolation (iop(5/22=20)), was used. CPU time required for each cycle of the SCF step (time took in link502 of G09 code) is given at the last column of Chart 1. (Steve: I shall explain about cpu after I finish couple of additional calculations)

Chart 1

		Basis Functions		Time (min) ^a
(10,0):	C ₂₈₀ H ₂₀	3960	1700 electrons	103.1 (51)
LA :	C ₁₈ H ₃₂ O ₂	344	156 electrons	0.5
(10,0)-LA-Orientation-1				
Full:	C ₂₉₈ H ₅₂ O ₂	4304	1856 electrons	189.7 (51)
SM-1:	C ₁₀₆ H ₃₂ O ₂	1576	684 electrons	27.0 (38)
(10,0) SM-1:	C ₁₀₀ H ₂₀	1440	620 electrons	17.7 (69)
SM-2:	C ₁₂₈ H ₃₄	1860	802 electrons	47.4 (106)
(10,0) SM-2:	C ₁₂₀ H ₂₀	1720	740 electrons	32.4 (129)
SM-3:	C ₈₄ H ₃₀	1236	534 electrons	13.6 (21)
(10,0) SM-3:	C ₈₀ H ₂₀	1160	500 electrons	11.0 (18)
				(=149.6)
(10,0)-LA-Orientation-2				
Full:	C ₂₉₈ H ₅₂ O ₂	4304	1856 electrons	191.6 (102)
				(=234.4)
SM-1:	C ₁₀₆ H ₃₂ O ₂	1576	684 electrons	32.6 (57)
(10,0) SM-1:	C ₁₀₀ H ₂₀	1440	620 electrons	25.1 (48)
SM-2:	C ₁₀₈ H ₃₄	1580	682 electrons	28.6 (26)
(10,0) SM-2:	C ₁₀₀ H ₂₀	1440	620 electrons	22.3 (21)
SM-3:	C ₈₄ H ₃₀	1236	534 electrons	15.3 (29)
(10,0) SM-3:	C ₈₀ H ₂₀	1160	500 electrons	10.9 (23)

^a cpu time (number of SCF cycles) for each cycle of the SCF step at the single point wB97XD/6-31G*(5d) calculations. 36 sharedprocessors and 20GB memory were used for all calculations at SGI HPC (Intel I5-2697v3, 2.6GHz) system.

We took advantage of the highly cost effective semiempirical PM6 method [Stewart, 2007 #3620] to obtain initial geometries of (10,0)-LA in two orientations. Optimized structures of (10,0)-LA, shown in Figure 2, indicate two possible arrangements: in orientation-1 (O-1), the linoleic acid remains flat at the surface along the axis of the tube, and in orientation-2 (O-2) LA partially wraps around the tube. In both cases, several C-H bonds of LA are directed to the surface of the tube, indicating C-H-- π interactions between adsorbate and adsorbent. Stretching of those C-H bonds in (10,0)-LA, compared to others away from the tube surface, confirms such interaction [Scheiner, 2001 #3008]. In both cases, the closest distance R between C-H and the center of C=C bond of the tube is about 2.9 Å. It may be worth mentioning that some of the C-H bonds are directed towards the center of hexagons

of the tube surface. Such intermolecular C-H-- π distances of about 2.9-3.4 Å, predicted by PM6 method, is slightly longer than the standard C-H-- π distance of 2.3-2.9 Å [Suezawa, 2004 #3623;Nishio, 2011 #3624]. Such a discrepancy is quite reasonable as PM6 omits explicit correlation as well as dispersion correction required for proper description of weak C-H-- π interaction. To correct such effects, we employed the ω B97XD functional [Chai, 2008 #3621] that includes an empirical dispersion correction. [As mentioned earlier, full optimization of \(10,0\)-LA at the \$\omega\$ B97XD/6-31G\(5d\) will be highly time-consuming due to large number of basis functions, number of electrons and number of variables in addition to SCF convergence problem during optimization, an alternative has been explored and verified in the next step of SAA.](#)

In this step, we performed a potential energy surface (PES) scan by bringing LA in orientation-1 from 3.0 Å to 2.4 Å towards the (10,0) tube by freezing the PM6 optimized geometry of both LA and the (10,0) tube at each intermolecular distance R. The variation of interaction energy (obtained using equation 1, where E is the electronic energy) with R is reported in Table 1.

$$E_{\text{int}} = E(\text{SWNT-LA}) - E(\text{SWNT}) - E(\text{LA}) \quad (1)$$

(It is worth noting that such a weak interaction has little or no effect on the internal geometrical parameters, which supports the freezing of LA and SWNT geometries during the PES scan). It can be seen that the PES is very flat (Figure 3) and exhibits a minimum around 2.6 Å at the ω B97XD/3-21G level. [That means a small change in intermolecular distance, that may results of full optimization, will have negligible effect on the interaction energy.](#) To further verify this finding, we also considered another method B3LYP/3-21G [Becke, 1993 #930;Lee, 1988 #3197] with GD3BJ dispersion correction [Grimme, 2010 #3622]. Although interaction energies at different R are slightly lower than that of ω B97XD energies, the functional behavior of the interaction energy at this level is similar and exhibits a minimum around 2.6 Å (see Figure 3). After establishing a reasonable geometry for (10,0)-LA, we calculated interaction energies of the entire system at different levels of theory and these energies are summarized in Table 2 under column Full.

The entire (10,0)-LA molecular model were then segmented into three smaller segments as shown in Figure 4. Such division (shown by vertical lines in the figure) creates dangling bonds at the newly created edge(s) of segment models. For example, in one edge

of SM-1 and SM-3, and in both edges of SM-2, and those edges were saturated by hydrogen atoms. Also dangling bonds created during the division of LA in segment models were saturated with hydrogen atoms using Chemcraft program [Zhurko #3259] that usages standard bond parameters. In this saturation process, positions of all atoms of tubes and LA, besides newly added H-atoms, remain same as in full system. The effects of optimization of those added atoms on interaction energies are discussed below.

Segmentation of entire (10,0)-LA mainly depends on the adsorbate molecules. In the present scenario, the emphasis was given to C-H bonds facing toward the tube surface as those mainly contribute to the stabilization of (10,0)-LA via C-H- π hydrogen bonds. The total number of such C-H bonds (13 in O-1, and 12 in O-2) in segment models was held fixed as in the full (10,0)-LA so that no C-H- π bonds of (10,0)-LA were left out of the smaller models. Also, both C=C units at the 9th and 12th positions from the carboxylic group end of the 18-carbon chain LA was considered in a smaller model (SM-2), to avoid any disruption of π -conjugation. After creating three smaller models, interaction energies were calculated at different levels of theory and the results are summarized in Table 2. All calculations were performed using Gaussian 09 program [Frisch, 2009 #3222].

III. Results and discussion

As mentioned earlier, the sole purpose of the present investigation is to explore this new approach of segmentation to study noncovalent functionalization of SWNTs by large molecules and assess the reliability of this protocol. The success of this approach depends on the interaction energies of segments and whether that correctly reproduces the Full interaction energy, not the total electronic energy. It may be noted that total electronic energy of the segmented models will differ from the energy of the full model due to additional hydrogen atoms in all three SMs, either at one edge or both edges. No additional calculations on the overlapping of fragments required in SAA.

Before moving to estimate energies at higher levels, we obtained PM6 interaction energies of three segment models shown in Figure 4, given in the first row of Table 2. In these calculations, PM6 optimized geometries of (10,0)-LA, (10,0) and LA were considered, and segmented models were created from those PM6 structures. If the sum of interaction energies of three small units is close to that of the full system, then segmentation

process is correct, and such models can be used for further investigation. Indeed for both orientations of (10,0)-LA Total energies of -5.7 kcal/mol (O-1) and -4.9 kcal/mol (O-2) are in very good agreement with corresponding Full value, supporting segment models are correct.

The second set of energy data, listed in Table 2, is obtained with the ω B97XD/3-21G method using $R=2.6 \text{ \AA}$ structures (as shown in Figure 2) of (10,0)-LA. At this level of theory, the interaction energy of the entire (10,0)-LA for both orientations is -31.3 kcal/mol. The sum of E_{Int} of the three smaller models overestimates this interaction energy by about 3-4 kcal/mol, i.e., about 10%. Since 3-21G is not an adequate basis set for dealing with such interactions, we considered larger 6-31G(5d) basis set, which seems a reasonable set. Moving from 3-21G to 6-31G(5d) basis set, an increase of 1500 basis functions, increases computation cost by almost five times in each cycle of SCF step of full (10,0)-LA. It is worth mentioning that inclusion of sp-diffuse functions, as in 6-31+G(5d), would be more appropriate for such noncovalent interactions. However, consideration of 6-31+G(5d) for (10,0)-LA increases the number of basis functions from 4304 to 5504, which will raise the computational cost significantly.

The third set of energies in Table 2 are from ω B97XD/6-31G(5d) calculations. At this level of theory, the interaction energy of orientation-1 is slightly greater than that of orientation-2, and the difference is 2.7 kcal/mol. The sum of segment model interaction energies is close to that of the full (10,0)-LA. For example, the difference in O-1 is only 0.5 kcal/mol, and for O-2 this quantity is 1.5 kcal/mol, where in both cases the total interaction energy is overestimated by a very small margin.

We also verified the effect of added hydrogen atoms at the edges of segmented models using graphics program by relaxing only those H-atoms in a partial optimization at the ω B97XD/6-31G(5d) method. It is quite natural that such optimization will lower electronic energies of the segmented complexes (SM-1 to SM-3) as well as their SWNT components, most likely by close magnitude than non-optimized structures. Thus, the error due to non-optimization of added hydrogen atoms will be canceled out to provide similar interaction energies. Indeed, the total interaction energies (given in the fourth row of Table 2) are almost same as that of non-optimized structures, lowered by only 0.5 and 0.2 kcal/mol in O-1 and O-2, respectively. Since in the present segmented models of (10,0)-

LA, edges of SWNT parts are not perturbed by fragmented LA, as C-H bonds of LA part are away from the tube surface at both edges of smaller models, changes in interaction energies are not infected. However, it is recommended to optimize the added hydrogen atoms if there are interactions of the adsorbate at the edges of the smaller models.

Interaction energies of non-optimized structures are then corrected for basis set superposition error (BSSE) using the counterpoise (CP) method [Boys, 1970 #3177], and BSSE corrected interaction energies, along with BSSE energies (in parenthesis), are given in the following row. It can be seen that sum of BSSE energies of SMs is exactly same to their full version, i.e., 9.0 kcal/mol in O-1 and 8.2 kcal/mol in O-2. Thus, the difference between total and full energies remain same after BSSE correction.

The energy of the C-H-- π hydrogen bond depends on the acidity of the C-H group [Shibasaki, 2007 #3619], and roughly varies in the order: alkyne C-H (~3.0 kcal/mol) > alkene C-H (~2.0 kcal/mol) > alkane C-H (~1.0 kcal/mol). Based on these energies and number of C-H bonds (12-13, mostly alkane type), BSSE corrected interaction energies of 22.6 and 20.7 kcal/mol of full (10,0)-LA seems quite reasonable.

Since we are not concern of total energies, but only interaction energies, the total number of carbon atoms in full and SMs may be different depending on the source of interaction between SWNTs and adsorbate. For example, full and smaller models (figure 4) contain the same number of carbon atoms (i.e., 280) in O-2 (14 layer/ring, 20 carbons in each layer/ring). Although full O-1 contains 280 carbon atoms, that number is 300 in combination of three segmented models, where SM-2 of O-1 contains additional one layer of carbon (on the right side connecting to SM-3) to accommodate all eight carbon atoms of LA part within the SWNT boundary, otherwise last C atom (near SM-3) would lay outside of the right-edge of SM-2. To verify the effect of SWNT model size, we extended SM-3 models of O-2 by adding one more layer of carbons at the right edge and interaction energy lowered by 0.2 kcal/mol. Thus, SAA is quite flexible in selecting segmented models.

(steve: I shall complete this para after couple of test runs and estimating cpu time)
Compared to the full system, computational time for each cycle of SCF step reduces by more than four times in SAA. So the cost-to-benefit ratio is optimal for the SAA at this level of theory.

Such encouraging results motivated us to consider other methods to verify the reliability of the SAA and to check that it is not only valid for ω B97XD. Local density approximation (LDA) [Hohenberg, 1964 #3625;Kohn, 1965 #3626;Slater, 1974 #3627;Vosko, 1980 #3628], a less computationally demanding method, has been in discussion for some time as an approach to noncovalent functionalization of SWNTs with small molecules [Tournus, 2005 #3156;Tournus, 2005 #3157]. The next set of energy data in Table 2 is derived from LDA(SVWN)/6-31G(5d) (BSSE energies obtained at the ω B97XD level is used) and the result is quite similar to that of ω 97XD energies. The sum of total BSSE corrected E_{Int} of three segmented models are -18.7 kcal/mol in O-1 and -18.4 kcal/mol in O-2 of (10,0)-LA, only slightly lower than the corresponding Full E_{Int} by 0.6 and 1.3 kcal/mol, respectively.

MP2 is probably the most commonly used method to study weak interactions in small molecules, but is out of reach for such large systems. However, the SAA protocol makes it possible to study smaller segments, as the number of basis functions and electrons (Chart 1) is reduced by a good margin. In the seventh row of Table 2, we present only interaction energies (without (with) BSSE correction) of segment models. Based on the findings of ω B97XD and LDA results, one may consider the sum of those energies will be close to full interaction energies. However, the contribution of each segment model differs from ω B97XD and LDA values. Analyses of such differences and also comparison among different methods are beyond the scope of the present investigation, as the main target of this study is identification of an alternative protocol to investigate noncovalent functionalization of SWNTs by large molecules.

If resources are available then parallel computing will certainly reduce wall-clock time of these calculations. For example, the MP2/6-31G(5d) calculations of SM-1 and SM-2 of O-2 structure completed in about 2.9 days when 8 nodes and 18 shared processors with 10GB RAM per processor were used. Same jobs completed in about 14 days when 36 shared processors with 20GB RAM were used. These smaller models are about one-third in size of the full system, and nonlinear scaling in such calculations restricts consideration of the full system at MP2 level. In comparison to SAA, Grafting procedure by Gadre group [Sahu, 2014 #3642] requires several additional calculations, such as interfragment sections beside 4-fold calculations on full/fragmented models using lower and higher basis sets.

Thus, overall wall-clock time may not reduce significantly, and need to verify computational cost for such large systems.

The Hartee-Fock (HF) method underestimates any weak interaction, as no correlation and dispersion required for the proper description of such interaction, is included. Nonetheless, we calculated energies using HF/6-31G(5d) at the same geometry used at ω B97XD, LDA, MP2 method, to check the reliability of the SAA method. Full E_{Int} value of +14.3 (O-1) and +11.6 (O-2) clearly indicates an endothermic/repulsive process and (10,0)-LA structures are not stable. What is important is that those energies are reproduced correctly for O-2 orientation, and only slightly underestimated for O-1 orientation. Overall, SAA reproduces interaction energy of the full system whether attractive or repulsive, and for any given level of theory.

In the above example we considered three segment models wherein the SAA reproduces the interaction energy of the entire system with reasonable accuracy. However, it is possible to partition even smaller models depending on the functional groups of adsorbates. For example in the present case, the carboxylic -COOH group of LA may contribute, besides C-H-- π bonds, to the interaction energy of entire (10,0)-LA. The SM-1 model does not provide such information. So we further partition the SM-1 model, separating the COOH segment from the rest of the carbon chain and created two models from SM-1 (in both orientations) as shown in Figure 5, where interaction energies are also corrected for BSSE. In these fragmented models, BSSE energies of SM-1 (4.4 kcal/mol) and SM-2 (3.6 kcal/mol) are 0.4 kcal/mol lower than the corresponding sum of smaller models. ω B97XD/6-31G(5d) BSSE corrected interaction energies of -9.6 kcal/mol (O-1) and -8.2 kcal/mol (O-2) are due to the interaction of each COOH unit with the tube and four C-H-- π bonds. Separation of both types of interactions into smaller models clearly provides a contribution from each group. Contributions from four C-H-- π bonds are close in both orientation and in the range of 3-4 kcal/mol, which is a reasonable number as each alkane C-H contributes about 1.0 kcal/mol. However, the contribution from each COOH group differs by about 2 kcal/mol, higher in O-1 than O-2 orientation, as the acid group of the former is closer to the tube.

To verify the effect of basis set on interaction energies, additional calculation were carried out using triple-zeta quality basis set (6-311G*) and those energies (with BSSE

correction) are summarized in the second set of results in Figure 5. Non-corrected interaction energy is only -0.2 kcal/mol lower in O-1 SM-1 with the extension of basis set, but higher by the same amount in O-2 SM-1. As expected BSSE energies are lower by about 1.0 kcal/mol at 6-311G(5d) level, which widens difference in BSSE corrected interaction energies by about 1.0 kcal/mol. Smaller segmented models reproduce Full (SM-1) interaction energies at both basis set with or without BSSE correction. For example, total BSSE corrected interaction energy of -9.4 (6-31G*) or -10.3 (6-311G*) of smaller models differ by only -0.2 kcal/mol than the full O-1 SM-1 model.

Thus depending on the adsorbates, one can partition the entire system into any number of smaller units to reduce computational costs and treat smaller units at a higher level of theory and estimate different properties. A note of caution is that one must properly assess the initial geometry of the entire system, which may be obtained at a lower level of theory, such as semiempirical methods or even using DFT method with a smaller basis set. The second important rule of SAA is to identify interaction types, such as C-H-- π , π - π etc., and preserve types and numbers of such interactions in smaller units, without breaking any π -network, and assess the reliability of segmented models before considering moderate to high levels of theory. For the SWNT component, a minimum requirement is to consider 6 x m number of carbons in (m,0) zigzag tubes and 5 x 2n in (n,n) armchair tubes. [Smaller than this size may lead to instability in the wavefunction \(due to the small singlet-triplet energy gaps\) as has been found in \[n\]cyclacenes \[Houk, 2001 #3643; Chen, 2007 #3645; Hirst, 2011 #3644\].](#)

IV. CONCLUSIONS

Although experimental studies on exohedral noncovalent functionalization of SWNTs by large molecules, such as polymers, bio-molecules, and large surfactants, continue apace, supporting computational studies are lacking, mainly due to the size of these systems which is highly computationally demanding and may not be doable at higher level of theory even with high performance computational facilities. The main purpose of the present study is to explore a new protocol named segmentation and additive approach (SAA) that can be comfortably applied to this demanding research arena. In this method, the large system is

segmented into smaller units such that the factors responsible for noncovalent interaction of the total system are preserved in segment units.

Linoleic acid (LA) at the surface of (10,0) is considered as a test case in the present investigation, and three segment models are created from the full system. Stabilization of (10,0)-LA is mainly due to C-H-- π hydrogen bonds in both orientations of functionalization. The sum of interaction energies of segment models is in very good agreement with the full interaction energy of the entire system at each level of theory. This finding supports the reliability of the SAA models and may be applicable to a wide range of noncovalent functionalization of SWNTs. This is the first attempt to establish this approach and further investigations will standardize this protocol and explore applications in estimating properties of the system other than the interaction energy. [This approach seems a reliable approach to study noncovalent interactions between large systems with pristine SWNTs.](#)

ACKNOWLEDGEMENTS

Ana M. Torres and John J. Bustamante thank to Colciencias (Colombia) for financial support, under project 513-2012. Andrés M. Garay would like to acknowledge financial support from CONACYT-Mexico CB-2012-1/182889 Grant. Computational facilities at Bioengineering Center (Medellín, Colombia), the Computational Thermodynamics Laboratory at CIMAV (Monterrey, México) and DoD HPC at AFRL (Dayton, Ohio, USA) are gratefully acknowledged.

Table 1. Variation of Interaction Energies (E_{Int} in kcal/mol) with Distance R^a.

R (Å)	E_{Int} (kcal/mol)		
	PM6	wB97XD/3-21G GD2 dispersion correction	B3LYP/3-21G GD3BJ dispersion correction
3.0	-5.70		
2.9	-5.76	-29.68	-26.53
2.8	-5.69	-30.45	-27.29
2.7	-5.47	-31.01	-27.79
2.6	-5.04	-31.28	-28.10
2.5	-4.34	-30.28	-28.02
2.4	-3.30	-28.67	-27.60

^a R is the closest distance between C-H of LA (as shown in Figure 2) and C=C bond of SWNT. Negative E_{Int} equates to a stable complex and the lowest E_{Int} in each method is shown in bold.

Table 2. Interaction Energies^a (E_{Int} , kcal/mol)

Method	SWNT-LA (O-1) ^b		SWNT-LA (O-2) ^b	
	Full	SM	Full	SM
PM6	-5.8	SM-1 -5.3 SM-2 -1.3 SM-3 +0.9 Total -5.7	-4.4	SM-1 -2.0 SM-2 -2.0 SM-3 -0.9 Total -4.9
ω B97XD/3-21G	-31.3	SM-1 -15.6 SM-2 -12.9 SM-3 -7.1 Total -35.6	-31.4	SM-1 -12.6 SM-2 -12.9 SM-3 -8.6 Total -34.1
ω B97XD/6-31G(5d) (without optimization of added H-atoms)	-31.6	SM-1 -14.0 SM-2 -11.7 SM-3 -6.3 Total -32.0	-28.9	SM-1 -11.8 SM-2 -11.5 SM3 -7.1 Total -30.4
ω B97XD/6-31G(5d) (optimization of added H-atoms)	-31.6	SM-1 -13.8 SM-2 -11.5 SM-3 -6.2 Total -31.5	-28.9	SM-1 -11.9 SM-2 -11.3 SM3 -7.0 Total -30.2
ω B97XD/6-31G(5d) BSSE corrected E_{Int} (BSSE energy)	-22.6 (9.0)	SM-1 -9.6 (4.4) SM-2 -8.9 (2.9) SM-3 -4.6 (1.7) Total -23.1 (9.0)	-20.7 (8.2)	SM-1 -8.2 (3.6) SM-2 -8.6 (2.8) SM-3 -5.4 (1.7) Total -22.2 (8.2)
LDA/6-31G(5d) ^c Without (with) BSSE corrected E_{Int}	-28.3 (-19.3)	SM-1 -14.1 (-9.7) SM-2 -8.4 (-5.5) SM-3 -5.2 (-3.5) Total -27.7 (-18.7)	-27.9 (-19.7)	SM-1 -11.2 (-7.5) SM-2 -9.6 (-6.8) SM-3 -5.8 (-4.1) Total -26.6 (-18.4)
MP2/6-31G(5d) ^c Without (with) BSSE corrected E_{Int}		SM-1 -7.3 (-2.9) SM-2 -11.3 (-8.4) SM-3 -6.0 (-4.3) Total -24.6 (-15.6)		SM-1 -5.3 (-1.6) SM-2 -15.1 (-12.3) SM3 -7.4 (-5.7) Total -27.8 (-19.6)
HF/6-31G(5d) Without BSSE corrected E_{Int}	+14.3	SM-1 +2.7 SM-2 +5.9 SM-3 +4.1 Total +12.7	+11.6	SM-1 +1.7 SM-2 +5.3 SM-3 +5.1 Total +12.1

^a Negative E_{Int} equates to a stable complex, otherwise unstable complex. ^bFull refers to entire (10,0)-LA; SM indicates segment models. ^c BSSE correction from ω B97XD/6-31G(5d) result.

Figure Captions

Figure 1. Structure of two orientations of linoleic acid (LA). C=C bonds are marked by arrows.

Figure 2. Structures and key geometric parameters of SWNT-LA in two possible orientations. Distances are in Angstrom (\AA). R at the PM6 method is close to 2.9 \AA . Color scheme: red – oxygen. Black – carbon, magenta – dummy atom at the center of C=C bond closest to C-H bond.

Figure 3. Normalized interaction energies as a function of distance R. See Figure 2 for the definition of R.

Figure 4. Segmentation models of SWNT-LA in two orientations.

Figure 5. Smaller segments of SM-1 models in both orientations (O-1 and O-2) of (10,0)-LA. Interaction energies (BSSE corrected energy) are in kcal/mol, and distances are in Angstrom. 1st set of interaction energy is from ω B97XD/6-31G* and 2nd set is from ω B97XD/6-311G* calculations.

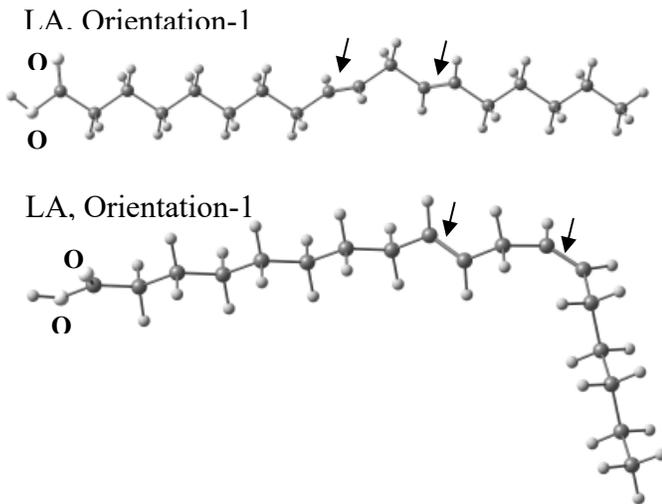
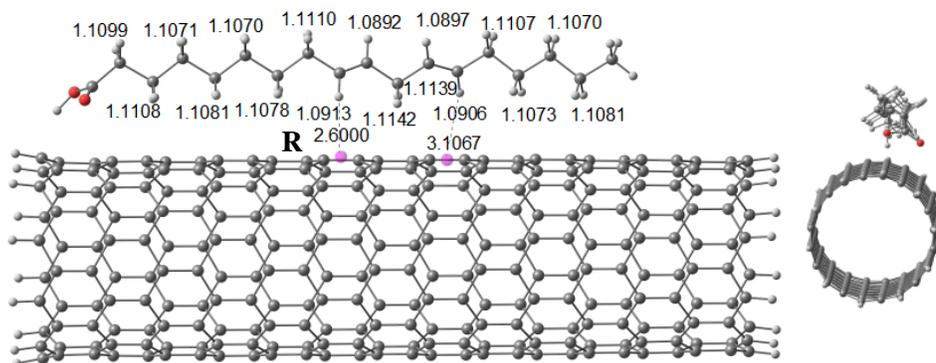


Figure 1. Structure of two orientations of linoleic acid (LA). C=C bonds are marked by arrows.

(10,0)-LA Orientation-1 (O-1)



(10,0)-LA Orientation-2 (O-2)

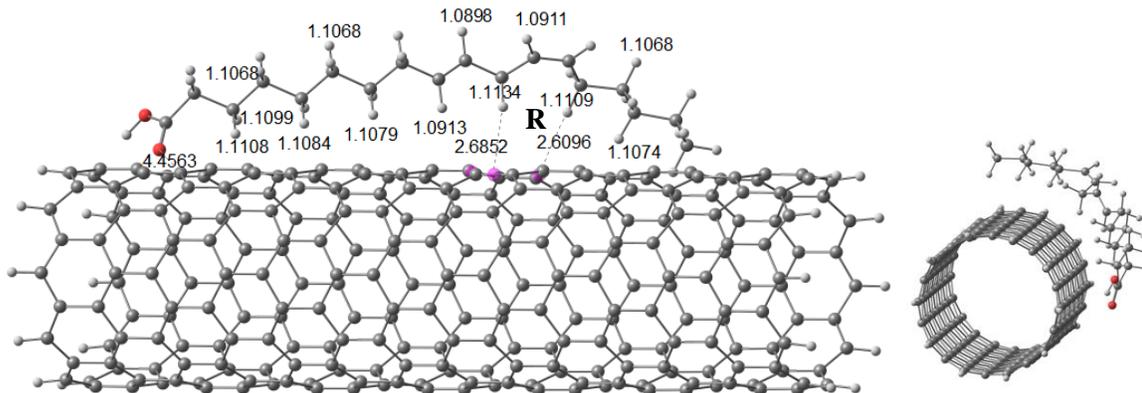


Figure 2. Structures and key geometric parameters of SWNT-LA in two possible orientations. Distances are in Angstrom (\AA). R at the PM6 method is close to 2.9 \AA . Color scheme: red – oxygen. Black – carbon, magenta – dummy atom at the center of C=C bond closest to C-H bond.

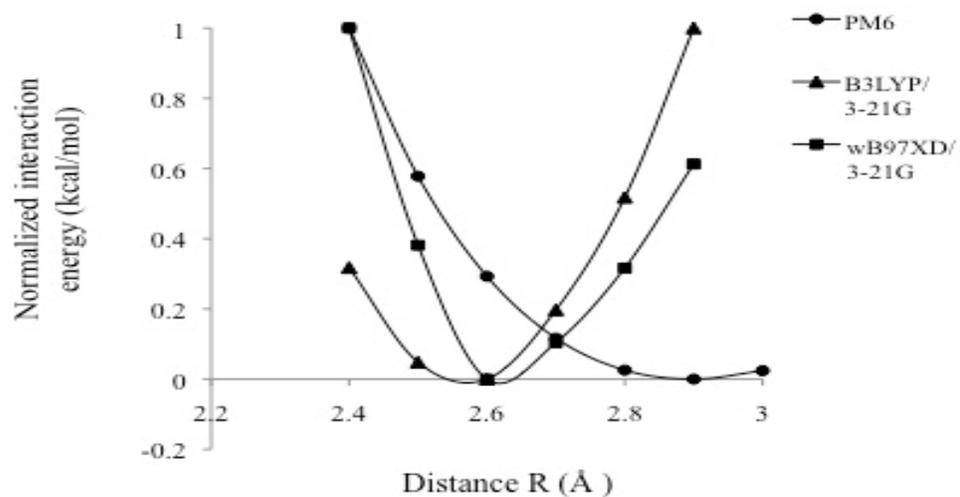


Figure 3. Normalized interaction energies as a function of distance R. See Figure 2 for the definition of R.

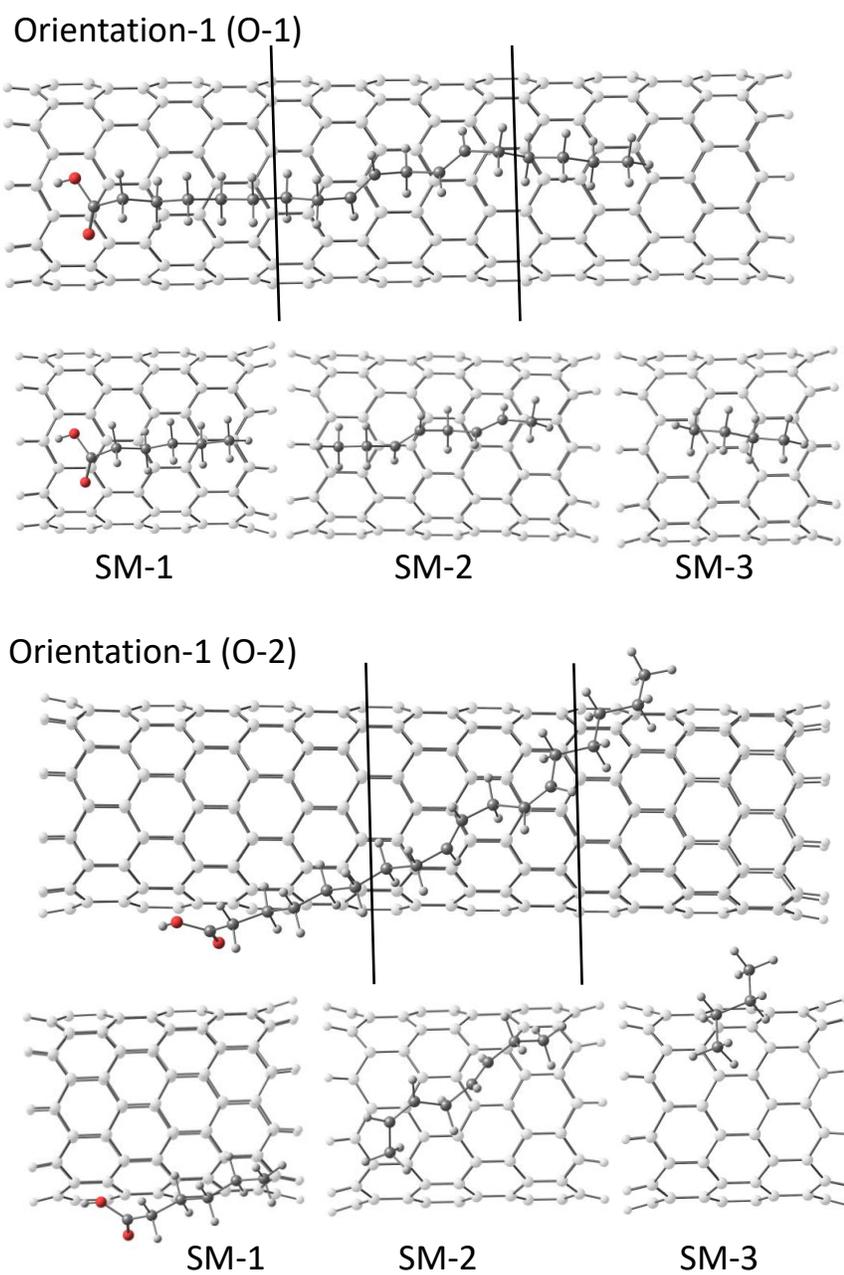


Figure 4. Segmentation models of SWNT-LA in two orientations.

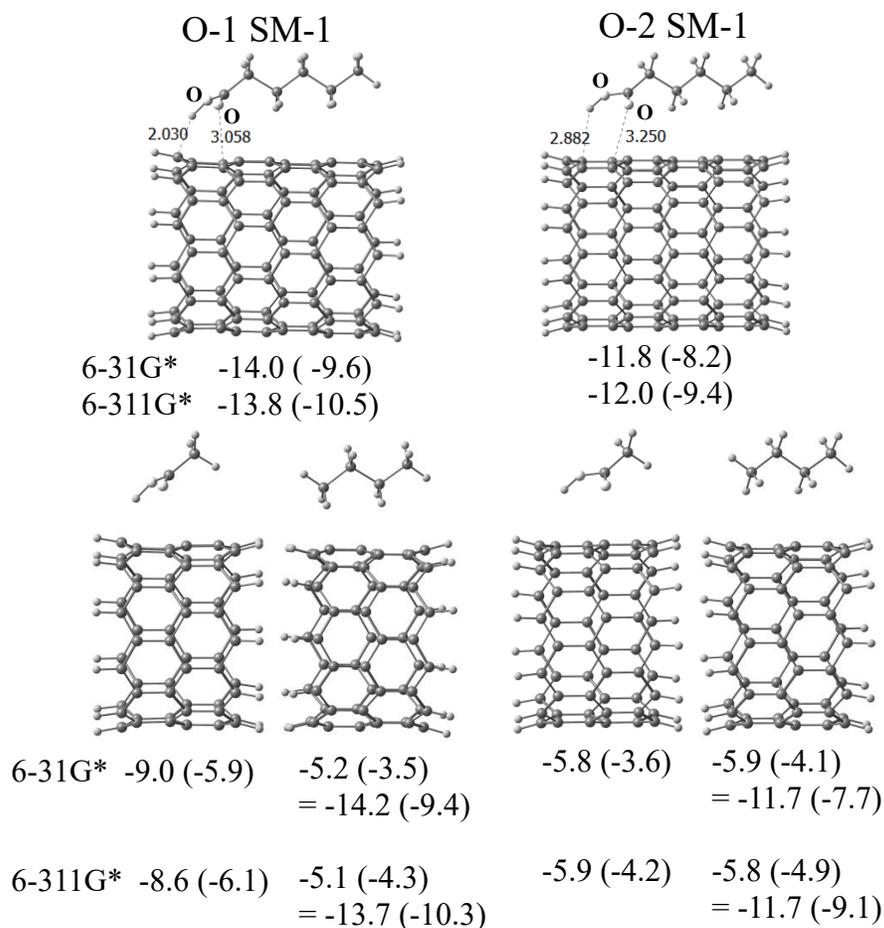
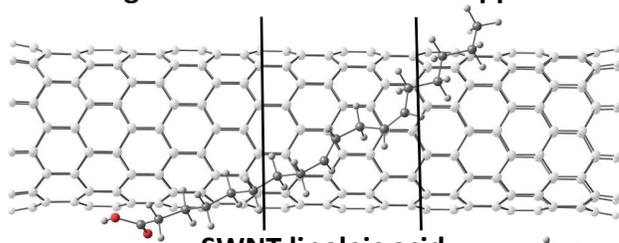


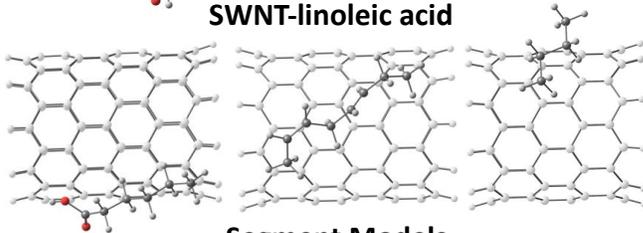
Figure 5. Smaller segments of SM-1 models in both orientations (O-1 and O-2) of (10,0)-LA. Interaction energies (BSSE corrected energy) are in kcal/mol, and distances are in Angstrom. 1st set of interaction energy is from wB97XD/6-31G* and 2nd set is from wB97XD/6-311G* calculations.

Graphical presentation

Segmentation and Additive Approach



SWNT-linoleic acid



Segment Models