Nanobiotechnology

DOI: 10.1002/smll.200600433

Optimal Atomistic Modifications of Material Surfaces: Design of Selective Nesting Sites for Biomolecules

*Boyang Wang and Petr Král**

In recent years, enormous progress has been made in understanding the structures and functions of biological molecules.[1] This effort is largely focused on deciphering the nature of biosystems,[2] designing novel drugs,[3] and treating diseases.[4] In numerous research areas, proteins, enzymes,[5] and various organic or inorganic molecules[6] are attached to materials surfaces. The goals of these challenging studies range from the realization of molecular-electronic[7] and bioelectronic systems[8] to the building of bionetworks functioning outside cells.[9,10]

The material substrates should allow stable nesting of the attached biomolecules in water and provide rich possibilities for their activation or sensing. Carbon compounds, such as diamond[11,12] or graphene-based nanostructures,[13,14] are ideal biosubstrates since they are water resistant, could be modified to be hydrophobic or hydrophilic, conducting or isolating, and allow activation of the biomolecules.[15] It is often crucial to attach the biomolecules to the substrates in specific configurations, in order to preserve and control their activity.[16,17] To minimize the molecule–surface interaction, one could covalently bind a selected atom in the biomolecule to the surface.[18] A robust and universal possibility, allowing the control of numerous molecular activities, is to position the molecules in selective nesting sites predesigned on the material surfaces.

In this work, we develop the methodology of "optimal modifications" of material surfaces, with the goal of designing highly selective nesting sites for the adsorbed molecules.[19,20] We resort to optimization approaches, based on genetic algorithms,[21] using similar principles to biological evolution. These universal methods have been widely applied in the design of proteins and drugs,[22,23] nanostructures,[24] and in optical control of chemical reactions[25,26] and nanoscale processes.[27]

This methodology is first developed in the absence of polar solvents and electrolytes. In such cases, one could build the selective molecular nests by doping the surfaces using carefully designed patterns of suitable atoms, which become partially charged upon doping. Their positions can be arranged in such a way that the local electrostatic potential formed above the doped surface selectively binds the chosen molecule. The methodology is illustrated on a graphene layer,[13,14] substitutionally doped with B and N atoms,[28] which are often used to change free-carrier concentrations in carbon systems. Nesting of the molecules on these modified surfaces is controlled by Coulombic and van der Waals coupling.

The nesting sites are designed by genetic algorithms that minimize the potential energy of the docked molecules. First, the local charges of each dopant and its neighbors is required. Ab initio calculations, done with the B3LYP exchange-correlation functional in the Gaussian03 software package,[29] show that when a single C atom in the graphene layer is substituted by an isolated B or N atom, it carries the charge $Q_B \approx -1.2$ e or $Q_N \approx 1.0$ e, respectively. We assume that this charge is neutralized within the first three neighboring C atoms, each carrying the charge $Q_C = -Q_{B(N)}/3$. When several dopants have a single C neighbor between them, the C atom is charged by summing contributions from those dopants. The same dopants are not allowed to be first neighbors. However, when different types of dopants (B and N) are first neighbors, we find that their charges are $Q_B \approx -1.0$ e and $Q_N \approx 1.0$ e. To simplify the simulations, it can be assumed that the charges of dopants are always $Q_B \equiv -1.1$ e and $Q_N \equiv 1.1$ e. This approximation allows implementation of simple genetic algorithms.

The total Coulombic interaction energy between the adsorbed molecule and the graphene layer doped by an isolated B (or N) atom is given by

$$E_{B(N)} = Q_{B(N)} \sum_{i=1}^{n} \frac{Q_i}{4\pi\varepsilon_0} \left(\frac{1}{r_{i0}} - \sum_{j=1}^{3} \frac{1}{3r_{ij}}\right) \quad (1)$$

where Q_i is the charge of the ith (out of n) atom of the molecule, r_{i0} is its distance to the doped site, r_{ij} is its distance to the j^{th} (out of three) neighbor of this site, and ε_0 is the vacuum permittivity.

The nest for the attached molecule can be designed by the following "optimization strategy". We assume that every C atom of the graphene layer is a potential doping site. It can be doped by B or N, depending on which of them decreases the Coulombic interaction energy in [Eq. (1)], where the surrounding atoms are assumed to be charged as if the dopants are isolated. For simplicity, the same approach is used even in the few rare cases where the B and N atoms are first neighbors. The final charges are assigned to the atoms correctly by the rules explained above. By limiting the number of dopants, the strength of the molecule–surface binding could be controlled and the selectivity of the nesting site can be tuned. To realize this idea a negative energy, E_m, called the *level of importance* is introduced, and a given site is doped only when the change of the Coulombic energy associated with this process is $E_{B(N)} < E_m$.

Figure 1 shows the electrostatic potential distribution obtained for the doping pattern designed, with $E_m = -0.02$ eV, on the graphene sheet (bottom) for the AYM peptide (top). This peptide consists of three amino acids, alanine (A), tyrosine (Y) and methionine (M), and has an acetylated N terminus and an amidized C terminus. It is de-

[*] B. Wang, Prof. P. Král
Department of Chemistry, University of Illinois at Chicago
Chicago, IL 60607 (USA)
Fax: (+1) 312-996-0431
E-mail: pkral@uic.edu

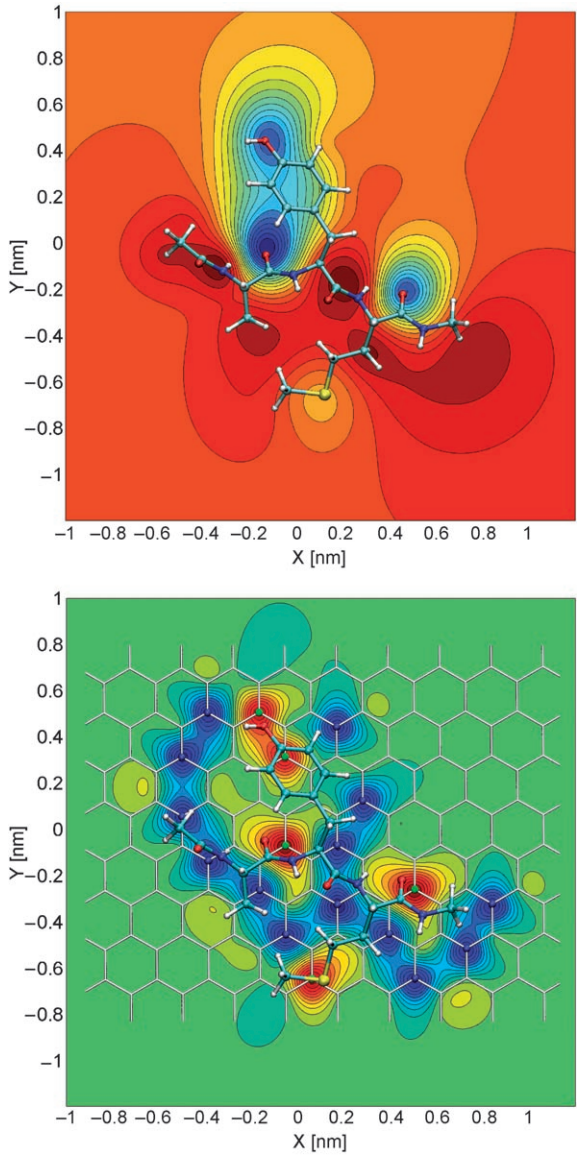

Figure 1. Top: The electrostatic potential distribution formed by the AYM peptide docked on the surface of the doped graphene sheet. Bottom: The "complementary" distribution formed by the doped graphene sheet.

signed (using PyMOL) with a beta-sheet secondary structure ($\phi = -135°$, $\psi = 135°$). The Figure displays the peptide nested above the center of the square graphene layer, formed by 684 C atoms.

The nesting of the peptides was studied by MD simulations[30] using the NAMD package,[31] based on the CHARMM27 force field.[32] The starting configuration of the AYM peptide is obtained from MD simulations at $T = 50$ K. After it is left to relax for 1 ns on the undoped static graphene layer, the plane defined by the three alpha C atoms of its backbone becomes nearly parallel to the layer. The resulting AYM configuration was used to design the initial doping pattern, and then run the MD simulations at $T = 300$ K for 100 ps and at $T = 50$ K for 1 ns, to further optimize the peptide's position. Then the doping profile for the peptide was recalculated, and the whole cycle was repeated once more. Finally, a snapshot of a typical peptide's configuration was taken, shown in Figure 1.

In this optimization process, the peptide adjusts its configuration on the surface spontaneously, and the trap is designed to minimize the total binding energy. We can also apply a different strategy, where one can control the peptide's final configuration by deliberately varying the doping pattern. This process of "configurational control" could be useful in many potential applications.

In Figure 1, we present the electrostatic potentials created by the peptide and the doped graphene sheet. The potentials are calculated in the plane that lies between the molecule and the surface, and is parallel to the surface at a distance of 1.1 Å from it. The positive and negative potentials above the B and N dopants, respectively, can clearly be seen. In almost all the regions where the peptide has a positive potential, the doped surface creates a negative potential, and vice versa. The complementarity of the two potential surfaces can be achieved because the polarization in both systems has a similar *local* character. It allows high-quality docking and immobilization of the peptide in the nesting site.

To quantify these features, Figure 2 (top) plots the Coulombic potential energy of the peptide, E_C, when it is shifted like a rigid body along the x and y coordinates from the docking site, shown in Figure 1. It has a deep minimum of $E_C \approx -1.25$ eV, when the peptide is docked, and it sharply increases and goes to $E_C \approx 0$ eV as the peptide is moved in either direction. In the inset, the full two-dimensional Coulombic potential energy surface of the shifted peptide is shown, where these two orthogonal cuts have been taken. It closely resembles funneling surfaces in protein-folding problems.[35]

Figure 2 (bottom) also presents the AYM peptide's average total binding energy,

$$\langle E_{tot} \rangle = \langle E_C \rangle + \langle E_{vdW} \rangle \quad (2)$$

as a function of the level of importance, E_m. The Coulombic (E_C) and van der Waals (E_{vdW}) energies are calculated for each frame as the difference of energy of the system when the molecule is docked and when it is lifted away from the surface (averaging is done over 1000 consecutive frames of the simulation trajectory, with a 1 ps time interval). Binding of the molecule in the nesting site increases as $E_m \to 0$, since less energetically significant B and N doping sites become also included. Therefore, the total number of dopants grows, as can be seen in the inset. This process allows optimization of the *strength* of the peptide binding by the choice of E_m.

The *selectivity* of the peptide binding can now be estimated and its dependence on the number of dopants can be evaluated. The ability of the nesting site to recognize the peptide for which the nest is designed is limited by the finite lattice constant of the graphene layer, the two (B and N) possible choices for charging of the doping sites, and the number of used dopants. The recognition of the peptide is limited by the recognition of its individual amino acids. To estimate the recognition of each amino acid, we dock in the

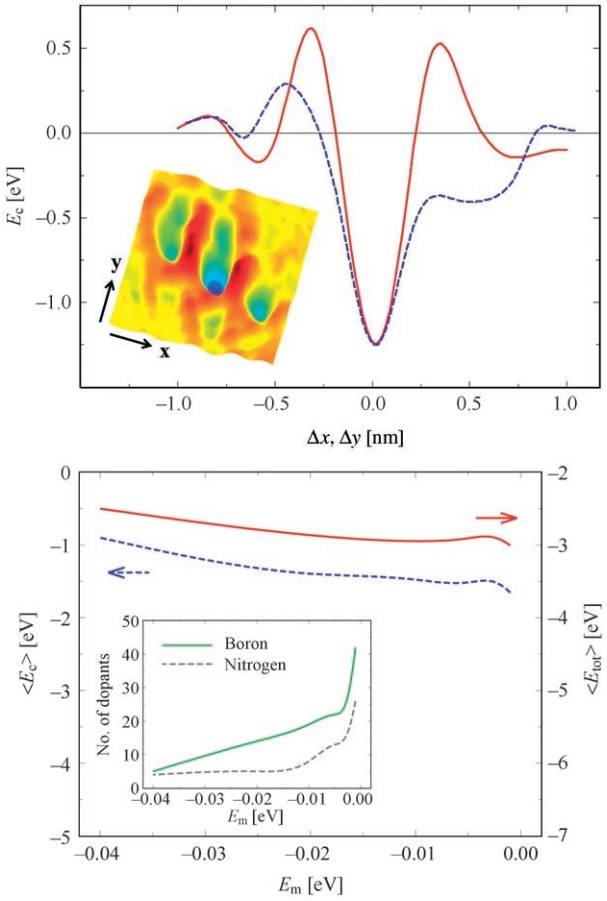

Figure 2. Top: The dependence of the Coulombic (E_C) potential energy of the system, when the peptide is shifted in the *x* (solid) and *y* (dashed) directions along the surface, away from its docking site. In the inset, the full two-dimensional potential surface of E_C is shown. Bottom: The average Coulombic ($\langle E_C \rangle$, dashed line), and total ($\langle E_{tot} \rangle$, solid line) binding energy of the AYM peptide to the doped surface. In the inset, the number of B (solid) and N (dashed) dopants are shown as a function of E_m.

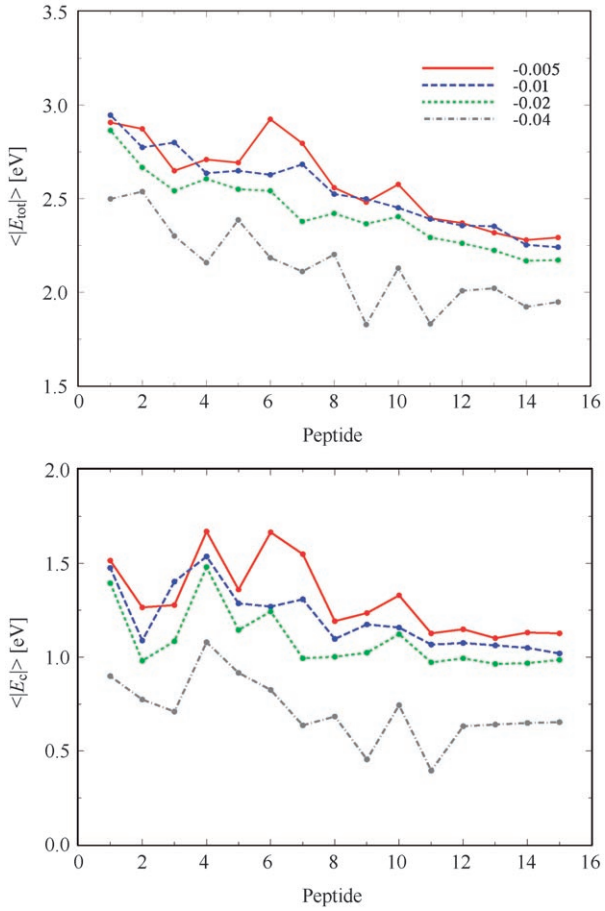

Figure 3. Top: The total binding energy of the 15 peptides at four levels of importance, E_m, shown in the upper-right corner. Bottom: The Coulombic binding energy of the peptides for these E_m values.

nesting site, formed for the AYM peptide, other peptides that differ from it just in the middle Y residue. We consider amino acids that have neutral sidechains, and calculate the binding energy of each peptide.

In Figure 3 (top and bottom), we plot the total ($\langle E_{tot} \rangle$) and Coulombic ($\langle E_C \rangle$) binding energies of the selected 1–15 peptides: AYM, AWM, AFM, ASM, AQM, ANM, AHM, AMM, ATM, ACM, ALM, AVM, AIM, AGM, and AAM, respectively. The results are calculated in the nesting site designed for the AYM peptide with the levels of importance: $E_m = -0.005$, -0.01, -0.02, and -0.04 eV. The results for the 15 peptides are ordered, as shown above, according to their total binding energies, $\langle E_{tot} \rangle$, obtained for $E_m = -0.01$ and -0.02 eV. We can see that the AYM peptide, and thus the Y residue, is very well recognized by the doping pattern designed for it, since its total binding energy, $\langle E_{tot} \rangle$, is 0.15–0.65 eV larger than in the other peptides. Even though its van der Waals binding energy, $\langle E_{vdW} \rangle$, is relatively large, the main contribution comes from the well-tuned Coulombic binding to the nesting site. It can also be seen in Figure 3 (bottom) that its $\langle E_C \rangle$values are among the largest under all the levels of importance.

Figure 4 displays the electrostatic potential distribution around the middle residues of the No. 2–5 peptides, docked in the nesting place. The S and Q residues have strong Coulombic binding to the surface, due to the N dopant located at $x \approx 0.275$, $y \approx 0.139$. On the other hand, the W and F residues have relatively strong van der Waals binding to the surface. In general, when there are too few dopants, the doping pattern does not have enough parameters to recognize the structure. When there are too many dopants, the structure can be recognized, but the chance that some groups in other peptides also bind to them increases.

Let us now generalize this methodology to designing selective nesting sites in polar solvents. Binding sites with larger charges are needed to keep the molecules attached in their nests. To achieve this goal, the previous algorithm is modified by replacing the B and N dopants by ammonium cations (NH_4^+) and carboxylate anions (CO_2^-), respectively, as charged ligands covalently attached to the surface.[36] The ligands are not allowed to be neighbors, because of their big sizes. This approach is applied to the graphene layer in water to calculate the charges of the covalently attached ligands by ab initio treatment on the Hartree–Fock level,

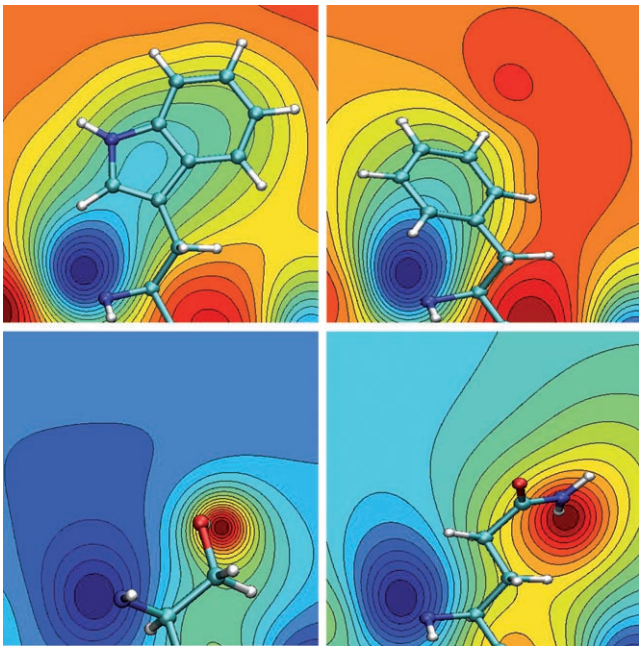

Figure 4. The electrostatic potential distribution around the residues W (upper left), F (upper right), S (lower left), and Q (lower right), plotted in the reduced region $-0.299 < x < 0.530$, $-0.180 < y < 0.649$.

with a polarized continuum model of the water solvent.[29] In the ammonium cation, the N and H atoms are charged by −0.78 and 0.467 e, respectively, while the carbon atom bonded to the ammonium group and its first neighbors are charged by −0.328 and 0.177 e, respectively. In the carboxylate anion, the C and O atoms are charged by 0.98 and −0.8 e, respectively, while the carbon atom bonded to the carboxyl group and its first neighbors are charged by −0.67 and 0.19 e, respectively.

The modified methodology was tested on small green fluorescence protein (GFP).[37] It is roughly formed by a "cylinder" made of beta sheet strands that protect a photoactive group in its interior. On the GFP surface, two neighboring strands are found (105–130 residues) that are relatively highly charged. Their spatial structure was obtained from the PDB Bank (ID 1W7S), and PyMOL was used to acetylate the N terminus of residue 105, and amidize the C terminus of residue 130.

The above-described approach was used to design on the graphene layer a nest for the exterior (charged) part of these two strands, instead of the whole GFP protein. The initial configuration of the strands was chosen in such a way that the plane of their backbones is approximately parallel to the surface, with a distance of about 10 Å from it. Then, the ligand pattern for this configuration was designed with the level of importance of $E_m = 1.0$ eV. A 10 ns MD simulation[30] of this system was run in a periodic cell of $56 \times 43 \times 45$ Å^3, with 2769 water molecules and at $T = 300$ K. During the simulation, the α carbons of the 105th and 130th residues were fixed to model the fact that these two residues should be connected to the rest of GFP. The graphene surface and the charged ligands were also fixed during the simulation.

Figure 5 (top) shows the initial and final configurations of the two GFP strands above the graphene layer with attached ligands in water (hidden). The charged residues in the initial configuration and the ligands on the graphene

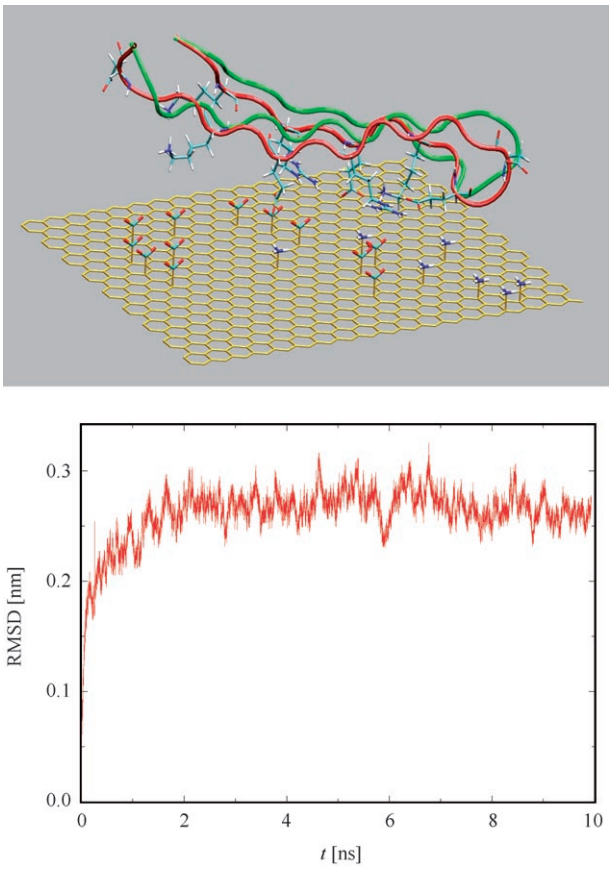

Figure 5. Top: Initial (red) and final (green) structures of the 105–130 residues from the GFP that are nested on the ligand-doped graphene layer. The peptide sequence is NY**KTR**A**E**V**K**F**E**G**D**TLVN**RIEL**K-GI**D**F, where the ten bold residues are charged and displayed in the initial structure. Bottom: The RMSD of this peptide obtained during the simulation.

layer are displayed in atomistic detail. During the simulation, the two strands shift their position somewhat, but they remain attached to the ligands. Figure 5 (bottom) shows a calculation by VMD[38] of the root-mean-square deviation (RMSD) of all atoms between the initial and the actual configurations of the two-strand peptide, obtained at each time frame. The RMSD stabilizes after 2.0 ns and fluctuates within 0.5 Å. Therefore, the peptide could be selectively nested above the predesigned ligand pattern on the graphene layer in water, without being dissolved or significantly modified.

The same approach could be used in diluted ionic solutions.[39] For larger concentrations of ions, more ligands will be blocked by ions, in accordance with the Langmuir-isotherm description.[40] In order to prevent the peptide from dissolving, due to the blockage of binding sites, the level of importance must be decreased in the design of the nest. Then, more ligands will be attached to the surface and some

will be left to bind the peptide, even if a significant fraction of them is blocked. Clearly, the recognition is getting worse, since the coverage of the ligands is random. In a more realistic modeling, one should also include the pH of the solution and match it to the protein and the charged ligands on the surface.

The presented examples demonstrate that one could design selective nesting sites for biomolecules on the surface of materials in water. Practical doping and functionalization of the surfaces by atoms and ligands, respectively, could be realized by STM and AFM techniques.[41] This methodology opens new avenues in the design of hybrid functional systems on many types of surfaces, with numerous potential applications.

Keywords:
ab initio calculations · biomolecules · doping · nanobiotechnology · peptides

[1] J. M. Thornton, A. E. Todd, D. Milburn, N. Borkakoti, C. A. Orengo, *Nat. Struct. Biol.* **2000**, *7*, 991 – 994.
[2] P. Aloy, R. B. Russell, *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 188 – 197.
[3] M. E. M. Noble, J. A. Endicott, L. N. Johnson, *Science* **2004**, *303*, 1800 – 1805.
[4] M. E. Anderson, L. S. Higgins, H. Schulman, *Nat. Clin. Pract. Cardiovasc. Med.* **2006**, *3*, 437 – 445.
[5] S. W. Lee, B.-K. Oh, R. G. Sanedrin, K. Salaita, T. Fujigaya, C. A. Mirkin, *Adv. Mater.* **2006**, *18*, 1133 – 1136.
[6] T. Aqua, R. Naaman, S. S. Daube, *Langmuir* **2003**, *19*, 10573 – 10580.
[7] M. Ratner, *Nature* **2005**, *435*, 575 – 577.
[8] H. Korri-Youssoufi, F. Garnier, P. Srivastava, P. Godillot, A. Yassar, *J. Am. Chem. Soc.* **1997**, *119*, 7388 – 7389.
[9] M. Sarikaya, C. Tamerler, A. K. Y. Jen, K. Schulten, F. Baneyx, *Nat. Mater.* **2003**, *2*, 577 – 585.
[10] V. Noireaux, R. Bar-Ziv, A. Libchaber, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 12672 – 12677.
[11] W. Yang, O. Auciello, J. E. Butler, W. Cai, J. A. Carlisle, J. E. Gerbi, D. M. Gruen, T. Knickerbocker, T. L. Lasseter, J. N. Russell, L. M. Smith, R. J. Hamers, *Nat. Mater.* **2002**, *1*, 253 – 257.
[12] J. Wang, M. A. Firestone, O. Auciello, J. A. Carlisle, *Langmuir* **2004**, *20*, 11450 – 11456.
[13] K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, Y. Zhang, S. V. Dubonos, I. V. Grigorieva, A. A. Firsov, *Science* **2004**, *306*, 666 – 669.
[14] S. Stankovich, D. A. Dikin, G. H. B. Dommett, K. M. Kohlhaas, E. J. Zimney, E. A. Stach, R. D. Piner, S. T. Nguyen, R. S. Ruoff, *Nature* **2006**, *442*, 282 – 286.
[15] S. Ferro, A. De Battisti, *Anal. Chem.* **2003**, ***75***, 7040 – 7042.
[16] B. Wang, P. Král, I. Thanopulos, *Nano Lett.* **2006**, *6*, 1918 – 1921.
[17] B. Wang, P. Král, *J. Am. Chem. Soc.* **2006**, *128*, 15984 – 15985.
[18] J. Madoz, B. A. Kuznetzov, F. J. Medrano, J. L. Garcia , V. M. Fernandez, *J. Am. Chem. Soc.* **1997**, *119*, 1043 – 1051.
[19] D. I. Meyer, D. I. E. Krause, B. Dobberstein, *Nature* **1982**, *297*, 647 – 650.
[20] D. B. Kitchen, H. Decornez, J. R. Furr, J. Bajorath, *Nat. Rev. Drug Discovery* **2004**, *3*, 935 – 949.
[21] S. Forrest, *Science* **1993**, *261*, 872 – 878.
[22] J. T. Pedersen, J. Moult, *Curr. Opin. Struct. Biol.* **1996**, *6*, 227 – 231.
[23] N. Pokala, T. M. Handel, *J. Mol. Biol.* **2005**, *347*, 203 – 227.
[24] G. L. W. Hart, V. Blum, M. J. Walorski, A. Zunger, *Nat. Mater.* **2005**, *4*, 391 – 394.
[25] D. J. Tannor, R. Kosloff, S. A. Rice, *J. Chem. Phys.* **1986**, *85*, 5805 – 5820.
[26] A. P. Peirce, M. A. Dahleh, H. Rabitz, *Phys. Rev. A* **1988**, *37*, 4950 – 4964.
[27] P. Král, *Phys. Rev. B* **2002**, *66*, 241401.
[28] O. Stephan, P. M. Ajayan, C. Colliex, P. Redlich, J. M. Lambert, P. Bernier, P. Lefin, *Science* **1994**, *266*, 1683 – 1685.
[29] M. J. Frisch et al., Gaussian 03, Revision C.02, Gaussian, Inc., Wallingford CT, **2004**.
[30] We estimate parameters of atoms in aliphatic groups from similar atom types, and add them to the CHARMM27 force field. The systems are equilibrated as an NPT ensemble (fixed number of particles, pressure and temperature, variable volume) with periodic boundary conditions, and the pressure is kept constant ($P = 1$ atm) using the Langevin–Piston method.[33] The temperature is kept constant by Langevin dynamics, with a damping coefficient of 10 ps^{-1} applied, and the time step is always 1 fs. The long-range electrostatic forces are computed by the particle-mesh Ewald method.[34]
[31] L. Kalé, R. Skeel, M. Bhandarkar, R. Brunner, A. Gursoy, N. Krawetz, J. Phillips, A. Shinozaki, K. Varadarajan, K. Schulten, *J. Comput. Phys.* **1999**, *151*, 283 – 312.
[32] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M. Karplus, *J. Phys. Chem. B* **1998**, *102*, 3586 – 3616.
[33] S. E. Feller, Y. H. Zhang, R. W. Pastor, B. R. Brooks, *J. Chem. Phys.* **1995**, *103*, 4613 – 4621.
[34] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, *98*, 10089 – 10092.
[35] P. E. Leopold, M. Montal, J. N. Onuchic, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 8721 – 8725.
[36] D. Tasis, N. Tagmatarchis, A. Bianco, M. Prato, *Chem. Rev.* **2006**, *106*, 1105 – 1136.
[37] J. J. van Thor, G. Y. Georgiev, M. Towrie, J. T. Sage, *J. Biol. Chem.* **2005**, *280*, 33652 – 33659.
[38] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graphics* **1996**, *14*, 33 – 38.
[39] F. Carlsson, E. Hyltner, T. Arnebrant, M. Malmsten, P. Linse, *J. Phys. Chem. B* **2004**, *108*, 9871 – 9881.
[40] I. Langmuir, *J. Am. Chem. Soc.* **1916**, *38*, 2221 – 2295.
[41] After the submission of this work a related study was published: A.-S. Duwez, S. Cuenot, C. Jérôme, S. Gabriel, R. Jérôme, S. Rapino, F. Zerbetto, *Nat. Nanotechnol.* **2006**, *1*, 122 – 125.

Received: August 20, 2006
Revised: December 8, 2006
Published online on February 27, 2007

