Endohedral Confinement of a DNA Dodecamer onto Pristine Carbon Nanotubes and the Stability of the Canonical B Form

Fernando J.A.L. Cruz,1,2,* Juan J. de Pablo2,3 and José P.B. Mota1

1Requimte/CQFB, Departamento de Química, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Caparica, 2829-516, Portugal. E-mail: fj.cruz@fct.unl.pt
2Department of Chemical and Biological Engineering, University of Wisconsin-Madison, Madison, Wisconsin, 53706, USA
3Institute of Molecular Engineering, University of Chicago, Chicago, Illinois, 60637, USA

Although carbon nanotubes are potential candidates for DNA encapsulation and subsequent delivery of biological payloads to living cells, the thermodynamical spontaneity of DNA encapsulation under physiological conditions is still a matter of debate. Using enhanced sampling techniques, we show for the first time that, given a sufficiently large carbon nanotube, the confinement of a double-stranded DNA segment, 5'-D(CP*GP*CP*GP*AP*AP*TP*TP*CP*GP*CP*G)-3', is thermodynamically favourable under physiological environments (134 mM, 310 K, 1 bar), leading to DNA-nanotube hybrids with lower free energy than the unconfined biomolecule. A diameter threshold of 3 nm is established below which encapsulation is inhibited. The confined DNA segment maintains its translational mobility and exhibits the main geometrical features of the canonical B form. To accommodate itself within the nanopore, the DNA's end-to-end length increases from 3.85 nm up to approximately 4.1 nm, due to a ~ 0.3 nm elastic expansion of the strand termini. The canonical Watson-Crick H-bond network is essentially conserved throughout encapsulation, showing that the contact between the DNA segment and the hydrophobic carbon walls results in minor rearrangements of the nucleotides H-bonding. The results obtained here are paramount to the usage of carbon nanotubes as encapsulation media for next generation drug delivery technologies.

I. INTRODUCTION

Deoxyribonucleic acid (DNA) and single-walled carbon nanotubes (SWCNTs) are prototypical one-dimensional structures; the former plays a central role in chemical biology and the latter holds promise for nanotechnology applications1, 3. From the point of view of biological purposes and DNA manipulation, carbon nanotubes have been proposed to be used as templates for DNA encapsulation 4, 5, intracellular penetration via endocytosis and delivery of biological payloads 6, 7, and ultrafast nucleotide sequencing 2, 8, 9. While structure in its natural form and environment is well established (e.g. B-DNA in aqueous solution), their interactions have been the subject of intense investigation 4, 5, 8, 10-16, nonetheless, the corresponding molecular-level phenomena remain rather unexplored. Is confinement thermodynamically spontaneous (free-energy)? How important are the conformational properties of the encapsulated double-strand (entropy)? How does the confinement process depend on nanotube diameter? Moreover, previous work has focused almost exclusively on exoadsorption of DNA on the external surface of SWCNTs 8, 11, 17-19, overlooking the possibility of endohedral confinement. Nonetheless, it is well known that the conformational properties of biopolymers under confinement are of crucial relevance in living systems (e.g., DNA packing in eukaryotic chromosomes, viral capsids).

DNA encapsulation is a phenomenon that remains utterly unmapped. Most of the earlier work focused on temperatures markedly distinct from the physiological value, precluding extrapolation of results to in vivo conditions. The pioneer work of Lau et al. 5 showed that a small DNA strand, initially confined in a 4 nm diameter nanotube, exhibits dynamics similar to the unconfined molecule, but that behaviour is drastically altered when diameter is decreased to 3 nm. Because the DNA had been artificially inserted, no information about the encapsulation process itself, e.g., its thermodynamical spontaneity and kinetics, was provided. Previous experiments and calculations indicate that the biomolecule can be confined onto D = 2.7 nm SWCNTs 13 and D = 3–4 nm multi-walled carbon nanotubes (MWCNTs) 20, 21, however, the corresponding working temperatures (350 – 400K) were far too high to have any physiological relevance. Furthermore, the experimental observations of Iijima et al. indicated that encapsulation of DNA onto MWCNTs was a competing mechanism with wrapping of the biomolecule around the external wall 20, their reported data failed to identify the relevant conditions upon which the confinement process is favoured, such as ionic strength of the media and temperature. Recently, Mogurampelly and Mati addressed 22 the encapsulation of dsDNA and siRNA onto SWCNTs, and established a critical diameter of D = 2.67 nm and D = 2.4 nm for the former and the latter, respectively, below which confinement was completely inhibited, and attributed it to a large free-energy barrier associated with the nanopore entrance.

Our results obtained under precise physiological conditions (310 K, 1 bar, [NaCl] = 134 mM) show that an atomically detailed DNA dodecamer can be encapsulated onto a D = 4 nm SWCNT, resulting in a decrease of the system’s
free energy. Encapsulation kinetics is fast (<16 ns) and the double strand retains its translational mobility within the nanotube. Very interestingly, our data indicate that the encapsulated molecule free-energy minima correspond to a DNA conformation similar to the bulk canonical B form, with a pitch length of 3.4 nm (10 bps repeating unit) and a double strand end-to-end length of approximately 4.1 nm. The canonical Watson-Crick H-bonds network is roughly maintained throughout confinement, exhibiting probability distributions essentially corresponding to more than 75% of hydrogen bonds existence. As far as we are aware these observations are the first of their kind, and they come to pave the way for the design of smart nanotube based devices for in vivo DNA encapsulation.

II. MODELS AND METHODS

A. Molecular Models

All molecules in this work are described using atomistically detailed force fields, including electrostatic charges in each atom. The dispersive interactions are calculated with the Lennard-Jones (12,6) potential, cross parameters between unlike particles described by the classical Lorentz-Berthelot mixing rules, and electrostatic energies described by Coulomb’s law. DNA is modelled as a completely flexible molecule within the framework of the AMBER99sb-ildn force-field, the corresponding potential energies associated with bond stretching and angle bending are calculated with harmonic potentials, whilst the dihedral angles, $U(\phi)$, are computed via Ryckaert-Bellemans functions, $U(\phi) = \sum_{dihedral} \sum_{i=0}^{5} C_i [\cos(\phi - 180\degree)]$; we have included in the potential the refinements recently proposed by Lindorff-Larsen et al., which result in improved accuracy of the DNA backbone dihedrals. The Na$^+$ and Cl$^-$ ions are described using the parameterization of Aquist and Dang and the water molecules by the TIP3P force field of Jorgensen and co-workers; a recent discussion on the NaCl force-field influence upon the static and dynamic properties of nucleic acids under physiological conditions can be found elsewhere. To maintain computational tractability, we have chosen the double-stranded B-DNA Dickerson dodecamer, exhibiting a pitch of ~3.4 nm corresponding to an average of 10–10.5 base-pairs per turn over the entire helix, and with a double-strand end-to-end length of ~3.8 nm measured between terminal (GC) base pairs; the well-known A-DNA form has a pitch length of ~2.6 nm with an average of 11 base-pairs per turn. Considering that the B-DNA backbone phosphorus atoms lie on a cylindrical surface, the average diameter of the double-strand corresponds to ~2 nm. Although explicitly smaller in length than genomic DNA, the Dickerson dodecamer main structural features resemble those of genomic E. coli DNA, namely in the radius of gyration and double-strand backbone diameter, $R_g \approx 0.7–1$ nm and $D \approx 2$ nm.

Recently, large diameter ($D \approx 4$ nm) single-walled carbon nanotubes have been prepared using supported-catalyst chemical vapour deposition. In order to probe the confinement of DNA into such large, hollow nanostructures, we have adopted two different diameter SWCNTs with zigzag symmetry, both with length $L = 8$ nm; the skeletal diameters, measured between carbon centres on opposite sides of the wall, are $D = 4$ nm (51,0) and $D = 3$ nm (40,0). The walls are built up of hexagonally-arranged $sp^2$ graphitic carbon atoms, with a C–C bond length of 1.42 Å, whose Lennard-Jones potential is given by Steele’s parameterization ($\sigma = 0.34$ nm, $\varepsilon = 28$ K). The positions of all solid atoms are fixed throughout the calculations.

B. Methodology and Algorithms

Large simulation cells with dimensions $(x, y, z) = 11.11 \times 15$ nm were built according to the following stepwise procedure: (i) initially the solid was placed inside the empty cell, aligned along the $z$-axis, and the DNA molecule inserted at a distance of 0.5 nm away from the nanopore entrance, (ii) the whole cell was solvated with $\text{H}_2\text{O}$ ($\rho = 1$ g/cm$^3$) and the appropriate number of $\text{Na}^+$ and $\text{Cl}^-$ ions were added to ensure physiological ionic strength, $[\text{NaCl}] = 134$ mM, (iii) then the whole system was energy-minimized and equilibrated during at least 0.8 ns, first in the canonical ensemble and then in the isothermal-isobaric ensemble, resulting in a fully equilibrated simulation cell, as observed by the constancy of the main thermodynamical parameters (potential energy, temperature, volume and pressure). During minimization and equilibration the DNA position was constrained; once these steps were completed, the biomolecule was unconstrained, production runs were started and data collected over a time interval of 0.07–0.1 μs. The box dimensions were chosen in order to accommodate a 3.5 nm solvation shell in all directions around the nanotube. Molecular dynamics (MD) simulations in the isothermal–isobaric ensemble (NpT) were performed using the Gromacs set of routines. Newton’s equations of motion were integrated with a time step of 1 fs and using a Nosé-Hoover thermostat and a Parrinello-Rahman barostat to maintain temperature and pressure at 310 K and 1 bar. A potential cut-off of 1.5 nm was employed for both the van der Waals and Coulombic interactions, and the long-range electrostatics were calculated with the particle-mesh Ewald method using cubic interpolation and a maximum Fourier grid spacing of 0.12 nm. Three-dimensional periodic boundary conditions were applied. The well-tempered metadynamics scheme of Barducci and Parrinello was employed to obtain the free-energy landscape associated with the confinement mechanism. The well-tempered algorithm biases Newton dynamics by adding a time-dependent Gaussian potential, $V(\xi, t)$, to the total (unbiased) Hamiltonian, preventing the system from becoming permanently trapped in local energy minima and thus leading to a more efficient exploration of the phase space. The potential $V(\xi, t)$ is a function of the so-called collective variables (or order parameters), $\xi(q) = [\xi_1(q), \xi_2(q), \ldots, \xi_n(q)]$, which in turn are related to the microscopic coordinates of the real system, $q$, according to equation (1):
\[ V(\mathbf{z}(q), t) = W \sum_{r=0}^{t_{\text{tot}}} \exp \left( -\frac{V(\mathbf{z}(q(r)), t)}{\Delta T} \right) \exp \left( -\frac{\sum_{i=1}^{n} (\mathbf{z}(q) - \mathbf{z}(q(r)))^2}{2\sigma_i^2} \right) \]  

where \( t \) is the simulation time, \( W = \tau_G \omega \) is the height of a single Gaussian, \( \tau_G \) is the time interval at which the contribution for the bias potential, \( V(\mathbf{z}, t) \), is added, \( \omega \) is the initial Gaussian height, \( \Delta T \) is a parameter with dimensions of temperature, \( \sigma_i \) is the Gaussian width and \( n \) is the number of collective variables in the system; we have considered \( \tau_G = 0.1 \text{ ps} \), \( \omega = 0.1 \text{ kJ/mol} \), \( \Delta T = 310 \text{ K} \) and \( \sigma = 0.1 \text{ nm} \). 

The parameter \( \Delta T \) determines the rate of decay for the height of the added Gaussian potentials and when \( \Delta T \rightarrow 0 \) the well-tempered scheme approaches an unbiased simulation. The SWCNTs are primarily one-dimensional symmetric, therefore we decided to construct the free-energy landscape in terms of two collective variables, that correspond to the positional vector of the centre of mass of the biomolecule \((\mathbf{R}_{\text{DNA}} - \mathbf{R}_{\text{SWCNT}})\) and of the nanotube \((\mathbf{R}_{\text{SWCNT}} - \mathbf{R}_{\text{GC}})\), projected along the \( z \)-axis, or of the terminal (GC) nucleobase pairs at the double-strand termini, \((\mathbf{R}_{\text{GC}} - \mathbf{R}_{1})\) and \((\mathbf{R}_{12} - \mathbf{R}_{\text{GC}})\). According to our definition of collective variables, \( \xi_1 \) corresponds to the \( z \)-distance between the biomolecule and the nanopore centre and \( \xi_2 \) can be interpreted as the DNA end-to-end length. The characteristic lengths of the nanotube and Dickerson dodecamer are, respectively, \( L = 8 \text{ nm} \) and \( L = 3.8 \text{ nm} \), and therefore any value \( \xi_1 = \Delta L = (L^{\text{DNA}} - L^{\text{SWCNT}})/2 < 2.1 \text{ nm} \) corresponds to a DNA–SWCNT hybrid in which the biomolecule is completely encapsulated within the solid; the threshold \( \xi_1 > 5.9 \text{ nm} \) obviously indicates the absence of confinement. At the end of the simulation, the three-dimensional free-energy surface is constructed by summing the accumulated time-dependent Gaussian potentials according to \( F(\mathbf{z}, t) = -\frac{T + \Delta T}{\Delta T} V(\mathbf{z}, t) \). A discussion of the algorithm’s convergence towards the correct free-energy profile is beyond the scope of this work and can be found elsewhere, \( \delta V(\mathbf{z}, t)/\delta t \rightarrow 0 \), the well-tempered method leads to a converged free-energy surface. An alternative approach to obtain the time-independent free-energy surface relies on integrating \( F(\mathbf{z}, t) \) at the final portion of the metadynamics run. The converged free-energy can thus be mathematically obtained from equation (2):

\[ F(\mathbf{z}) = -\frac{1}{\tau_{\text{tot}}} \int_{t_{\text{tot}}-\tau}^{t_{\text{tot}}} F(\mathbf{z}, t) \]  

where \( t_{\text{tot}} \) is the total simulation time and \( \tau \) is the time window over which averaging is performed. We have implemented a convergence analysis for each collective variable, \( \xi_1 \) and \( \xi_2 \), splitting the last 40 ns of simulation time into \( \tau = 4 \text{ ns} \) windows, and the results show that the bias potential \( V(\mathbf{z}, t) \) has converged, and thus the three-dimensional surface of Figure 2 is a good estimator of the free-energy changes associated with molecular encapsulation.

Independent calculations were performed using the umbrella sampling technique \( 35, 46 \). For a system composed of \( N \) particles, the method consists in biasing the classical (unbiased) Hamiltonian that depends on the potential, \( U(\mathbf{r}_N) \), and kinetic energies, \( E_{\text{kin}}(\mathbf{p}_N) \), by introducing a time independent harmonic potential, \( V(\mathbf{z}_i) = c/k (\mathbf{z}_i - \mathbf{z}_i^0)^2 \), according to \( H(\mathbf{r}_N, \mathbf{p}_N, \mathbf{z}_i) = U(\mathbf{r}_N) + E_{\text{kin}}(\mathbf{p}_N) + V(\mathbf{z}_i) \); \( k \) is the harmonic force constant, \( \mathbf{z}_i \) is an order parameter and \( \mathbf{z}_i^0 \) corresponds to the position of the umbrella restrain; in the present case, \( \mathbf{z}_i = \mathbf{R}_{\text{DNA}} - \mathbf{R}_{\text{SWCNT}} \) corresponds to the absolute three-dimensional distance between the centres of mass of the double-strand and the SWCNT, and \( \mathbf{z}_i^0 = 4.1 \text{ nm} \), in direct analogy with the collective variables defined in the well-tempered metadynamics algorithm, \( \xi_1 \) and \( \xi_2 \). When such a biasing potential is used, the biased probability distribution of the system, \( \mathbf{p}^b(\mathbf{z}_i) \), can be obtained from a Boltzmann weighted average along the \( \mathbf{z}_i \) order parameter and, therefore, assuming that the system is ergodic \( 46 \):

\[ p^b(\mathbf{z}_i) = \frac{\exp[-\beta(U(\mathbf{r}) + V(\mathbf{z}_i(\mathbf{r})))]}{\int \exp[-\beta(U(\mathbf{r}) + V(\mathbf{z}_i(\mathbf{r})))]d^N \mathbf{r}} \]  

where \( \beta = 1/k_BT \), \( k_B \) is the Boltzmann constant, \( \delta \) is the Dirac delta function, and \( N \) is the total number of particles in the system. Because the biasing potential depends only on the order parameter \( \mathbf{z}_i \), and the integration in the numerator is performed over all degrees of freedom except \( \mathbf{z}_i \), the unbiased probability of the real system, \( \mathbf{p}^u(\mathbf{z}_i) \), can be evaluated from equation (4):

\[ p^u(\mathbf{z}_i) = p^b(\mathbf{z}_i) \exp^{b\mathbf{z}_i(\mathbf{z}_i)} \Gamma \]  

where \( \Gamma \) contains the normalisation factor and the term \( b\mathbf{z}_i(\mathbf{z}_i) \) corresponds to the number of windows used in the calculation.
where $\Gamma = -(1/\beta) \ln \left( e^{-\beta V(\Omega)} \right)$ is independent of $\Omega$, and the triangular brackets denote an ensemble average. The reconstruction of the true (unbiased) free energy profile or potential of mean force $47,\, 48$, consistent with the Gibbs free energy, $\text{PMF}(\Omega) = -k_B T \ln P^\theta(\Omega)$, is accomplished using the weighted histogram analysis method $47-49$.

C. Parametric Analysis

The radius of gyration, $R_g$, gives a measure of a molecule’s compactness and is defined by equation 5:

$$ R_g = \sqrt{\frac{1}{N} \sum_{i=1}^{N} r_i^2 m_i} $$

where $N = 758$ is the total number of atoms in the DNA molecule, $m_i$ is the mass of atom $i$ and $r_i$ is the positional vector of the atom relative to the molecular center of mass. The root mean squared deviation, $\text{RMSD}$, is obtained by calculating the distance $r_{ij}$ between atoms $i$ and $j$ at time $t$, and comparing with the same distance observed at time $t = 0$, according to:

$$ \text{RMSD} = \sqrt{\frac{1}{n} \sum_{i,j=1}^{N} \sum_{t=1}^{T} (r_{ij}(t) - r_{ij}(0))^2} $$

### III. RESULTS AND DISCUSSION

Confinement of double-stranded DNA (dsDNA, $30$) into a $(51,0)$ nanotube ($D = 4$ nm) is fast and becomes complete within the first 16 ns of observation time. Initially, the double strand is in the bulk ($0-2$ ns) and as it diffuses towards the SWCNT entrance undergoes structural rearrangements leading to minor increases in pitch length, $P$, and end-to-end distance, $L$ (Fig. 1A). After 2 ns, the dodecamer is already at the nanopore entrance, where it experiences strong van der Waals attractions towards the solid $30$, resulting in complete encapsulation at 15.4 ns, after which the double-strand relaxes to $P \approx 3.4$ nm and $L \approx 4 - 4.1$ nm. It is very interesting to observe that confinement appears to be permanent, i.e., the DNA fragment never returns to the bulk solution during the observation time window, maintaining direct local contact with the solid wall at a distance of closest approach of ca. 2.6 Å. Nonetheless, the encapsulated molecule clearly retains its translational mobility, diffusing freely along the nanopore main axis (Fig. 1B). The encapsulation mechanism can be divided into a three-step process: (i) fast diffusion of DNA towards the nanopore entrance ($0-2$ ns), (ii) strong van der Waals attractions towards the solid, leading to confinement of terminus 1 at 2.33 ns, followed by structural rearrangements of the whole double strand occurring in bulk solution ($t < 15$ ns) and finally (iii) penetration of terminus 2 into the confining volume resulting in complete encapsulation of the biomolecule.

In contrast, the narrowness of a $(40,0)$ topology ($D = 3$ nm) completely inhibits encapsulation even over an observation time of 0.1 μs. Instead, the DNA contact with the smaller nanopore results in the occurrence of two distinct states: (i) exoesorption onto the external surface of the nanotube via a π-π stacking mechanism of a terminal (GC) nucleobase pair onto the graphic surface, and threading of the biomolecule along the hydrophobic cylinder similar to previous observations $11$, or (ii) trapping of the DNA at the nanopore entrance, with partial melting of the double-strand.
metadynamics encapsulation is probed by the free-energy barrier located at the nanopore entrance. The decomposition of the interaction energies between DNA and the surrounding environment, solid and solution, can help to throw some light onto this issue. Upon confinement onto the (51,0) topology, the dodecamer becomes less solvated by the H2O molecules than in the bulk, as indicated by a decrease of the DNA/H2O interaction energy of 1327 kJ/mol, and leading to a thin hydration shell of ca. 1 ± 2 water molecules between the biomolecule and the solid walls. However, this instabilization is roughly compensated by an increase of the DNA/ions interactions of ~ 851 kJ/mol and also by the intrinsic effect exerted by the hydrophobic solid upon the DNA van der Waals cloud of ~ 442 kJ/mol. On the other hand, molecular epxoadsorption onto the (40,0) nanotube prevents stabilization of the biomolecule from the dispersive interaction with the graphitic walls, leading to a diminished DNA/(40,0) interaction energy of ~ 170 kJ/mol, clearly insufficient to overcome the energetic penalty of a decreased solvation effect caused by encapsulation.

At physiological conditions the canonical B-form is dsDNA most stable configuration, however, little is known when the molecule is confined into a strongly hydrophobic solid, such as a carbon nanotube. The data recorded in Fig. 1A suggest that this is also the case for encapsulated dsDNA. In fact, when simulation results are used to determine distribution histograms, the corresponding frequencies are well correlated by Gaussian statistics exhibiting a pitch length of $P = 3.4$ nm, for a 10 nucleobase pair repeating unit, and a double strand length of $L = 4.1$ nm (Fig. 1C), consistent with the geometrical characteristics of B-DNA in bulk solution.

We have determined the biomolecular characteristic lengths—radius of gyration ($R_g$) and its projection along the nanotube main axis ($R_g^z$) as well as the root-mean squared deviation (RMSD, equation 6) from the B-DNA Dickerson dodecamer used as starting configuration and plotted them in Fig. 1D. Because the RMSD compares the structure at any time $t$ with the original DNA structure ($t = 0$), the blue curve in Fig. 1D indicates a minor relaxation of the double-strand from the crystal structure to accommodate liquid state flexibility, whilst maintaining the relative average distance between each atom in the double-strand. After 20 ns, the data converged smoothly to average values of RMSD $= 0.36 ± 0.002$ nm and $R_g^z = 1.02 ± 0.0015$ nm; the latter value indicates an alignment of the biomolecule along the axial axis of the nanopore. The $R_g^z$ of 0.8 nm depression observed at 56.6 ns is transient and matches a total number of Watson-Crick H bonds of 28.

The thermodynamical stability of molecular encapsulation is probed by the free-energy ($F$) differences associated with the process, using well-tempered metadynamics. We have chosen two order parameters to construct the free-energy landscape, $\xi_1$ and $\xi_2$, where $\xi_1$ is the distance between centres of mass of the dsDNA and nanotube, projected along the nanopore main axis ($z$), and $\xi_2$ is the absolute distance between the (GC) termini on opposite sides of the double-strand, equivalent to the DNA end-to-end length (cf. Models and Methods). An inspection of the resulting 3-D surface (Figure 2) reveals the existence of five distinct free-energy minima, sharing in common the fact that all are located at discrete positions along the internal volume of the nanopore, $\xi_1 < 1.8$ nm; the absolute minimum at $\xi_2 = 0.117$ nm indicates that the center of the nanotube is the most energetically stable region for the encapsulated biomolecule, which results in the strongest concentration of molecular density at that location.

Owing to their thermodynamical similarity, the five free-energy minima along $\xi_2$ provide a minimum free-energy path along which DNA can translate within the pore, visiting maximum probability configurations as indicated by the dotted line in Figure 2. In order to escape from those deep free-energy valleys, $F(\xi_1, \xi_2) \sim 40$ kJ/mol, DNA has to overcome large energetic barriers, rendering the exit process towards the bulk solution thermodynamically expensive. The reversibility of encapsulation is discussed in the Conclusions, where different ejection mechanisms are tackled to externalize the biomolecule.

Figure 2. Free energy landscape of DNA@(51,0) SWCNT hybrid. The thermodynamical surface is built using two macroscopic descriptors, $\xi_1$ and $\xi_2$, where $\xi_1$ is the distance between centres of mass of the dsDNA and nanotube, projected along the nanopore main axis ($z$), and $\xi_2$ is the absolute distance between the (GC) termini on opposite sides of the double-strand, equivalent to the dsDNA end-to-end length. The several adjacent free-energy minima along $\xi_1$ demonstrate that the molecule is relatively mobile to translocate along the nanotube, however, the absolute minimum at $\xi_2 = 0.117$ nm indicates that the nanopore center is the energetically favoured region upon confinement. All the $\xi_2$ minima are located along a quasi-linear path defined by $\xi_2 = 4.1$ nm highlighting the enhanced thermodynamic stability associated with the canonical B form. The snapshots were taken at different time intervals corresponding to ($\xi_1, \xi_2$) nm: A) (0.117, 4.112), B) (0.621, 4.164), C) (1.307, 4.164) and D) (1.796, 4.115). H2O molecules and Na$^+$ and Cl$^-$ ions are omitted for clarity sake.

It is remarkable to observe that the exact position of the free-energy minima is almost invariant at $\xi_2 \sim 4.1$ nm, consistent with the end-to-end length of $L = 3.8$ nm of a B-
DNA conformation. We have performed independent umbrella sampling calculations, using an harmonic bias, to determine the potential of mean force (PMF) and the system’s thermodynamical probability distribution, \( P(\Omega) \), employing two order parameters to describe the distance between c.o.m.s of DNA and the solid, \( \Omega_1 \), and the end-to-end length of the double-strand, \( \Omega_2 \). The calculations with \( \Omega_1 \) clearly corroborate the main findings revealed by the metadynamics analysis, namely that the system has probability maxima at \( \Omega_1 = (0.21, 1.62, 1.92) \) nm (Figure 3); the DNA molecule is fairly mobile inside the nanopore, easily moving from one free-energy minimum to an adjacent one. The thermodynamical robustness of the canonical B-form under confinement is illustrated by the Gaussian profile assumed by the probability distribution regarding the DNA end-to-end length, Figure 3. \( P(\Omega_2) \) is well described by

\[
P(\Omega_2) = q \exp\left[ -\frac{1}{2} \left( \Omega_2 - \Omega_2^0 / \sigma \right)^2 \right],
\]

with a Gaussian peak width at half height of \( \sigma = 0.16 \) nm, \( \phi = 4.32 \cdot 10^{-2} \), and a peak centered at \( \Omega_2^0 = 4.01 \) nm \( \pm 0.001 \) nm corresponding to the equilibrium (unbiased) end-to-end distance of encapsulated dsDNA, \( L \). It now becomes clear that a perturbation of the double strand towards non-equilibrium values of \( L \), leading to either a contraction \( \Omega_2 < 4.01 \) or a stretching \( \Omega_2 > 4.01 \) of B-DNA, results in a rapid increase of the associated potential of mean force rendering the perturbation process thermodynamically unstable.

\[
\phi_1 = \phi_{24} = \angle(r_2 \cdot C \cdot r_{24}) \quad \text{and} \quad \phi_2 = \phi_{214} = \angle(r_{12} \cdot C \cdot r_{14}),
\]

which for the crystalline form of pure B-DNA assume the values \( \phi_1 = 55.9^\circ \) and \( \phi_2 = 56.6^\circ \) as measured by Dickerson and co-workers \(^\text{29}^\). The anisotropy of the terminal angles is simply given by \( \Delta \phi = (\phi_1 - \phi_2) \). The local elevation of \( \Delta \phi \) observed in the 10–20 ns time window, an interval during which DNA is being confined, falls back to negligible values (\( \Delta \phi \approx 0.35^\circ \)) indicating that the anisotropic deformation of the double strand is reversible once confinement is complete. Because the molecule undergoes encapsulation from the \( \phi_2 \) termini, the fact that \( \phi_1 > \phi_2 \) while confinement takes place indicates a slight compression at the \( \phi_2 \)-end, which is replicated on the other side of the chain (\( \phi_1 \)) as it penetrates into the SWCNT. This previously unobserved compression phenomenon is an entropic effect arising from the constriction caused by the pore walls, but also an energetic effect given that the interaction energy between DNA and the nanotube only stabilizes once encapsulation is complete \(^\text{30}^\). To accommodate itself in the endohedral volume, the double-strand skeletal diameter slightly decreases and is balanced by an end-to-end length increase from \( L = 3.85 \) nm to \( L = 4\sim 4.1 \) nm (Figure 1A).

Upon encapsulation the dodecamer maintains its kinetic mobility, exploring a region whose boundaries are located at the nanotube termini and correspond to minima in the overall free-energy landscape (Figure 2). Translocation within the solid occurs via a self-translational diffusion process along the central axis and also via a mechanism of self-rotation about the biomolecular axis. The conformational ensemble of individual strand axes recorded in Figure 4 resembles a toroid in the \( yz \)-plane (parallel to the nanotube main axis), whose centre is largely unpopulated. On the other hand, even though the double strand is flexible, it cannot be overstretched along the \( z \)-axis without a drastic increase of the corresponding PMF, and thus the \( xy \)-projection shows symmetrical opposite regions at the boundaries (corresponding to domains close to the wall), where the density of the axes is smaller.
The conformational stability of DNA is markedly influenced by the canonical Watson-Crick H-bonds network, which is also paramount to maintain double-strand integrity and avoiding melting into single strands. In crystals, the distance between a donor (D) and the acceptor (A), $d_{D,A}$, is maximum for the N6–O4 H-bond of (AT) pairs corresponding to $d_{D,A} = 2.95$ Å (Figure 6A), however, in the liquid phase $d_{D,A}$ slightly increases to accommodate transient thermal fluctuations and molecular flexibility. Accordingly with previous studies, we postulate that an H-bond exists when the donor and acceptor are not separated by more than $d_{D,A} = 3.5$ Å and with a characteristic angle $\phi$ of $\angle_{DIA} \leq 35^\circ$. In order to probe the complete Watson-Crick network, we have determined the number of H-bonds of each donor-acceptor type occurring in the DNA dodecamer and represented them in terms of the corresponding probability of occurrence computed by histogram weighting. The results in Figure 6A are represented in terms of the normalized probability of occurrence, $P(\%)$, where the normalization $\int_0^{100} P(\%) d\% = 1$ is performed for the maximum number of allowed H-bonds for a particular donor-acceptor pair. The H-bond between N1 and N3, characteristic of both (AC) and (GC) basepairs, has a maximum occurrence allowance of 12, and therefore the peak at 92% with a normalized probability of $P(92\%) = 0.574$ corresponds to the existence of 11 H-bonds; the other occurrences are located at 100% with $P(100\%) = 0.095$ (12 H-bonds), at 83.5% with $P(83.5\%) = 0.264$ (10 H-bonds), at 75% with $P(75\%) = 0.061$ (9 H-bonds) and at 67% with $P(67\%) = 0.006$ (8 H-bonds). Occurrences below 50% have negligible probabilities for all H-bonds in the double-strand. For occurrences higher than 75%, the accumulated probabilities correspond to $P(\% > 75) \geq 0.87$–0.91; thus, Figure 6A indicates that the canonical Watson-Crick H-bond network is essentially conserved when the molecule becomes encapsulated under physiological conditions. This comes to show that the contact between the DNA dodecamer and the hydrophobic inner surface of the carbon nanotube results in rather minor rearrangements of the nucleotides H-bonding, which is of utmost relevance to maintain double-strand integrity for drug delivery techniques currently exploring SWCNT-based media as encapsulating agents.
in the encapsulated DNA molecule, considering the last 20 ns of observation time: for the N1–N3 pair, %H bonds = 100% corresponds to the existence of 12 H-bonds throughout the dodecamer, %H bonds = 92% to 11 H-bonds, %H bonds = 83.5% to 10 H-bonds and so on. The probability maxima are indicated in the graph by the corresponding % value: (dark blue) H-bonds between N2–O2, (red) H-bonds between N4–O6, (green) H-bonds between N6–O4 and (black) H-bonds between N1–N3. B) Intrastrand stacking geometry. (Dark red) strand 1, (grey) strand 2. The stacking distance (helical rise) is measured between the geometrical centres of the molecular planes belonging to two consecutive nucleobases of the same strand (e.g., stacked NB pair 1 corresponds to C1–G2 and also G24–C23), and the stacking angle is calculated between the normal vectors of those two planes (two perfectly stacked bases, with their rings parallel to each other, have a stacking angle of 0° corresponding to cos (0°) = 1).

The N2–O2 bond in (GC) pairs shows an accumulated probability of $P(\%>75) \sim 0.87$, suggesting that H-bond breaking/formation occurs essentially in the double-strand *termini*, richer in (GC) moieties, leaving intact the (AT) rich inner tract. This is clearly consistent with the intrastrand data recorded in Figure 6B, where consecutive nucleotides have been probed for their stacking distance (helical rise) and normal vector angle between molecular planes. From the corresponding distribution histograms, the maximum probability of occurrence has been obtained for both geometric parameters and plotted as function of the stacked pair index. The crystallographic studies of Vargason *et al.* indicate that the transition from pure B-DNA to an A-DNA conformation is accomplished via a monotonical decrease of the intrastrand stacking distance from 3.4 Å to ca. 2.6 Å for the pure A-form. Apart from the severe deformation induced by basepairs 1 and 11 upon the double-strand, our results indicate an helical rise between nucleotides of ~3.8 Å, resulting in $L = 4.18$ nm. We have observed (Figure 1A) that encapsulation of DNA leads to a slight increase of $L$ from the pure B-form, and are now able to attribute this increase to a 1D anisotropically stretching of the molecule located essentially at the end of the (GC) tracts. From an energetical point of view this is somehow unexpected, bearing in mind that (AT) pairs form two H-bonds instead of the characteristic three H-bonds of (GC) duos; the interaction energies between purines and pyrimidines favour the stability of (GC) pairs against their (AT) counterparts, $E_{GC} = -291.9 \text{ kJ/mol}$ and $E_{AT} = -110.2 \text{ kJ/mol}$.

The explanation for this apparent inconsistency lies in entropic causes. Because the *termini* are more flexible than the inner core, they are more prone to adaption to the local environment, thus being more able to accommodate elastic deformation.

**IV. CONCLUSIONS AND PROSPECTS**

DNA can be encapsulated onto the purely hydrophobic (51,0) topology but not so at the (40,0) analogue, indicating that a 1 nm decrease of diameter might prevent confinement. This observation is encouraging for technologies using pristine SWCNTs as drug delivery agents, however, needs to be wisely put in perspective. It is known that nanotubes can be electrically charged, either using an AFM tip and applying a voltage bias or by chemically doping the solids with p-type dopants to obtain positively charged nanotubes. The effect of charge density upon the energetics and dynamics of confinement needs to be carefully addressed in the future; because DNA’s outer surface is negatively charged (phosphates), its interaction with a positively charged (40,0) solid might indeed lead to the occurrence of encapsulation with enhanced thermodynamical stability. The latter is of paramount importance, for any technological application to find its way into the industrial production line the confinement of DNA molecules needs to be a thermodynamically reversible process, and subsequent ejection possible towards the nanotube exterior. Xue et al. used filler agents (C$_{60}$) and mechanical actuators (Ag) to eject ssDNA from hydrophobic SWCNTs, and related the feasibility of the ejection process with the enhanced dispersive interactions resulting from DNA externalization; their C$_{60}$ agents evidenced interaction energies with the nanotube of $\sim 1800$ kJ/mol ($\sim 7000$ kJ/mol for the Ag nanowires), an order of magnitude higher than the ones we observe for the (51,0) encapsulated DNA, leading us to believe that similar externalization mechanisms can be employed to revert DNA encapsulation.

**ACKNOWLEDGEMENTS**

The authors would like to acknowledge Requimte/CQFB (Universidade Nova de Lisboa, Portugal) and NSEC (University of Wisconsin – Madison, USA) for generous CPU time. This work was partially supported by grant PTDC/CTM/104782/2008 (Portugal) and makes use of results produced with the support of the Portuguese National Grid Initiative (https://wiki.ncg.ingrid.pt). F.J.A.L. Cruz gratefully acknowledges financial support from FCT/MCTES (Portugal) through grant SFRH/BPD/45064/2008.

**REFERENCES**

1. Kumar, H.; Lansac, Y.; Glaser, M. A.; Maiti, P. K., Biopolymers in Nanopores: Challenges and Opportunities. *Soft Matter* 2011, 7, 5898–5907.
2. Venkatesan, B. M.; Bashir, R., Nanopore Sensors for Nucleic Acid Analysis. *Nature Nano.* 2011, 6, 615-624.
3. Franklin, A. D.; Luisier, M.; Han, S-J.; Tulevski, G.; Breslin, C. M.; Gignac, L.; Lundstrom, M. S.; Haensch, W., Sub-10 nm Carbon Nanotube Transistor. *Nano Lett.* 2012, 12, 758–762.
4. Gao, H.; Kong, Y.; Cui, D., Spontaneous Insertion of DNA Oligonucleotides into Carbon Nanotubes. *Nano Letters* 2003, 3, 471-473.
5. Lau, E. Y.; Lightstone, F. C.; Colvin, M. E., Dynamics of DNA Encapsulated in a Hydrophobic Nanotube. *Chem. Phys. Letters* 2005, 412, 82–87.
