Supporting Information for

Molecular-Basis of Single-Walled Carbon Nanotube Recognition by Single-Stranded DNA

Daniel Roxbury, Jeetain Mittal, and Anand Jagota

S1. Computing DNA strand end-to-end distance distributions

As another method to compare equilibrium structures amongst simulated configurations, the three-dimensional DNA end-to-end distances, i.e. H3T (3' terminal hydrogen) to H5T (5' terminal hydrogen), are calculated throughout the trajectory for each individual DNA strand. Histograms of the distribution of end-to-end distances are constructed for each configuration, e.g. 100-400 ns for $(TAT)_4$-(6,5) configuration. The solid black bar graphs in Figure S1 shows the histogram for the entire trajectory. Additionally, we computed the end-to-end distance histograms of only the conformations that belong to the clusters shown in Figure 2a of the main text. These have been overlaid on the histogram of the entire trajectory, plotted as red lines. Observe that the peaks in the probability distribution functions (PDFs) computed on the cluster do not entirely match those computed using the full trajectory. The end-to-end distance is an alternative to the RMSD of backbone atom positions as a way to cluster conformations. While the latter better identifies unique motifs, the former might be easier to relate to experimental measurements.

Compare Figure S1 and Figure 2a of the main text. Observe that the end-to-end distance distribution of $(TAT)_4$-(6,5) has one dominant peak at 2.3 nm, especially when the distribution is computed on conformations that belong to the cluster. This peak corresponds therefore to the helical structure in which short DNA strand ends line up effectively ligating into one long strand. The end-to-end distance distributions for $(T)_{12}$-(6,5) are clearly bimodal. The green and blue strands (Figure 2(a) of the main text) adopt a β-barrel and thus have much more outstretched configurations, corresponding to a peak in their end-to-end distribution peaks at 4.2 nm. The third strand (red) is not able to participate in the β-barrel, and thus forms a more tightly bound wrap with end-to-end distance of 2.1 nm. For the $(TAT)_4$-(8,7) combination, because individual DNA strands are not able to fully wrap the SWCNT, the end-to-end distributions are populated by looped and disordered DNA structures.

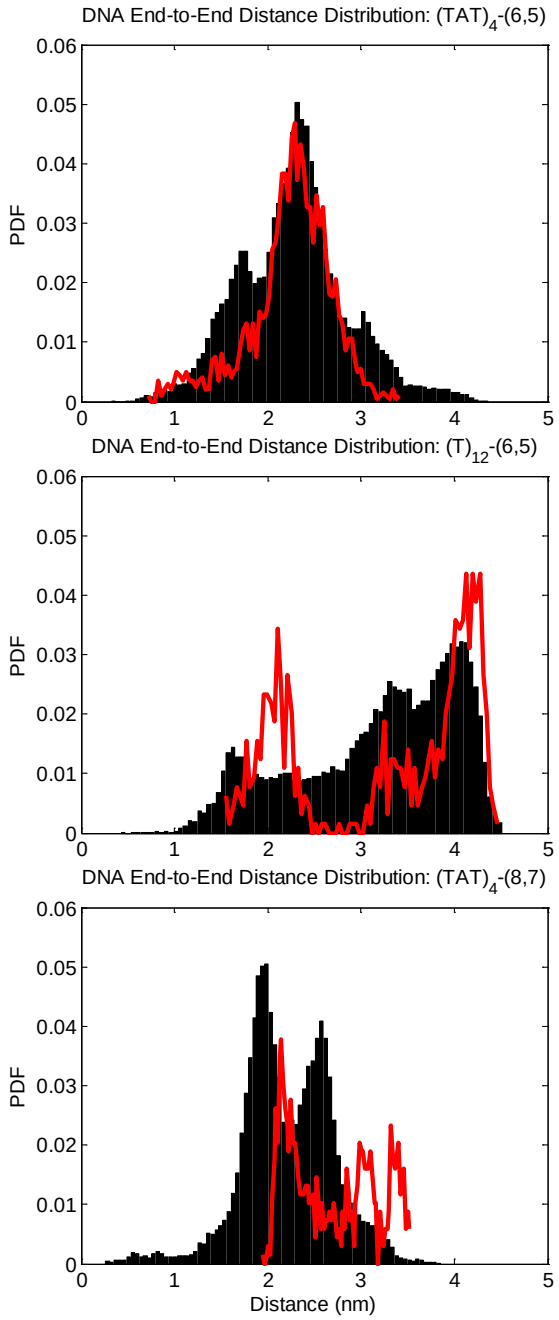
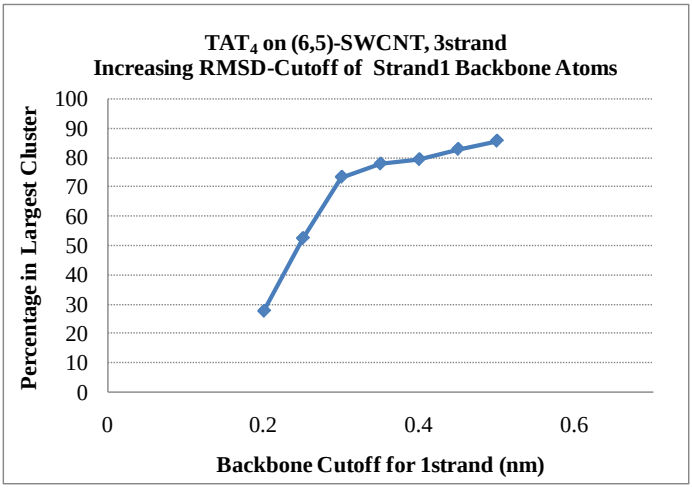


Figure S1. End-to-end distance distribution for DNA strands on the various DNA-SWCNT simulated combinations. Solid black bar histograms are computed using the full equilibrium trajectory while the overlaid red lines are computed using only those conformations that belong to clusters shown in Figure 2a of the main text.

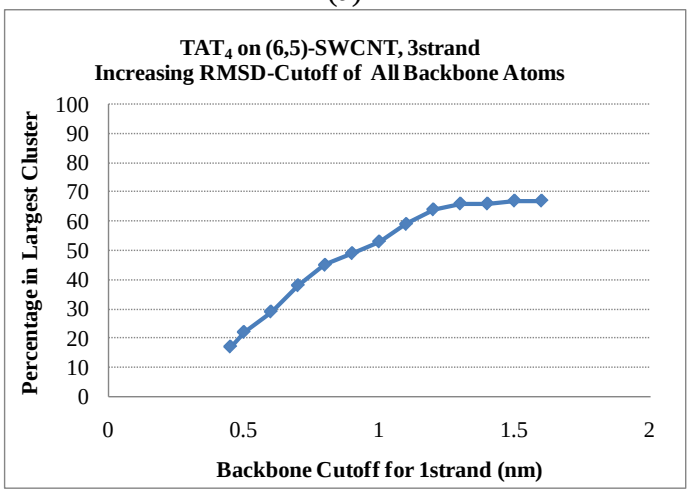
S2. Choosing root mean squared deviation (RMSD) cutoff value for clustering analysis

When examining a trajectory, it is important to group structures into clusters based on universal predetermined criteria. As mentioned in the main text, DNA backbone atoms were subjected to RMSD cutoffs of 0.45 nm and 1.3 nm for individual-strand and multi-strand

clustering, respectively. Choosing an appropriate RMSD cutoff for clustering is also of significant importance. If the RMSD value is too low, percentages in the largest cluster drop off significantly, as well as the contrast between hybrids of different DNA composition. For individual-strand clustering of $(TAT)_4$ on (6,5)-SWCNT, systematically increasing the RMSD cutoff from 0.2 to 0.5 nm, we find that by ~0.4 nm, the percentage of structures in the largest cluster essentially remains at ~80%, (Figure S2). To compare, the percentage of trajectory in the largest cluster was 83%, 39%, and 59% for $(TAT)_4$-(6,5), $(TAT)_4$-(8,7), and $(T)_{12}$-(6,5) configurations, respectively. A greater population in the largest cluster should signify a more stable structure, as DNA is in a more-confined configuration. Likewise, in multi-strand clustering of $(TAT)_4$ on (6,5), for an RMSD cutoff of ~1.3 nm, the asymptotic limit of ~67% of the trajectory is reached. As such, this was the chosen cutoff. Comparing multi-strand configurations, $(TAT)_4$ on (6,5) still has the greatest population in the largest cluster (67%, 12%, and 17% of the trajectory for $(TAT)_4$-(6,5), $(TAT)_4$-(8,7), and $(T)_{12}$-(6,5) hybrids, respectively).



(a)



(b)

Figure S2. **a,** Plot showing process for choosing individual-strand clustering criteria via backbone RMSD. The backbone RMSD cutoff values for strand1 were increased from 0.2 – 0.5 nm to find the optimum cutoff. **b,** In a similar fashion, RMSD cutoff was increased for multi-strand clustering.