

Supplementary Information

Sébastien Côté,^{*a,c} Delphine Bouilly^{*a,b} and Normand Mousseau^{*a}

^a Département de physique, Faculté des Arts et des Sciences, Université de Montréal, Montréal, Canada.

^b Institut de recherche en immunologie et oncologie (IRIC), Université de Montréal, Montréal, Canada.

^c Département de physique, Cégep de Saint-Jérôme, Saint-Jérôme, Canada.

*sebastien.cote.4@umontreal.ca; delphine.bouilly@umontreal.ca; normand.mousseau@umontreal.ca

Parameterization of the linkers to the nanotube

In this section, the parameterization protocol of the anchors between the biomolecules and the carbon nanotube is presented. This protocol follows the standard procedure for the AMBER force fields to determine the partial charges using the RESP protocol¹ as well as the bonded and Lennard-Jones parameters using the Generalized Amber Force Field (GAFF).² We closely follow the procedure of AMBER in order to generate parameters that are compatible with AMBER14sb (proteins) and AMBER-OL15 (nucleic acids) force fields. The RESP protocol has been applied using AmberTools20³ on the electrostatic potential (ESP) data obtained from *ab initio* HF/6-31G*//HF/6-31G* calculations with Gaussian16.⁴ The topology files compatible with Gromacs were produced using ACPYPE.⁵

Lysozyme. The lysozyme is attached to the carbon nanotube using a pyrene-maleimide anchor that forms π - π interactions on the surface of the nanotube (Figure S3A).^{6,7} The pyrene-maleimide molecule is covalently attached to S90C of the lysozyme. To prepare this configuration, the 3D chemical structure of N-(1-pyrenyl)-maleimide was downloaded from PubChem,⁸ then one of the two C-H of the maleimide ring was covalently linked to the sulfur atom (HG removed) of the cysteine, moving the H to the other C-H of the maleimide ring.

For the parametrization of the partial charges, the cysteine is capped with ACE and NME groups as was done for AMBER when parameterizing amino acids.¹ The cysteine is made into two configurations with different backbone dihedral angles: c_5 ($\omega_1 = 180^\circ$, $\phi = 206^\circ$, $\psi = -141^\circ$ and $\omega_2 = 180^\circ$) and α_r ($\omega_1 = 180^\circ$, $\phi = -60^\circ$, $\psi = -40^\circ$ and $\omega_2 = 180^\circ$) to respectively represent an extended and a helix configurations, exactly as was done for AMBER. Once these two configurations are prepared, ACPYPE is used to generate the bonded and Lennard-Jones parameters, setting all charges to zero. Then, the configuration is energy minimized, while restraining the ω , ϕ and ψ dihedral angles. An ensemble of rotamers is then produced by varying the dihedral angles implicated between the cysteine and the maleimide ring. The configurations having pairs of non covalently bonded atoms within 1.5 nm are removed from the parametrization set.

Ab initio calculations are performed using HF/6-31G*//HF/6-31G* to optimize the structures and to determine the electrostatic potential they generate, as was done for AMBER.¹ During the optimization, constraints are applied on the ω , ϕ and ψ dihedral angles to the values previously mentioned. The ESP values are output at 4 layers of points – 1.4, 1.6, 1.8 and 2.0 times the van der Waals radii – with a resolution of 1 point per Å^2 for each atom in the configuration.

The lowest energy structure – without polar contact – for c_5 and that for α_r are used in a multiconfiguration RESP optimization with default parameters to determine the partial charges that best reproduce the electrostatic potential (ESP) obtained from the *ab initio* calculations. During RESP, the partial charges of the cysteine (except C_α and H_α because they are involved in dihedral angles with the side chain atoms), ACE and NME are constrained to their values in AMBER.^{1,9} The total charge of the molecule is constrained to zero. Shown in Figure S4A, the optimized partial charges yield an ESP with a relative RMS of 0.112 compared to the *ab initio* ESP, which is deemed acceptable.¹

DNA. The DNA strand is attached covalently to the carbon nanotube using an amide linker (Figure S3B).^{10,11} The C=O group of the amide forms a covalent bond with the carbon nanotube, while its N-H group forms a covalent bond with the C5' atom of the DNA 5'-end.

For the parametrization of the partial charges, the amide is capped with two methyl groups and ACPYPE is used to generate the bonded and Lennard-Jones parameters, setting all charges to zero. Then, the configuration is energy minimized. *Ab initio* calculations are performed using HF/6-31G*//HF/6-31G* to optimize the structures and to determine the electrostatic potential they generate, as was done for AMBER.¹ The ESP values are output at 4 layers of points – 1.4, 1.6, 1.8 and 2.0 times the van der Waals radii – with a resolution of 1 point per Å^2 for each atom in the configuration.

The lowest energy structure is used in a RESP optimization with default parameters to determine the partial charges that best reproduce the electrostatic potential (ESP) obtained from the *ab initio* calculations. During RESP, the partial charges of the methyl attached to N-H are constrained to a total charge of $+0.1896e$, which corresponds to the total charge of the 5'-end hydrogen (H5T) and oxygen (O5') in AMBER. This has been similarly done when parameterizing the bases along with the phosphate group for AMBER: the

phosphate was capped with two methyl and their total charge were respectively constrained to that of the 5'-end H5T-O5' atoms and 3'-end H3T-O3' atoms.¹ This allows the total charge of the system to stay the same when the 5'-end (no H5T and O5') is covalently linked to the amide group (no methyl). The total charge of the molecule is constrained to zero. Shown in Figure S4B, the optimized partial charges yield an ESP with a relative RMS of 0.104 compared to the *ab initio* ESP, which is deemed acceptable.¹

In all these *ab initio* calculations, we neglected the presence of the nanotube as its size (infinite length and radius of 0.695 nm) makes it computationally too demanding to study at the HF level required by AMBER. We do not expect that to significantly affect the biomolecules–nanotube interactions observed in the MD simulations because the charge transfer is expected to be small and localized near the anchor/graft point.^{12?}

Convergence intervals

In this section, the intervals of convergence chosen for the analysis of the simulations are presented. To do so, structural parameters are monitored as a function of time and convergence is assessed when these indicate that the system fluctuates within the same structural ensemble.

For the simulations on the lysozyme, the convergence interval of each simulation has been determined by looking at the root mean square deviation (RMSD) on the backbone atoms (N, C α , C and O), the secondary structure and the probability of contact similarity against the experimental structure as a function of time as shown in Figure S5 for the simulations in solution (denoted S# lysozyme/lysozyme-ligand), in Figure S6 for the simulations of the lysozyme without the ligand with the nanotube (denoted C# lysozyme) and in Figure S7 for the simulations of the lysozyme with the ligand with the nanotube (denoted C# lysozyme-ligand). We monitor these structural parameters because they are used to determine the stability of the lysozyme in the presence of the carbon nanotube (Table 1). We decided to discard the first 250 ns for all simulations. For others, the convergence interval is chosen to start later in light of the data as a function of time : C2 lysozyme starts from 400 ns and C5 lysozyme starts from 600 ns (see Figure S6) and C2 lysozyme-ligand starts from 400 ns and C4 lysozyme-ligand starts from 550 ns (see Figure S7). We then confirmed that the overall orientation of the lysozyme and the overall number of lysozyme–nanotube contacts are also converged on those intervals.

For the simulations on the DNA, the convergence interval of each simulation has been determined by looking at the number of residue-residue contacts, the number of base-base H-bonds and the radius of gyration as a function of time as shown in Figure S8 for the simulations in solution (denoted S# ssDNA/dsDNA), in Figure S9 for the simulations of the single-stranded DNA with the nanotube (denoted C# ssDNA) and in Figure S10 for the simulations of the double-stranded DNA with the nanotube (denoted C# dsDNA). We monitor these structural parameters because they are used to determine the stability of the DNA systems in the presence of the carbon nanotube (Table 2). We decided to discard the first 250 ns for all simulations. For others, the convergence interval is chosen to start later in light of the data as a function of time : S1 and S2 ssDNA starts from 500 ns (see Figure S8) and C2 ssDNA starts from 450 ns and C4 ssDNA starts from 500 ns (see Figure S9). Overall, the ssDNA experiences more fluctuations than the dsDNA in solution and in the presence of the carbon nanotube because ssDNA is a lot less structured. We then confirmed that the overall number of contacts with the nanotube is also converged on those intervals.

Electrostatic potential on the nanotube

In this section, we compare the electrostatic potential (ESP) maps obtained for each simulation and we justify why simulations C4 of lysozyme and C1 of ssDNA were treated separately.

Lysozyme. For the lysozyme without the ligand, the region with the highest ESP values is localized to the left of the anchor point at an angle of $\sim 50^\circ$ around the nanotube (Figure S13). As expected the simulation C4 is different because the lysozyme interacts differently with the nanotube (Figure S11). More precisely, the ESP covers a larger surface area because more charged residues are in contact with the nanotube. Moreover, negative ESP values are observed due to the presence of negatively charged residues, such as Asp-127, nearer of the nanotube. For the lysozyme with the ligand, the ESP covers approximately the same region in all simulations as shown by the small standard deviation (Figure S14). This region is localized within 2 nm to the left of the anchor point at a positive angle of $\sim 50^\circ$ around the nanotube.

In terms of overall ESP, the distribution in simulation C4 is shifted towards more positive values (Figure S17) because many more charged residues are near the nanotube (Figure S11). As a result, the difference from the lysozyme-ligand system becomes greater. While the ESP values are overall more positive, the same trend is observed as a function of the concentration.

DNA. For the single-stranded DNA, the ESP covers a larger surface area due to its proximity to the nanotube (Figure S15). The region covered in each simulation shows some difference due to the high flexibility of ssDNA on the surface of the nanotube: it covers mainly a region of ± 2 nm on each side of the graft point and $\sim 150^\circ$ around the nanotube. The ESP map of simulation C1 is more localized because residues 5 to 8 are stacked and thus not directly on the surface of the nanotube (Figure S12). For the double-stranded DNA, the ESP covers a smaller region of less than 2 nm by 50° in all simulations. While the ESP is on the negative angle side for C1, C2 and C3, it is on the positive angle side for C4 (mostly) and C5.

In terms of overall ESP, the distribution in simulation C1 is shifted towards less negative values (Figure S18) because residues 5 to 8 are farther from the nanotube (Figure S12). As a result, the difference from the dsDNA system becomes smaller, but stays significantly different. While the ESP values are overall less negative, the same trend is observed as a function of the concentration.

Notes and references

- 1 P. Cieplak, W. D. Cornell, C. I. Bayly and P. A. Kollman, *J Comput Chem*, 1995, **16**, 1357–1377.
- 2 J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J Comput Chem*, 2004, **25**, 1157–1174.
- 3 D. Case, H. Aktulga, K. Belfon, I. Ben-Shalom, S. Brozell, D. Cerutti, I. T.E. Cheatham, G. Cisneros, V. Cruzeiro, T. Darden, R. Duke, G. Giambasu, M. Gilson, H. Gohlke, A. Goetz, R. Harris, S. Izadi, S. Izmailov, C. Jin, K. Kasavajhala, M. Kaymak, E. King, A. Kovalenko, T. Kurtzman, T. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, M. Machado, M. M. V. Man, K. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, K. O’Hearn, A. Onufriev, F. Pan, S. Pantano, R. Qi, A. Rahnamoun, D. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. Simmerling, N. Skrynnikov, J. Smith, J. Swails, R. Walker, J. Wang, H. Wei, R. Wolf, X. Wu, Y. Xue, D. York, S. Zhao and P. Kollman, *Amber 2021*, 2021, University of California, San Francisco.
- 4 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *Gaussian~16 Revision C.01*, 2016, Gaussian Inc. Wallingford CT.
- 5 A. W. S. da Silva and W. F. Vranken, *BMC Research Notes*, 2012, **5**, 367.
- 6 Y. Choi, I. S. Moody, P. C. Sims, S. R. Hunt, B. L. Corso, I. Perez, G. A. Weiss and P. G. Collins, *Science*, 2012, **335**, 319–324.
- 7 Y. Choi, T. J. Olsen, P. C. Sims, I. S. Moody, B. L. Corso, M. N. Dang, G. A. Weiss and P. G. Collins, *Nano Lett*, 2013, **13**, 625–631.
- 8 S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, Q. Li, B. A. Shoemaker, P. A. Thiessen, B. Yu, L. Zaslavsky, J. Zhang and E. E. Bolton, *Nuc. Acids Res.*, 2021, **49**, D1388–D1395.
- 9 J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, *J Chem Theory Comput*, 2015, **11**, 3696–3713.
- 10 S. Sorgenfrei, C. Y. Chiu, J. Gonzalez, R. L., Y. J. Yu, P. Kim, C. Nuckolls and K. L. Shepard, *Nat Nanotechnol*, 2011, **6**, 126–132.
- 11 S. Sorgenfrei, C. Y. Chiu, M. Johnston, C. Nuckolls and K. L. Shepard, *Nano Lett*, 2011, **11**, 3739–3743.
- 12 D. Bouilly, J. L. Janssen, J. Cabana, M. Côté and R. Martel, *ACS Nano*, 2015, **9**, 2626–2634.
- 13 Schrödinger LLC, *The PyMOL molecular graphics system*, <https://pymol.org>.

Supplementary Figures

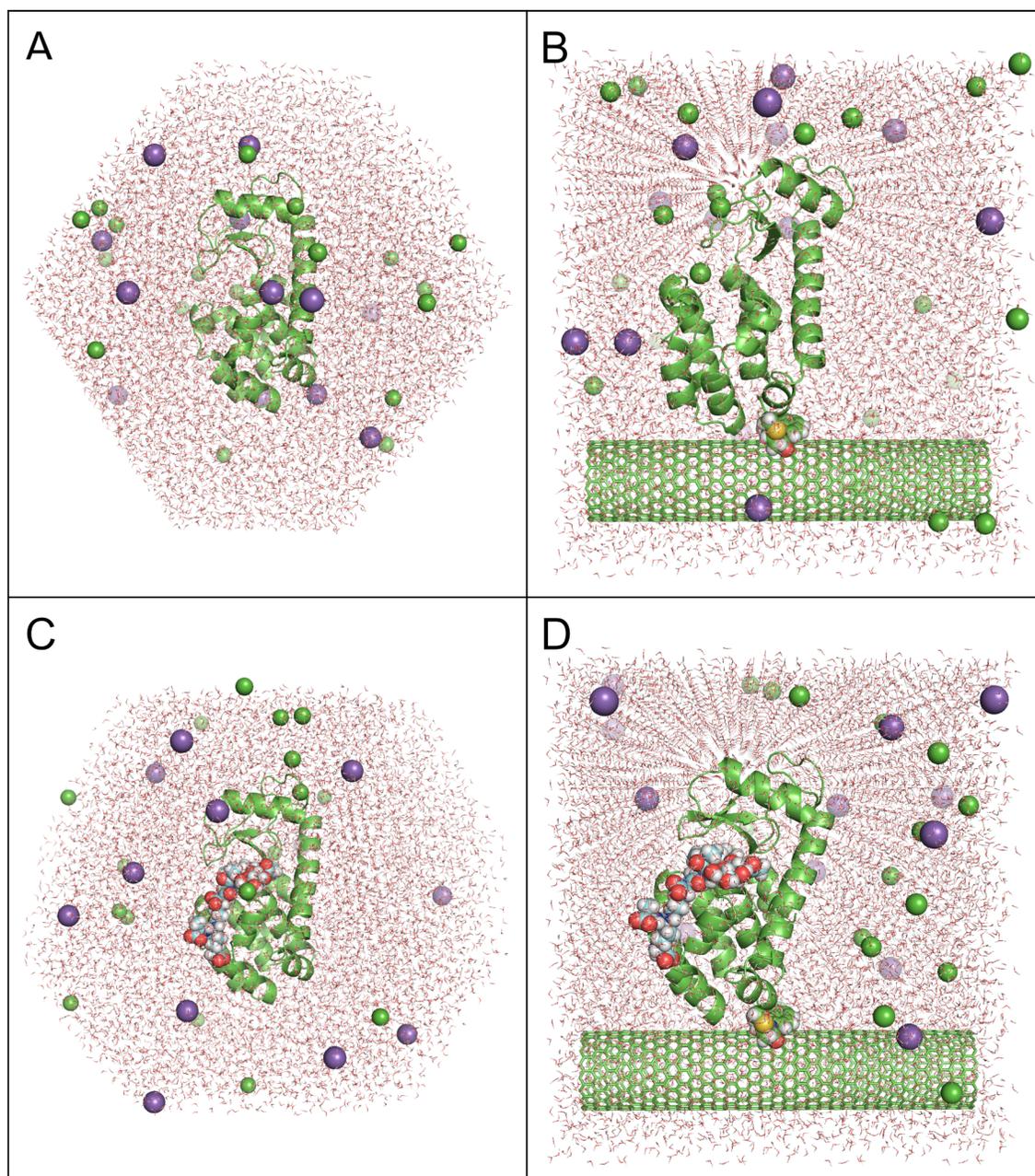


Fig. S1 Initial systems of the lysozyme. (A-B) Lysozyme without the ligand (PDB: 1QTV). (C-D) Lysozyme with the peptidoglycan ligand shown with spheres and carbon atoms in teal (PBD: 148L). (B-D) The pyrene-maleimide anchor to the nanotube is covalently linked to S90C of the lysozyme and is shown with spheres. The sodium and chloride ions are respectively shown as purple and green spheres. The images were produced with PyMOL.¹³

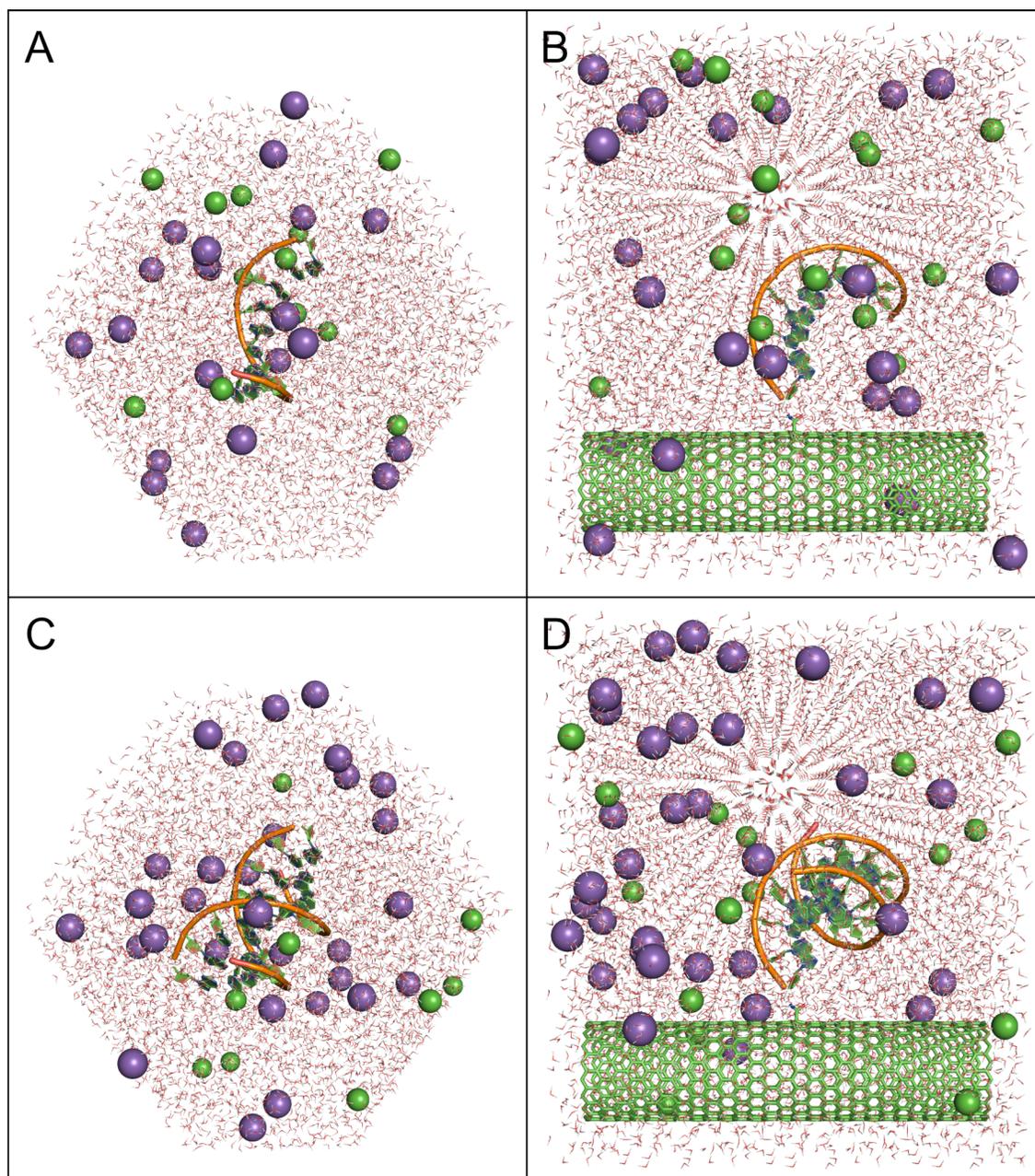


Fig. S2 Initial systems of the DNA sequence. (A-B) The single-stranded 5'-GTGAGTTGTT-3' DNA sequence. (C-D) The double-stranded B-DNA sequence. (B-D) The sequence is covalently linked to the carbon nanotube with an amide group. The sodium and chloride ions are respectively shown as purple and green spheres. The images were produced with PyMOL.¹³

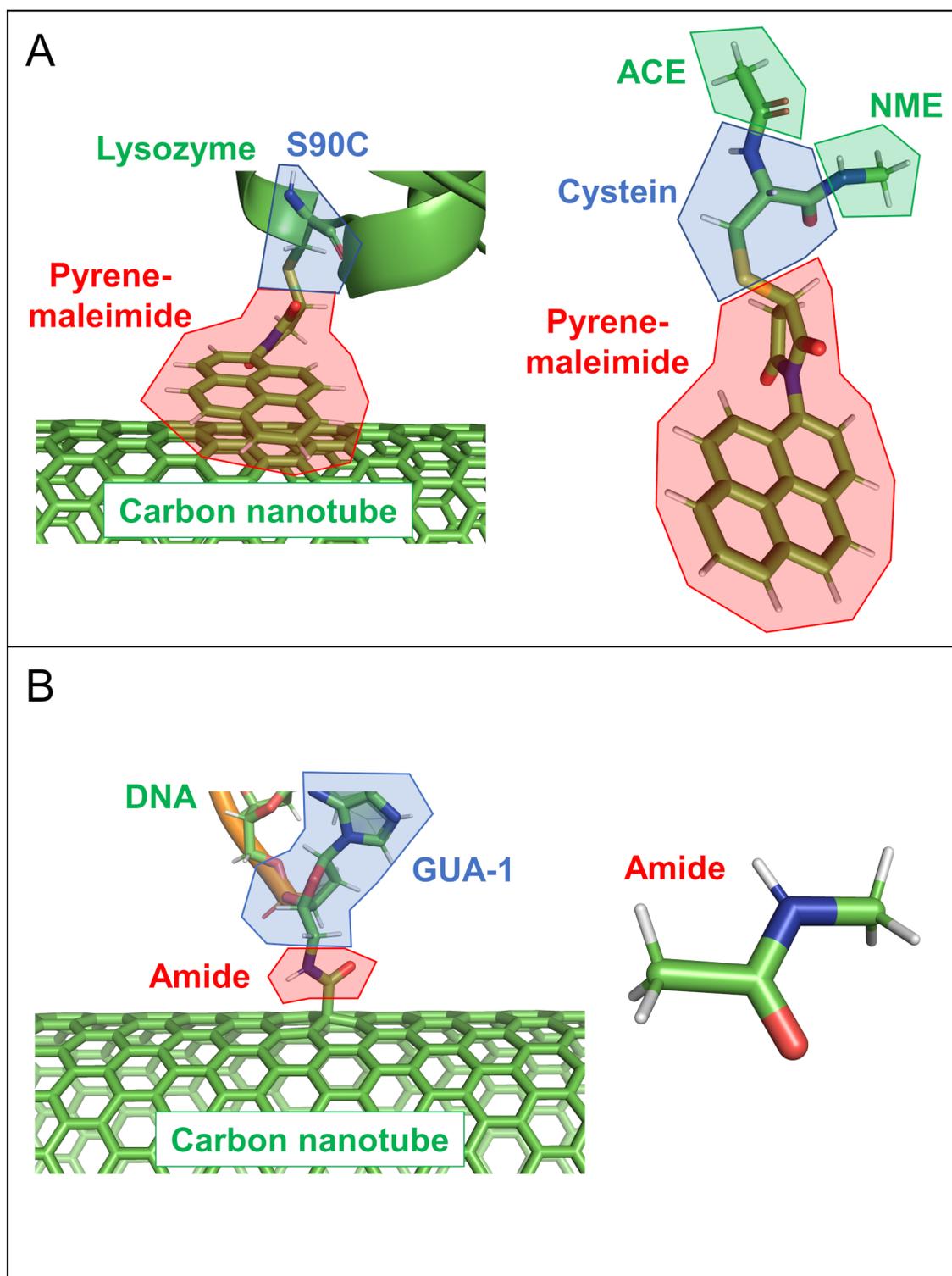


Fig. S3 Anchors to the carbon nanotube. Structure of the anchors to the carbon nanotube : (A) the pyrene-maleimide anchor of the lysozyme and (B) the amide covalent link with the DNA strand.

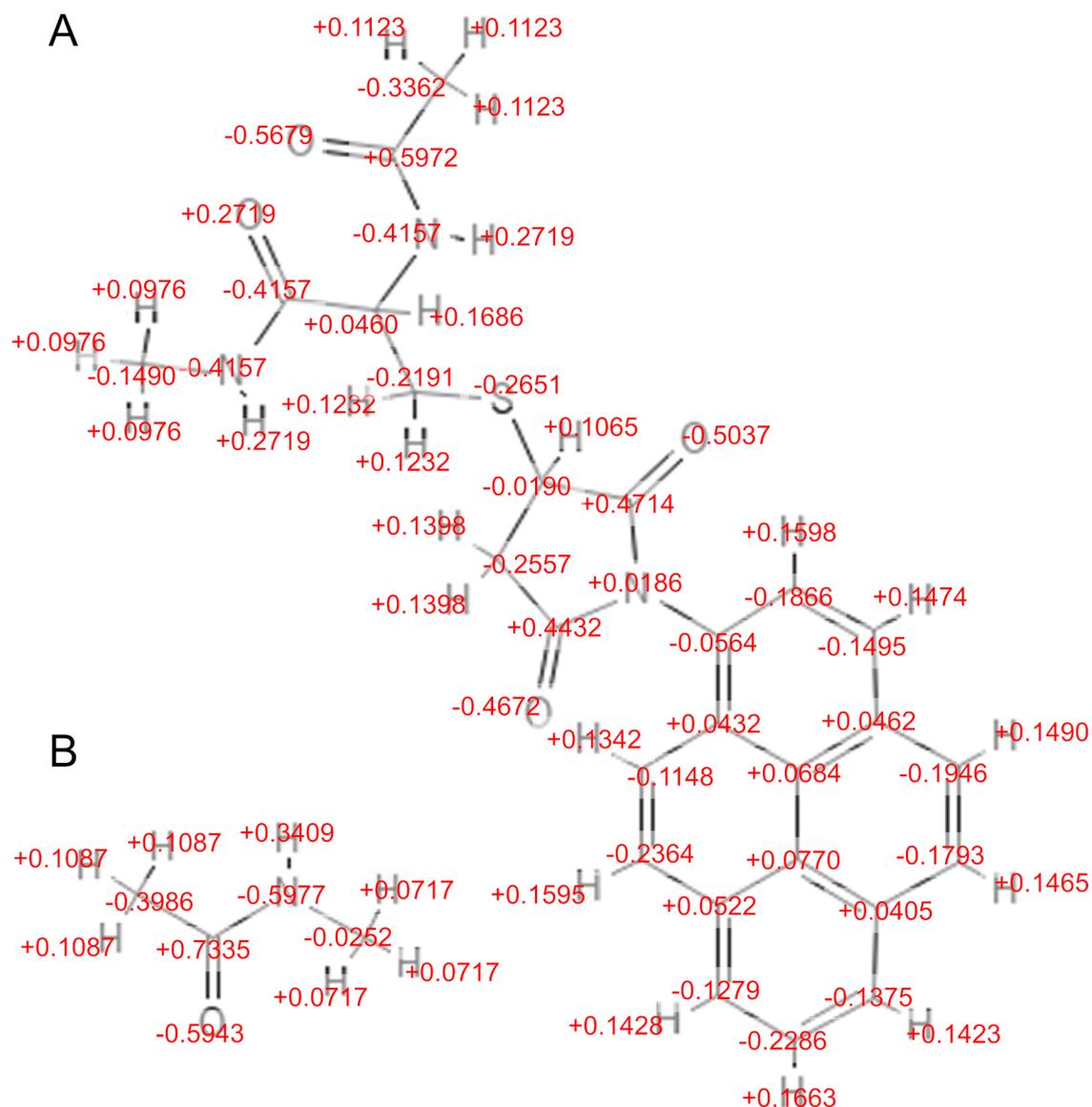


Fig. S4 Partial charges of the anchors. Partial charges in e determined from the RESP protocol on electrostatic potential maps determined using *ab initio* calculations. (A) The molecule used for the *ab initio* calculations on the lysozyme anchor : the pyrene-maleimide compound is linked to the cystein that is capped by ACE and NME. (B) The molecule used for the *ab initio* calculations on the DNA anchor : the amide compound is capped by two methyl groups.

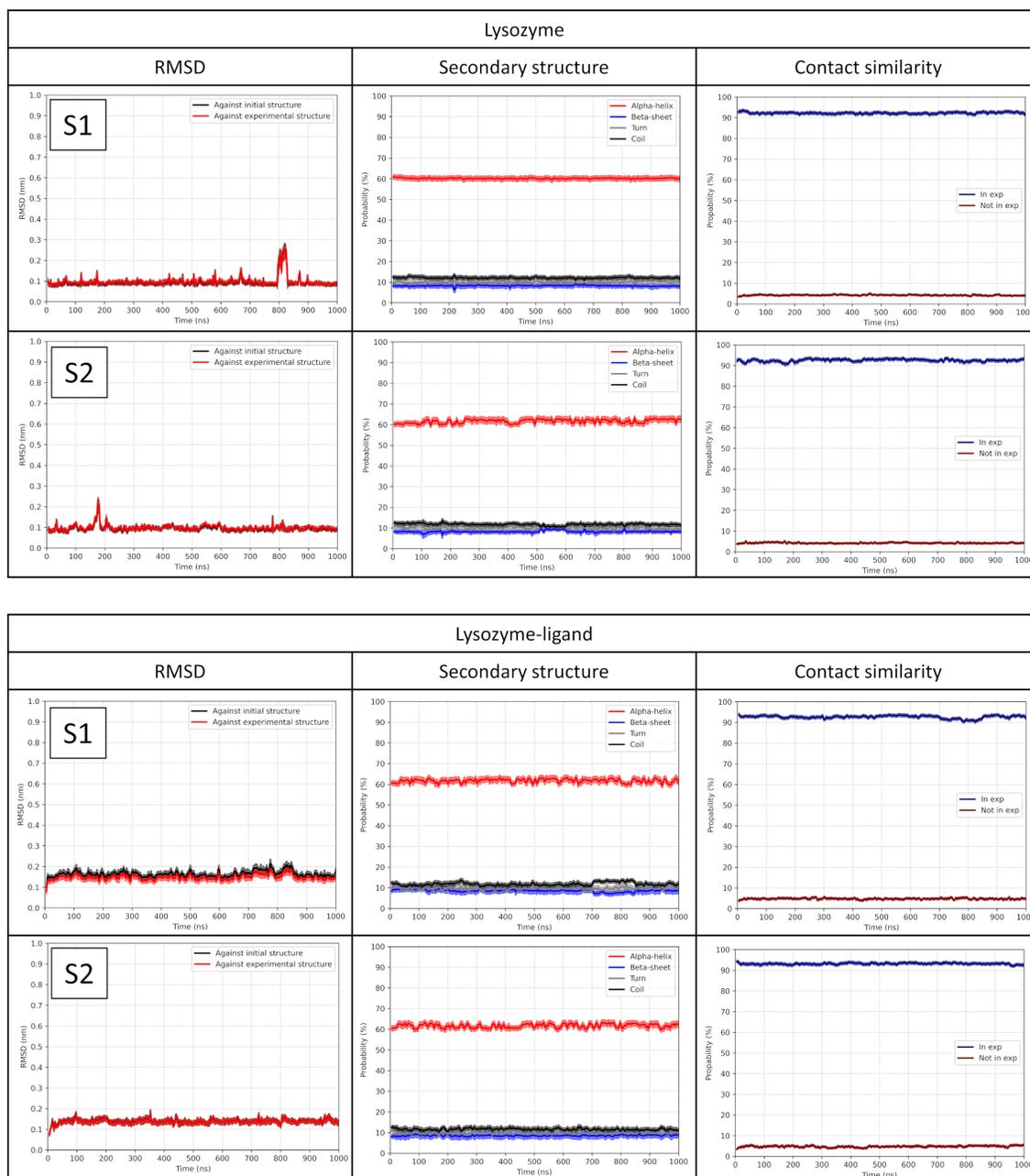


Fig. S5 Convergence intervals of the Lysozyme simulations in solution. The root mean square deviation (RMSD) on the backbone atoms (N, C α , C and O), the secondary structure probability and the probability of contact similarity against the experimental residues-residues contacts are shown for the lysozyme (TOP) without the ligand and (BOTTOM) with the ligand. For the RMSD, the comparison is made against the experimental structure (red) and against the initial structure (black). For the secondary structure, the probability of the main structural features are shown: α -helix (red), β -sheet (blue), turn (gray) and coil (black). For the contact similarity, the percentage of residue-residue experimental contacts preserved in the simulation is shown in blue, while the percentage of contacts not in the experimental structure is shown in red. The curve is the running time average using a 5-ns window and its envelop represents the 1- σ interval.



Fig. S6 Convergence intervals of the Lysozyme simulations with the nanotube. The root mean square deviation (RMSD) on the backbone atoms (N, α , C and O), the secondary structure probability and the probability of contact similarity against the experimental residues-residues contacts are shown for the lysozyme without the ligand. For the RMSD, the comparison is made against the experimental structure (red) and against the initial structure (black). For the secondary structure, the probability of the main structural features are shown: α -helix (red), β -sheet (blue), turn (gray) and coil (black). For the contact similarity, the percentage of residue-residue experimental contacts preserved in the simulation is shown in blue, while the percentage of contacts not in the experimental structure is shown in red. The curve is the running time average using a 5-ns window and its envelop represents the 1- σ interval.

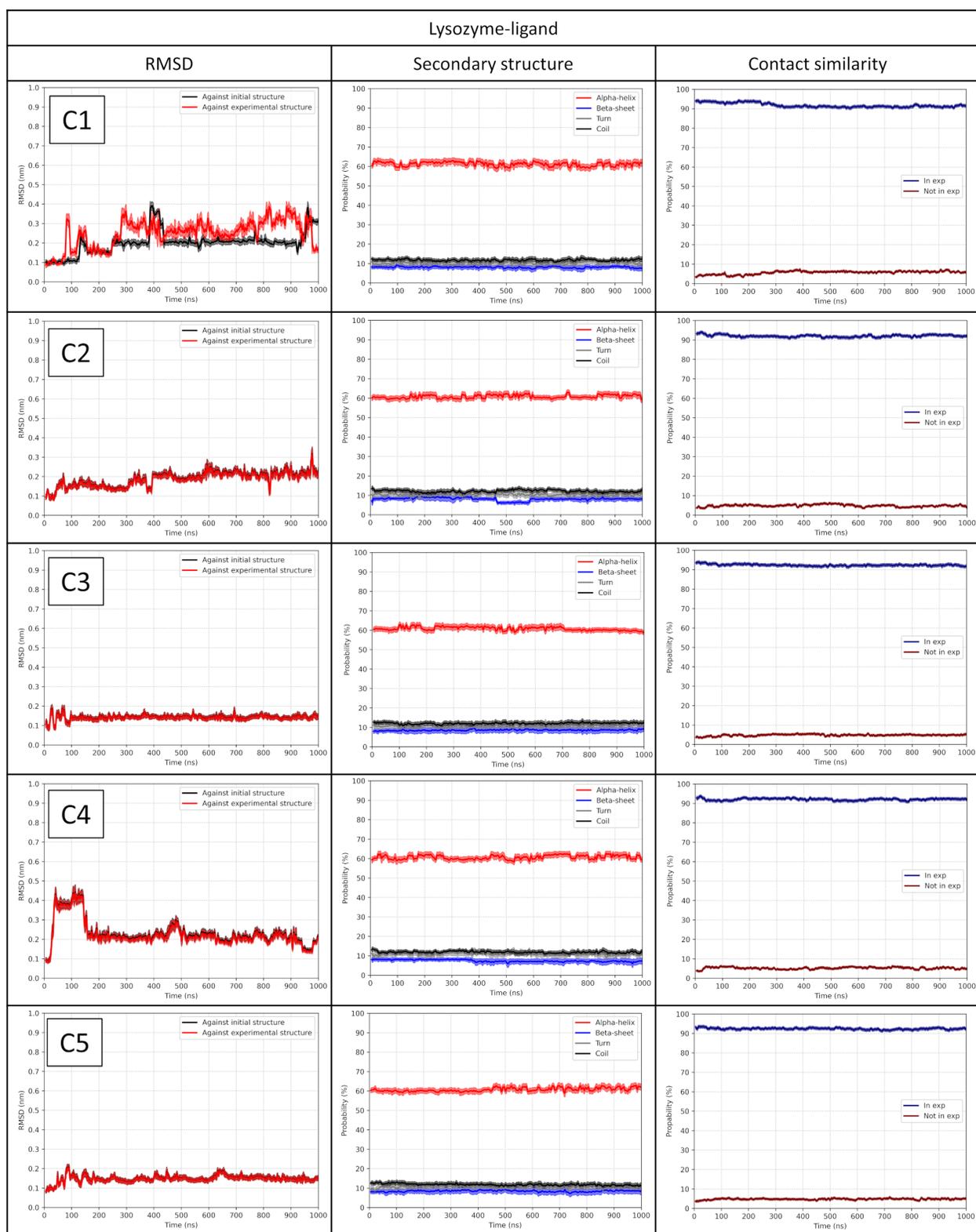


Fig. S7 Convergence intervals of the Lysozyme-ligand simulations with the nanotube. The root mean square deviation (RMSD) on the backbone atoms (N, C α , C and O), the secondary structure probability and the probability of contact similarity against the experimental residues-residues contacts are shown for the lysozyme with the ligand. For the RMSD, the comparison is made against the experimental structure (red) and against the initial structure (black). For the secondary structure, the probability of the main structural features are shown: α -helix (red), β -sheet (blue), turn (gray) and coil (black). For the contact similarity, the percentage of residue-residue experimental contacts preserved in the simulation is shown in blue, while the percentage of contacts not in the experimental structure is shown in red. The curve is the running time average using a 5-ns window and its envelop represents the 1- σ interval.

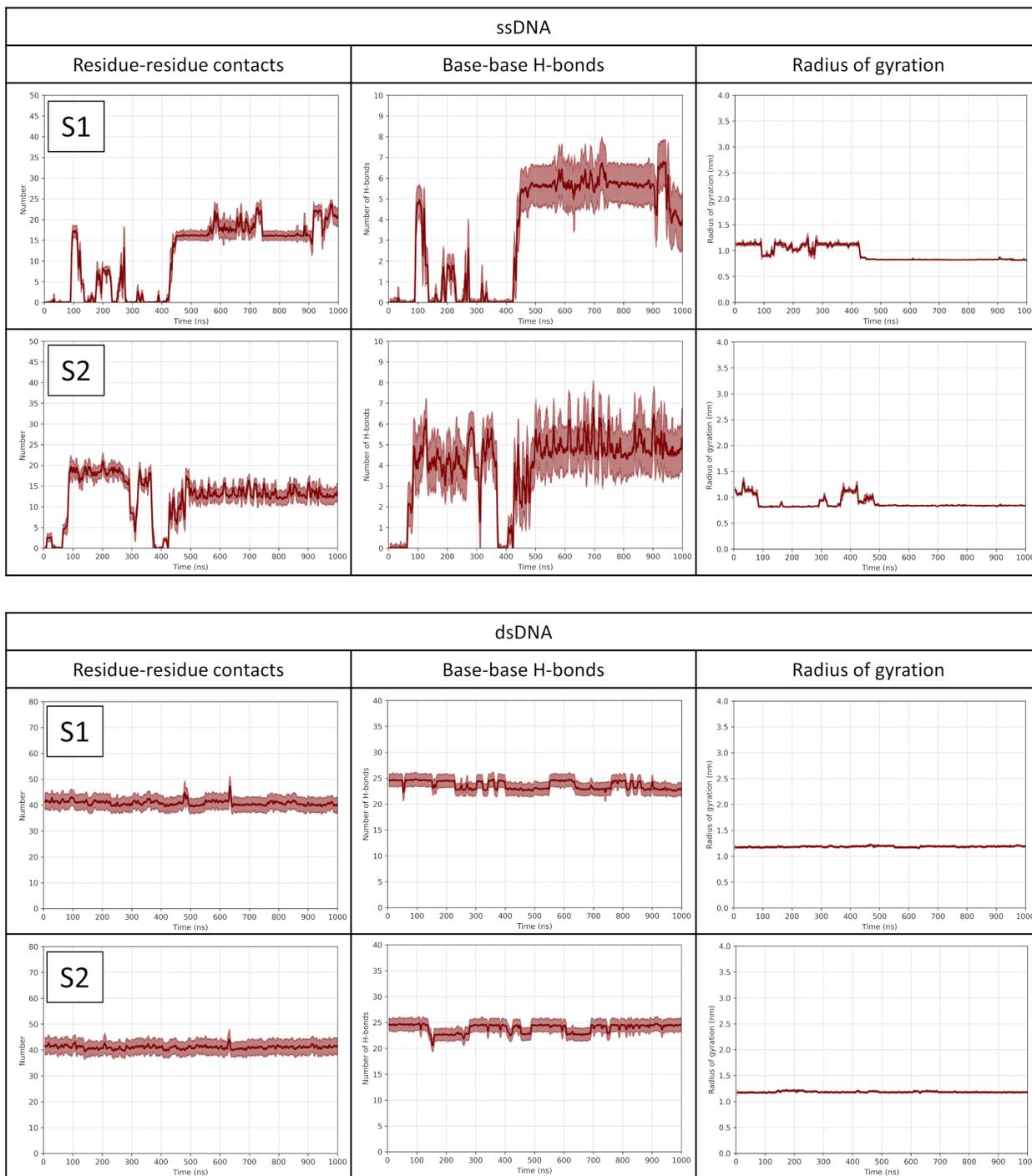


Fig. S8 Convergence intervals of the DNA simulations in solution. The number of residue-residue contacts, the number of base-base H-bases and the radius of gyration are shown for (TOP) the single-stranded DNA (BOTTOM) and the double-stranded DNA. The curve is the running time average using a 5-ns window and its envelop represents the 1- σ interval.

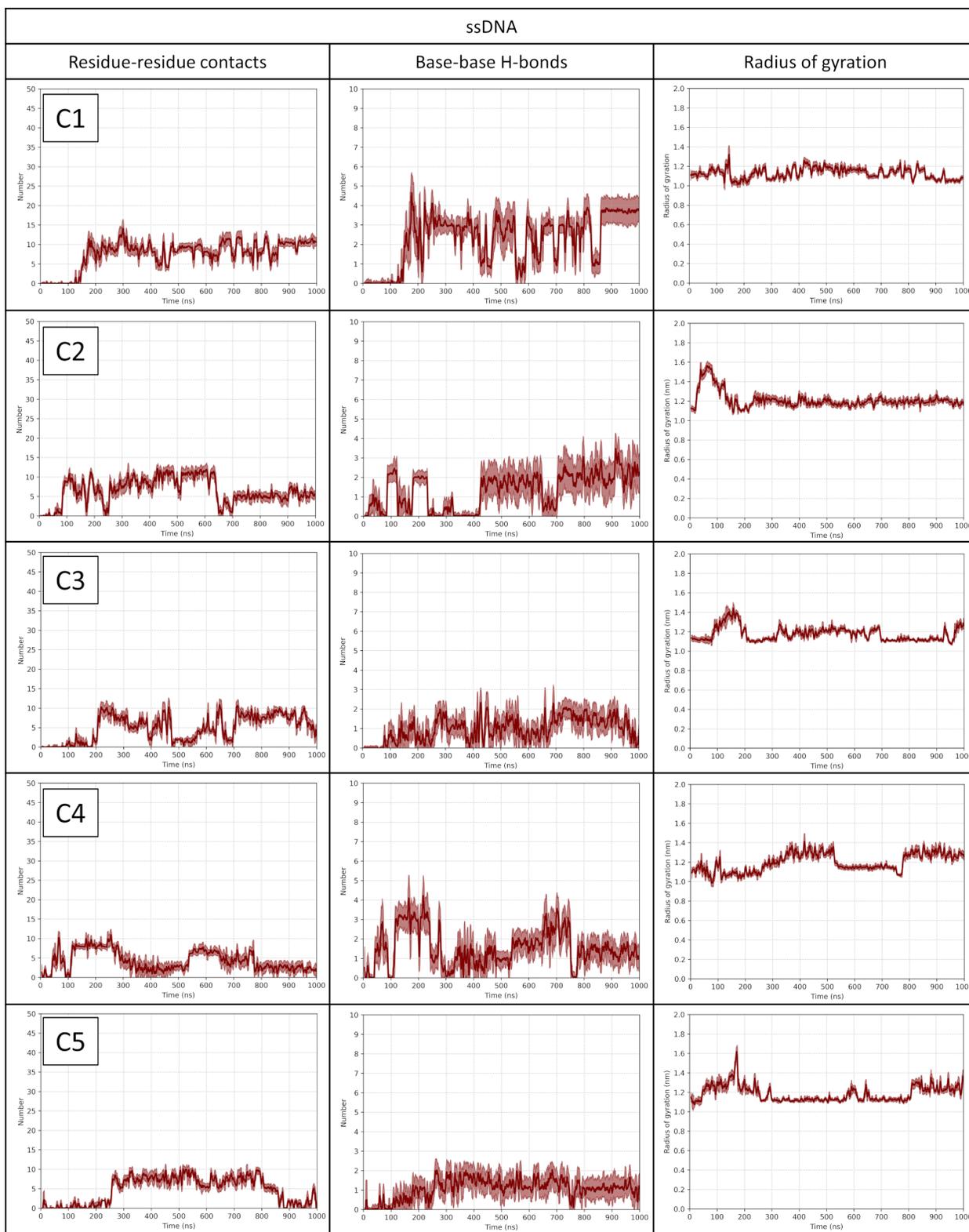


Fig. S9 Convergence intervals of the ssDNA simulations with the nanotube. The number of residue-residue contacts, the number of base-base H-bases and the radius of gyration are shown for the single-stranded DNA. The curve is the running time average using a 5-ns window and its envelop represents the $1-\sigma$ interval.

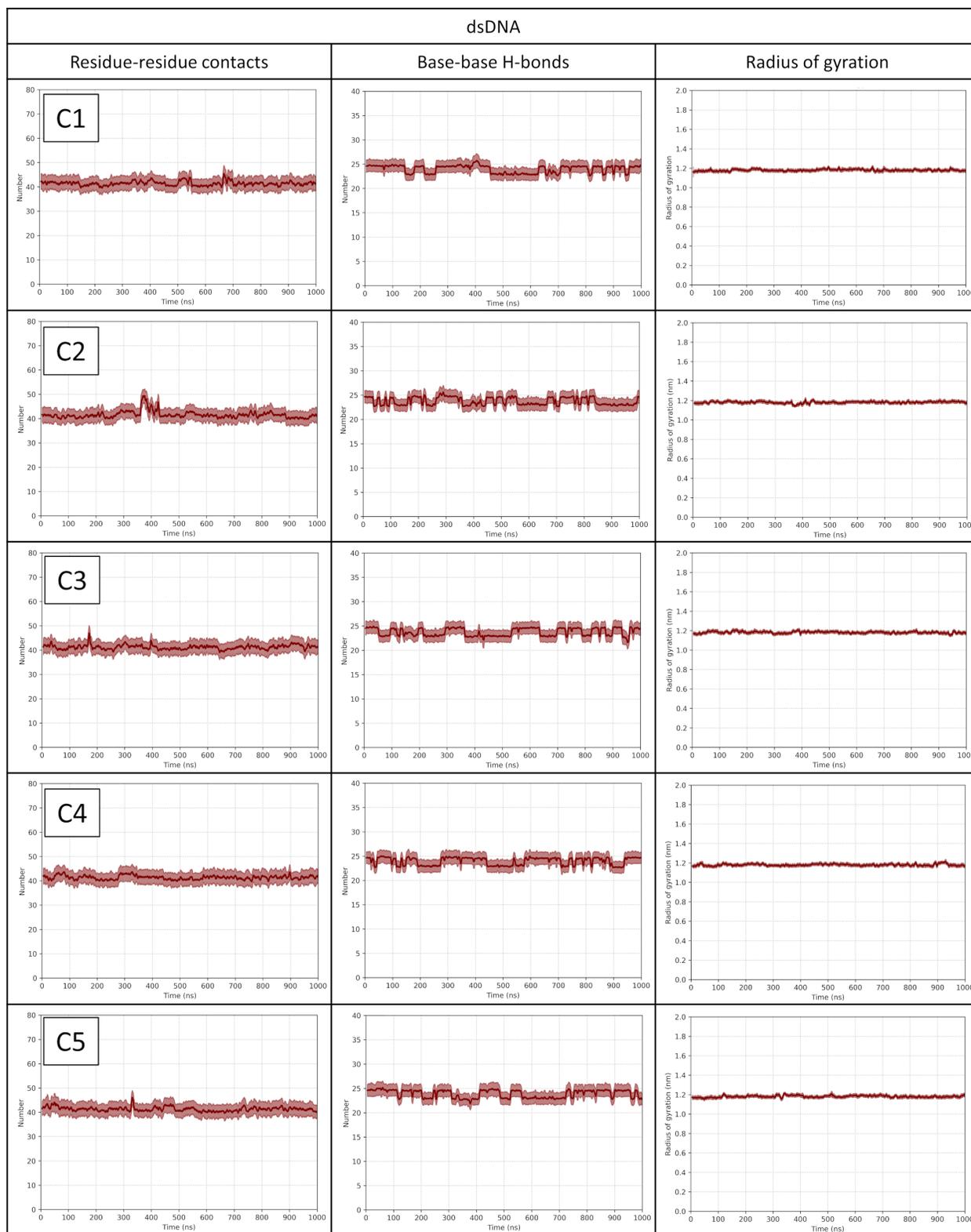


Fig. S10 Convergence intervals of the dsDNA simulations with the nanotube. The number of residue-residue contacts, the number of base-base H-bases and the radius of gyration are shown for the single-stranded DNA. The curve is the running time average using a 5-ns window and its envelop represents the 1- σ interval.

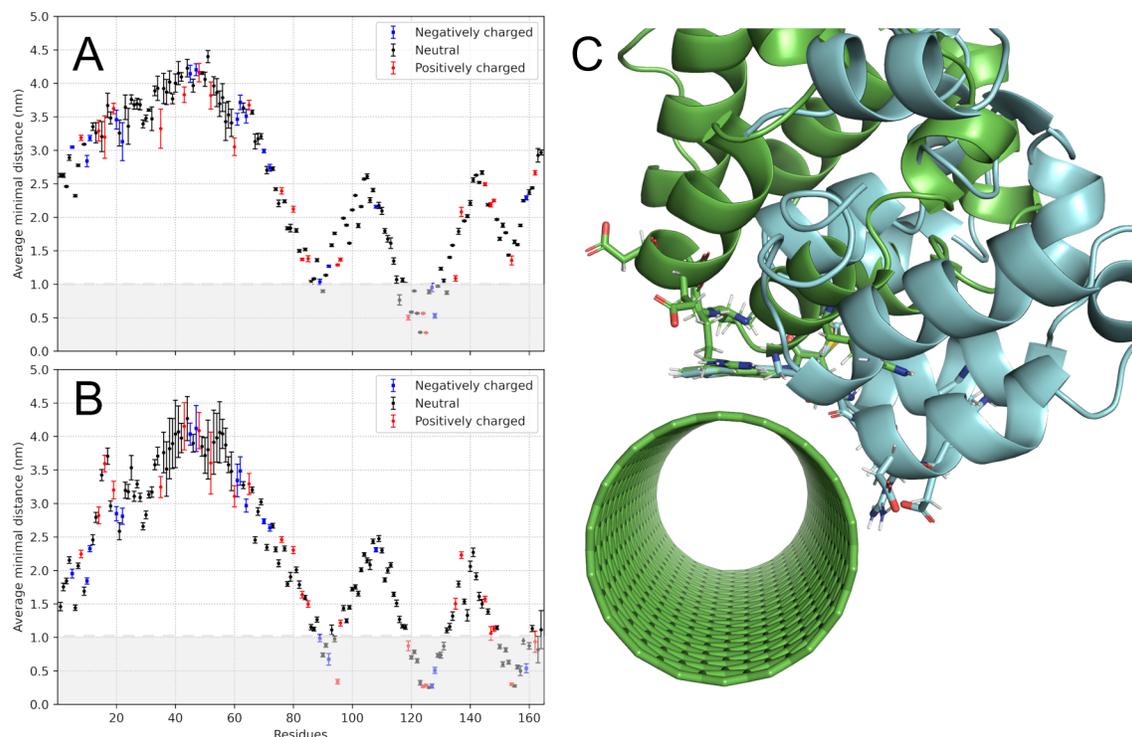


Fig. S11 Simulation C4 of the lysozyme. Comparison of simulation C4 of the lysozyme without the ligand in the presence of the carbon nanotube against the other four simulations. Average minimal distance between the carbon nanotube and the amino acids of the lysozyme considering (A) all simulations except C4 and (B) only simulation C4. (C) Main cluster for simulation C1 (green) vs that for simulation C4 (teal). The clusters were determined using a threshold of 0.5 nm on the backbone atoms and aligning only the nanotube.

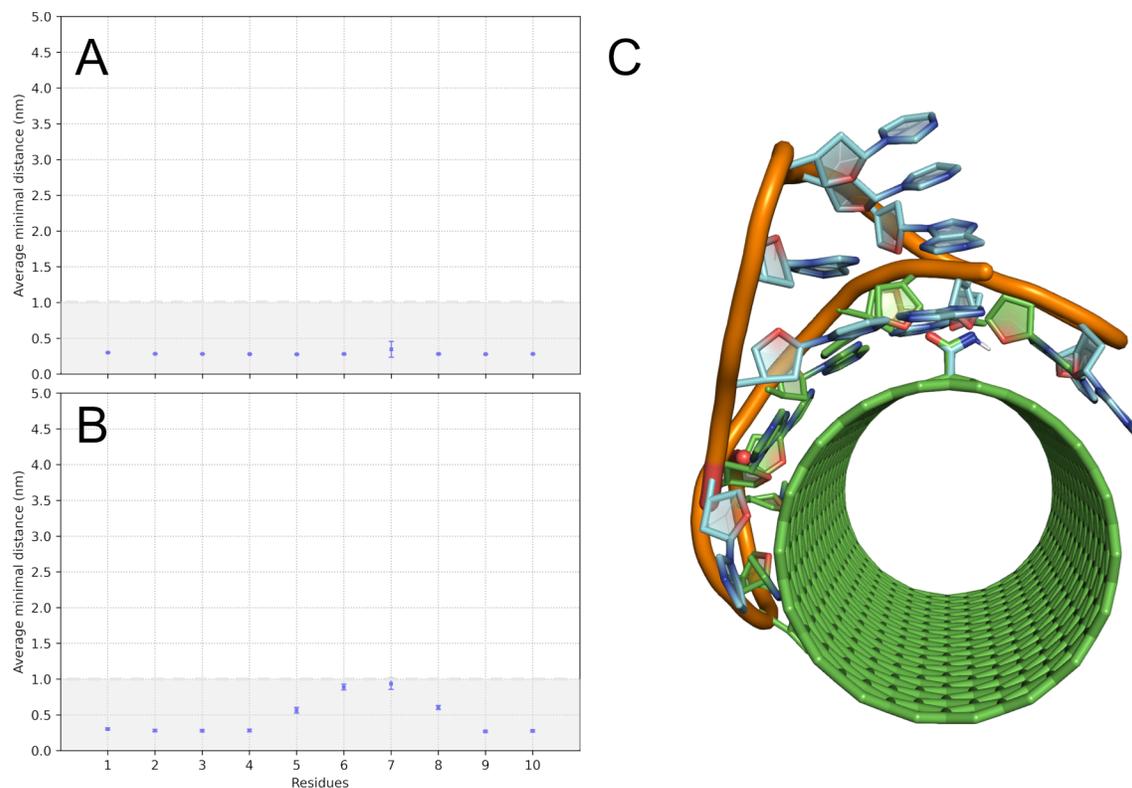


Fig. S12 Simulation C1 of ssDNA. Comparison of simulation C1 of the single-stranded DNA in the presence of the carbon nanotube against the other four simulations. Average minimal distance between the carbon nanotube and the nucleic acids of ssDNA considering (A) all simulations except C1 and (B) only simulation C1. (C) Main cluster for simulation C2 (green) vs that for simulation C1 (teal). The clusters were determined using a threshold of 0.5 nm on the backbone atoms (P, O5', C5', C4', C3' and O3') and aligning only the nanotube.

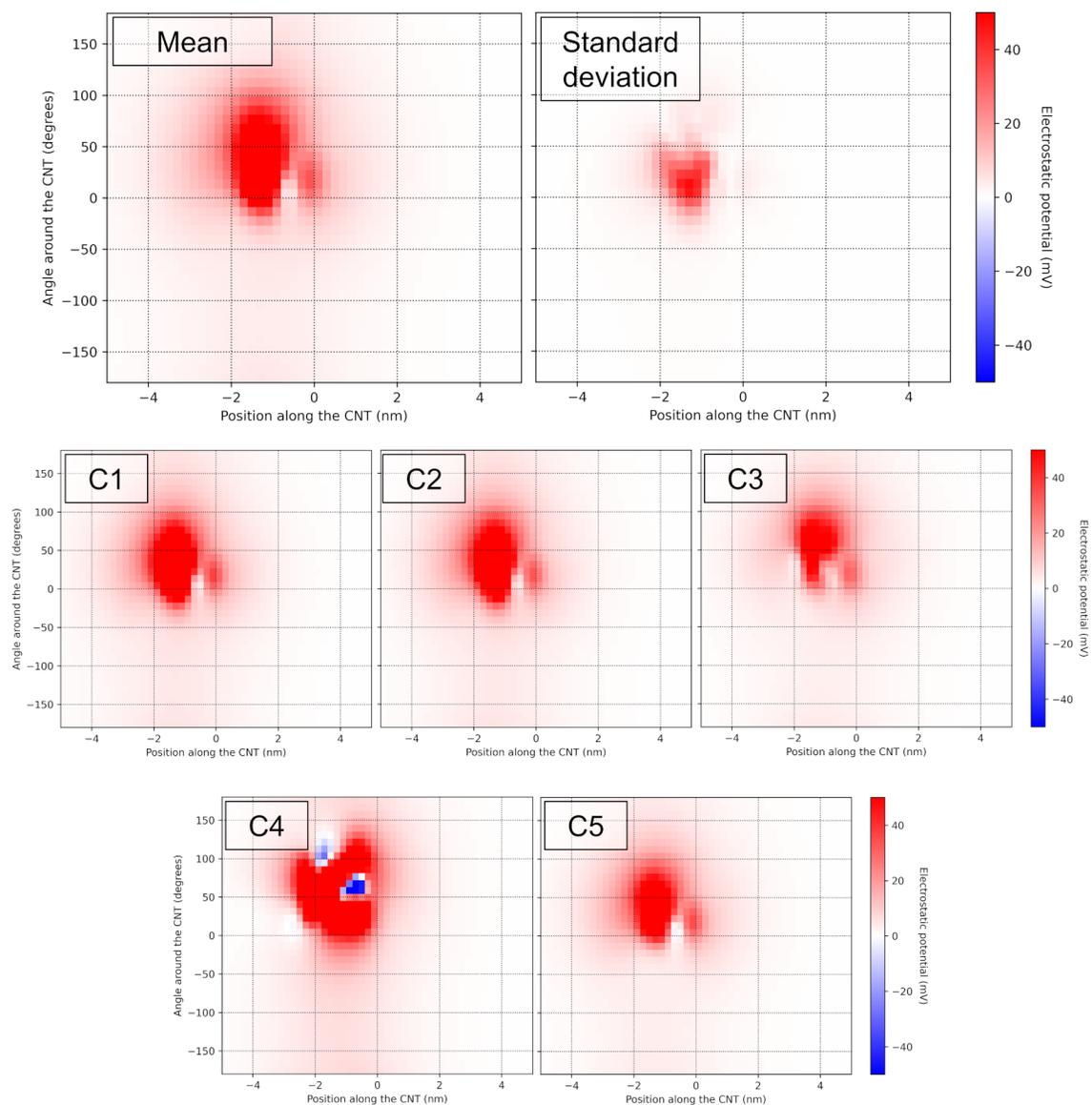


Fig. S13 Electrostatic potential generated by lysozyme. Electrostatic potential (ESP) generated by the lysozyme without the ligand on the surface of the nanotube. The mean and the standard deviation ESP maps are computed over all simulations, except C4. The individual maps represent the mean values on the convergence interval of each simulation.

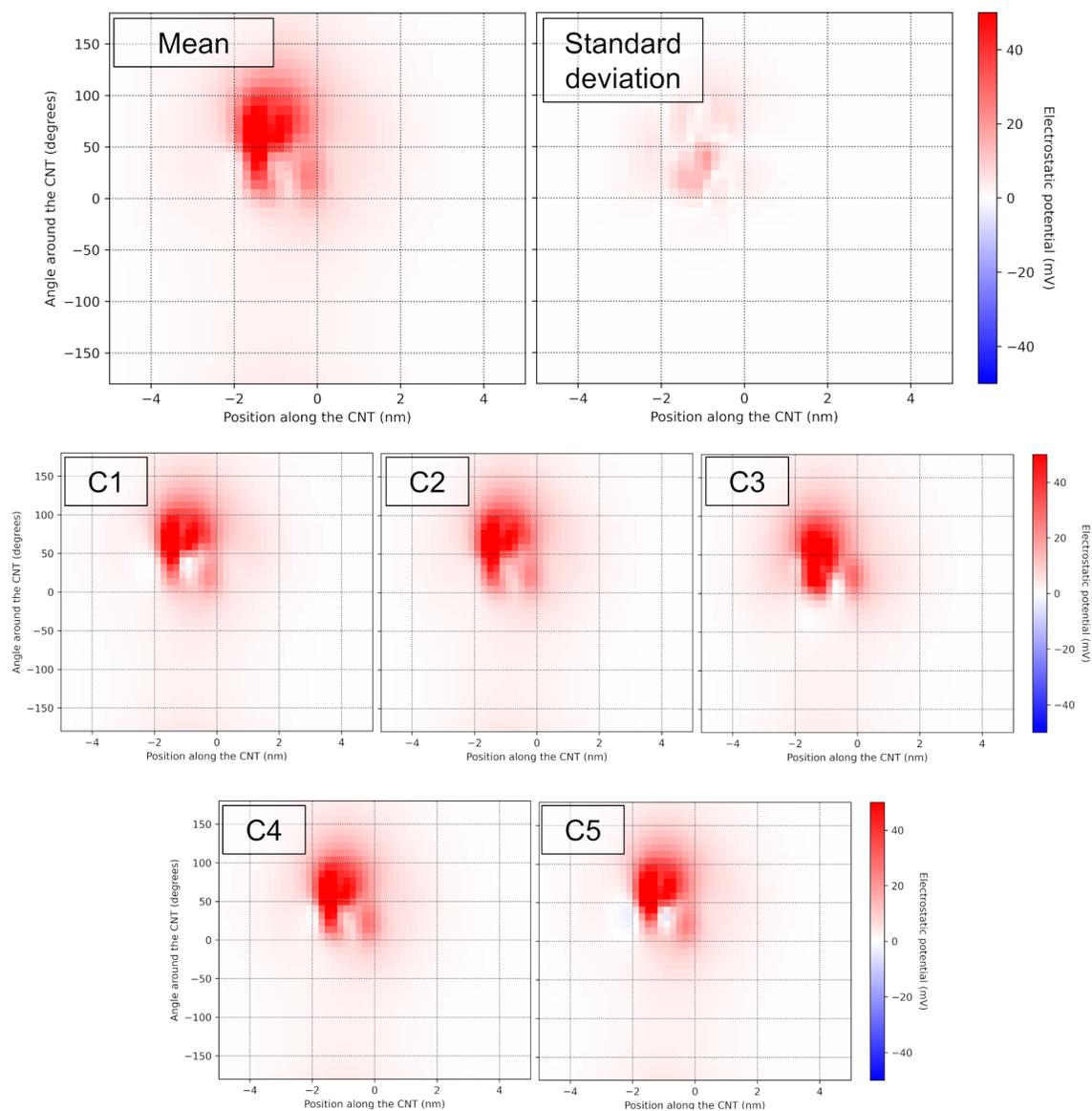


Fig. S14 Electrostatic potential generated by lysozyme-ligand. Electrostatic potential (ESP) generated by the lysozyme with the ligand on the surface of the nanotube. The mean and the standard deviation ESP maps are computed over all simulations. The individual maps represent the mean values on the convergence interval of each simulation.

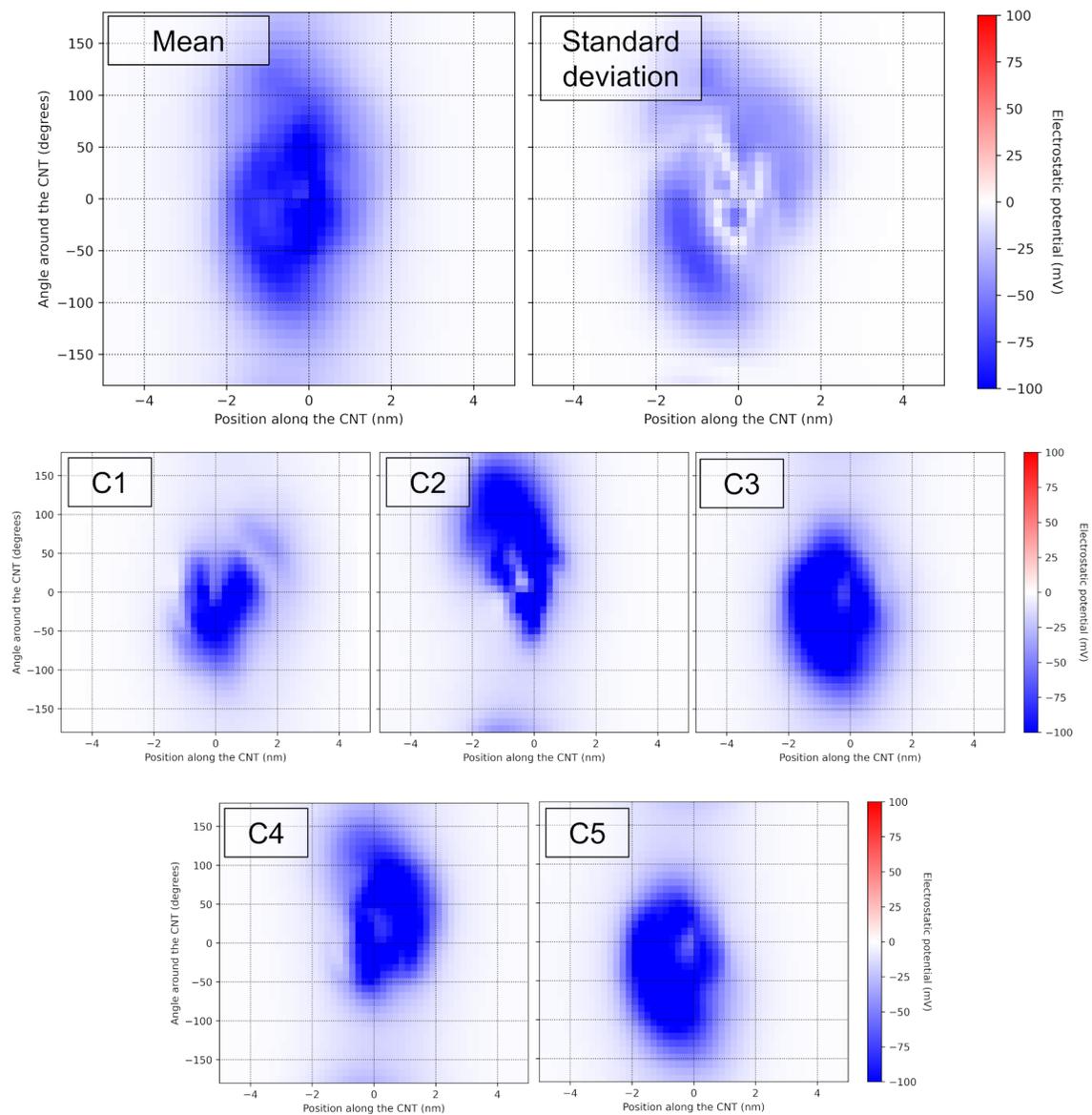


Fig. S15 Electrostatic potential generated by ssDNA. Electrostatic potential (ESP) generated by the single-stranded DNA on the surface of the nanotube. The mean and the standard deviation ESP maps are computed over all simulations, except C1. The individual maps represent the mean values on the convergence interval of each simulation.

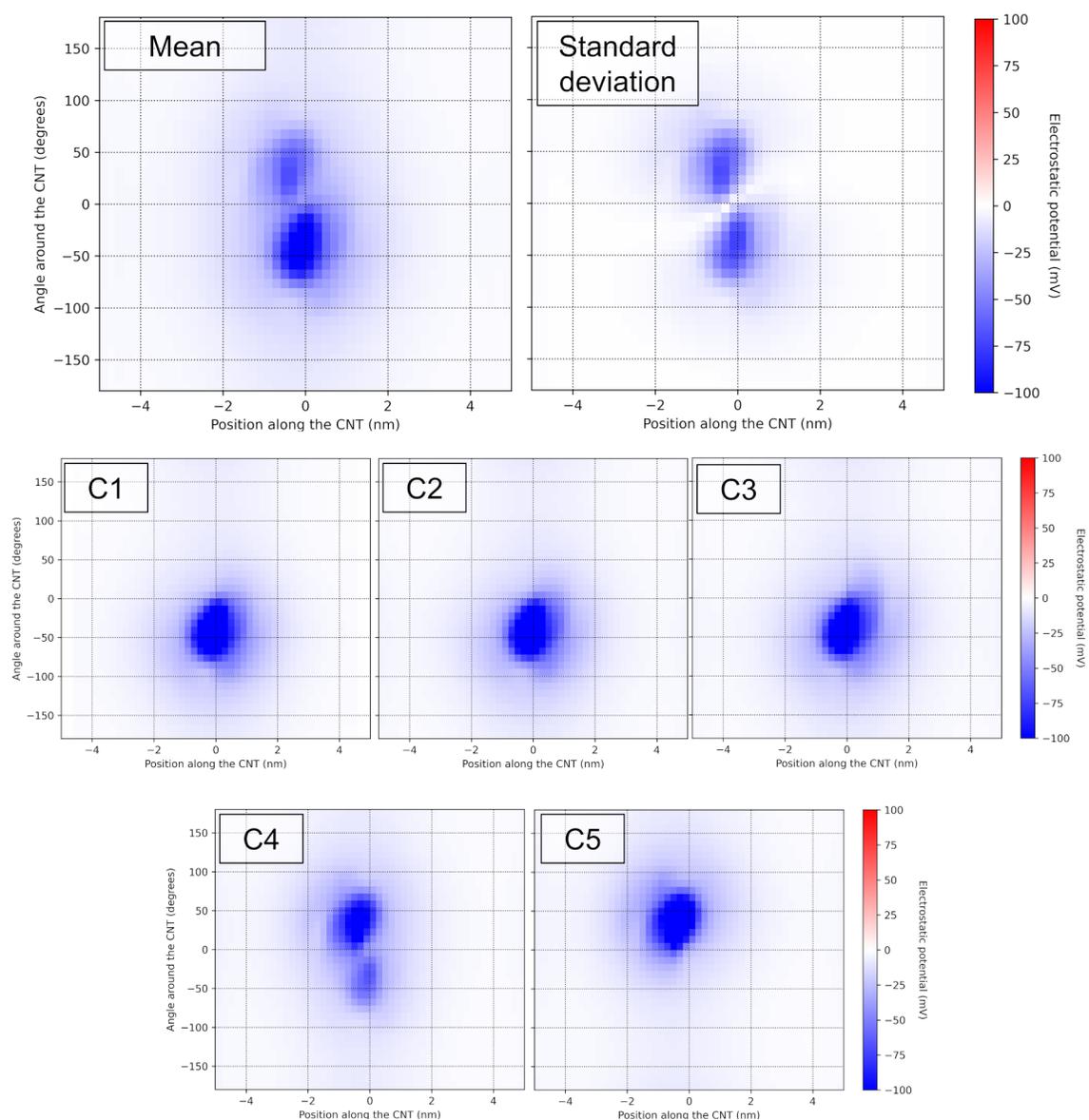


Fig. S16 Electrostatic potential generated by dsDNA. Electrostatic potential (ESP) generated by double-stranded DNA on the surface of the nanotube. The mean and the standard deviation ESP maps are computed over all simulations. The individual maps represent the mean values on the convergence interval of each simulation.

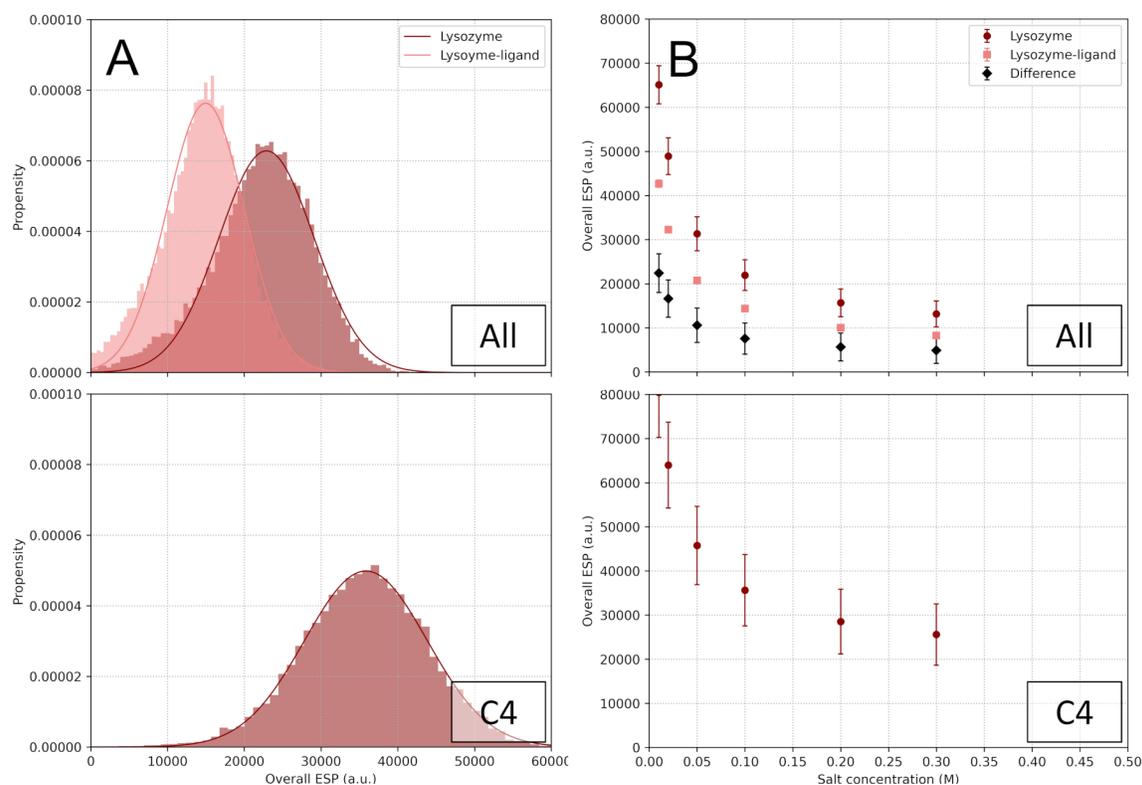


Fig. S17 Overall electrostatic potential generated by lysozyme. The overall electrostatic potential (ESP) generated by the lysozyme without the ligand on the surface of the nanotube in (B) simulation C4 is compared to (A) the other simulations. The error bars represents $1-\sigma$ (A) over the average from the simulations and (B) over all configurations in simulation C4. Statistics are performed on the converged interval of each simulation.

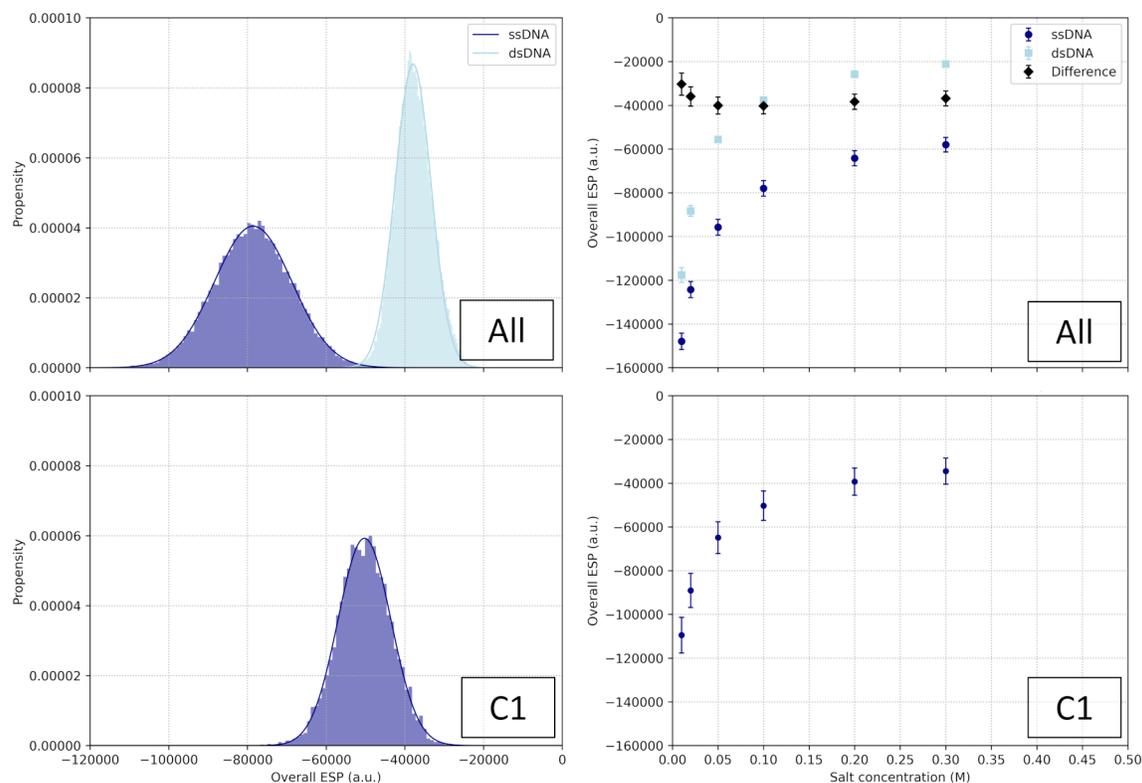


Fig. S18 Overall electrostatic potential generated by ssDNA. The overall electrostatic potential (ESP) generated by the single-stranded DNA on the surface of the nanotube in (B) simulation C1 is compared to (A) the other simulations. The error bars represents $1-\sigma$ (A) over the average from the simulations and (B) over all configurations in simulation C1. Statistics are performed on the converged interval of each simulation.

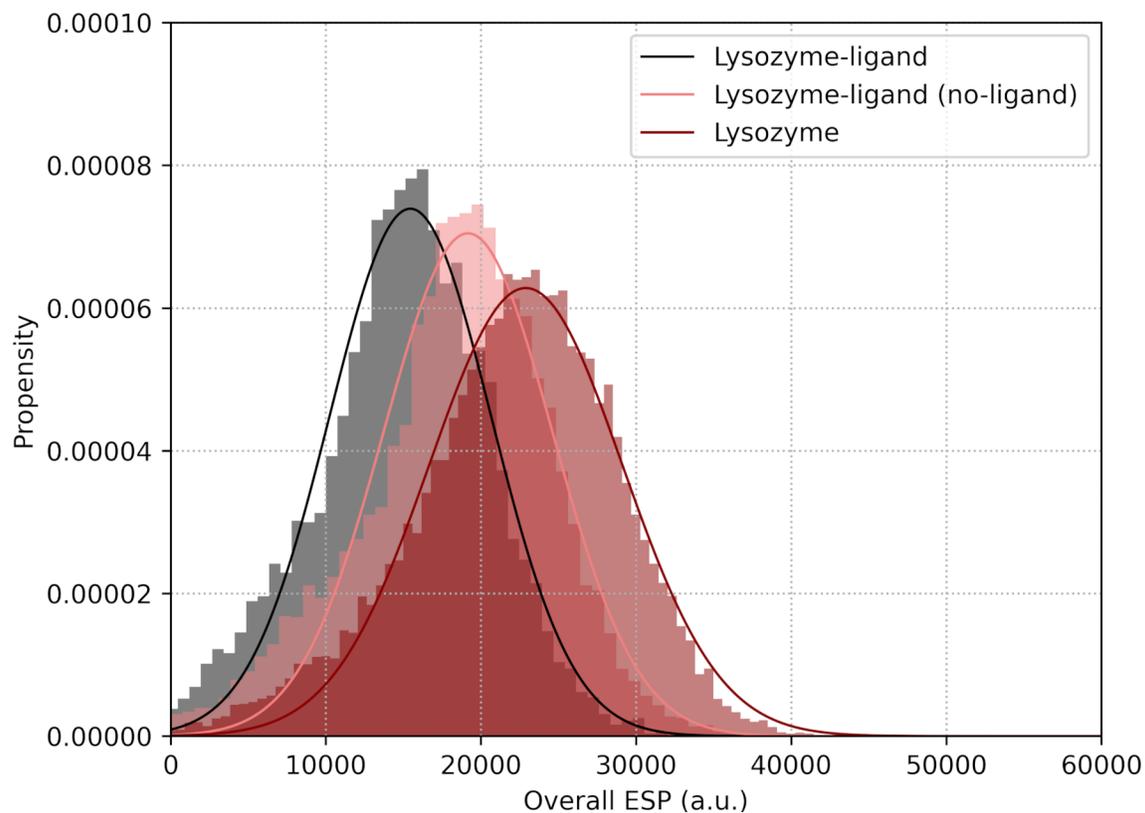


Fig. S19 The effect of the ligand on the overall electrostatic potential generated by lysozyme-ligand. The overall electrostatic potential (ESP) generated by the lysozyme-ligand system (black) is compared to that of the lysozyme-ligand system without considering the ligand atoms (pale red) and to that of the lysozyme system (dark red). Only simulation C3 is considered for these ESP calculations on the lysozyme-ligand system because the ESP distribution of all lysozyme-ligand simulations are similar (Figure 4).