CHARMM Molecular Simulation Observations of Carbonaceous Materials as Possible Templates for Prebiotic Nucleic Acid Oligomer Formation

Draft Manuscript

Anna Du

Barnas Monteith

DPI Polymer Research Group

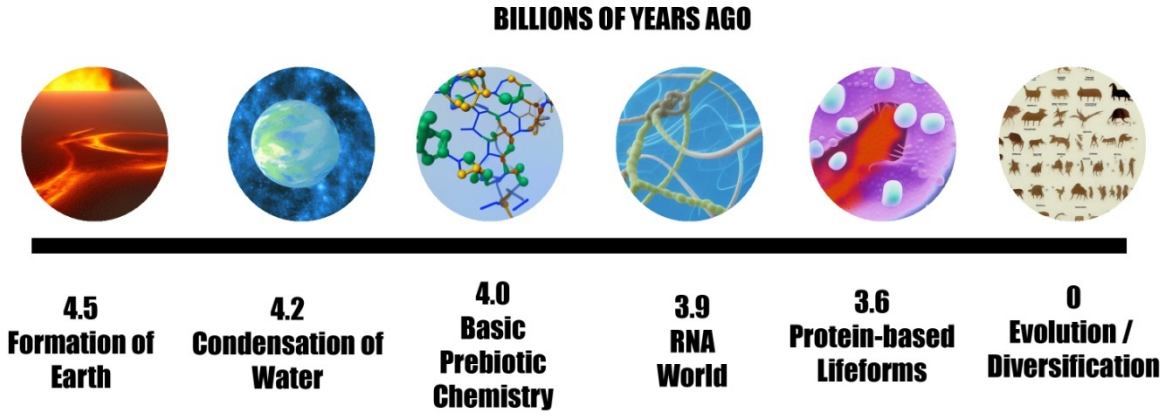
ABSTRACT

This research utilizes CHARMM molecular simulations for the purposes of an initial qualitative observational study of diamond-like carbon and related carbonaceous substrates as potential templates for the abiotic formation of RNA polymers,. While a number of research group have looked at nanodiamond, graphite and other carbon and non-carbon-based materials (especially clays) as possible solid templates for early life, we propose that diamond-like carbon (DLC, also known as amorphous diamond or aD) may have been a plausible ideal template. This is due largely to the fact that DLC is a heterogeneous material, essentially an aggregation of various constituent carbon species (like grapheme, graphite, CNT, etc.), depending on formation conditions. Further, DLC exhibits a range of superlative material properties

shared by diamond and graphitic materials It possesses high thermal diffusivity and conductivity along all planes due to its inherent anisotropism. Internal and surface structure variability could have played a crucial role in facilitating the molecular interactions that accelerated the abiotic formation of RNA oligomers, potentially supporting the RNA World hypothesis. Significant outside energy is not needed to enable oligomer formation. Entropy, in the form of heat, is a driving force that provides conditions necessary for oligomer formation. To reduce computational and financial costs, we investigate artificial intelligence-driven methodologies to inform optimal MD input parameters with the goal of generating key qualitative endpoints. As MD simulations are extremely computationally expensive due to the large quantity of electronegativity and force field interactions for trajectory calculations, we initially used a lightweight MD system combined with a transformer model to better predict the likelihood of achieving collisions between target molecules, and thus increase the chance of observing exemplar endpoints. Various key endpoints were observed, including Van der Waals oligomer-template attachment, translocation/rotation potentially leading to more complex configurations, quasi-elongation and bridging, as well as detachment). The observations of these endpoints suggest that carbonaceous materials such as DLC could plausibly have served as templates for the formation of early nucleic acid oligomers. Further simulations and physical experiments are needed to confirm these results, and to develop a supporting statistical model.

Background

The question of life's origin has driven researchers to unravel the mysteries underlying the transition from non-life to life. This pursuit is not merely an intellectual endeavor but has profound implications for our understanding of the universe, our place within it, and the potential for extraterrestrial life. It also casts a light on how humans can deal with future problems, such as diseases or environmental disasters. While numerous theories and experimental approaches have been proposed, a comprehensive understanding of how life might emerge from simple chemical compounds remains elusive.



Visual representation of Earth's history across billions of years for the formation of life (diagram by authors, after Walker, 2017).

Among the more prominent of current abioogenesis theories is the concept of the RNA world, which suggests that RNA may have been the first informational molecule, and it played a pivotal role in the transition from simple pre-biotic chemistry to complex, self-replicating life forms. The Miller-Urey experiment of the 1950s demonstrated that various amino acids (precursors of RNA) could be synthesized from simple gases and minimal applied input energy under conditions mimicking those of the early Earth (Miller & Urey, 1959), and presumably other locations throughout the universe.

This is further supported by recent discoveries of organic compounds from extraterrestrial sources. Uracil, one of the key nucleobases of RNA, was discovered on the surface of the near-Earth asteroid Ryugu (2023). The presence of uracil on an extraterrestrial body holds profound implications for the plausibility of abiogenetic nucleotide synthesis, not just on Earth but throughout the universe. The identification of uracil on an asteroid surface, far removed from Earth's biosphere, suggests that the formation of these vital molecules might have occurred through natural processes and available precursor materials prevalent in outer space.

Various ionic materials have garnered attention as potential catalysts in prebiotic chemistry. Over the years, a significant body of research has also focused on the potential of solid catalysts, particularly clays, to support prebiotic reactions. It is commonly accepted that out of the 300+ clays that are common today, which likely existed towards the beginning of Earth, one of the most promising has been and continues to be montmorillonite. While on the one hand, it has been observed that amino acids and oligomers readily attach to the surface of clays, the mechanism of attachment of long chain biopolymers to inorganic templatizing surfaces tends to be parallel, or quasi-parallel, rather than normal, to the surface (note simulation images in Greathouse, et atl, 2014). This may inhibit longer chains from forming, and would make detachment much more difficult.

There have been other confirmatory experiments, both in vitro and in-vivo , which have validated the idea that abiotic development of life is plausible, purely through a combination of entropy and random stochastic chemical interactions (Prigogine 1975, England, 2013). Recent advances in force-field based MD and economically plausible large scale studies of organic chemical formation, have led to a number of newer theories proposed for the progression of life between the early inorganic formation of basic organic molecules and the last universal common ancestor (LUCA) during a period referred to as the RNA World period. Current theories consider plausible prebiotic origins in hydrothermal vents, intertidal pools, warm ponds, and even extraterrestrial sources. The existence of abiogenetically formed informational molecule precursors outside of Earth is validated by the presence of confirmed extraterrestrial amino acids and nucleobases extracted from a number of meteorites.

Hypothesis

This work presents a plausible carbonaceous template concept for abiogenesis, departing from the assumptions that relatively homogenous materials like montmorillonite or nanodiamonds were the only abundant likely templates for organic polymer growth. We propose that heterogenous carbon materials

with non-uniform crystal lattice structures, like amorphous diamond / DLC (diamond-like carbon), could have been ideal, both structurally, and chemically, for the temporary, yet robust, attachment, growth and detachment of polymeric organic molecules. Structures that contain predominantly sp3 hybridized and sp2 bonded carbon atoms could have served as plausible binding and nucleation sites for stable biopolymer growth, leading to a localized, saturated condition where adjacent oligomers could interact, and form complex, functional 3D structures, and ultimately serve the function of informational (genetic) molecules in early precellular/protocellular organisms. Robust seeding/anchoring points allowed for the maximum high density growth of stable biopolymers, at lengths that are more biologically relevant (100's of mers versus a few dozen as had been typically observed in lab experiments in the past), as well as subsequent release into the local aqueous environment (Totani, 2020). This could have plausibly achieved a saturation point of sufficient biopolymers to make the various subsequent abiogenesis theories statistically possible, if not likely, given enough time.

This work aims to explore carbonaceous-template based mechanisms that could have driven the early stages of chemical evolution. In particular, diamond-like carbon (DLC) and related materials are viewed as potential catalysts for biopolymer growth, due to the stochastic distribution of different constituent carbon species, anisotropism, and superlative thermal conductivity/diffusivity. Among the diverse array of materials investigated for their role in prebiotic chemistry, diamond-like carbon emerges as a unique candidate with many benefits. DLC, composed of carbon atoms bonded in a mixture of sp2 and sp3 hybridization states, possesses a distinct nanocrystalline structure characterized by its variability and heterogeneity (synthetic DLC is highly tunable based on deposition formula). DLC is known for its varied surface nanostructures, which can take various forms (including needle-like shapes that have been utilized in technology industries for their field emission properties), which can be modulated based on synthesis recipe/conditions. This distinctive morphological variability feature gives rise to the possibility of a range of potential binding sites, making DLC an attractive template for seeding and catalyzing the growth of biopolymers.

Based on today's knowledge of synthetic DLC deposition, it is thought that the conditions and materials present within and surrounding early Earth would have made the presence of DLC likely. The fact that DLC could have been abundant throughout the universe makes it an appealing candidate for playing a role in prebiotic chemistry. The prevalence of methane in outer space and the high vacuum environment, coupled with radiation and energy sources, resemble the conditions that can lead to the formation of DLC. These are the same conditions used to produce synthetic DLC in lab environments. There is also evidence that instantiated high pressure environments, which were plentiful within the early Earth crust, could have formed both amorphous diamond/DLC and fully sp3 hybridized diamond crystals, which were subsequently exposed on the surface through constant upheaval of tectonic/volcanic activity, which was prevalent throughout early Earth. Whether or not aluminosilicate clays/minerals were abundant, the likely presence of heterogeneous crystal lattice structures as well as mixtures of homogeneous carbonaceous materials would have been highly efficient template for seeding and biopolymerization of long chain informational molecules. It is reasonable to infer that DLC must have existed on a prebiotic Earth, possibly exposed to the surface through kimberlite pipes or other such mechanisms, but also potentially originating from extraterrestrial sources. Furthermore, the discovery of asteroids containing nanodiamonds on Earth provides additional evidence of the potential widespread distribution of diamond-like carbon in the cosmos, fostering intriguing connections between space chemistry and the emergence of life throughout the universe. Serpentinite xenoliths have been found in Sicily containing carbonaceous materials, demonstrating the presence of organic compounds during the Mesozoic era (Simakov, et al, 2020).

Unlike many crystalline structures, such as diamond, with uniform lattice structures, DLC can be a combination of both crystals and non-crystalline carbon-based structures, representing a mumber of discrete and hybrid carbon species. This forms a heterogeneous structure comprised of an amalgamation of closely packed intermingled nanostructures. This inherent variability provides a range of potential binding sites for biomolecules. Modulation of the formation conditions results in a wide range of potential

material properties, ranging from more graphitic to more diamond-like (and when highly graphitic, it may be referred to as GLC or graphite-like carbon). This extensibility extends to surface features, including different dimensions of both convex and concave structures, which could serve as anchor points to accommodate the growth of biopolymers of varying lengths, enhancing the possibility of long-chain polymerization on DLC templates. Thus, DLC emerges as a compelling candidate to serve as a template for prebiotic RNA oligonucleotides/ Imperfect crystalline structures found in DLC will theoretically provide more potential sites for nucleotides or various functional groups to anchor via VDW forces.

The field of molecular simulations has progressed gradually over the years, yet simulations remain computationally expensive and generally slow. Recent advances in AI promise to accelerate the field of ab initio molecular dynamics, allowing for greater efficiency and throughput of simulation experiments. In this work we use AI (a Tensorflow transformer model) as a pre-processing step to predict likely molecular collisions, given a set of starting parameters and target molecules. This information is then used to formulate the initial CHARMM setup parameters, in order to achieve a higher likelihood to observe key molecular interaction endpoints.

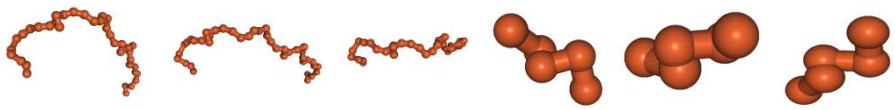
Methods

Polychrom/OpenMM Simulations

Achieving accurate results for MD simulations necessitates calculating the trajectories of thousands to tens of thousands or more molecules, in simulations such as those conducted for this research. These interactions, which are dictated by force fields, electronegativity, and other factors, inherently introduce complexity, which poses challenges to simulation speed, accuracy, and computational efficiency. It is imperative to explore avenues for reducing complexity or otherwise improving efficiencies to expedite simulations, especially when conducting open-ended observational studies. This can be achieved through a number of methods; in the case of this experiment, we do so through parametric optimization. This strategy optimizes simulation starting parameters and choice of

target molecules to enhance computational efficiency. The goal is to identify configurations that are most likely to create a collision event (potentially representative of a key endpoint) while preserving the integrity of the experiment and results (within a realist range of simulated environmental constraints). The ramifications of enhanced simulation efficiency extend beyond speed. By mitigating complexity, simulations become more cost-effective and accessible, allowing for a larger volume of MD simualation runs, and thus more opportunity to observe interactions between nucleotides, oligomers, and carbonaceious templates.

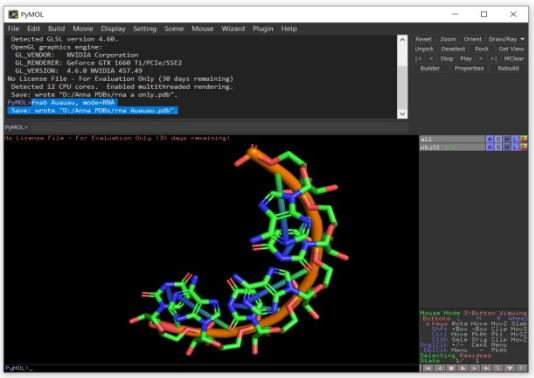
In order to predict likely collisions we used the Polychrom/OpenMM simulation package developed at MIT. This is a lightweight MD modeling tool, to rapidly explore key parameters including polymer density, simulation waterbox temperature, polymer chain amount, and length of polymer. These parameters were methodically modulated to generate an extensive dataset, which would serve as the foundation for training our Tensorflow transformer-based model for starting parameter prep/optimization. Polychrom simulations, which use representative geometric structures such as beads and chains, to emulate biological structures such as nucleotides and phosphate or pentose groups, acted as a rapid first-order proxy for the more comprehensive CHARMM simulations. This approach delivers an accelerated pathway to obtaining a dataset of plausible outcomes based on given starting parameters. The ability of Polychrom to run complex simulations at an accelerated speed, make it an ideal platform for the generation of sufficient AI training data.



3D representations of RNA were generated by using the 'ngutils' visualization tool, from data collected through the Polychrom simulations. The left set of images demonstrates the bead and spring depiction of a polymer of 40 nucleotides, while the right set shows a closer view of a smaller polymer of 6 nucleotides. In this visualization, each individual bead corresponds to a single nucleotide.

Polychrom-generated data, including collision information, kinetic and potential energy, polymer length, and density were then used to train a Tensorflow transformer model, which was subsequently used to aid in determining ideal starting parameters for the lengthier and more costly CHARMM simulations. Chosen starting parameters for exemplar observed results below are noted in the supplemental data materials. Full details of this parameter optimization method are being prepared for publication in a separate paper.

PyMol, an open source molecular visualization tool, was utilized to generate the .PDB (protein data bank) files of the oligonucleotides based on the scaled down parameters from the transformer output. The resultant .PDB files categorized each individual atom as a 'HETATM' (heterogeneous atom), and furthermore, PDB manipulation software packages tended to alter the formats and atom definitions, causing simulations to fail. To address these compatibility issues and ensure accurate representation, the PDBFixer application was employed. PDBFixer redesignates the individual target molecules as separate RNA strands, standardizing the positions of all atoms. Furthermore, other necessary conditions were normalized, rendering the resultant .PDB file fully compatible with CHARMM-GUI.

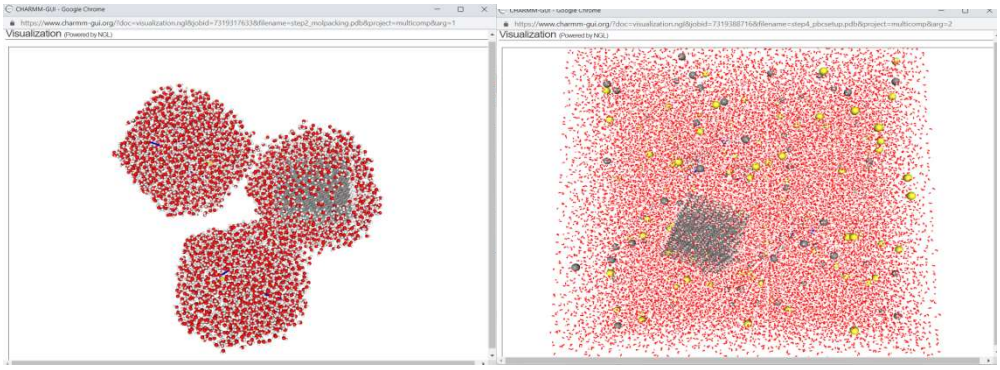


Using PyMol to generate oligonucleotide strands with specific sequences to use in CHARMM simulations

The generated and refined PDB files were then entered into CHARMM-GUI, a suite of tools for MD simulation file preparation (Jo, et al, 2008). Specifically, the Solution Builder module within CHARMM-GUI was used. This step involved the creation of an emulated waterbox surrounding the

chosen oligonucleotides, thus providing the aqueous environment for simulations. Normalized results from the transformer step mentioned above were used to inform setup parameters in CHARMM-GUI.

The outputs from PDBFixer were used as PDB structure file inputs for CHARMM-GUI's Nanomaterial Modeler. The carbonaceous species such as graphene, carbon nanotubes (CNT), and graphite were also prepared, within waterbox solutions. Subsequently, the Multicomponent Assembler module was utilized to combine these various components along with the refined oligonucleotide .PDB files. This process yielded a composite system encompassing all components. Furthermore, the system was ionized using a default concentration of sodium chloride (NaCl) to simulate a saline environment.



(Left) A preview of a simulation with a sample of graphite and two oligonucleotide strands on CHARMM-GUI prior to encasement within a waterbox. (Right) A single waterbox is constructed around the individual experiment components. The sheets of graphite, two oligonucleotides, and 0.15 M NaCl ions are then scattered throughout the final simulation environment using the Monte-Carlo method.

CHARMM Simulations

CHARMM (Chemistry at HARvard Molecular Mechanics) is an open source software to run molecular simulations. CHARMM utilizes sophisticated force fields, energy and other data to plot trajectories and orientations of each molecule, one frame at a time. These calculations include parameters that define bond lengths, angles, dihedral angles, non-bonded interactions, and more. They are created based on experimental data and derived calculations, effectively translating complex molecular

interactions into realistic 3D visualizations of highly controlled experiments, with ideal, repeatable conditions. Individual simulation run times for the experiments conducted within this research range from many hours for simple initial tests, to multiple days for full simulation runs, with the hardware configurations described below.

The resulting file packages from CHARMM-GUI were used to create setup files/environments for CHARMM C47B1 OpenMM 7.6 to create trajectory files (.DCD). The following parameters were used:

- Fortran 90
- CUDA 9.2
- CMake 2.8.12.2
- C++ 7.3.1

Most simulations were conducted on AWS cloud-based on-demand processing computing services; some testing was conducted on a small cluster of Lenovo i7 laptops with a RTX-class NVIDIA GPU running the latest version of Ubuntu Linux.

1. Large Server n=1
    - 64 GB GPU
    - 16 core CPU, 64 RAM
2. Small Server n=3
    - 16 GB GPU
    - Quad core CPU, 16 RAM

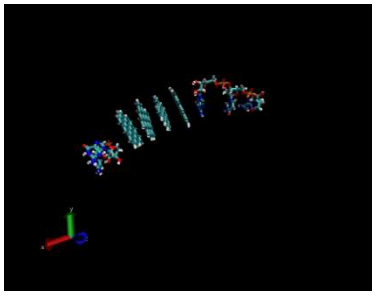
Optimal Configuration Testing			
GPU Type	GPU	Swap for CPU Mem	Average time to finish 1 iteration of production step (hours)
None	0	Yes	> 50
NVIDIA T4	1	Yes	49
NVIDIA T4	4	Yes	23.2
NVIDIA T4	24	Yes	21.3
NVIDIA T4	1	No	13.8
NVIDIA T4	4	No	8.1
NVIDIA T4	24	No	7.8

Initial test models were first run to gain a preliminary understanding of how long it would take for a single simulation to run until exemplar endpoints could be achieved, given typical setup parameters, and whether or not the runs would fail due to configuration errors. Occasionally, simulations would fail due to out-of-bounds-type errors, and could be restarted by deleting the current steps/frames and restarting the simulation from the point of failure; very few simulations experienced a fatal failure and had to be abandoned. Following simulations, the resulting .DCD trajectory files were then visualized and analyzed using Visual MD (VMD). Note that DCD files from the exemplar results below are provided as downloadable file links in the supplementary data materials.

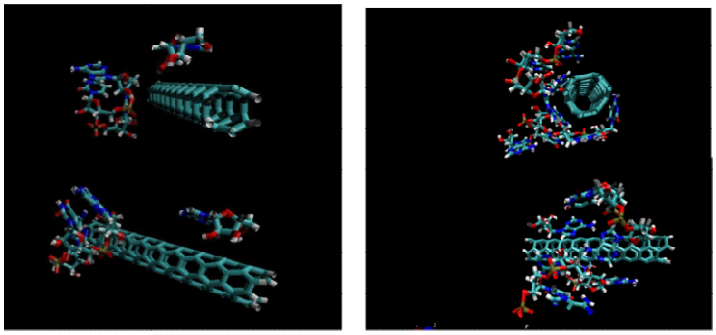
Based on the CHARMM simulations conducted followed by VMD observations, a number of key qualitative endpoints were observed, unveiling intriguing molecular interactions and behaviors. Among the notable outcomes were:

- Oligomer Bridging via Carbon Sheet Structures: The simulations revealed instances in which oligomers exhibited bridging behavior, connecting to opposing sides of a series of graphene sheets. While this represents a discontinuos polymerization process, it points to the possibility

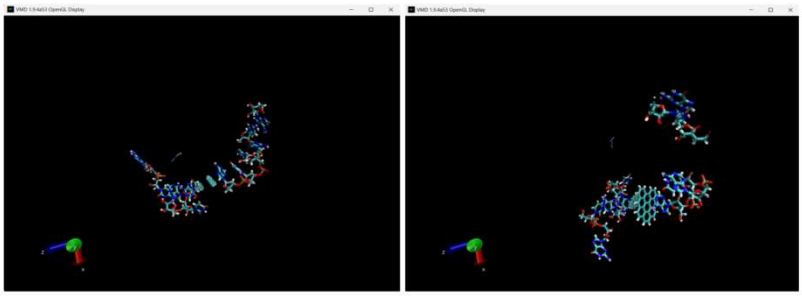
that the carbon sheets may act as a means of drawing multiple oligomers together into closer proximity, facilitating subsequent interactions.



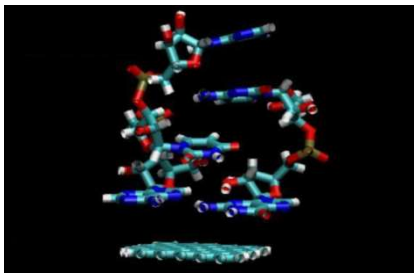
- Oligomer Attachment to Carbon Nanotubes: The simulations observed captured the attachment of oligonucleotides to the surface of carbon nanotubes. The oligonucleotides generally bonded in a parallel fashion to the surface, along the length of the cylindrical CNT.



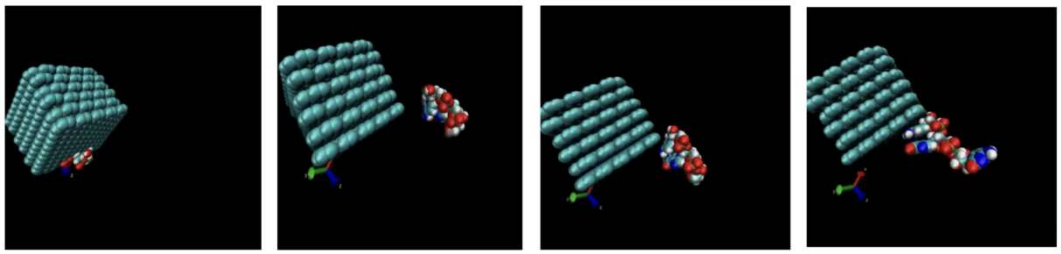
- Detachment of Dimer from Oligonucleotide Complex: In this simulation, a dimer dissociated from a complex containg a much larger, 5-mer oligonucleotide.



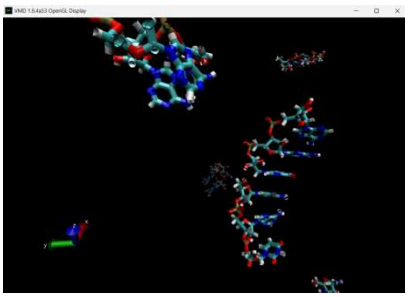
- Close proximity, parallel configurations: This scenario demonstrates two oligonucleotides in close proximity on a graphene sheet, that have rotated into positions that could potentially lead to hydrogen bonding and ultimately, double stranded molecules. This shows potential for these dimers to initiate the formation of more complex organized structures if given sufficient simulation time, or perhaps given different temperatures, initial trajectories, or other starting input parameters.



- Surface Attachment and Translocation: This simulation yielded multiple instances of oligonucleotide VDW bonding to the surface of graphite sheets, representing multiple dimensions.



- Quasi-elongation: The simulations captured instances resembling polymer elongation, where multiple oligonucleotides formed a Van der Waals bond along the ends of the oligomers, demonstrating the possibility for the formation of longer chain oligonucleotides.



These observations underscore the complexity and potential for molecular interactions facilitated by carbonaceous substrates. The transformer outputs not only enabled the identification of key endpoints but also provided a comprehensive view of how starting parameters and changes in simulated conditions influenced the observed behaviors.

An initial round of physical experiments were conducted at the UMASS Boston Integrated Sciences Complex Analytics facilities, using both nucleotides and short oligomers, as well as a variety of carbonaceous substrates. Experiments were conducted and analyzed using high performance liquid chromatography (HPLC) and liquid chromatography combined with tandem mass spectroscopy (LC-MS-MS). Initial positive results were obtained. Further tests are planned, and are being prepared for separate publication

Discussion

This work presents a rudimentary first-order look at the possibilities and mechanisms of oligomer and nucleotide interactions with carbonaceous substrates. The carbon based substrates (graphite, grapheme and CNT) were chosen out of convenience due to their ready availability within the CHARM-GUI model file preparation system (and share many characteristics with certain forms of DLC). Additional force field and structure files representative of different types of DLC and nanodiamond were prepared with the assistance of researchers at the Interface Force Field team. These will be incorporated into future work. As mentioned in the results, multiple instances of oligomer attachment, detachment, translocation/rotation on carbonaceous template surfaces and other key endpoints were observed. Evidence of plausible

elongation was also noted, and it is possible that proper covalent and hydrogen bonding could ultimately form, with additional simulation time.

For the most basic lifeforms to exist, oligonucleotide chains would ideally exceed 300 nucleotides; complex proteins require lengthy RNA strands and continuous elongation enables the possibility of sufficient genetic information.

The detachment events observed in these experiments likely occurred due to an increase of temperature in the CHARMM simulations, emulating temperature cycles in the natural environment. While the detachment observed in this case was not a direct result of a breakdown of the VDW bond at the surface of a carbon-based template, it points to the possibility that a solution of free oligomers, due to increased heat, is possible. Thus, in a higher heat environment where a high volume of oligomers are forming, then a hyper-saturated solution of oligonucleotides could have existed within aqueous conditions such as a "warm little pond".

The movement of oligonucleotides, including lateral translocation across the surface of graphite and graphene, demonstrated the mag-lev-like mobility of these molecules, enabling interactions such as rapid rotation and docking with other molecules, in various orientations. Also, with sufficient simulation time, it is possible that hydrogen bonds may form double stranded configurations of the oligomers, with additional simulation time.

It is thought that the VDW attachment observed in these experiments can provide sufficiently robust anchoring of RNA oligomers, enabling polymer growth, yet at the same time permitting rapid detachment at a later point. Under the right cyclical conditions, this could enable the production of a locally concentrated pool of RNA, which can then be incorporated into proto-cellular lifeforms.

The simulations conducted in this research were simplified, using relatively ideal conditions, and would not necessarily reflect the conditions in a real world environment. It is clear that much more work needs to be done to create a larger sample size of longer, more complex simulations, in order to determine the likelihood of the various endpoint occurrences over time. Furthermore, it is known that hydrogen and covalent bonds can take a considerably longer time to form, in such molecular simulations. As such, with

proper funding, or further improvements in the time efficiency through the use of novel AI-enhanced methodologies, longer duration simulations could be conducted. It is important that additional evidence be gathered, to confirm the repeatability of these endpoints at scale. Furthermore, additional work must be done to ascertain the probabilities that these events could occur under a range of starting conditions, similar to those thought to have existed billions of years ago, in consideration of the various Earth-based and extraterrestrial abiogenesis theories. In addition, as mentioned above, physical experiments are being conducted in addition to a new set of simulations, to confirm these results and develop a larger endpoint dataset,

References

1. England, J. L. (2013). Statistical physics of self-replication. *The Journal of Chemical Physics*, *139*(12), 121923. https://doi.org/10.1063/1.4818538

2. Ilya Prigogine (1975). Dissipative structures, dynamics and entropy. *The Quantum Chemistry, 19/25 January 1975* https://onlinelibrary.wiley.com/doi/10.1002/qua.560090854

3. Rothstein, D. M. (2012, May 1). Solid-Phase supports for oligo synthesis. *GEN - Genetic Engineering and Biotechnology News*. https://www.genengnews.com/magazine/solid-phase-supports-for-oligo-synthesis/

4. Swadling, J. B. (2010, September 15). *Clay minerals mediate folding and regioselective interactions of RNA: A large-scale atomistic simulation study*. ACS Publications. https://pubs.acs.org/doi/abs/10.1021/ja104106y

5. S. Jo, T. Kim, V.G. Iyer, and W. Im (2008) CHARMM-GUI: A Web-based Graphical User Interface for CHARMM. J. Comput. Chem. 29:1859-1865

6. B.R. Brooks, C.L. Brooks III, A.D. MacKerell, Jr., L. Nilsson, R.J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A.R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R.W. Pastor, C.B. Post, J.Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D.M. York, and M. Karplus (2009). CHARMM: The Biomolecular Simulation Program. J. Comput. Chem. 30:1545-1614

7. J. Lee, X. Cheng, J.M. Swails, M.S. Yeom, P.K. Eastman, J.A. Lemkul, S. Wei, J. Buckner, J.C. Jeong, Y. Qi, S. Jo, V.S. Pande, D.A. Case, C.L. Brooks III, A.D. MacKerell Jr, J.B. Klauda, and W. Im (2016). CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations using the CHARMM36 Additive Force Field. J. Chem. Theory Comput. 12:405-413

8. S. Jo, X. Cheng, S.M. Islam, L. Huang, H. Rui, A. Zhu, H.S. Lee, Y. Qi, W. Han, K. Vanommeslaeghe, A.D. MacKerell, Jr., B. Roux, and W. Im (2014) CHARMM-GUI PDB Manipulator for Advanced Modeling and Simulations of Proteins Containing Non-standard Residues. Adv. Protein Chem. Struct. Biol. 96:235-265

9. Y.K. Choi, N.R. Kern, S. Kim, K. Kanhaiya, S.H. Jeon, Y. Afshar, S. Jo, B.R. Brooks, J. Lee, E.B. Tadmor, H. Heinz, and W. Im (2022) CHARMM-GUI Nanomaterial Modeler for Modeling and Simulation of Nanomaterial Systems. J. Chem. Theory Comput. in press

10. Dhindsa, G. K., Bhowmik, D., Goswami, M., O'Neill, H., Mamontov, E., Sumpter, B. G., Hong, L., Ganesh, P., & Chu, X. (2016). Enhanced Dynamics of Hydrated tRNA on Nanodiamond Surfaces: A Combined Neutron Scattering and MD Simulation Study. *The Journal of Physical Chemistry B*, *120*(38), 10059–10068. https://doi.org/10.1021/acs.jpcb.6b07511

11. Ferris, J. P. (2006). Montmorillonite-catalysed formation of RNA oligomers: The possible role of catalysis in the origins of life. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *361*(1474), 1777–1786. https://doi.org/10.1098/rstb.2006.1903

12. Greathouse, Jeffery & Johnson, Karen & Greenwell, Chris. (2014). Interaction of Natural Organic Matter with Layered Minerals: Recent Developments in Computational Methods at the Nanoscale. Minerals. 4. 519-540. 10.3390/min4020519.

13. Kopidakis, G., Remediakis, I. N., Fyta, M. G., & Kelires, P. C. (2007). Atomic and electronic structure of crystalline–amorphous carbon interfaces. *Diamond and Related Materials*, *16*(10), 1875–1881. https://doi.org/10.1016/j.diamond.2007.07.013

14. Olawanle, J. (2022, October 5). Big O cheat sheet – time complexity chart. *freeCodeCamp.Org*. https://www.freecodecamp.org/news/big-o-cheat-sheet-time-complexity-chart/

15. Ren, H., Wang, J., Zhao, W. X., & Wu, N. (2021, August 14). RAPT: Pre-training of time-aware transformer for learning robust healthcare representation. *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. http://dx.doi.org/10.1145/3447548.3467069

16. Seif, A., Hafezi, M., & Jarzynski, C. (2020). Machine learning the thermodynamic arrow of time. *Nature Physics*, *17*(1), 105–113. https://doi.org/10.1038/s41567-020-1018-2

17. Sergei K. Simakov, Vittorio Scribano, Nikolai N. Mel'Nik, Germana Barone. Sicilian serpentinite xenoliths containing abiotic organics with nanodiamond clusters as key model for prebiotic processes[J]. Geoscience Frontiers, 2021, 12(1): 255-263. DOI: 10.1016/j.gsf.2020.04.008

18. Sitapure, N., & Kwon, J. S.-I. (2023). CrystalGPT: Enhancing system-to-system transferability in crystallization prediction and control using time-series-transformers. *Computers & Chemical Engineering*, *177*, 108339. https://doi.org/10.1016/j.compchemeng.2023.108339

19. Totani, T. Emergence of life in an inflationary universe. Sci Rep 10, 1671 (2020). https://doi.org/10.1038/s41598-020-58060-0

20. Yu, H., Zhong, Y., Ji, J., Gong, X., & Xiang, H. (2022). *Time-reversal equivariant neural network potential and Hamiltonian for magnetic materials*. American Chemical Society (ACS). http://dx.doi.org/10.26434/chemrxiv-2022-h6f69

21. Zhang, S., Murray, N., Wang, L., & Koniusz, P. (2022). Time-rEversed DiffusioN tEnsor Transformer: A New TENET of Few-Shot Object Detection. In *Lecture Notes in Computer Science* (pp. 310–328). Springer Nature Switzerland. http://dx.doi.org/10.1007/978-3-031-20044-1_18

22.

Data files

Supplemental data tables and raw DCD (trajectory) files are linked in attachment. These files can be visualized and analyzed in the freely available software tool known as VMD. Note that these files are extremely large.