




E2EDNA 2.0: Python Pipeline for Simulating DNA Aptamers with Ligands

Michael Kilgour^{*1}, Tao Liu^{†1}, Ilya S. Dementyev¹, and Lena Simine^{‡1}

¹ Department of Chemistry, McGill University, Montreal, Quebec, Canada

DOI: [10.21105/joss.04182](https://doi.org/10.21105/joss.04182)

Software

- [Review](#) 
- [Repository](#) 
- [Archive](#) 

Editor: [Charlotte Soneson](#) 

Reviewers:

- [@JoaoRodrigues](#)
- [@rvhonorato](#)

Submitted: 31 January 2022

Published: 18 February 2022

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).

Summary

DNA aptamers are short sequences of single-stranded DNA with untapped potential in molecular medicine, drug design, and materials design due to their strong and selective and most importantly tunable binding affinity to target molecules (O. Tucker et al., 2012; Zhou & Rossi, 2017). For instance, DNA aptamers can be used as therapeutics (Corey et al., 2021) for a wide range of diseases such as epilepsy (Zamay et al., 2020) and cancer (Morita et al., 2018). They can also be used to detect a wide variety of molecular ligands, including antibiotics (Mehlhorn et al., 2018), neurotransmitters (Sinha & Das Mukhopadhyay, 2020), metals (Qu et al., 2016), proteins (Kirby et al., 2004), nucleotides (Shen et al., 2007) and metabolites (Dale, 2021; Dauphin-Ducharme et al., 2022) in real time, even in harsh environments (McConnell et al., 2020).

We present E2EDNA 2.0: End-2-End DNA 2.0, a python simulation pipeline which offers a unified and automated solution to computational modeling of DNA aptamers with molecular ligands. It is broadly aimed at researchers developing therapeutics and sensors based on DNA aptamers who require detailed atomistic information on the behavior of aptamers and ligands in realistic media. Similar to its predecessor E2EDNA (Kilgour et al., 2021), E2EDNA 2.0 predicts DNA aptamers' secondary and tertiary structures, and if a ligand is present, the configuration of the solvated aptamer-ligand complex.

Statement of Need

With E2EDNA 2.0 our goal is to create an all-python simulation package for single-stranded DNA with small ligands that is easy to install and use in python-based workflows. The pipeline is automated, yet flexible, taking us from DNA aptamer sequence to folded aptamer, and aptamer-ligand complex. Currently, available softwares designed for computationally studying DNA aptamers predominantly focus on partial feature analyses of RNA and DNA aptamers. For example, APTANI (J. Caroli et al., 2016) and APTANI2 (Jimmy Caroli et al., 2020), commonly used by the aptamer community, select potentially relevant aptamers from SELEX (Systematic Evolution of Ligands by EXponential enrichment)(McKeague & DeRosa, n.d.; Tuerk & Gold, 1990) experimental data sets through a sequence-structure analysis, and AEGIS, a platform equipped with a generative deep learning model to propose novel aptamer sequences (Biondi & Benner, 2018). These approaches are aimed at fast black-box analysis of large numbers of sequences. A detailed analysis of a small number of high-promise candidate sequences is often required but no automated and easy-to-use simulation package exists in the computational space aside from E2EDNA (Kilgour et al., 2021) and E2EDNA 2.0 with

^{*}co-first author

[†]co-first author

[‡]corresponding author

the latter preferable for ease of installation and user-friendly implemented eight modes of simulation.

The gap that E2EDNA family of programs addresses is the absence of a non-black-box one-stop-shop aptamer simulation packages, which are capable of providing in silico predictions of 2D structure, 3D structure, and aptamer-ligand binding simulation while keeping methodology flexible and analysis open-ended. E2EDNA 2.0 achieves this in two key ways: an all-python implementation which makes installation and access easier for users than the original E2EDNA package, and the automation of the dozens of tasks required in setting up, running, and interpreting atomistic aptamer-ligand simulations.

Figure

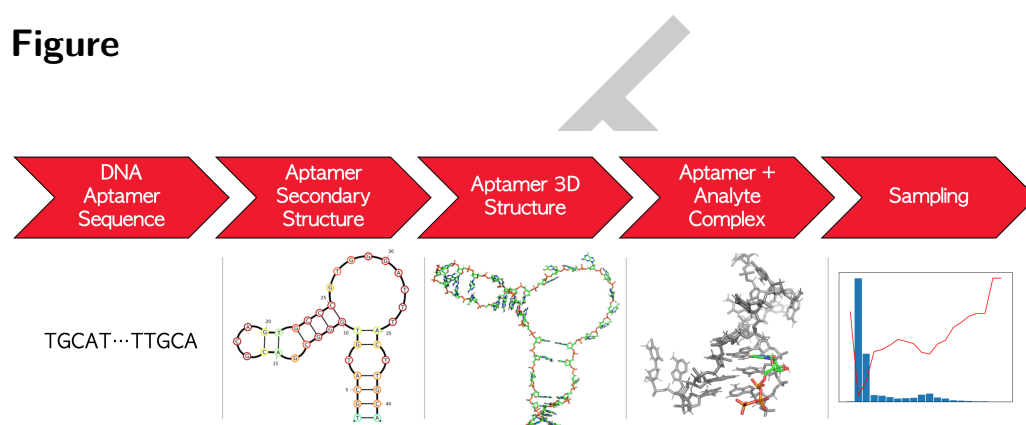


Figure 1: Schematic workflow of E2EDNA 2.0 pipeline.

Components and Features

As shown in Figure 1, the complete simulation pipeline in E2EDNA 2.0 consists of the following main steps: 1. Secondary structure prediction, 2. Tertiary structure prediction, 3. Molecular dynamics simulation, and 4. Aptamer-ligand docking. The key inputs into E2EDNA 2.0 are the DNA sequence in the FASTA format, the structure of the ligand in the pdb format (optional), and the choice of simulation mode. The output includes secondary structure in dot-bracket notation, tertiary structures in pdb format of free aptamer and aptamer-ligand complex (optional), and simulation trajectories in the dcd format. It is worth noting that there are other parameters besides “key inputs” that can be customized in the main.py, such as solvent ionic strength. Detailed analysis of the generated trajectories is not performed by E2EDNA 2.0, though is straightforward to set-up for a particular workflow using built-in or user-specified functions.

Next we briefly discuss the external software packages engaged in the E2EDNA 2.0 pipeline. For developers we point out that the pipeline is modular and any of these packages may be drop-in replaced with any equivalent or competing software. The first module, NUPACK (Zadeh et al., 2011), generates a predicted secondary structure given DNA FASTA sequence, temperature, and ionic strength. It can output explicit probability of observing the most likely secondary structure for a given sequence, as well as suboptimal structures and their probabilities. The second module, MarcoMoleculeBuilder (MMB) (Flores et al., 2011) is a multifunctional software from simTK which allows for rapid directed folding of oligonucleotides and peptides via straightforward inputs on various platforms. MMB initializes a given ssDNA sequence in a single-helix configuration, and folds it according to user-specified base-pairing conditions via simulation with fictitious forces which pull the respective bases together. We include in the software scripts which automatically take in secondary structure instructions

from NUPACK in the form of a list of paired bases, and generate MMB command files accordingly. The MMB outputs, as initial structures, are then passed to MD simulation for relaxation. The third module, OpenMM (Eastman et al., 2017), is the molecular dynamics engine powering E2EDNA 2.0. OpenMM is used to sample 3D structures of both the ‘free’ aptamer and its aptamer-target complex. The representative DNA aptamer structure is chosen from the MD trajectory via a principle component analysis on backbone dihedrals: a free energy is constructed in the space of top-ranked principal components (5) as reaction coordinates and the lowest free energy structure is passed on to the docking with ligand step. The fourth module, LightDock (Jiménez-García et al., 2018; Roel-Touris et al., 2020), automates the docking between a free DNA aptamer and a given target ligand. LightDock uses a glowworm swarm algorithm; we compute the number of swarms required as being proportional to the approximate surface area of the aptamer and use the best-scored glowworm as the docked complex structure.

We conclude the presentation of E2EDNA 2.0 by listing the available simulation modes. We refer the reader to Documentation for more details on the modes, as well as for installation and running the simulation instructions.

1. Mode: ‘2d structure’; Description: predicts secondary structure.
2. Mode: ‘3d coarse’; Description: ‘2d structure’ + quick folding into a 3D structure.
3. Mode: ‘3d smooth’; Description: ‘3d coarse’ + MD relaxation.
4. Mode: ‘coarse dock’; Description: ‘3d coarse’ structure + molecular docking.
5. Mode: ‘smooth dock’; Description: ‘3d smooth’ structure + molecular docking.
6. Mode: ‘free aptamer’; Description: ‘3d coarse’ + MD sampling search for representative aptamer folded conformation.
7. Mode: ‘full dock’; Description: ‘free aptamer’ + molecular docking.
8. Mode: ‘full binding’; Description: ‘full dock’ + MD sampling of aptamer-ligand complex.

Acknowledgements

Funding from the NSERC Discovery grant RGPIN-201924734 and an NSERC PDF for M.K. is greatly appreciated. Computations were made on the supercomputer Beluga, managed by Calcul Quebec (<https://www.calculquebec.ca/>) and Compute Canada (<https://www.computecanada.ca/>). The operation of this supercomputer is funded by the Canada Foundation for Innovation (CFI).

References

- Biondi, E., & Benner, S. (2018). Artificially Expanded Genetic Information Systems for New Aptamer Technologies. *Biomedicine*, 6(2), 53. <https://doi.org/10.3390/biomedicine6020053>
- Caroli, Jimmy, Forcato, M., & Bicciato, S. (2020). APTANI2: Update of aptamer selection through sequence-structure analysis. *Bioinformatics*, 36(7), 2266–2268. <https://doi.org/10.1093/bioinformatics/btz897>

- 110 Caroli, J., Taccioli, C., De La Fuente, A., Serafini, P., & Bicciato, S. (2016). APTANI: A
111 computational tool to select aptamers through sequence-structure motif analysis of HT-
112 SELEX data. *Bioinformatics*, 32(2), 161–164. [https://doi.org/10.1093/bioinformatics/
113 btv545](https://doi.org/10.1093/bioinformatics/btv545)
- 114 Corey, D. R., Damha, M. J., & Manoharan, M. (2021). Challenges and Opportunities for
115 Nucleic Acid Therapeutics. *Nucleic Acid Therapeutics*, nat.2021.0085. [https://doi.org/
116 10.1089/nat.2021.0085](https://doi.org/10.1089/nat.2021.0085)
- 117 Dale, N. (2021). Real-time measurement of adenosine and ATP release in the central
118 nervous system. *Purinergic Signalling*, 17(1), 109–115. [https://doi.org/10.1007/
119 s11302-020-09733-y](https://doi.org/10.1007/s11302-020-09733-y)
- 120 Dauphin-Ducharme, P., Ploense, K. L., Arroyo-Currás, N., Kippin, T. E., & Plaxco, K. W.
121 (2022). Electrochemical Aptamer-Based Sensors: A Platform Approach to High-Frequency
122 Molecular Monitoring In Situ in the Living Body. In *Biomedical Engineering Technologies*
123 (Vol. 2393, pp. 479–492). Springer US. https://doi.org/10.1007/978-1-0716-1803-5_25
- 124 Eastman, P., Swails, J., Chodera, J. D., McGibbon, R. T., Zhao, Y., Beauchamp, K. A.,
125 Wang, L.-P., Simmonett, A. C., Harrigan, M. P., Stern, C. D., Wiewiora, R. P., Brooks,
126 B. R., & Pande, V. S. (2017). OpenMM 7: Rapid development of high performance
127 algorithms for molecular dynamics. *PLOS Computational Biology*, 13(7), e1005659. <https://doi.org/10.1371/journal.pcbi.1005659>
128
- 129 Flores, S. C., Sherman, M. A., Bruns, C. M., Eastman, P., & Altman, R. B. (2011). Fast Flex-
130 ible Modeling of RNA Structure Using Internal Coordinates. *IEEE/ACM Transactions on*
131 *Computational Biology and Bioinformatics*, 8(5), 1247–1257. [https://doi.org/10.1109/
132 TCBB.2010.104](https://doi.org/10.1109/TCBB.2010.104)
- 133 Jiménez-García, B., Roel-Touris, J., Romero-Durana, M., Vidal, M., Jiménez-González, D.,
134 & Fernández-Recio, J. (2018). LightDock: A new multi-scale approach to protein–protein
135 docking. *Bioinformatics*, 34(1), 49–55. <https://doi.org/10.1093/bioinformatics/btx555>
- 136 Kilgour, M., Liu, T., Walker, B. D., Ren, P., & Simine, L. (2021). E2EDNA: Simulation
137 Protocol for DNA Aptamers with Ligands. *Journal of Chemical Information and Modeling*,
138 61(9), 4139–4144. <https://doi.org/10.1021/acs.jcim.1c00696>
- 139 Kirby, R., Cho, E. J., Gehrke, B., Bayer, T., Park, Y. S., Neikirk, D. P., McDevitt, J.
140 T., & Ellington, A. D. (2004). Aptamer-Based Sensor Arrays for the Detection and
141 Quantitation of Proteins. *Analytical Chemistry*, 76(14), 4066–4075. [https://doi.org/10.
142 1021/ac049858n](https://doi.org/10.1021/ac049858n)
- 143 McConnell, E. M., Nguyen, J., & Li, Y. (2020). Aptamer-Based Biosensors for Environmental
144 Monitoring. *Frontiers in Chemistry*, 8, 434. <https://doi.org/10.3389/fchem.2020.00434>
- 145 McKeague, M., & DeRosa, M. C. (n.d.). *Aptamers and SELEX: Tools for the Development*
146 *of Transformative Molecular Recognition Technology*. 5.
- 147 Mehlhorn, A., Rahimi, P., & Joseph, Y. (2018). Aptamer-Based Biosensors for Antibiotic
148 Detection: A Review. *Biosensors*, 8(2), 54. <https://doi.org/10.3390/bios8020054>
- 149 Morita, Y., Leslie, M., Kameyama, H., Volk, D., & Tanaka, T. (2018). Aptamer Thera-
150 peutics in Cancer: Current and Future. *Cancers*, 10(3), 80. [https://doi.org/10.3390/
151 cancers10030080](https://doi.org/10.3390/cancers10030080)
- 152 O. Tucker, W., T. Shum, K., & A. Tanner, J. (2012). G-quadruplex DNA Aptamers and their
153 Ligands: Structure, Function and Application. *Current Pharmaceutical Design*, 18(14),
154 2014–2026. <https://doi.org/10.2174/138161212799958477>
- 155 Qu, H., Csordas, A. T., Wang, J., Oh, S. S., Eisenstein, M. S., & Soh, H. T. (2016). Rapid
156 and Label-Free Strategy to Isolate Aptamers for Metal Ions. *ACS Nano*, 10(8), 7558–7565.
157 <https://doi.org/10.1021/acs.nano.6b02558>

- 158 Roel-Touris, J., Bonvin, A. M. J. J., & Jiménez-García, B. (2020). LightDock goes
159 information-driven. *Bioinformatics*, 36(3), 950–952. [https://doi.org/10.1093/](https://doi.org/10.1093/bioinformatics/btz642)
160 [bioinformatics/btz642](https://doi.org/10.1093/bioinformatics/btz642)
- 161 Shen, L., Chen, Z., Li, Y., Jing, P., Xie, S., He, S., He, P., & Shao, Y. (2007). A chrono-
162 coulometric aptamer sensor for adenosine monophosphate. *Chemical Communications*,
163 21, 2169. <https://doi.org/10.1039/b618909a>
- 164 Sinha, K., & Das Mukhopadhyay, C. (2020). Quantitative detection of neurotransmitter
165 using aptamer: From diagnosis to therapeutics. *Journal of Biosciences*, 45(1), 44. <https://doi.org/10.1007/s12038-020-0017-x>
166 <https://doi.org/10.1007/s12038-020-0017-x>
- 167 Tuerk, C., & Gold, L. (1990). Systematic Evolution of Ligands by Exponential Enrichment:
168 RNA Ligands to Bacteriophage T4 DNA Polymerase. *Science*, 249(4968), 505–510. <https://doi.org/10.1126/science.2200121>
169 <https://doi.org/10.1126/science.2200121>
- 170 Zadeh, J. N., Steenberg, C. D., Bois, J. S., Wolfe, B. R., Pierce, M. B., Khan, A. R., Dirks,
171 R. M., & Pierce, N. A. (2011). NUPACK: Analysis and design of nucleic acid systems.
172 *Journal of Computational Chemistry*, 32(1), 170–173. <https://doi.org/10.1002/jcc.21596>
- 173 Zamay, T. N., Zamay, G. S., Shnayder, N. A., Dmitrenko, D. V., Zamay, S. S., Yushchenko, V.,
174 Kolovskaya, O. S., Susevski, V., Berezovski, M. V., & Kichkailo, A. S. (2020). Nucleic Acid
175 Aptamers for Molecular Therapy of Epilepsy and Blood-Brain Barrier Damages. *Molecular*
176 *Therapy - Nucleic Acids*, 19, 157–167. <https://doi.org/10.1016/j.omtn.2019.10.042>
- 177 Zhou, J., & Rossi, J. (2017). Aptamers as targeted therapeutics: Current potential and
178 challenges. *Nature Reviews Drug Discovery*, 16(3), 181–202. [https://doi.org/10.1038/](https://doi.org/10.1038/nrd.2016.199)
179 [nrd.2016.199](https://doi.org/10.1038/nrd.2016.199)