

IN SILICO BINDING OF A NOVEL PURINE-BASED TRIMER

Madelyn Tarvin* and Roslyn Lampkins†

Department of Chemistry, University of Evansville, Evansville, IN 47722

Abstract

Regulation of specific biological processes, such as DNA replication or transcription, occurs through the binding of small molecules or proteins to DNA. Distamycin and mithramycin are examples of small molecules that bind to A-/T-rich and G-/C-rich sections of the minor groove of DNA, respectively. Similarly, we have designed a novel purine-based trimer to bind to the minor groove of DNA, spanning a distance of six base pairs, specifically targeting C-/G- rich sequences. This project computationally investigates the binding between our trimer and the CGG trinucleotide repeat section of the FMR1 gene, using ligand receptor docking programs through Schrödinger's Small Molecule Suite. We have successfully designed procedures for using Glide Docking and Quantum-Polarized Ligand Docking (QPLD) to dock our ligand in its dimeric form to a small section of DNA.

*Corresponding author: rl106@evansville.edu

†Undergraduate researcher and co-author

Keywords: Computational chemistry, ligand-DNA interactions, molecular docking, quantum mechanics, molecular mechanics

Received: August 19, 2026 Accepted: August 27, 2026 Revision received: September 1, 2026 Published: September 16, 2026

Introduction

Many molecules bind to DNA in order to regulate specific biological processes, such as DNA replication, transcription, etc. A classic example of DNA-binding proteins are transcription factors, which block or promote other transcriptional proteins¹. Additionally, distamycin is a small molecule that binds to the minor groove of DNA. Specifically, distamycin binds to A- and T-rich sequences to block transcription factors and prevent specific viruses from replicating their DNA². Another known DNA binder is mithramycin, which binds to C- and G-rich sequences, although less is known about its binding properties³. In order to analyze a molecule's binding abilities, computational approaches can be used. With computational molecular docking simulations, the binding affinity of a ligand for a specific receptor can be estimated for a variety of generated conformations of the ligand. Molecular docking simulations are also often used in drug discovery to model how a drug may bind to its target receptor protein. This allows researchers to determine the capabilities of drug candidates to find which molecules bind to the designated receptor with higher affinity⁴. This helps reduce the time and cost of synthesizing all possible candidates by providing them with mathematical-based estimations of each candidate's binding capabilities.

Schrödinger is a physics-based computational platform used for chemical modeling and molecular docking in many fields. Within Schrödinger is Glide, a molecular mechanics docking program, that allows for ligand-receptor docking, virtual screening, binding mode prediction, and 3D molecular design. Effective docking requires preparation of both the receptor and the ligand. The protein structure is first refined using Protein Preparation Workflow to ensure proper protonation states, resolve structural inconsistencies, and optimize the hydrogen bonding network¹⁷. A receptor grid is then defined to specify where the ligand will bind on the protein, followed by ligand preparation with LigPrep to generate relevant states and conformations within the given information¹⁵. After the necessary preparation, Ligand Docking is performed to evaluate potential binding poses, which are ranked by GlideScores and Docking scores that estimate binding affinity^{5,13}. The GlideScore ranks poses based on contributions from different energies, while the Docking score also takes into account

Epik state penalties^{5,12}. Glide docking uses molecular mechanics (MM) which assumes non-flexible proteins during ligand binding and fixed dielectric charges of both the ligand and the receptor⁶. While this is a useful estimation for many systems, it can lead to less accurate results due to its failure to account for charge transfer and the flexibility of the receptor.

Quantum mechanical (QM) methods can help improve results because they account for polarization, charge transfer, and more phenomena that cannot be described by classical mechanics⁶. The primary limitation to applying QM calculations to molecular docking is the computational power and time required to complete the calculations. To address this, hybrid quantum mechanics/molecular mechanics (QM/MM) approaches are often utilized, enabling improved accuracy while maintaining more practical computational demands. In QM/MM approaches, the ligand is often treated with quantum mechanical theories while the protein is treated with molecular mechanics. The main QM/MM program in Schrödinger is QSite; however, this is protein specific and cannot be used with non-amino acid based receptors¹⁸. Another approach to QM/MM in Schrödinger is to use Jaguar to optimize the ligand with QM and then dock the QM-optimized ligand to the receptor using Glide docking¹⁴. Most of these docking methods have been thoroughly studied with typical protein receptors, but there is significantly less research on methods to dock a ligand to DNA in Schrödinger.

We have designed a purine-based ligand to bind to C- and G-rich sections of the minor groove of DNA, spanning 6 base pairs (Figure 1A). The binding of this ligand to DNA, specifically a section of the FMR1 gene due to its high C/G content, was investigated through both MM and QM/MM ligand docking approaches (Figure 1C). Since ligand docking is primarily used to dock a ligand to a protein, this research identifies a method of ligand docking to a DNA receptor, using both Glide and Jaguar QM optimization. After extensive computational analysis, we have found that QM/MM approaches result in more favorable binding affinities relative to the MM Docking scores.

Methods

Building DNA in Schrödinger's Maestro. From the task dropdown,

the task “build biopolymer from sequence” was used to create a double-stranded section of DNA. In this task, “DNA” was selected as the biopolymer type and the sequence was given a title. For the nucleotide sequence, a section of the FMR1 gene, containing 19 CGG trinucleotide repeats and a total of 59 nucleotides, was used (Figure 1B and C). The DNA strand was built double-stranded by checking the “build double stranded” box in the Build Biopolymer from Sequence window.

Building a ligand. The novel trimer used for this research was built using the 2D Sketcher tool (Figure 1A).

Dimerization of ligand. A dimer of the novel ligand was created using the following procedure. First, the original ligand was duplicated in the project table. The atoms of the duplicated structure were highlighted and, using the “Build” panel, moved to be antiparallel to the original ligand structure. After the duplicated structure was moved, both ligands were included in the workspace, and the duplicated ligand was moved as necessary to ensure that it was antiparallel and not directly overlapped with the original ligand. Once the ligands are in a preferable position, the workspace entries for the two ligands were merged into one new workspace entry. To make this a metal ion-coordinated dimer, a metal center was added to the dimer entry from the “Build” window. Using the “Draw” tool within the “Build” window, eight bonds were drawn from the ligand to the metal center—four bonds from each monomer. The angles and specifics of bonds do not matter here, as that will be corrected with LigPrep¹⁵ in the following paragraph. Dimerization using zero order bonds was also done instead of adding a metal center, in attempts to have more favorable dimer conformations. However, using zero order bonds did not result in any favorable docking scores.

Glide docking. The process of ligand docking includes four main

parts: ligand preparation, protein preparation, Glide grid creation, and Glide docking. The ligand to be docked to the DNA is prepared using the LigPrep tool found in the task bar¹⁵. The settings used are as follows: pH of 7.4 +/- 1.0 (see note in results section) and all combinations were generated. Other settings were left as the default. Multiple prepared ligands were analyzed for comparison. Next, the protein was prepared using the Protein Preparation Workflow¹⁷. The protein used for this research project is the double-stranded section of the FMR1 gene that was built in the first paragraph of this section. In the Protein Preparation Workflow window, the protein is prepared using the default settings¹⁷. Then, a glide grid on the prepared protein was created using the Receptor Grid Generation tool. In the Receptor Grid Generation window under the “receptor” tab, the “pick to identify the ligand” option was unchecked since the receptor in the workspace did not include a ligand. Under the “site” tab, the “center of selected residues” option was selected, and the specific residues were manually selected or the coordinates were specified to center the ligand diameter midpoint box over the minor groove. All residues selected were in the minor groove, but not all grids had the same selected residues (Table 1). The size of the enclosing box was kept at 20 Å, but the size of the ligand diameter midpoint box was changed for each grid (Table 1). All other settings were kept as default. All grids used the same prepared protein. To dock the ligand using Glide, the specific receptor grid file was used, and the ligands were included from the project table¹³. All settings were kept as default, except the number of poses per ligand that were included in the output. Docking scores were generated for each ligand docking calculation that was done, and the results are included in the right-most column of Table 1.

Quantum Polarized Ligand Docking. In order to obtain more optimal docking scores, quantum polarized ligand docking (QPLD) was employed, treating the ligand with QM models and the receptor with MM models. QM optimization of the ligand was performed using Schroödinger’s Jaguar program¹⁴. The QM theories used were either M06-2X or B3LYP-D3 along with density functional theory (DFT) calculations (Table 2). The optimizations were run for up to 300 steps with loose or normal convergence, using the Poisson-Boltzmann finite (PBF) solvation model in water to reflect biologically relevant conditions. Once the ligand was optimized, it was docked to the same grids as described previously using Glide, inputting the partial charges of the optimized ligand. Some ligands that yielded the most favorable results were optimized a second or third time with Jaguar to give more favorable ligand conformations¹⁴. Results are summarized in Table 2.

Controls. These methods were also repeated with distamycin, a well-known DNA minor-groove binder (PDB 378D). The structure of distamycin was input into Maestro via the 2D sketcher tool and then prepared with LigPrep¹⁵. An A-/T-rich section of DNA was also built and prepared with Protein Preparation Workflow¹⁷. A 15 Å ligand diameter midpoint box was used to design the receptor grid and centered over the minor groove. The prepared distamycin structure was docked using both Glide docking and QPLD. Additionally, the structure of distamycin bound to a section of DNA was imported from the Protein Data Bank (PDB 378D), and a docking score was calculated for this structure using Glide docking to serve as another point of comparison for the proposed docking method.

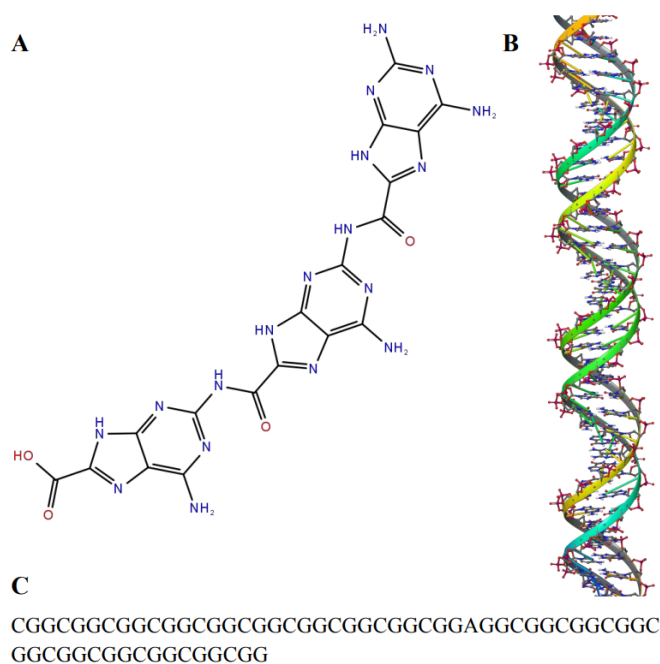


Figure 1. Summary of background information for structures and sequences used in the present research. (A) Structure of the designed minor-groove binder. (B) Section of the FMR1 gene built in Maestro. (C) Nucleotide sequence of the section of FMR1 gene used in this study. NCBI accession number NC_000023.11:g.133_192.

Results

Ligand creation. Since there is no current experimental data on this ligand, multiple ligand dimers were made and docked to the FMR1 gene to see which resulted in the best score. Furthermore, to make the ligand into a dimer, multiple techniques were attempted: four bonds to the metal center instead of eight, making the bonds from different atoms, etc. While it does not affect the computational studies, our current hypothesis is that this dimer would be formed via a magnesium ion; however, there is no current data on this. Dimers can also be created with zero order bonds using the Build tool. Zero order bonds are a computational model to represent a noncovalent interaction between molecules in an atom⁷. However, with this method, the ligands resulted in no favorable conformations or docking scores. It is also important to note that the pH range used in the LigPrep settings was specified as 7.4 +/- 1.0 in the methods section above¹⁵. In an attempt to have more biological relevance, LigPrep was also employed using a pH range of 7.4 +/- 0.5; however, this resulted in no usable results¹⁵. The ligand with the best conformation yielded the most favorable docking scores and was used for the rest of the ligand docking simulations.

Receptor grid generation and glide docking. During the receptor grid generation process, multiple residues and sizes of ligand diameter midpoint boxes were attempted. The results are summarized in Table 1. The four grids that gave promising results are 3, 8, 9, and 11. Each of these grids was generated in a different way. Grid 3 was created using specific residues and a 15 Å ligand diameter midpoint box. Grid 8 used the coordinates $x=15$, $y=3.21$, $z=60.19$ and a 15 Å ligand diameter midpoint box. Grid 9 used the same coordinates as grid 8, but the ligand diameter midpoint box was 15x15x20 Å. Grid 11 was also made using specified coordinates and a 15x15x15 Å ligand diameter midpoint box. There was one ligand that led to relatively favorable docking scores with each of these grids. To determine which type of grid worked best, each of these favorable grids were duplicated with the exact same settings, and used to dock each ligand. It was determined that a 15x15x15 Å ligand diameter midpoint box was the most ideal size. To get the ideal placement, the box was centered over the minor groove using the coordinates rather than selecting specific residues. Since some of the ligands still docked to the major groove, this may indicate an issue with the receptor grid, but grid 11 docked the ligand to the minor groove a majority of the time, indicating that it is likely the best grid design.

Each glide dock was run with the same settings, the only difference being the grid and the ligands used. While the resulting docking scores were not strongly favorable (Table 1, Figure 2), these values should be interpreted in the context of known limitations of the computational model. In the absence of experimental data for this ligand-DNA system, accurately defining the ligand's position in the minor groove remains challenging, which may contribute to the less favorable binding energies. Additionally, the docking protocol did not include explicit water molecules within the minor groove. Given their established role in mediating ligand binding in the minor groove through hydrogen bonding networks and structural bridging, the incorporation of a solvation model in future work could further refine these binding energy estimates^{8,9}. Furthermore, the use of a molecular mechanics framework in

Glide, which treats the molecules with fixed charges, further limits its ability to capture effects such as charge polarization. To address these constraints and improve representation of ligand-DNA interactions, QM approaches were also employed.

Quantum Polarized Ligand Docking. The first attempt at QPLD was through QSite; however, QSite is protein specific, so it did not yield any results¹⁸. Jaguar optimization of the ligand must be used, and then the ligand must be docked back to the DNA receptor using Glide^{13,14}. Both B3LYP-D3 and M06-2X theories were used, but M06-2X provided the most optimal results (Table 2). To achieve successful convergence during geometry optimization, a maximum of 300 optimization steps and loose convergence criteria were employed. The need for these settings likely reflects the size and complexity of the ligand, as optimization did not converge within the default iterations. PBF solvation was also

Table 1. Summary of Glide grid performance in Glide Docking simulations

Grid	Center of Grid	Ligand Diameter Midpoint Box*	Enclosing Box*	Docking Score
1	Selected residues	15 Å	20 Å	-4.193
2	Selected residues	20 Å	20 Å	N/A**
3	Selected residues	15 Å	20 Å	-6.043
4	Selected residues	10 Å	20 Å	N/A**
5	Selected residues	10 Å	20 Å	N/A**
6	Selected residues	20 Å	20 Å	N/A**
7	Selected residues	15 Å	20 Å	N/A**
8	Used coordinates to center box over minor groove	15 Å	20 Å	-5.194 -4.677
9	Used coordinates to center box over minor groove	15x15x20 Å	20 Å	-4.342
10	Used coordinates to center box over minor groove	15 Å	20 Å	N/A**
11	Selected residues	15 Å	20 Å	-5.718

*this number refers to the size of each side of the box (i.e. 20x20x20 Å) unless all three dimensions are listed

**N/A refers to results that were not usable due to binding in the major groove, poor docking score, etc.

used to create more biologically relevant conditions. The default

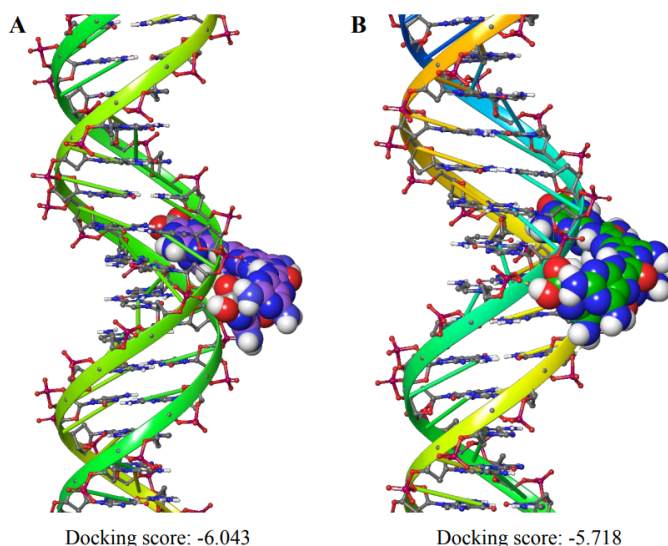


Figure 2. Top 2 Glide docking poses. (A) Image of ligand docking with grid 3 that yielded a docking score of -6.043. (B) Image of ligand docking with grid 11 that yielded a docking score of -5.718.11.

without adding a solvation theory is to treat the ligand in the gas phase which is not an accurate representation. PBF is an implicit solvent technique, so it treats the ligand in solvation, but does not take into account any interactions with water molecules that may alter the conformation of the ligand or affect the ligand binding to DNA¹⁰. Each optimized ligand was then docked to the DNA receptor using Glide docking, and the results are included in Table 2. The most optimal docking score obtained was -7.331 (Figure 3A). This docking score is better than the scores achieved through MM Glide docking. This most favorable docking score was achieved through triple optimization of the ligand (Table 2, row 10). In the first optimization, loose convergence criteria was used to achieve an initial structure, and then the next two optimizations used normal convergence to achieve a more favorable conformation. This technique did in fact provide better docking results.

In addition to the docking scores, Prime MM-GBSA was used to calculate ligand binding energies for each docked pose¹⁶. In Table 2, both MM-GBSA dG bind and dG (NS) are included¹⁶.

Table 2. Summary of QPLD Docking Simulations.

Theory	Grid	Glide gscore	Docking Score	MMGBSA dG bind	MMGBSA dG (NS)
B3LYP-D3	11	-5.799	-5.617*	-15.84	-40.77
B3LYP-D3	11	-6.015	-5.833	-8.46	-23.16
B3LYP-D3	8	-6.085	-5.903	-.84	-16.27
M06-2X	11	-7.384	-7.202	-7.39	-24.71
M06-2X	11	-6.819	-6.637	-6.05	-19.26
M06-2X	11	-7.006	-6.824	9.51	-25.59
M06-2X	8	N/A**	N/A**	N/A**	N/A**
M06-2X	3	N/A**	N/A**	N/A**	N/A**
M06 2X	11	-7.514	-7.331***	-17.90	-48.45
M06-2X	12	-6.821	-6.639***	-17.09	-40.69

*did not input partial charges when docking

**N/A refers to results that were not usable due to binding in the major groove, poor docking score, etc.

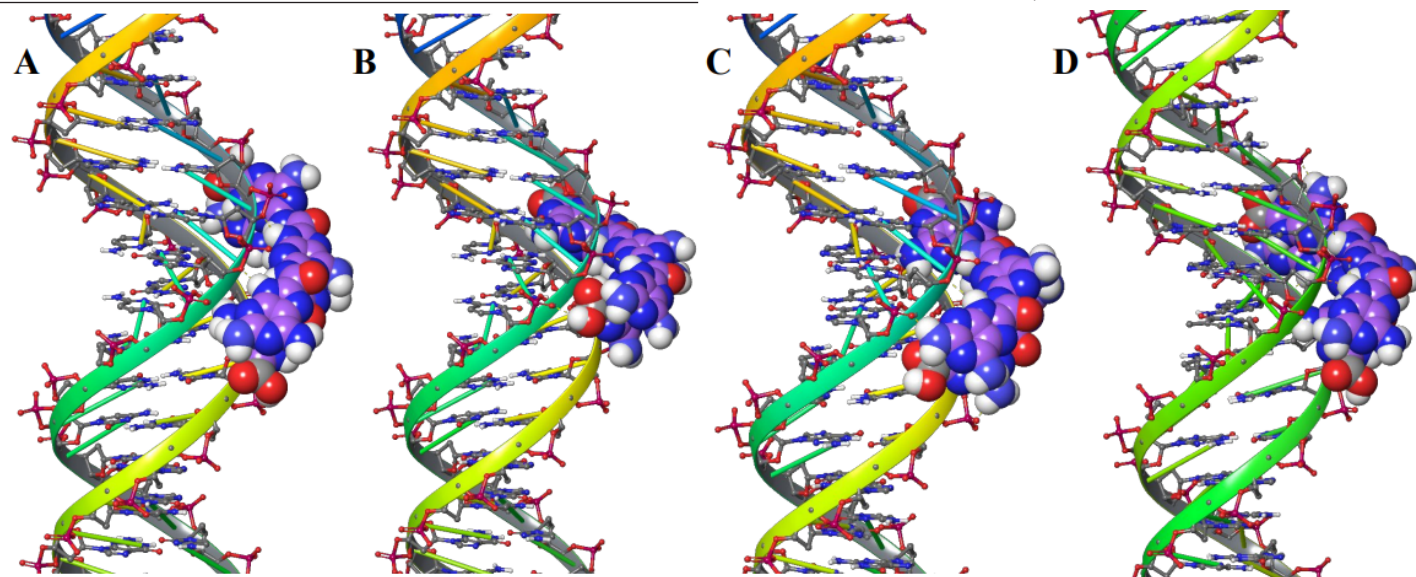
***triple optimization of a previous ligand

The dG bind calculation represents the total binding free energy of the ligand-receptor complex, while the dG bind (NS) calculation is the binding free energy without accounting for ligand-receptor strain¹¹. Based on the unfavorable dG bind scores and the more favorable dG bind (NS) scores, it can be hypothesized that the reason for the less favorable binding scores is strain between the ligand and the minor groove. Despite these scores, this method does show that the QM approach provides more optimal ligand poses and binding, even though there is still work to be done in order to better represent the interactions between the ligand and the minor groove of the DNA. Although experimental validation has not yet been performed, the computational results provide a valuable comparative assessment of the docking approaches investigated. Future experimental studies will be necessary to determine which method most accurately reflects the biological binding behavior of the ligand-DNA system.

Controls. Creating distamycin in the 2D Sketcher and docking it to an A-/T-rich section of DNA with Glide docking yielded a Docking score of -11.439. Jaguar optimization with B3LYP gave a docking score of -11.564, while optimization with M06-2X gave a score of -11.951¹⁴. These results of the distamycin control show that this proposed docking method yields similar results to the known binding structure of distamycin.

Conclusion and future work

There is still future work needed to further optimize these results and obtain experimental validation of the findings. For further optimization, new receptor grid generation methods may be designed. However, the best grid created was grid 11 which yielded the most optimal docking scores and the least amount of docking to the major groove. Additionally, more optimization of the ligand needs to be done to obtain lower docking scores. For example, exploring different QM theories could provide more favorable results. Furthermore, different solvation models need to be



Docking score: -7.331

Docking score: -7.202

Docking score: -6.824

Docking score: -6.639

Figure 3. Top 4 QPLD docking poses. (A) Docking of the ligand optimized with M06-2X and docked with grid 11 to yield a docking score of -7.331. This ligand was optimized three times to provide the most favorable conformation. (B) Docking of the ligand optimized with M06-2X and docked with grid 11 to yield a docking score of -7.202. (C) Docking of ligand optimized with M06-2X and docked with grid 11 to yield a docking score of -6.824. (D) Docking of ligand optimized with M06-2X and docked with grid 12. This ligand was optimized three times to yield a docking score of -6.639.

built to most accurately represent the role of water molecules in the binding site. An explicit solvation model may be explored to analyze water molecules' interactions within the binding site. This can be done using Desmond to build a solvent system around the DNA¹⁹. Then, Desmond molecular dynamics would allow analysis of water molecules within the binding site to determine which molecules may play a crucial role in the interaction of the ligand and DNA¹⁹. While these simulations were beyond the scope of the present study due to software licensing limitations, they are an important avenue for further refinement of the docking model.

Additionally, including an explicit solvation model could possibly help simulate the water displacement that occurs during ligand binding, which may better account for changes in entropy and result in a more favorable docking score^{20, 22}. Glide docking simply accounts for entropy in the docking score by rewarding water molecules that are outside of a hydrophobic pocket with a more negative value in more lipophilic ligands and adding energy penalties based on the number of rotatable bonds in the ligand^{24, 25}. However, this is unable to account for the water displacement that occurs when the ligand binds to the receptor, which typically contributes to the overall increase in entropy^{20, 22}. The effect of water displacement on the change in entropy has proven very difficult to accurately simulate in computational programs²³. Using an explicit solvation model might account more for the entropy changes and, more importantly, provide more realistic results.

Of course, experimental validation will also be needed to ensure this methodology and the results obtained are representative of the physical properties and binding abilities of this ligand. Currently, synthesis of this ligand is in progress, and after successful synthesis, the ligand will undergo dimerization and DNA binding assays to verify whether or not this ligand can bind DNA and provide physical meaning to the computational workflow described in this paper.

Although Schrödinger's Maestro is not commonly used for DNA-ligand docking studies, this work establishes a computational workflow for investigating DNA-binding ligands in the absence of experimental data. Using nucleotides 133-192 of the FMR1 gene as a model system, quantum mechanical-polarized ligand docking (QPLD), which combines Jaguar-based QM optimization of the ligand with Glide docking to the DNA receptor, produced the most favorable docking results among the approaches evaluated. While the performance of this workflow will likely depend on ligand size, structure, and chemical properties, this methodology provides a foundation for future studies of DNA-ligand interactions within the Schrödinger platform. These results demonstrate the advantages of incorporating QM-based methods into DNA docking workflows and offer a framework that can be further refined as experimental data become available.

Acknowledgements

We thank the University of Evansville Wargel Department of Chemistry and the University of Evansville CiSM program for supporting this project.

References

1. Wang, G.; Wang, F.; Huang, Q.; Li, Y.; Liu, Y.; Wang, Y. *BioMed Research International* **2015**, 2015, 757530.
2. Broyles, S. S.; Kremer, M.; Knutson, B. A. *Journal of Virology* **2004**, 78 (4), 2137–2141.
3. Weidenbach, S.; Hou, C.; Chen, J.-M.; Tsodikov, O. V.; Rohr, J. *Journal of inorganic biochemistry* **2016**, 156, 40–47.
4. Morris, G. M.; Lim-Wilby, M. *Methods in molecular biology (Clifton, N.J.)* **2008**, 443, 365–382.
5. Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelly, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. *Journal of Medicinal Chemistry* **2004**, 47 (7), 1739–1749.
6. Asadi, P.; Ghadamali Khodarahmi; Hossein Farrokhpour; Hossainzadeh, F.; Lotfollah Saghaei. *Research in Pharmaceutical Sciences* **2017**, 12 (3), 233–233.
7. Clark, A. M. *Journal of Chemical Information and Modeling* **2011**, 51 (12), 3149–3157.
8. Nguyen, B.; Neidle, S.; Wilson, W. D. *Accounts of Chemical Research* **2008**, 42 (1), 11–21.
9. Satange, R.; Hou, M.-H. *RSC Chemical Biology* **2025**.
10. Decherchi, S.; Masetti, M.; Vyalov, I.; Rocchia, W. *European Journal of Medicinal Chemistry* **2014**, 91, 27–42.
11. Genheden, S.; Ryde, U. *Journal of Chemical Theory and Computation* **2011**, 7 (11), 3768–3778.
12. Schrödinger Release 2026-2: Epik, Schrödinger, LLC, New York, NY, 2025.
13. Schrödinger Release 2026-2: Glide, Schrödinger, LLC, New York, NY, 2025.
14. Schrödinger Release 2026-2: Jaguar, Schrödinger, LLC, New York, NY, 2025.
15. Schrödinger Release 2026-2: LigPrep, Schrödinger, LLC, New York, NY, 2025.
16. Schrödinger Release 2026-2: Prime, Schrödinger, LLC, New York, NY, 2025.
17. Schrödinger Release 2026-2: Protein Preparation Workflow; Epik, Schrödinger, LLC, New York, NY, 2024; Impact, Schrödinger, LLC, New York, NY; Prime, Schrödinger, LLC, New York, NY, 2025.
18. Schrödinger Release 2026-2: QSite, Schrödinger, LLC, New York, NY, 2025.
19. Schrödinger Release 2026-2: Desmond Molecular Dynamics System, D. E. Shaw Research, New York, NY, 2024. Maestro-Desmond Interoperability Tools, Schrödinger, New York, NY, 2025.