Submitted: 01/14/21

Research Article

Ilya Dementyev[1], Ashkan Karimi[1]

[1]Deparment of Chemistry, McGill University, Montreal, QC, Canada

A Molecular Dynamical Investigation of the 7,8-dihydro-8-oxoguanine Mutation in dsDNA

Keywords

Hoogsteen, Watson-Crick, Metadynamics, Oxoguanine, GROMACS

Email Correspondence

ilya.dementyev@mail.mcgill.ca

Abstract

Background: The oxidization of a Guanine (G) base pair to 7,8-dihydro-8-oxoguanine (OG) is one of the most common DNA mutations. OG mutations can undergo a regular Watson-Crick base-pairing, or a reverse Hoogsteen (HG) base-pairing, especially in OG:A mismatches. While the causes of these mutations are well-understood, the kinetic and energetic characteristics of this new pseudo-base have never been fully investigated, especially at temperatures around biological function (17-37°C).

Methods: We created a simulation to derive the Free Energy Surface (FES) of OG:C and OG:A Hoogsteen to Watson-Crick base-pair (bp) transitions under multiple temperatures, relative to 2 collective geometric variables: the dihedral Chi and the pseudo-dihedral CPD angle. To make the simulation, we used the relatively recent Metadynamics algorithms in conjunction with GROMACS 2020.2.

Results: The lowest free energy increased linearly with increasing temperatures (17-37°C). Major Chi and CPD rotations at these minima varied heavily for 27°C and 32°C (the largest was seen in the former), but stayed relatively similar for other temperatures, indicating a highly sensitive relationship to temperature, likely due to DNA flexibility, quantum mechanical (QM) effects, and hydrogen bonding. Free energies had a weak negative linear relationship, and free energy hypersurfaces were given for studied temperatures of 17-37°C. Human body temperature (37°C) results were also included and explained. The simulations showed why OG:A Hoogsteen bps often occur in organisms and are energetically preferable to standard Watson-Crick. OG:C HG base pairings are determined to likely be not as common as OG:A HG.

Limitations: Future investigations must focus on discovering rate constants of these base-pairs, as time constraints did not permit them to be done here, as well as more QM-focused simulations.

Introduction

Genetic information storage in DNA, encoded by base-pair sequences of Adenine, Guanine, Cytosine, and Thymine, is extremely sensitive to change. Cancer and other genetic diseases, such as Sickle-cell Anemia, are all caused by mutations in this polymer, leading to a change in function due to information distortion. The conversion of a Guanine (G) base pair to 7,8-dihydro-8-oxoguanine (OG) via oxidation (1) is among the most common and biologically relevant DNA mutations. This mutation is caused by the oxidation of guanine, by a reactive oxygen species (ROS), for instance with HOOO, O2-, and OH. (2) While these mutations and the mechanism of their formation is well-studied, dynamical characteristics of the new pseudo-base have not been fully investigated. HG pairing is also a required characteristic for proper DNA strand replication. (3) Hence, HG is necessary for some genomic loci, but is detrimental in others.

The 8-oxoguanine-Cytosine (G:C) base pair, mutated from the standard G:C base pair, is the mutation that will be covered, as well as OG:A (mutated from G:A). Upon mutating, OG is capable of forming not only a Watson-Crick base-pair, but also the less-common Hoogsteen base-pair after propeller rotation (when one base (OG) turns using its glycosidic bond with respect to the other (C or A)). It is the DNA bending local to the mutation, caused by this glycosidic bond rotation, which causes the human OGG1 enzyme to identify and repair the mismatch. (4) Therefore, understanding the kinetic details would help scientists develop better diagnostic tools and treatments for cancer, especially at the human body temperature (approximately 37°C). (5)

The software PLUMED was used in conjunction with GROMACS. For PLUMED, GROMACS v. 2019.6 was used (6) (later releases not yet supported for PLUMED). For non-Metadynamics simulations, GROMACS v. 2020.2 was used (29)—a software for molecular dynamics (MD) simulations. Metadynamics is a computational method where one is able to sample conformations with high-energy barriers for any molecular process (from a geometric transition to a chemical reaction) that are normally not reached with regular non-biased methods. Metadynamics is especially useful to study the energetics of conformational changes, as more than one collective variable (CV) can be used to study the energy. CVs are differentiable functions of vectors of 3N atomic cartesian coordinates, (7) which may be anything from dihedral angle torsion, to the centre-of-mass distance between 2 nucleotides. (8) CVs were chosen at the researcher's discretion and are constrained only by the software's capabilities. The most important CVs that were used for this project were the Chi dihedral and CPD pseudo-dihedral angles defined by Pak et. al.—the outlined boxes represent the centre of mass of atoms within the box (Fig. 1B). (9)

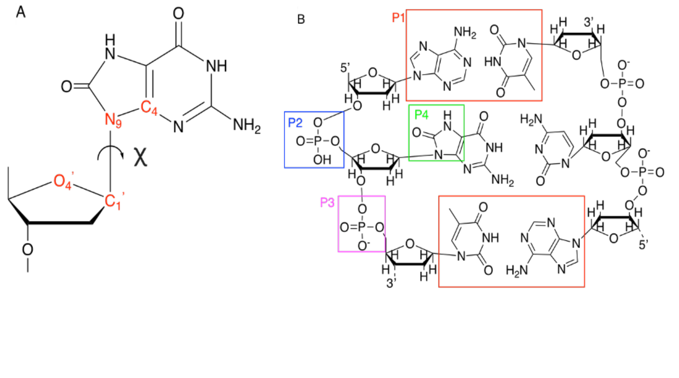

Figure 1. A. The definition of the glycosidic dihedral angle, chi. B. The definition of the pseudo-dihedral angle CPD. Both modified diagrams from Pak et. al. to fit this paper's context. (6)

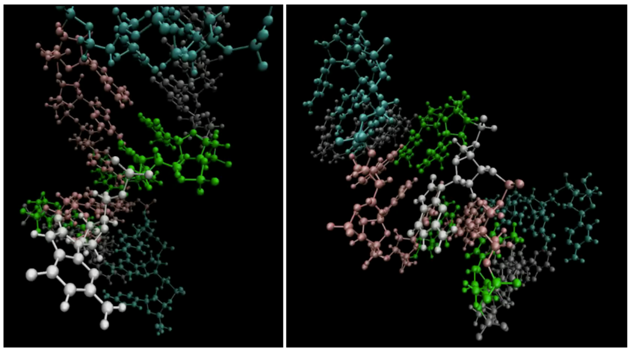

Figure 2. (left to right) A. A nitrogenous base flipped out of the helix. B. The nitrogenous base engaging in Hoogsteen base pairing.

Contrary to older energetic sampling methods, no "a priori knowledge of the [energy] landscape is required" (8) for accurate sampling with Metadynamics. The only requirement is to have a sufficiently long simulation that will allow the sampling to explore every niche of the energy surface. According to Pak et. al., 20 ns is sufficient, (9) although 500 ns was used to ensure convergence specific for the outlined systems. This was decided based on simulations performed from 50 to 200 ns. Metadynamics is inherently parallelizable, allowing GPU acceleration to speed up the sampling. (8) Specifically, Well-tempered Metadynamics was used, in which a biased sampling filled up a region, and became progressively less biased the "higher" it sampled, ensuring a relevant description of the potential landscape. Energetics plays an important role as seen in its direct mathematical relationship to probability (10):

$$\lim_{t\to\infty} P(s,t) \propto e^{\frac{-F(s)}{T+\Delta T}} \tag{1}$$

Where P(s, t) is the probability of the system having CV "s" at time "t", e is Euler's number, F(s) is the Free Energy, T is the absolute temperature (17-37°C), and ΔT is the raised temperature, defined as such (10):

$$\Delta T = T(\gamma - 1) \tag{2}$$

Where γ is the bias factor (γ = 15). (9) Hence, the lower a conformation's free energy, the higher the likelihood of that conformation existing at any time, which is relevant when predicting the equilibrium rate constants for the WC to HG transition. A system's free energy is defined as:

$$\Delta G = \Delta G° + k_B T \ln Q \tag{3}$$

$$\Delta G = \Delta H - T\Delta S \tag{4}$$

Where ΔG° is the Gibbs free energy (GFE) at SATP, k_B is Boltzmann's constant, T is the absolute Kelvin temperature, Q is the reaction quotient, ΔH is enthalpy, and ΔS is entropy. At equilibrium, Q = K, whose formula is expressed as:

$$K = \frac{k_{WC\to HG}}{k_{HG\to WC}} \tag{5}$$

Because of DNA's complex chemistry, some genomic loci will have different energies than others. (11) Propeller rotation does not happen in a regular simulation because of the high energy barrier between WC and HG conformations. For instance, one calculated value of kWC->HG = 16.7 s^{-1} showed that the HG conformation occurs infrequently. (12) Although the nucleotides used by Alvey et. al. differ from the ones we studied, around the same values for the OG:A and OG:C were expected. The forward rate constant for OG:A may be slightly greater than OG:C, due to OG:A being the preferred mismatch. To analyze the error, block sampling was used. Sometimes, CVs were correlated, as can be displayed by the autocorrelation function (13):

$$R(\tau) = \frac{(X_t - <X>)(X_{t+\tau} - <X>)}{\sigma^2} \tag{6}$$

Where X is a variable, <X> is the variable's average, σ^2 is the variance, t is time, and τ is the time lag. Between 2 variables, if the function is 1 at x = 0, and 0 elsewhere, the variables are perfectly independent. Within these simulations, CVs had non-zero values at x ≠ 0, displaying correlation. This implies that the central limit theorem does not hold, and taking the variance of all the values is not enough for calculating error bars as it underestimates such errors. (14) However, block sampling circumvents this issue by taking averages of groups of data ("blocks"), giving an accurate value for the error while taking correlation into account:

$$\sigma^2 = \frac{1}{N_b - 1}\sum_{i=1}^{N_b}(s_i' - < S >)^2 = 1 \tag{7}$$

Where N_b is the number of blocks, s_i' the block's average, and <S> is the average of all blocks summed together.

Methods

System Preparation

The system designed was an NPT canonical ensemble. PDB files for the B-DNA were made on Avogadro (15) and edited using Notepad++. (16) The charge entries for 8-oxo-guanine in residuetypes.dat were taken from Miller et. al (17) to ensure a charge close to zero (-0.0001 < Qsys < 0.0001). (18) The Particle-Mesh Ewald (PME) algorithm was used for long-range interactions, which introduced new species into the system if the net charge was non-zero. (18) Miller's data did not include the sugar, but the charge became an integer once the partial charges for the deoxyribose were included (Appendix I). The resonance structures shown in Fig. 2A demonstrate that the base's 5-membered ring is planar. For the simulated system, the most recent and reliable AMBER-OL15 force field used (19) was available on GROMACS' user-generated site. The SPCE model of water was used for solvation. (20) VMD software was used for the modelling to check every GROMACS output file. (21) To neutralize the system, the 5.69 nm cubic unit cell was programmed to contain 0.2 M NaCl. This salt was chosen per the supervisors' advice since they designed probes to analyze Hoogsteen base-pairings in NaCl solution. A Van-der-Waals cut-off was used for short-range atomic interactions. (22) The algorithms used for the barostat and thermostat were Parrinello-Rahman and Nose-Hoover respectively. (23, 24) These algorithms were chosen based on accuracy. Results would have been different if different parameters were used. The dsDNA used was an 11-nucleotide sequence (11-mer) recommended by Mr. Karimi, shown below:

5'-GGGTT(OG)ATGGG-3'
3'-CCCAA(T/A)TACCC-5'

The 3 GC pairs at each end increased the structural stability of the strand by decreasing fraying. An 11-mer was chosen since an 11-nucleotide strand is the minimum length required to be stable over long simulation periods, as suggested by Mr. Karimi.

The simulation process was divided into 5 sections: System Design and Preparation (including solvation and addition of ions), Energy Minimization, NVT Equilibration, NPT Equilibration, and the canonical production MD. (25) NVT was done first as it is less calculation-dense and avoids the travelling of the dsDNA molecule between unit cells. NPT then equilibrates for pressure and prepares the system for NPT-style Production MD, more accurate to real-life experiments. The equilibration took no longer than an hour, with a time step of 1 fs. The simulations ran on the Cedar cluster located in Vancouver, CA. To speed up the simulation, the Nvidia P100 GPU was used to perform CUDA acceleration, an optimization method used for parallelizing heavy computational tasks. (26) Metady namics was used to analyze the kinetics of OG propeller rotation. Since a larger time-length gives more accurate results, (8) 500 ns simulations were conducted on OG:C and OG:A base pairs.

All MD parameter and topology files were displayed in the supporting information. The following summarization steps were modified from Dr. Lemkul's tutorials on GROMACS. (30) The following steps assume that GROMACS and PLUMED were already installed, and that the topology file was saved after every step.

System Design

To begin the simulation, the pre-edited .pdb file was converted into a GROMACS file. AMBER99-sb-il and SPCE were chosen for the force field and water model respectively. Now that the file was created, the system shape was defined to be a cubic unit cell of side length 5.69 nm. This number was taken from Pak et al., (9) who had previously done calculations for 8-oxoguanine in a different dsDNA system. The dimension was also chosen due to it tightly fitting around the dsDNA, while also maintaining enough space so that the dsDNA did not immediately leave the unit cell.

Solvation

Now that the system was constructed, it needed to be solvated with water, with the gmx solvate command. Several other commands were used that chose the spc216.gro GROMACS water model file corresponding to SPCE.

Addition of Ions

To add ions, the grompp command was used to prepare the file (ions.mdp must be created beforehand; see Supporting Information). This created a .tpr (portable binary run input) file, which contains coordinate, trajectory, and parameter information about the system. (29) Then, the actual ion generation was performed using the gmx genion command. When prompted, "SOL" was chosen, which would correspond to a number—the same process as choosing the force field. This replaced some solvent molecules with ions.

Energy Minimization

Energy minimization was performed using the "grompp" command, which outputted file_em.tpr, used for energy minimization. It used the steepest-descent minimization algorithm to adequately minimize the system's energy. Graphs made from the descent simulation showed convergence. A conjugate gradient method would have also increased the speed of the initial convergence, but at the time of creating the simulation, we had not received help from supervisors with regards to minimization algorithms, and were not aware of this method. Hence, a good quantity of information was found by searching through past online MD tutorials, none of which made mention of the conjugate gradient method.

The simulation ran using the gmx mdrun command. The important file was "file_em.tpr" as that gave the specificity to the MD simulation type that ran with mdrun. Using "-v" (verbose), every step of the process was listed out. This simulation took no longer than 30 minutes. The output file was "file_em.edr". To analyze the results, we used the gmx energy command. Any analysis that was listed by GROMACS was able to be performed in this step. The most important would be the potential energy. Its graph appears as exponential decay for most systems. The graph was acquired by typing "xmgrace analysis.xvg" into the terminal (assuming xmgrace is installed), and the results were displayed.

NVT Equilibration

The first step of the 2-step equilibration was performed using gmx grompp, which was a similar process to the previous step, outputting the file "file_nvt.edr". The second simulation was performed using the same gmx mdrun code syntax. Again, this did not take more than an hour. The file was analyzed using the same command, "gmx energy".

NPT Equilibration

The second step was done with gmx grompp with the file_nvt.gro file as the checkpoint and the gmx mdrun command as usual:

```
gmx grompp -f file_npt.mdp -c file_nvt.gro -r file_nvt.gro -t file_nvt.cpt -p
topol.top -o file_npt.tpr
```

The "-t" command included the checkpoint file (file_nvt.cpt) that contained the required variables needed for the equilibration to continue from the previous one. The results were again analyzed using "gmx energy", and the "xmgrace" software.

Production MD

The final simulation was longer than the previous 3, because they always ran for at least 1 ns, corresponding to a 500,000 time-step simulation (the previous ones had 50,000). Nevertheless, the syntax was very similar to the NPT equilibration step, except the file_npt.gro and file_npt.cpt files were used as checkpoint files. The final step was done using the gmx mdrun and -plumed command, with "plumed.dat" written in text after the latter command. The production MD was NPT, to ensure that the GFE was being calculated (10)

Post-Production Analysis

There were a lot of possible analyses that could be done after the final MD was finished. However, correcting for periodicity is good for all analyses. The studied solute would sometimes cross the boundary between system boxes (i.e. unit cells). To avoid re-analysing that periodicity, the "trjconv" command was used as per GROMACS guidelines. Now, any type of analysis could've been performed using analytical commands found in the GROMACS handbook. (29)

Metadynamics – PLUMED Analysis

For correct PLUMED utilization, the steps were completed until NPT Equilibration. Then, the template file "plumed.dat" (Supporting Info) generated necessary variables. Chi, CPD, bias-factors, and sigma values were found in published papers related to the system investigated. (9) For OG-related simulations with dsDNA, only atom index editing was necessary. Chi and CPD values were cross-referenced with the most recent coordinate (.gro) file to ensure that the correct atoms are used. Finally, the simulation ran with the mdrun and -plumed commands. When the simulation finished, dihedral angles (and their energy) were extracted using plumed sum_hills, which generated a file called "fes.dat" which was then converted to Excel format and plotted.

Error Analysis

Error analysis was performed using simulation post-analysis. To do this, the PLUMED "driver" function was used. However, before any inputs were given for the function, we created a mass-charge (mc) file for the driver to use. To do this, the most recent .gro file was taken (usually the NPT one) and analyzed via gmx editconf. This created the necessary .pdb file that was used with the "plumed driver" utility and plumed.dat. This gave COLVAR_ERR, which would then be used to extract the error biases for each data point (the hashtags in the beginning denote sentences that should not be written into the terminal). The following code is modified from https://www.plumed.org/doc-v2.5/user-doc/html/trieste-4.html:

```
# find maximum value of bias
bmax=`awk ‘BEGIN{max=0.}{if($1!=”#!” && $4>max)max=$4}
END{print max}’ COLVAR_ERR`
```

```
# print phi values and weights
awk '{if($1!="#!") print $2,exp(($4-bmax)/kbt)}' kbt=2.494339 bmax=$bmax COLVAR_ERR > chi.weight
```

This created an error file for the Chi variable (chi.weight). This file was then processed using a python script to get the results for each block size:

```
for i in `seq 1 10 x`; do python3 do_block_fes.py chi.weight 1 -3.141593 3.018393 51 2.494339 $i; done
```

Variable "x" was chosen for the length of the simulation, as it can vary. This printed out one data file for each block size containing the error value. To collect the data into one text file, the following input was given:

```
for i in `seq 1 10 x`; do a=`awk '{tot+=$3}END{print tot/NR}' fes.$i.dat`; echo $i $a; done > err.blocks
```

This outputs an Excel-compatible ".blocks" file. The same steps can be repeated for as many CVs as necessary. The error for a specific temperature was given as the average between the corresponding Chi and CPD errors.

Results

All data was analysed and visualized using the Python Pandas package in Spyder 4.0 and Matplotlib.

In Fig. 3, OG:A energetics have a much smaller range, indicating more stability. This was expected, as OG tends to get mismatched more with A during replication. Both energies have a positive correlation, as expected, with high R-squared values. However, while the R-squared value was relatively strong, neither lines will likely give an accurate description of the process' enthalpy and entropy, due to the non-Boltzmann interactions that took place. (12)

Lastly, for both base-pairs, Chi was determined to change in a polynomial fashion across the temperature scale chosen. The curves bore a resemblance to sinusoidal functions, but such a relation can only be confirmed from multiple analyses of each temperature. All Chi and CPD values had an error of ± 0.0561 rad (= ± 3.2°).

Discussion and Conclusion

Firstly, the "blue" area for all simulations had the largest area at the highest temperature, due to an increase in area with an increase in temperature (Fig. 3 – Only OG:A bp is shown). This makes sense, as each increase in temperature steadily approached DNA's melting point. It also confirms that HG base-pairing is more preferential (on average) in OG:A as opposed to OG:C bp, evidenced by the larger frequency of blue areas in $-\pi/2 = -90° < \text{Chi} < \pi/2 = 90°$.

Next, all of the minimum FE angles lied within the HG range, indicating that both systems were relatively more stable in that state rather than a WC one. However, the absolute transition energy was quite small, indicating a weakly spontaneous rotation can occur, except for OG:A 27°C (Fig. 3). Furthermore, OG:C showed an almost-constant angle for the most stable WC angle, which interestingly lies exactly on the upper angle bound for HG to WC transition. Lastly, the rotational energy for the more stable OG:A bp at $T_{\text{human body}}$ = 37°C (310 K) (Fig. 4 and 5) showed an almost double increase relative to the OG:C transition, implying that they were likely more common and more relevant to biological effects of HG bp. Smaller error within OG:A points implies a greater stability of that conformation relative to OG:C.

For OG:C, although the general positive trend was expected, the low strength of it was not (Fig. 5 – Only OG:A bp is shown). This was very likely due to the non-Boltzmann effects on the system, such as quantum mechanical (QM) effects that were not taken into consideration for the interest of fast time and data processing. QM effects of the glycosidic and CPD angles potentially contributed to the discrepancy. Simulations in the future should focus on finding them.

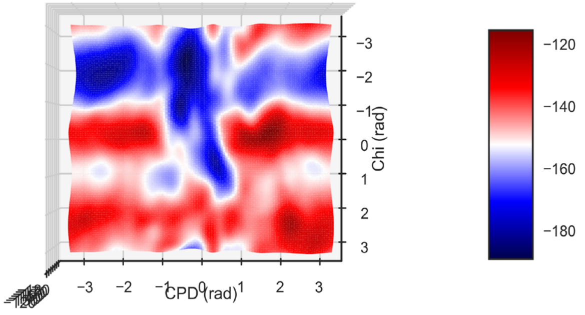

Figure 3. A. The Gibbs FES (in kJ/mol) of OG:A at 27°C

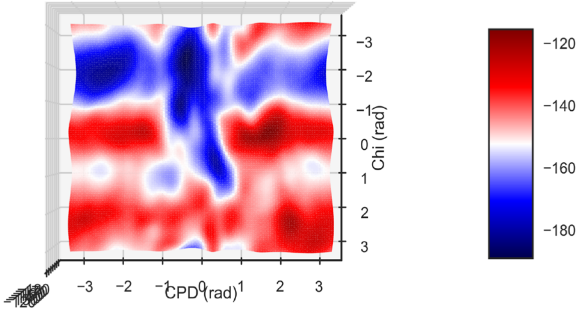

Figure 4. A. The Gibbs FES (in kJ/mol) of OG:A at 37°C

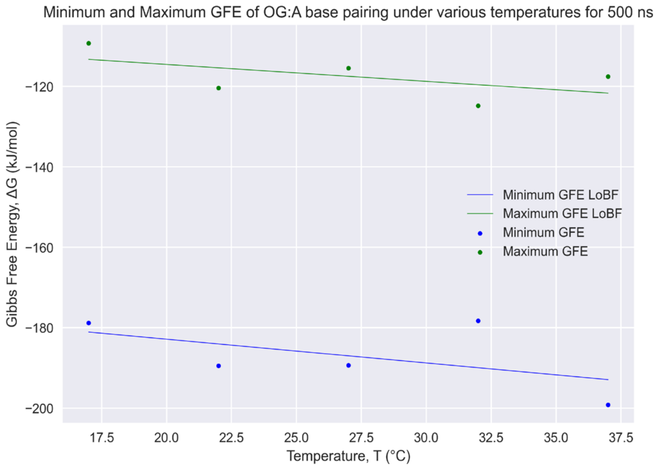

Figure 5. The minimum and maximum GFE of OG:A base pairing under various temperatures for 500 ns, with their corresponding lines of best fit (LoBF). The R-squared values for the maximum and minimum lines are 0.57 and 0.53 respectively. This shows decent correlation with the minimum GFE from standard thermodynamic principles. Error bars are too small to be significant in visual format.

In Fig. 5, most of the data trends were expected. The relatively high R-squared of the minimum value indicates that the MD was partially successful in replicating the results without QM assistance. However, while the R-squared value was relatively strong, neither lines will likely give an accurate description of the process' enthalpy and entropy, due to the non-Boltzmann interactions that took place. (12) In Fig. 6, it is interesting that the major rotations from WC to Hoogsteen occur between 300-305 K (27-32°C), just before reaching the original position at the human body temperature of 310 K (37°C). This suggests a very sensitive relationship between the stability of the Hoogsteen conformation to temperature, within just a few degrees Celsius, which is the same as the CPD angle, although the latter shows less dramatic changes between temperatures. It is likely that at these temperatures, the dihedral angle rotations would occur the most frequently, but this can only be confirmed through more Metadynamics studies of this temperature range. More data points at or above the minimum GFE could have been sampled to create a better representation of the angle. However, assuming that the sampling of the minimum GFE is sufficient, one reason for this phenomenon could be the DNA's intrinsic flexibility providing more support for the Chi angle to turn. At lower temperatures, the DNA may not be as flexible, but at higher temperatures, the flexibility could cause the CPD angle to change significantly, indirectly affecting Chi. This would not explain the regular Chi at 310 K (37°C) though.

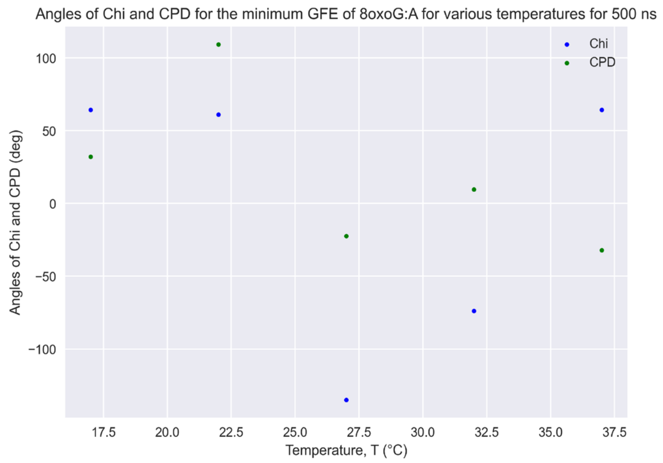

Figure 6. Chi and CPD angles for minimum GFE of OG:A for various temperatures for 500 ns.

Lastly, since the rotations of Chi in both base pairs are periodic for the temperature scales chosen, it is likely that a sinusoidal function (or some form of a Taylor approximation) can represent it well (Fig. 6 – only OG:A is shown). However, the CPD angle's deviation from a sinusoid in OG:C implies backbone contributions to base-pair flipping are more complex than trigonometric expressions, and potentially better represented with polynomials, which is why they were fitted and analyzed as such. Again, QM likely played a large role in influencing this data that Classical MD could not account for. The maximum errors listed were all smaller than $k_B T$, further supporting that the simulations converged well. Error graphs also showed excellent convergence too, with very small error. Future simulations should focus on confirming the near-sinusoidal relationship between temperature and Chi rotation, as well as finding ways to incorporate a more QM-leaning analysis for the sake of improved accuracy, as simulation technology advances to the point where such is possible.

Acknowledgements

This research was enabled in part by support provided by WestGrid (www.westgrid.ca), ACENET (www.ace-net.ca) and Compute Canada (www.computecanada.ca). This work was carried out using the open-source, community-developed PLUMED library,(27) version 2.5.4.(28). According to supervisors, the author is the first person in McGill Chemistry to perform MD simulation research, so some of the techniques/algorithms chosen may not be up-to-standard with the latest research due to lack of guidance. For access to more figures, they can be provided free-of-charge by contacting the author at ilya.dementyev@mail.mcgill.ca.

References

1. Pinak M. CHAPTER 10 - Enzymatic recognition of radiation-produced oxidative DNA lesion. Molecular dynamics approach. In: Starikov EB, Lewis JP, Tanaka S, editors. Modern Methods for Theoretical Physical Chemistry of Biopolymers. Amsterdam: Elsevier Science; 2006. p. 191-210.

2. Bruskov VI, Malakhova LV, Masalimov ZK, Chernikov AV. Heat-induced formation of reactive oxygen species and 8-oxoguanine, a biomarker of damage to DNA. Nucleic Acids Res. 2002;30(6):1354-63.

3. Johnson RE, Prakash L, Prakash S. Biochemical evidence for the requirement of Hoogsteen base pairing for replication by human DNA polymerase ι. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(30):10466.

4. Hashiguchi K, Stuart JA, de Souza-Pinto NC, Bohr VA. The C-terminal alphaO helix of human Ogg1 is essential for 8-oxoguanine DNA glycosylase activity: the mitochondrial beta-Ogg1 lacks this domain and does not have glycosylase activity. Nucleic Acids Res. 2004;32(18):5596-608.

5. Obermeyer Z, Samra JK, Mullainathan S. Individual differences in normal body temperature: longitudinal big data analysis of patient records. BMJ. 2017;359:j5468.

6. Berendsen HJC, van der Spoel D, van Drunen R. GROMACS: A message-passing parallel molecular dynamics implementation. Computer Physics Communications. 1995;91(1):43-56.

7. Fiorin G, Klein ML, Hénin J. Using collective variables to drive molecular dynamics simulations. Molecular Physics. 2013;111(22-23):3345-62.

8. Barducci A, Bonomi M, Parrinello M. Metadynamics. Wiley Interdiscip Rev: Comput Mol Sci. 2011;1(5):826-43.

9. Yang C, Kim E, Pak Y. Free energy landscape and transition pathways from Watson–Crick to Hoogsteen base pairing in free duplex DNA. Nucleic Acids Res. 2015;43(16):7769-78.

10. Barducci A, Bussi G, Parrinello M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. Physical Review Letters. 2008;100(2):020603.

11. Khandelwal G, Lee RA, Jayaram B, Beveridge DL. A statistical thermodynamic model for investigating the stability of DNA sequences from oligonucleotides to genomes. Biophys J. 2014;106(11):2465-73.

12. Alvey HS, Gottardo FL, Nikolova EN, Al-Hashimi HM. Widespread transient Hoogsteen base pairs in canonical duplex DNA with variable energetics. Nat Commun. 2014;5:4786.

13. Nounou MN, Bakshi BR. Chapter 5 - Multiscale Methods for Denoising and Compression. In: Walczak B, editor. Data Handling in Science and Technology. 22: Elsevier; 2000. p. 119-50.

14. Ross SM. Chapter 7 - Distributions of Sampling Statistics. In: Ross SM, editor. Introductory Statistics (Fourth Edition). Oxford: Academic Press; 2017. p. 297-328.

15. Hanwell MD, Curtis DE, Lonie DC, Vandermeersch T, Zurek E, Hutchison GR. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. Journal of Cheminformatics. 2012;4(1):17.

16. Berman H, Henrick K, Nakamura H. Announcing the worldwide Protein Data Bank. Nature Structural & Molecular Biology. 2003;10(12):980-.

17. Miller JH, Fan-Chiang C-CP, Straatsma TP, Kennedy MA. 8-Oxoguanine Enhances Bending of DNA that Favors Binding to Glycosylases. Journal of the American Chemical Society. 2003;125(20):6331-6.

18. Hub JS, de Groot BL, Grubmüller H, Groenhof G. Quantifying Artifacts in Ewald Simulations of Inhomogeneous Systems with a Net Charge.

Journal of Chemical Theory and Computation. 2014;10(1):381-90.

19. Zgarbová M, Šponer J, Otyepka M, Cheatham TE, Galindo-Murillo R, Jurečka P. Refinement of the Sugar–Phosphate Backbone Torsion Beta for AMBER Force Fields Improves the Description of Z- and B-DNA. Journal of Chemical Theory and Computation. 2015;11(12):5723-36.

20. Berendsen HJC, Grigera JR, Straatsma TP. The missing term in effective pair potentials. The Journal of Physical Chemistry. 1987;91(24):6269-71.

21. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. Journal of Molecular Graphics. 1996;14(1):33-8.

22. De Leeuw SW, Perram JW, Smith ER. Simulation of electrostatic systems in periodic boundary conditions. I. Lattice sums and dielectric constants. Proc R Soc London, Ser A. 1980;373(1752):27-56.

23. Hoover W. Canonical Dynamics: Equilibrium Phase-Space Distributions. Phys Rev A: At, Mol, Opt Phys. 1985;31:1695.

24. Parrinello M, Rahman A. Polymorphic transitions in single crystals: A new molecular dynamics method. J Appl Phys. 1981;52(12):7182-90.

25. Zheng L, Alhossary AA, Kwoh C-K, Mu Y. Molecular Dynamics and Simulation. In: Ranganathan S, Gribskov M, Nakai K, Schönbach C, editors. Encyclopedia of Bioinformatics and Computational Biology. Oxford: Academic Press; 2019. p. 550-66.

26. Hwu W, Rodrigues C, Ryoo S, Stratton J. Compute Unified Device Architecture Application Suitability. Computing in Science & Engineering. 2009;11(3):16-26.

27. Bonomi M, Bussi G, Camilloni C, Tribello GA, Banáš P, Barducci A, et al. Promoting transparency and reproducibility in enhanced molecular simulations. Nature Methods. 2019;16(8):670-3.

28. Tribello GA, Bonomi M, Branduardi D, Camilloni C, Bussi G. PLUMED 2: New feathers for an old bird. Computer Physics Communications. 2014;185(2):604-13.

29. Lindahl, Abraham, Hess, & van der Spoel. (2020, January 1). GROMACS 2020 Manual (Version 2020). Zenodo. http://doi.org/10.5281/zenodo.3562512

30. Lemkul J. From Proteins to Perturbed Hamiltonians: A Suite of Tutorials for the GROMACS-2018 Molecular Simulation Package [Article v1.0]. Living Journal of Computational Molecular Science. 2018;1.