Ilya Sergeevich Dementyev
Luedtke Research Group (Supervision under Dr. Nathan Luedtke and Mr. Ashkan Karimi)

SUPPORTING INFORMATION

YouTube Tutorial: Installing GROMACS

This video was followed: https://www.youtube.com/watch?v=kA3SmMnphmI

FILES

All files listed here were added directly to the active simulation directory, or else GROMACS would have not been able to find them.

GROMACS MD

ions.mdp – plaintext file used to neutralize the system and create the .tpr file

; ions.mdp - used as input into grompp to generate ions.tpr

; Parameters describing what to do, when to stop and what to save

integrator = steep ; Algorithm (steep = steepest descent minimization)

emtol = 1000.0 ; Stop minimization when the maximum force < 1000.0 kJ/mol/nm

emstep = 0.01 ; Minimization step size

nsteps = 50000 ; Maximum number of (minimization) steps to perform

; Parameters describing how to find the neighbors of each atom and how to calculate the interactions

nstlist = 1 ; Frequency to update the neighbor list and long range forces

cutoff-scheme = Verlet ; Buffered neighbor searching

ns_type = grid ; Method to determine neighbor list (simple, grid)

coulombtype = PME ; Treatment of long range electrostatic interactions

rcoulomb = 1.0 ; Short-range electrostatic cut-off

rvdw = 1.0 ; Short-range Van der Waals cut-off

pbc = xyz ; Periodic Boundary Conditions in all 3 dimensions

file_em.mdp – plaintext file needed for the energy minimization

; minim.mdp - used as input into grompp to generate em.tpr

; Parameters describing what to do, when to stop and what to save

integrator = steep ; Algorithm (steep = steepest descent minimization)

emtol = 1000.0 ; Stop minimization when the maximum force < 1000.0 kJ/mol/nm

emstep = 0.01 ; Minimization step size

nsteps = 50000 ; Maximum number of (minimization) steps to perform

; Parameters describing how to find the neighbors of each atom and how to calculate the interactions

nstlist = 1 ; Frequency to update the neighbor list and long range forces

cutoff-scheme = Verlet ; Buffered neighbor searching

ns_type = grid ; Method to determine neighbor list (simple, grid)

coulombtype = PME ; Treatment of long range electrostatic interactions

rcoulomb = 1.0 ; Short-range electrostatic cut-off

rvdw = 1.0 ; Short-range Van der Waals cut-off

pbc = xyz ; Periodic Boundary Conditions in all 3 dimensions

file_npt.mdp – plaintext file needed for NPT equilibration (NVT is similar; simply add or set pcoupl=no, continuation=no, gen_vel=yes, gen_temp=300, gen_seed=-1)

```
title                = OP15 DNA with 8-oxoguanine NPT equilibration
define               = ; for specific position restraints

; Run parameters
integrator           = md       ; leap-frog integrator
nsteps               = 50000    ; 2 * 50000 = 100 ps
dt                   = 0.002    ; 2 fs

; Output control
nstxout              = 500      ; save coordinates every 1.0 ps
nstvout              = 500      ; save velocities every 1.0 ps
nstenergy            = 500      ; save energies every 1.0 ps
nstlog               = 500      ; update log file every 1.0 ps

; Bond parameters
```

continuation = yes ; Restarting after NVT

constraint_algorithm = lincs ; holonomic constraints

constraints = h-bonds ; bonds involving H are constrained

lincs_iter = 1 ; accuracy of LINCS

lincs_order = 4 ; also related to accuracy

; Nonbonded settings

cutoff-scheme = Verlet ; Buffered neighbor searching

ns_type = grid ; search neighboring grid cells

nstlist = 10 ; 20 fs, largely irrelevant with Verlet scheme

rcoulomb = 1.0 ; short-range electrostatic cutoff (in nm)

rvdw = 1.0 ; short-range van der Waals cutoff (in nm)

DispCorr = EnerPres ; account for cut-off vdW scheme

; Electrostatics

coulombtype = PME ; Particle Mesh Ewald for long-range electrostatics

pme_order = 4 ; cubic interpolation

fourierspacing = 0.16 ; grid spacing for FFT

; Temperature coupling is on

tcoupl = nose-hoover ; Nose-Hoover thermostat

tc-grps = DNA Water_and_Ions ; two coupling groups - more accurate

tau_t = 0.4 0.4 ; time constant, in ps

ref_t = 300 300 ; reference temperature, one for each group, in K

; Pressure coupling is on

pcoupl = Parrinello-Rahman ; Pressure coupling on in NPT

pcoupltype = isotropic ; uniform scaling of box vectors

tau_p = 2.0 ; time constant, in ps

ref_p = 1.0 ; reference pressure, in bar

compressibility = 4.5e-5 ; isothermal compressibility of water, bar^-1

refcoord_scaling = com

; Periodic boundary conditions

pbc = xyz ; 3-D PBC

; Velocity generation

gen_vel = no ; Velocity generation is off

file_md.mdp – needed for the production MD simulation; values may be changed as seen fit.

title = OL15 DNA with 8-oxoguanine production MD simulation

; Run parameters

integrator = md ; leap-frog integrator

*nsteps = 500000 ; 2 * 500000 = 1000 ps (1 ns)*

dt = 0.002 ; 2 fs

; Output control

nstxout = 0 ; suppress bulky .trr file by specifying

nstvout = 0 ; 0 for output frequency of nstxout,

nstfout = 0 ; nstvout, and nstfout

nstenergy = 5000 ; save energies every 10.0 ps

nstlog = 5000 ; update log file every 10.0 ps

nstxout-compressed = 5000 ; save compressed coordinates every 10.0 ps

compressed-x-grps = System ; save the whole system

; Bond parameters

continuation = yes ; Continuing from NPT

constraint_algorithm = lincs ; holonomic constraints (LINCS Algorithm)

constraints = all-bonds ; all bonds are constrained

lincs_iter = 1 ; accuracy of LINCS

lincs_order = 4 ; also related to accuracy

; Neighbor searching

cutoff-scheme = Verlet ; Buffered neighbor searching

ns_type = grid ; search neighboring grid cells

nstlist = 10 ; 20 fs, largely irrelevant with Verlet scheme

rcoulomb = 1.0 ; short-range electrostatic cutoff (in nm)

rvdw = 1.0 ; short-range van der Waals cutoff (in nm)

; Electrostatics

coulombtype = PME ; Particle Mesh Ewald for long-range electrostatics

pme_order = 4 ; cubic interpolation

fourierspacing = 0.16 ; grid spacing for FFT

; Temperature coupling is on

tcoupl = nose-hoover ; Nose-Hoover thermostat

tc-grps = DNA Water_and_Ions ; two coupling groups - more accurate

tau_t = 0.4 0.4 ; time constant, in ps

ref_t = 300 300 ; reference temperature, one for each group, in K

; Pressure coupling is on

pcoupl = Parrinello-Rahman ; Pressure coupling on in NPT

pcoupltype = isotropic ; uniform scaling of box vectors

tau_p = 2.0 ; time constant, in ps

ref_p = 1.0 ; reference pressure, in bar

compressibility = 4.5e-5 ; isothermal compressibility of water, bar^-1

; Periodic boundary conditions

pbc = xyz ; 3-D PBC

; Dispersion correction

DispCorr = EnerPres ; account for cut-off vdW scheme

; Velocity generation

gen_vel = no ; Velocity generation is off

Metadynamics - PLUMED

plumedGC.dat – PLUMED text file that executes the Metadynamics simulation (for OG:C)

 # Variable creation and definition, naming conventions kept from Yang et. al., # for 8oxoGC

p1: COM ATOMS=142-155,208-221,554-567,492-505

p2: COM ATOMS=195-199

p3: COM ATOMS=161-165

p4: COM ATOMS=174-179,189

 # Restraints to avoid DNA backbone-breaking, distances taken to simulate #hydrogen bonds between A and T in neighboring bps.

d1: DISTANCE ATOMS=217,503

d2: DISTANCE ATOMS=216,501

d3: DISTANCE ATOMS=151,562

d4: DISTANCE ATOMS=153,563

 # Call and restrain the restraints

RESTRAINT ARG=d1,d2,d3,d4 AT=1.0,1.0,1.0,1.0 KAPPA=150.0,150.0,150.0,150.0

LABEL=restraint

 # Define angles to analyze

chi: TORSION ATOMS=171,172,174,189

cpd: TORSION ATOMS=p1,p2,p3,p4

metad: METAD ARG=chi,cpd PACE=500 HEIGHT=1.0 SIGMA=0.16,0.16 BIASFACTOR=15

TEMP=300.0 GRID_MIN=-pi,-pi GRID_MAX=pi,pi GRID_SPACING=0.1,0.1 FILE=HILLS

 # monitor the two variables and the metadynamics bias potential

PRINT STRIDE=10 ARG=chi,cpd,metad.bias FILE=COLVAR

This will output a “HILLS” and “COLVAR” file in the same directory after the simulation is complete.

plumedGA.dat – PLUMED text file that executes the Metadynamics simulation (for OG:A) – atom numbers will vary for other researchers

p1: COM ATOMS=142-155,208-221,492-505,556-569

p2: COM ATOMS=195-199

p3: COM ATOMS=161-165

p4: COM ATOMS=174-179,189

d1: DISTANCE ATOMS=217,502

d2: DISTANCE ATOMS=216,501

d3: DISTANCE ATOMS=153,565

d4: DISTANCE ATOMS=151,564

cpd: TORSION ATOMS=p1,p2,p3,p4

chi: TORSION ATOMS=171,172,174,189

metad: METAD ARG=chi,cpd PACE=500 HEIGHT=1.0 SIGMA=0.16,0.16 FILE=HILLS_NEW

GRID_MIN=-pi,-pi GRID_MAX=pi,pi BIASFACTOR=15.0 TEMP=300.0

PRINT STRIDE=10 ARG=chi,cpd,metad.bias FILE=COLVAR_NEW

errGC.dat – for analyzing the error (OG:C)

RESTART

 # Variable creation and definition, naming conventions kept from Yang et. al.

p1: COM ATOMS=142-155,208-221,554-567,492-505

p2: COM ATOMS=195-199

p3: COM ATOMS=161-165

p4: COM ATOMS=174-179,189

 # Restraints to avoid DNA backbone-breaking, distances taken to simulate hydrogen bonds between A and T in neighboring bps.

d1: DISTANCE ATOMS=217,503

d2: DISTANCE ATOMS=216,501

d3: DISTANCE ATOMS=151,562

d4: DISTANCE ATOMS=153,563

 # Call and restrain the restraints

RESTRAINT ARG=d1,d2,d3,d4 AT=1.0,1.0,1.0,1.0 KAPPA=150.0,150.0,150.0,150.0 LABEL=restraint

#Define angles to study

chi: TORSION ATOMS=171,172,174,189

cpd: TORSION ATOMS=p1,p2,p3,p4

metad: METAD ARG=chi,cpd PACE=10000000 HEIGHT=1.0 SIGMA=0.16,0.16

BIASFACTOR=15 TEMP=300.0 GRID_MIN=-pi,-pi GRID_MAX=pi,pi

GRID_SPACING=0.1,0.1 FILE=HILLS # The pre-existing HILLS file will be analyzed

 # monitor the two variables and the metadynamics bias potential

PRINT STRIDE=1 ARG=chi,cpd,metad.bias FILE=COLVAR_ERR

errGA.dat – for analyzing the error (OG:A)

RESTART

 # Variable creation and definition, naming conventions kept from Yang et. al.

p1: COM ATOMS=142-155,208-221,554-567,492-505

p2: COM ATOMS=195-199

p3: COM ATOMS=161-165

p4: COM ATOMS=174-179,189

 # Restraints to avoid DNA backbone-breaking, distances taken to simulate hydrogen bonds between A and T in neighboring bps.

d1: DISTANCE ATOMS=217,503

d2: DISTANCE ATOMS=216,501

d3: DISTANCE ATOMS=151,562

d4: DISTANCE ATOMS=153,563

 # Call and restrain the restraints

RESTRAINT ARG=d1,d2,d3,d4 AT=1.0,1.0,1.0,1.0 KAPPA=150.0,150.0,150.0,150.0

LABEL=restraint

#Define angles to study

chi: TORSION ATOMS=171,172,174,189

cpd: TORSION ATOMS=p1,p2,p3,p4

metad: METAD ARG=chi,cpd PACE=10000000 HEIGHT=1.0 SIGMA=0.16,0.16

BIASFACTOR=15 TEMP=300.0 GRID_MIN=-pi,-pi GRID_MAX=pi,pi

GRID_SPACING=0.1,0.1 FILE=HILLS

 # monitor the two variables and the metadynamics bias potential

PRINT STRIDE=1 ARG=chi,cpd,metad.bias FILE=COLVAR_ERR

Local Desktop Specifications

OS: Windows 10 (MobaXTerm SSH/SFTP Client used to connect to Compute Canada clusters)

RAM: 16 GB

CPU: Intel® Core™ i7-4770 CPU @ 3.40 GHz

GPU: NVIDIA GeForce GTX 645

Local GROMACS simulations were run on Ubuntu 20.04 for Windows

Python code for atomic charge calculations for OG:

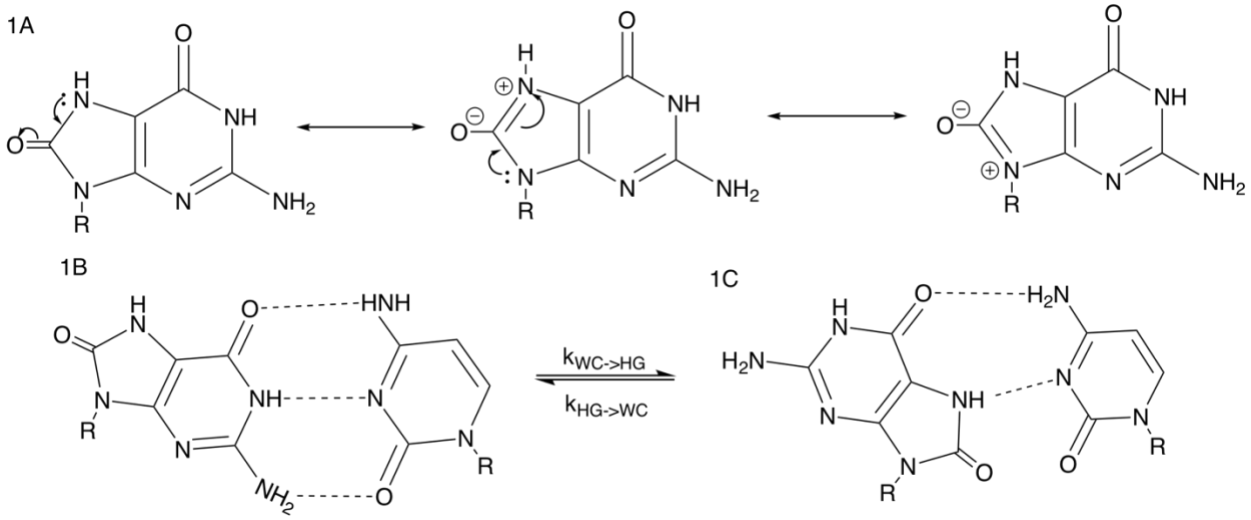
```
QoxoG = [-0.4025, 0.3266, 0.7208, -0.9625, 2*0.4371,
    -0.6118, 0.2108, -0.0211, 0.4299, -0.55,
    -0.5129, 0.4077, 0.4468, -0.5558, 0.111]

deoxy = [1.1659, -0.7761, -0.7761, -0.4954, -0.0069, 0.0754, 0.0754, 0.1629, 0.1176, -0.3691,
0.0358, 0.1746, 0.0713, 0.0985, -0.0854, 0.0718, 0.0718, -0.5232]

sigma_QdG_nuc = sum(QoxoG) + sum(deoxy)

print('Before deoxyribose sugar charges:', sum(QoxoG))
print('After deoxyribose sugar charges:', sigma_QdG_nuc)
```

OUTPUT:

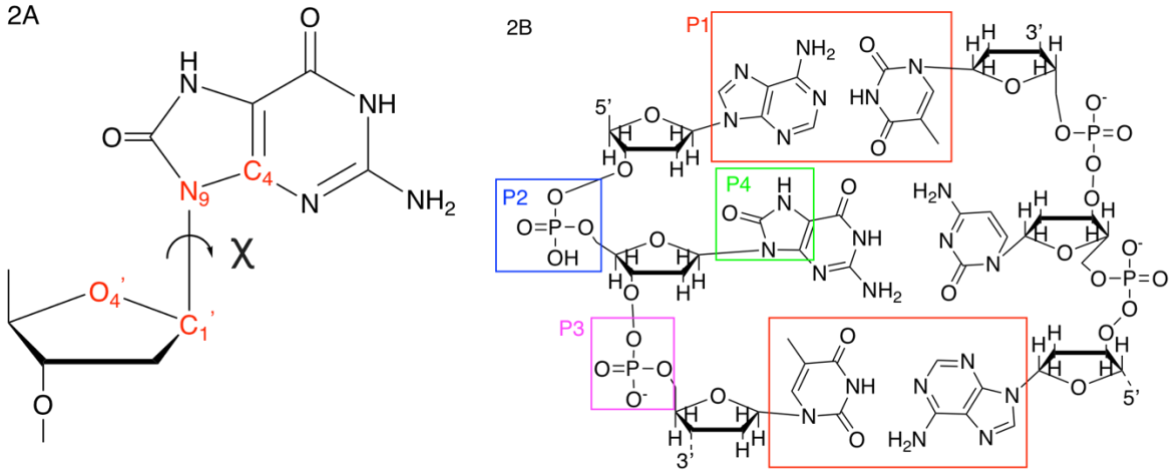
```
Before deoxyribose sugar charges: -0.08880000000000003
After deoxyribose sugar charges: -1.0000000000000002
```

Of course, the floating point in the second line of the output is analytically insignificant.

BACKGROUND INFO FIGURES



Figures: **1A)** The OG nitrogenous base, with resonance structures displayed. **1B)** An OG:C base pair, both nucleotides in regular anti conformation. **1C)** An OG:C base pair, with the OG rotated to its syn conformation to form a HG base-pair with C. The R-group represents the rest of the nucleotide



Figures: **2A)** The definition of the glycosidic dihedral angle, chi. **2B)** The definition of the pseudo-dihedral angle CPD. Both modified diagrams from Pak et. al. to fit this paper's context.(6)

RESULTS FIGURES

Contour graphs

The Gibbs FES of the OG:C bp 500 ns simulation at 290 K

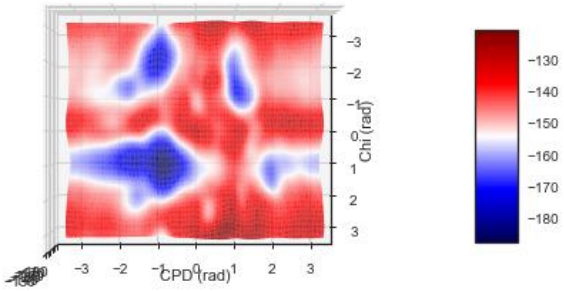


Fig. I

The Gibbs FES of the OG:A bp 500 ns simulation at 290 K

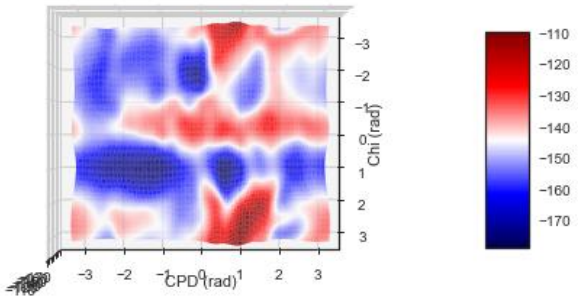


Fig. II

The Gibbs FES of the OG:C bp 500 ns simulation at 295 K

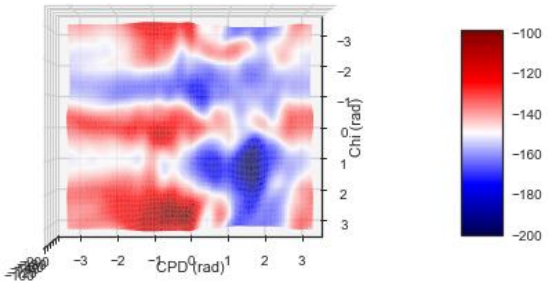


Fig. III

The Gibbs FES of the OG:A bp 500 ns simulation at 295 K

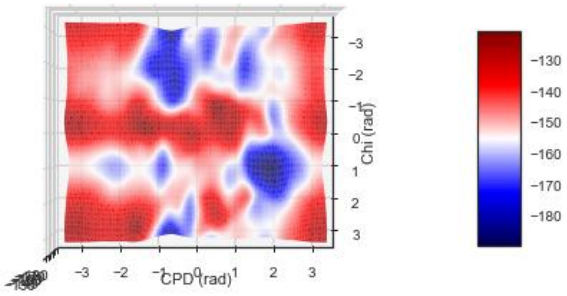


Fig. IV

The Gibbs FES of the OG:C bp 500 ns simulation at 300 K

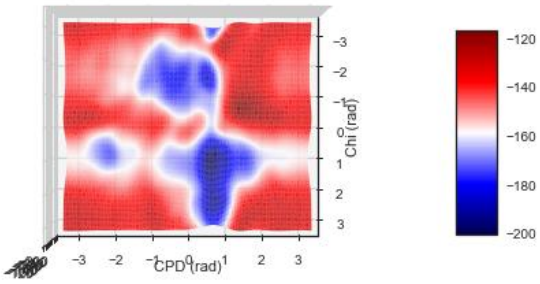


Fig. V

The Gibbs FES of the OG:A bp 500 ns simulation at 300 K

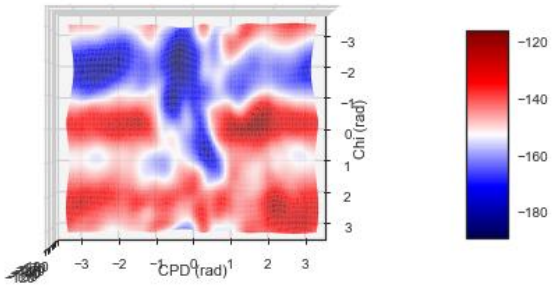


Fig. VI

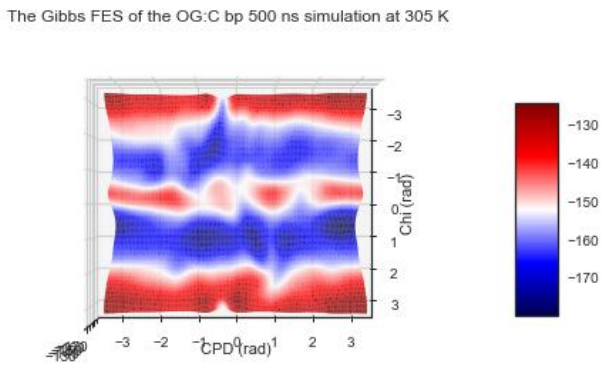


Fig. VII

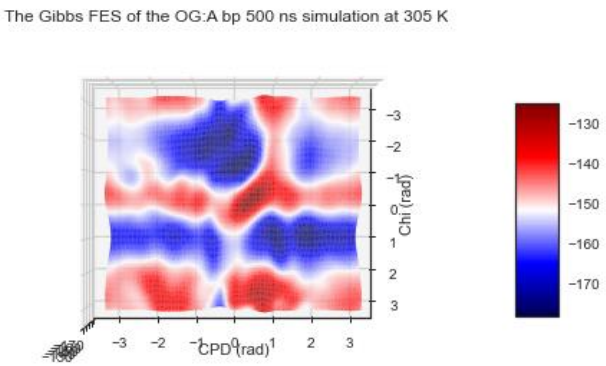


Fig. VIII

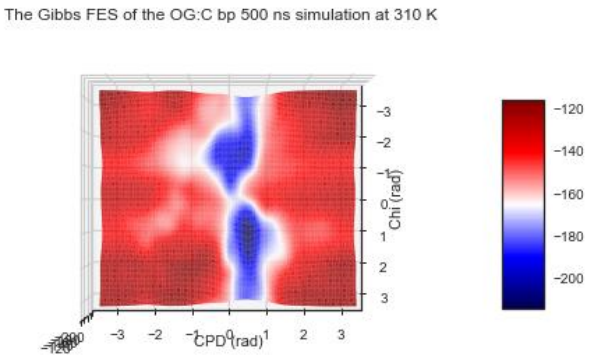


Fig. IX

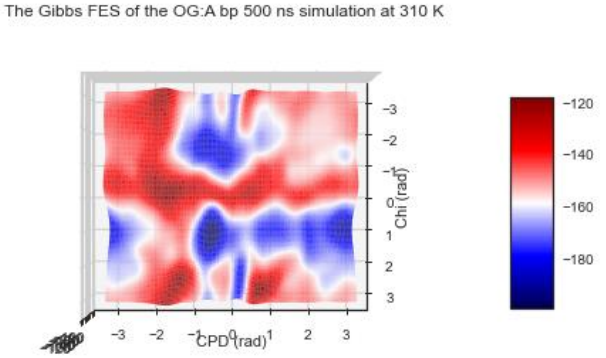


Fig. X

Fig. I – X: The Gibbs FES (in kJ/mol) of OG:C and OG:A systems at various temperatures

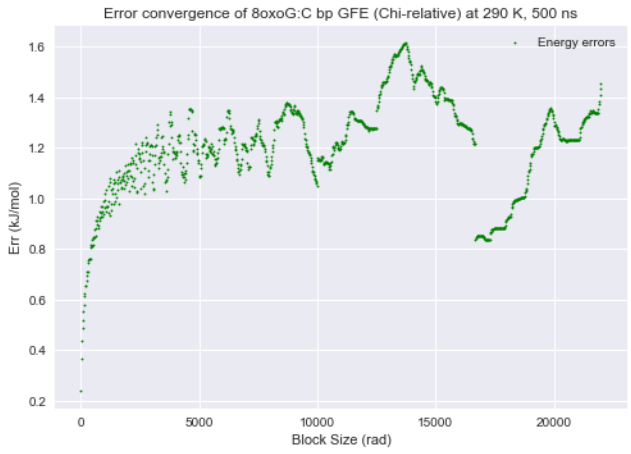


Fig. XI

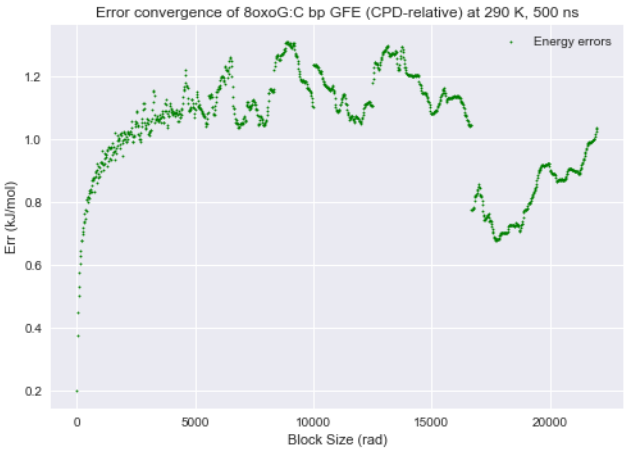


Fig. XII

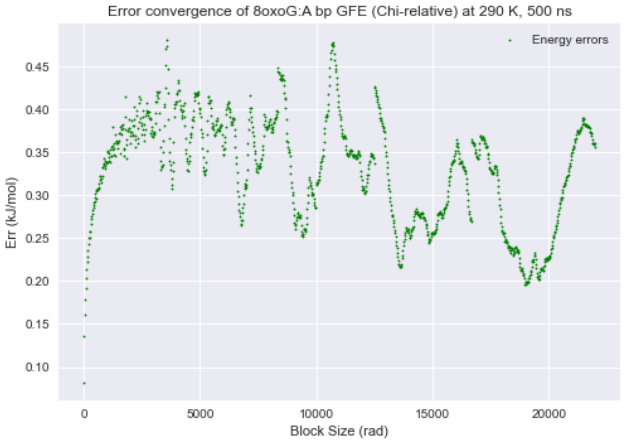


Fig. XIII

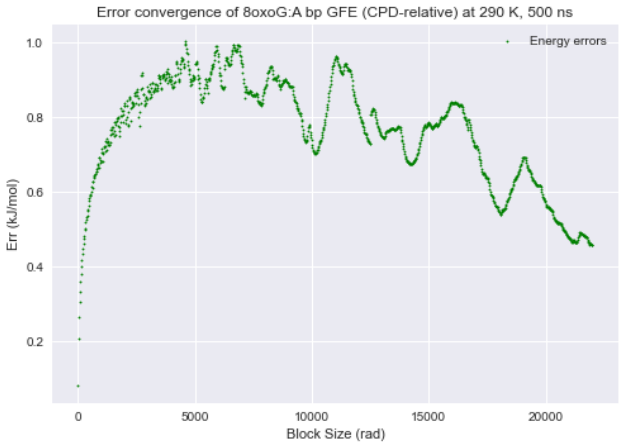


Fig. XIV

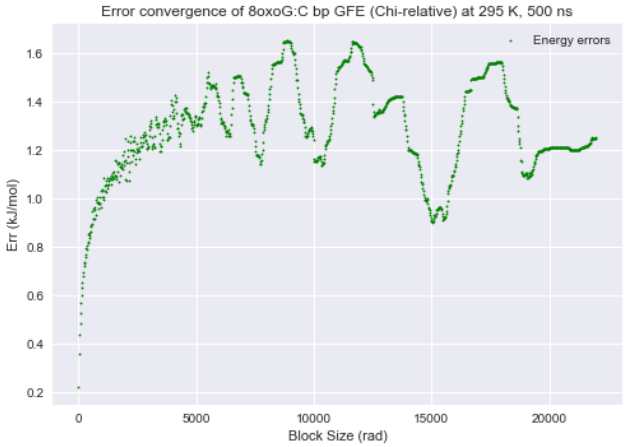


Fig. XV

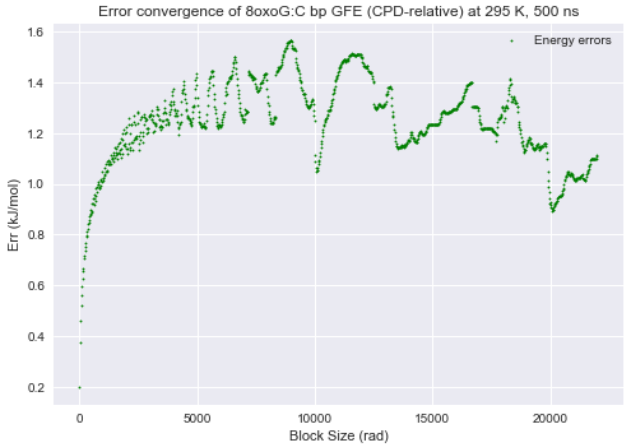


Fig. XVI

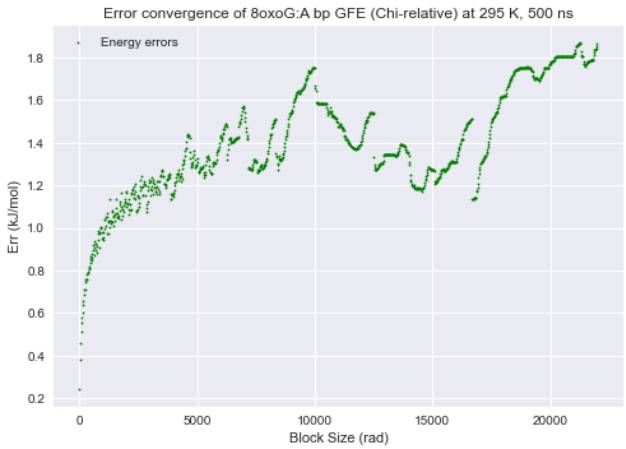


Fig. XVII

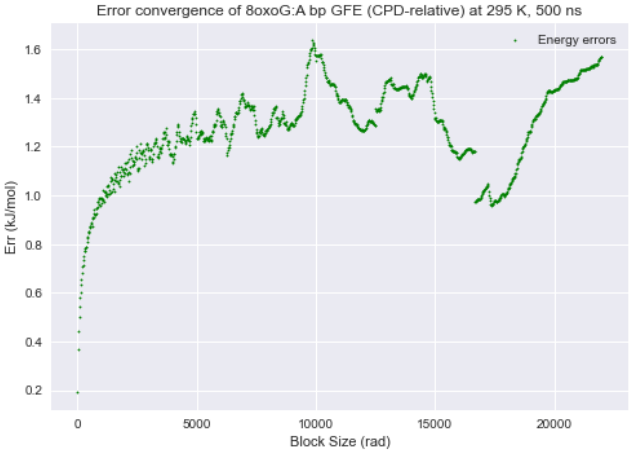


Fig. XVIII

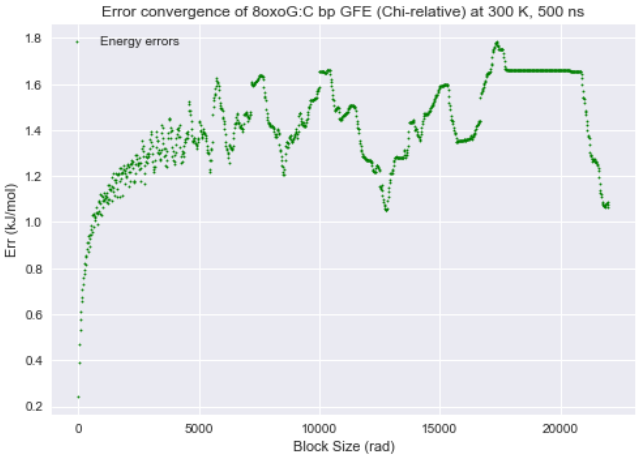


Fig. XIX

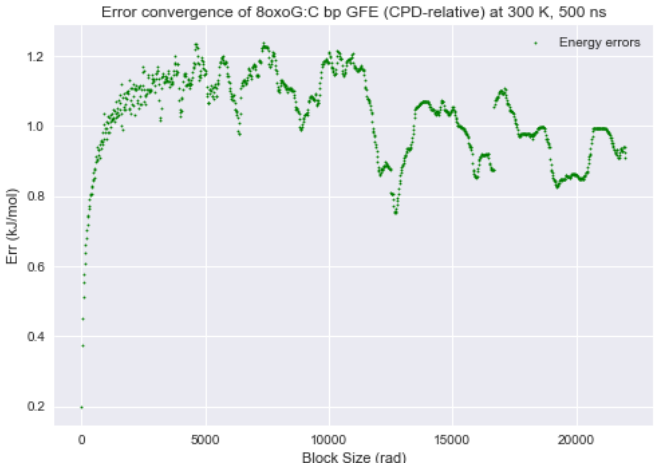


Fig. XX

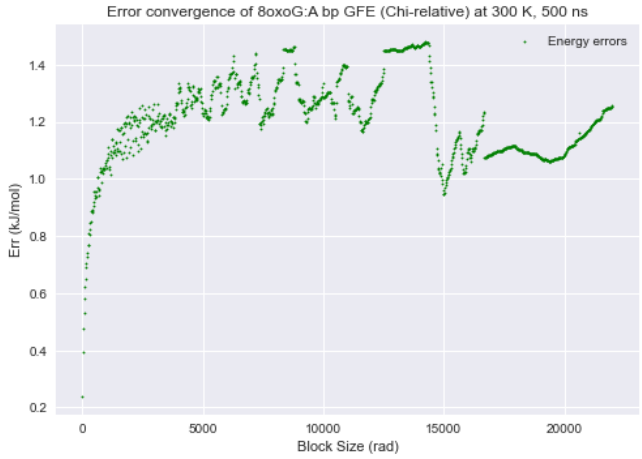


Fig. XXI

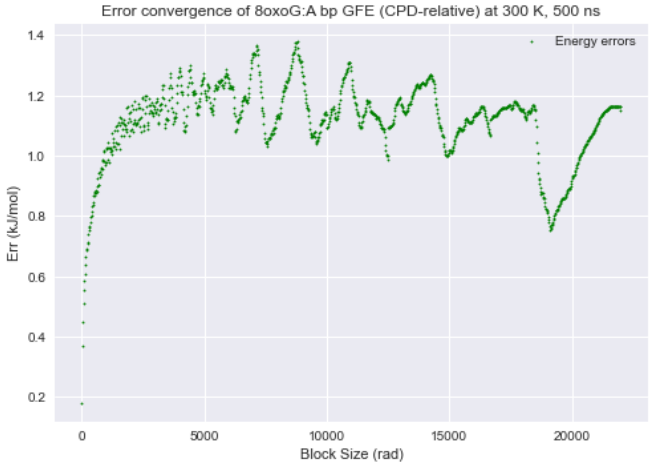


Fig. XXII

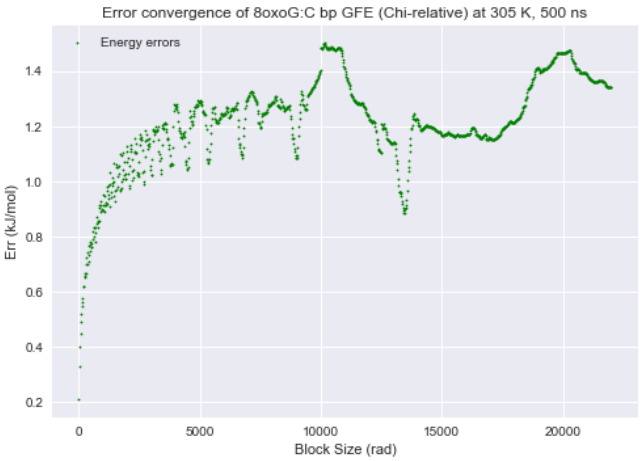


Fig. XXIII

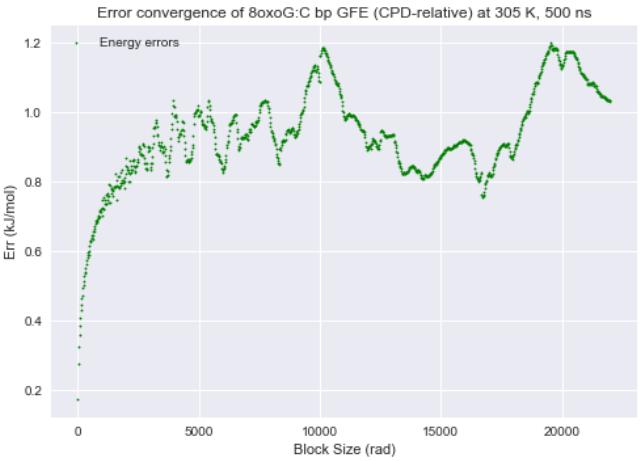


Fig. XXIV

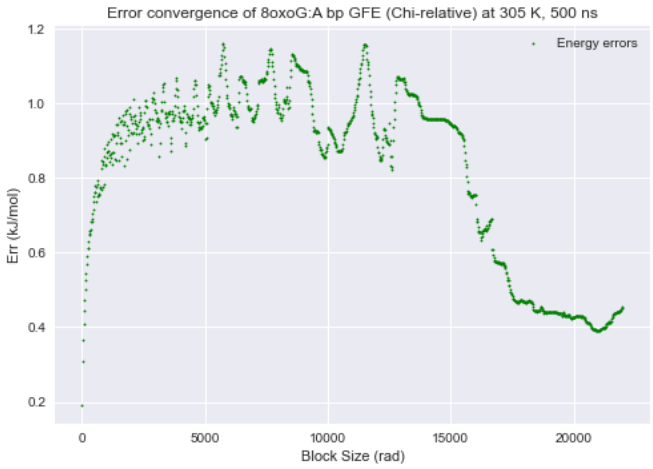


Fig. XXV

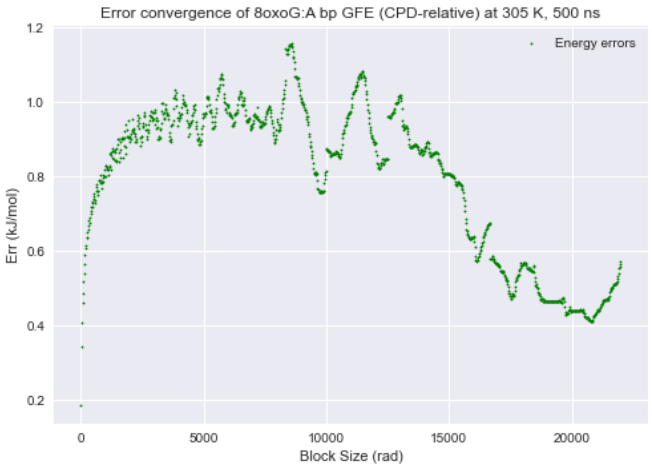


Fig. XXVI

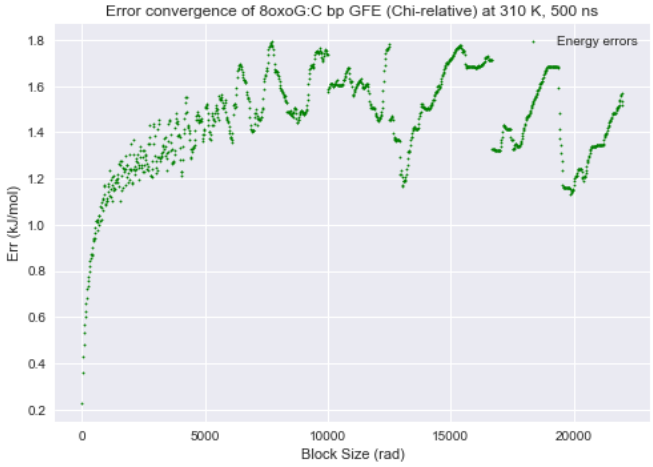


Fig. XXVII

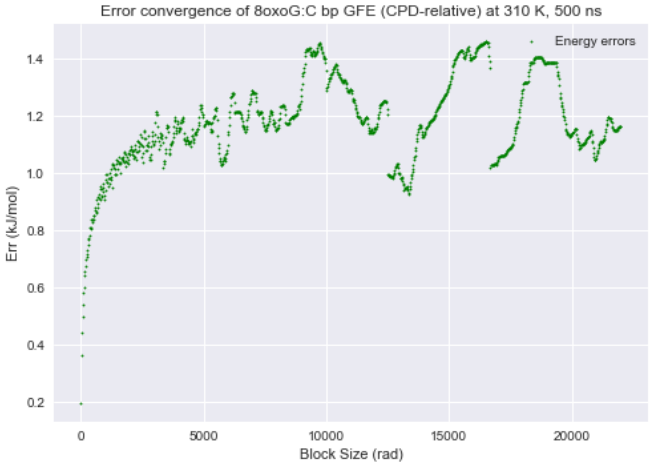


Fig. XXVIII

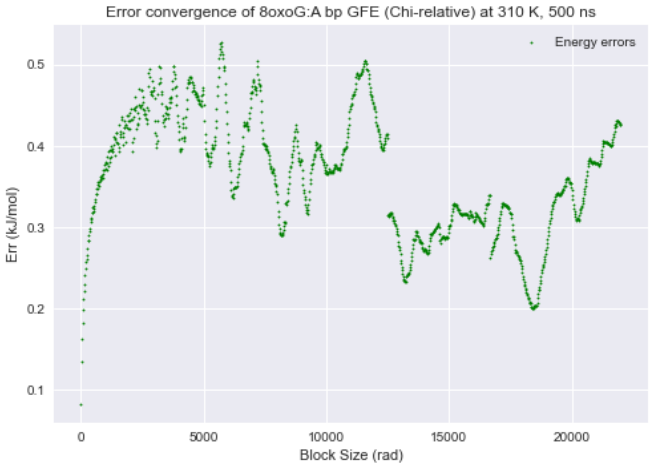


Fig. XXIX

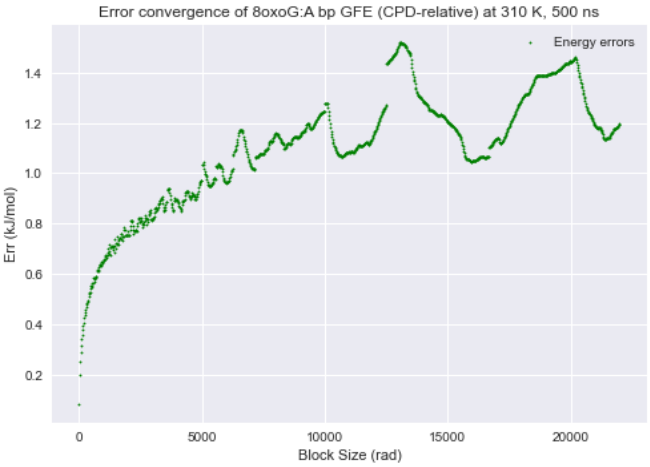


Fig. XXX

Fig. XI – XXX: GFE Error for each simulation. “Block size” refers to the size of the blocks (in rad) used in the error analysis to achieve these results. The error between angles was averaged to attain the respective error for each temperature.

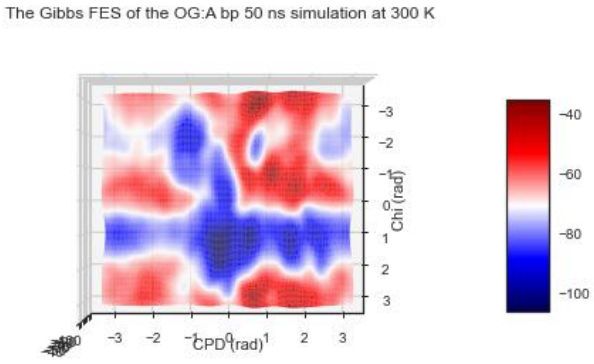


Fig. XXXI

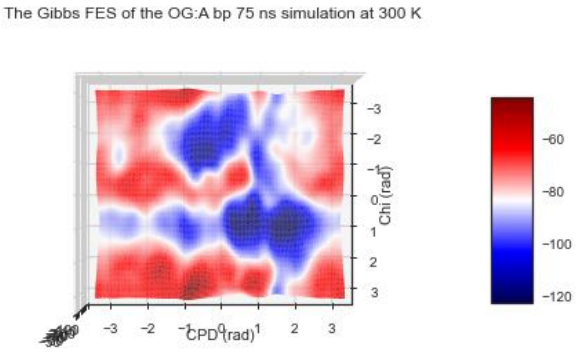


Fig. XXXII

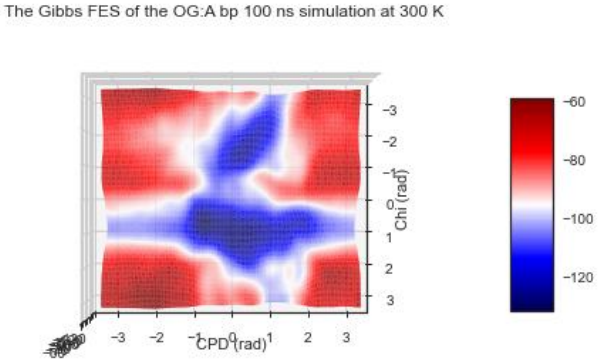


Fig. XXXIII

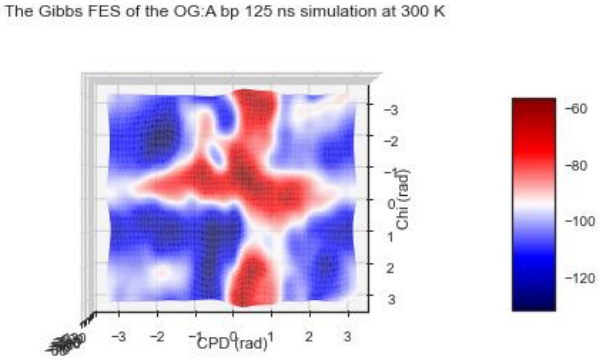


Fig. XXXIV

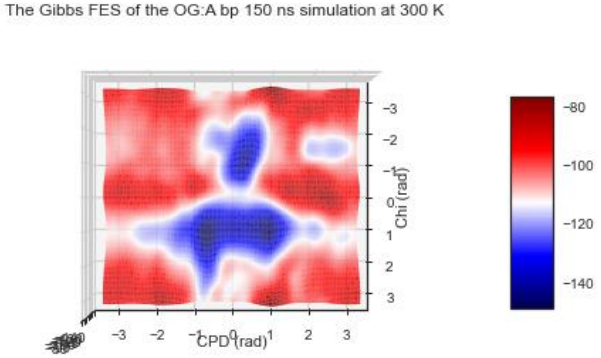


Fig. XXXV

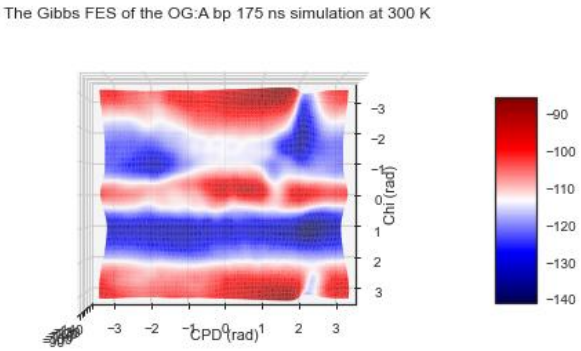


Fig. XXXVI

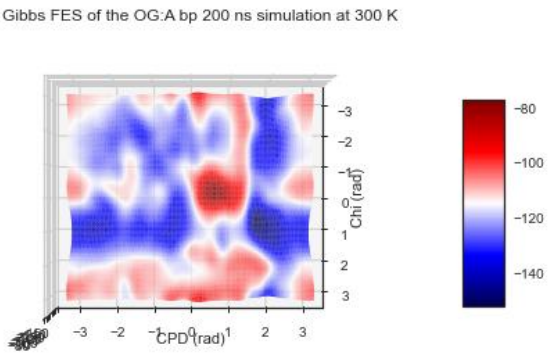


Fig. XXXVII

Fig. XXXI – XXXVII: Gibbs Free Energy surfaces of the OG:A system at 27°C simulations with various times. These graphs were used to approximate the appropriate time-length for simulation convergence.

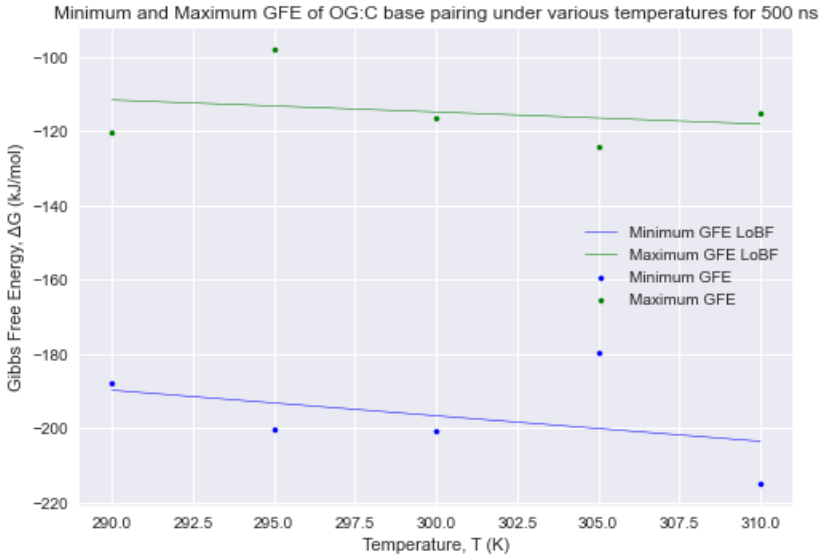


Fig. XXXVIII: Minimum and maximum GFEs of OG:C system at various temperatures. The lines of best fit (LoBF) for each data set are shown as well, assuming linear relationship to temperature by Eq. 4. R-squared values are 0.38 and 0.27 for maximum and minimum data respectively. Error bars are too small to be significant in visual format.

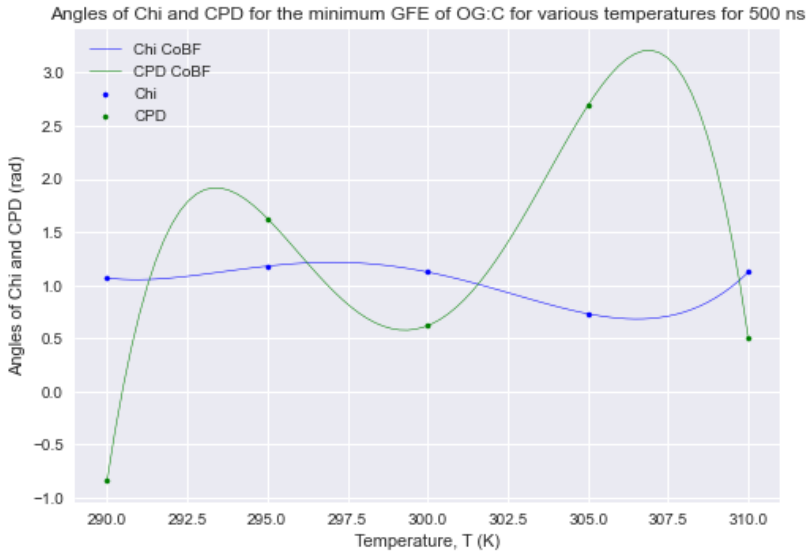


Fig. XXXIX: Chi and CPD angles of minimum GFE in OG:C systems at various temperatures. Curves of best fit (CoBF) are also shown. Both were taken to be 4th-degree polynomials and were chosen based on the best fit possible with the lowest degree. Their equations are $\theta_{chi} = .0000860T^4 - 0.103T^3 + 45.9T^2 - 9110T + 678813$ and $\theta_{CPD} = -0.000928T^4 + 1.11T^3 - 500T^2 + 100000T - 7491151$

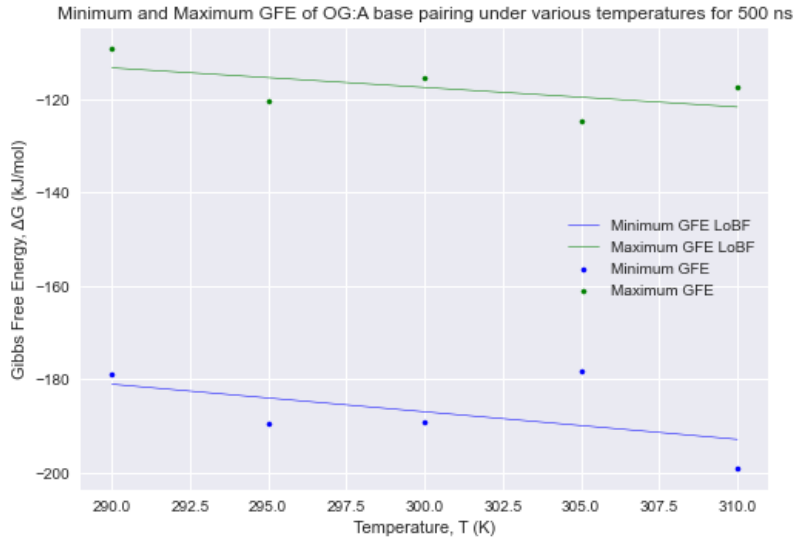


Fig. XXXX: The minimum and maximum GFE of OG:A base pairing under various temperatures for 500 ns, with their corresponding lines of best fit (LoBF). The R-squared values for the maximum and minimum lines are 0.57 and 0.53 respectively. This shows decent correlation with the minimum GFE from standard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

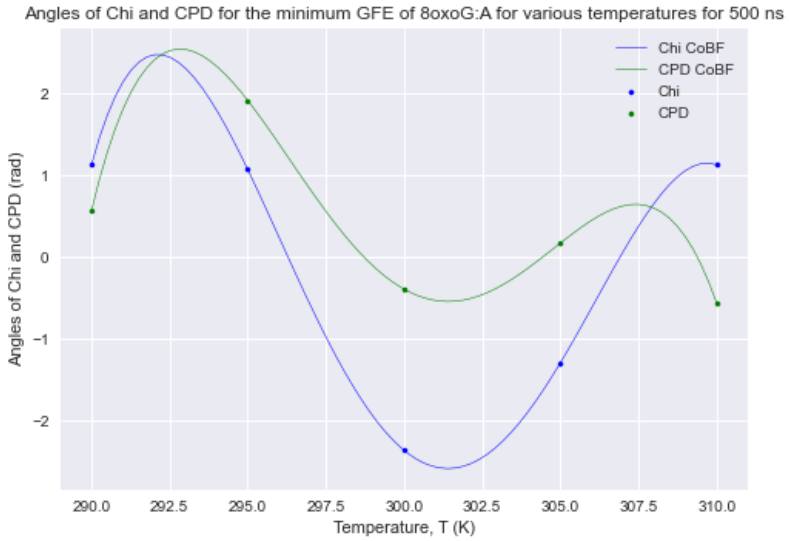


Fig. XXXXI: The angles of Chi and CPD for the minimum GFE of OG:A for various temperatures for 500 ns, with their respective curves of best fit (CoBF). The curves are 4th degree polynomials and were calculated as such to ensure consistency with previous analytical methods. θ_{chi} = -0.000733T^4 + 0.883T^3 - 399T^2 – (8.00×10^4)T - 6012614 and θ_{CPD} = -0.000711T^4 + 0.854T^3 - 385T^2 + 77100T - 5791446

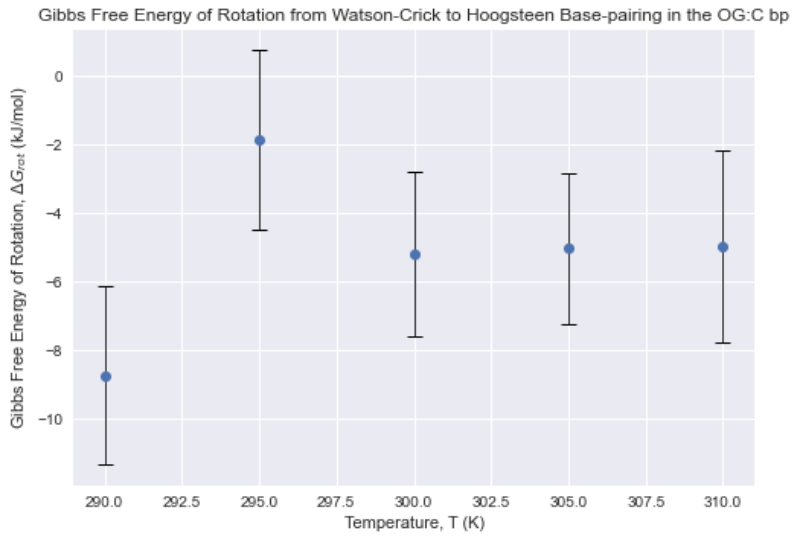


Fig. XXXXII: ΔG_{rot} of the OG:C system at various temperatures, intended as a visual representation of Table 3.

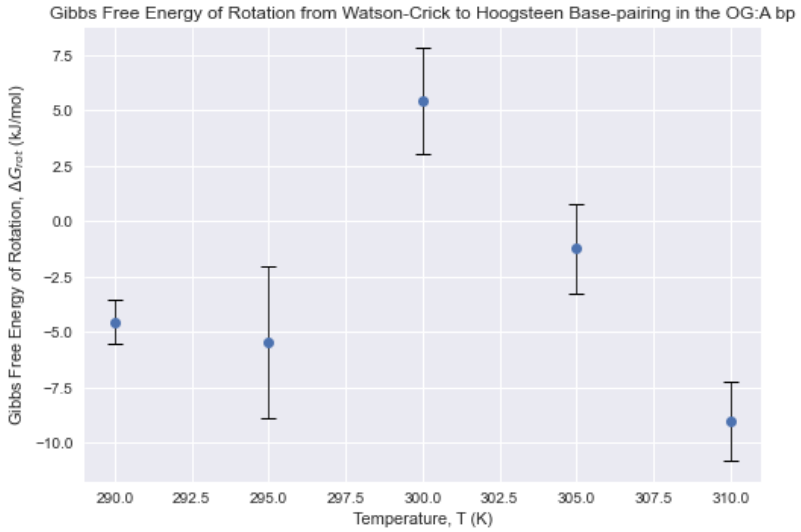


Fig. XXXXIII: ΔG_{rot} of the OG:A system at various temperatures, intended as a visual representation of Table 4.

TABLES

Table 1: Temperature-dependent energetics of the OG:C system*

Temperature (K)	GFE_{min} (kJ/mol)	GFE_{max} (kJ/mol)	Chi_{min} (rad)	CPD_{min} (rad)	Chi_{max} (rad)	CPD_{max} (rad)
290	-187.7 ± 1.3	-120.3 ± 1.3	1.07	-0.841	2.86	0.729
295	-200.2 ± 1.3	-98.1 ± 1.3	1.18	1.62	2.58	-0.168
300	-200.7 ± 1.2	-116.4 ± 1.2	1.12	0.617	-0.561	1.35
305	-179.9 ± 1.1	-124.1 ± 1.1	0.729	2.69	2.97	3.09
310	-215.1 ± 1.4	-115.4 ± 1.4	1.12	0.505	-2.97	2.58

Table 2: Temperature-dependent energetics of the OG:A system*

Temperature (K)	GFE_{min} (kJ/mol)	GFE_{max} (kJ/mol)	Chi_{min} (rad)	CPD_{min} (rad)	Chi_{max} (rad)	CPD_{max} (rad)
290	-178.8 ± 0.5	-109.3 ± 0.5	1.12	0.561	2.92	0.898
295	-189.5 ± 1.7	-120.4 ± 1.7	1.07	1.91	-0.673	0.505
300	-189.3 ± 1.2	-115.5 ± 1.2	-2.36	-0.393	-0.168	1.51
305	-178.3 ± 1.0	-124.8 ± 1.0	-1.29	0.168	-0.224	0.449
310	-199.2 ± 0.9	-117.5 ± 0.9	1.12	-0.561	2.64	0.673

*Energy errors were estimated based on the convergence in the error graphs made with PLUMED.

See Supporting Info (Fig. XI-XXX)

Table 3: GFE of Rotation (OG:C) from WC to HG bp and their associated Chi angles**

Temperature (K)	ΔG_{rot} (kJ/mol)	$\theta_{chi\ (WC)}$ (rad)	$\theta_{chi\ (HG)}$ (rad)
290	-8.7 ± 2.6	1.57	1.07
295	-1.9 ± 2.6	1.57	1.18
300	-5.2 ± 2.4	1.57	1.12
305	-5.0 ± 2.2	-1.74	0.729
310	-5.0 ± 2.8	1.57	1.12

Table 4: GFE of Rotation (OG:A) from WC to HG bp and their associated Chi angles**

Temperature (K)	ΔG_{rot} (kJ/mol)	$\theta_{chi\ (WC)}$ (rad)	$\theta_{chi\ (HG)}$ (rad)
290	-4.7 ± 1.0	-1.80	1.12
295	-5.5 ± 3.4	-3.03	1.07
300	5.4 ± 2.4	-2.36	-1.01
305	-1.3 ± 2.0	-2.13	-1.29
310	-9.0 ± 1.8	1.57	1.12

**Energies were chosen by finding the minimum FE for each bp type and using their corresponding values for ΔG_{rot} and θ_{chi}