Conformers of Substituted Cyclohexane Rings

A Validation of Electronic Structure and Molecular Dynamics Methods

Jonathan Chan, Joseph Chang, Sahaan Sarveswaran, Ali Shoja

Electronic structures method (ES) and molecular dynamics method (MD) were compared using recreated empirical data of substituted cyclohexane systems. Two main comparisons were made: the energies and conformer structures of the cyclohexanes, as well as the ratios between chair and twist-boat conformers. This was done through four different substituted cyclohexanes: cyclohexane (CY), 1-methylcyclohexane (MCY), 1-phenyl-1-methylcyclohexane (PMCY), and cis-1,4-di-tert-butylcyclohexane (DtCY). In ES, Gaussian was used to calculate the seven conformer structures and energies of each molecule, including their transition states. The chair to twist-boat ratios were calculated using thermodynamics equations using the resulting energies. In MD, Gromacs minimization was used to determine the optimized structures and their energies, while simulations of 1000 CY and MCY molecules were used to calculate the chair to twist-boat ratios. Gaussian was able to replicate literature results in its structural and energy determination for all four molecules. Calculated chair to twist-boat ratios further supported those results. Gromacs minimization was only able to produce correct structures for CY and MCY, though their energies were of unexpected values. As for the simulations, although chair and twist-boat conformers were determined to be non-dichotomous in the simulation, it was able to accurately determine the preferred position of the methyl substituent when the correct conformers were isolated as inputs. Overall, ES Gaussian was able to produce accurate results through Gaussian, but due to challenges in MD's Gromacs, the data between the two methods were not thoroughly compared.

Introduction

There are many aspects of chemistry which are difficult, and sometimes even impossible, to observe experimentally. An example would be transition states of molecules which cannot be physically isolated and examined using physical analytical methods. In these situations, computational methods can be employed to simulate reality in aid of scientific discovery and deeper understanding of the universe. Two well-known computational methods are the electronic structure method (ES) and molecular dynamics method (MD). Both of these methods are often used to determine local or global minimum energies in molecules and atoms, which are useful not only for determining chemical structures, but even chemical interactions through simulations.

ES methods use approximations of solutions to Schrodinger's equation to compute the energy of a system. Within the electronic structure method exists many sub-methods that calculate the Schrodinger equation differently. One of the more popular methods is the Density Functional Theory (DFT), which as the name suggests, uses electron density as its means of calculation. As for the MD method, it relies on simplified models of electronic potentials that can be computed repeatedly to simulate the movement of particles, especially under the effects of surrounding particles. In other words, MD uses classical mechanics as opposed to ES's quantum mechanics approach.

Given that both methods aim to accurately represent reality using fundamentally different approaches, it is important for them to reach the same conclusion for any given problem. An interesting system one might use to compare the methods is cyclohexane conformations.

1

Cyclohexanes are a well studied conformational system given its relatively many possible bond angles that are reasonably stable. Yet, its symmetry allows it to fall neatly into two major stable conformations that are distinct and recognizable. This allows for more than one point of comparison within one given system.

Here, we aim to recreate empirical data by studying cyclohexane systems using ES and MD computational methods to show that both methods are indeed reliable as a reflection of reality. Specifically, the methods will be compared by their recreation of several substituted cyclohexane's conformer energies and structures, as well as the methods' determinations of chair to twist-boat ratios.

Methodology

Cyclohexanes

Four cyclohexanes were chosen for this study. Cyclohexane (CY) and 1-methylcyclohexane (MCY) were chosen as baseline molecules, whereas 1-phenyl-1-methylcyclohexane (PMCY) and cis-1,4-di-tert-butylcyclohexane (DtCY) were chosen for their unique properties.

For PMCY, the most stable conformation is a chair with the phenyl substituent in the axial position.[1] This may seem counterintuitive at first when rationalized using the steric bulk of the substituent; a bulky group should prefer the equatorial position to reduce steric interaction with the cyclohexane ring. Though through further analysis and as shown in this study, this is true. On the other hand, DtCY was shown to be disfavoured when it is in a chair conformation, which most other cyclohexane derivatives prefer. Instead, DtCY is most stable as a twist-boat.[2]

Electronic Structure Energy Calculations

We used gaussian[3] to find the potential energy surface of the conformations, starting with the transition state (TS). First, we will need to find structures very close to the TS's structures via geometry optimization with constraint.[5] If we perform optimization without constraint, we would go all the way down to a more locally stable well. So, we fixed the dihedral angle of carbons on the ring and used the B3LYP/6-31G level of theory on all molecules to perform the geometry optimization.

Next step is to find the TS. We used the TS keyword and frequency option to find and validate TSs on gaussian. There are several ways to validate TSs, one of which is to check for negative frequency vibration. Negative frequency is an imaginary frequency that is related to the vibration that the system wants to release. The presence of this signifies the presence of a TS. The other way to check the TS is by visual confirmation; the structure should make sense as a TS and it must be higher in energy compared to the starting material and product (Figure 1).
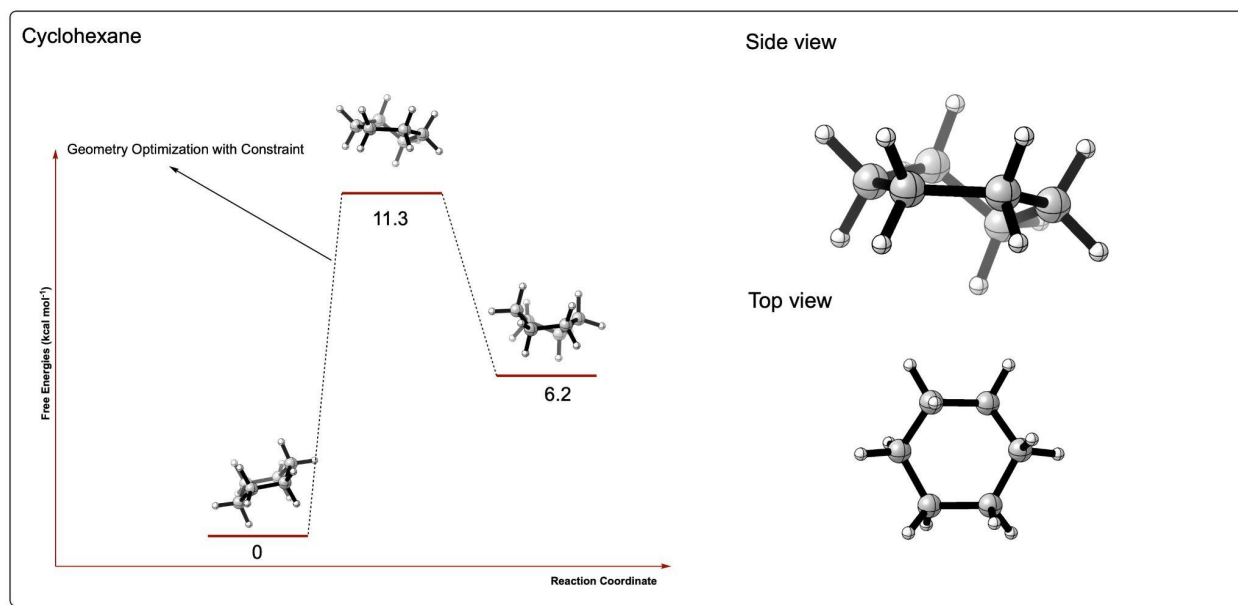
Figure 1. Transition state calculation steps

But, the best way to validate the TS is using IRC calculation[5] (Figure 2). In IRC calculation, the system tries to find the wells by going through both directions. This can give us very close geometres to the chair and twist-boat conformers, but it won't go all the way down to the minima. So, by using the two last points on the IRC calculation and optimizing them to get to the final chair and twist-boat structure, we can validate the TS.
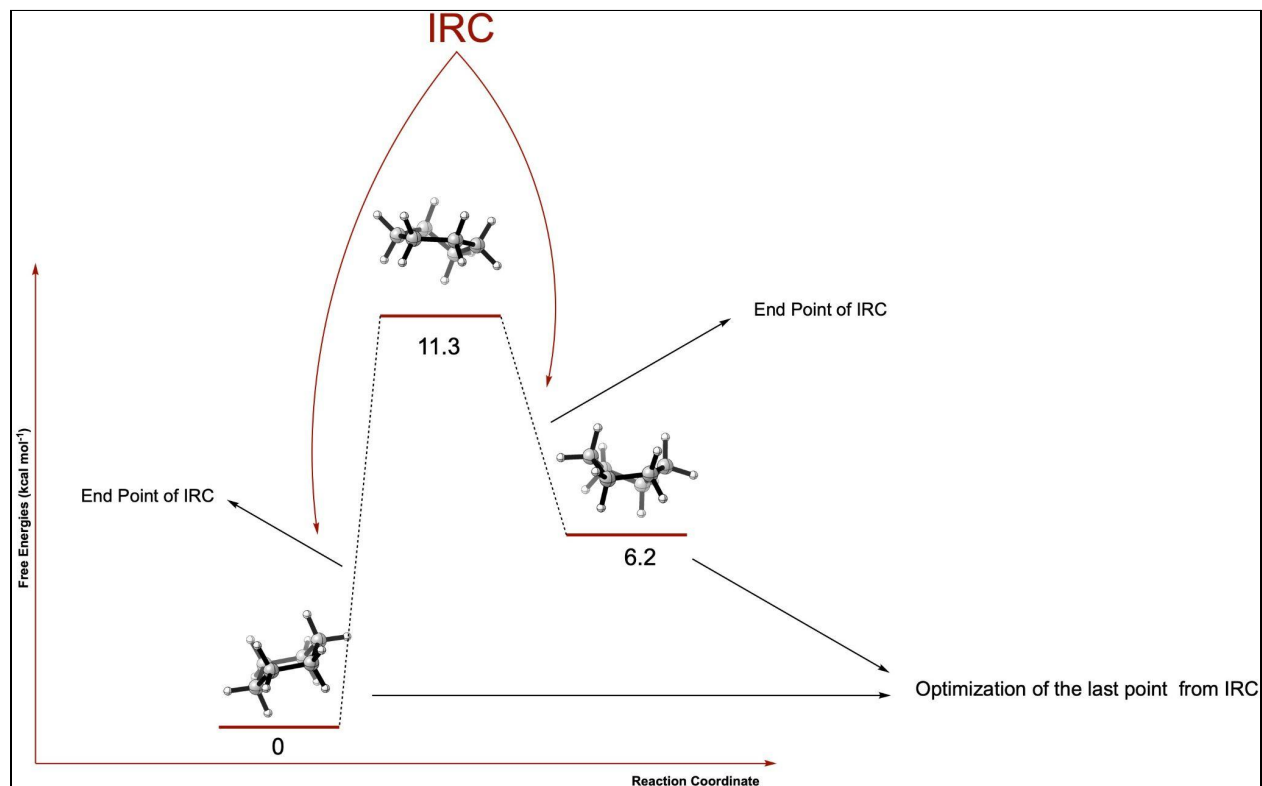

Figure 2. IRC calculation

Once all the conformer and their energies have been determined, the expected equilibrium constant can be determined with the Gibbs free energy equation for chemical equilibria:

$$K \ = \ e^{\frac{-\Delta G}{RT}} \tag{1}$$

where K is the equilibrium constant, R is the gas constant, T is the temperature, and ΔG is the energy difference between two states. The K constant in turn provides information on the ratios of each state.

Molecular Dynamics Energy Calculations

Energies in molecular dynamics were calculated through the minimization function in Gromacs, which finds the optimized structures of a given input.[6] To initiate, the structural file, topology file, and the MD parameters files were prepared. The structural input file is a pdb file from Gaussian optimization runs, which provides the ground state structure for the desired conformer. This helps Gromacs when it is looking for structures that are local minima rather than global minima. The topology files for PMCY and DtCY were generated with the Jorgensen Lab's LigParGen OPLS/CM1A Parameter Generator for Organic Ligands at Yale.[7] The remaining two topology files were generated with Gromacs' x2top. As for the MD parameters file, all three integrators for minimization (steepest descent, conjugate gradient, l-bfgs) were tested, with the coulomb type setting set to PME and periodic boundary conditions set to all directions.

Energies and optimized structures were obtained from energy minimization. Though Gromacs optimized energy values are not directly comparable to the ones from Gaussian, the energy differences between each conformers are. Consequently, this can then be for comparison to ES's results. The subsequent structures were qualitatively evaluated.

Molecular Dynamics Geometry Analysis

To show that simulation not only calculates the free energy difference correctly, but simulates the equilibrium ratios as expected, we compared the ratios from the simulation to those expected from the Gibbs free energy equation. We simulated CY and MCY in the gas phase and sorted them into the appropriate conformers with geometry analysis.

While there is empirical conformational data for CY up to 800 K,[8] we recognize that a limitation of geometry analysis is that there is low precision and that small variations in geometry identification can have a relatively large error when measuring a small quantity of conformer. Thus, we simulated the molecules at an exceedingly high temperature that will induce the relative quantity of the less favorable conformer to increase, decreasing the impact of error. It is important to note that these conditions are not possible in reality, as hydrocarbons begin decomposing beyond about 800 K.[9] However, this will show that simulations can isolate the effects of intermolecular and intramolecular interactions from bond potentials yet, still be accurate in principle.

A simulation with 1000 molecules was chosen because of the convenience of setting up such a starting input and because of computational resource considerations, with more molecules adding precision but using more time to process. A frame from the simulation after reaching equilibrium was analyzed with geometric analysis to provide direct evidence of the presence of

each conformer. We used the Gromacs package[6] to perform these molecular dynamics simulations.

Importantly for geometric analysis, we must recognize that the defining geometry of each conformer comes from whichever geometry that results in the lowest energy locally and not on any other idealized geometry. For this algorithm, we relied on characteristics found in an DFT optimized model of the ground state structure as parameters for analysis. In the interest of time, of many identified characteristics, we focused on one criterion: the alternating direction of C-H axial bonds. Two main algorithms are needed in the analysis program: One that identifies whether a CY is a chair or twist-boat, and one that identifies whether a substituted CY that is a chair has the substituted group in the equatorial or axial position.

We wrote an analysis algorithm in Python 3.[10] The program takes output coordinate files and the connectivity information in a topology file as inputs. We started by encapsulating data such that bond connectivity such as C-H bonds, and carbons along a CY ring could be systematically recognized as one object.

To begin our analysis, we identified axial C-H bonds, or more accurately, a list of possible axial bonds. In principle, all axial bonds point either "directly up" or "direction down" relative to the CY ring, in contrast to equatorial bonds that point outward. This means that there must exist a set of six bonds such that their vectors are parallel to one another. We accounted for variation by tolerating angles within a range given as an input. This identification was done recursively with an arbitrary search tree. The breadth of which represents every bond as a vector, and the depth of which represents each of the six carbons around a CY ring.

The next step was to recognize relative direction. We did this by comparison of axial vectors to each other, given that parallel vectors can only have two distinct directions. The ring conformation function takes the output and attempts to find at least one set of vectors that alternate directions as axial bonds should. The substituent conformation function then takes the ring conformation function's outputs that are determined to be chair, and attempts to find at least one set of vectors that alternate with respect to the bond of the substitution. This determines whether the substituent is in the axial or equatorial position. Each function will "mark" a molecule as being one in conformation or the other.

As stated earlier, many geometric "parallel" judgements depend on a given acceptable range to account for imperfect geometries. We wanted to be sure that the ratio obtained was independent of the given tolerance. Thus, the final part of the program runs the algorithm on every molecule, at every angle of tolerance from 0° to 90°, giving the ratio at each angle in integers with each iteration, and as a graph at the end.

A function was written to output the selections made by the algorithm as viewable molecule coordinate files to qualitatively confirm the efficacy of the geometry algorithm.

Results and Discussion

Electronic Structure Energy Calculations

We calculated the full potential energy diagrams for all four substituted cyclohexanes. For CY (Figure 3), since it has a symmetrical structure, we got two identical TSs and conformers. The half-chair TS has its four carbons of the ring on the same plane and as a result, the hydrogen-hydrogen steric interactions increase their energy around 11.3 kcal/mol compared to the chair structure. The boat TS also has four carbons of the ring on a same plane, but since there's less steric interaction between hydrogens, it is lower in energy compared to the half-chair TS.
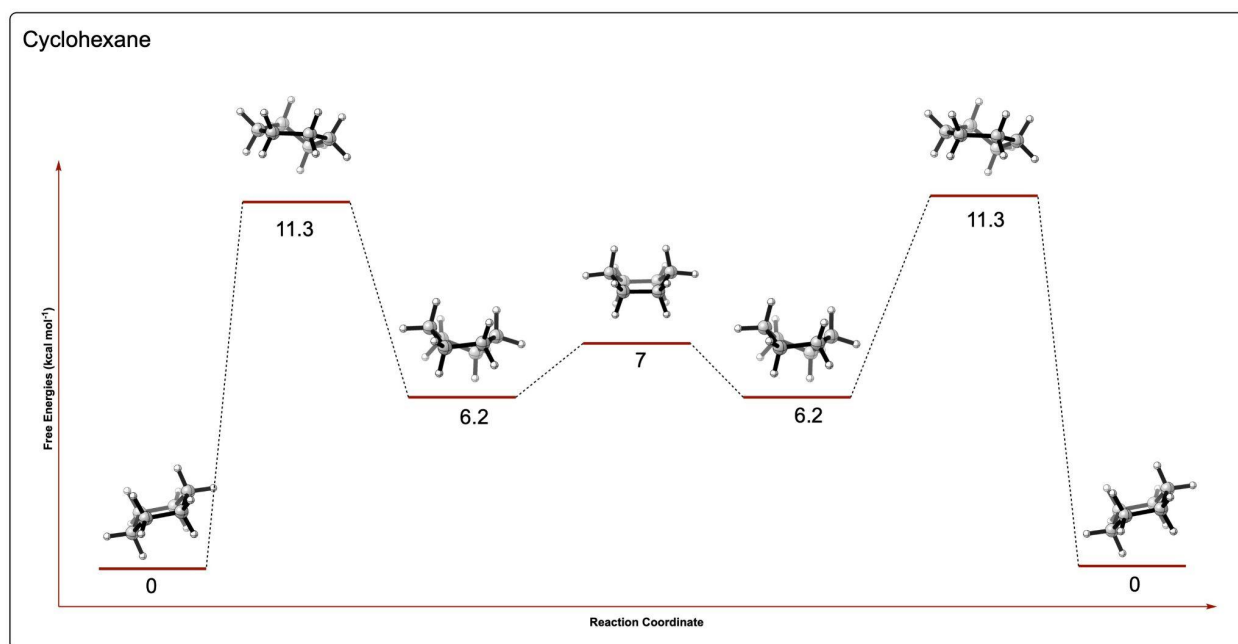
Figure 3. Potential energy diagram for Cyclohexane

In the second case, we calculated the energies of MCY (Figure 4). It is more complicated compared to CY since it has less symmetrical structures, meaning we had to calculate all the possible conformers. As we expected, the methyl group prefers to be in equatorial position since it is larger than hydrogen and has more 1,3 interactions with other hydrogens.
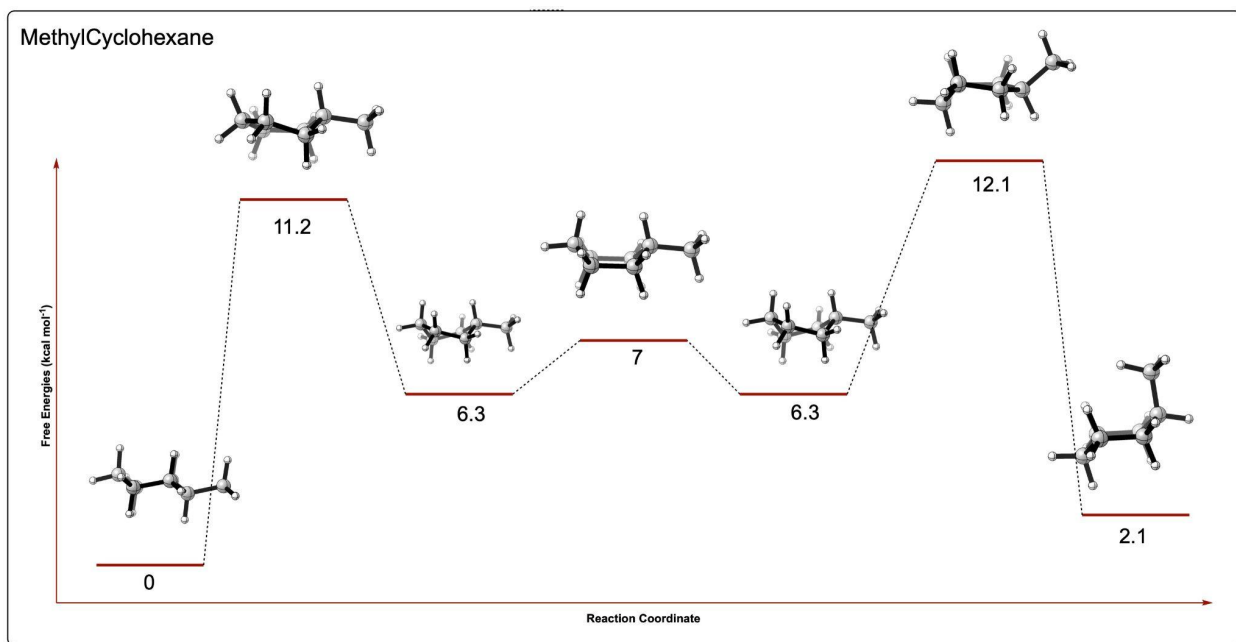
Figure 4. Potential energy diagram for MethylCyclohexane

For the next molecule, we tried an interesting cyclohexane derivative. Based on the experimental data, cis-1,4-di-tert-butylcyclohexane prefers twist-boat structure due to the greater 1,3 sterics interaction compared to the simpler MCY in its chair structure (Figure 5). On the other hand, in the twist-boat conformers, both tert-butyl groups are outside of the ring, minimizing the repulsive interaction with hydrogens.
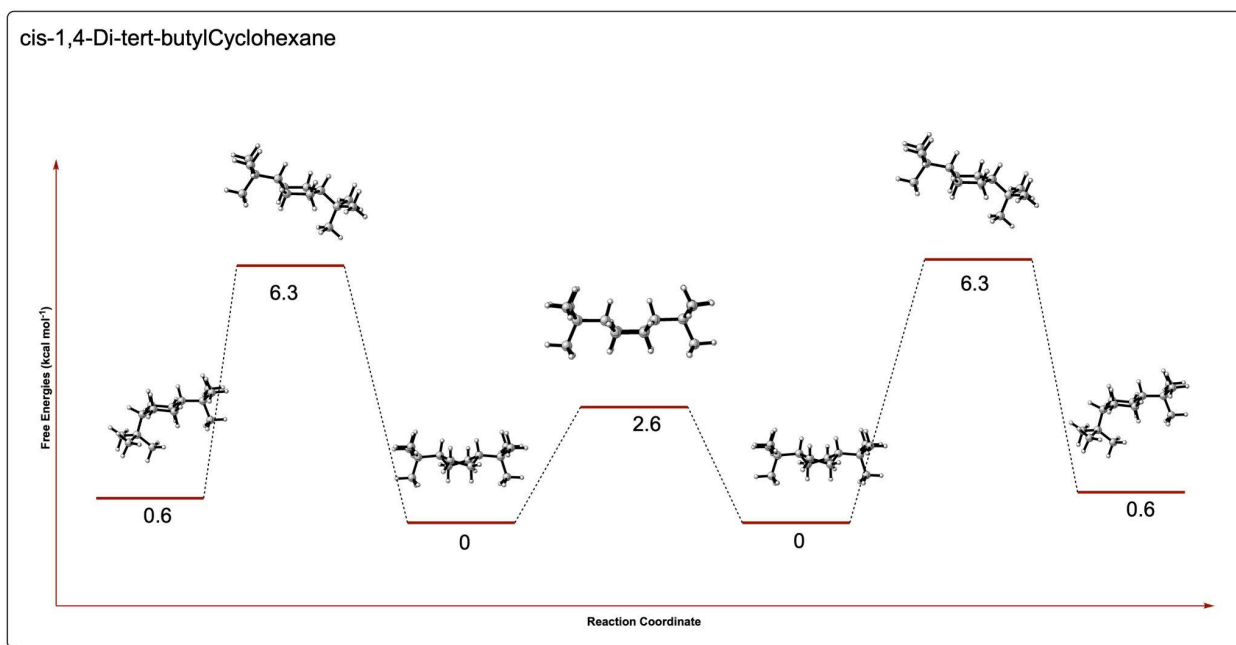

Figure 5. Potential energy diagram for cis-1,4-Di-tert-butylCyclohexane

For the last case of study, we chose PMCY which phenyl prefers the axial position and breaks the trend of equatorial preference. It is due to the steric interaction between methyl and phenyl

groups when phenyl is on equatorial. And again our prediction is consistent with experimental data (Figure 6).
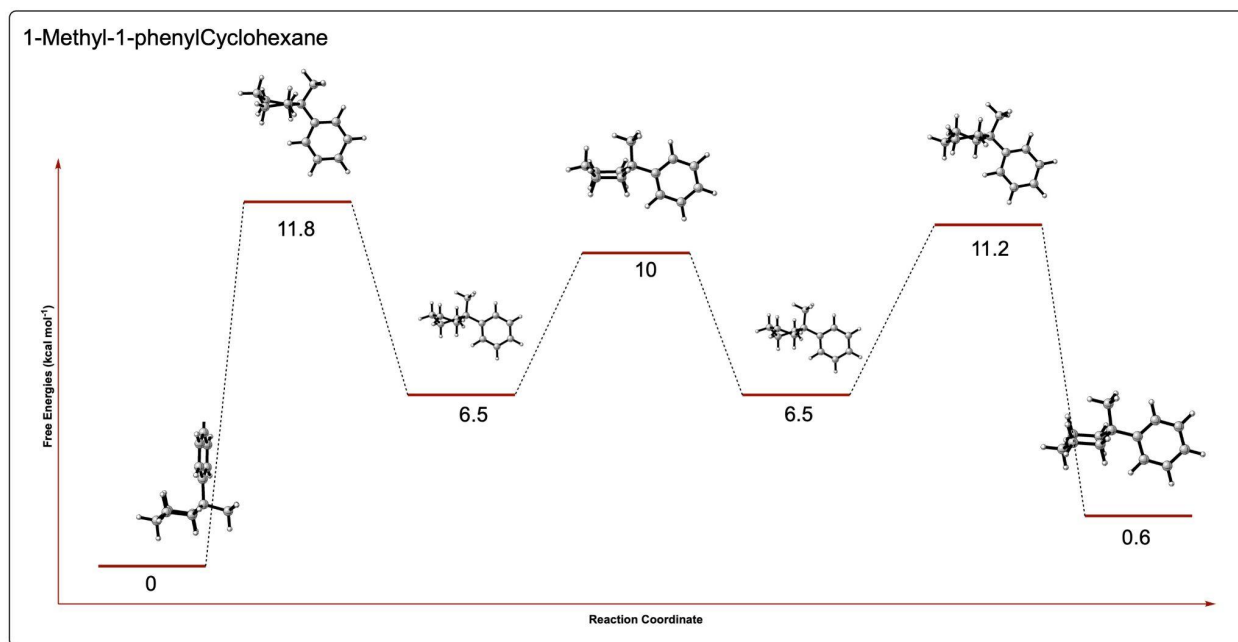

Figure 6. Potential energy diagram 1-methyl-1-phenylCyclohexane2.46

Tabulated in Table 1 are the structural results as well as the calculated chair to twist-boat ratios from Equation (1). It is apparent that the preferred conformers shown in the figures above correspond to the highest ratio number, further supporting the validity of Gaussian's calculations.

Table 1. Gaussian output summary.

	Chair to Twist-Boat Ratio (Chair : Twist-Boat)	Preferred Conformer
CY	60.0 : 40.0	Chair
MCY	56.9 : 43.1 | 60.2 : 39.8 (Ax | Eq)	Chair (Eq)
PMCY	60.5 : 39.5 | 59.6 : 40.4 (Ph-Ax | Ph-Eq)	Chair (Ph-Ax)
DtCY	49.0 : 51.0	Twist-Boat

Molecular Dynamics Energy Calculations

Our initial energy minimization calculation trials in Gromacs[6] did not yield correct structures. Whether the input .pdb structure was a chair or twist-boat conformer, the minimized structure would be half-chair-like. We knew this because the resulting structures from Gromacs closely resembled the half-chair structures generated in Gaussian (Figure 7, 8). To investigate the source of the error, we changed the integrator in the .mdp file from the steepest descent (steep) algorithm to conjugate gradient (cg) and then to quasi-Newtonian (l-bfgs). The same problems

occurred with all three algorithms, so we concluded that the error was likely from something other than the .mdp files.
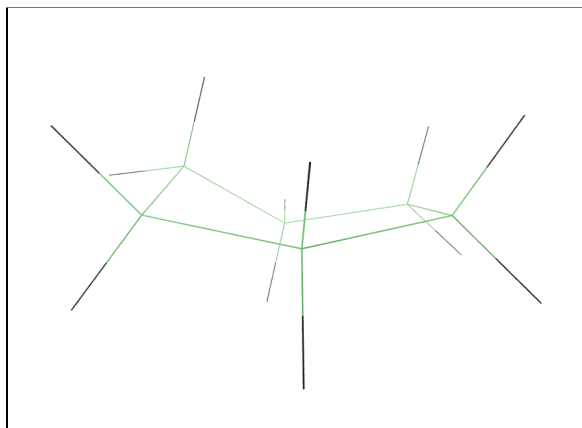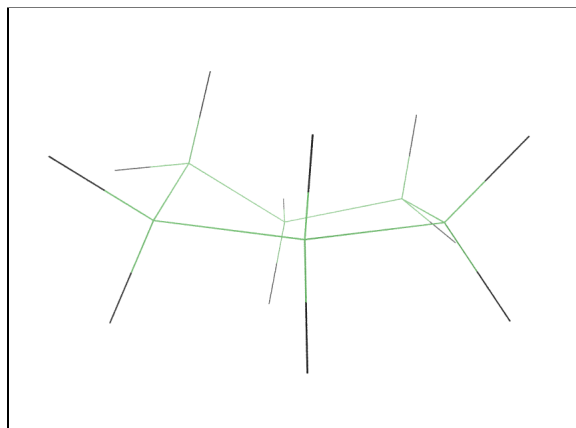

Figure 7. Gromacs minimized structure.

Figure 8. Gaussian generated half-chair

It is also worth noting that the energy values from these trials were not accurate either. The energy differences between conformers were less than 1 kJ/mol, or less than 0.24 kcal/mol. This was much lower than the expected value of around 25 kJ/mol (or ~6 kcal/mol). The inaccurate results here were not too surprising, given that both conformers ended up producing half-chair like structures instead of staying as their respective conformers.

Next, we examined our topology files to see if something there was causing the errors. We found that our .top files included inputs under the c0, c1, c2, and c3 columns by default (Figure 9), which is usually supposed to be empty in newly generated files. The inputs acted as constraints which prevented our molecules from being affected by the selected OPLS force field. Essentially, the inputs were overriding the force field interactions, possibly causing the false minimization results.

```
[ bonds ]
;  ai   aj funct           c0             c1             c2             c3
    1    2    1  1.560000e-01  4.000000e+05  1.560000e-01  4.000000e+05
    1    6    1  1.570000e-01  4.000000e+05  1.570000e-01  4.000000e+05
    1    7    1  1.100000e-01  4.000000e+05  1.100000e-01  4.000000e+05
    1    8    1  1.100000e-01  4.000000e+05  1.100000e-01  4.000000e+05
    2    3    1  1.540000e-01  4.000000e+05  1.540000e-01  4.000000e+05
    2    9    1  1.100000e-01  4.000000e+05  1.100000e-01  4.000000e+05
    2   10    1  1.100000e-01  4.000000e+05  1.100000e-01  4.000000e+05
    3    4    1  1.540000e-01  4.000000e+05  1.540000e-01  4.000000e+05
    3   11    1  1.100000e-01  4.000000e+05  1.100000e-01  4.000000e+05
    3   12    1  1.100000e-01  4.000000e+05  1.100000e-01  4.000000e+05
    4    5    1  1.540000e-01  4.000000e+05  1.540000e-01  4.000000e+05
    4   13    1  1.100000e-01  4.000000e+05  1.100000e-01  4.000000e+05
    4   14    1  1.100000e-01  4.000000e+05  1.100000e-01  4.000000e+05
    5    6    1  1.560000e-01  4.000000e+05  1.560000e-01  4.000000e+05
    5   15    1  1.100000e-01  4.000000e+05  1.100000e-01  4.000000e+05
    5   16    1  1.100000e-01  4.000000e+05  1.100000e-01  4.000000e+05
    6   17    1  1.100000e-01  4.000000e+05  1.100000e-01  4.000000e+05
    6   18    1  1.100000e-01  4.000000e+05  1.100000e-01  4.000000e+05
```

Figure 9. Topology file with inputs in columns c0-c3.

The topology files were then revised to exclude the constraints, allowing OPLS an attempt to construct the correct structures. These changes provided us with middling results, as there were some improvements but still not perfect. The PMCY and DtCY structures were now unable to be minimized in Gromacs entirely, as .tpr files would not be produced from the .gro files at the grompp step, as the calculations would error. Thus, we could not yield energy values for those two cyclohexanes using MD methods.

Fortunately, calculations were successfully completed for the CY and MCY cyclohexanes. Additionally, the energy minimizations yielded the correct conformers, as the final .gro structures appeared as chairs and twist-boats as expected (Figure 10), instead of as half-chair like structures as they did before. The computed energy values are listed below in Table 2.
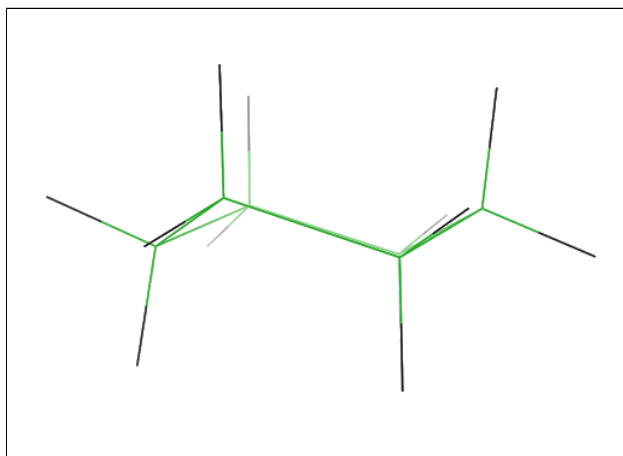

Figure 10. Output .gro file for CY Chair in correct conformation

Table 2. Energy differences from Gromacs minimization

	Conformer	Potential Energy (kJ/mol)	Energy Difference (kJ/mol)	Energy Difference (kcal/mol)
CY	Chair	21.3	8.37	2.00
	Twist-boat	29.6		
MCY	Axial Chair	1.23×10^3	702	168
	Axial Twist-boat	532		
	Equatorial Chair	16.0	9.32	2.23
	Equatorial Twist-boat	25.4		

From these results, it appears that the chair conformation is preferred over twist-boat for a regular (unsubstituted) cyclohexane, as it is lower in energy. This agrees with our result from ES methods.

It also appears that for an equatorial methyl substituted cyclohexane, that the chair conformer is again preferred over the twist-boat. This agrees with our ES method results as well. It is important to note too that the equatorial energy values are significantly lower than the axial ones for MCY. This was an expected result as it would be much better for sterics to have the methyl group in an equatorial position.

One potentially inaccurate and unreasonable result from the table above is the extremely high energy value calculated for both axial MCY molecules. Thus we made one more modification to attempt to improve the data. Specifically, this was by adding a constraint specification to all bonds in the MD configuration, and running the same energy minimization calculations with the constrained files. Without specifying constraints, bonds are given harmonic potentials. The changes produced the following results.

Table 3. Energy differences from Gromacs minimization, with constrained .mdp files

	Conformer	Potential Energy (kJ/mol)	Energy Difference (kJ/mol)	Energy Difference (kcal/mol)
CY	Chair	21.6	8.67	2.07
	Twist-boat	30.3		
MCY	Axial Chair	1.73×10^{10}	1.72×10^{10}	4.12×10^{9}
	Axial Twist-boat	7.79×10^{6}		
	Equatorial Chair	16.6	9.64	2.30
	Equatorial Twist-boat	28.3		

From these results, it appears that the added constraints did not do much to improve our results, and in fact, the axial MCY energy values are even more unreasonable. It is notable though that the same trends are otherwise observed, such as the chair being the preferred conformer for both CY and the equatorial MCY. The energy differences for these did not change by much either, as the conformers again differ by ~2 kcal/mol for CY, and ~2.2-2.3 kcal/mol for MCY.

Molecular Dynamics Geometry Analysis

The algorithm that processes CY and MCY simulations generally works as intended but is not error-free. Starting with chair conformer determinations, after manually sifting through molecules to find errors in conformer determination, a rare few examples were found to be sorted incorrectly. This is demonstrated in the bottom two structures in both Figures 11 and 12.
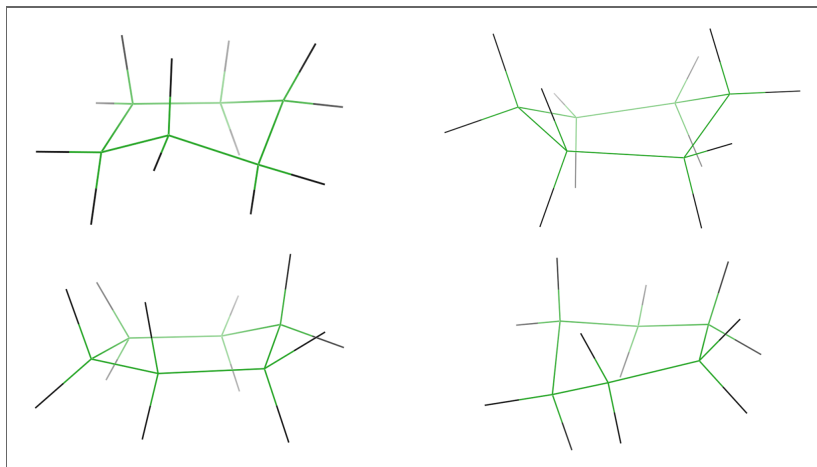
Figure 11. A sample of molecules algorithmically identified as chair conformers (left) and non-chair conformers (right) from a 1830 K simulation of CY with a 60° tolerance for identification. The upper structures are correctly assigned, while the lower structures are incorrectly assigned.
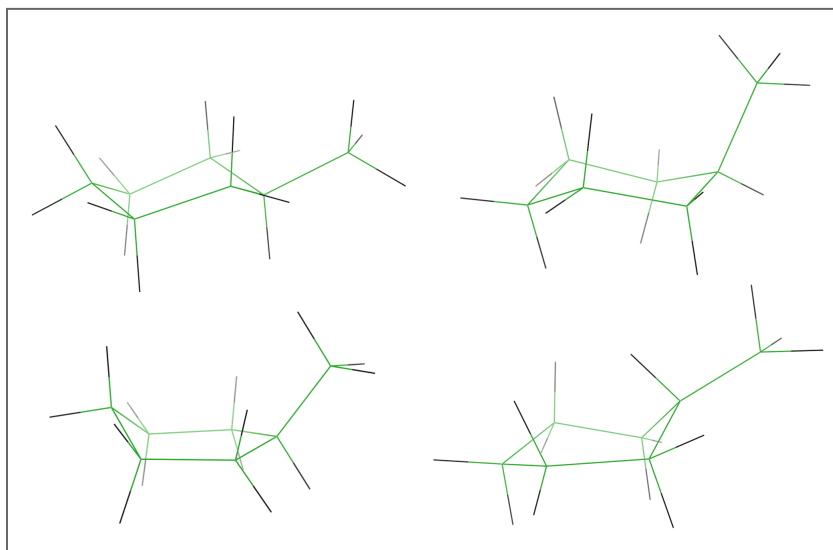

Figure 12. A sample of molecules algorithmically identified as equatorial conformers (left) and axial conformers (right) from a 1830 K simulation of MCY with a 60° tolerance for chair identification and a 60° tolerance for axial identification. The upper structures are correctly assigned, while the lower structures are incorrectly assigned.

In Figure 11, the lower left structure, a half-chair conformer is incorrectly identified as a chair conformer. Similarly, in Figure 12, the lower left structure, a substituted half-chair is incorrectly identified as an equatorial chair conformer. Both these cases deal with the half-chair conformers that are not only similar, but miniscule angles of rotation away from being chair conformers

In Figure 11, on the lower right, a chair is incorrectly identified as a non-chair conformer. Similarly, in Figure 12, on the lower right, an equatorial chair conformer is incorrectly identified as an axial chair conformer. It appears an extreme angle of rotation of hydrogen or a substituent relative to the ring is necessary for this determination to be made incorrectly.

The problematic structures are either conformationally similar to the point of being difficult to distinguish or conformationally extreme structures. These are examples of edge cases however,

13

and are not representative of the majority of conformer determinations. They are only a small source of error. After qualitative examination of a range of angles of tolerance, we found that 60° was a reasonable tolerance, balancing the accuracy of conformer determination while not being overly selective.
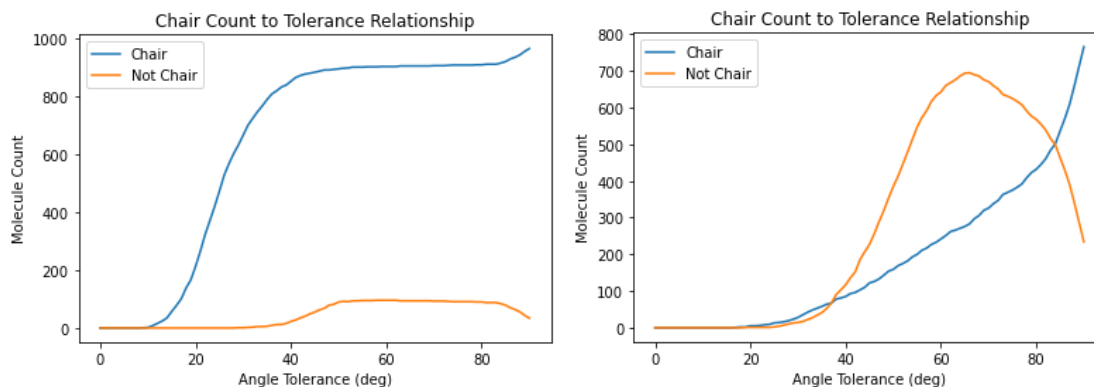

Figure 13. CY ring conformer count vs tolerance graphs for 298 K (left) and 1830 K (right) simulations.

There were some difficulties obtaining ring conformer ratios with our method. From Figure 13, we can see that it is possible to identify a distinct ring conformation ratio in CY at 298 K in the simulation but not necessarily at 1830 K because of the lack of a plateau. This is likely because the molecules uniformly take on well defined, low energy chair conformations at 298 K making them clearly sortable. Simulating at higher temperatures the molecules adopt a range of varied higher energy conformations that become less clearly sorted. Not only is there no clear plateau to draw an average ratio from, the ratio between two conformers should not surpass 1. We expand on this below.

Computational resources for multiple simulations primarily spent on MCY as opposed to CY. A temperature series analysis will thus focus on MCY (Figure 14).
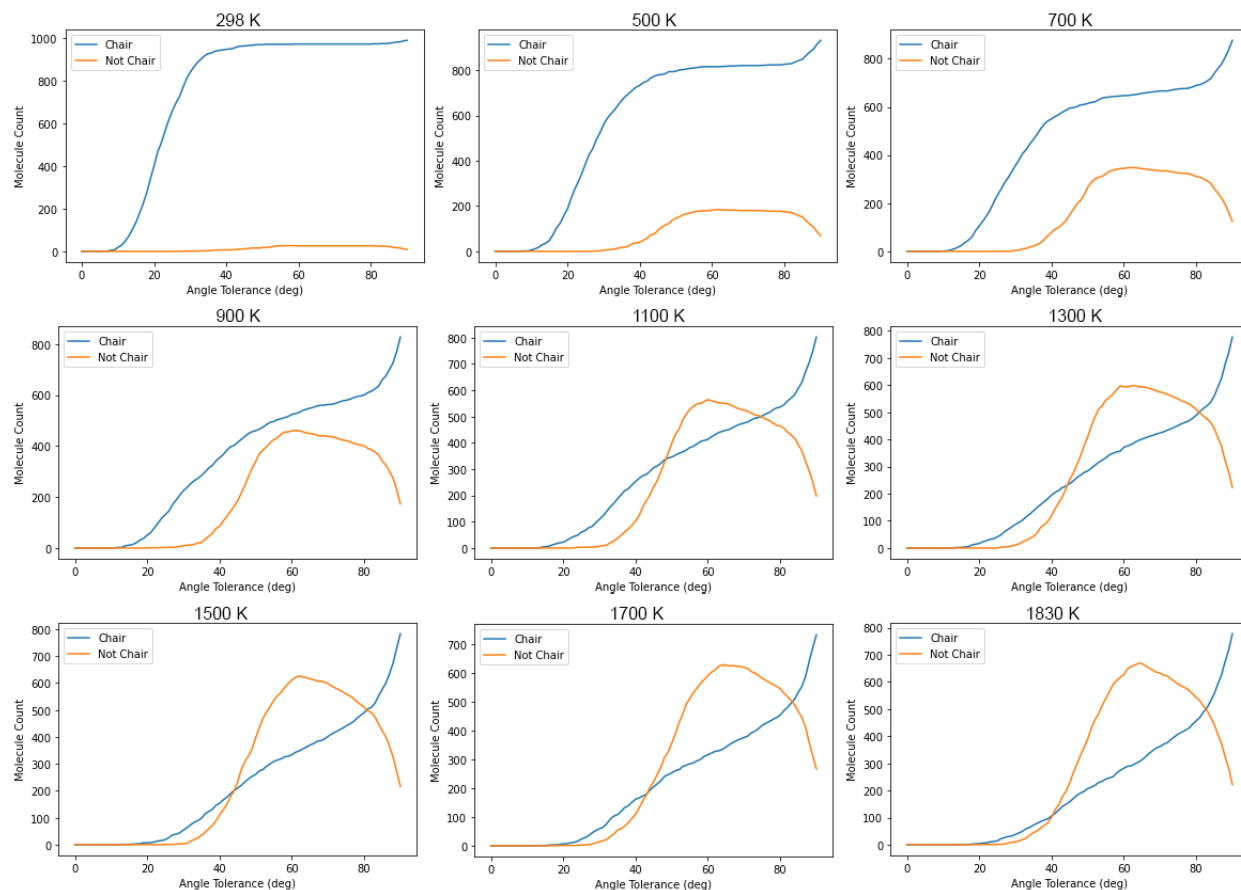
Figure 14. MCY ring conformer count vs tolerance graphs for temperatures 298 K to 1830 K.

In Figure 14, the initial plateau quickly disappears after 700 K. The non-chair categorization rises quickly with temperature, surpassing the chair categorization at 1100 K.

Again, if there are only two states present, this ratio should not surpass 1. Thus the non-chair categorization likely includes multiple other conformational states. We initially expected that all CY rings would virtually completely fall into the two main conformational minima: the chair and the twist-boat, but data shows unexpectedly large quantities of neither. The algorithm only looks to identify a structure as a chair or not a chair. It does not also additionally identify something as a twist-boat. Instead, by only having one criterion, it groups all non-chairs, including twist-boats together into one category, Thus, no direct comparison could be made between the two conformers. This could be corrected with an additional algorithm function that sorts twist-boats from non-chairs.

We were not able to evaluate the simulation via ring conformer ratios, but we found that it was still possible to evaluate the simulation via substituent conformer ratios. By using both algorithms in combination, the ring conformation algorithm, by its filtering effect, passes unambiguous chair conformers to the substituent conformation algorithm, which can then clearly sort the given input. This in turn produced meaningful data for MCY simulated at a range of temperatures from 298 K to 1830 K (Figure 15).
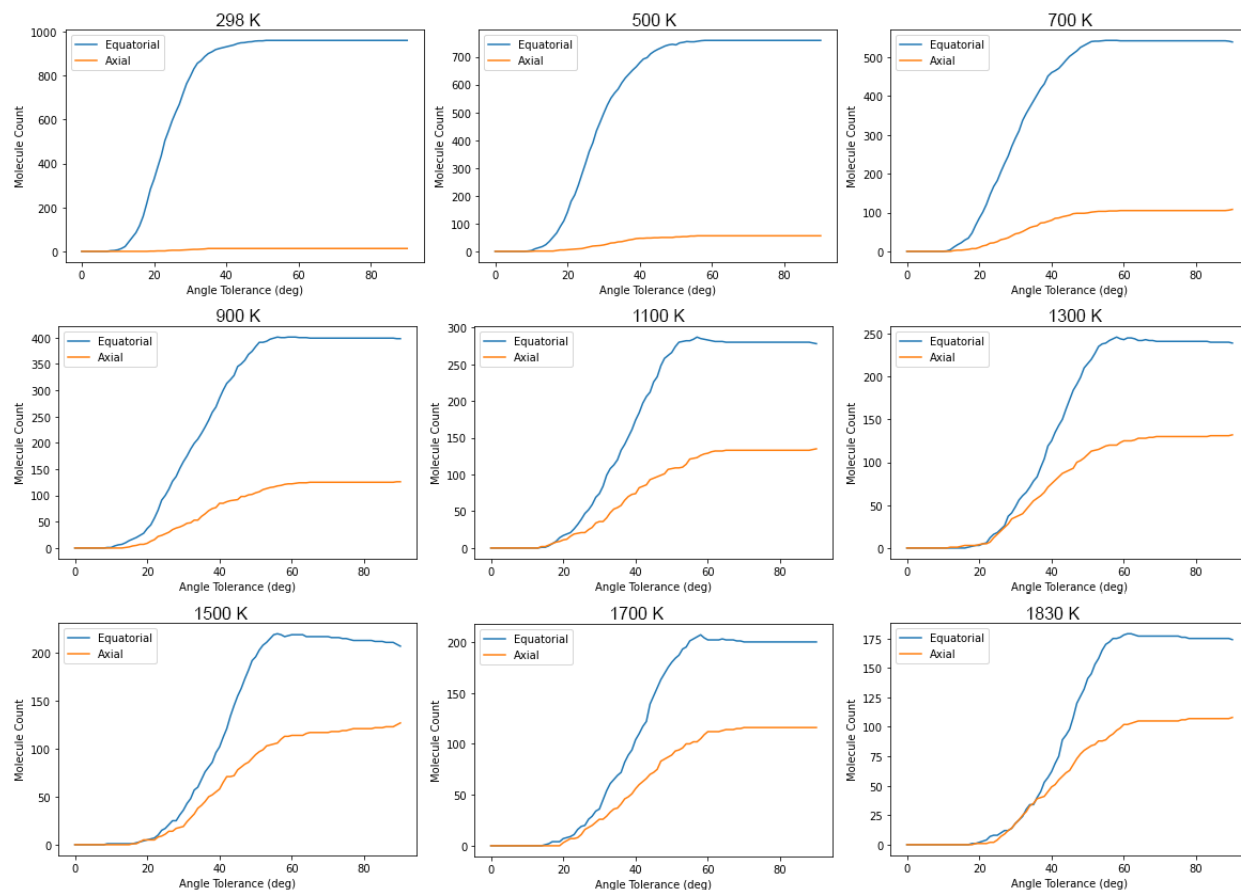
Figure 15. MCY substituent conformer count vs tolerance graphs for temperatures 298 K to 1830 K.

In Figure 15, there are distinct plateaus for the count of each category, showing that the ratio obtained past a certain point is independent of the tolerance used. The quantity of axial conformers rises with temperature but never surpasses the quantity of equatorial conformers. Additionally, we notice that the total number of evaluable molecules (ie. chair conformers given as input) decreases as temperature increases even while the trend in ratio between two conformers remains.

We consider this is consistent with the idea that molecules will begin occupying all other higher energy conformational states as they have more kinetic energy. Proportionally fewer of the total molecules will be in the lower energy states. On an imaginary conformational coordinate axis, these molecules are likely in states part way along one of the many transition state barrier curves, though not likely at the vertex. We remark that this quantity appears unexpectedly high at the higher temperature range.

Averaging the quantities of the plateaus, the simulated equilibrium constant - temperature dependence was compared to that of Equation (1) (Figure 16).
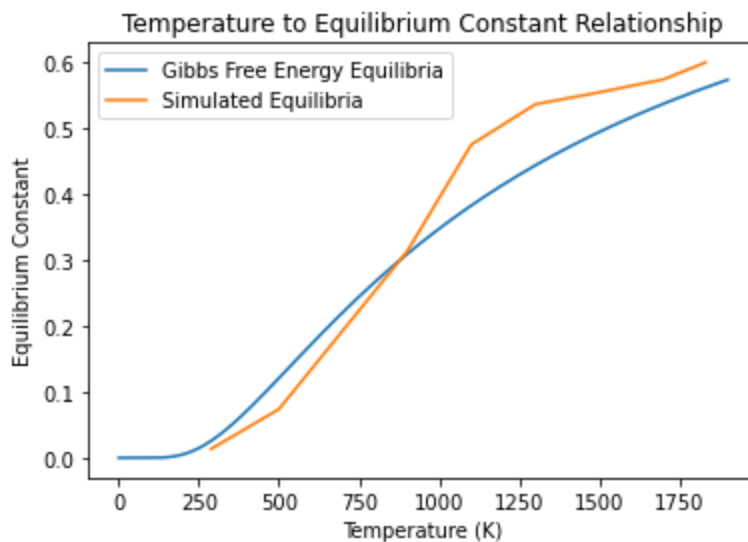
Figure 16. A comparison of equilibrium constants K, measured algorithmically from MD simulation, and calculated from Equation 1 with ΔG = 2.1 kcal/mol, R = 8.314.

Looking at Figure 16, the simulated equilibria increases with temperature, starting at close to 0 and ending approximately 0.6 between 298 K to 1830 K in gentle sigmoid fashion. The simulated equilibrium curve generally follows the shape of the curve of Equation 1 with some small deviations, mostly in the intensity of the sigmoid shape.

The MD simulation appears to somewhat accurately recreate the relationship between two conformations across a range of temperatures. It is difficult to attribute the largest source of error as flaws of MD simulation or the flaws in the geometry measurement method but it is more likely that improvements in the measurement algorithm are needed.

Conclusions

Using Gaussian, ES was able to recreate literature results for all conformers' structures and energies. Notably, it successfully determined the optimized structures for PMCY and DtCY. In addition, the calculated ratios of chair to twist-boat further supported Gaussian's output.

For MD, Gromacs minimization was able to produce correct conformer geometries for CY and MCY, but not PMCY or DtCY. The results were also not meaningful in terms of energy differences, thus the data from the two methods were not compared thoroughly. MD simulations have been shown to accurately simulate conformational ratios that reflect the relative energies of substituent conformational states in MCY through multiple temperatures. Directly through geometry, we observed that a surprising proportion of molecules occupy conformational states that are not local minima at high temperatures, emphasizing that conformational states of CY and its derivatives are not a dichotomy. We are looking to complete exploration of the MD simulation conformational ratio and temperature relationship in PMCY and DtCY. Currently, alterations to the algorithm will be required to add support for CY di-substitutions and substitution priority recognition.

To improve the MD results, future work would be to investigate the faults in Gromacs minimization. In addition, the post-simulation sorting algorithm could be upgraded to also identify other non-chair conformers such as twist-boats. This may be better achieved through other means such as machine learning. Alternatively, to achieve better chair to twist-boat ratios via simulated annealing. This protocol works by relaxing the structures that are energetically in between chairs and twist-boats by slowly decreasing temperatures. Overtime, the sample distribution would be sorted into only chairs or twist-boats while maintaining their original conformer ratios.

References

1. Wiberg, K. B.; Castejon, H.; Bailey, W. F.; Ochterski, J. Conformational Studies in the Cyclohexane Series. 2. Phenylcyclohexane and 1-Methyl-1-phenylcyclohexane. *J. Org. Chem.* **2000***, 65*, 1181-1187. https://pubs.acs.org/doi/10.1021/jo9917386 (accessed October 27, 2021)

2. Clayden, J.; Greeves, N.; Warren, S. G. *Organic chemistry;* Oxford University Press: New York; Oxford, 2012; .

3. Gaussian 16, Revision C.01: Gaussian, Inc.: Wallingford, 2016. https://gaussian.com/gaussian16 (accessed November 20, 2021)

4. *Optimization | Gaussian.com*. https://gaussian.com/opt/ (accessed November 20, 2021)

5. *IRC | Gaussian.com*. https://gaussian.com/irc/ (accessed November 20, 2021)

6. Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: A High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics*, **2013**, 29, 845–854. DOI: 10.1093/bioinformatics/btt055. https://academic.oup.com/bioinformatics/article/29/7/845/253065 (accessed 2021-12-19).

7. Dodda, L. S.; Cabeza de Vaca, I.; Tirado-Rives, J.; Jorgensen, W. L. LigParGen web server: an automatic OPLS-AA parameter generator for organic ligands. *Nucleic Acids Res.* **2017***, 45*, W331-W336. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5793816/?tool=pmcentrez&report=abstract (accessed November 22, 2021)

8. Squillacote, M.; Sheridan, R. S.; Chapman, O. L.; Anet, F. A. L. Spectroscopic Detection of the Twist-Boat Conformation of Cyclohexane. Direct Measurement of the Free Energy Difference between the Chair and the Twist-Boat. *J. Am. Chem. Soc.*, **1975**, 97, 3244–3246. DOI: 10.1021/ja00844a068. https://pubs.acs.org/doi/abs/10.1021/ja00844a068 (accessed 2021-12-19).

9. Herndon, L. R.; Reid, E. E. THE DECOMPOSITION OF ORGANIC COMPOUNDS AT HIGH TEMPERATURES AND PRESSURES. *J. Am. Chem. Soc.*, **1928**, 50, 3066–3073. DOI: 10.1021/ja01398a033. https://pubs.acs.org/doi/10.1021/ja01398a033 (accessed 2021-12-19).

10. Van Rossum, G. *Python tutorial, Technical Report*; CS-R9526; Centrum voor Wiskunde en Informatica (CWI): Amsterdam, 1995. https://ir.cwi.nl/pub/5007 (accessed 2021-12-19).