

Supplementary Information

Machine Learning Classification of Disrotatory IRC and Conrotatory Non-IRC Trajectory Motion for Cyclopropyl Radical Ring Opening

Steven M. Maley, Jesse Melville, Spencer Yu, Matthew S. Teynor, Ryan Carlsen, Cal Hargis, R. Spencer Hamilton, Benjamin O. Grant, and Daniel H. Ess *

Department of Chemistry and Biochemistry, Brigham Young University, Provo, Utah, 84602, USA.

Table of Contents

<i>Thermodynamic Data</i>	S2
<i>Jupyter Notebook Information</i>	S2
<i>Additional Data</i>	S3
<i>Xyz Coordinates</i>	S6

Thermodynamic Data

Table S1. Electronic energy (E), enthalpy (H), Gibbs free energy (G), and zero-point correction (ZPE) for the reactant, transition state (**TS1**), and product. All values in hartree.

	E	ZPE <i>correction</i>	H	G
Reactant	-117.1559171	0.0679620	-117.0836150	-117.1127580
TS1	-117.1062407	0.0646980	-117.0332110	-117.0820590
Product	-117.1915897	0.0665630	-117.1202810	-117.1501560

Jupyter Notebooks

The Jupyter notebook “Example_Machine_Learning_Code.ipynb” contains an example Python code for the initial training of machine learning models, hyperparameter optimization, application of optimized random forest classifier to data set, and feature importance. A second notebook “Individual_Trajectory_Accuracy.ipynb” contains an example Python code to determine the accuracy of individual trajectories.

Additional Data

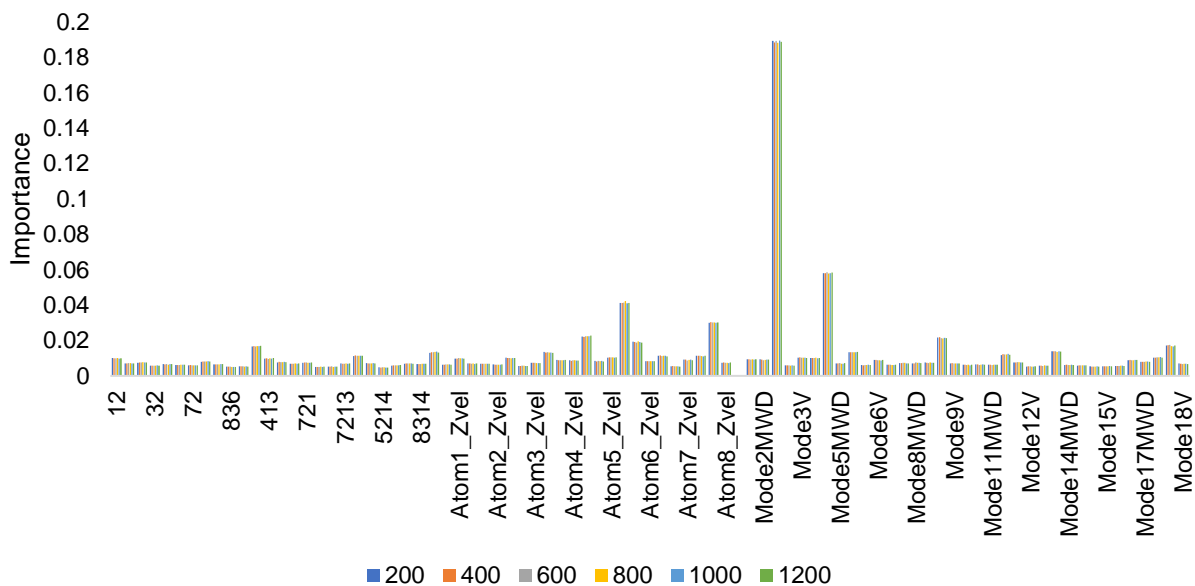


Figure S1. Feature importance determined using the optimized random forest model for dataset samples of 200, 400, 600, 800, 1000, and 1200 trajectories. The ratio of conrotatory:disrotatory trajectories is 1:1 to maintain a baseline accuracy of 50%.

Table S2. Mean accuracy and confidence intervals of optimized random forest model for dataset samples of 200, 400, 600, 800, 1000, and 1200 trajectories. The ratio of conrotatory:disrotatory trajectories is 1:1 to maintain a baseline accuracy of 50%.

# Trajectories	Accuracy	Upper_CI	Lower_CI
200	0.76	0.82	0.70
400	0.78	0.82	0.74
600	0.79	0.82	0.76
800	0.81	0.85	0.78
1000	0.83	0.85	0.80
1200	0.82	0.84	0.80

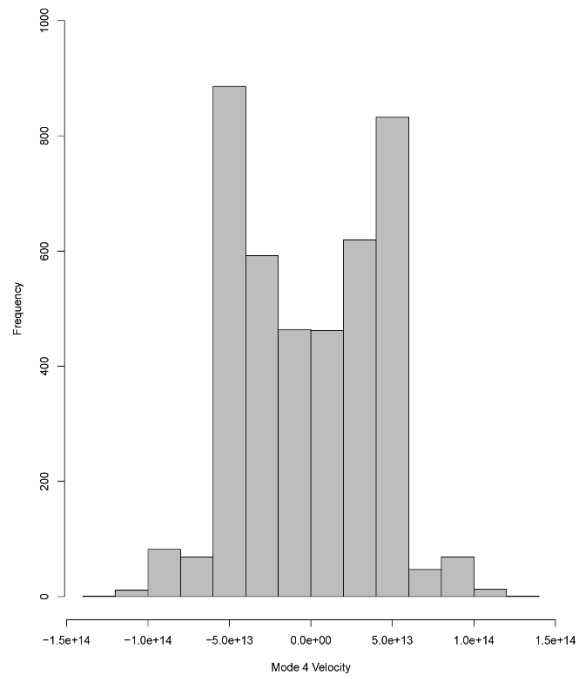
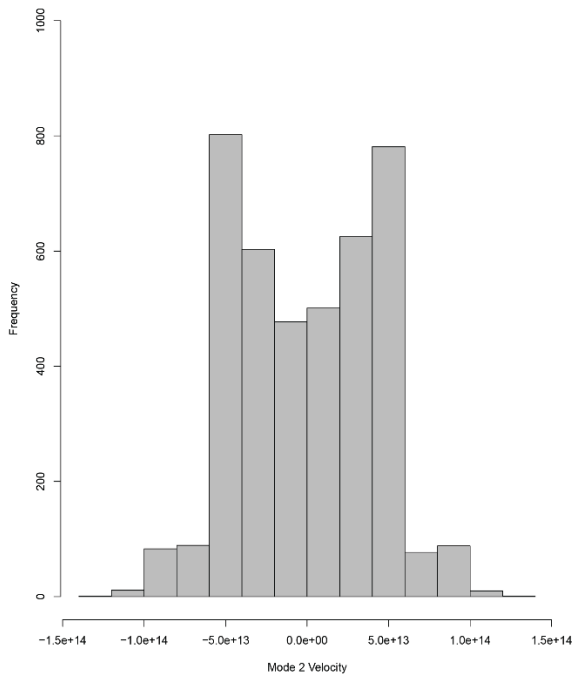
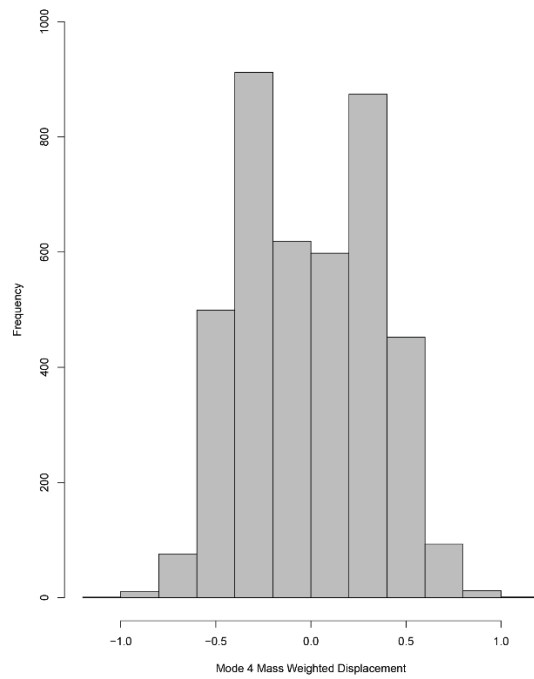
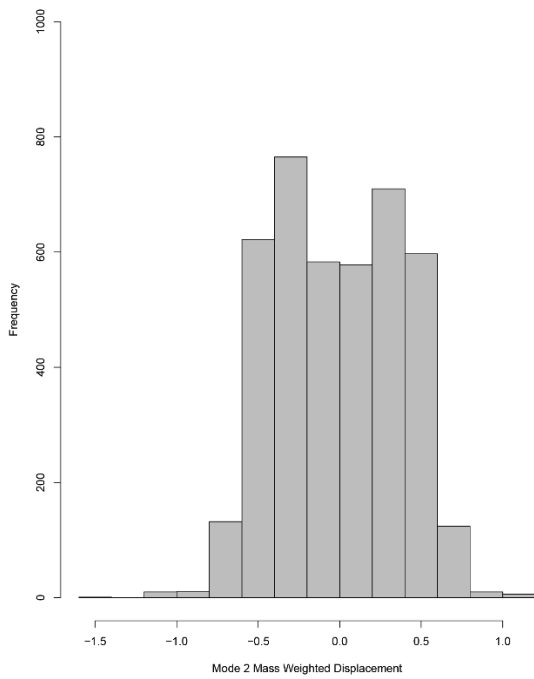


Figure S2. Histograms of mode 2 and 4 mass weight displacement and velocity.

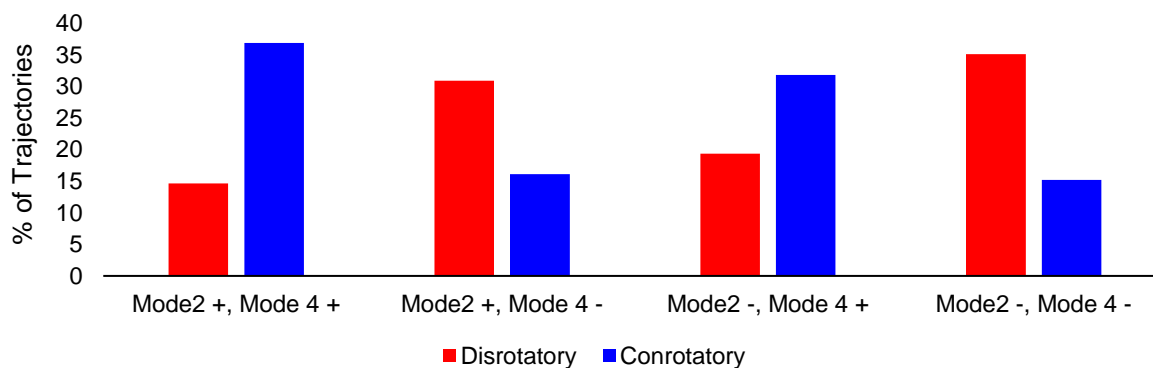


Figure S3. Percentages of disrotatory and conrotatory trajectories with directional combination of vibrational mode 2 and 4 velocities. The percentages are relative to the total number of overall disrotatory or conrotatory trajectories. Mode2 +, Mode 4 + refers to vibrational modes 2 and 4 both having positive directional velocity. Mode2 +, Mode 4 - refers to vibrational mode 2 with positive directional velocity and mode 4 with negative vibrational mode velocity. Mode2 -, Mode 4 + refers to vibrational mode 2 with positive directional velocity and mode 4 with negative vibrational mode velocity. Mode2 -, Mode 4 - refers to vibrational modes 2 and 4 both having negative directional velocity. Importantly, the definition of positive and negative velocity is consistent for all trajectories.

Note on trajectories that are difficult to classify by random forest:

As discussed in the manuscript, ~250 trajectories could never be classified correctly. Because the trajectories were initialized with both geometric and velocity components, we considered the possibility that the random forest algorithm might be unable to distinguish between velocities that are close to reversing their direction versus velocities that are far from reversing their direction. Therefore, we analyzed the RMS geometrical displacement in modes 2 and 4. For mode 2, the average RMS displacement is 0.319 Å for trajectories that can never be correctly classified (0% accuracy) and 0.310 for trajectories that are always classified correctly (100% accuracy). For mode 4, the RMS averages are 0.318 and 0.292 for 0% and 100% accuracy. For both modes 2 and 4 these values are not statistically different.

Note on random forest model using DynSuite trajectories:

It is also important to note that while the 979 DynSuite trajectories gave a nearly identical ratio of disrotatory to conrotatory outcomes, a random forest model built using these trajectories has a diminished prediction accuracy. This decrease in accuracy is potentially the result of less overall model information because vibrational mode energy is initiated only as atomic velocities rather than a mixture of geometry displacement (potential energy) and atomic velocities (kinetic energy).

XYZ Coordinates

Reactant

6	0.00004	0.87306	-0.17449
6	0.76209	-0.36347	0.03074
6	-0.76214	-0.36338	0.03075
1	0.00016	1.79809	0.38340
1	1.25085	-0.82362	-0.82451
1	-1.25083	-0.82344	-0.82459
1	1.28523	-0.51416	0.97186
1	-1.28535	-0.51416	0.97181

TS1

6	0.08486	0.69599	-0.22116
6	0.95590	-0.33687	0.02968
6	-1.03842	-0.19732	0.06444
1	0.20194	1.69837	0.17110
1	0.89089	-1.27334	-0.51900
1	-1.45645	-0.85782	-0.68852
1	1.79390	-0.27381	0.73085
1	-1.44427	-0.26421	1.06779

Product

6	0.00043	0.87295	-0.17386
6	0.76184	-0.36401	0.03072
6	-0.76241	-0.36313	0.03080
1	0.00122	1.79928	0.38173
1	1.25035	-0.82316	-0.82515
1	-1.25097	-0.82193	-0.82529
1	1.28608	-0.51471	0.97117
1	-1.28584	-0.51432	0.97157