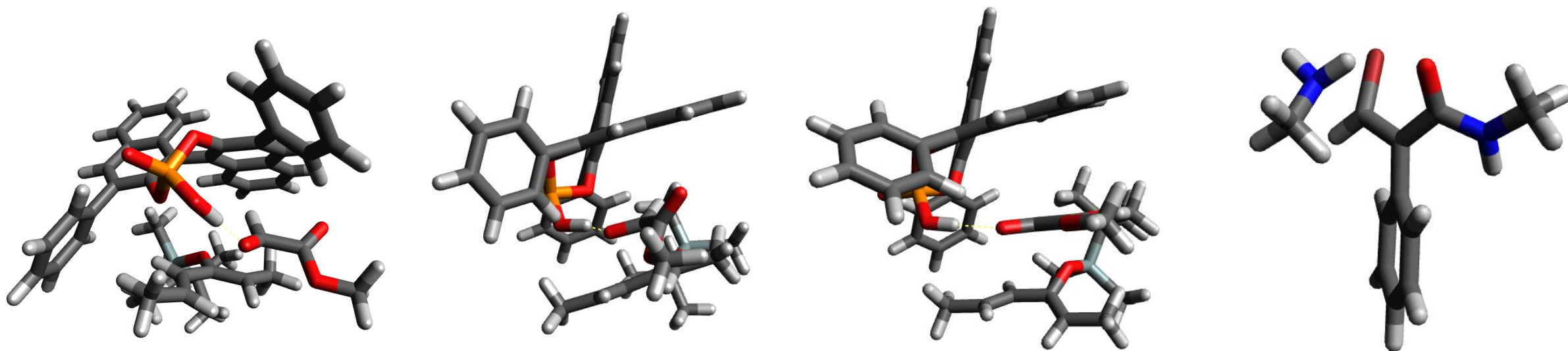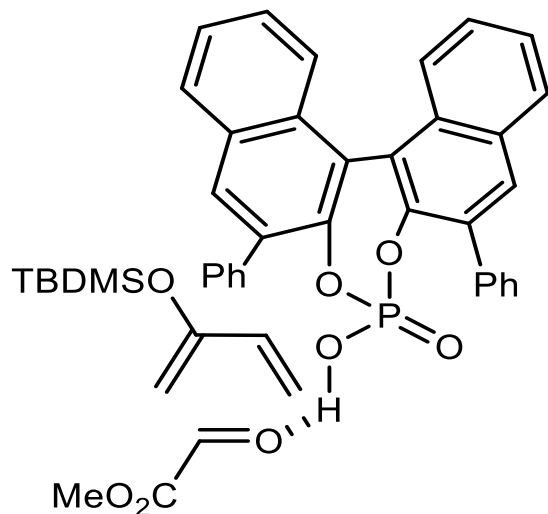# TS searching with mixed method Hessians

January 2024

Pre-optimising at a lower level of theory

BPA_TS_1
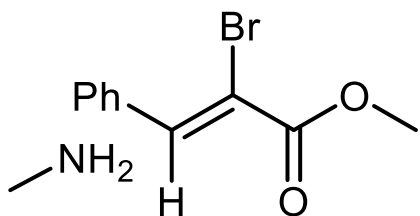(catalysed hetero Diels-Alder)

SQM (PM6 PCM(toluene)):
4x optimisation cycles, 153 steps, 11h-17m-47.3s
HF (HF 6-31G* PCM(toluene))
8x optimisation cycles, 374 steps, 15d-14h-22m-54.5s
LDA (SVWN def2-SVP PCM(toluene))
10+ optimisation cycles, 500+ steps, 16d+

TS_68
(uncatalysed aza-Michael)

SQM (PM6 PCM(toluene)):
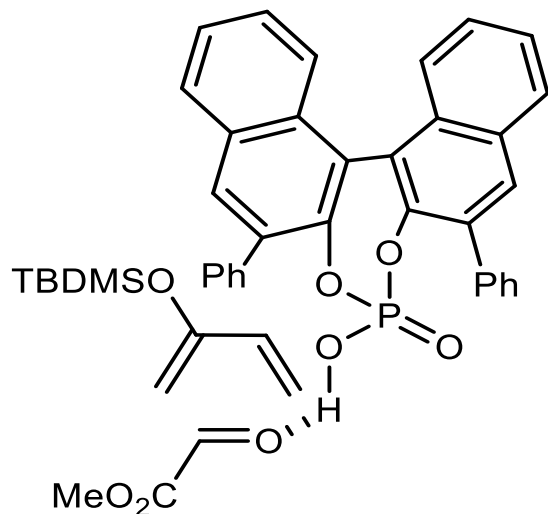1x optimisation cycle, 36 steps, 19m-55.3s
HF (HF 6-31G* PCM(toluene))
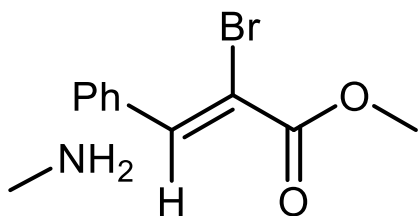1x optimisation cycle, 23 steps, 1h-15m-11.6s
LDA (SVWN def2-SVP PCM(toluene))
3x optimisation cycles, 131 steps, 9h-38m-31.7s

Re-optimising lower-level structures to DFT (wB97XD def2-SVP PCM(toluene))

BPA_TS_1
(catalysed hetero Diels-Alder)

SQM → DFT :
?x optimisation cycle, ? steps, ?
HF → DFT :
?x optimisation cycle, ? steps, ?
LDA → DFT :
?x optimisation cycles, ? steps, ?
DFT:
6x optimisation cycles, 73 steps, 17d-8h-17m-43s

TS_68
(uncatalysed aza-Michael)

SQM → DFT :
4x optimisation cycles, 161 steps, 22h-43m-54.8s
HF → DFT :
1x optimisation cycle, 16 steps, 1h-59m-55.9s
LDA → DFT :
4x optimisation cycles, 181 steps, 21h-59m-30.1s
DFT:
1x optimisation cycle, 27 steps, 4h-11m-38.1s
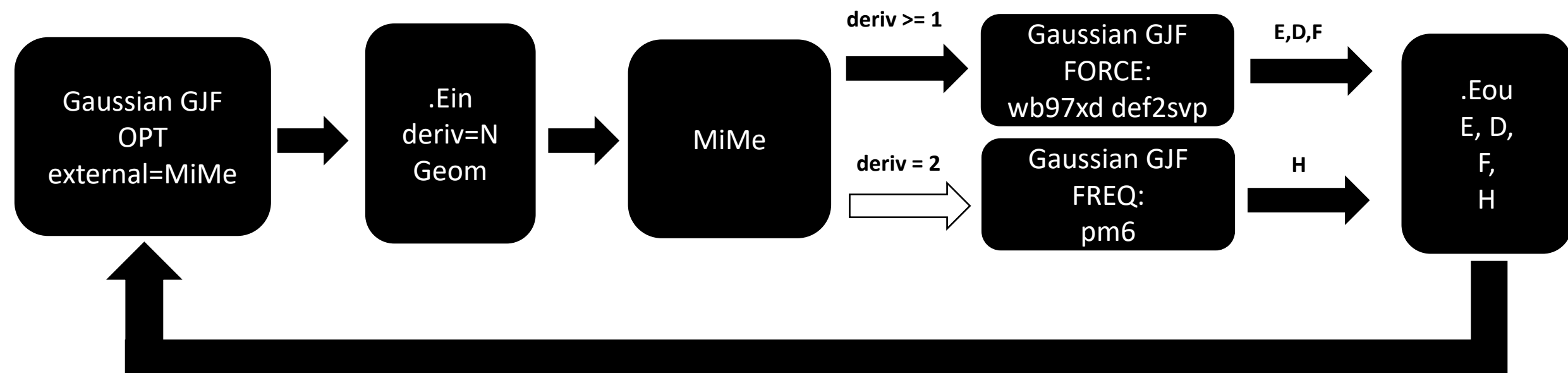
External keyword for Gaussian interface

Previously:

1. Run separate Freq=noraman calculation at lower level of theory (e.g. PM6, HF 6-31G*, SVWN def2-SVP) with `%chk=freq.chk` keyword

2. Move checkpoint file from frequency calculation to folder with optimisation Gaussian input file

3. Use `%chk=freq.chk` with `opt=(readfc,ts,noeigen,maxcyles=N,…)` keyword arguments for optimisation

4. This is designed to load in the low-level hessian into a higher-level TS optimisation, with new hessians recalculated every N steps

External keyword:

1. Run TS optimisation all in one Gaussian input file using:
`external=`"`~/scripts/MiMe -F 'pm6' -EF 'wb97xd def2svp' -c 12 -m 48GB`" `opt=(recalcfc=N,ts,noeigen,…,nomicro)`

2. No need for additional frequency calculations, but does require a working script to interface with G16 via `External`

3. G16 sends the python script (MiMe) a `Gau-12345.EIn` "input" file containing [a] deriv = d where d is the derivatives of Energy, and [b] the molecule specification (charge, spin and geometry in standard orientation in Bohrs)

4. Python script reads in `Gau-12345.EIn` and sends off a new G16 calculation of Energy, Force, or Freq (if d = 0,1,2) and returns those values to the main G16 task via the `Gau-12345.EOu` "output" file

External keyword for Gaussian interface

External keyword for Gaussian interface

1. Currently MiMe does not replicate the results obtained by using an external frequency calculation, reading in the Hessians via the `opt=readfc` keyword argument. The calculations begin very similarly, with the **identical** Forces and **similar** Eigenvalues but these numbers **diverge** the longer the calculation goes on for – leading to slightly more steps being taken with External and therefore taking longer:

2. Currently testing on DA-1_2 (acrolein Diels-Alder dimerization) at **23 steps vs 21 steps** and on TS_68 (uncatalysed aza-Michael reaction) at **42 steps vs 42 steps** for `calcfc`

3. One possibility is that each Freq calculation involves a putative Berny optimisation step, and the Hessian after optimisation is what is used by `readfc` whereas the Hessian before optimisation is used by `External` (or vice versa). i.e. the Hessians are incorrect from the start (or from the 2nd step onwards)

4. Another explanation is that the Hessians begin the same but are **updated** in different ways by the Berny / LBFGS update scheme over the course of optimisation.

5. It would be worth getting the `External` keyword to work, since this would allow us to intercept the low-level Hessian and improve it with ML before passing it back to G16, and because the process is more self-contained (no explicit Freq calculations).

6. The problem is likely to involve the use of the Hessian and Forces in converting between Cartesian coordinates and Internal Coordinates – when mixed Hessians and Forces are used the change of coordinates introduces an error

External keyword for Gaussian interface

Once the `external` keyword is able to replicate the optimisation results for `readfc`:

1. Investigate different frequencies of Hessian calculation using `opt=(calcfc,recalcfc=N,…)` for N = {1=`calcall`, 2, 5, 10, 20, 50=`calcfc`}

2. Create a script to intercept the Geometry and Hessians from MiMe and use delta-ML to improve the low-level Hessian towards the higher-level (DFT) Hessian

3. Test ML Hessians with different combinations of low and high level of theory (or select one high level of theory and try different low levels e.g. SMQ-wB97xd, HF-wB97xd, LDA-wB97xd

TS structures: OPLS3e initial geometries (reacting bonds constrained)

Phosphoric-acid catalysed Hetero Diels-Alder (two different conformers)
BPA_1, BPA_2
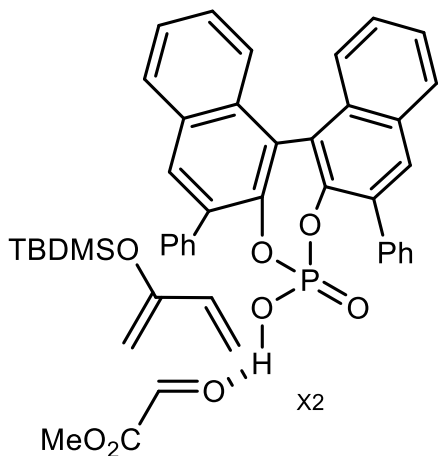
Aza-Michael reactions
AM_249, AM_898,
AM_325, AM_68
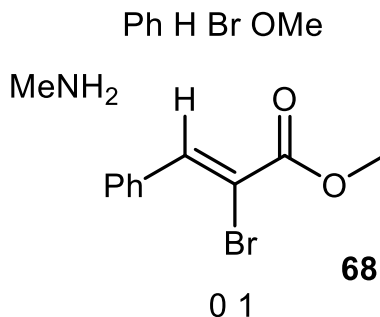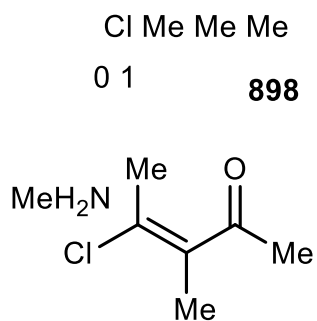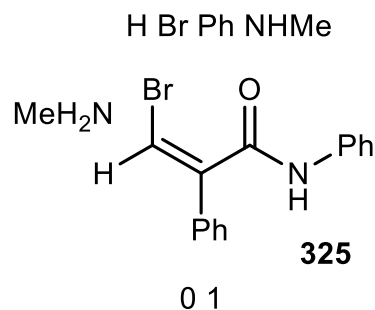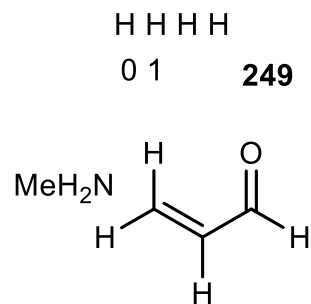
Diels-Alder reactions
IMDA_1, IMDA_2,
DA_1, DA_2
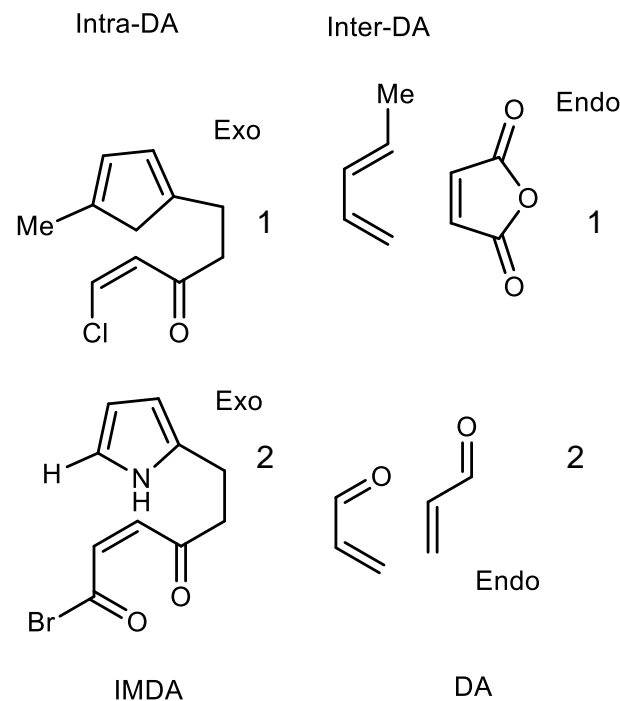
1,3-Dipolar cycloadditions
allylic, cyclic
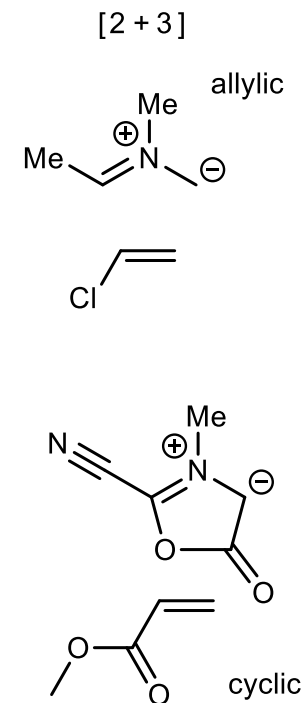
Hessian levels of theory: SQM, HF, DFT

	DFT (wB97xd/def2-SVP)	HF (HF/6-3G(d))	SQM (PM6)

ReadFC	`%chk=BPA_TS_1_opt_dft.chk`
	`%nprocshared=12 %mem=48GB`
	`# wb97xd def2svp geom=allchk`
	`scrf=(iefpcm,solvent=toluene)`
	`opt=(readfc,ts,noeigen,maxcycles=50,maxstep=10)`
		`%chk=BPA_TS_1_opt_hf.chk`
		`%nprocshared=12 %mem=48GB`
		`# wb97xd def2svp geom=allchk`
		`scrf=(iefpcm,solvent=toluene)`
		`opt=(readfc,ts,noeigen,maxcycles=50,maxstep=10)`
		
		`(BPA_TS_1_hf.chk obtained from previous opt)`
			`%chk=BPA_TS_1_opt_sqm.chk`
			`%nprocshared=12 %mem=48GB`
			`# wb97xd def2svp geom=allchk`
			`scrf=(iefpcm,solvent=toluene)`
			`opt=(readfc,ts,noeigen,maxcycles=50,maxstep=10)`
			
			`(BPA_TS_1_opt_sqm.chk obtained from previous opt)`

CalcFC	`%chk=BPA_TS_1_opt_dft_re.chk`
	`%nprocshared=12 %mem=48GB`
	`# wb97xd def2svp geom=allchk`
	`scrf=(iefpcm,solvent=toluene)`
	`opt=(calcfc,ts,noeigen,maxcycles=50,maxstep=10)`
		`%chk=BPA_TS_1_opt_hf.chk`
		`%nprocshared=12 %mem=48GB`
		`# wb97xd def2svp geom=allchk`
		`scrf=(iefpcm,solvent=toluene)`
		`opt=(readfc,ts,noeigen,maxcycles=50,maxstep=10)`
		
		`(BPA_TS_1_hf.chk obtained from`
		`a new hf 6-31g(d) hessian calculation)`
			`%chk=BPA_TS_1_opt_sqm.chk`
			`%nprocshared=12 %mem=48GB`
			`# wb97xd def2svp geom=allchk`
			`scrf=(iefpcm,solvent=toluene)`
			`opt=(readfc,ts,noeigen,maxcycles=50,maxstep=10)`
			
			`(BPA_TS_1_opt_sqm.chk obtained from`
			`a new pm6 hessian calculation)`

BPA_1
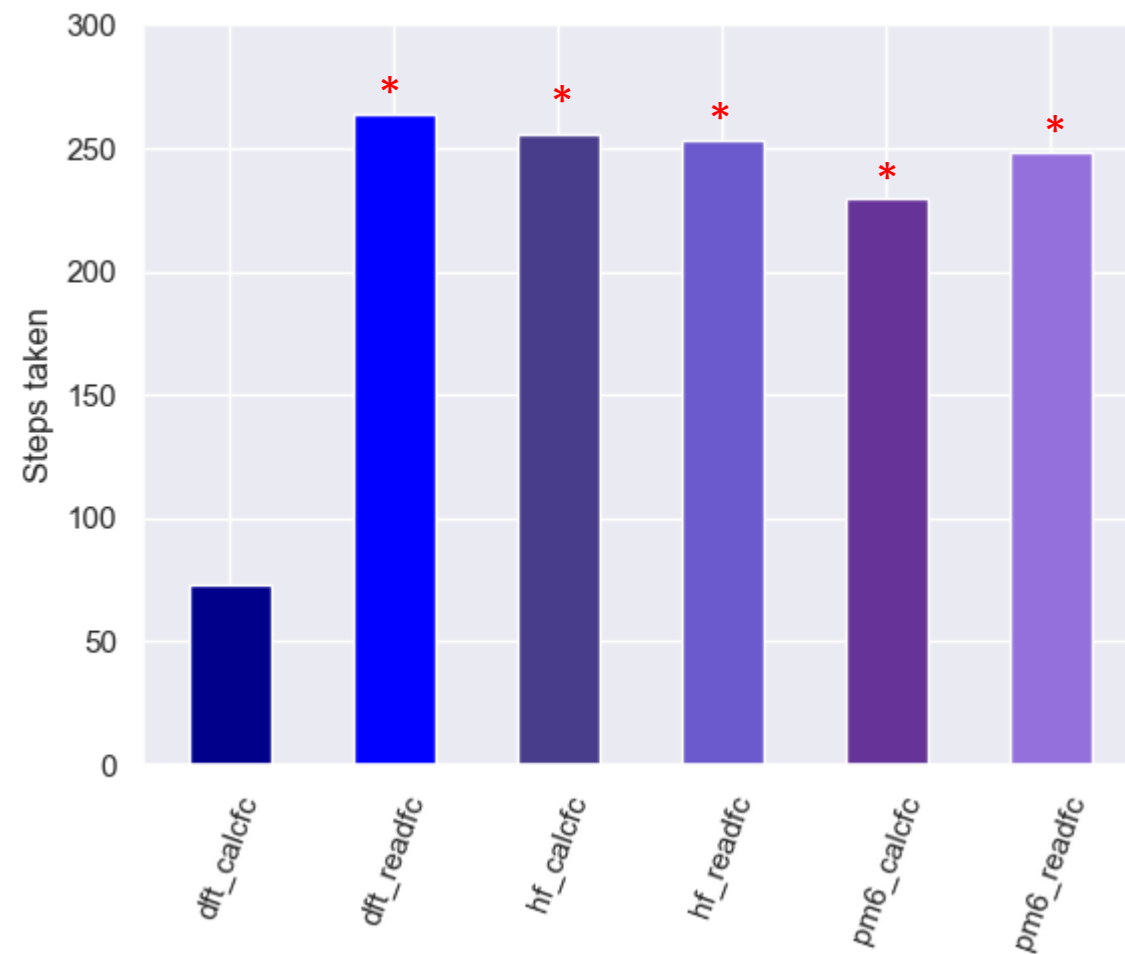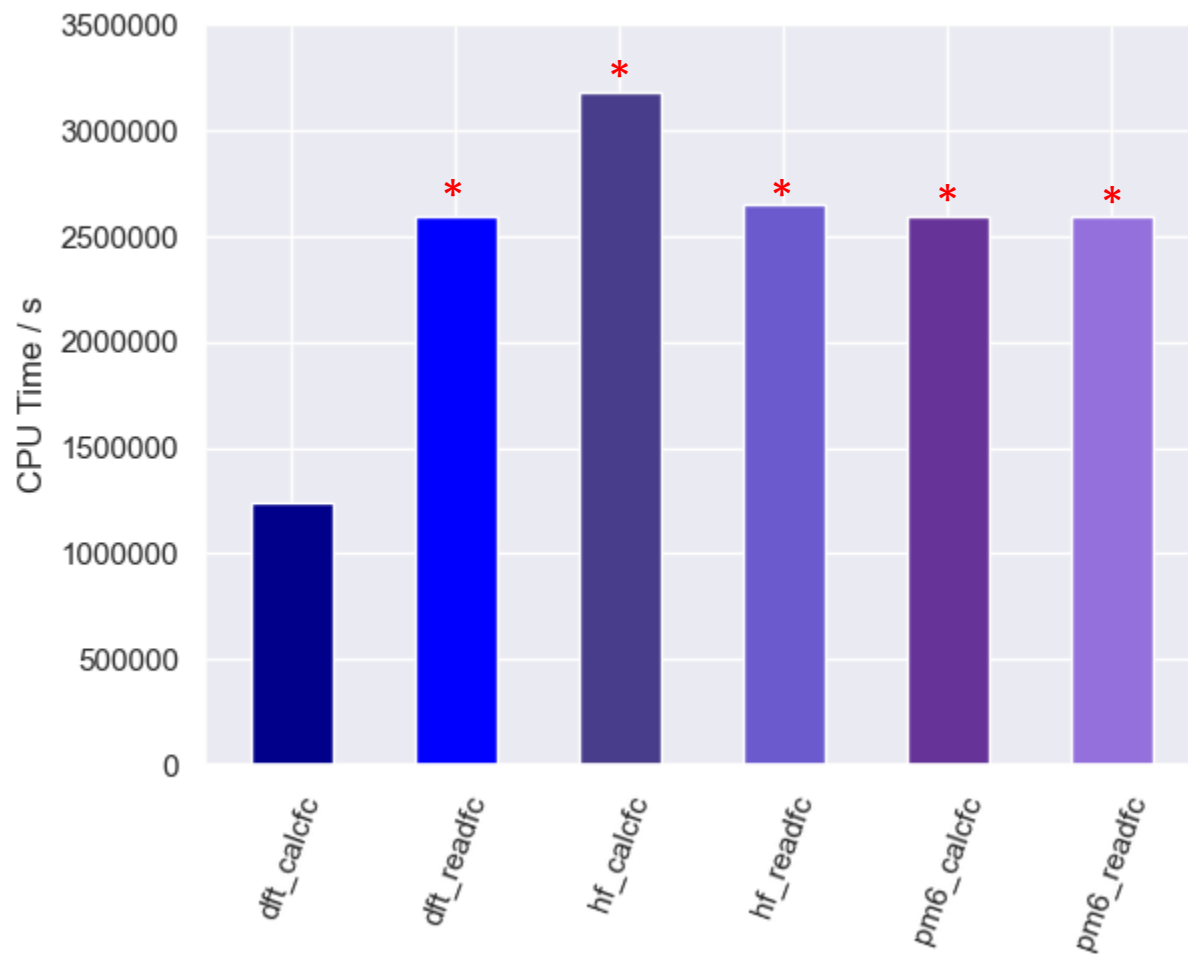
* TS structure did not optimise after 10 job cycles

BPA_1: conformation
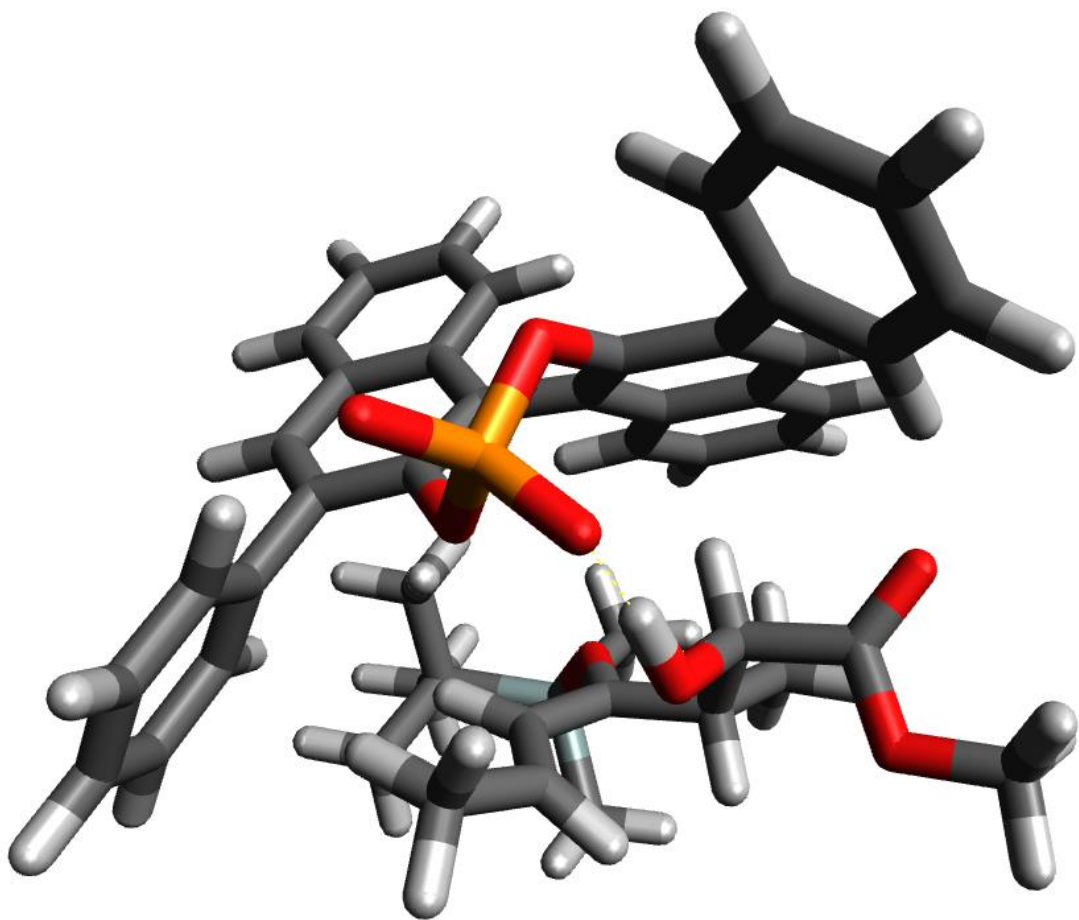
DFT calcfc
(no other jobs converged)
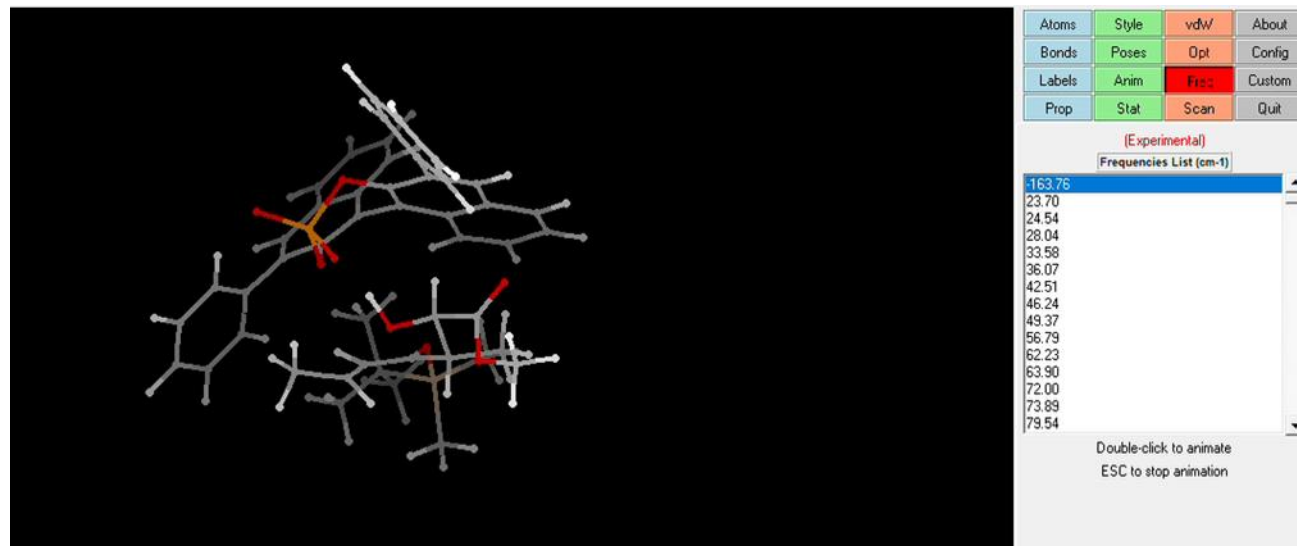
BPA_1
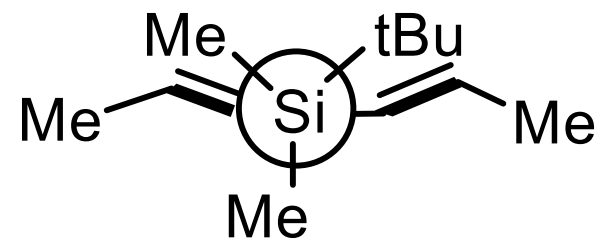
BPA_2
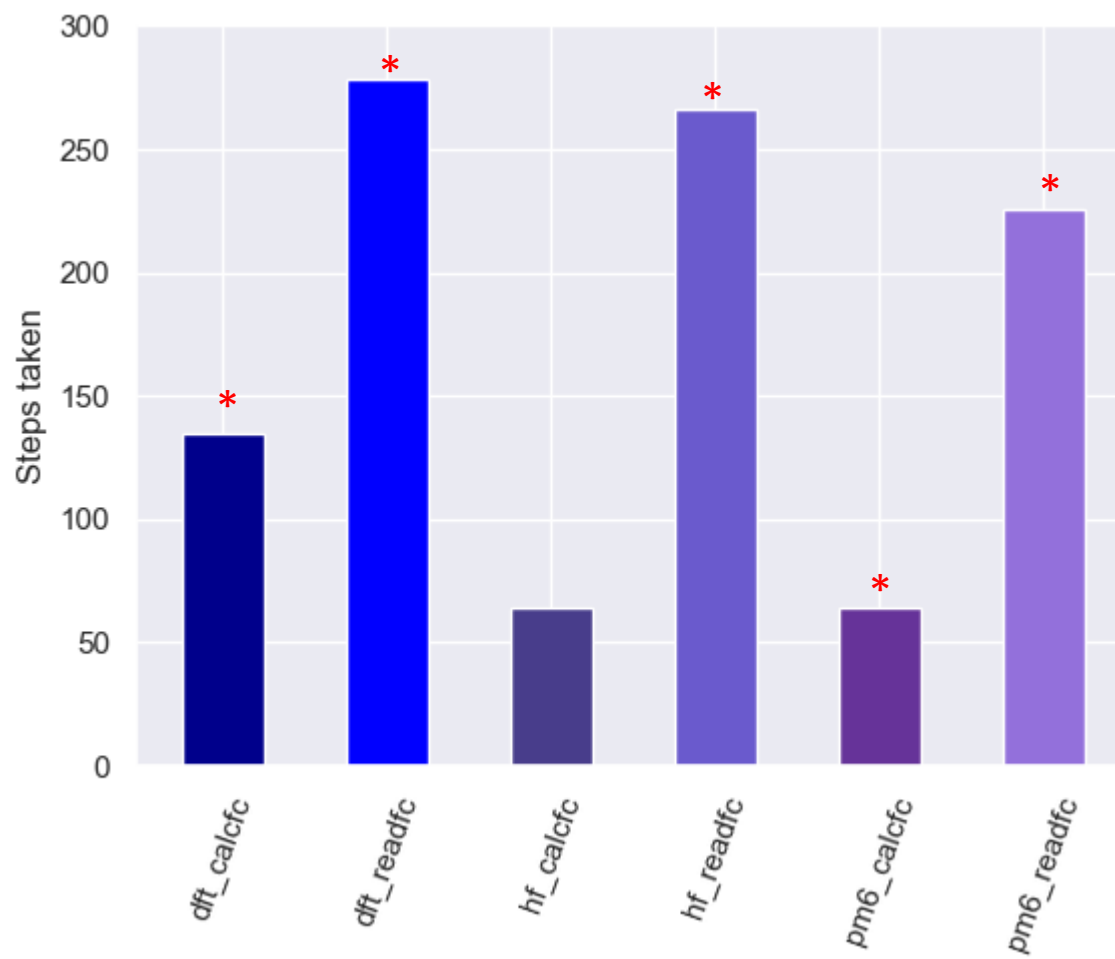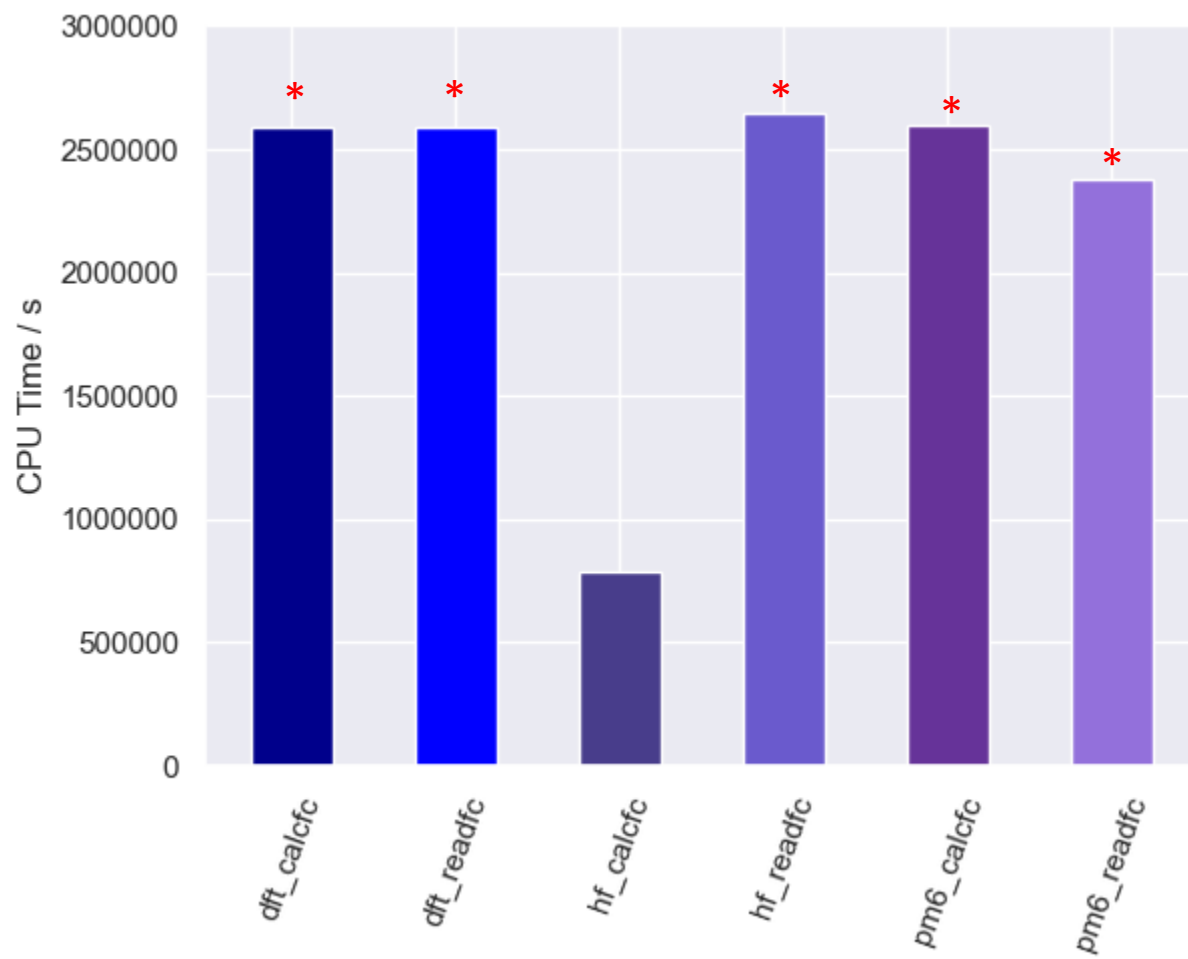
BPA_2: conformation
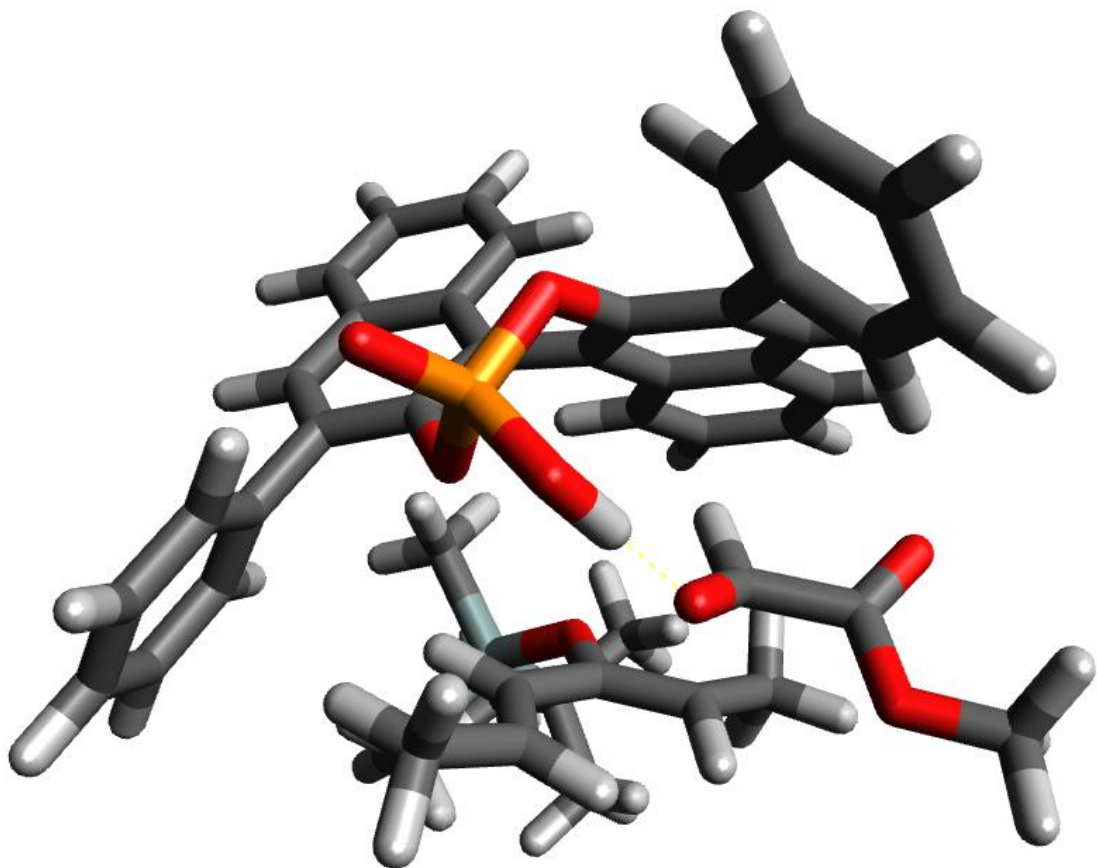
HF calcfc
(no other jobs converged)
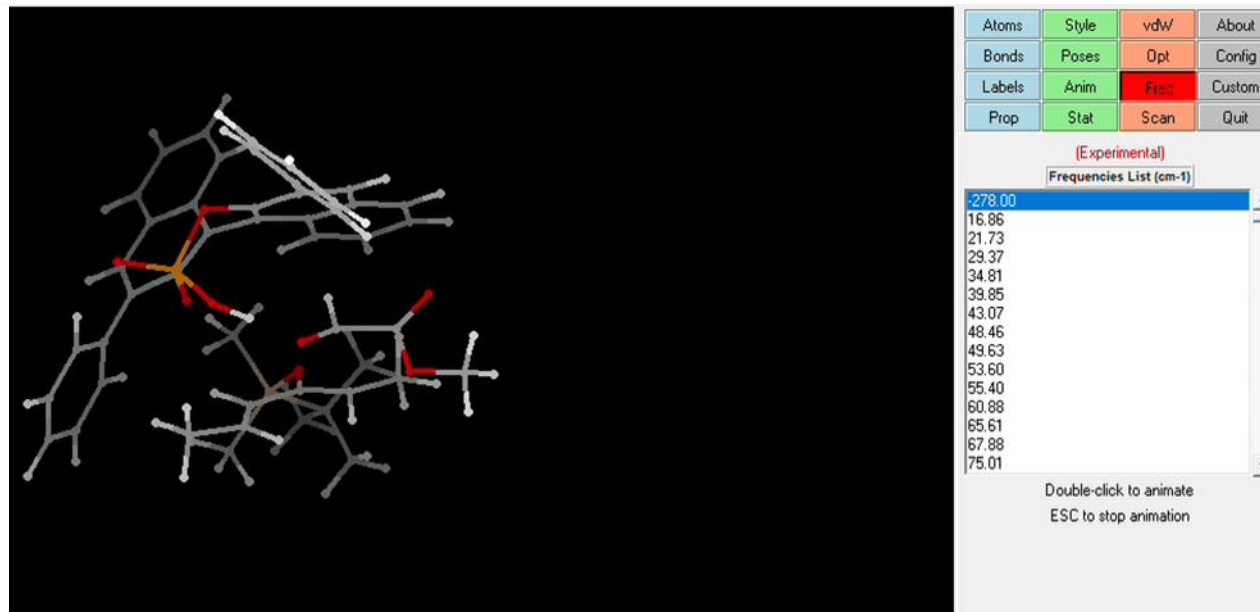
BPA_2
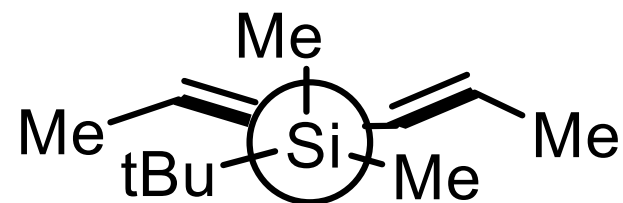
AM_249
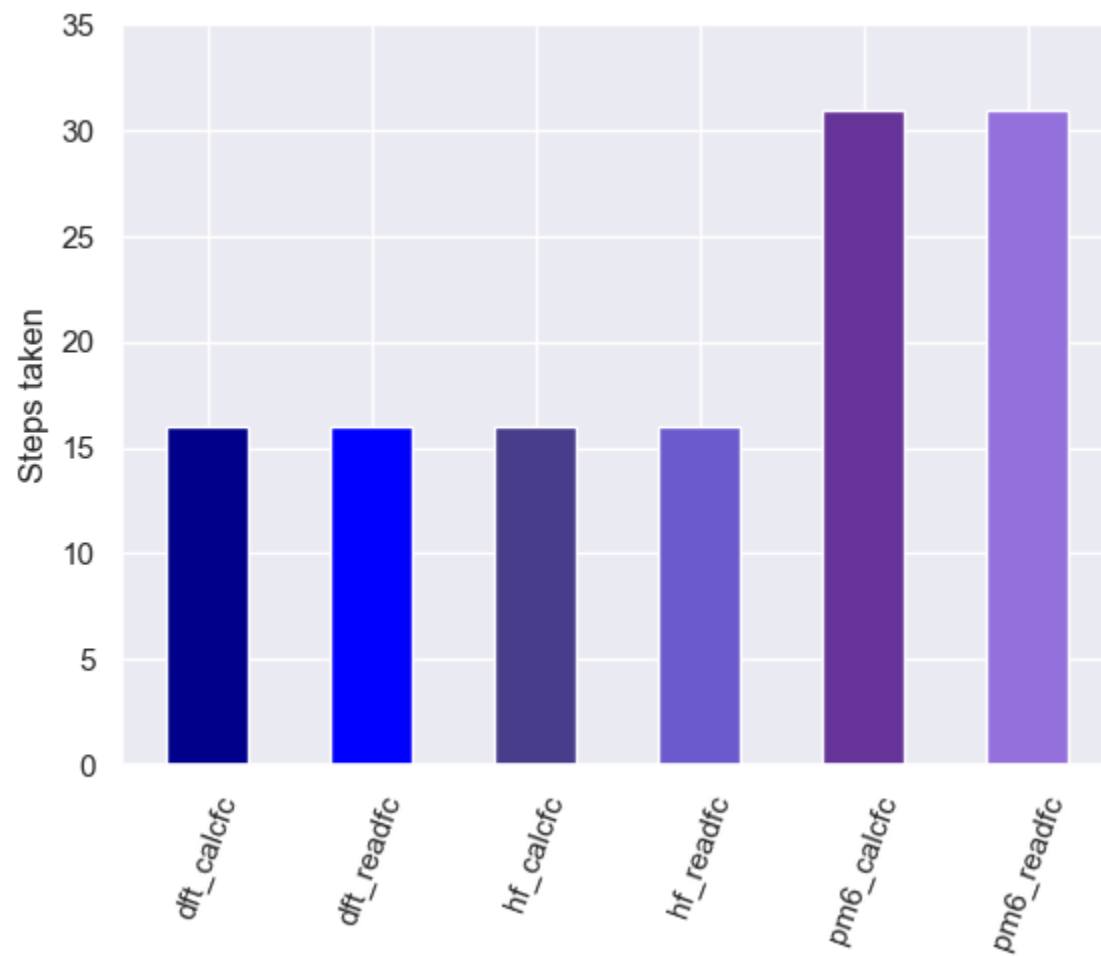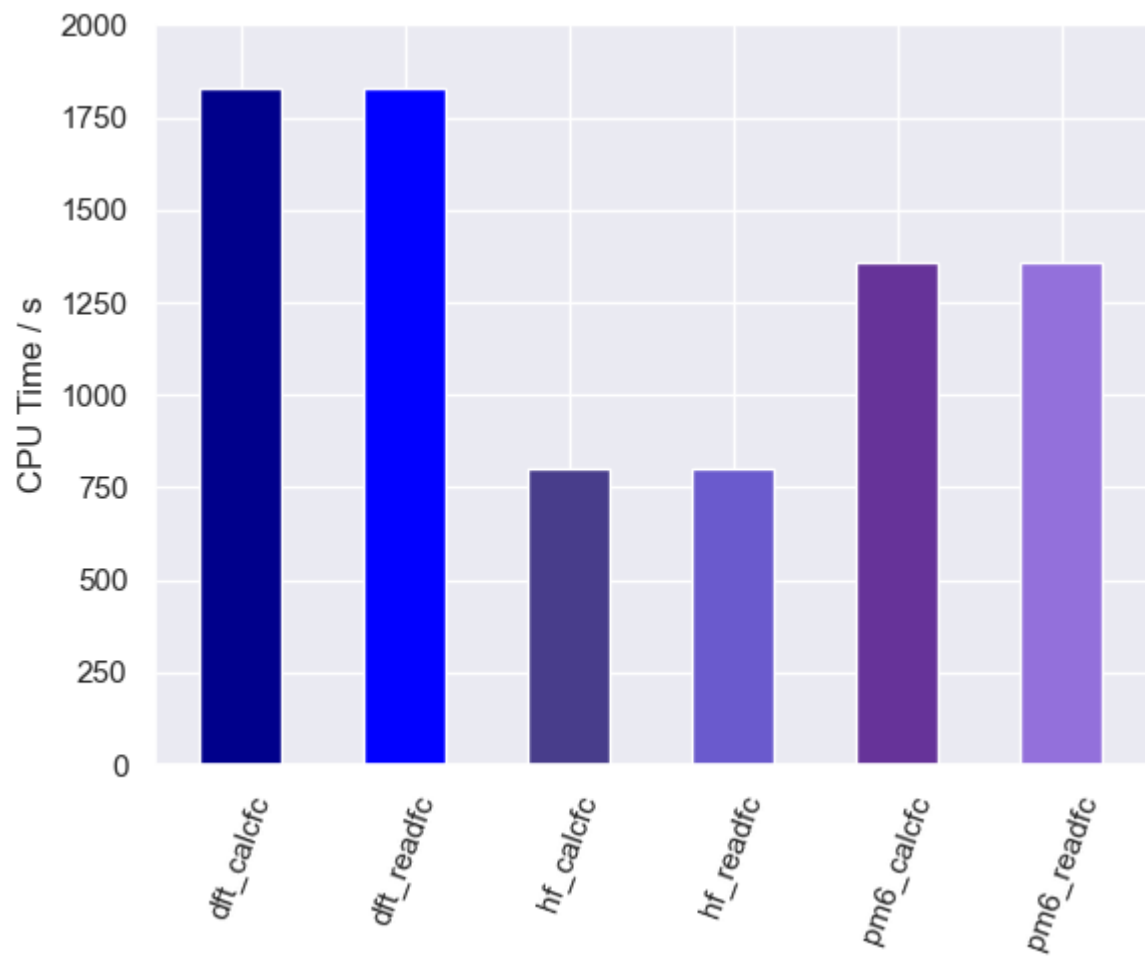
AM_249: conformations
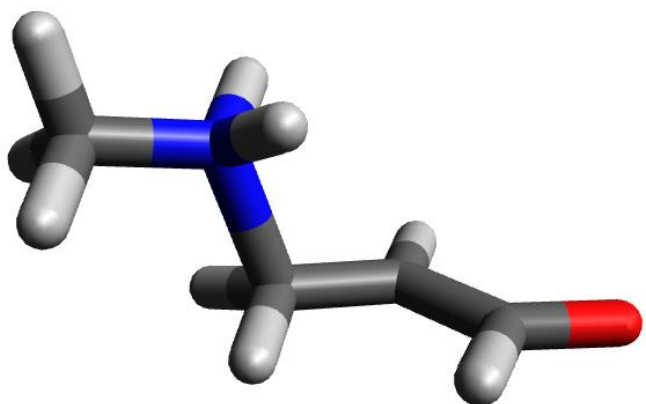
DFT Hessian
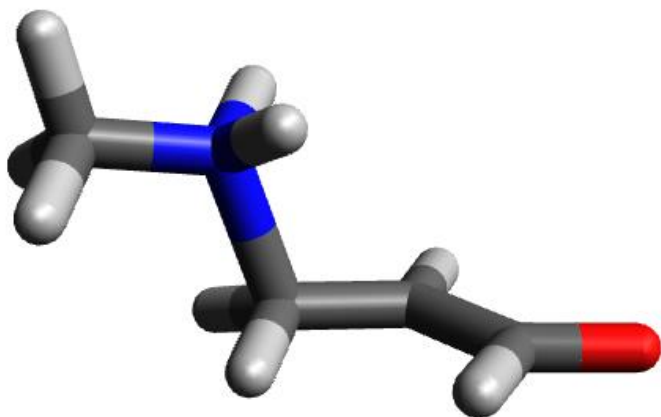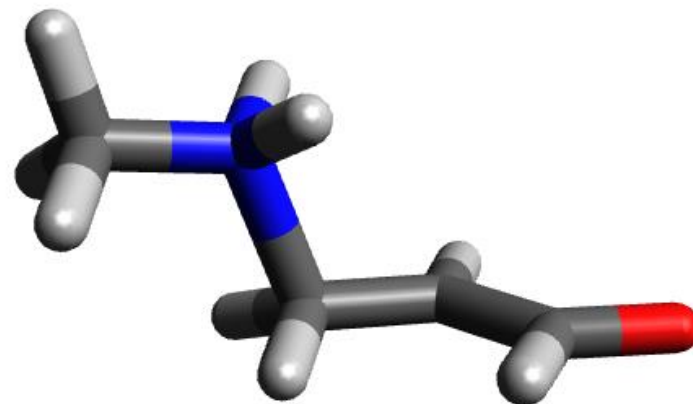
HF Hessian

SQM Hessian

AM_325
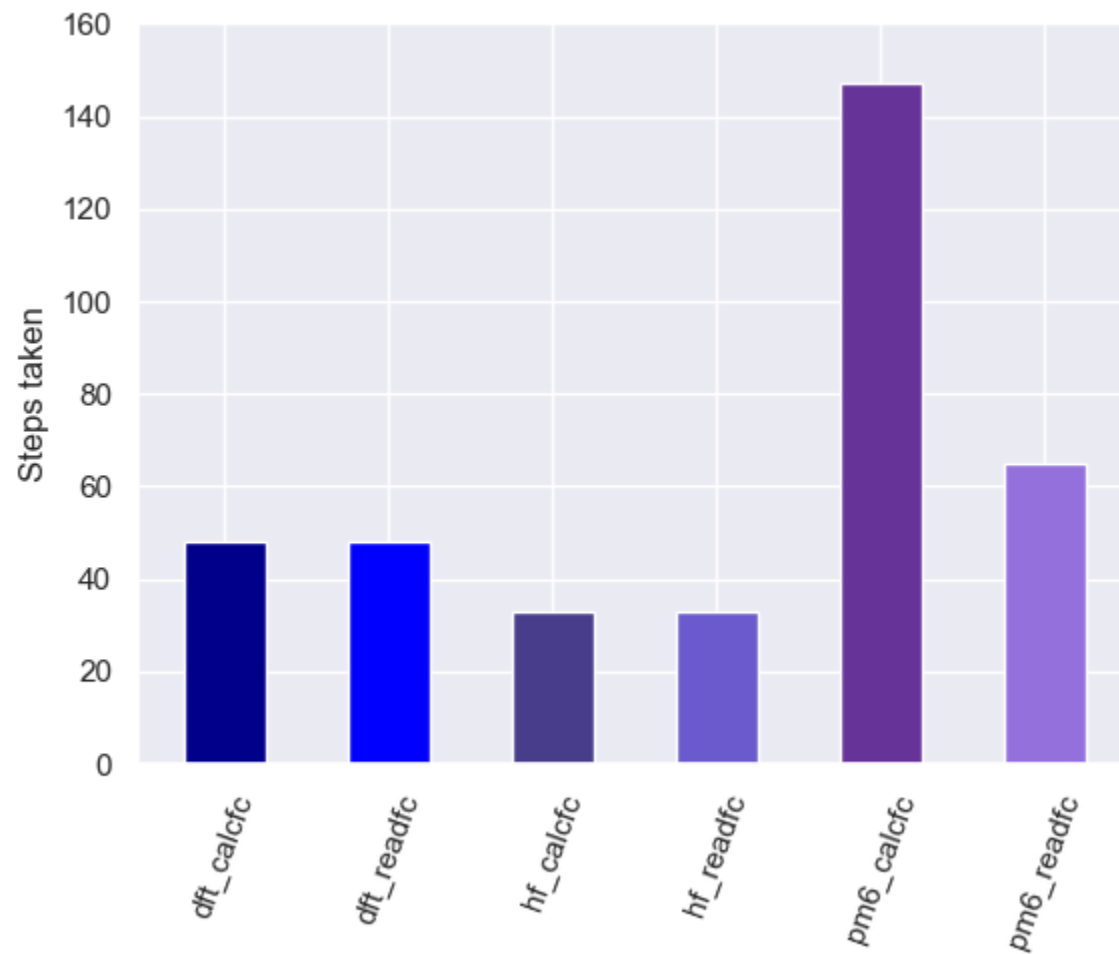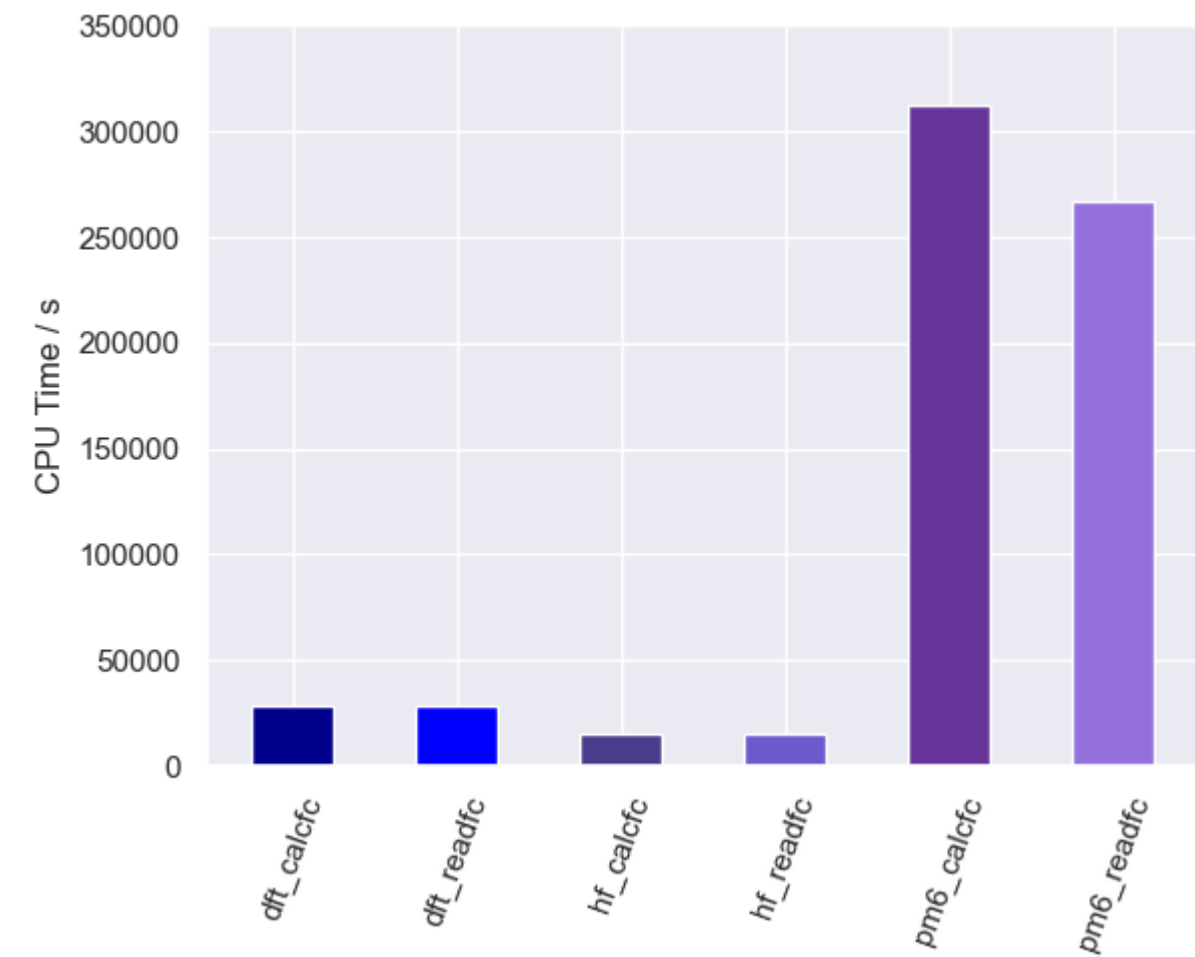
AM_325: conformations
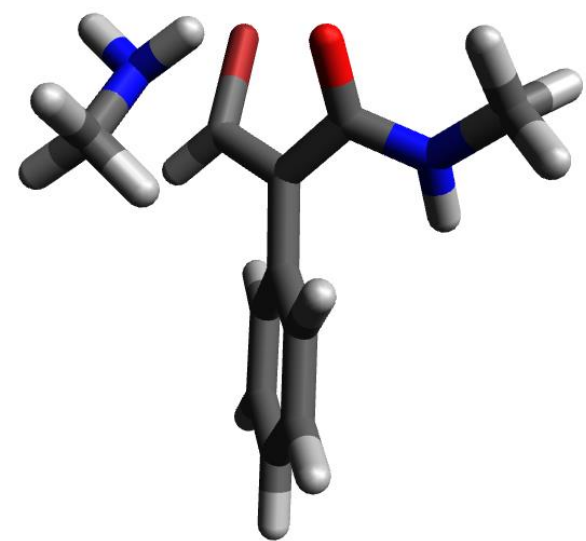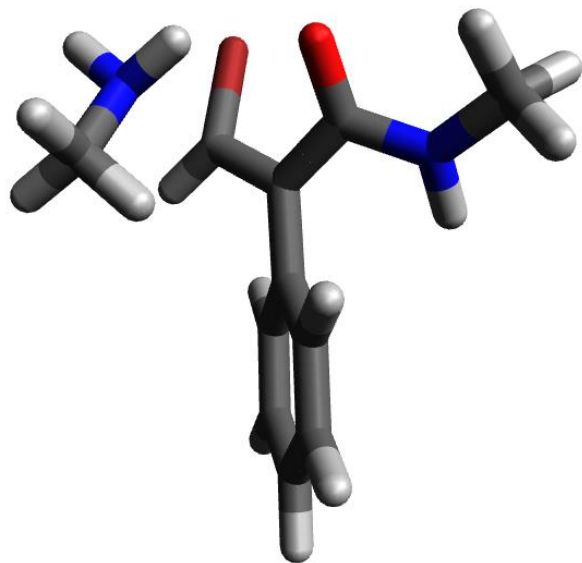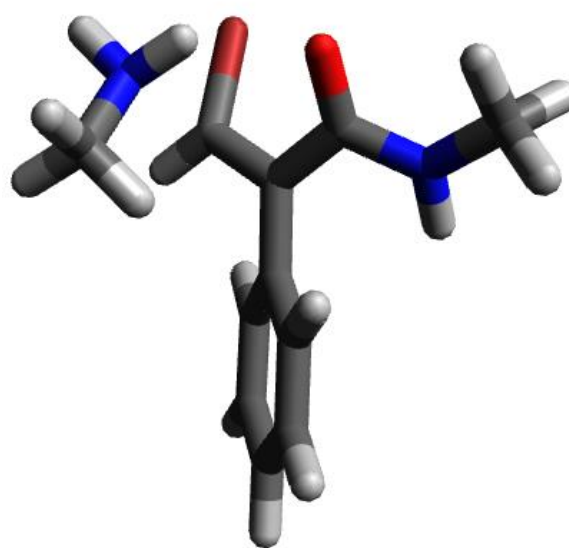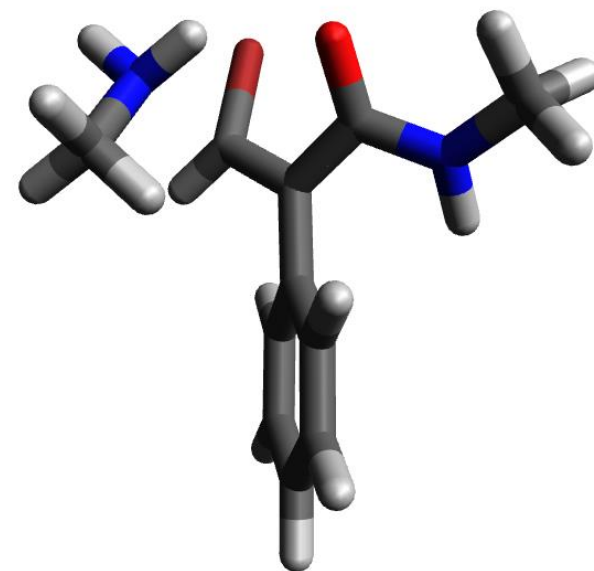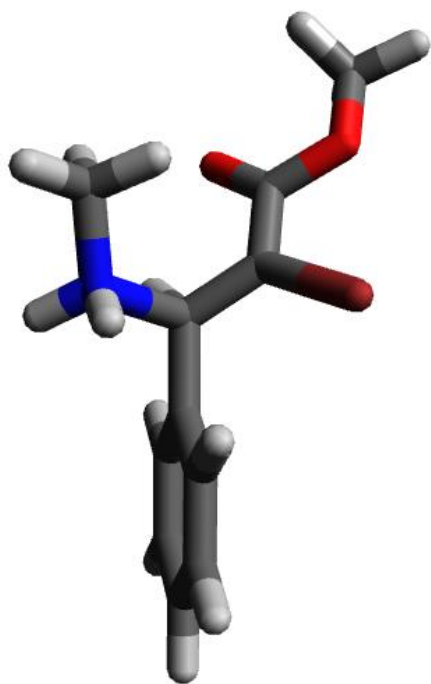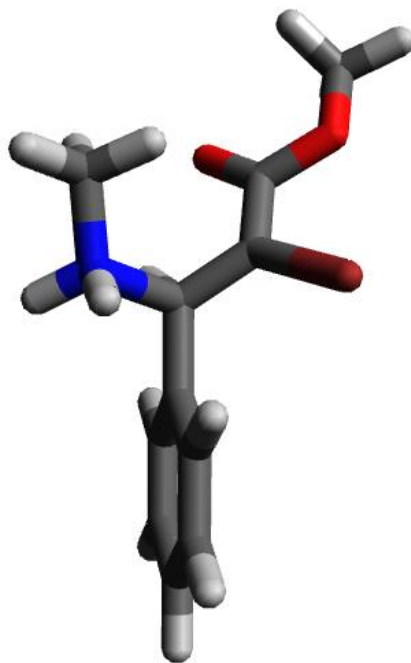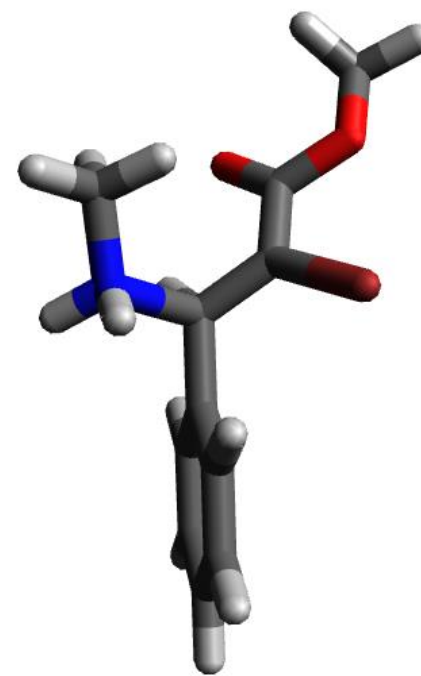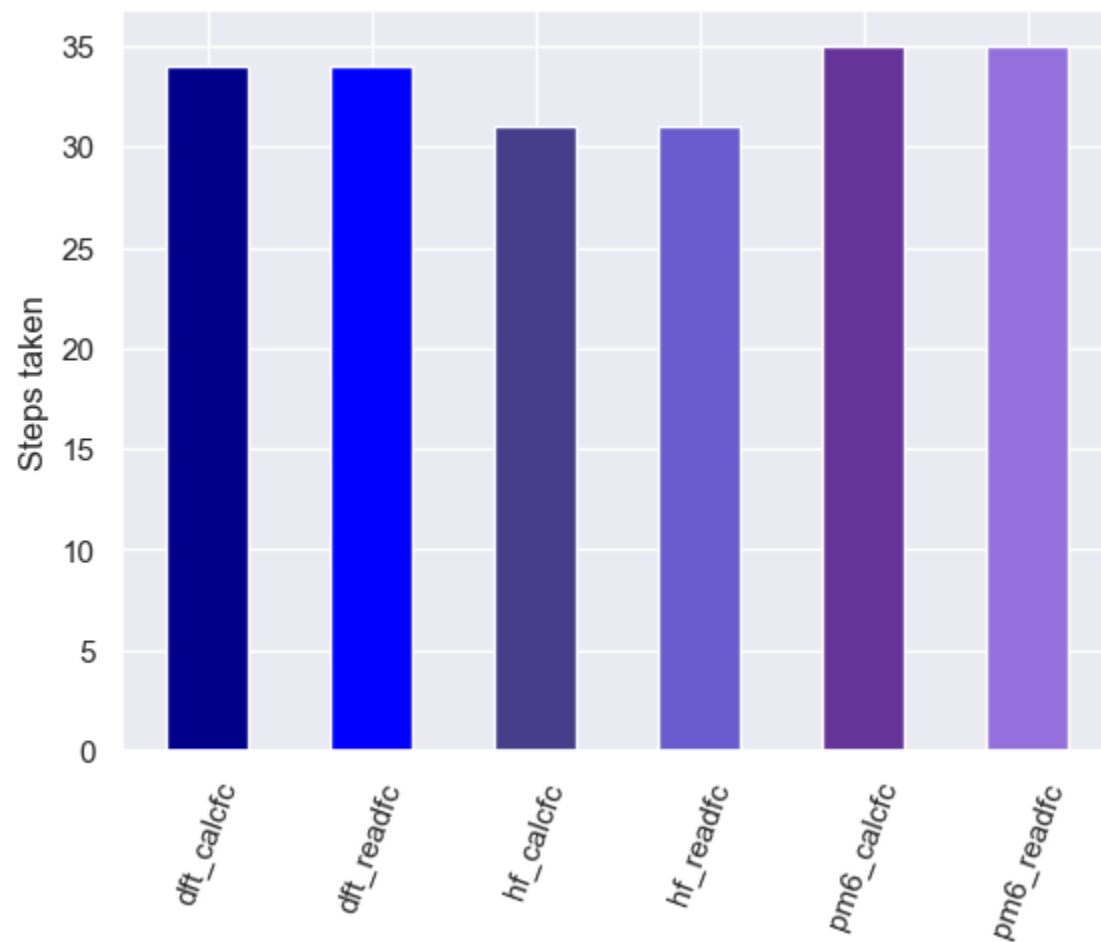

DFT Hessian   HF Hessian   SQM_calcfc   SQM_readfc
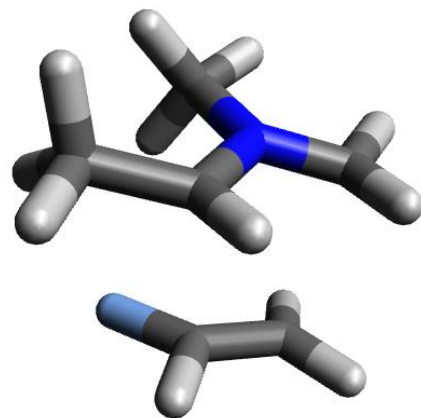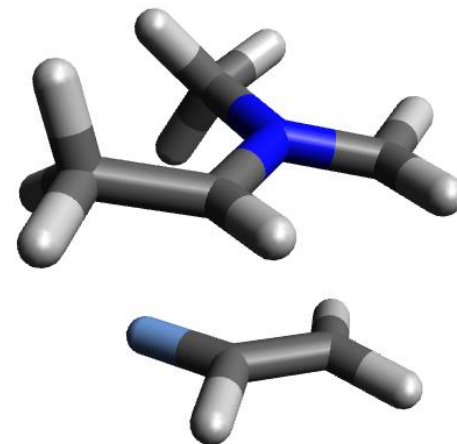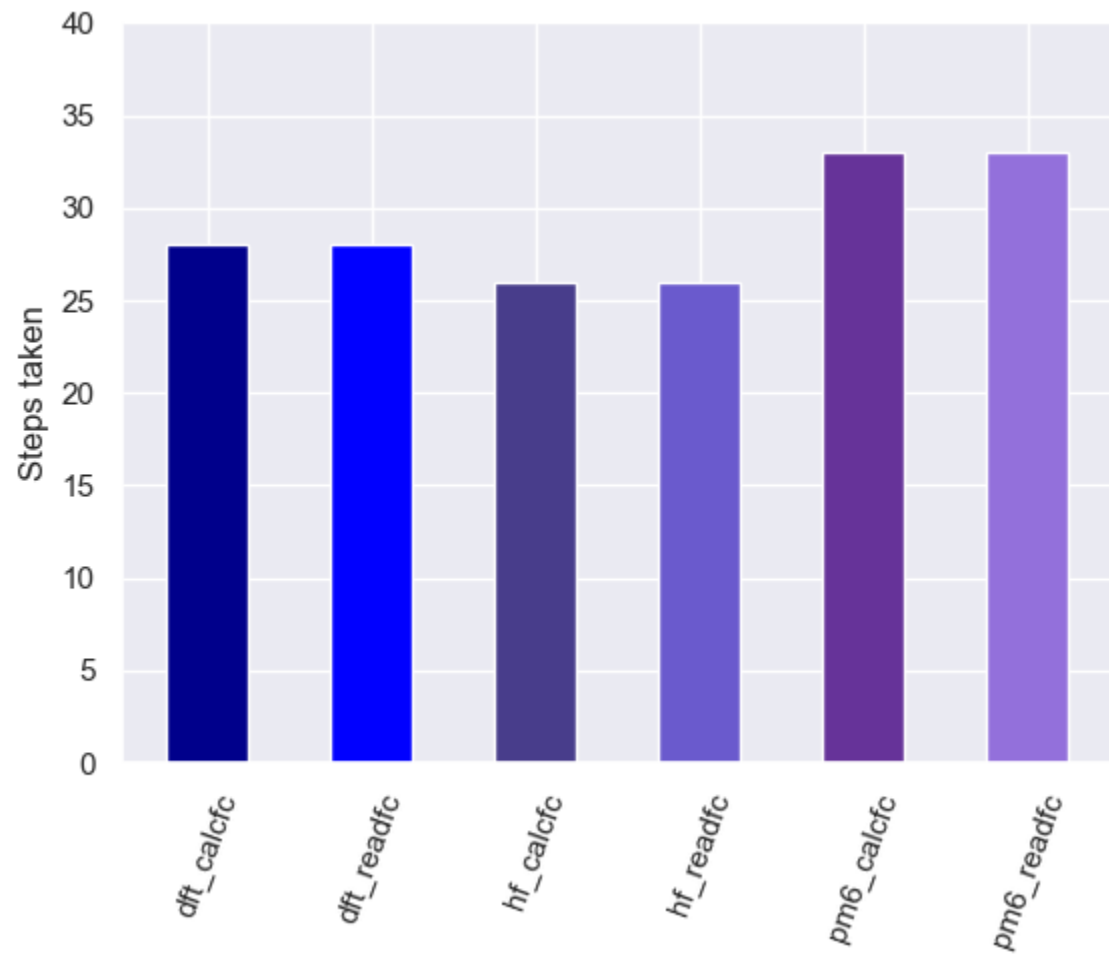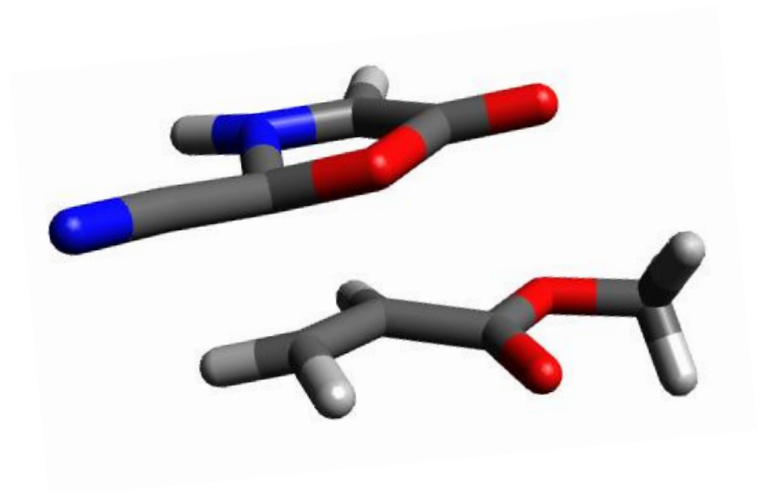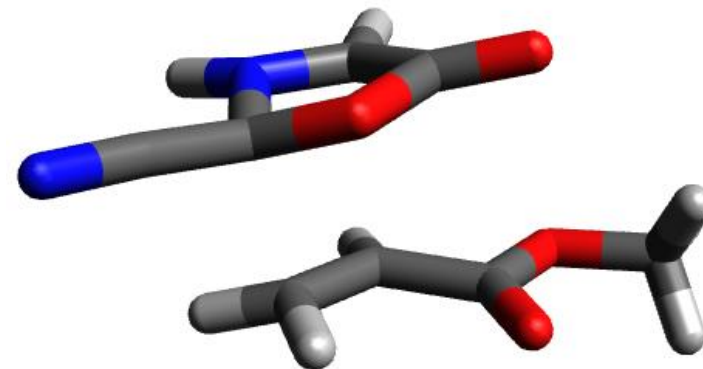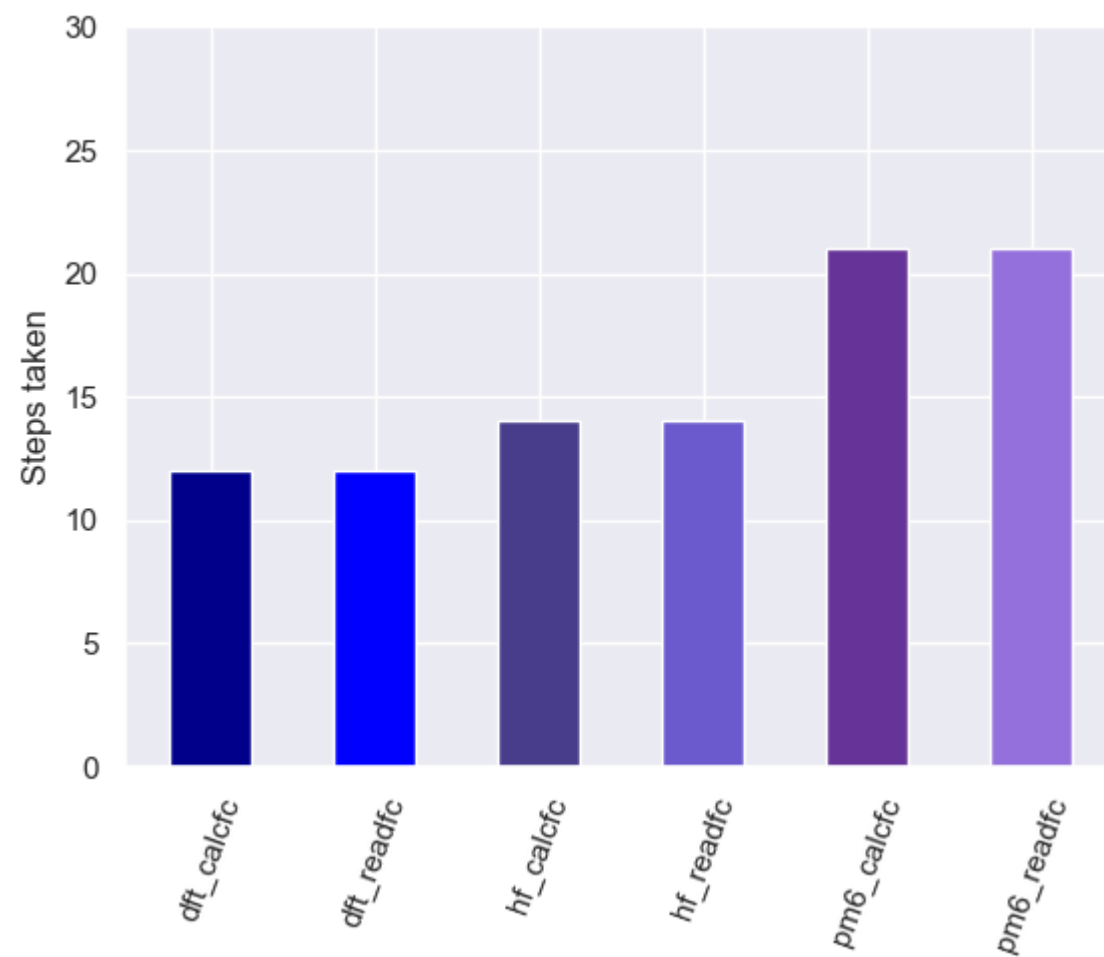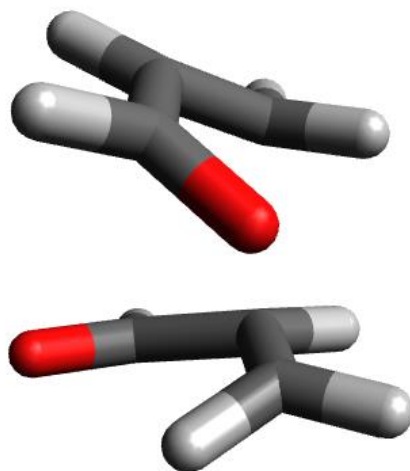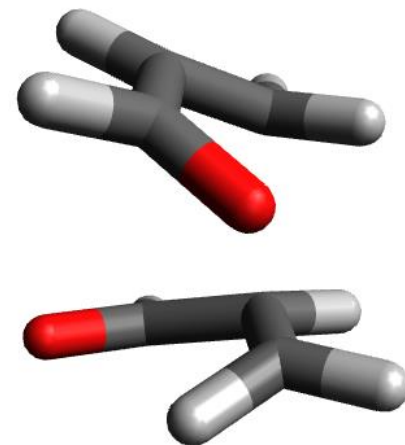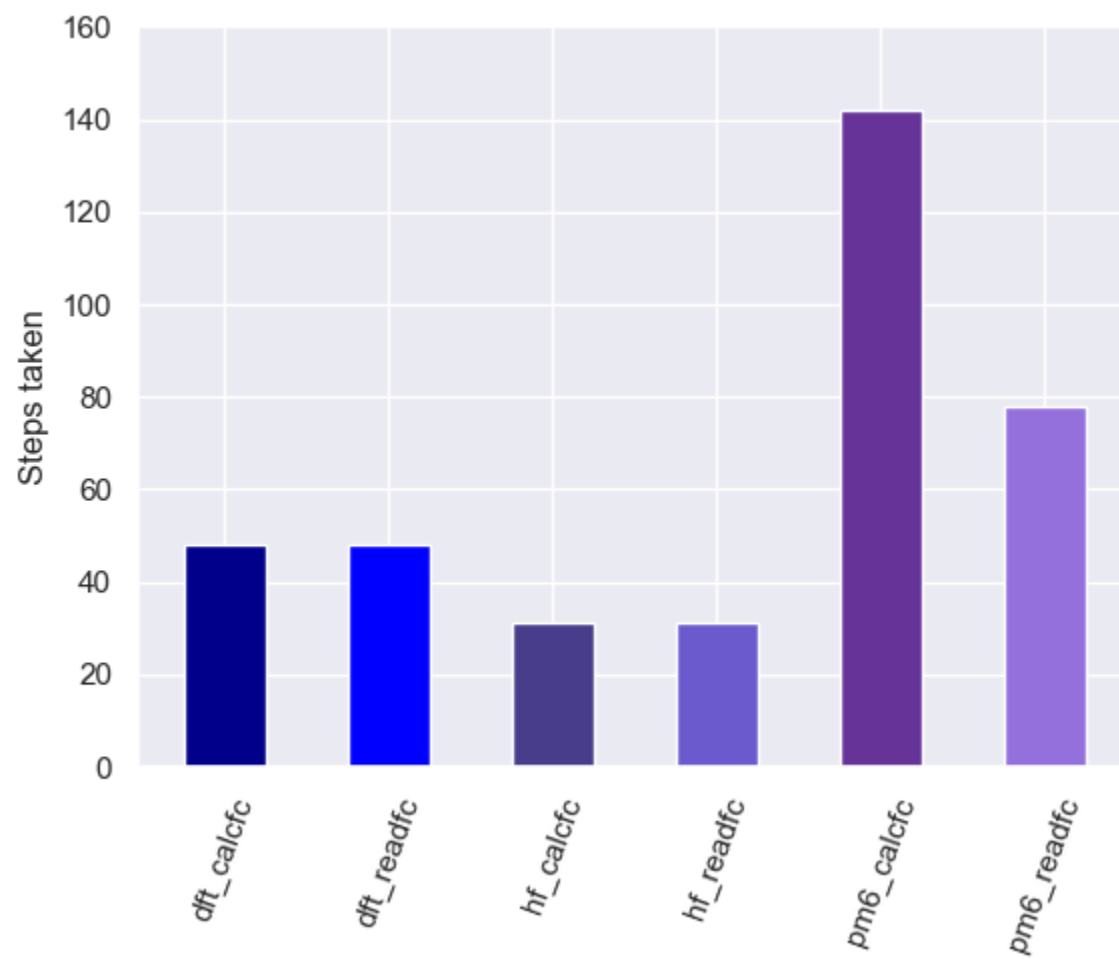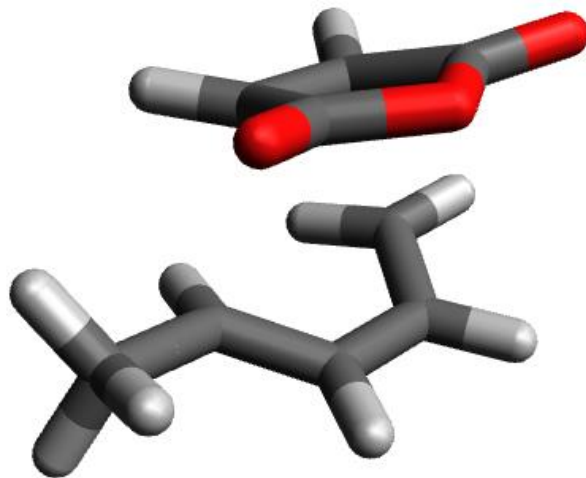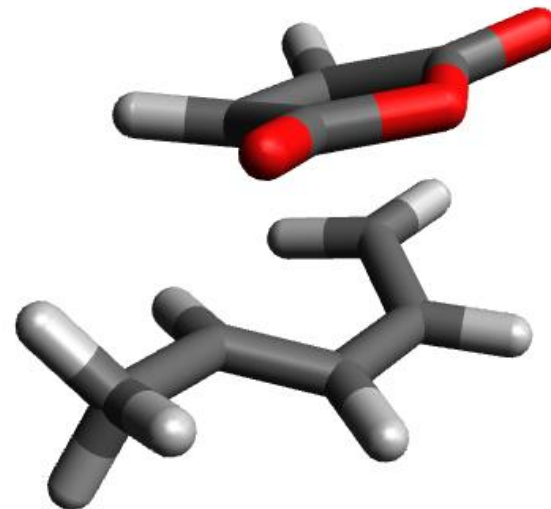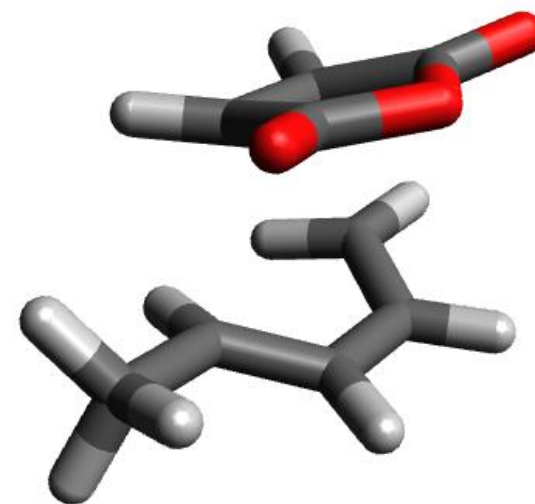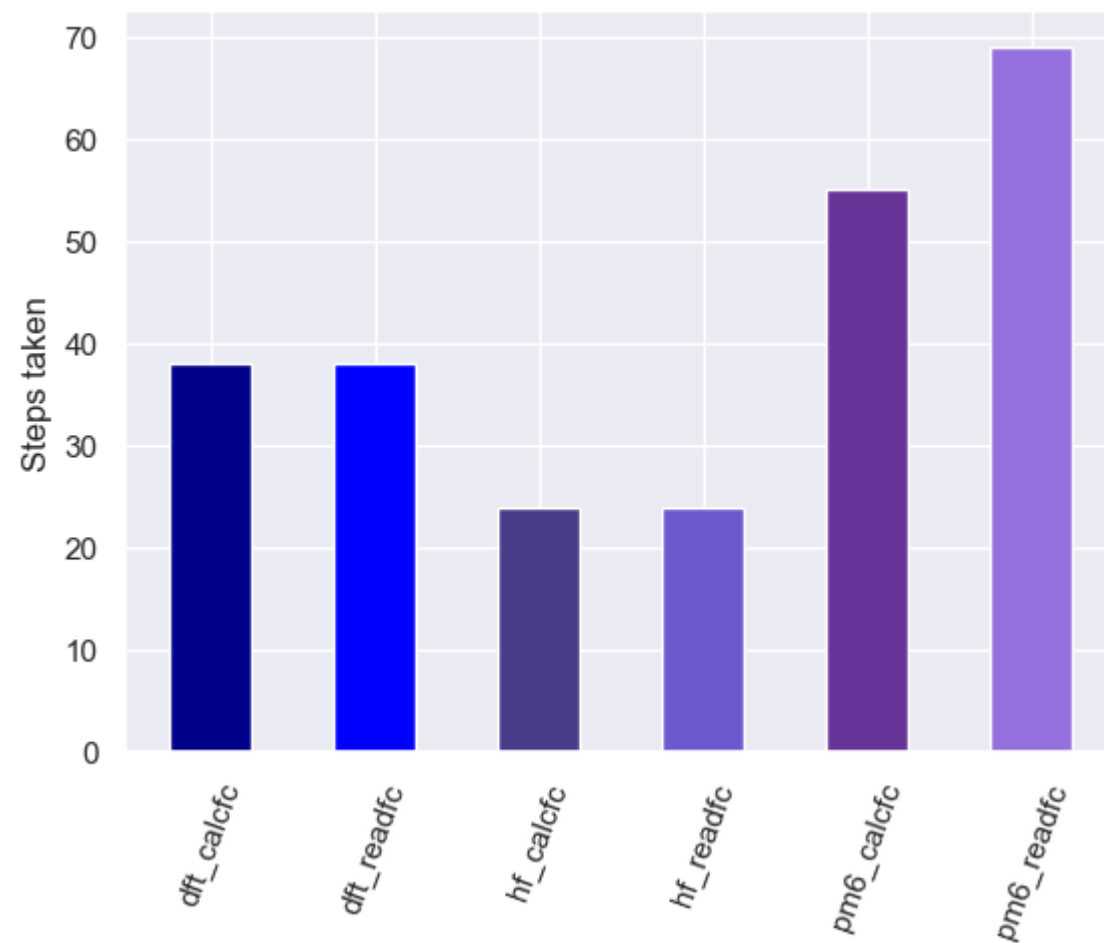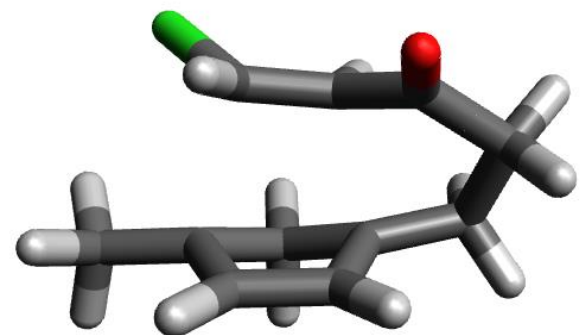
AM_898
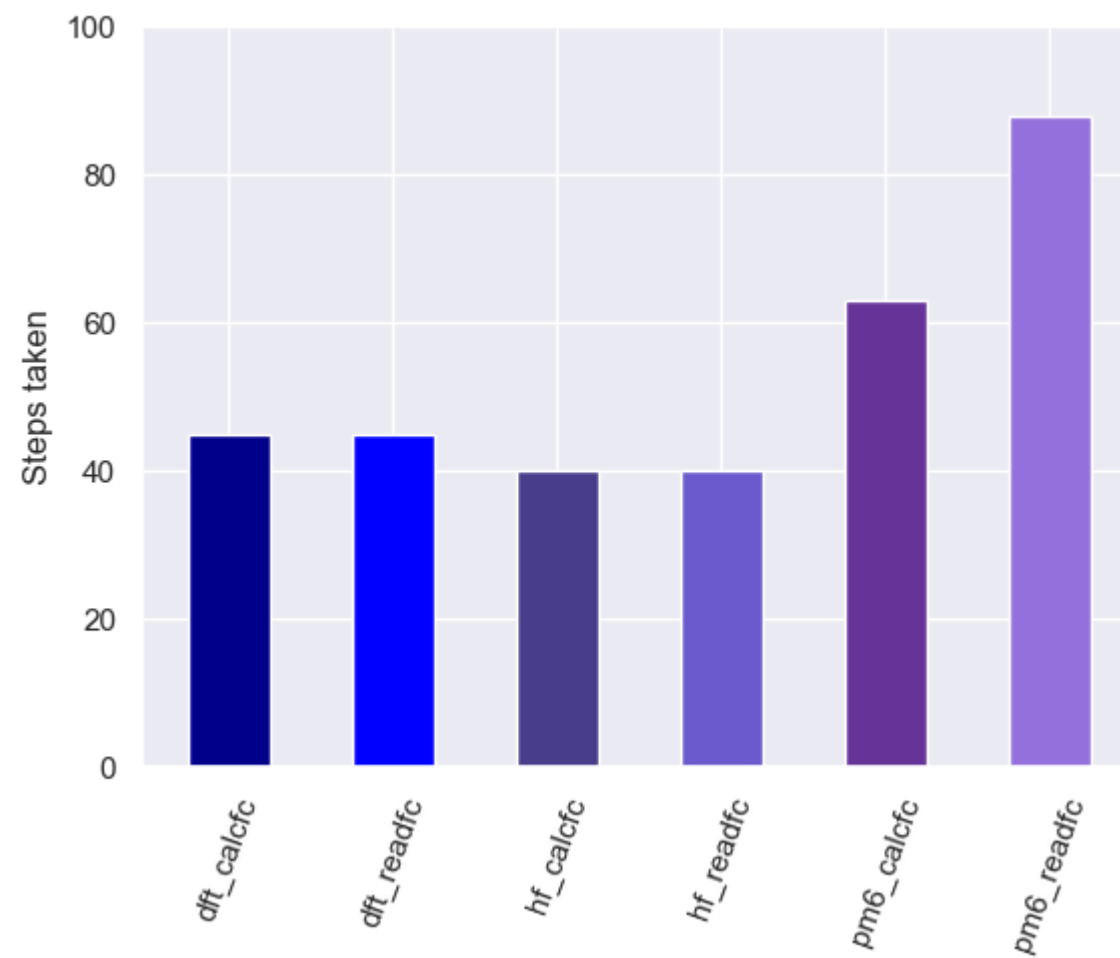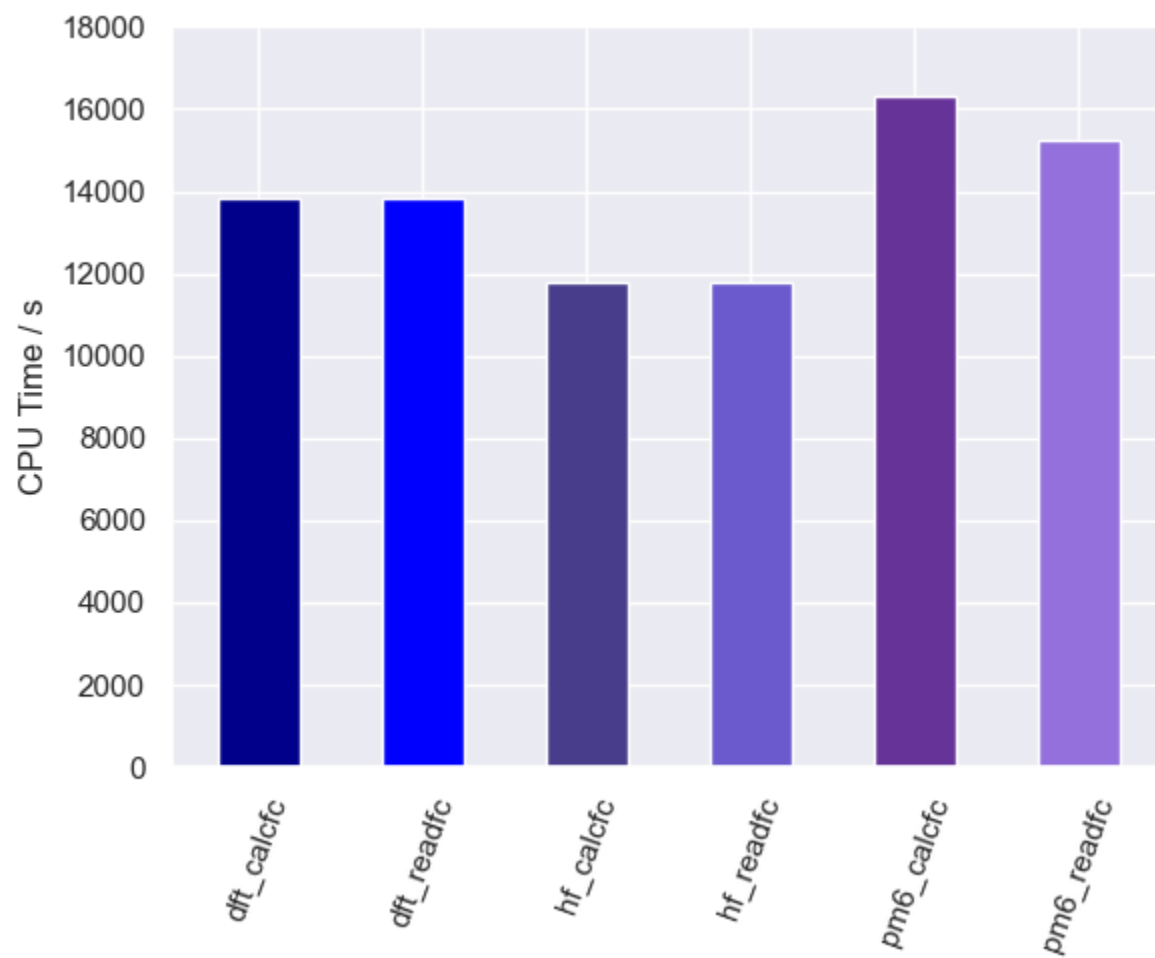
AM_898: conformations
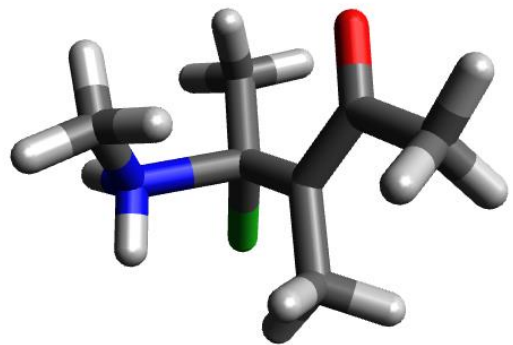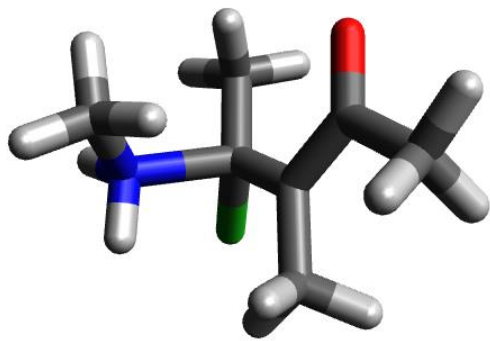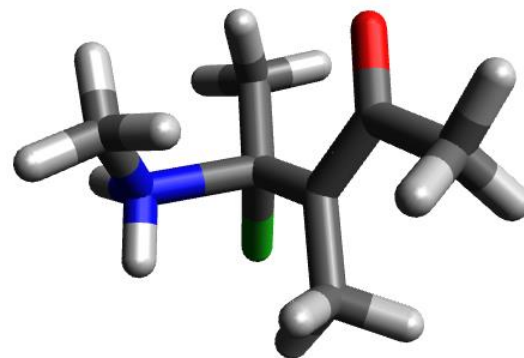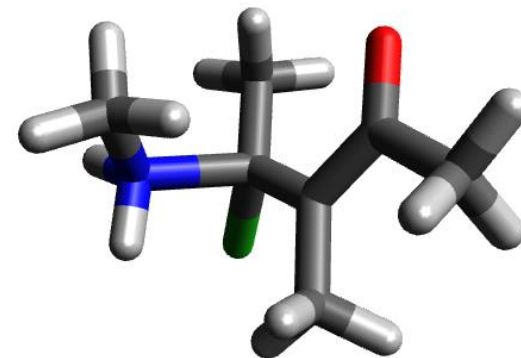
DFT Hessian   HF Hessian   SQM_calcfc   SQM_readfc

AM_68
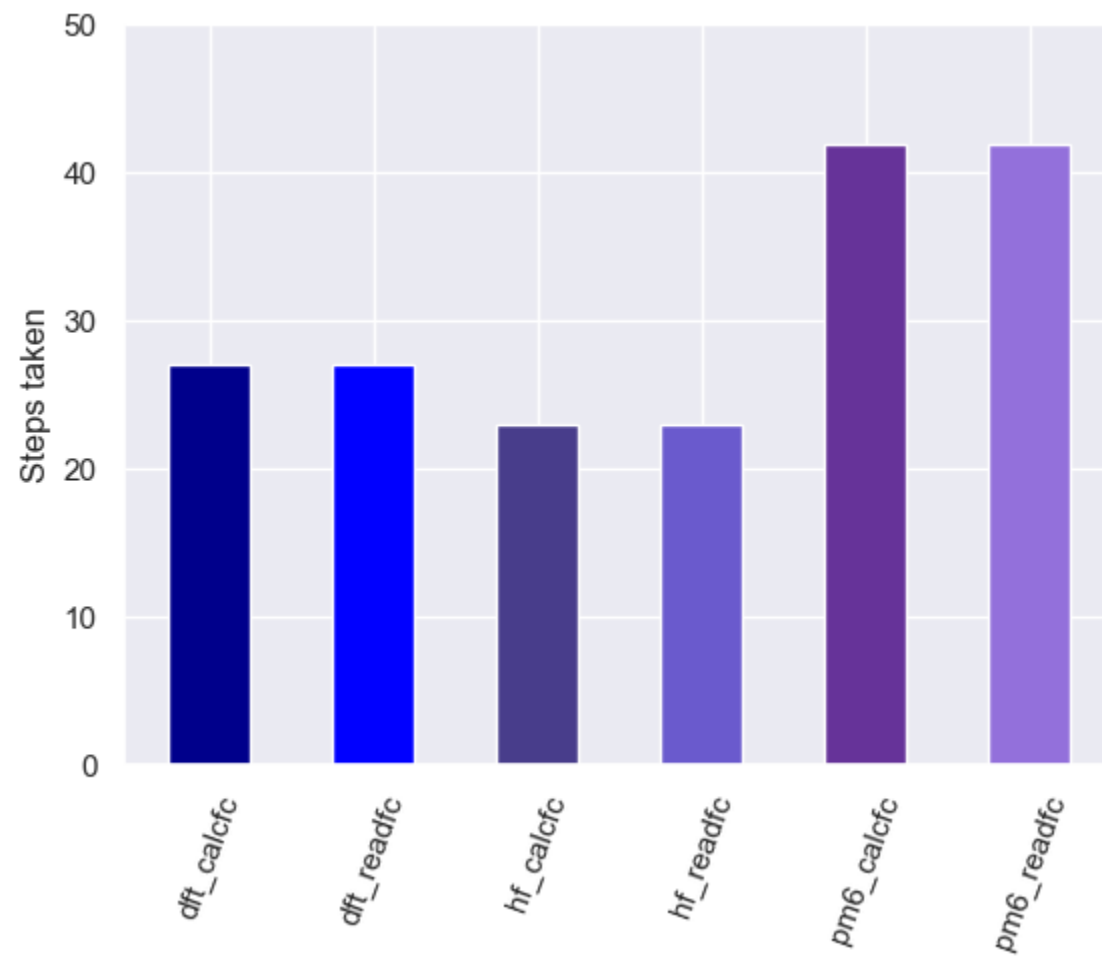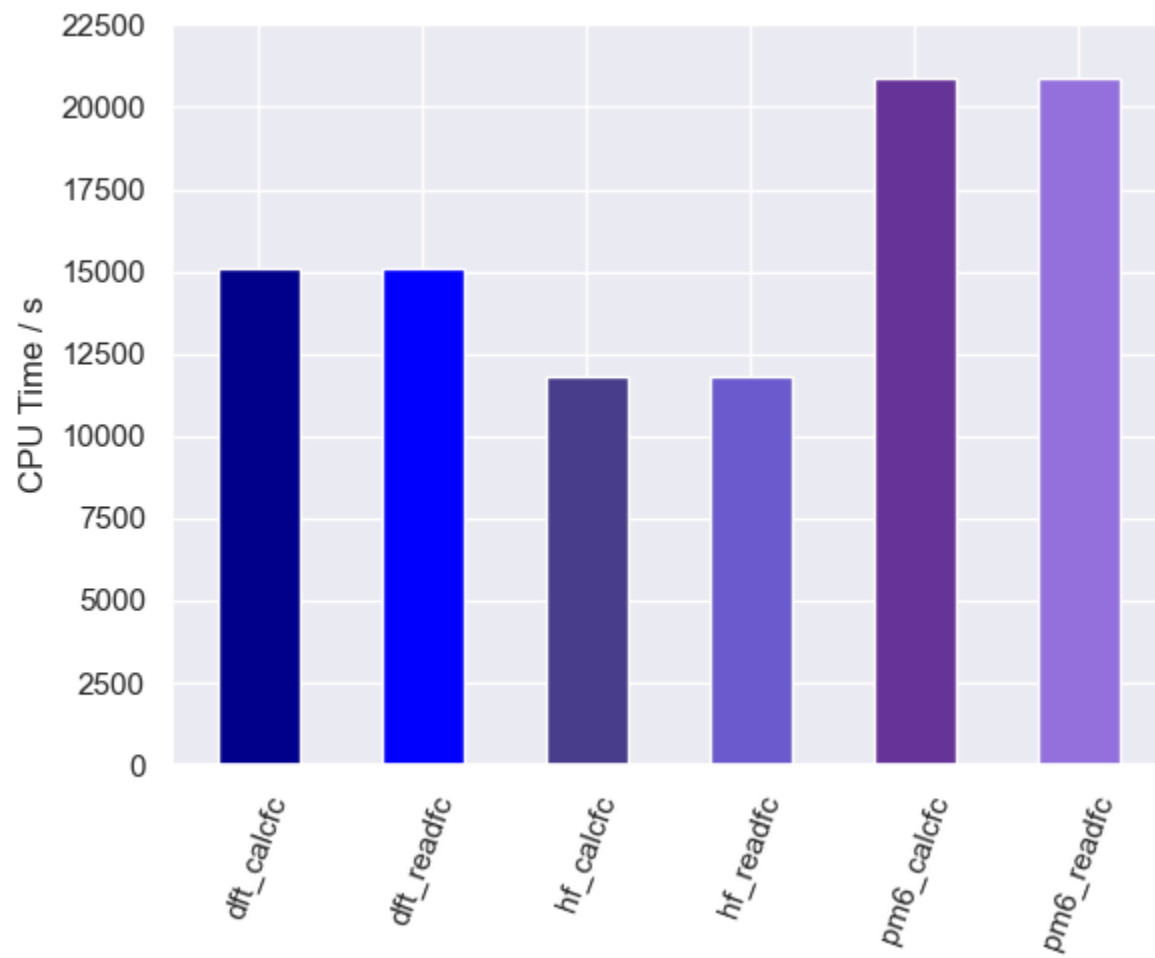
AM_68: conformations

DFT Hessian   HF Hessian   SQM Hessian

allylic
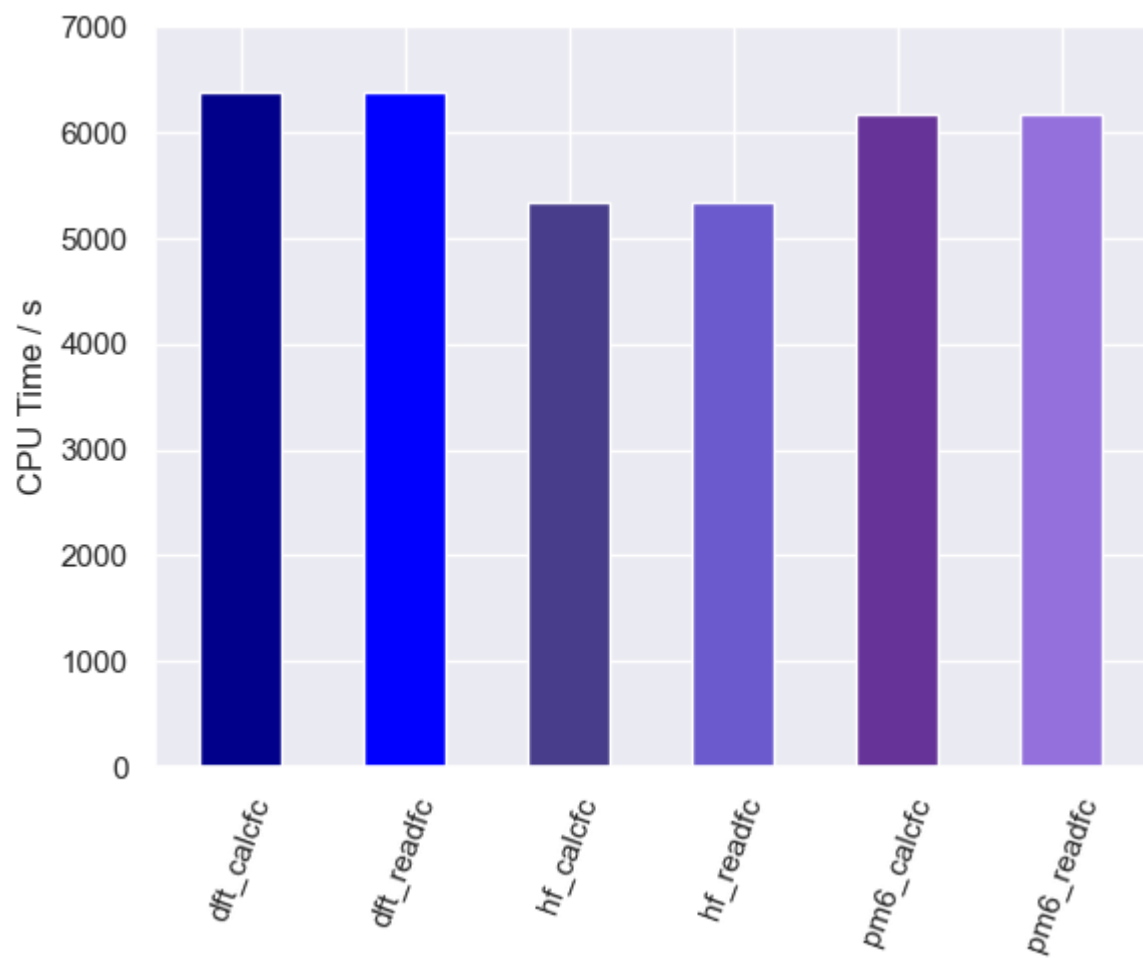
allylic: conformations
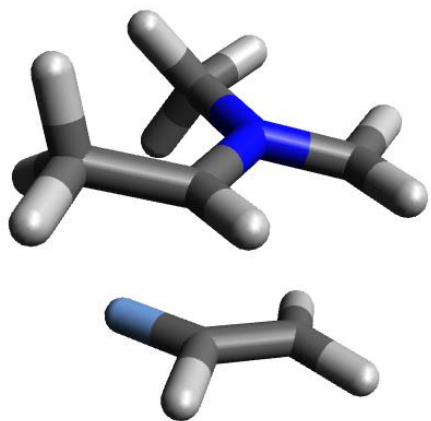
DFT Hessian

HF Hessian

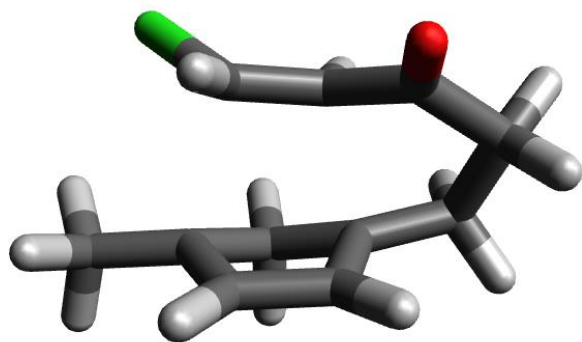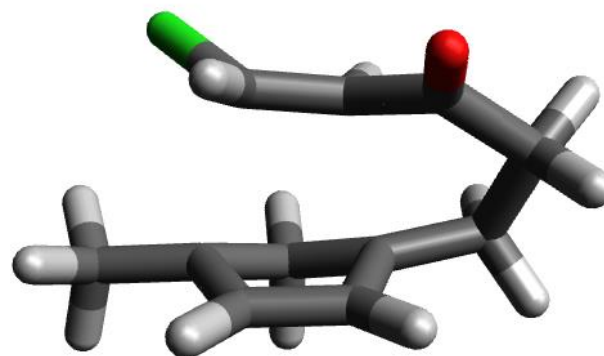SQM Hessian

cyclic
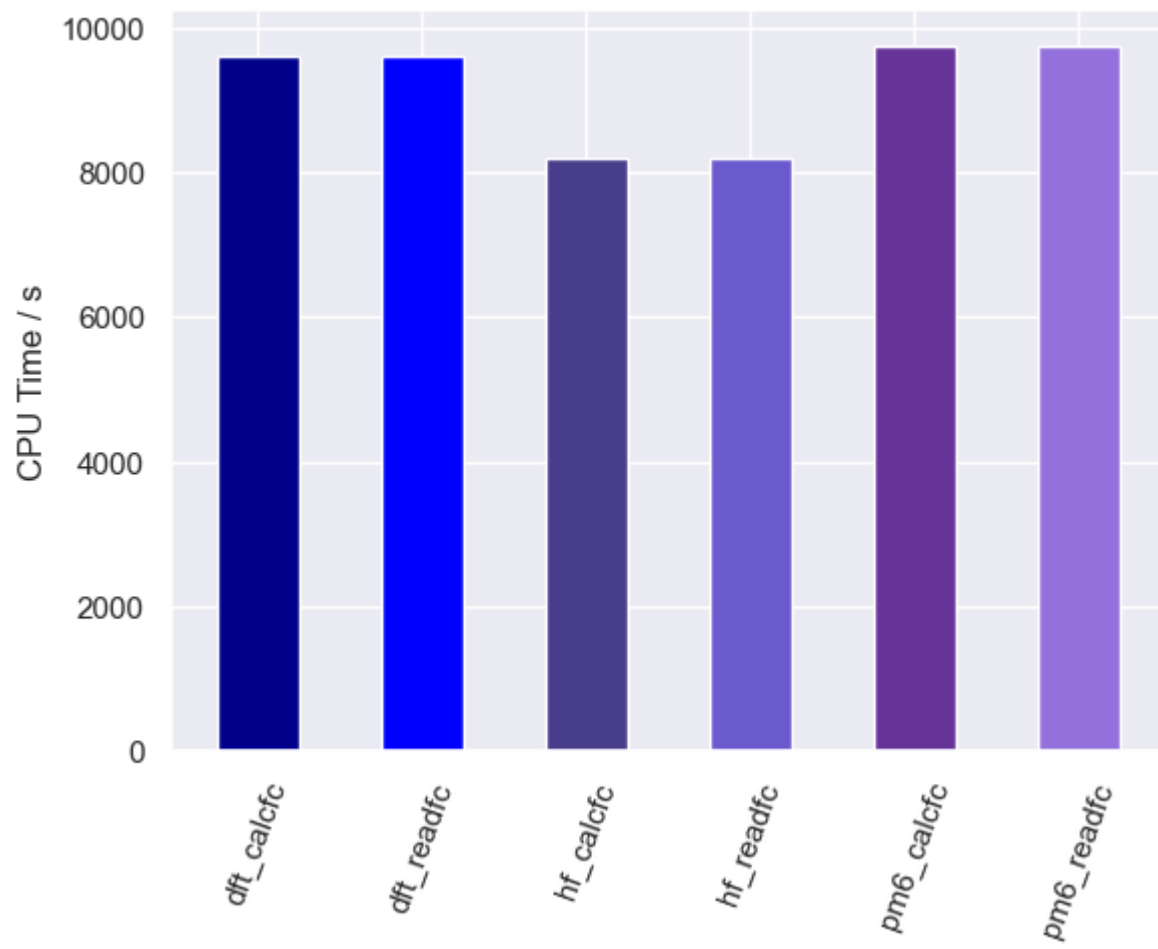
cyclic: conformations
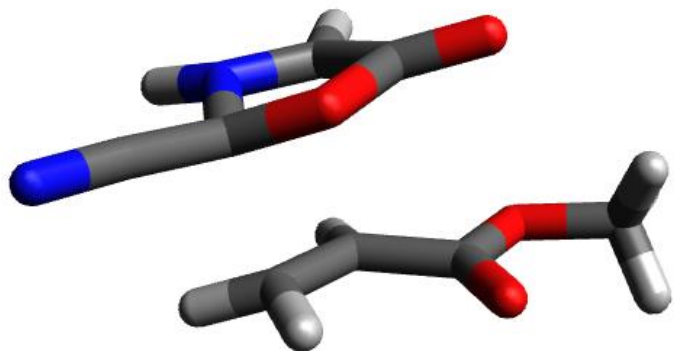
DFT Hessian

HF Hessian

SQM Hessian

DA_1
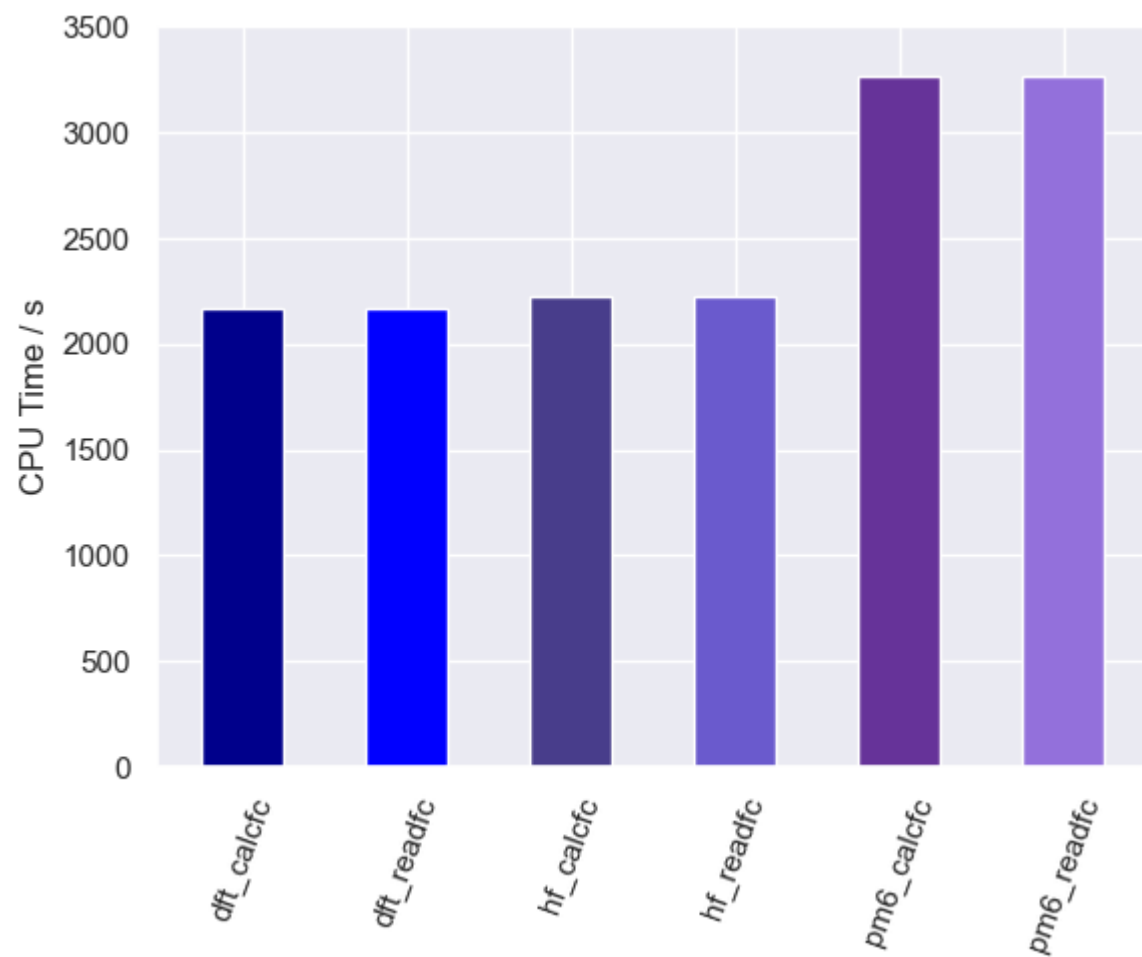
DA_1: conformations
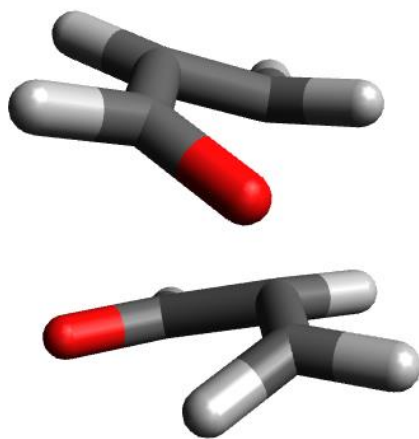
DFT Hessian

HF Hessian

SQM Hessian

DA_2
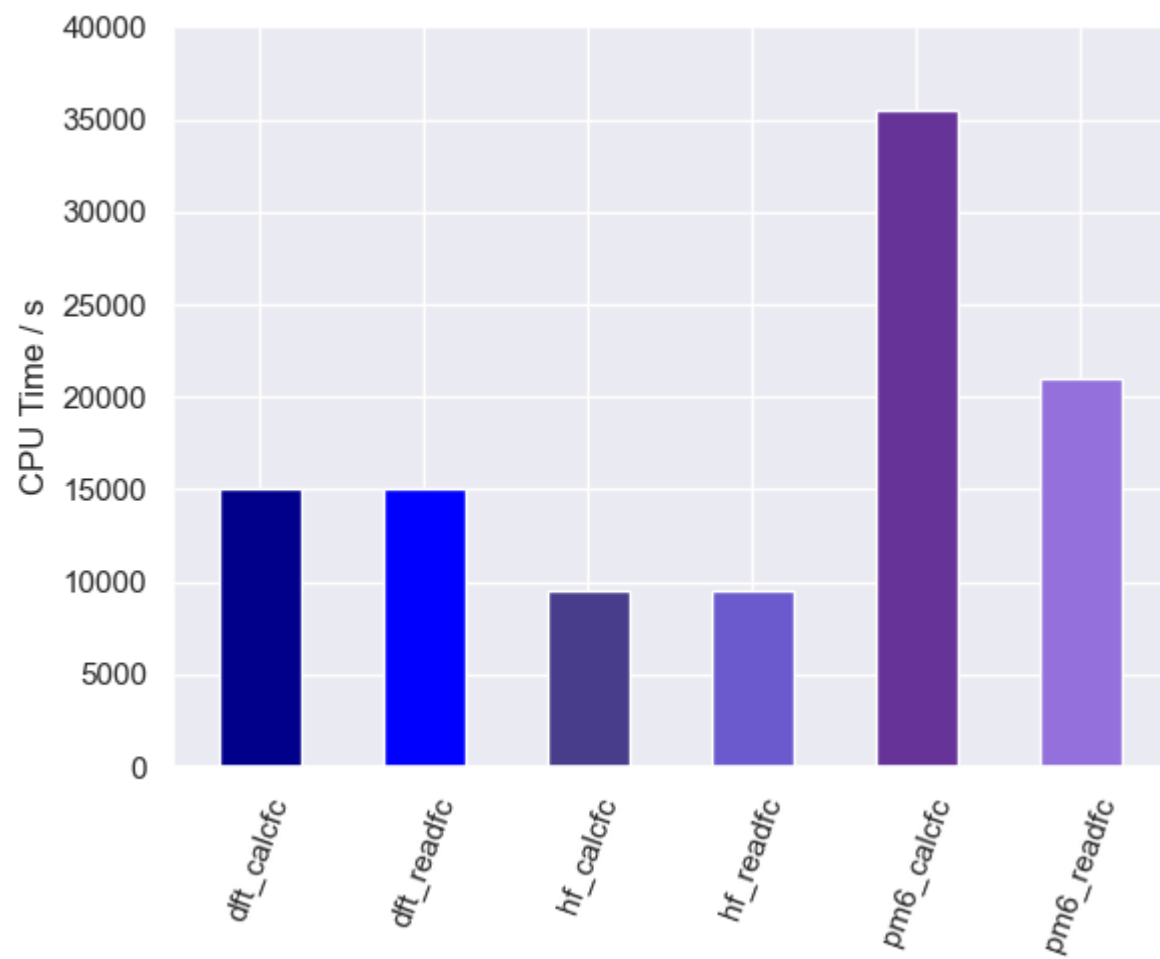
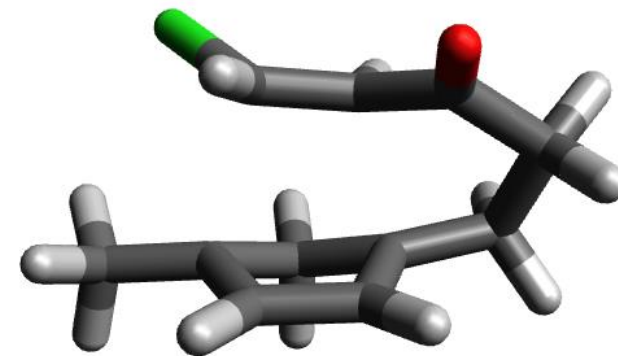# DA_2: conformations
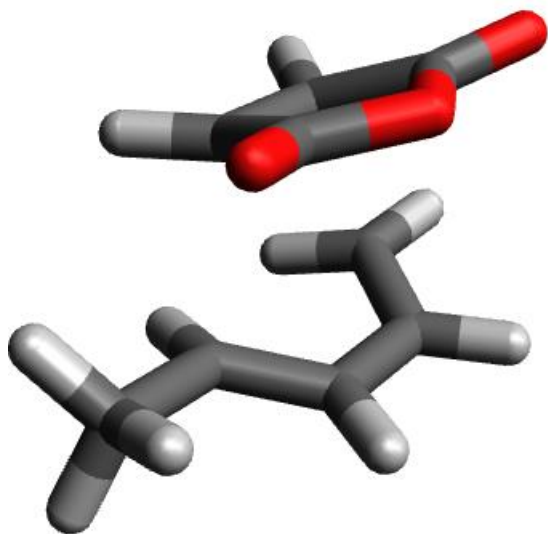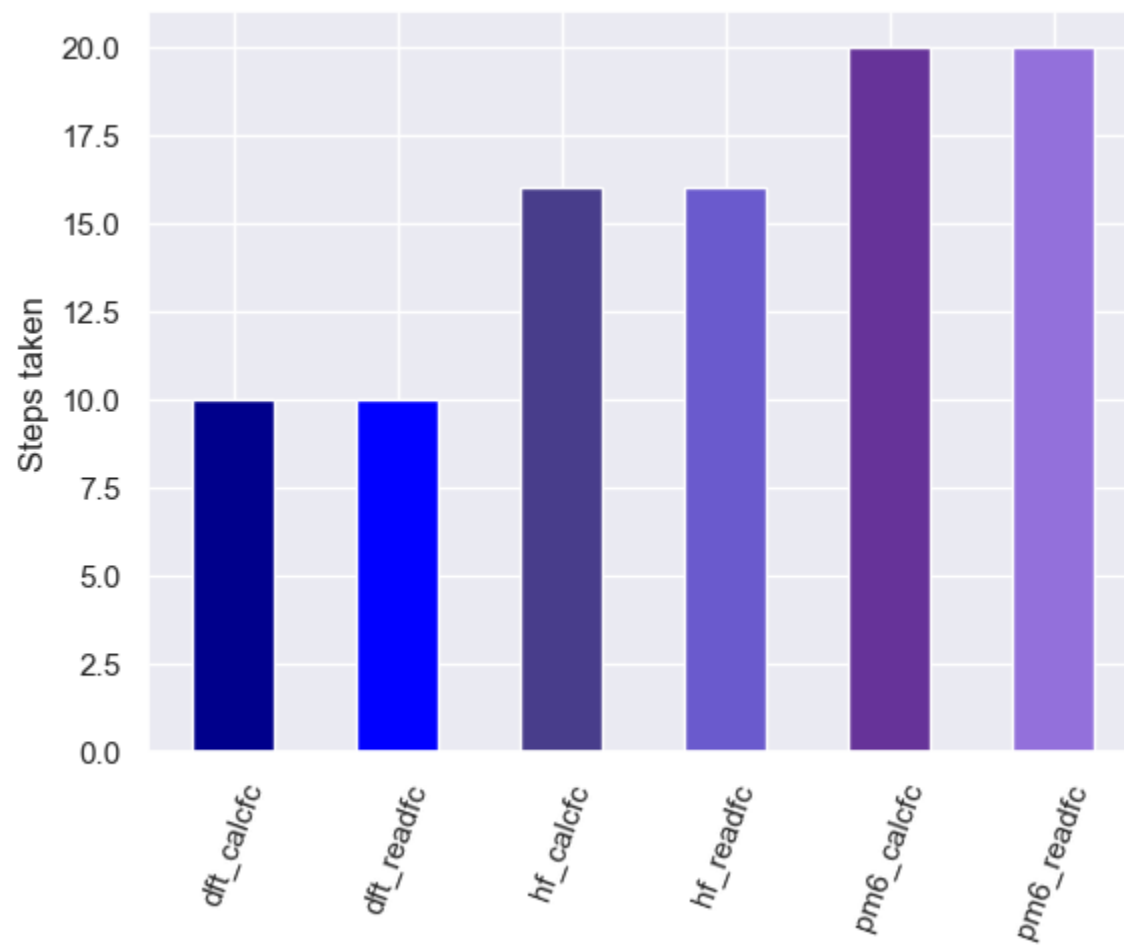

DFT Hessian   HF Hessian   SQM_calcfc   SQM_readfc

IMDA_1
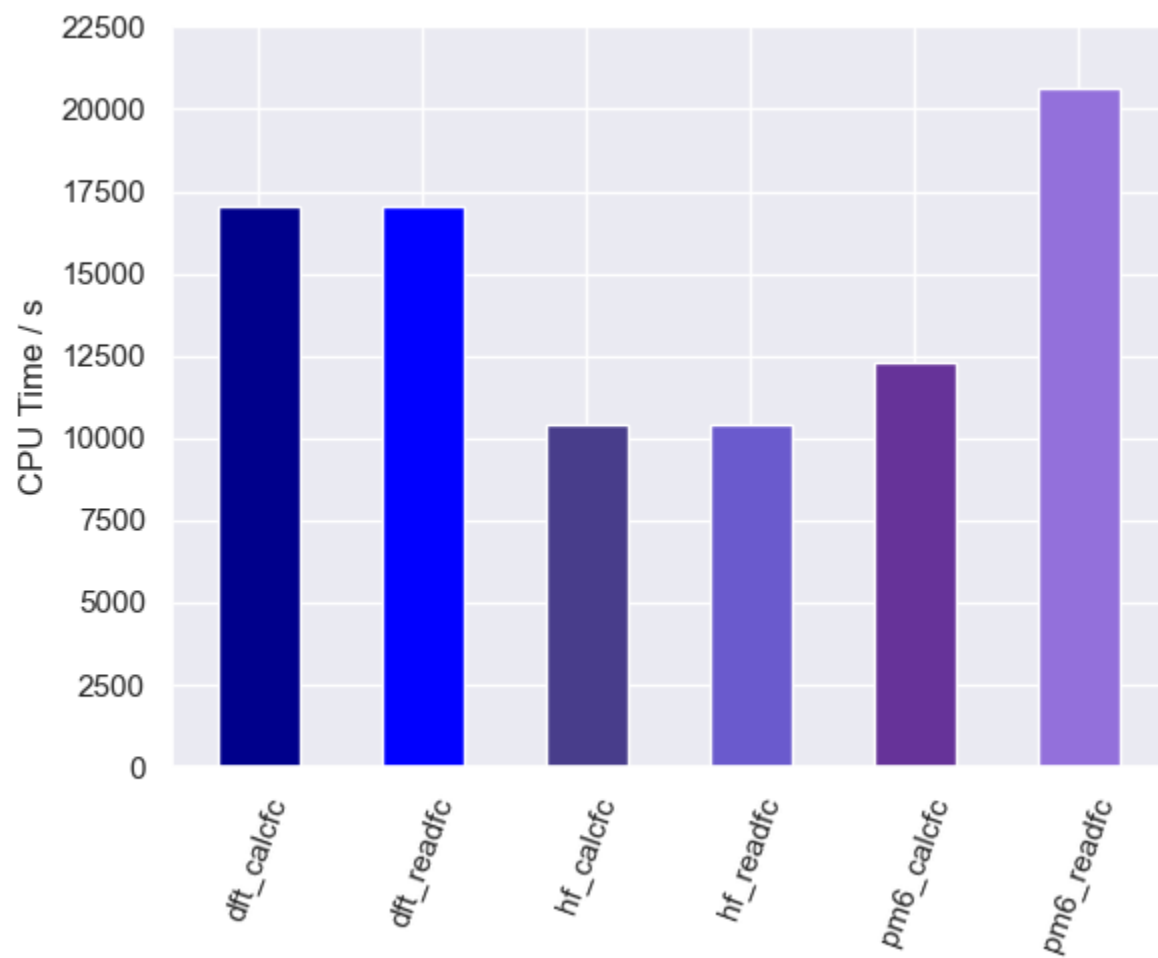
IMDA_1: conformations

DFT Hessian   HF Hessian   SQM_calcfc   SQM_readfc
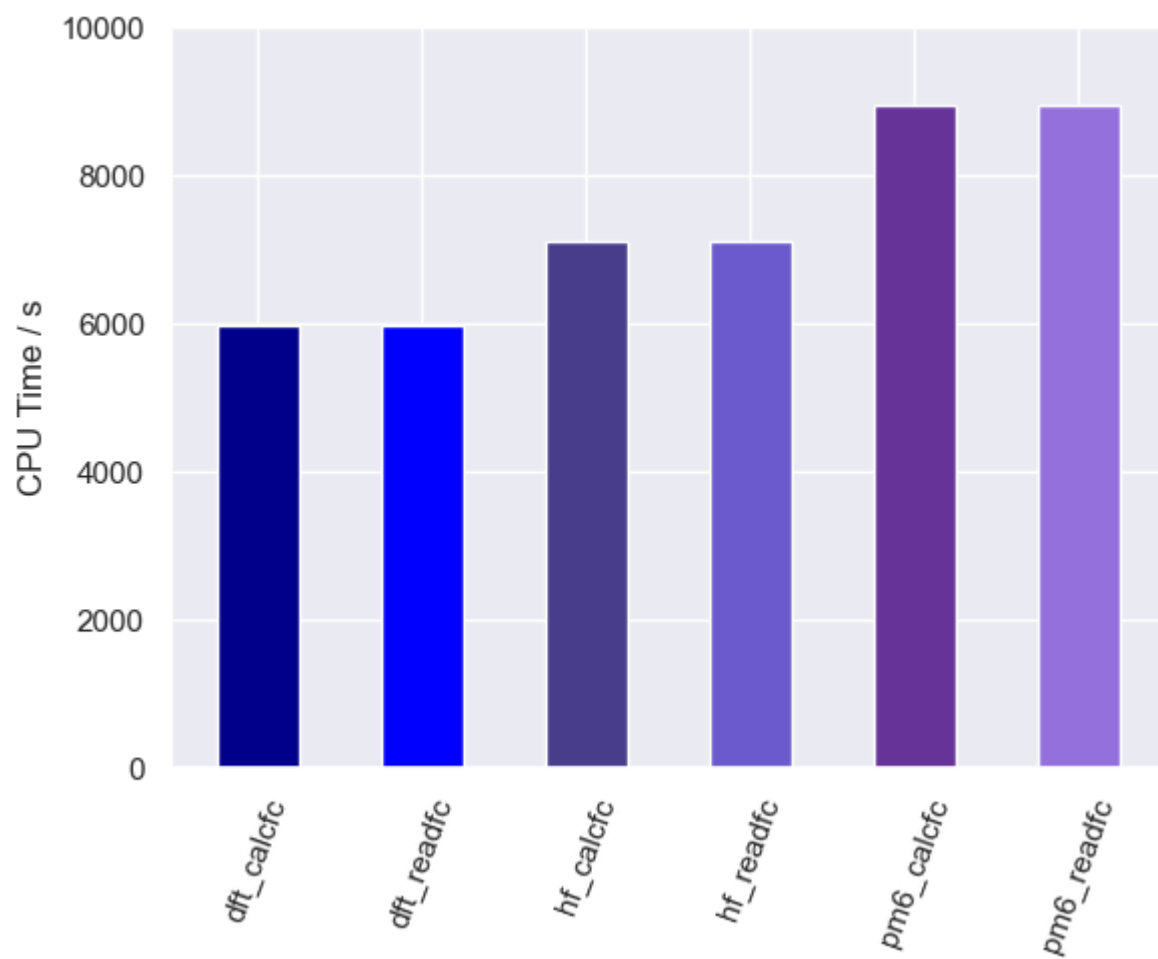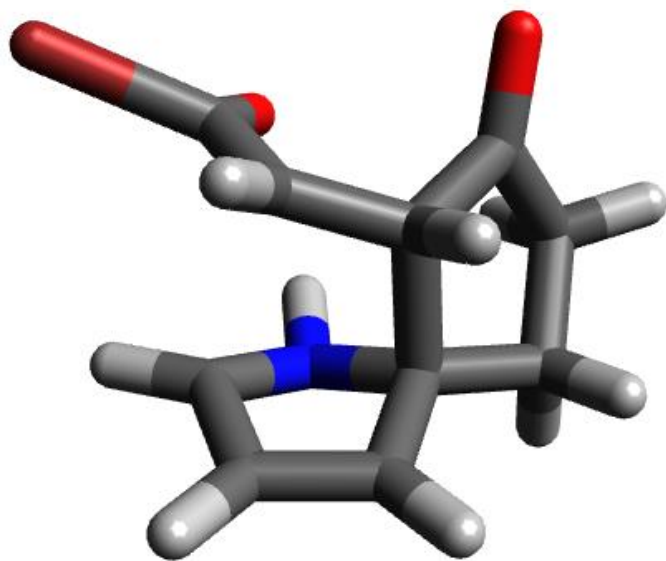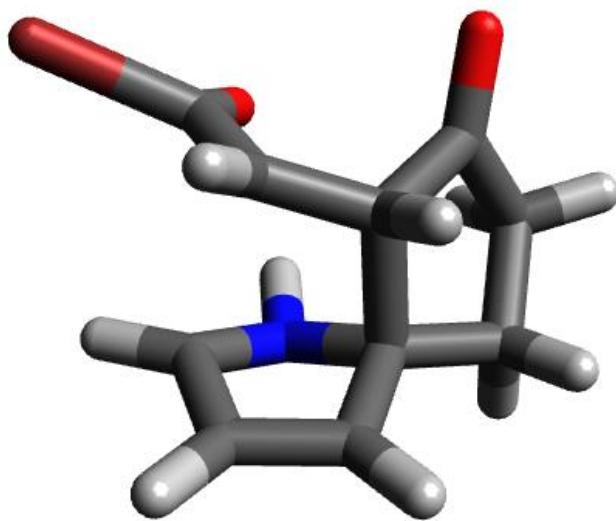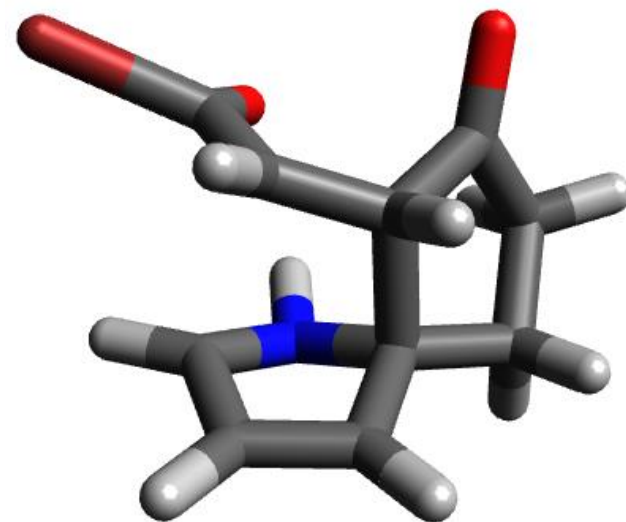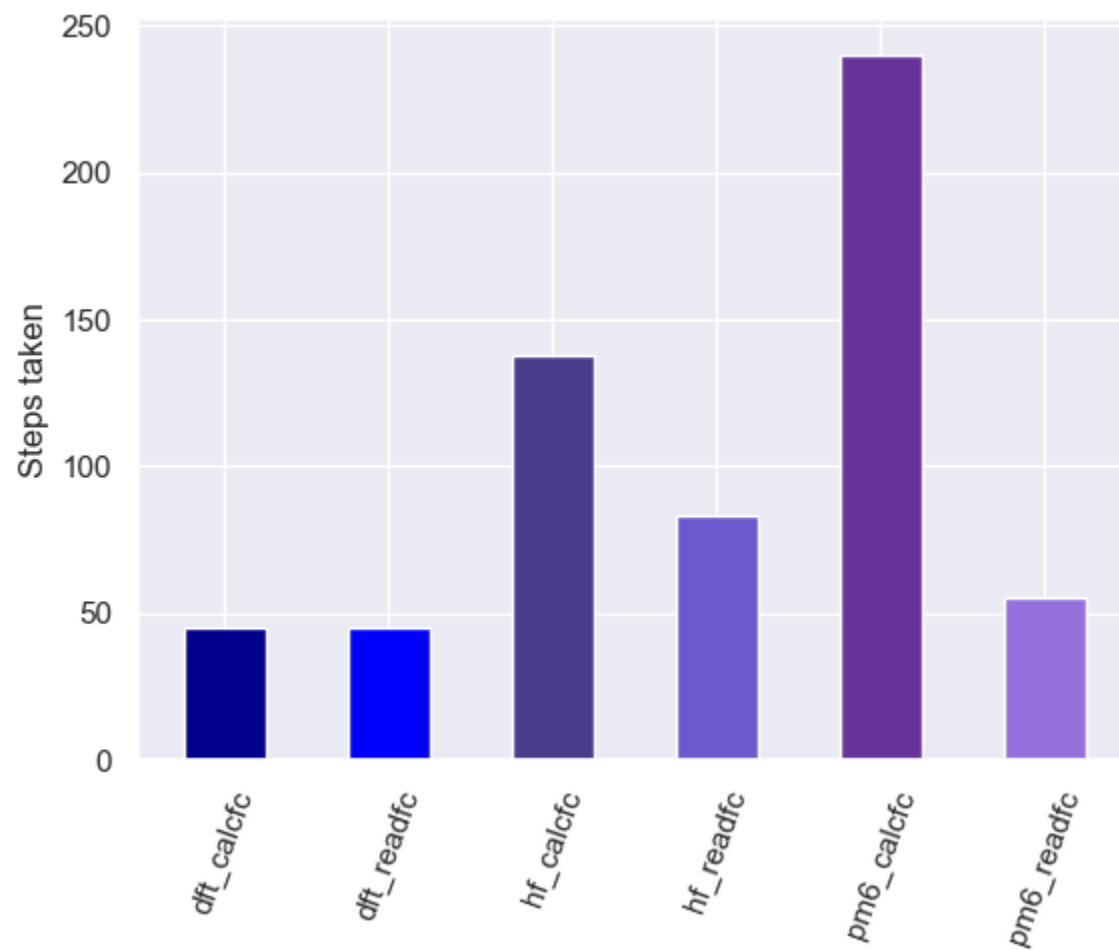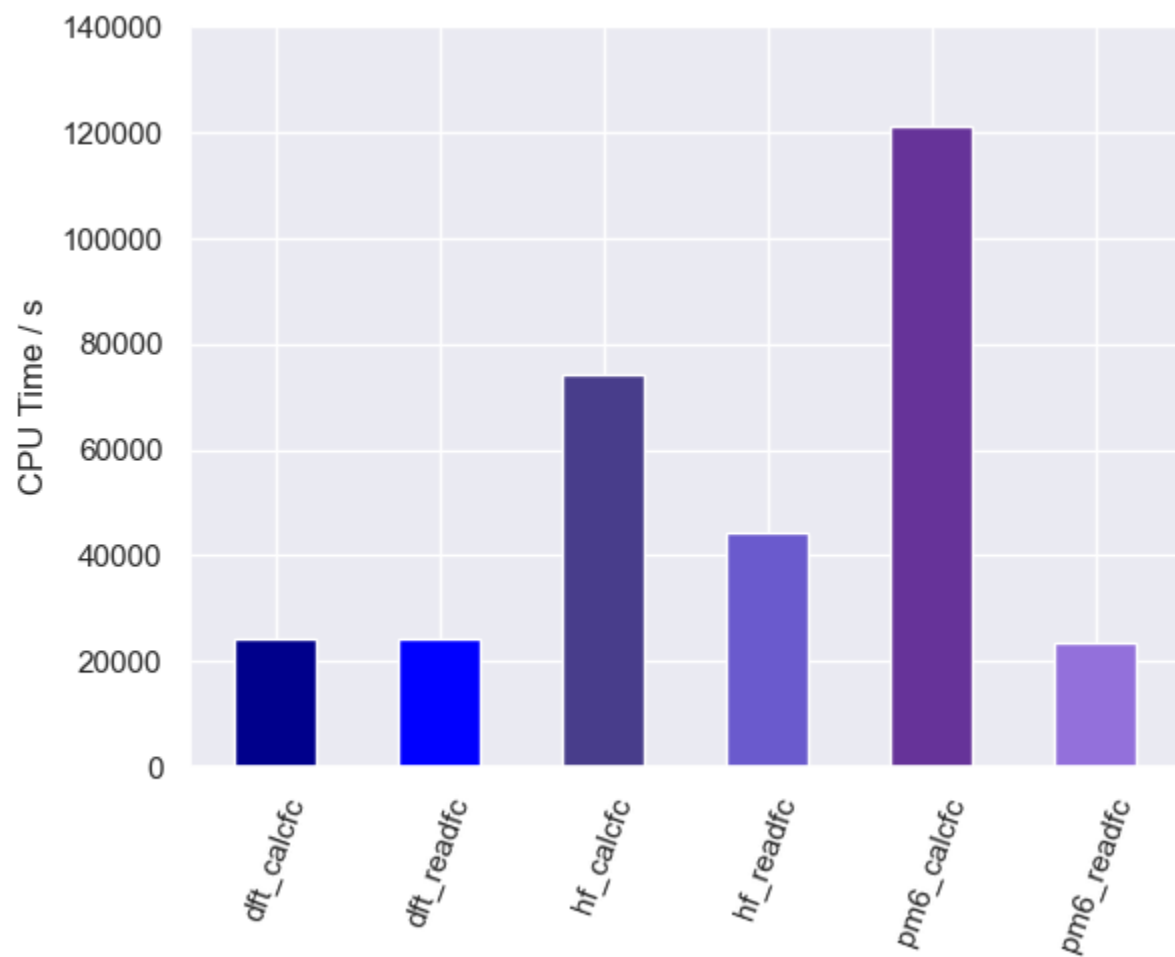
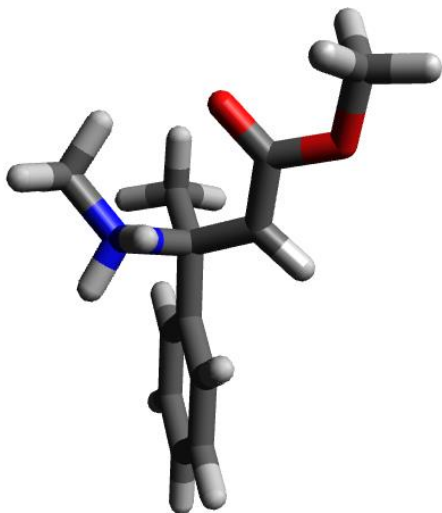# IMDA_2

IMDA_2: conformations

DFT Hessian
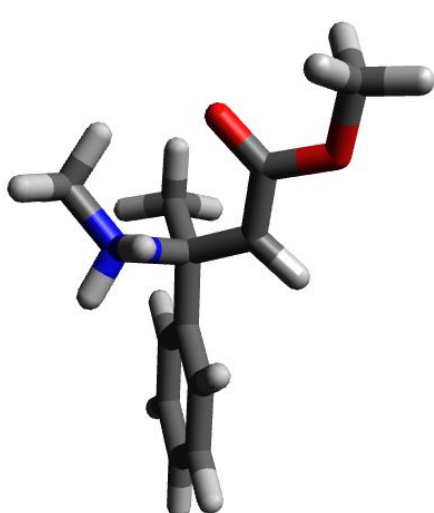
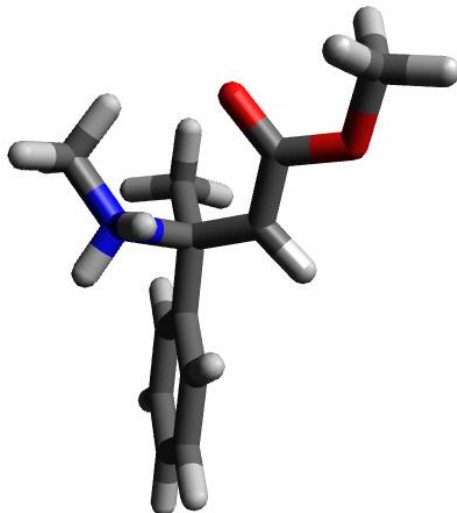HF Hessian

SQM Hessian

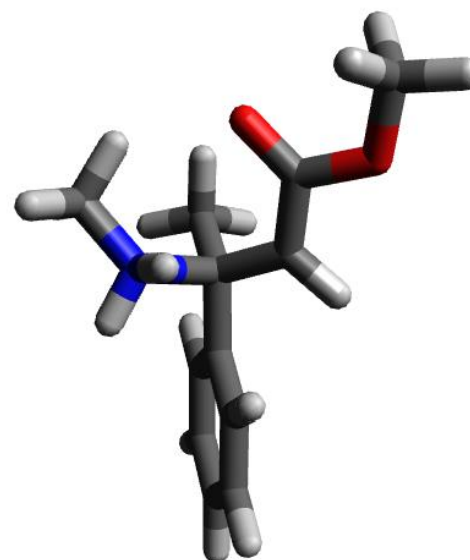# AM_92

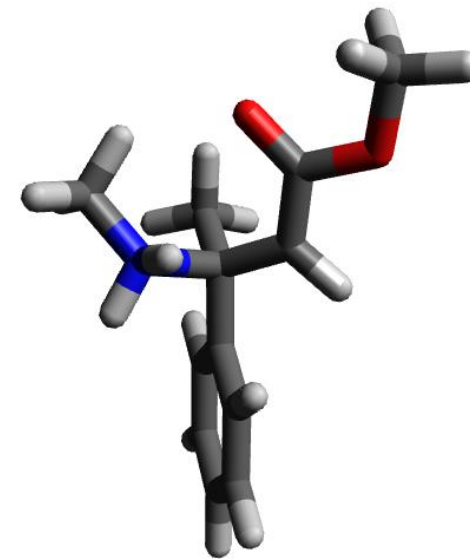# AM_92: conformations

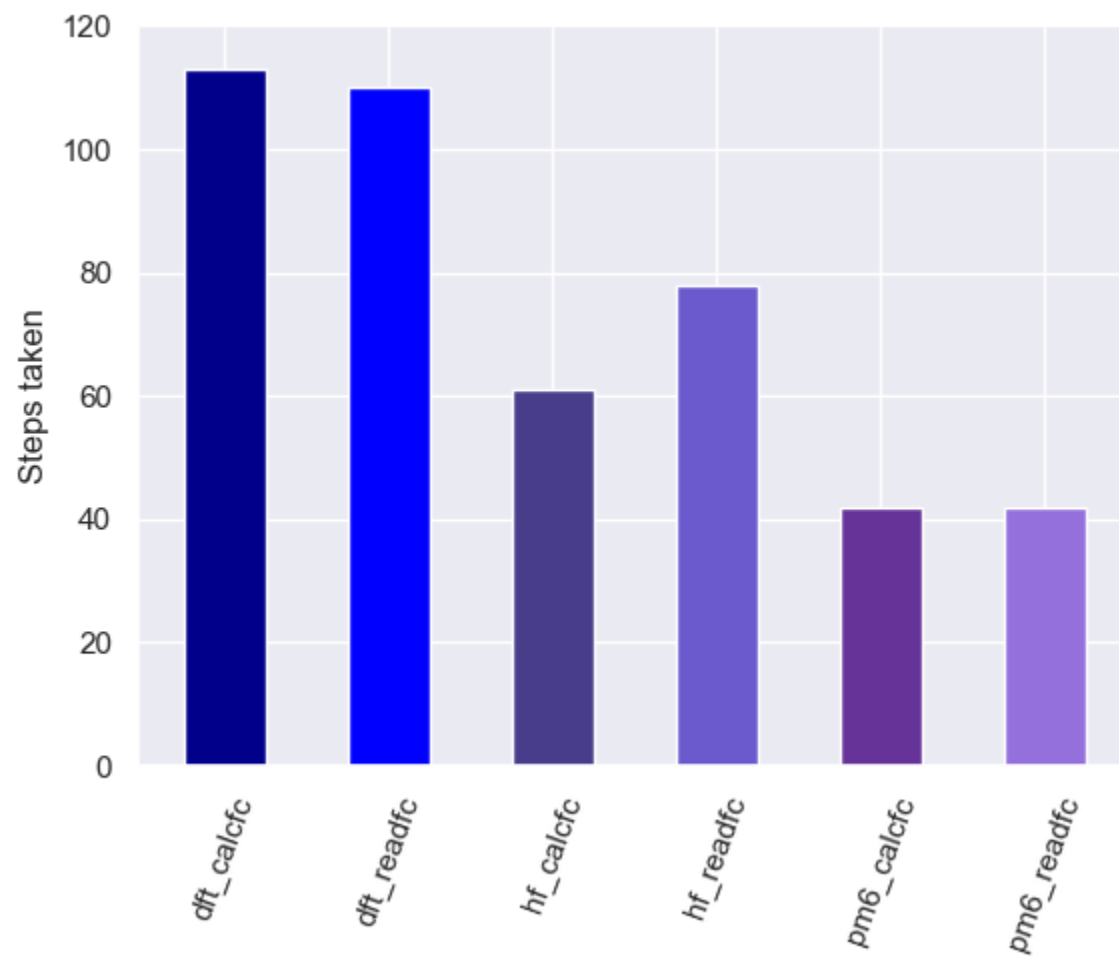DFT Hessian   HF calcfc   HF readfc   SQM calcfc   SQM readfc
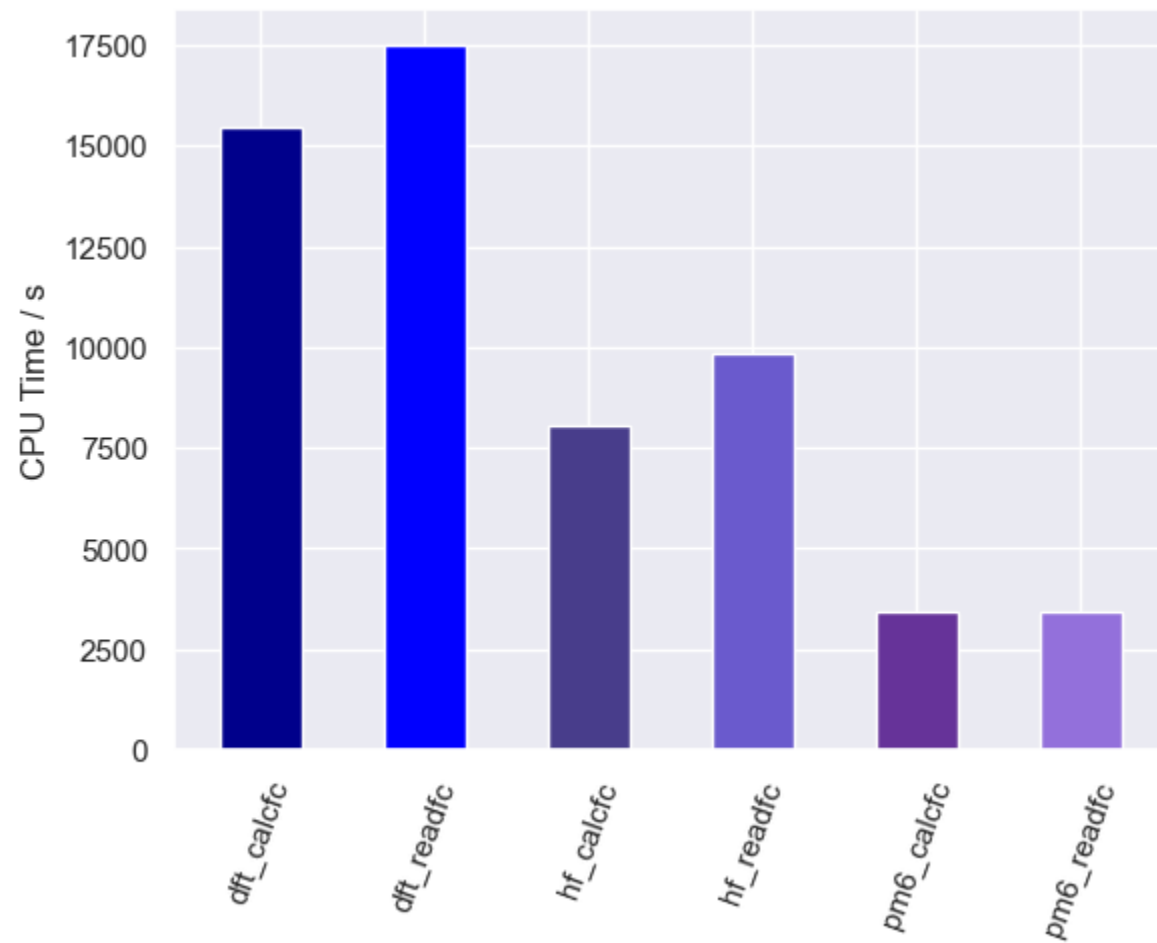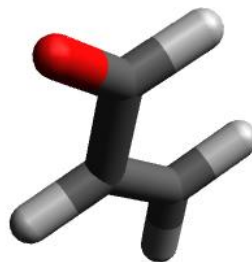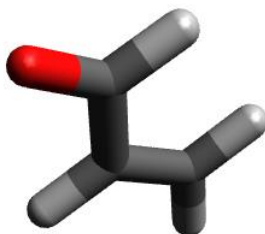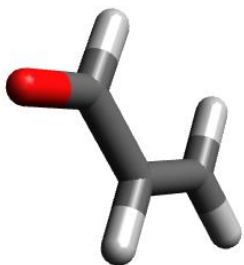
Incorrect TS:
All: N---C reacting bond rotation at around 77 1/cm

DA_1

DA_1: conformations

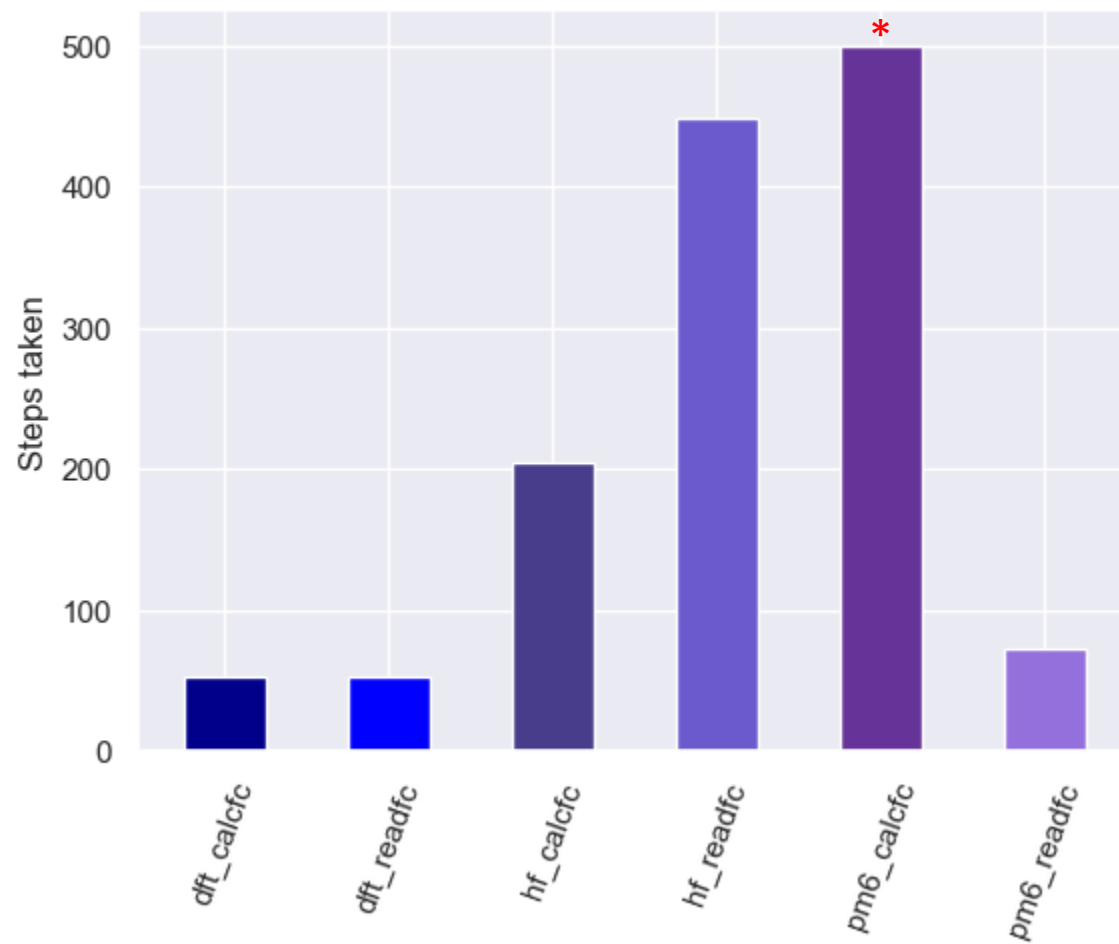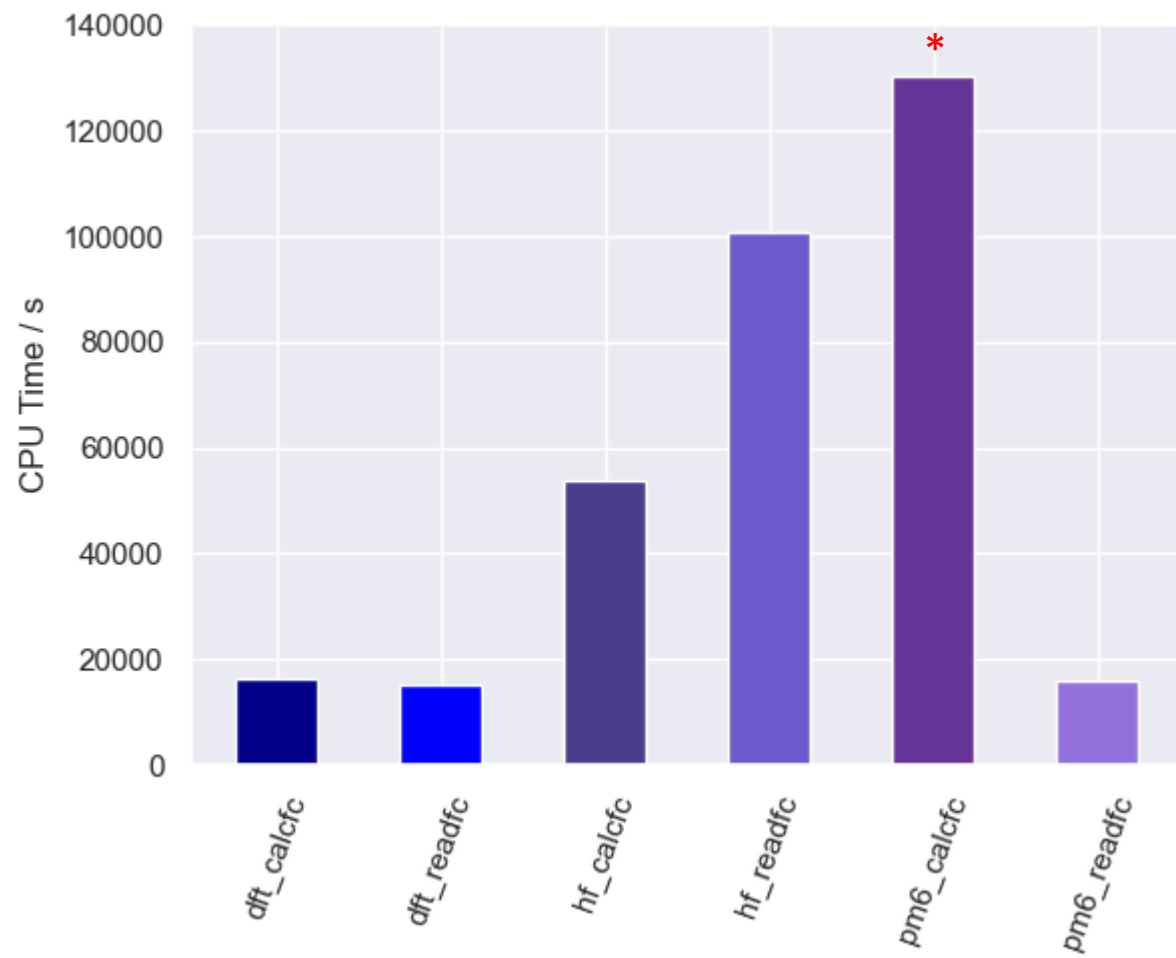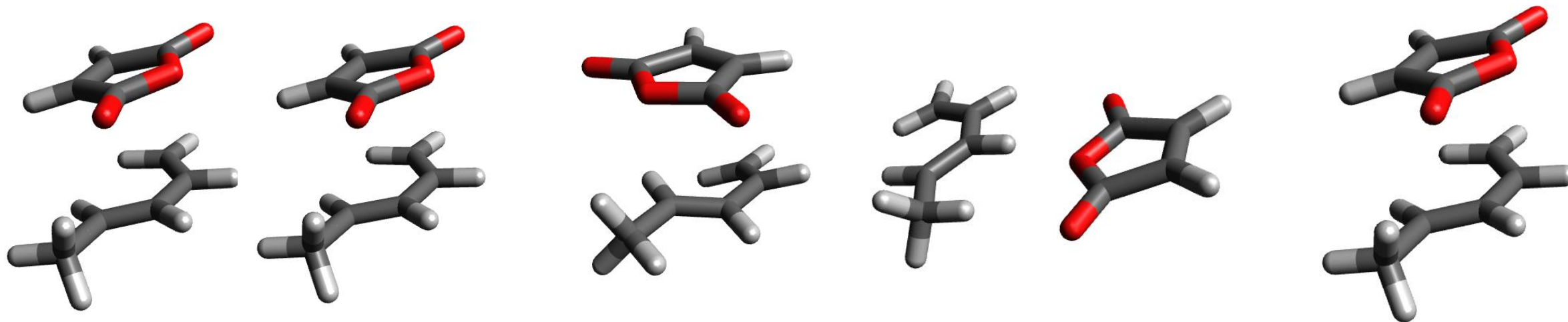| DFT calcfc | DFT readfc | HF calcfc | HF readfc | SQM Hessian |

Incorrect TS:
All: Symmetric rocking at around 20-30 1/cm

DA_2

DA_2: conformations

DFT calcfc   DFT readfc   HF calcfc   HF readfc   SQM readfc
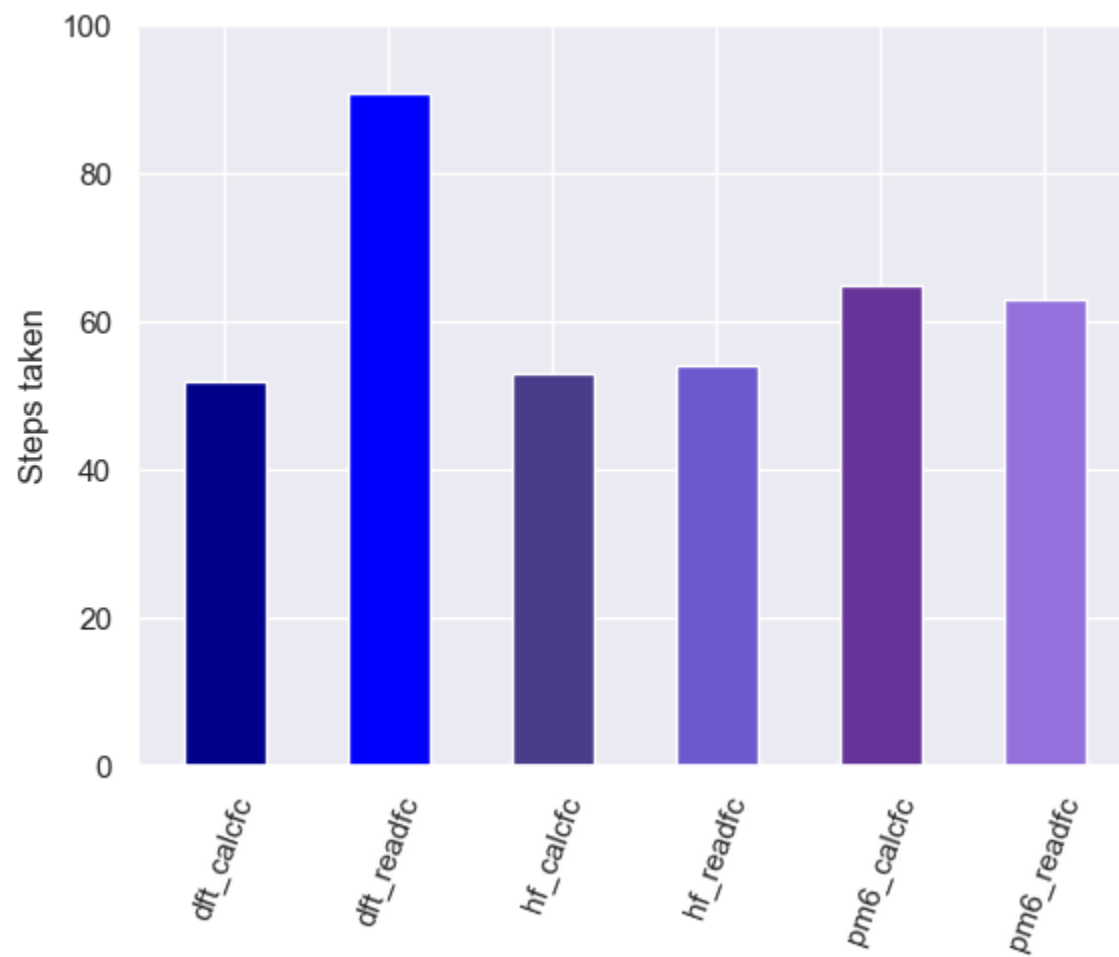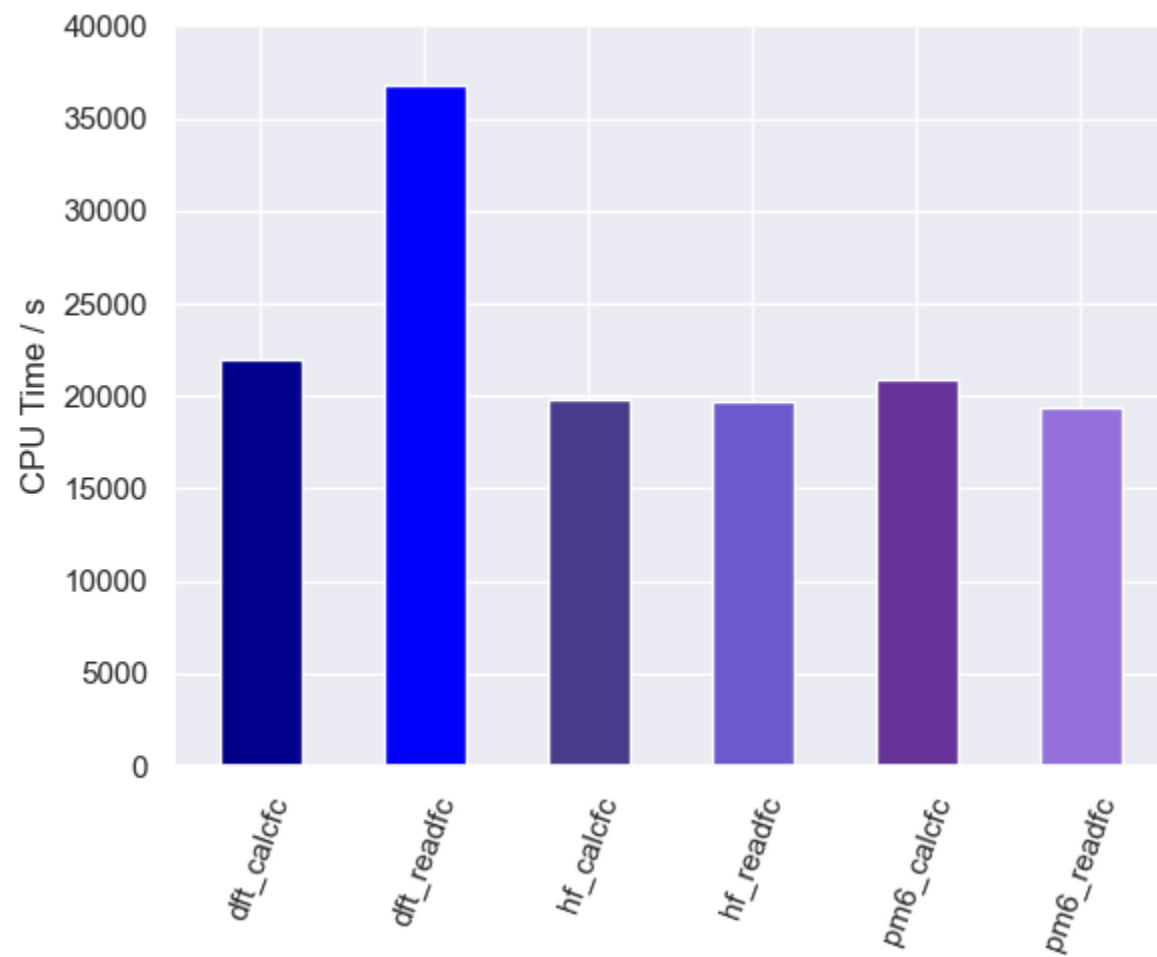
(SQM calcfc did not converge)

Incorrect TS:
DFT_calcfc, DFT_readfc and SQM_readfc: Anti-Houk to Houk RHC-Me rotation at around 180 /1cm
HF_calcfc, HF_readfc: Diene rocking at around 25-85 1/cm

IMDA_1

IMDA_1: conformations
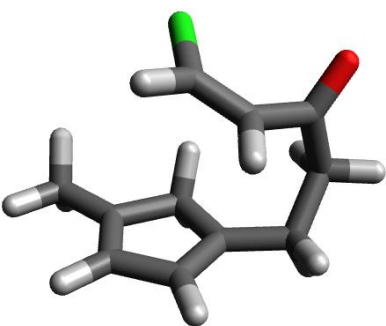
DFT calcfc   DFT readfc   HF calcfc   HF readfc   SQM calcfc   SQM readfc

Incorrect TS:
R(O)C-CH2(R) bond rotation at around 28 1/cm

IMDA_2

IMDA_2: conformations
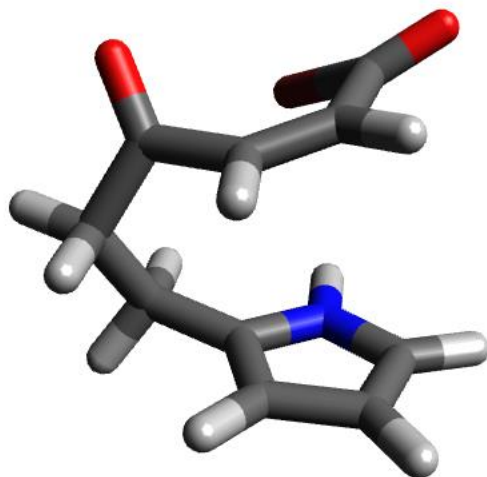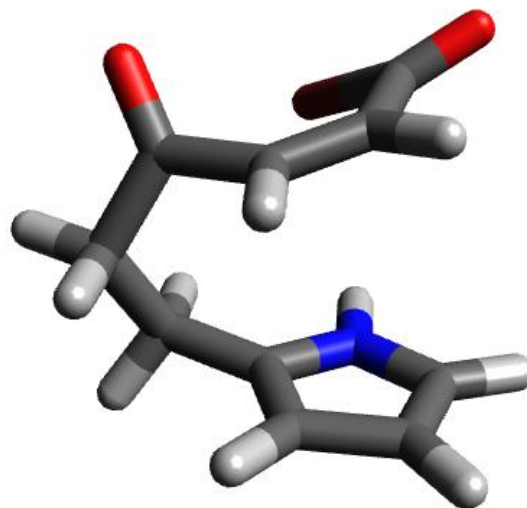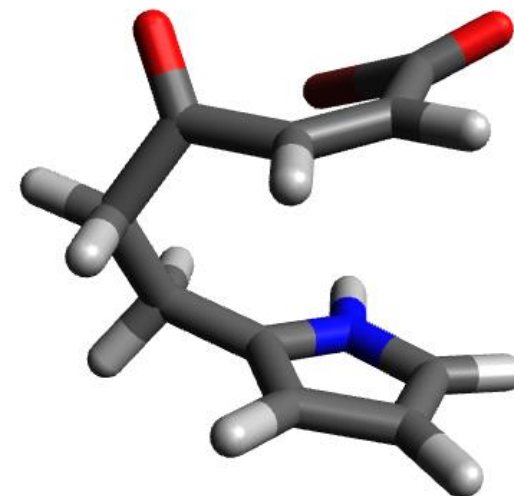
DFT Hessian   HF Hessian   SQM calcfc   SQM readfc

Incorrect TS:
Br(O)C-CH2(R) bond rotation at around 10 1/cm

cyclic
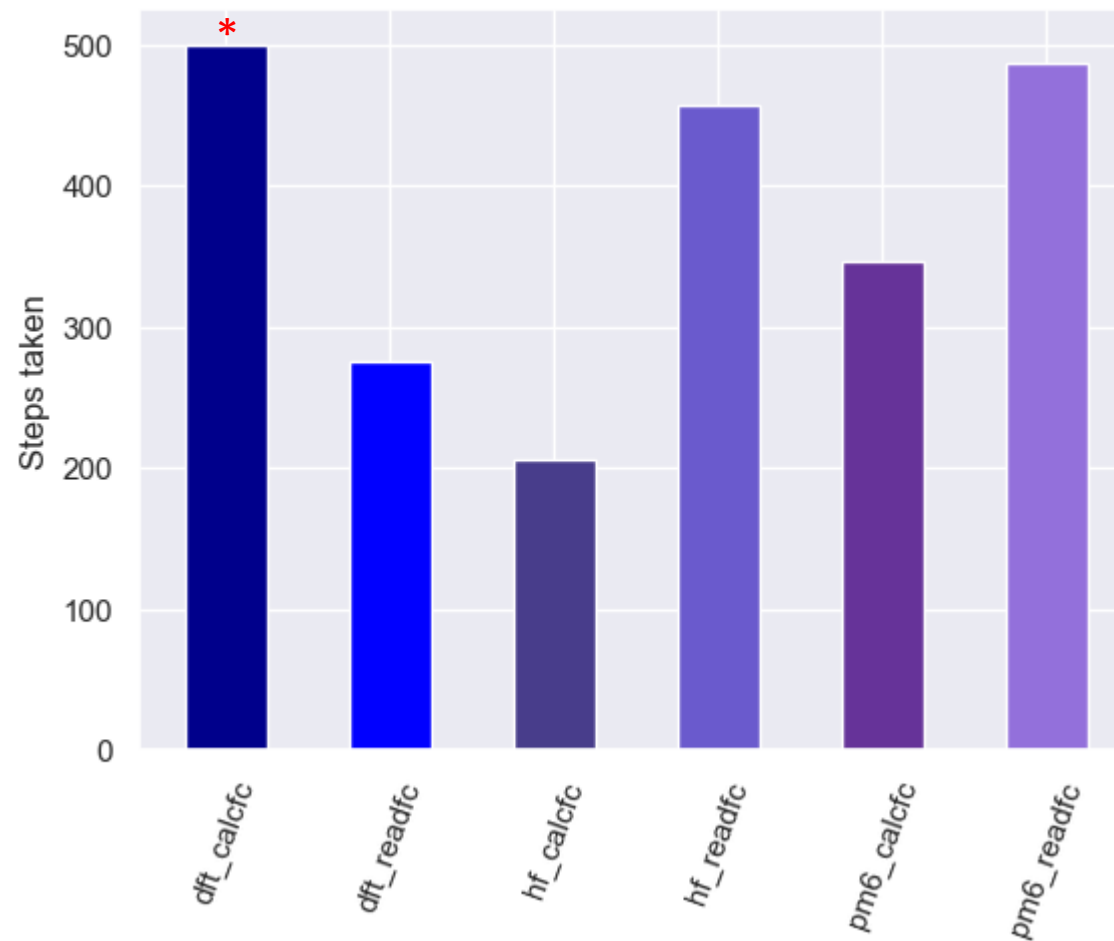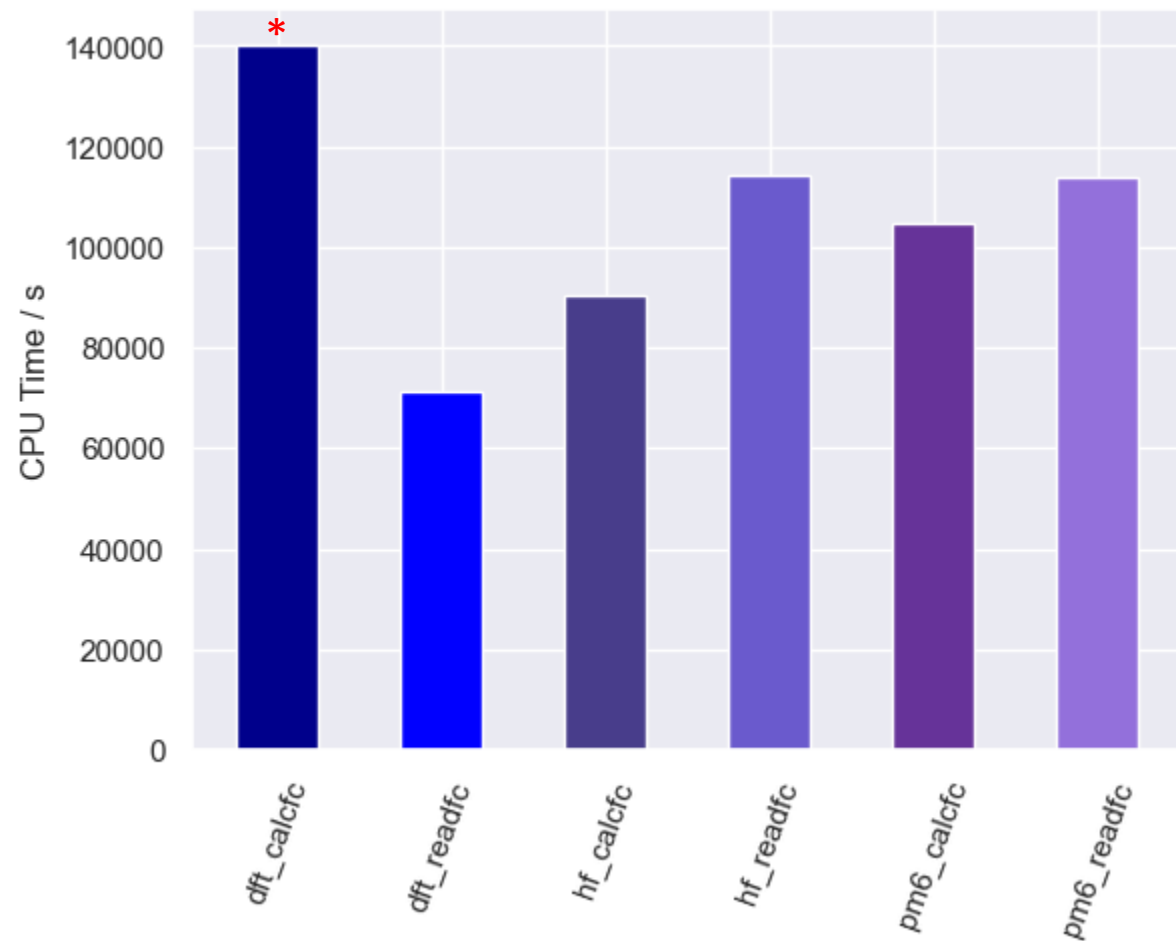
cyclic: conformations
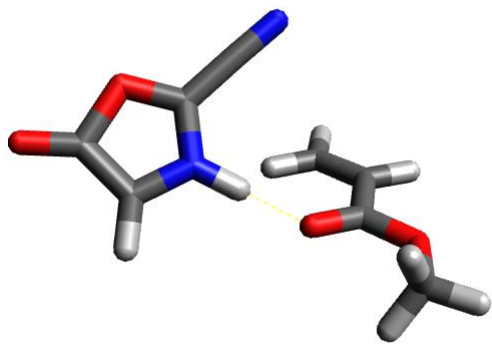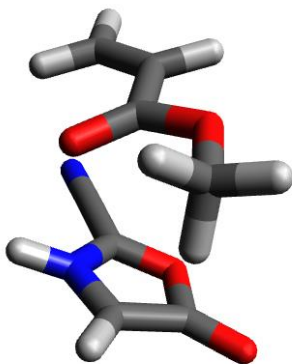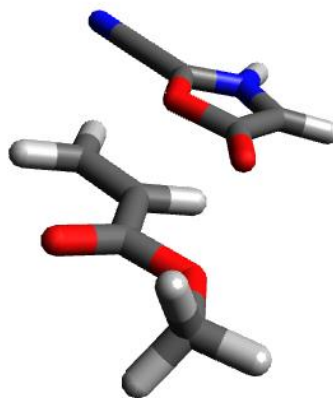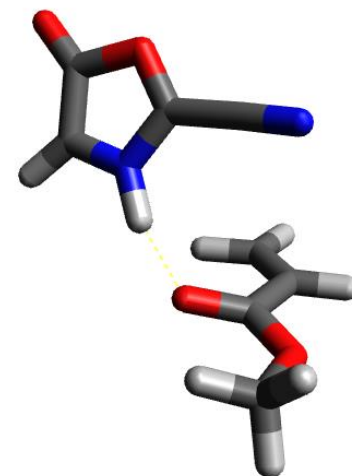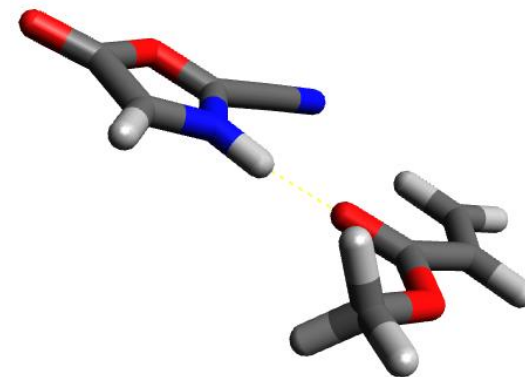
DFT readfc　　　HF calcfc　　　HF readfc　　　　　SQM calcfc　　　SQM readfc
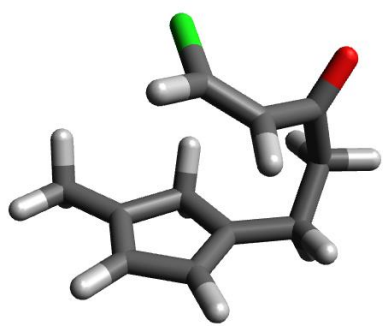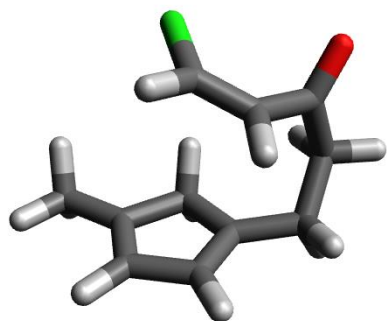
(DFT calcfc did not converge)

Incorrect TS:
DFT_readfc and SQM_readfc: (R)2NH--O=C(OMe)R H-bond rotation at around 10 1/cm
HF_calcfc: dipole rocking at around 42 1/cm
HF_readfc: not a TS (second order saddle point)
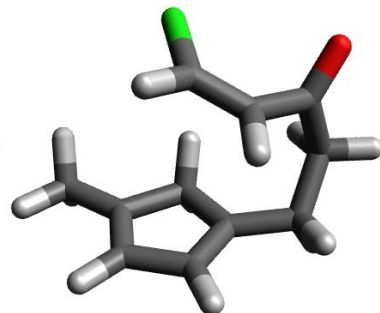SQM_calcfc: R(O)C-OMe bond rotation at around 116 1/cm
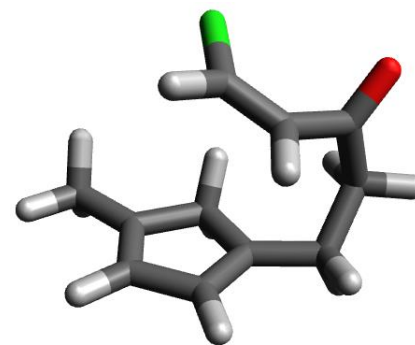
allylic
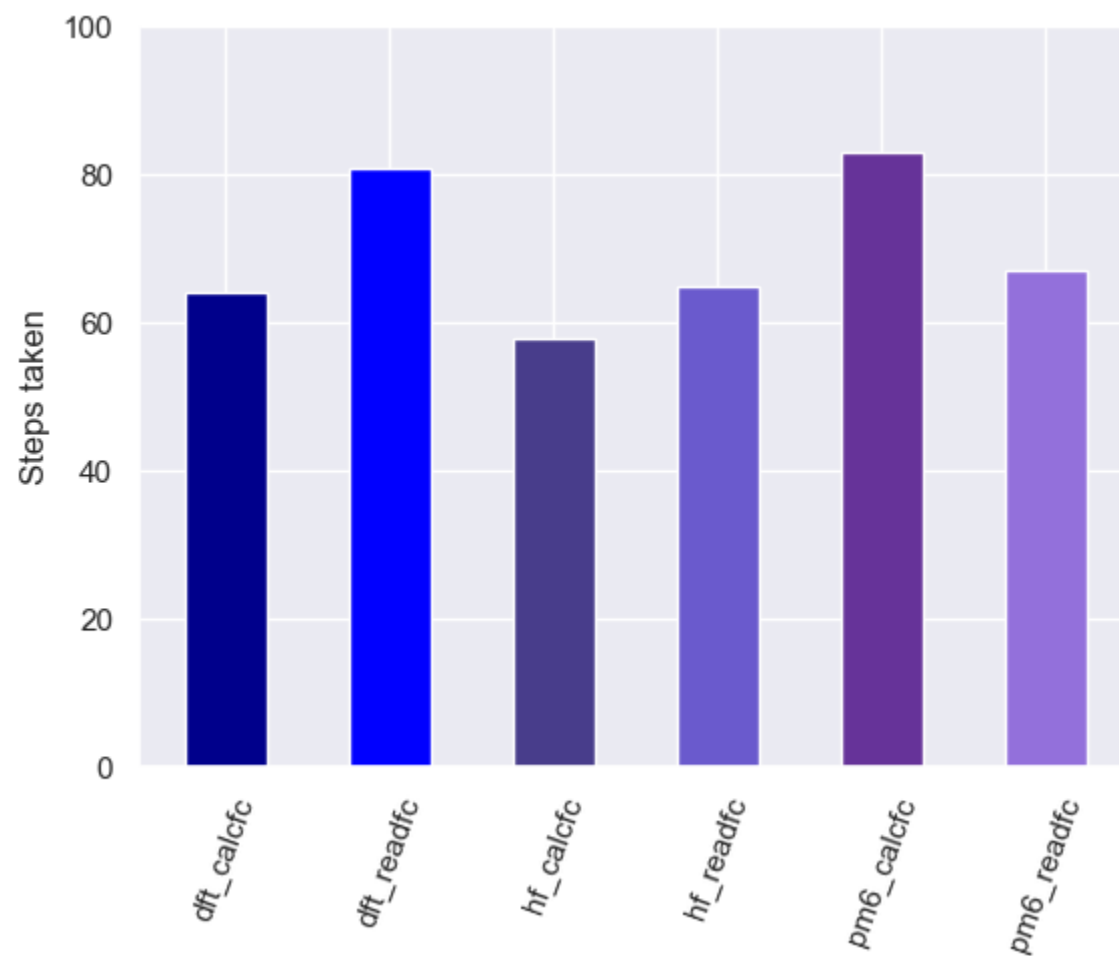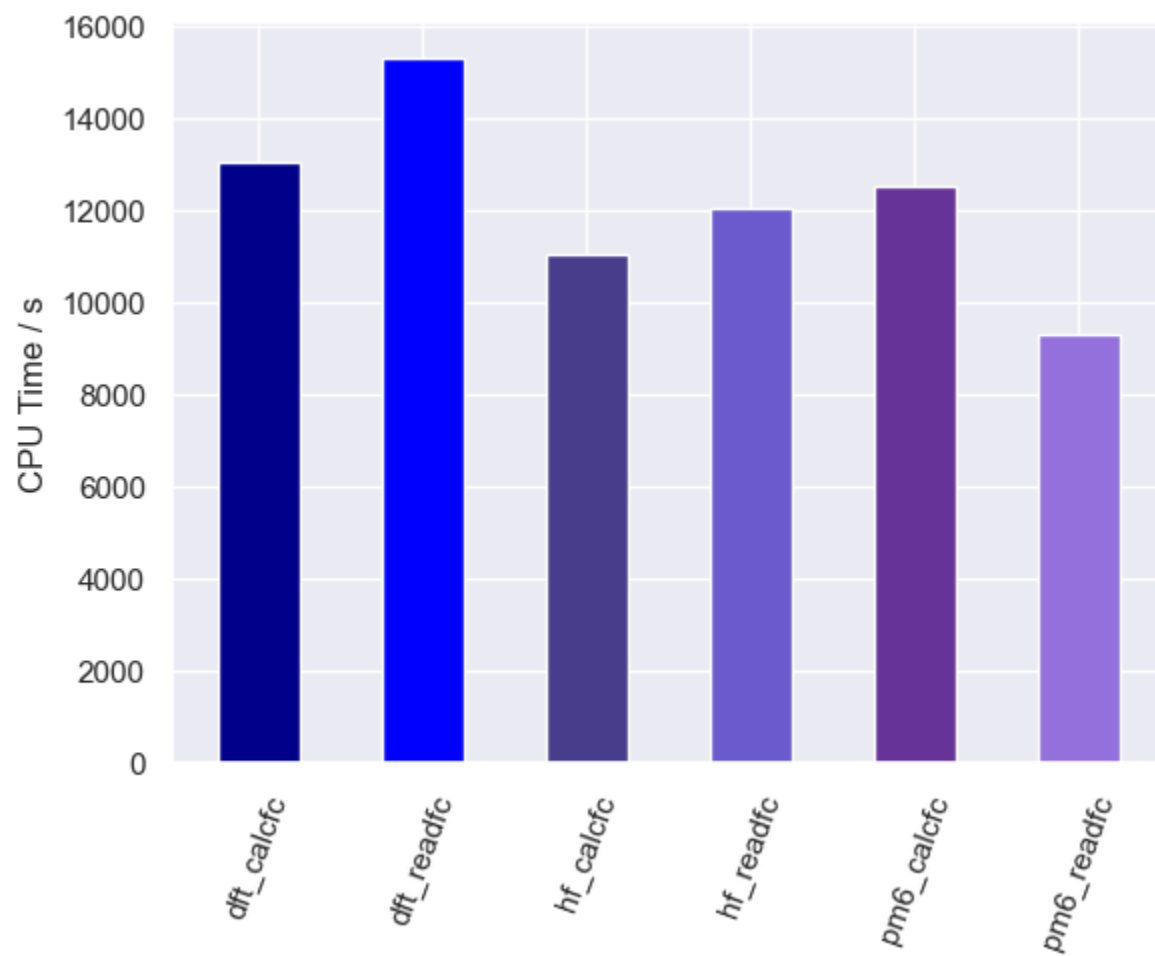
allylic: conformations
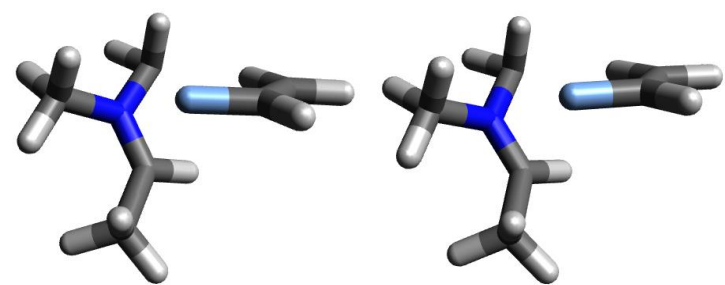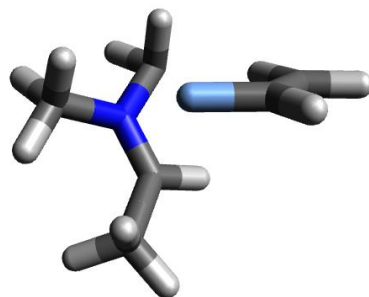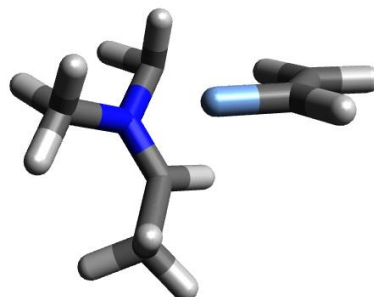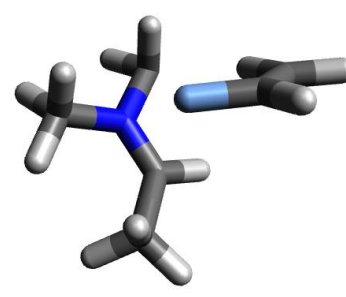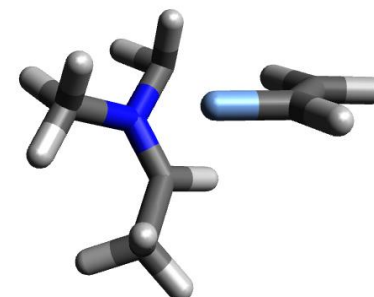
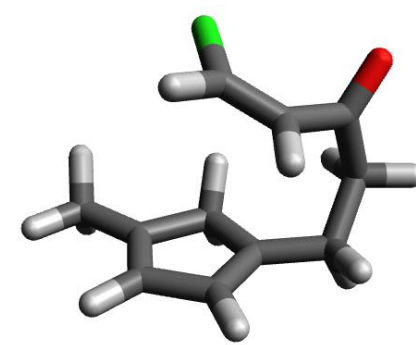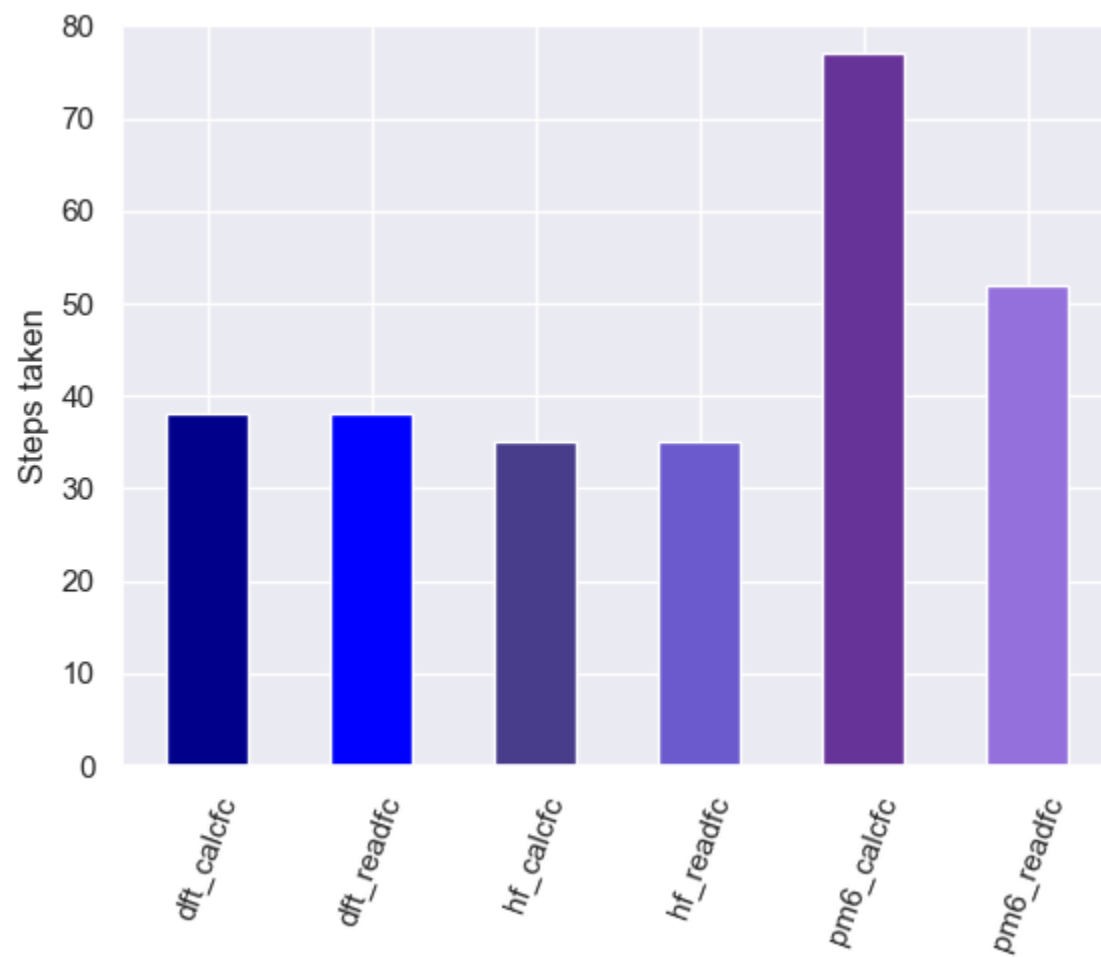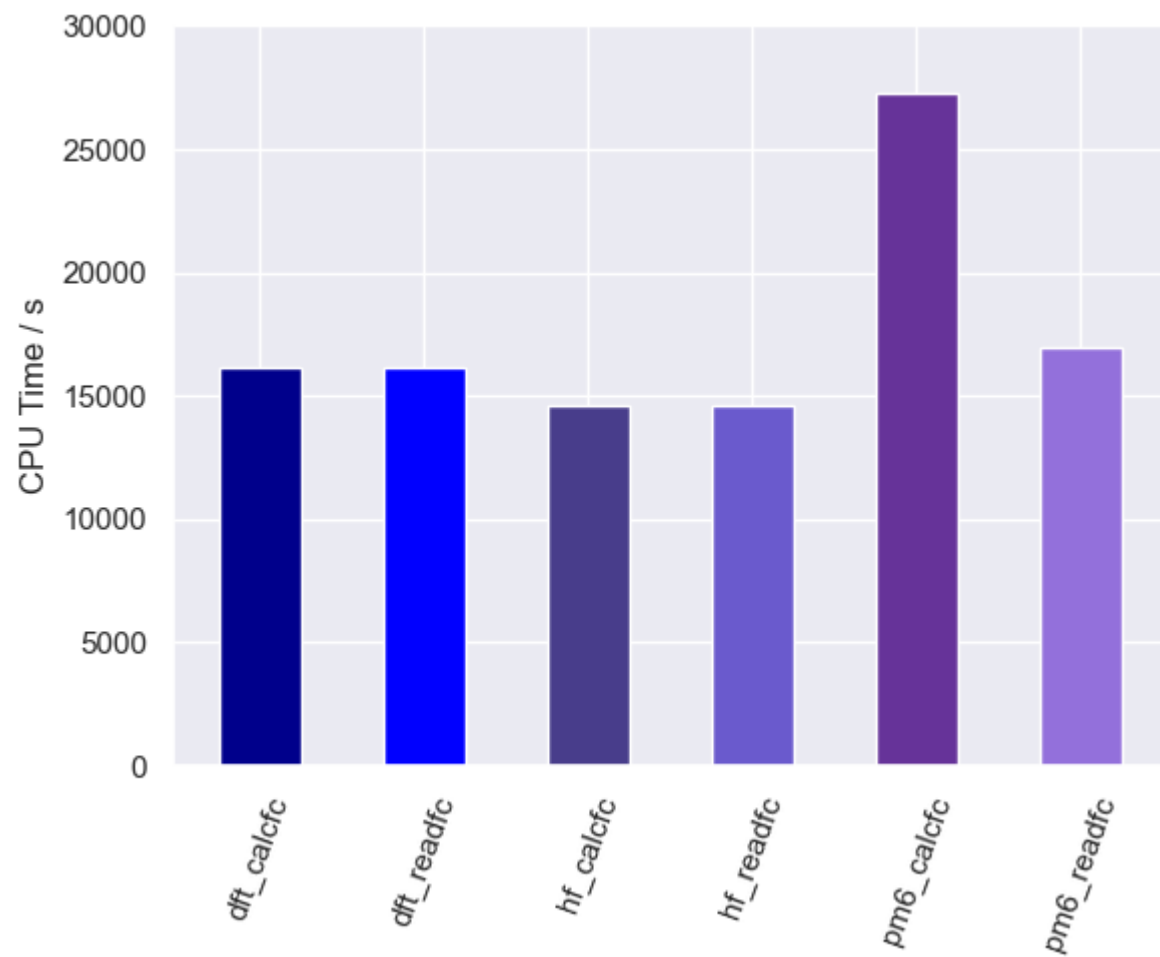DFT calcfc DFT readfc HF calcfc HF readfc SQM calcfc SQM readfc
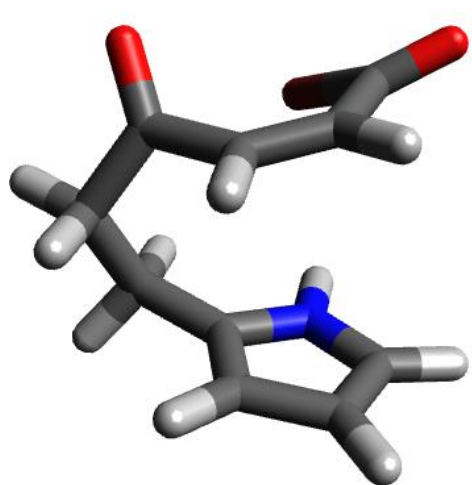
Incorrect TS:
All: (CH2)-N(CH3)R bond rotation at around 360 1/cm

Conclusions

- DFT_calcfc and DFT_readfc are usually equally fast (almost none required reoptimisation)

- HF_calcfc and HF_readfc are usually equally fast (almost none required reoptimisation)

- SQM_calcfc is usually similar to SQM_readfc, or slighltly slower

- HF initial Hessians were generally the fastest methods, (closely) followed by DFT Hessians and then SQM (PM6) Hessians. This could be because the HF Hessian balances accuracy with calculation speed better than the SQM Hessian, and therefore fewer optimisation steps are needed with a HF Hessian compared to an SQM Hessian

- Computational speedup is not very large compared to the traditional DFT_calcfc method (HF_calcfc is approximately 80% the CPU time of DFT_calcfc, on average)

Alternative project: Human-guided TS conformation ranking by eigenvector analysis

Lam, Y. H., Grayson, M. N., Holland, M. C., Simon, A., & Houk, K. N. (2016). Theory and modeling of asymmetric catalytic reactions. *Accounts of Chemical Research, 49*(4), 750-762.
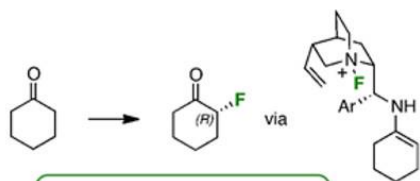
MECHANISTIC HYPOTHESIS
• enamine activation of ketone
• Lewis activation of NFSI ("F+")

(1) Hypothesis

(2) Conformational Search
(Force Field Calculations)

(3) Locate Transition Structures
(Density Functional Calculations)

$\Delta\Delta G^{\ddagger}$ (kcal/mol)

(R)-TS 0
(S)-TS 6.9

computed > 99% ee
experimental 99% ee

(R)-TS   (S)-TS

(4) Optimize Geometries, Compute Energies, Evaluate Hypothesis
(Density Functional Calculations)

Cα-Si face exposed

Cα-Re face exposed

(R)-TS   (S)-TS
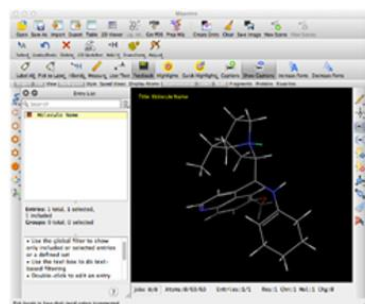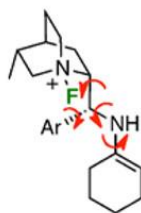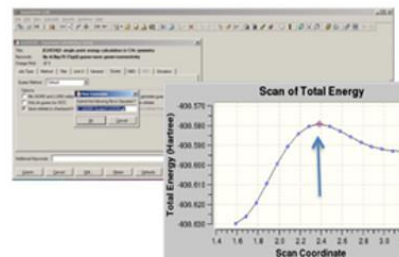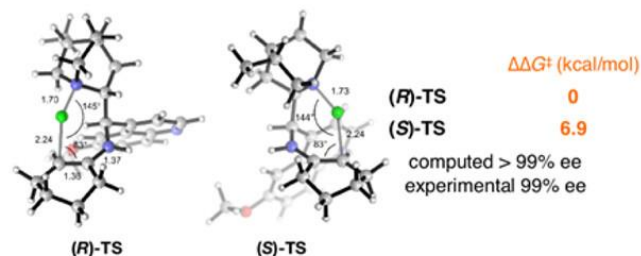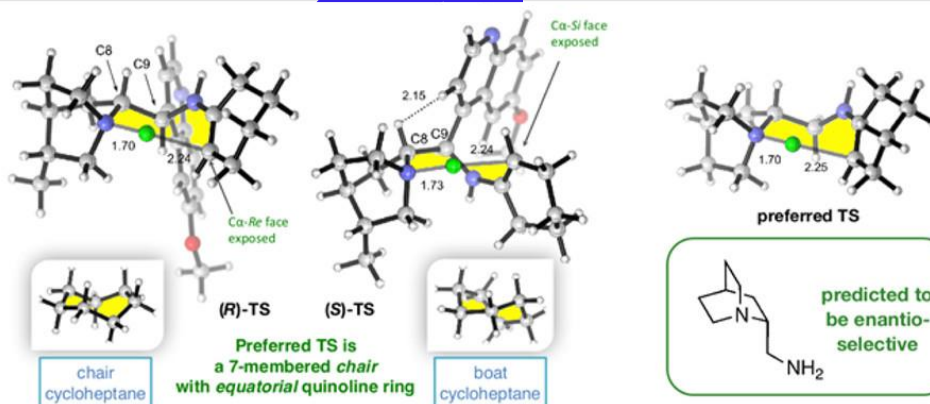
chair cycloheptane

boat cycloheptane

Preferred TS is a 7-membered *chair* with *equatorial* quinoline ring

(5) Acquire Insights, Develop Models

preferred TS

predicted to be enantio-selective

(6) Make New Predictions

(2.5) Conformational ranking/filtering (Density Functional Frequency Calculations)

- Compute Frequencies with `freq` keyword (Hessian eigenvalues and eigenvectors)
- Human selects internal coordinates they expect to contribute most to the final TS eigenvector (e.g. C—N bond in Aza-Michael)
- Automatically rank TS conformations by contribution of this internal coordinate to the eigenvector whose eigenvalue is the most negative (this is the eigenvector used in the eigenvector-following algorithm)
- These Hessians can then be used at the start of TS searching (they are required in any case – therefore the only computational cost is in the eigenvector calculations)

C_1 C_2

C_4

C_3

freq

$$\begin{bmatrix} -\lambda_1 \\ -\lambda_2 \\ -\lambda_3 \\ 0 \\ 0 \\ 0 \\ \dots \end{bmatrix}$$

$-\lambda_1$ evect

$$\begin{bmatrix} C_1\text{-}C_3 \\ C_2\text{-}C_4 \\ Q_i \\ Q_j \\ \dots \end{bmatrix}$$

FF TS conformation