

Virtual Chemist Suite 2019

QUEMIST, CONSTRUCTS, and ACE

Tutorial

Department of Chemistry
McGill University
Montréal, Québec, Canada



Molecular Forecaster Inc.
Montréal, Québec, Canada



24/08/2019

Table of Contents

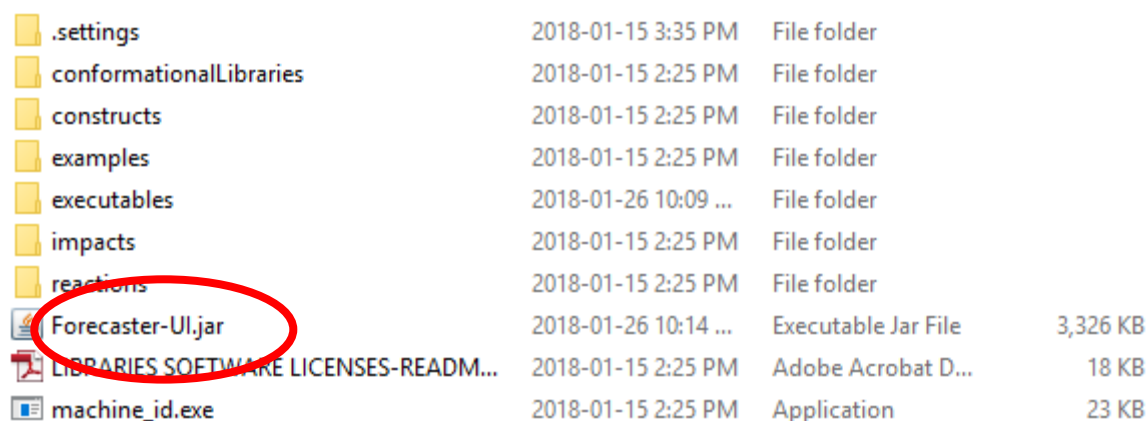
I.	Running Forecaster with the User Interface	3
II.	Quemist	3
1.	Basics	4
2.	Platform.....	5
3.	Basis sets.....	5
III.	Analyzing Quemist results.....	7
1.	Single point energy	7
2.	Geometry optimization.....	9
3.	Obtaining force field parameters	11
IV.	Quemist and Reactants and Products.....	14
V.	Constructs.....	18
1.	Diethyl zinc addition to aldehydes	18
2.	Asymmetric Dihydroxylation.....	33
VI.	References.....	43

I. RUNNING FORECASTER WITH THE USER INTERFACE

The user interface (UI) can be started by double clicking on the `Forecaster-UI.jar` file in the `Forecaster` folder. This will open the main window.

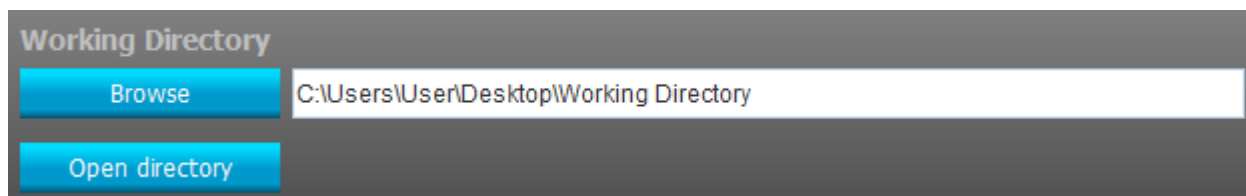
Under Linux and Mac OSX, it is recommended to launch it from a terminal window by typing the command below. Make sure that you are located in the folder where this jar file is.

```
Forecaster@Linux/Forecaster:~$ java -jar Forecaster-UI.jar
```



.settings	2018-01-15 3:35 PM	File folder	
conformationalLibraries	2018-01-15 2:25 PM	File folder	
constructs	2018-01-15 2:25 PM	File folder	
examples	2018-01-15 2:25 PM	File folder	
executables	2018-01-26 10:09 ...	File folder	
impacts	2018-01-15 2:25 PM	File folder	
reactions	2018-01-15 2:25 PM	File folder	
Forecaster-UI.jar	2018-01-26 10:14 ...	Executable Jar File	3,326 KB
LIBRARIES SOFTWARE LICENSES-READM...	2018-01-15 2:25 PM	Adobe Acrobat D...	18 KB
machine_id.exe	2018-01-15 2:25 PM	Application	23 KB

The first step is to set the working directory. Click **Browse**, under **Working Directory**. You will be prompted to navigate to the folder where you will save various files while working with FORECASTER. If this is the first time you use FORECASTER, you may need to create a new folder. This folder can have any name and can be located anywhere you want..



Working Directory

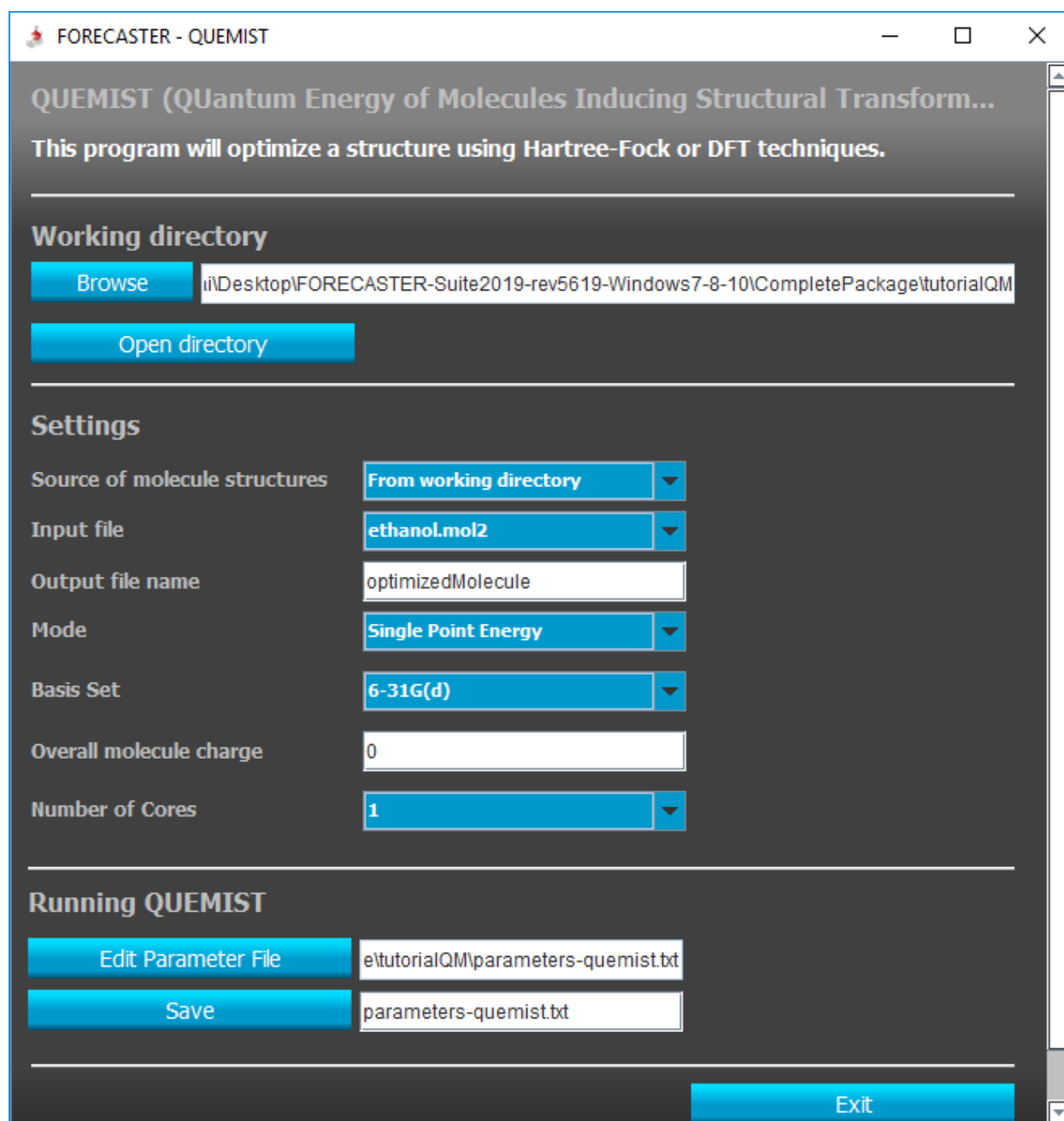
Browse C:\Users\User\Desktop\Working Directory

Open directory

You may then click **Start Forecasting** to expand the “*Modular Workflow*” section.

II. QUEMIST

As soon as the working directory is set, copy the file of your small molecule of interest in this folder. Make sure this file is in mol2 format. Then make a modular workflow that includes QUEMIST. As soon as a modular workflow is ready (see corresponding tutorial), open QUEMIST.



1. Basics

QUEMIST can run HF (Windows, Linux, MacOS) and DFT (Linux, MacOS) calculations, including single point energy calculations and geometry optimizations (in Cartesian coordinates). Moreover, if one wishes to parametrize their own reaction in ACE, QUEMIST can run geometry optimizations of reactants/products necessary for TS template building followed by computation of force field parameters using the Seminario method¹ later improved by Allen *et. al.*²

Some important information about QUEMIST:

- Running it on Linux/MacOS will always be significantly faster than running it on Windows because of the usage of the libint library for the computation of two-electron integrals. From our experience, we recommend running only very small molecules (<

40 atoms) with small basis sets (STO-3G, minix, 6-31G) on Windows. If you have Windows 10, you can download the Linux subsystem for Windows 10 and can run the Linux version of QUEMIST on this subsystem. We highly recommend using the Linux version!

- For now, DFT calculations are only possible on Linux/macOS. We use the libxc library for functionals which is only available for Linux systems.
- Geometry optimizations/force field parameter generation is only available on Linux.

Options:

After significant testing we have come up with defaults for various options available within QUEMIST. These are encoded into the platform and will be automatically written to the parameter file. **DO NOT CHANGE THESE DEFAULTS** (unless you know the potential impact of these changes).

2. Platform

You can run single point energies, geometry optimizations and geometry optimizations that involve computing force field parameters (WARNING: this last option involves computing the Hessian of your molecule, which will take a very long time (ie, days!)). For the first part of the tutorial, we will run a single point energy calculation on ethanol, both on Windows and Linux. We will use the STO-3G basis set and will be running the calculation on 4 cores. First, change the name of the output file name to ethanol-QM-Windows.

Settings

Source of molecule structures	From working directory
Input file	ethanol.mol2
Output file name	ethanol-QM-Windows
Mode	Single Point Energy
Basis Set	Single Point Energy Geometry Optimization Geometry Optimization and FF parameters
Overall molecule charge	0
Number of Cores	1

The QUEMIST code is parallelized so you can select as many cores as you can spare for your calculations.

Number of Cores

4

3. Basis sets

Below are listed the available basis sets:

- 6-31G (Elements w/ atomic numbers 1-30, 34-35 and 53 taken from 6-311G)

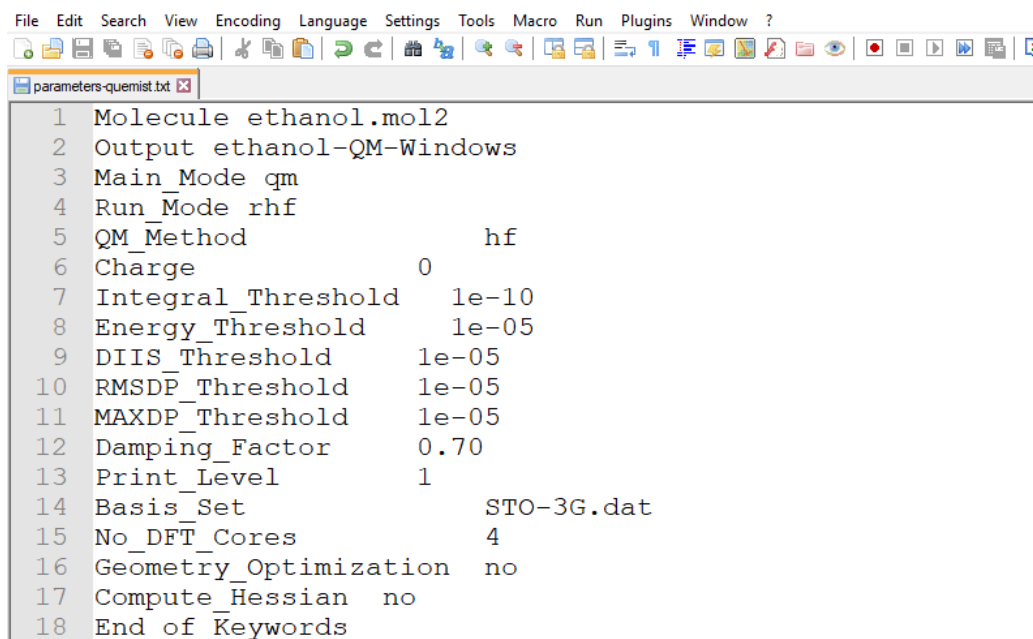
- 6-31G(d) = 6-31G* (Elements w/ atomic numbers 1-30, 34-35 and 53 taken from DZVP)
- 6-31G(d,p) = 6-31G** (Elements w/ atomic numbers 1-30, 34-35 from 6-31G* and 53 taken from DZVP)
- 6-311G (Elements w/ atomic numbers 1-20, 34-35 and 53)
- Augmented-dzp (Elements: H (1), B (5), C (6), N(7), O(8), F(9), Al(13), Si(14), P(15), S(16), Cl(17), Fe(26), Zn(30), Se(34), Br(35), Rh(45), I(53), Os(76), Ir(77))
- DZP (Elements: H (1), B (5), C (6), N(7), O(8), F(9), Al(13), Si(14), P(15), S(16), Cl(17), Fe(26), Zn(30), Se(34), Br(35), I(53), Os(76))
- Def2-SVP (def2-SVP basis set for elements 1-36. Iodine, Rhodium and Osmium taken from augmented-DZP)
- Def2-TZVP (def2-TZVP basis set for elements 1-36. Iodine, Rhodium and Osmium taken from augmented-DZP)
- STO-3G (Elements w/ atomic numbers 1-53 + 76 from WTBS)
- Minao (Elements w/ atomic numbers 1-36 + 45, 53 and 76 from ANO-RCC, 77 from WTBS)
- Minix :

H-He, B-Ne	MINIS
Li-Be	MINIS + 1(p)
Na-Mg	MINIS + 1(p)
Al-Ar	MINIS + 1(d)
K-Zn	SV
Ga-Kr	SVP
I, Rh, Os and Ir	taken from augmented-DZP
- Forecaster-BS:

Elements:	1-36, 45, 53, 76 and 77
H - Al	def2-SVP
Si - Cl	def2-TZVPD
Ar - Cu	def2-SVP
Zn	def2-TZVPD
Ga - Kr	def2-SVP
Rh, I, Os, Ir	augmented-DZP

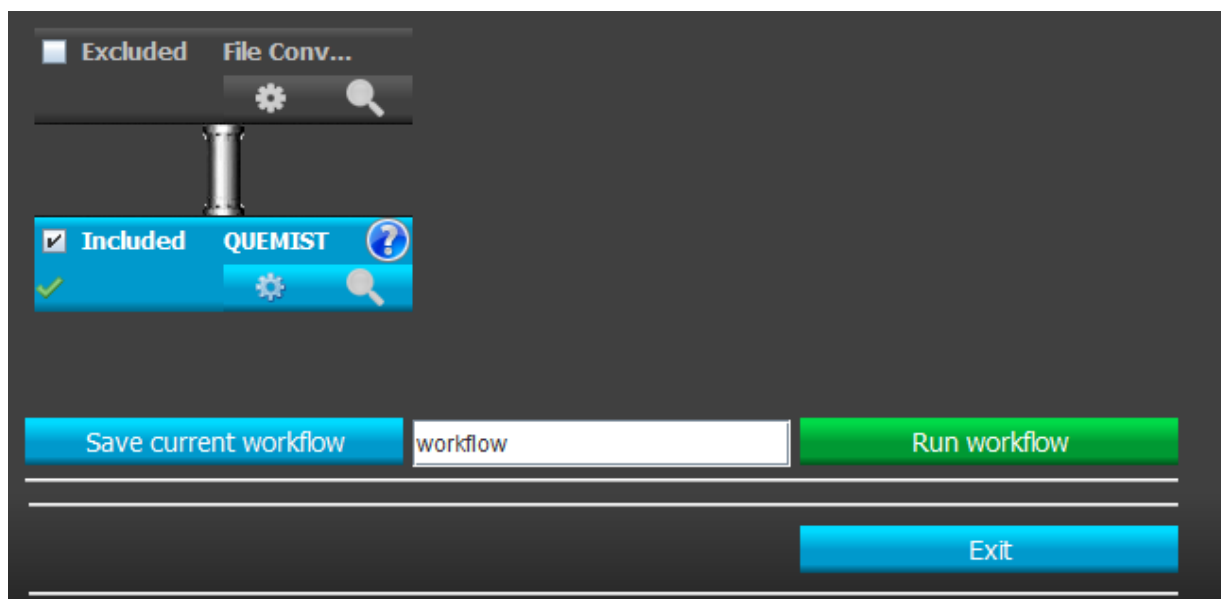
Basis Set	STO-3G
Overall molecule charge	6-31G(d)
Number of Cores	6-311G
	STO-3G
	6-31G
	6-311G
	dzp
	def2-SVP
Running QUEMIST	x2c-SVP

Once the selections is made, save the parameter file. Below is the parameter file generated.



```
1 Molecule ethanol.mol2
2 Output ethanol-QM-Windows
3 Main_Mode qm
4 Run_Mode rhf
5 QM_Method hf
6 Charge 0
7 Integral_Threshold 1e-10
8 Energy_Threshold 1e-05
9 DIIS_Threshold 1e-05
10 RMSDP_Threshold 1e-05
11 MAXDP_Threshold 1e-05
12 Damping_Factor 0.70
13 Print_Level 1
14 Basis_Set STO-3G.dat
15 No_DFT_Cores 4
16 Geometry_Optimization no
17 Compute_Hessian no
18 End of Keywords
```

At this stage, you are ready to Run the workflow.



The calculation should not take more than a few minutes.

III. ANALYZING QUEMIST RESULTS.

1. Single point energy

Open the file ethanol-QM-Windows.out using your favorite text editor. Here are a few things we are interested in:

SCF ITERATIONS					
It. #	Energy (Eh)	deltaEn (Eh)	RMS (D, P)	[F, P]	MAX-DP
0	-151.486373924	0.000000	0.030397	0.230774	0.209471
1	-151.535097379	-0.048723	0.025150	0.184925	0.172083
2	-151.572420964	-0.037324	0.043483	0.147579	0.249489
3	-151.700402562	-0.127982	0.003610	0.007969	0.023590
4	-151.700988245	-0.000586	0.000782	0.001969	0.006618
5	-151.701013676	-0.000025	0.000207	0.000598	0.001886
6	-151.701015506	-0.000002	0.000067	0.000172	0.000600
7	-151.701015680	-0.000000	0.000019	0.000041	0.000128
8	-151.701015694	-0.000000	0.000006	0.000013	0.000038
9	-151.701015695	-0.000000	0.000003	0.000006	0.000024
SCF CONVERGED AFTER 10 ITERATIONS					
FINAL SINGLE POINT ENERGY					
-151.701015695304					
Last energy change = 1.53830e-009 Tolerance = 1.00000e-005 Last RMS change = 3.23117e-006 Tolerance = 1.00000e-005 Last MAX-DP change = 2.38863e-005 Tolerance = 1.00000e-005 Last DIIS error = 6.19049e-006 Tolerance = 1.00000e-005					
Ekinetic = 1.523185e+002 Eh Enuclattr = -5.264914e+002 Eh Enuclear = 8.449362e+001 Eh Eone = -3.741728e+002 Eh Ecoul = 1.590333e+002 Eh Eel = -2.361946e+002 Eh E(XC) = -2.105507e+001 Eh Etotat = -1.517010e+002 Eh Virial factor= 9.95946e-001					

The calculation converged after 10 iterations. Most important, the virial factor is 0.99 (~1.00), which is what you would expect from a converged calculation.

Now, if you have the Linux version, run the same parameter file (make sure to change the output file name to ethanol-QM-Linux) using the Linux version. Below find the results you should obtain:

Total time elapsed : 0.0015 minutes.					
4	-151.701015094	-0.000036	0.000078	0.000344	0.000620
Total time elapsed : 0.0019 minutes.					
5	-151.701016516	-0.000001	0.000027	0.000060	0.000178
Total time elapsed : 0.0022 minutes.					
6	-151.701016628	-0.000000	0.000006	0.000013	0.000035
Total time elapsed : 0.0026 minutes.					
7	-151.701016635	-0.000000	0.000004	0.000007	0.000027
Total time elapsed : 0.0030 minutes.					
SCF CONVERGED AFTER 8 ITERATIONS					
FINAL SINGLE POINT ENERGY					
-151.701016635474					
Dispersion correction = -0.162855473696 Eh					
Total energy (including corrections) = -151.863872109170 Eh					
Last energy change	= 7.30179e-09	Tolerance	= 1.00000e-05		
Last RMS change	= 3.83300e-06	Tolerance	= 1.00000e-05		
Last MAX-DP change	= 2.66312e-05	Tolerance	= 1.00000e-05		
Last DIIS error	= 6.54883e-06	Tolerance	= 1.00000e-05		

As you can see, the results files are different. First, using libint (a library used in QUEMIST when running on Linux/MacOS) reduces the number of iterations from 10 to 8. Moreover, for the Linux version, there is no detailed energy breakdown (including virial factor computation) but a special dispersion correction is included. However, as you can see, the energy difference between the two computations is as low as -0.00000094017 Hartrees, approx. -0.000589 kcal/mol.

2. Geometry optimization

WARNING: Geometry optimizations are only available on Linux.

In the platform, select "Geometry optimization" instead of "Single point energy". Change the name of the output file to ethanol-QM-Linux-GeomOpt. The parameter file should look like this:

```

File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
parameters-queimist.bt ethanol-QM-Linux.out
1 Molecule ethanol.mol2
2 Output ethanol-QM-Linux-GeomOpt
3 Main_Mode qm
4 Run_Mode rhf
5 QM_Method hf
6 Charge 0
7 Integral_Threshold 1e-10
8 Energy_Threshold 1e-05
9 DIIS_Threshold 1e-05
10 RMSDP_Threshold 1e-05
11 MAXDP_Threshold 1e-05
12 Damping_Factor 0.70
13 Print_Level 1
14 Basis_Set STO-3G.dat
15 No_DFT_Cores 4
16 Geometry_Optimization yes
17 Compute_Hessian no
18 End_of_Keywords

```

Run the calculation. You should obtain a results file named ethanol-QM-Linux-GeomOpt.out, a file ethanol-QM-Linux-GeomOpt.xyz that shows all the steps of the geometry optimization (you can open it with Avogadro for example) and a file called ethanol-QM-Linux_qm.mol2 which will contain the optimized structure in mol2 format. Here are a few guidelines to interpreting the results file:

- At every step we evaluate convergence:

726		-----	
727			
728		--- STEP 1 --- COMPLETE	
729			
730		-----	
731		***** CONVERGENCE CRITERIA *****	
732		-----	
733		dEn. = 2.80e-01 Tolerance : 5.00e-04 Converged : NO	
734		Fmax = 1.91e-01 Tolerance : 1.50e-04 Converged : NO	
735		Frms = 1.30e-01 Tolerance : 4.50e-04 Converged : NO	
736		Drms = 1.26e-01 Tolerance : 1.20e-03 Converged : NO	
737		Dmax = 1.83e-01 Tolerance : 1.80e-03 Converged : NO	
738			

As the optimization runs the convergence criteria will start being fulfilled until all "Converged" fields will say "YES".

```

-----
          --- STEP 39 --- COMPLETE
-----
          ***** CONVERGENCE CRITERIA *****
-----
dEn.  = 1.11e-05      Tolerance : 5.00e-04      Converged : YES
Fmax  = 1.29e-04      Tolerance : 1.50e-04      Converged : YES
Frms  = 8.65e-05      Tolerance : 4.50e-04      Converged : YES
Drms  = 3.40e-04      Tolerance : 1.20e-03      Converged : YES
Dmax  = 5.64e-04      Tolerance : 1.80e-03      Converged : YES

          Geometry optimization complete.
-----

```

3. Obtaining force field parameters

In the platform, select Geometry optimization and FF parameters instead of Geometry optimization. Change the name of the output file to ethanol-QM-Linux-GeomOpt-FF. The parameter file should look like this:

```

parameters-queemist.txt
1 Molecule ethanol.mol2
2 Output ethanol-QM-Linux-GeomOpt-FF
3 Main_Mode qm
4 Run_Mode rhf
5 QM_Method hf
6 Charge 0
7 Integral_Threshold 1e-10
8 Energy_Threshold 1e-05
9 DIIS_Threshold 1e-05
10 RMSDP_Threshold 1e-05
11 MAXDP_Threshold 1e-05
12 Damping_Factor 0.70
13 Print_Level 1
14 Basis_Set STO-3G.dat
15 No_DFT_Cores 4
16 Geometry_Optimization yes
17 Compute_Hessian yes
18 End_of_Keywords

```

Run the calculation. You should obtain a results file titled ethanol-QM-Linux-GeomOpt-FF.out , a file called ethanol-QM-Linux-GeomOpt-FF.xyz that shows all the steps of the geometry optimization (you can open it with Avogadro for example), a file called ethanol-QM-Linux-GeomOpt-FF_customized_ff_parameters.txt which contains the custom force field parameters and a file called ethanol-QM-Linux-GeomOpt-FF_hessian.txt that contains the Hessian of the molecule. Here are a few guidelines to interpreting the results file:

```

-----
--- STEP 39 --- COMPLETE

***** CONVERGENCE CRITERIA *****
-----
dEn. = 1.11e-05      Tolerance : 5.00e-04      Converged : YES
Fmax = 1.29e-04      Tolerance : 1.50e-04      Converged : YES
Frms = 8.65e-05      Tolerance : 4.50e-04      Converged : YES
Drms = 3.40e-04      Tolerance : 1.20e-03      Converged : YES
Dmax = 5.64e-04      Tolerance : 1.80e-03      Converged : YES

      Geometry optimization complete.
-----

```

The geometry optimization successfully completed, as such the Hessian can now be computed.

```

-----
The flag for computing the Hessian matrix has been found. Will proceed.
-----
Performing (+) displacement on coordinate # 0 ... done
Computing (+) SCF for coordinate # 0 ... done
Computing (+) gradient for coordinate # 0 ... done
Performing (-) displacement on coordinate # 0 ... done
Computing (-) SCF for coordinate # 0 ... done
Computing (-) gradient for coordinate # 0 ... done
Computing Hessian elements for coordinate # 0 ... done
--- 26 coordinates left to compute. ---
Performing (+) displacement on coordinate # 1 ... done
Computing (+) SCF for coordinate # 1 ... done
Computing (+) gradient for coordinate # 1 ... done
Performing (-) displacement on coordinate # 1 ... done
Computing (-) SCF for coordinate # 1 ... done
Computing (-) gradient for coordinate # 1 ... done
Computing Hessian elements for coordinate # 1 ... done
--- 25 coordinates left to compute. ---
Performing (+) displacement on coordinate # 2 ... done
Computing (+) SCF for coordinate # 2 ... done
Computing (+) gradient for coordinate # 2 ... done
Performing (-) displacement on coordinate # 2 ... done
Computing (-) SCF for coordinate # 2 ... done
Computing (-) gradient for coordinate # 2 ... done
Computing Hessian elements for coordinate # 2 ... done

```

WARNING: depending on the size of the system, the Hessian computation can be very lengthy. Below find the end of the Hessian computation:

```

| ---      0 coordinates left to compute. ---
| Hessian eigenvalues:
| Eigenvalue      0 ---      0.000000
| Eigenvalue      1 ---      0.000000
| Eigenvalue      2 ---      0.000000
| Eigenvalue      3 ---      0.000011
| Eigenvalue      4 ---      0.000026
| Eigenvalue      5 ---      0.000074
| Eigenvalue      6 ---      0.002644
| Eigenvalue      7 ---      0.007837
| Eigenvalue      8 ---      0.030980
| Eigenvalue      9 ---      0.041964
| Eigenvalue     10 ---      0.070608
| Eigenvalue     11 ---      0.090664
| Eigenvalue     12 ---      0.113989
| Eigenvalue     13 ---      0.126167
| Eigenvalue     14 ---      0.155349
| Eigenvalue     15 ---      0.172208
| Eigenvalue     16 ---      0.199503
| Eigenvalue     17 ---      0.239940
| Eigenvalue     18 ---      0.280157
| Eigenvalue     19 ---      0.392493
| Eigenvalue     20 ---      0.743094
| Eigenvalue     21 ---      1.038901
| Eigenvalue     22 ---      1.130815
| Eigenvalue     23 ---      1.178609
| Eigenvalue     24 ---      1.265350
| Eigenvalue     25 ---      1.325864
| Eigenvalue     26 ---      1.431079
| Hessian contains only positive eigenvalues, optimized structure
| is a minimum !!!                                HURRAY
|
| The Hessian matrix has been successfully computed.
| -----
| Will now create a customized force field of the optimized structure.
| File name ... ethanol-QM-Linux-GeomOpt-FF customized ff parameters.txt

```

The generation of the force field parameters should be fairly fast. The message “*Force field file created successfully.*” will appear once the parameter computation finishes.

To visualize the force field parameters, open the file ethanol-QM-Linux-GeomOpt-FF_customized_ff_parameters.txt. Here is how the file should look like:

```
ethanol-QM-Linux-GeomOpt-FF_customized_ff_parameters.txt
```

```

1  -----
2  # ATOM TYPES
3  -----
4  1  C  3  2  3  4  5  00  00  -0.1910  -4.3341  0.5951  0.0846
5  2  C  3  1  6  7  8  00  00  0.0088  -3.0707  1.4731  -0.0757
6  3  H  0  1  00  00  00  00  00  0.0647  -5.0096  0.7579  -0.7495
7  4  H  0  1  00  00  00  00  00  0.0555  -4.8593  0.8409  1.0035
8  5  H  0  1  00  00  00  00  00  0.0669  -4.0646  -0.4561  0.1132
9  6  H  0  2  00  00  00  00  00  0.0477  -3.3678  2.5280  -0.1317
10 7  H  0  2  00  00  00  00  00  0.0636  -2.5736  1.2235  -1.0174
11 8  O  3  2  9  00  00  00  00  -0.2861  -2.1023  1.2627  0.9613
12 9  H  0  8  00  00  00  00  00  0.1699  -2.5815  1.5162  1.7912
13 10 *****
14 -----
15 # K is given in mdyn/A
16 # If Rel = 0 K is completely reliable
17 # If Rel = 1 an imaginary eigenvalue > 0.1 was detected for the bond - K is moderately reliable
18 # If Rel = 2 a negative eigenvalue was detected for the bond - K is unreliable since there is no bond between the atoms
19 # Bond length (L0) is given in Angstrom
20 -----
21 # BOND-GENERAL
22 -----
23 #
24 # I-J      REL.      K      L0      Bond moment  Bond Order  Comments
25 -----
26 BM-1      1  2      0      4.8873  1.5469  -0.0025  0.9882  None
27 BM-2      1  3      0      7.2385  1.0856  -0.0596  0.9821  None
28 BM-3      1  4      0      7.1822  1.0865  -0.0511  0.9832  None
29 BM-4      1  5      0      7.2614  1.0856  -0.0617  0.9829  None
30 BM-5      2  6      0      6.8059  1.0974  -0.0434  0.9653  None
31 BM-6      2  7      0      7.0552  1.0937  -0.0582  0.9715  None
32 BM-7      2  8      0      6.3991  1.4344  0.0810  0.9879  None
33 BM-8      8  9      0      9.9018  0.9912  -0.1714  0.9501  None
34 -----
35 # K is given in kcal / mol*deg ^ 2
36 # If Rel = 0 K is completely reliable
37 # If Rel = 1 an imaginary eigenvalue > 0.1 was detected for the bond - K is moderately reliable
38 # If Rel = 2 a negative eigenvalue was detected for the bond - K is unreliable since there is no bond between the atoms
39 # Angle value (A0) is given in Degrees
40 -----
41 # ANGLE-GENERAL
42 -----
43 #
44 # I-J-K      REL.      K      A0      Comments
45 -----
46 AM-1      1  2  6      0      1.1970  109.2579  None
47 AM-2      1  2  7      0      1.3709  109.3284  None
48 AM-3      1  2  8      0      2.9966  113.1566  None
49 AM-4      2  1  3      0      0.9692  110.0844  None
50 AM-5      2  1  4      0      0.9884  110.7348  None
51 AM-6      2  1  5      0      0.9781  110.4616  None
52 AM-7      2  8  9      0      1.2389  103.9680  None
53 AM-8      3  1  4      0      1.2219  108.3722  None
54 AM-9      3  1  5      0      1.2358  108.6550  None
55 AM-10     4  1  5      0      0.6631  108.4691  None
56 AM-11     6  2  7      0      1.0774  107.3655  None
57 AM-12     6  2  8      0      2.4782  111.1390  None
58 AM-13     7  2  8      0      1.1706  106.3878  None
59 -----

```

IV. QUEMIST AND REACTANTS AND PRODUCTS

In this section, we describe how to apply QUEMIST to more complex molecules with the ultimate objective to use these as templates for CONSTRUCTS and generated force field parameters with ACE. To start, create the modular workflow shown below.



We will first use reported structures to build models of reactants and products for the Sharpless dihydroxylation reaction. To do so, open the “File converter” menu by clicking on the cog on the first box of the workflow. We will use the Cartesian coordinates provided by Houk and co-workers shown below.³

S6

```

BECKE3LYP/GEN PSEUDO=READ
Os: 341/321/21 with ECP
all other elements: 6-31G*

E(RB+HF-LYP) = -566.383753477

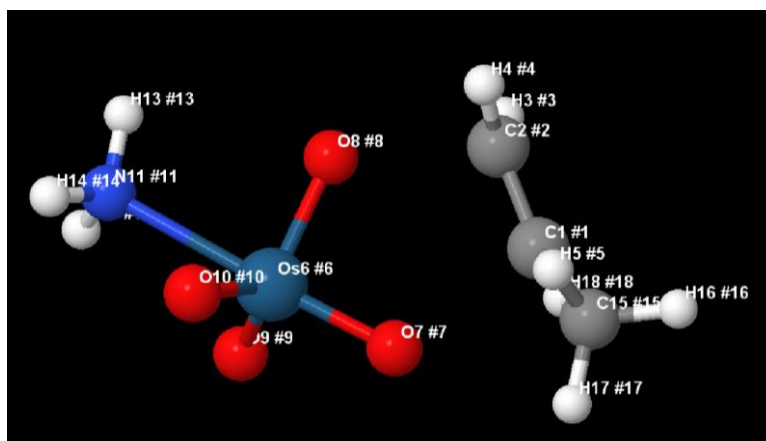
Zero-point correction=                .133976 (Hartree/Particle)
Thermal correction to Energy=          .145801
Thermal correction to Enthalpy=        .146745
Thermal correction to Gibbs Free Energy= .095318
Sum of electronic and zero-point Energies= -566.249777
Sum of electronic and thermal Energies=  -566.237953
Sum of electronic and thermal Enthalpies= -566.237009
Sum of electronic and thermal Free Energies= -566.288436

              E              CV              S
            KCAL/MOL      CAL/MOL-KELVIN  CAL/MOL-KELVIN
TOTAL              91.491              41.242              108.237

Standard orientation:
-----
Center      Atomic      Coordinates (Angstroms)
Number      Number           X             Y             Z
-----
  1           6         2.800552        .283553        .388893
  2           6         2.325823        1.543624        .085438
  3           1         2.428505        1.941977       -.918403
  4           1         2.122397        2.262903        .870033
  5           1         2.941296        .027668        1.435018
  6          76        -.475420       -.167099        .062250
  7           8         1.039055       -.934166        .524284
  8           8         .257573        1.409297       -.251936
  9           8        -.900363       -1.050549       -1.366843
 10           8        -1.385725       -.257608        1.533398
 11           7        -2.528883        .975162       -.591331
 12           1        -2.922272        .521619       -1.414645
 13           1        -2.347938        1.954243       -.807331
 14           1        -3.199871        .923602        .173815
 15           6         3.489683       -.597112       -.613250
 16           1         4.579755       -.477558       -.537419
 17           1         3.258273       -1.651406       -.433261
 18           1         3.193274       -.345843       -1.637195
-----

```

These coordinates will be entered in the “File Converter” menu as shown below. As you will see, some reformatting may be necessary. This transition state also needs some changes to convert it alternatively into reactants or into a product. Click on “Fix structure” then on “Preview”. The structure will be opened in Jmol as shown below.

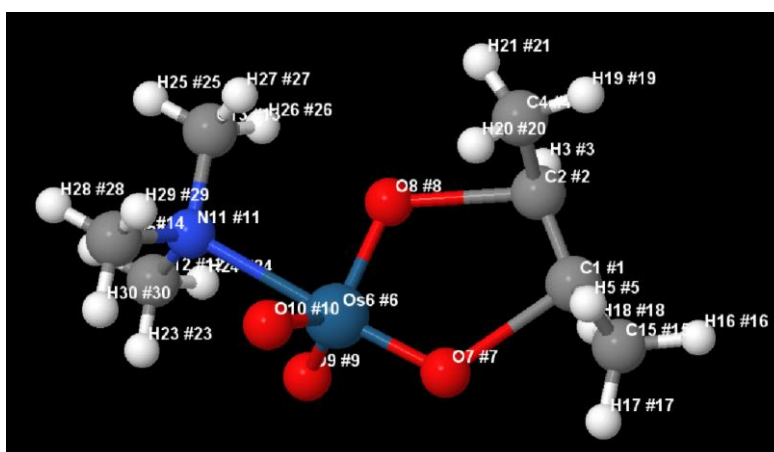


First, the NH_3 group used as OsO_4 ligand will be converted into Me_3N and the propene will be converted into butene by converting 4 hydrogen atoms (#4, 12, 13 and 14) into methyl groups. Bonds will then be added between atoms 2 and 8 and between atoms 1 and 7.

Settings

Source of ligand structures	To be Entered Below ▼
Output file name	molecule_mol2
Molecule Name	Dihydroxylation_Product
Cartesian Coordinates	<pre> C 2.8005 0.2835 0.3888 C 2.3258 1.5436 0.0854 H 2.4285 1.9419 -0.9184 H 2.1223 2.2629 0.8700 H 2.9412 0.0276 1.4350 Os -0.4754 -0.1670 0.0622 O 1.0390 -0.9341 0.5242 O 0.2575 1.4092 -.2519 O -0.9003 -1.0505 -1.3668 O -1.3857 -0.2576 1.5333 N -2.5288 0.9751 -0.5913 H -2.9222 0.5216 -1.4146 H -2.3479 1.9542 -0.8073 H -3.1998 0.9236 0.1738 C 3.4896 -0.5971 -0.6132 H 4.5797 -0.4775 -0.5374 H 3.2582 -1.6514 -0.4332 H 3.1932 -0.3458 -1.6371 </pre>
	Fix structure
Add Methyl groups	4 1 12 13
Add Phenyl groups	
Adding Bonds	2 8 1 1 7 1
	Preview
Rotate bonds	
Adjust bond length	2 8 1.4 1 7 1.4

Clicking on Preview will open the modified structure as shown below.



The change in bond length to convert the TS into either reactants or product will be carried out when the workflow will be launched.

You can now save the parameters.

Running FILECONVERTER

Edit Parameter File

Save

parameters-fileconverter.txt

Now we are ready to open QUEMIST. In the “Source of molecule structures”, select “From File Converter Box”. This will ensure that we use the Cartesian coordinates we introduced previously. We will be using geometry optimization coupled with FF parameter generation. For this reaction, we will use the “Forecaster-BS” basis set, which is a combination of several basis sets designed for speed and accuracy. Select also the number of CPU cores you would like to use (4 in our case) and save the parameters.

Settings

Source of molecule structures: From File Converter box (Wor...

Output file name: Dihydroxylation_Reactants_QM

Mode: Geometry Optimization and F...

Basis Set: Forecaster-BS.dat

Overall molecule charge: 0

Number of Cores: 4

This will produce a parameter file which should look the one below.

```
Molecule dihydroxylation_reactants.mol2
Output Dihydroxylation_Reactants_QM
Main_Mode qm
Run_Mode rhf
QM_Method hf
Charge 0
Integral_Threshold 1e-10
Energy_Threshold 1e-05
DIIS_Threshold 1e-05
RMSDP_Threshold 1e-05
MAXDP_Threshold 1e-05
Damping_Factor 0.70
Print_Level 1
Basis_Set Forecaster-BS.dat
No_DFT_Cores 4
Geometry_Optimization yes
Compute_Hessian yes
End_of_Keywords
```

The workflow is now ready to be started.

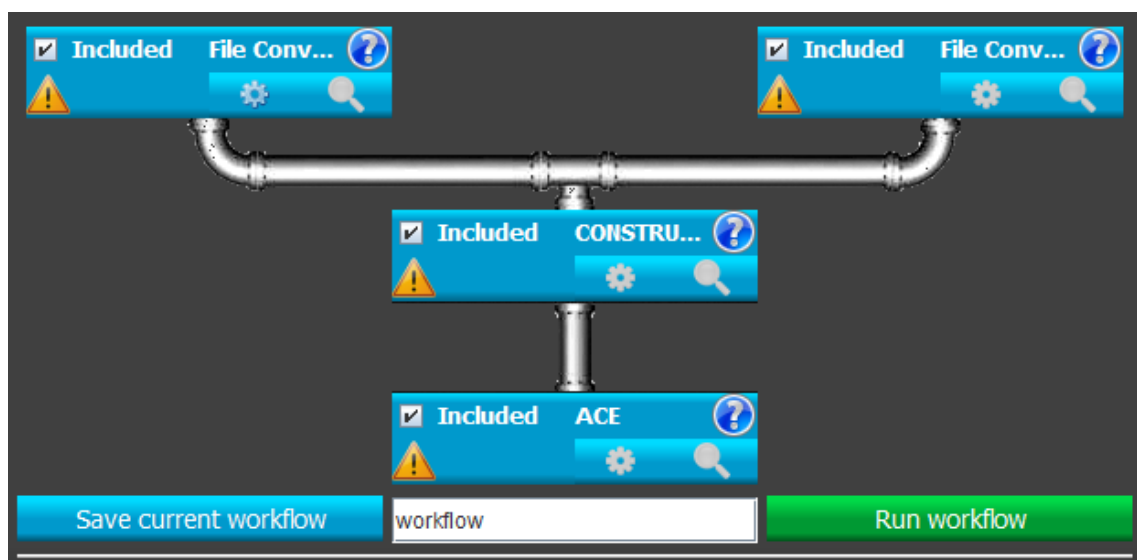
WARNING: Run this calculation only on Linux. This calculation will take some time, ranging from a few hours to a few days depending on the PC power and number of cores used (if you use more cores it will take less time). Also, computing the Hessian needed to generate the FF parameters is very time-expensive! So be prepared to not use your PC for quite some time.

This first workflow will generate the files for the product. Run the same workflow with adjusted parameters for the reactants. At the end of these computations, you will have a new reaction parameterized for ACE.

V. CONSTRUCTS

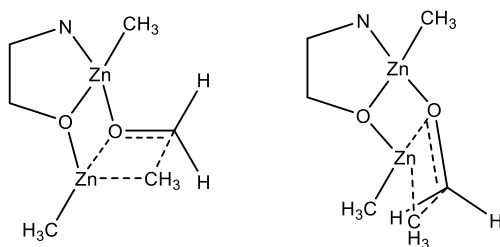
1. Diethyl zinc addition to aldehydes

We now wish to use the optimized structures with CONSTRUCTS to generate templates with associated force field parameters and use them to create transition state structures ready for stereoselectivity prediction with ACE. We start using the following workflow

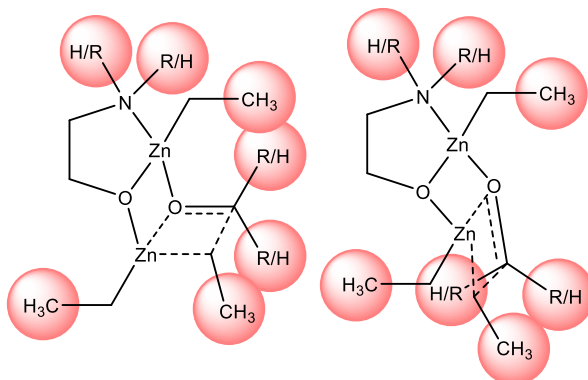


The *anti* and *syn* transition states reported by Norrby and co-workers will be used.⁴ The first “File Converter” menu will be open. The Cartesian coordinates available as supporting information to the report from Norrby and co-workers will then simply be copy/pasted into the Cartesian coordinate box after the source of ligand structures is set to “To be entered below”

a) Transition state from Norrby et al.

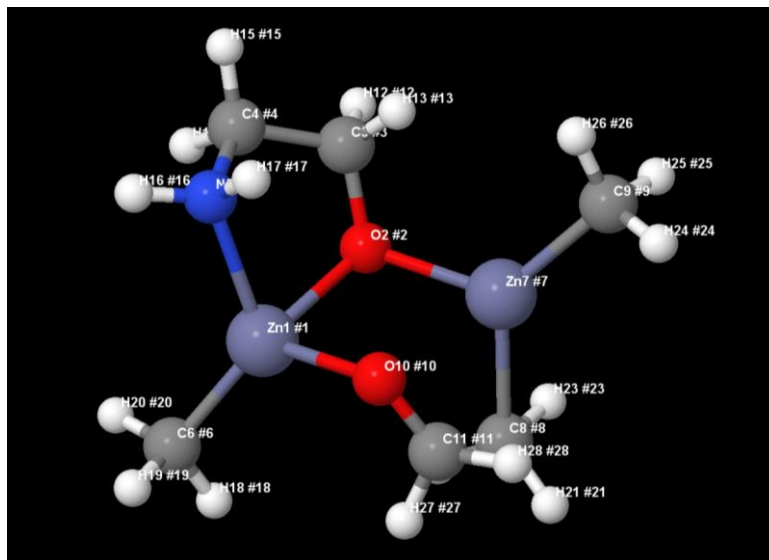


b) Templates developed herein. Templates for the 2 rings flips will also be developed as well as the enantiomers.

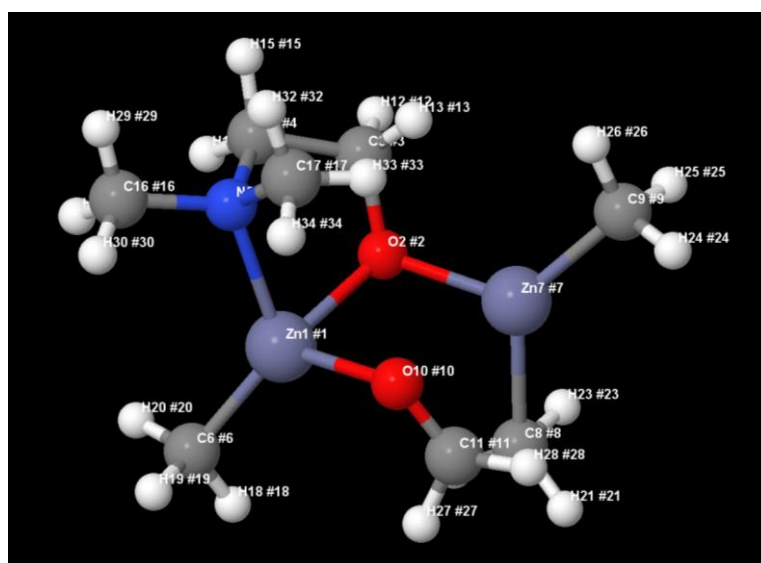


Source of ligand structures	To be Entered Below
Output file name	TS-Anti
Molecule Name	TS-Anti
Cartesian Coordinates	<pre>Zn -1.44340 -0.47728 0.03998 O -0.08044 0.76329 -0.85089 C -0.34536 2.13468 -0.65064 C -1.79105 2.36225 -0.19104 N -2.09632 1.40812 0.90368 C -2.66954 -1.90150 -0.58914 Zn 1.71967 0.13572 -0.19425 C 1.79150 -1.85011 -1.01312 C 3.10586 1.32739 0.57730 O 0.10284 -0.95001 1.40506 C 0.86209 -1.91054 1.04724 H -0.18562 2.70558 -1.57842 H 0.33259 2.56440 0.10666 H -2.48581 2.14182 -1.00915 H -1.93731 3.40858 0.11112 H -3.07700 1.45322 1.17341 H -1.54073 1.62063 1.73323 H -2.13450 -2.65801 -1.17633 H -3.15075 -2.41868 0.25050 H -3.46657 -1.50387 -1.22979 H 2.39608 -2.69636 -0.67638 H 0.89907 -2.20859 -1.53199 H 2.41096 -1.33763 -1.77423 H 3.72241 0.79753 1.31362 H 3.78240 1.70570 -0.19958 H 2.66870 2.19842 1.08127 H 0.43043 -2.82539 0.62894 H 1.82159 -2.03661 1.56242</pre>
	Fix structure

To visualize the structure, click on “Fix structure”. In this current case, we are converting the NH_2 of the catalyst reported into $\text{N}(\text{Me})_2$ which we believe would be more representative of the catalysts under investigation. To do so, two hydrogens (16 and 17) are converted into methyl. To identify the atom number, you may click on “Preview”. Before the instruction to add two methyl groups, the structure looks like this:



As soon as “16 17” is added to the “Add methyl groups” text box, clicking on preview shows:



Once this is done, save the parameters.

The second “File converter” box can now be setup similarly with the *syn* TS also adding two methyl groups.

Source of ligand structures	To be Entered Below ▼
Output file name	TS-Syn
Molecule Name	TS-Syn
Cartesian Coordinates	<pre> Zn -1.39243 -0.66756 0.04334 O -0.12779 0.57259 -0.95067 C -0.50953 1.92617 -0.95696 C -1.99564 2.07925 -0.61021 N -2.28095 1.29429 0.61567 C -2.61314 -2.17114 -0.36141 Zn 1.70750 -0.01017 -0.34749 C 2.91975 -1.21387 -1.34487 C 2.14547 1.39552 1.26463 O 0.23349 -1.08966 1.32253 C 0.98017 -0.38742 2.07361 H -0.33871 2.37936 -1.94611 H 0.07863 2.51174 -0.22988 H -2.60882 1.65398 -1.41227 H -2.26012 3.14059 -0.50213 H -3.28210 1.23409 0.79139 H -1.85895 1.73976 1.43057 H -2.07074 -2.98695 -0.85348 H -3.05584 -2.58008 0.55551 H -3.43650 -1.87853 -1.02496 H 3.54549 -1.81063 -0.67067 H 2.34803 -1.90515 -1.97450 H 3.59094 -0.64551 -2.00081 H 2.74357 1.88157 0.47152 H 1.40044 2.13127 1.57890 H 2.84839 1.21768 2.08255 H 1.91446 -0.83026 2.43379 H 0.54846 0.38463 2.72152 </pre>
	Fix structure
Add Methyl groups	16 17
Add Phenyl groups	
Adding Bonds	
	Preview

By default, the parameters are saved into parameters-fileconverter.txt. If this name is not changed, you will be warned that the keyword file has already been used by the other “File converter” box (there are two of them in the workflow) and should not be used again as it would overwrite the parameters of the other box and affect the calculations. Change the file name into for example parameters-fileconverter2.txt and save.

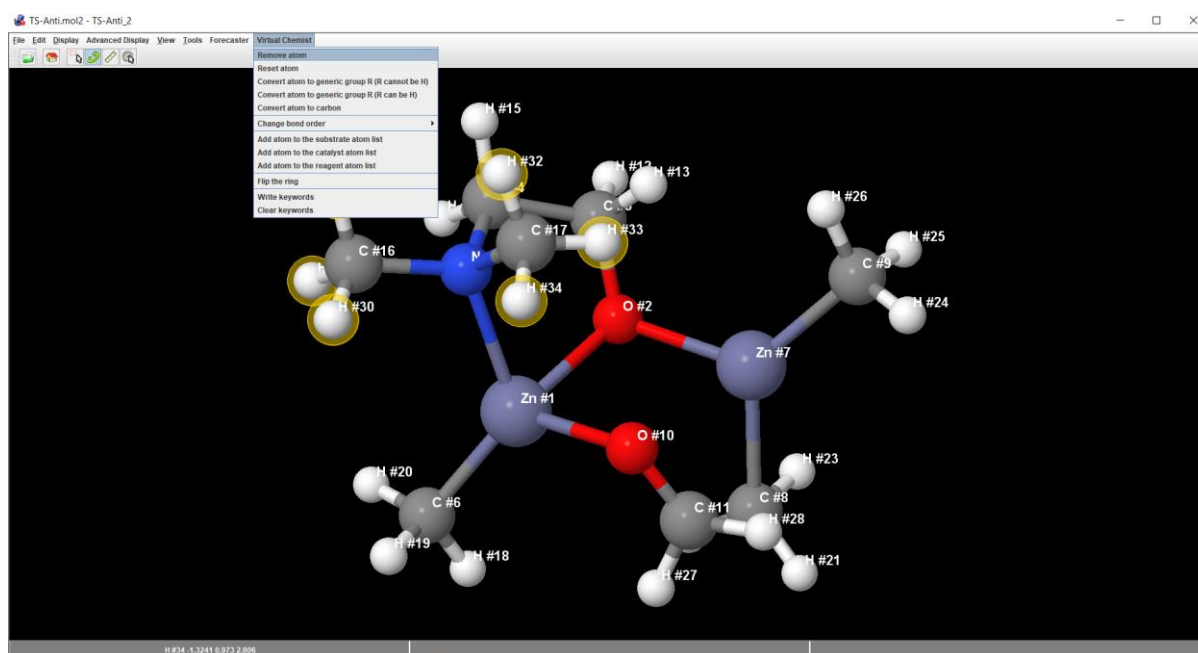
At this stage, these two TS structures will be used to prepare TS templates to assemble the catalyst-substrate containing TS structures. To do so, the “CONSTRUCTS” box should be setup. In a first step, we recommend building the templates and checking them before assembling the TS structures. To do so, select “Prepare template only” as show below. As CONSTRUCTS has identified that two “File converter” boxes are above in the workflow, it automatically fills the menu with the corresponding files (TS-Anti.mol2 and TS-Syn.mol2 in this case).

Settings

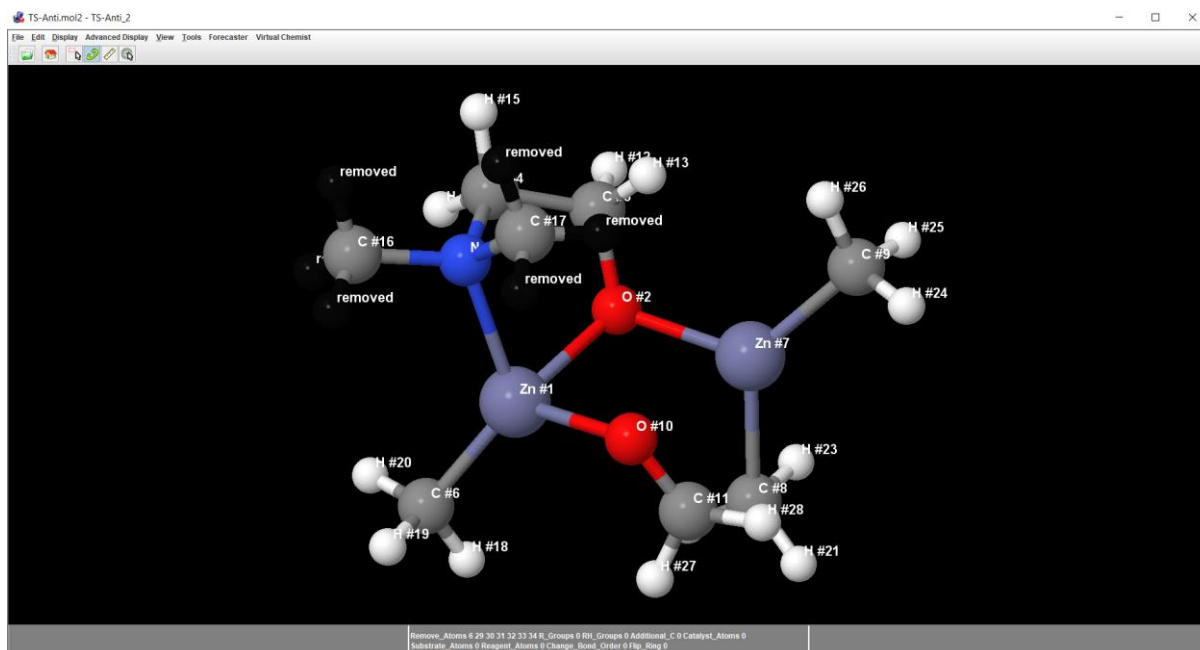
Source of catalyst structures	From File Converter box above ▼
Preset reactions	No ▼
Transition state template	Build from TS file (mol2) ▼
Prepare TS templates	Prepare templates only ▼
Transition states	2 transition states ▼
Transition state file #1	TS-Anti.mol2 ▼

Load template structure

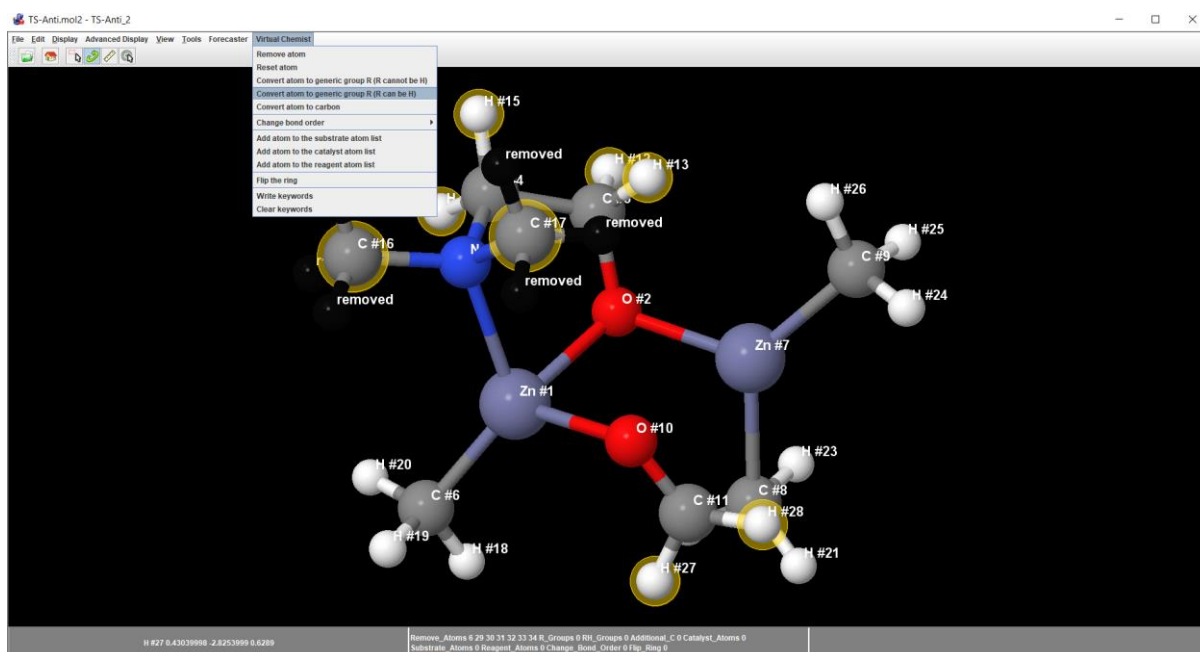
These templates can now be built by clicking on “Load template structure”. The files will be loaded and we will now characterize the templates from the TS structures. First, the catalyst part O-CH₂-CH₂-NMe₂ group will be converted into a generic O-CR₂-CR₂-NR₂. To do so, first select the 6 hydrogens of the methyl groups by clicking on them and remove them as shown below.



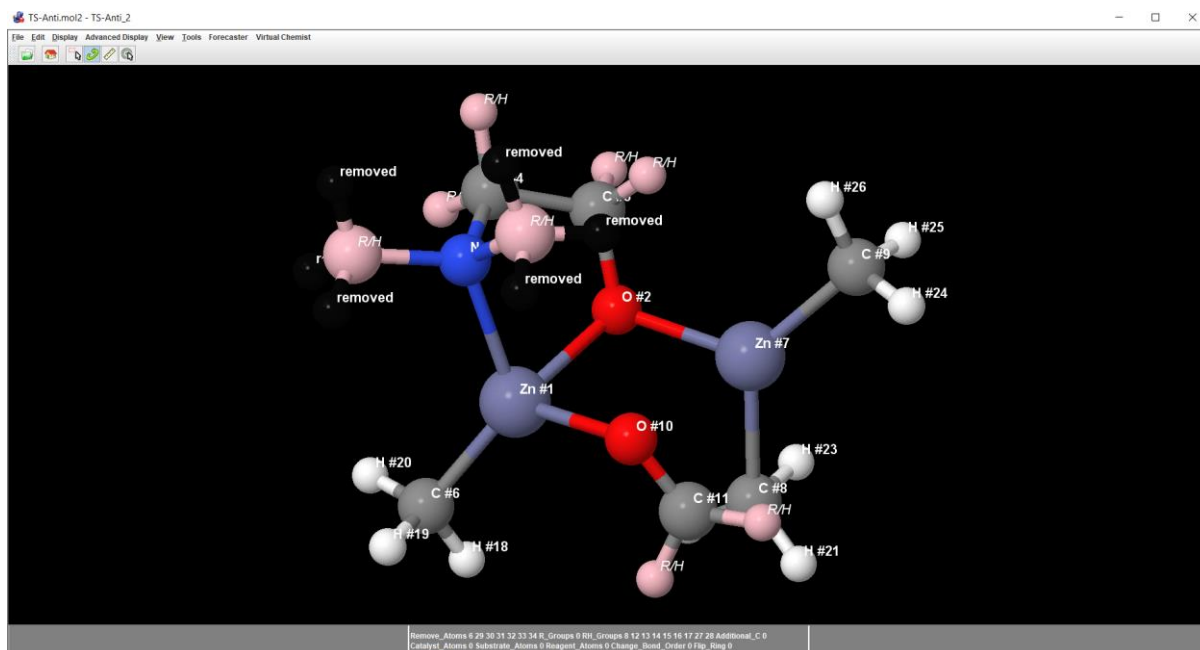
At this stage, the structure will now look like this:



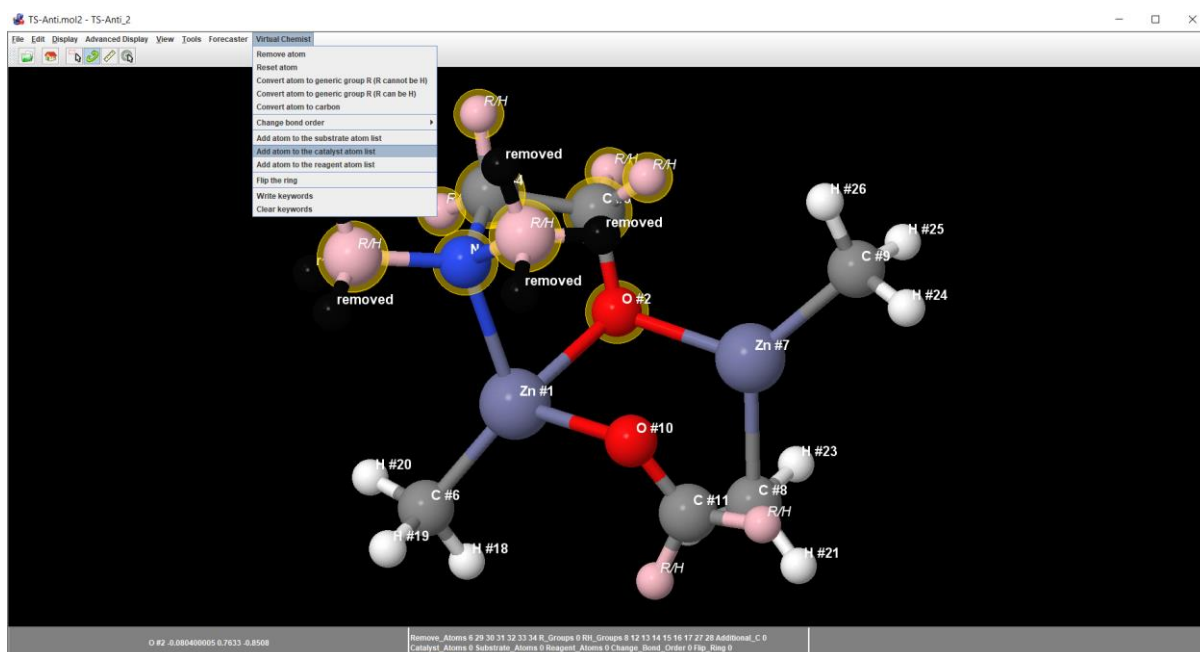
Then select the different atoms which will be converted into R (can be H) using the menu shown below:



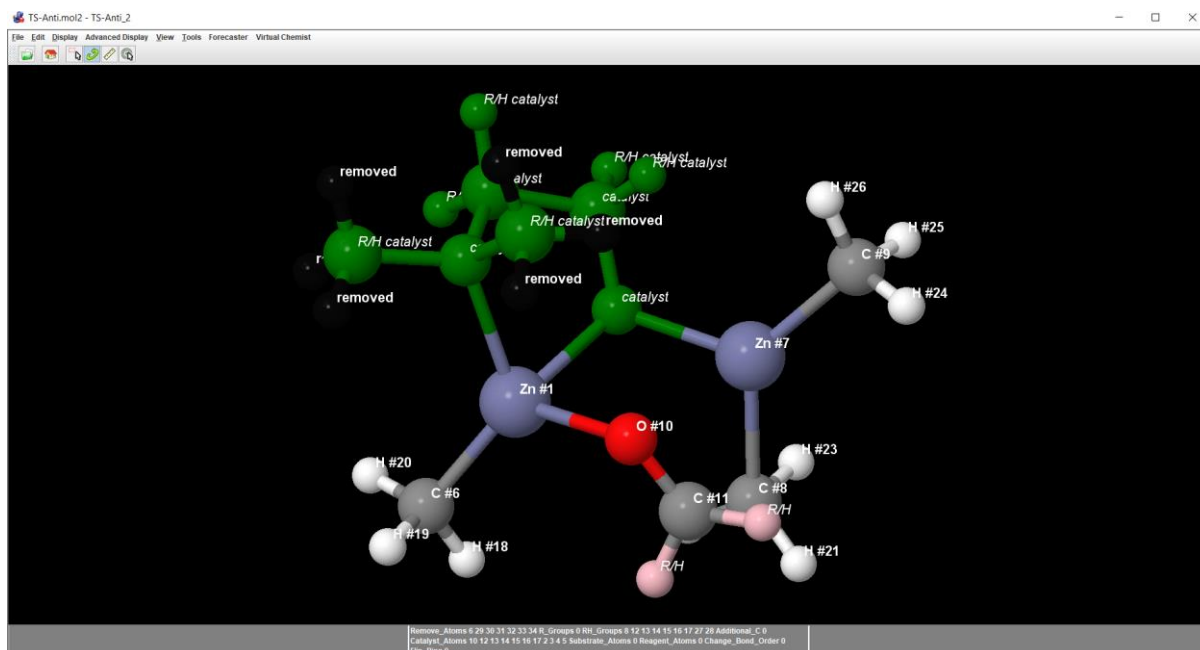
At this stage, the structure will look like this:



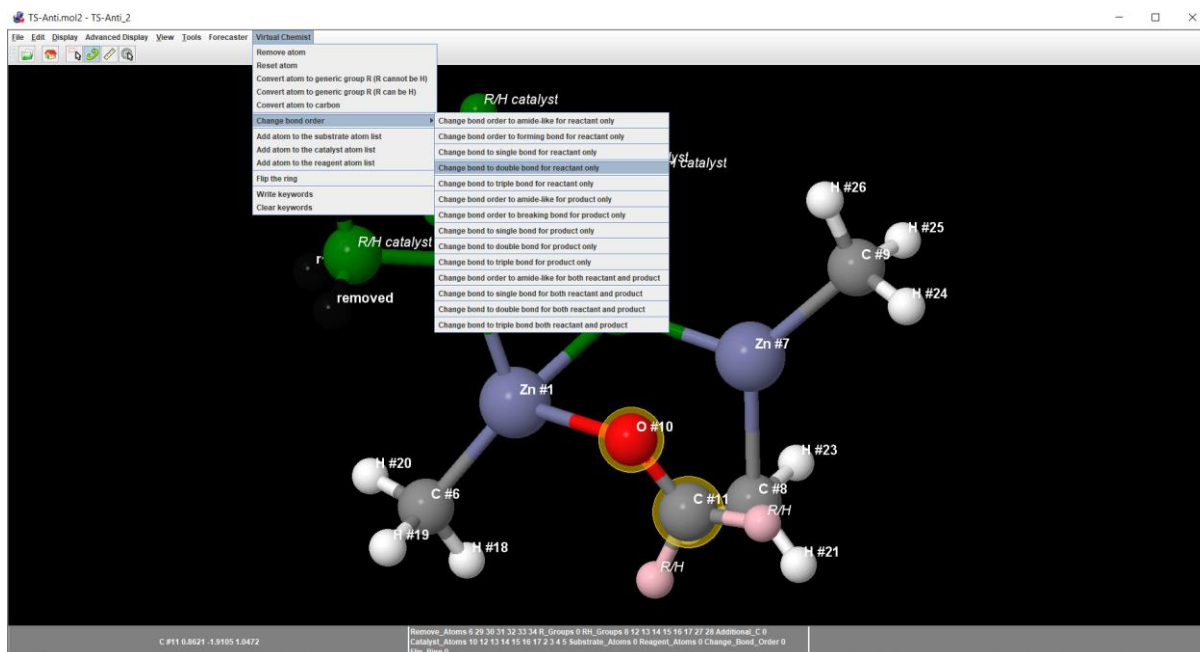
Now, the different atoms making up the catalyst (amino alcohol in this case) will be identified as shown below:



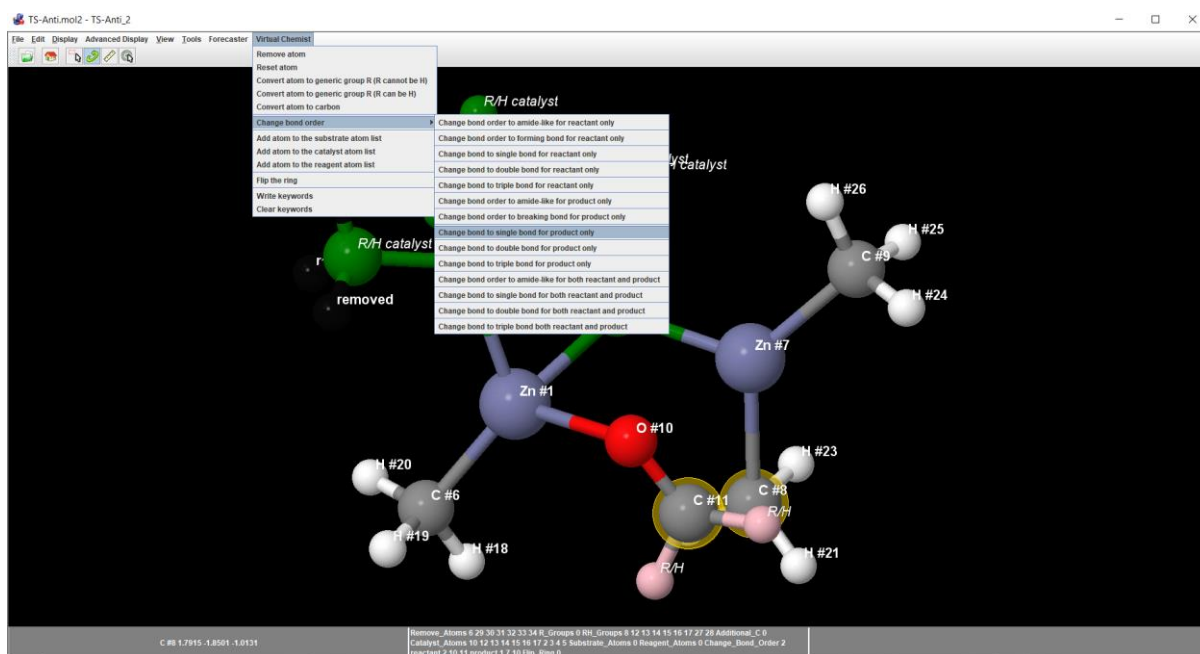
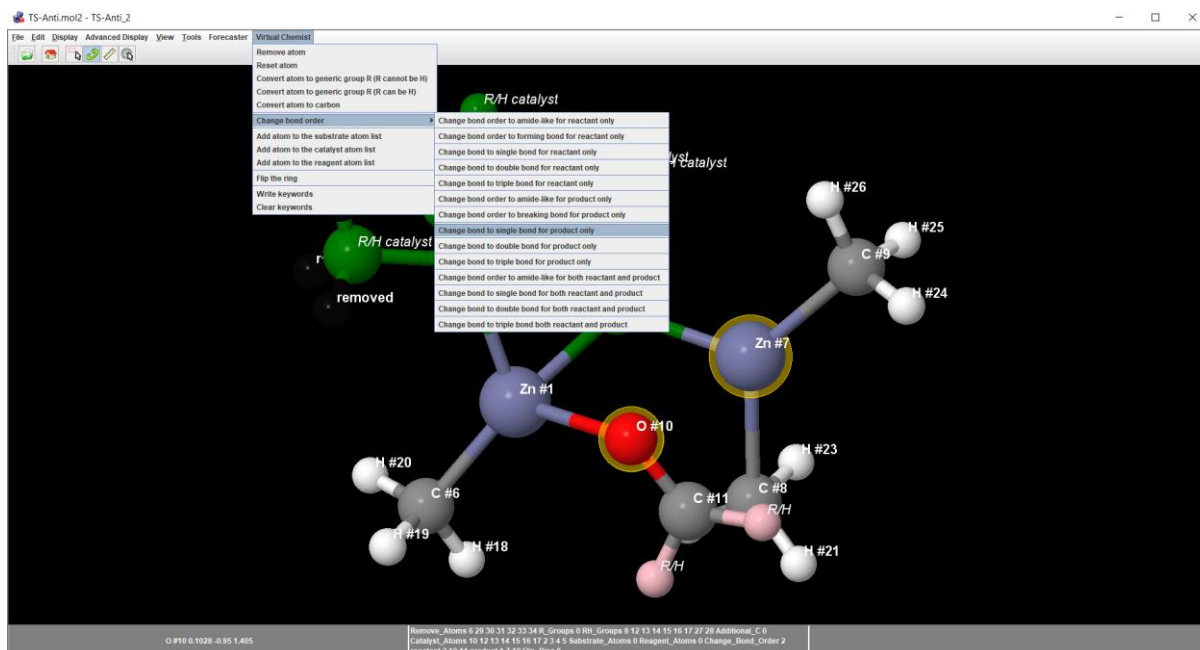
These atoms will be used to place the actual catalyst when assembling the TS structures. At this stage, the structure looks like this:



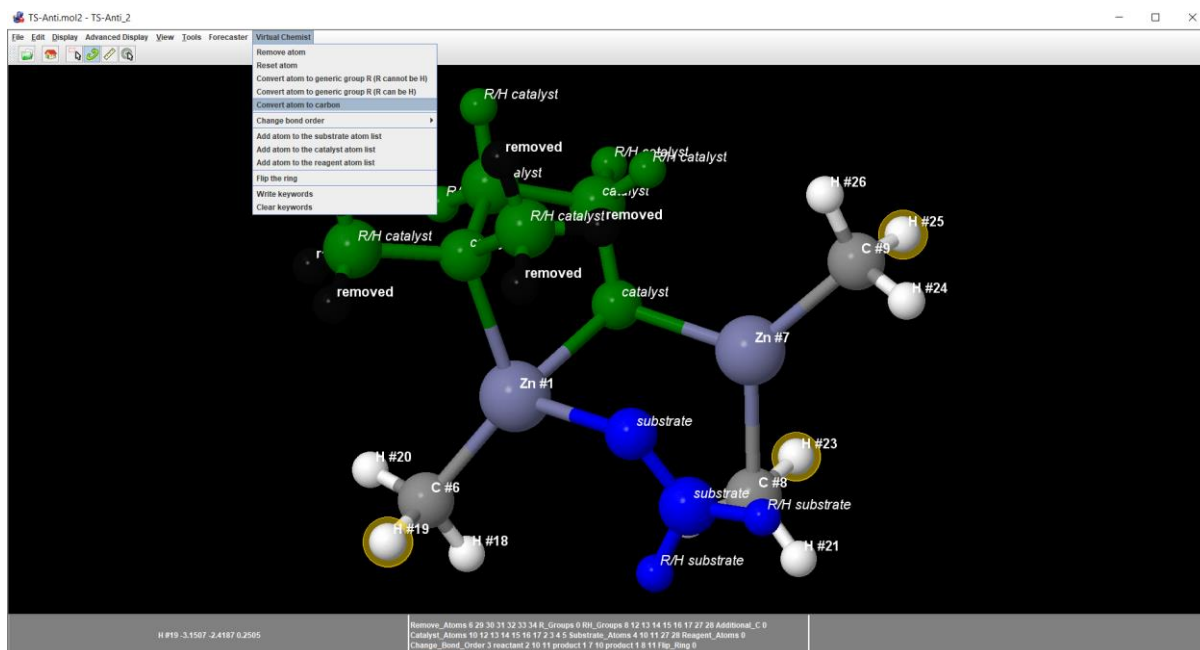
The bonds order should also be corrected. For instance, through the course of the reaction, the carbonyl (C=O) group will be converted into an alcohol (C-O bond). Select the two atoms of this carbonyl group and change the bond into a “double bond for reactant only”. The program will understand that the bond remains a single bond in the product.



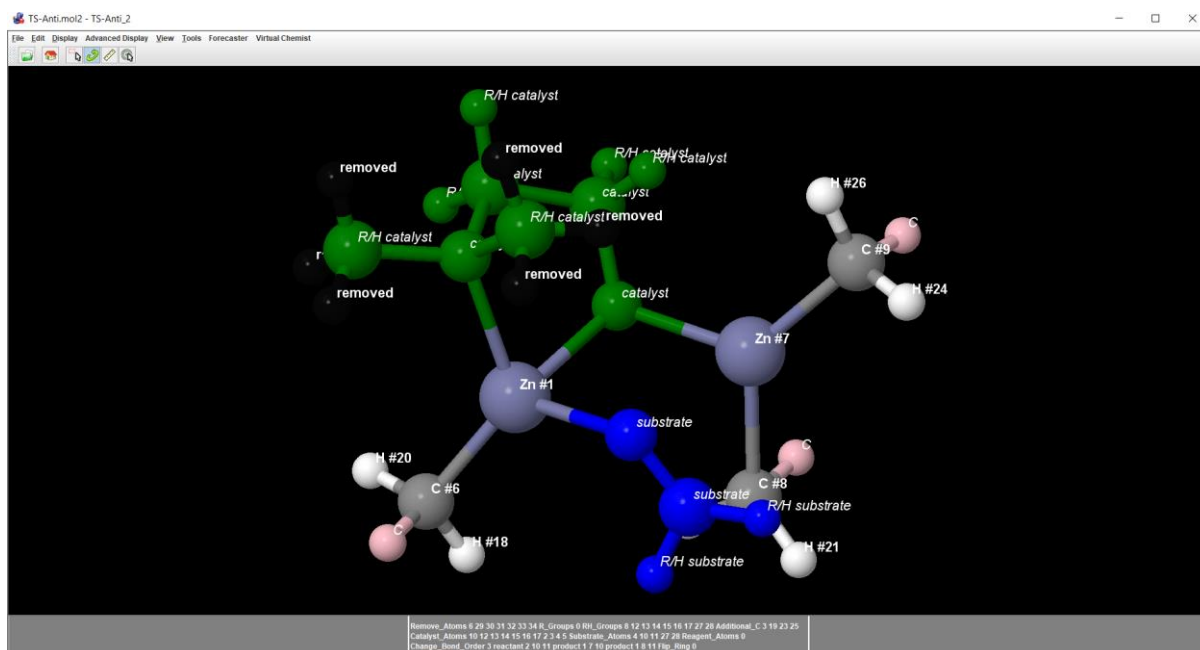
Do the same with the O-Zn and C-C bonds which will form during the course of the reaction.



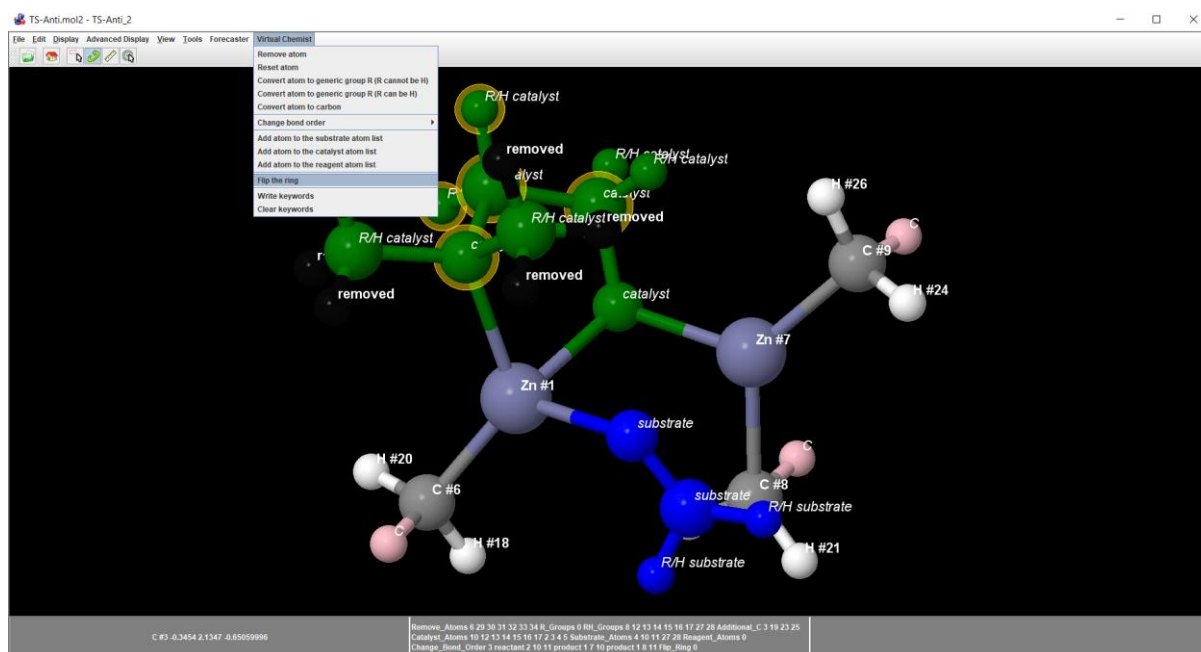
And the Zn-C bond breaking should be identified as follows:



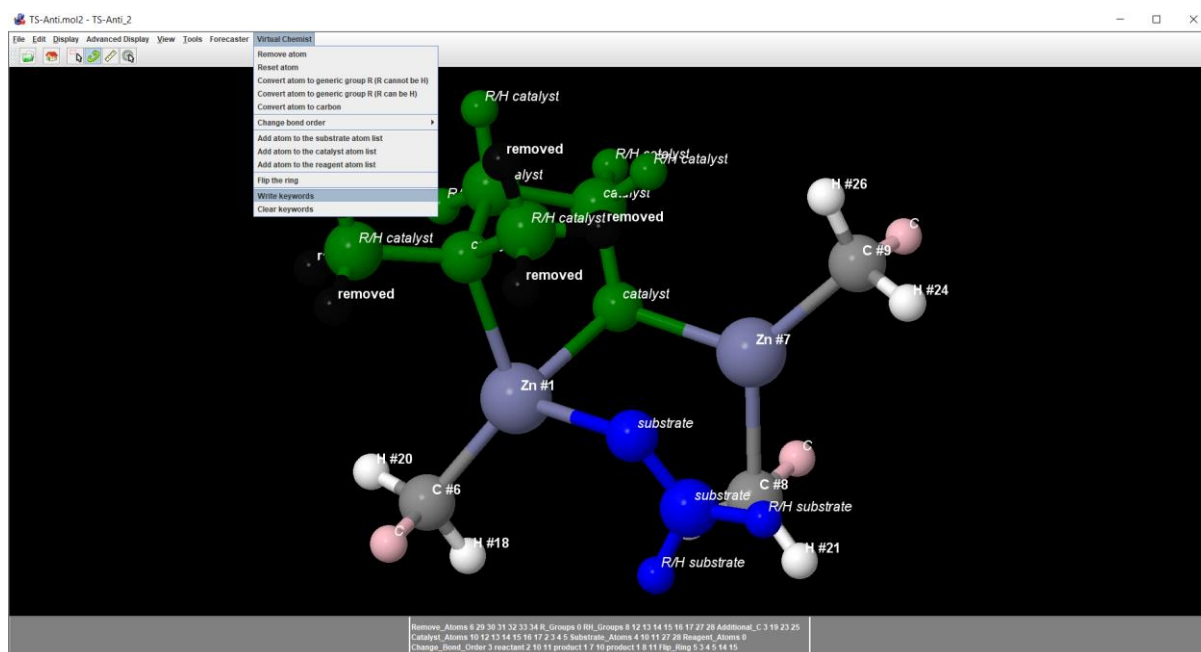
At this stage, the structure should look like this:



A last step is to prepare templates with ring flipped for optimal TS structure development. To do so, select an atom of the ring to flip and the four atoms directly connected and select "flip the ring". With these instructions, both flipped templates will be produced.



At this stage, the instructions are complete, and the parameters can be saved as shown below.



This will automatically add these instructions to the CONSTRUCTS menu.

Transition state file #2 TS-Syn.mol2

Load template structure

Parameters for template 2

```

Remove_Atoms 6 29 30 31 32 33 34
R_Groups 0
RH_Groups 8 12 13 14 15 16 17 27 28
Additional_C 3 19 23 24
Catalyst_Atoms 10 12 13 14 15 16 17 2 3
4 5
Substrate_Atoms 4 10 11 27 28
Reagent_Atoms 0
Change_Bond_Order 4 reactant 2 10 11
product 1 7 10 product 1 9 11 product 0
7 9
Flip_Ring 5 3 4 5 14 15

```

Only this enantiomer

Catalyst residue name CAT1

Follow the same steps with the other TS (TS-Syn.mol2 in this case). As you can see below, you may also instruct the program to derive both enantiomeric forms of the templates by selecting "Both enantiomers".

Transition state file #2 TS-Syn.mol2

Load template structure

Parameters for template 2

```

Remove_Atoms 6 29 30 31 32 33 34
R_Groups 0
RH_Groups 8 12 13 14 15 16 17 27 28
Additional_C 3 19 23 24
Catalyst_Atoms 10 12 13 14 15 16 17 2 3
4 5
Substrate_Atoms 4 10 11 27 28
Reagent_Atoms 0
Change_Bond_Order 4 reactant 2 10 11
product 1 7 10 product 1 9 11 product 0
7 9
Flip_Ring 5 3 4 5 14 15

```

Only this enantiomer

Catalyst residue name CAT1

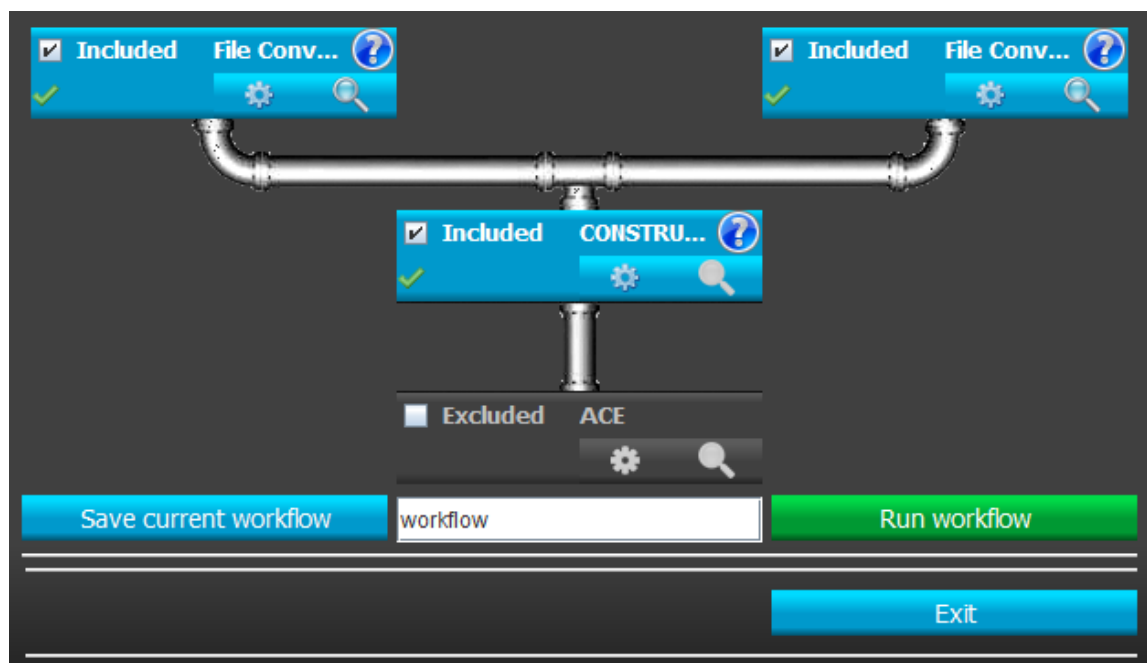
Finally, to ensure that all the TS structures eventually produced with all the catalysts and substrates will not be in a single file, you may indicate how many files (ie, how many cores you will use for computation); 200 is requested below. You may also request a minimum number of structures per file (16 here). The default for the other parameters can be used. To complete the process, save your parameters.

Catalyst residue name	<input type="text" value="CAT1"/>
Substrate residue name	<input type="text" value="REA1"/>
Reagent residue name	<input type="text" value="REA2"/>
Output File Name	<input type="text" value="constructs_output"/>
ACE to be ran on multiple cores	<input type="button" value="Multicore or cluster"/>
Number of cores	<input type="text" value="200"/>
Number of TS configurations	<input type="text" value="16"/>

Running CONSTRUCTS

<input type="button" value="Edit Parameter File"/>	<input type="text"/>
<input type="button" value="Save"/>	<input type="text" value="parameters-constructs.txt"/>

At this stage, 3 of the 4 boxes of the workflow are ready. Exclude ACE for now and click on “Run workflow”. This will take a few seconds to complete. You may now want to open the produced templates (constructs_output-autoTS.mol2 here) to check whether they correspond to your needs.



When this is done, select “Prepare template and TS for ACE” to provide the catalysts and substrates file names as shown below.

Settings

Source of catalyst structures	From File Converter box above ▼
Library of Catalysts	Et2ZnAddition-catalysts.sdf ▼
Library of Substrates	Et2ZnAddition-substrates.sdf ▼
Library of Reagents	none ▼
Preset reactions	No ▼
Transition state template	Build from TS file (mol2) ▼
Prepare TS templates	Prepare template and TS for ... ▼
Transition states	2 transition states ▼
Transition state file #1	TS-Anti.mol2 ▼

Load template structure

Parameters for template 1

```
Remove_Atoms 6 29 30 31 32 33 34
R_Groups 0
RH_Groups 8 12 13 14 15 16 17 27 28
Additional_C 3 19 23 25
Catalyst_Atoms 10 12 13 14 15 16 17 2 3
```

Your workflow is now ready to assemble the TS structures. It is not recommended to run ACE with too many catalysts and substrates on a regular Windows PC as it may take a lot of time (ca. 1 to 12 hours per catalyst-substrate pair). Also be aware that laptops are not necessarily set up to run heavy calculations which require ventilation (heat released by the processor) often good enough in workstations or desktop PC/Mac but not always on laptops.

To prepare the ACE keywords, use (the force field parameters have been developed in a separate workflow presented earlier in this tutorial):

Settings

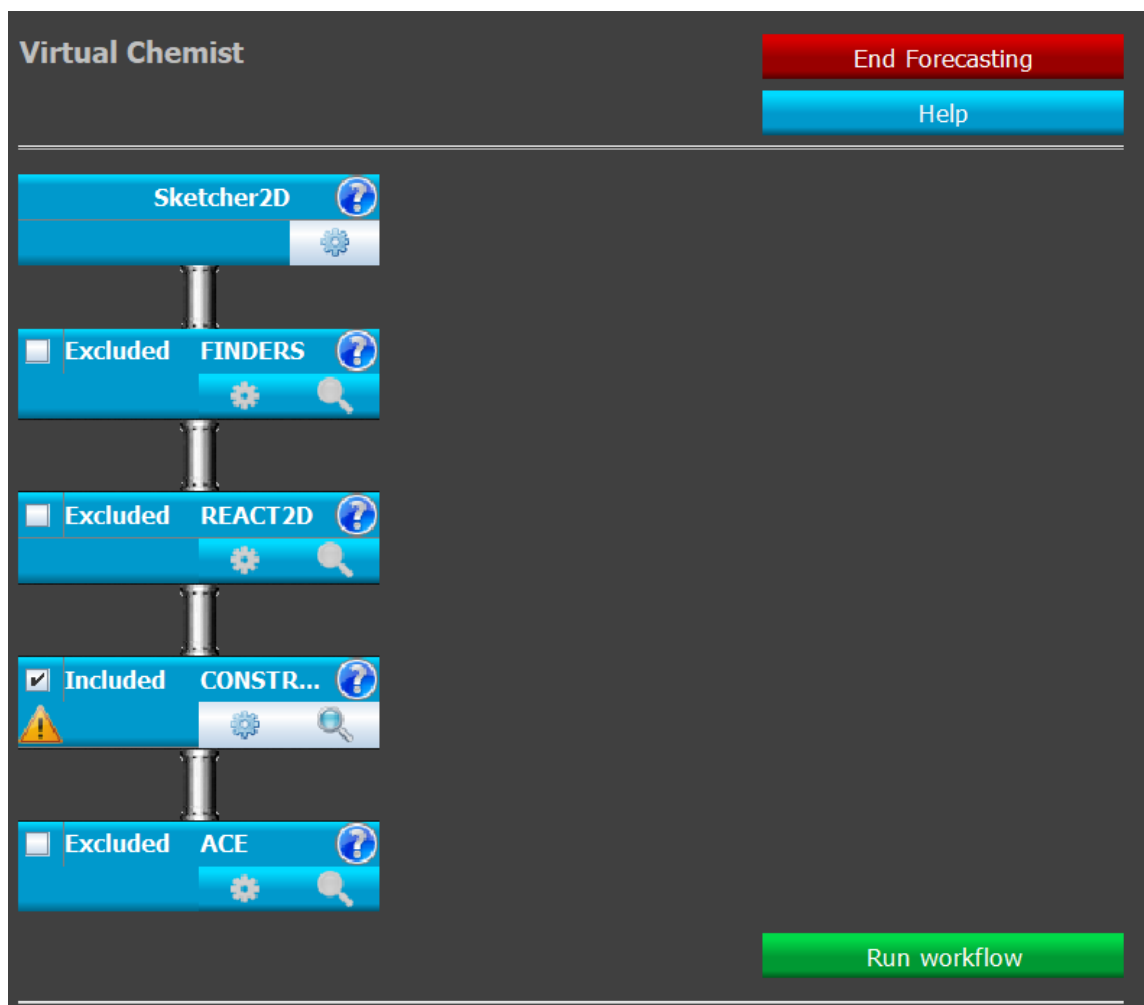
Source of TS structures	From CONSTRUCTS box abov...	
Output File Name	ace_output	
ACE or Q2MM	ACE	
Ground states parameters	2 reactants and 2 products	
Q2MM parameters	ACE_Zn_Anti_Products_resul...	
ACE_Zn_Reactants_Anti_res...	ACE_Zn_Products_Anti_resul...	
ACE_Zn_Reactants_Syn_res...	ACE_Zn_Products_Syn_resul...	
Number of lambda values(s)	1 - Synchronous	
Lambda #1	0.5	* - Generic Lambda
Temperature (in Celcius)	0	
Epsilon (dielectric constant)	4.0	
Number of TS configurations	4	

You can now run the workflow again which will generate all the necessary files ready to be used by ACE.

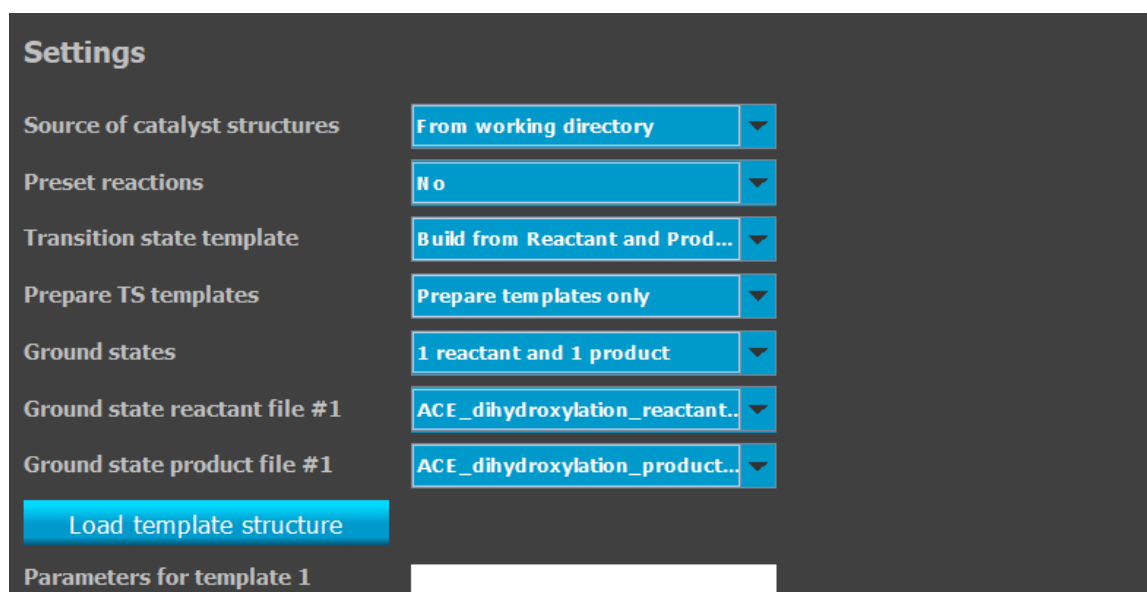
2. Asymmetric Dihydroxylation

For this example, we will use the optimized reactant and product structure files to build the templates and the VIRTUAL CHEMIST preset workflow. These optimized structures were obtained using QUEMIST and TS previously reported following a strategy similar to the one presented earlier.

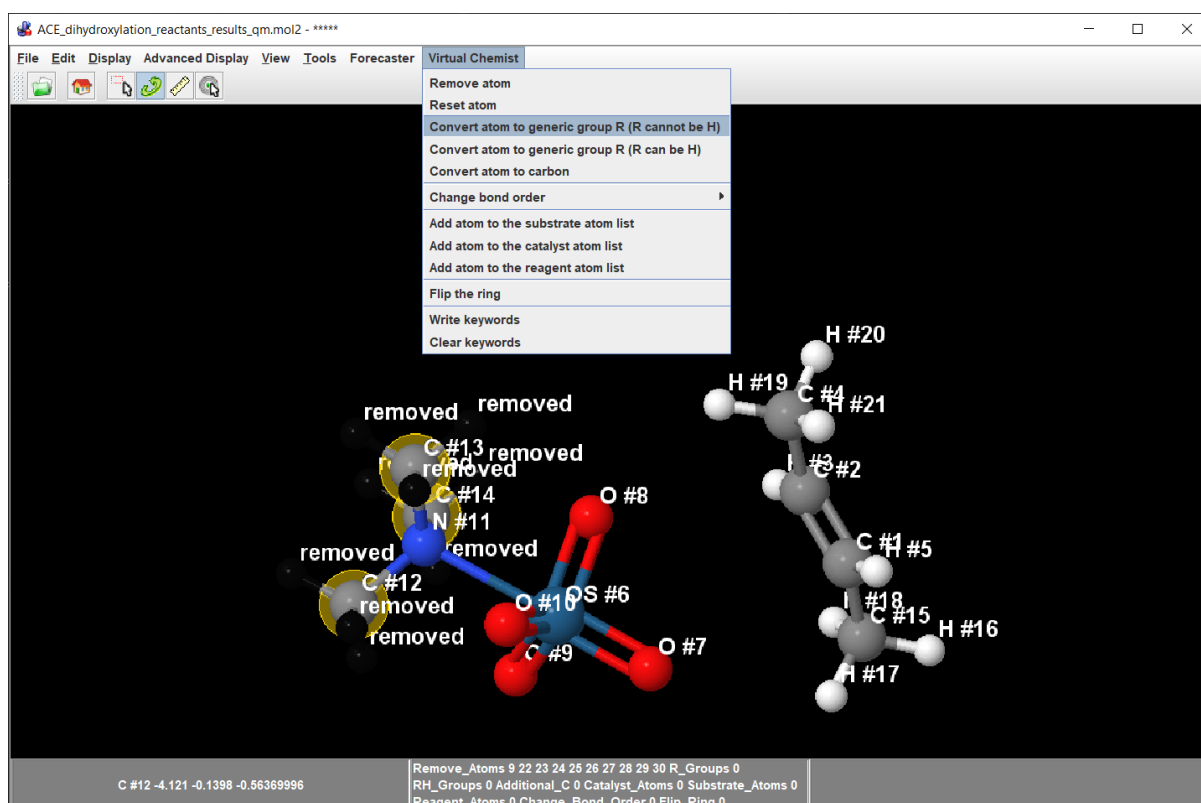
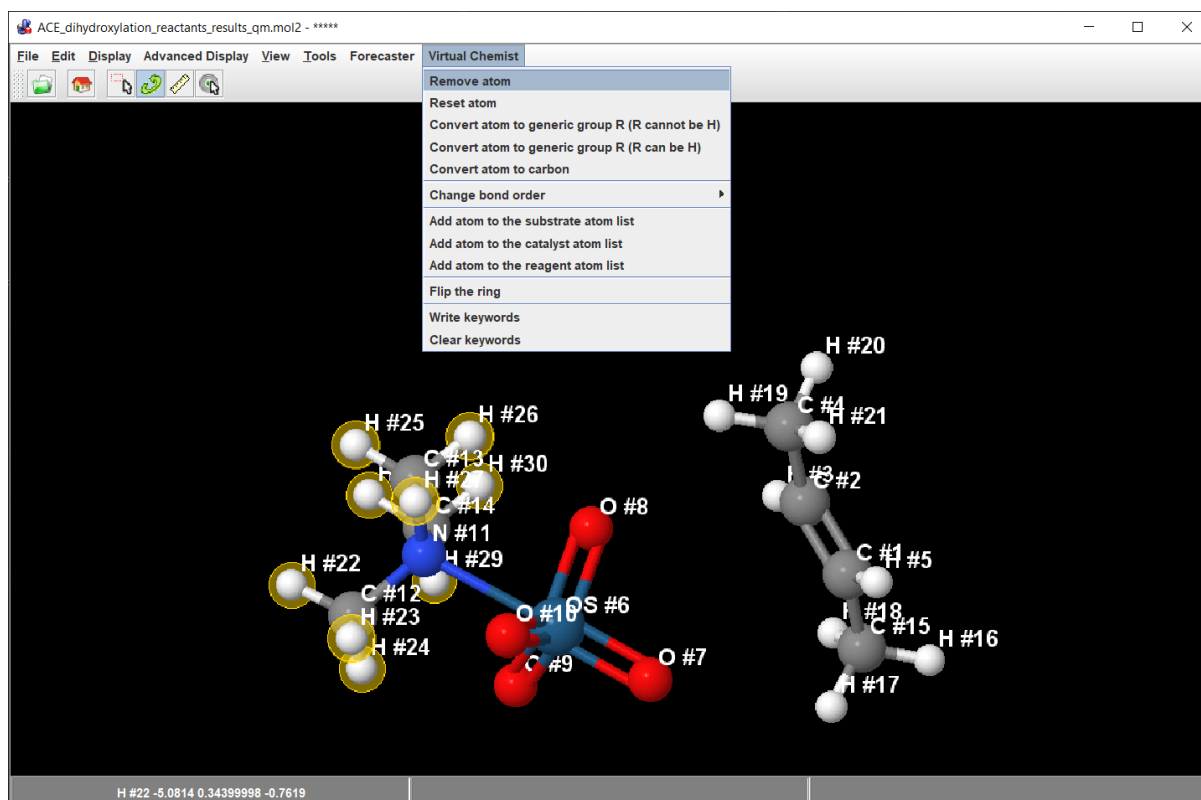
First exclude FINDERS, REACT2D and ACE.



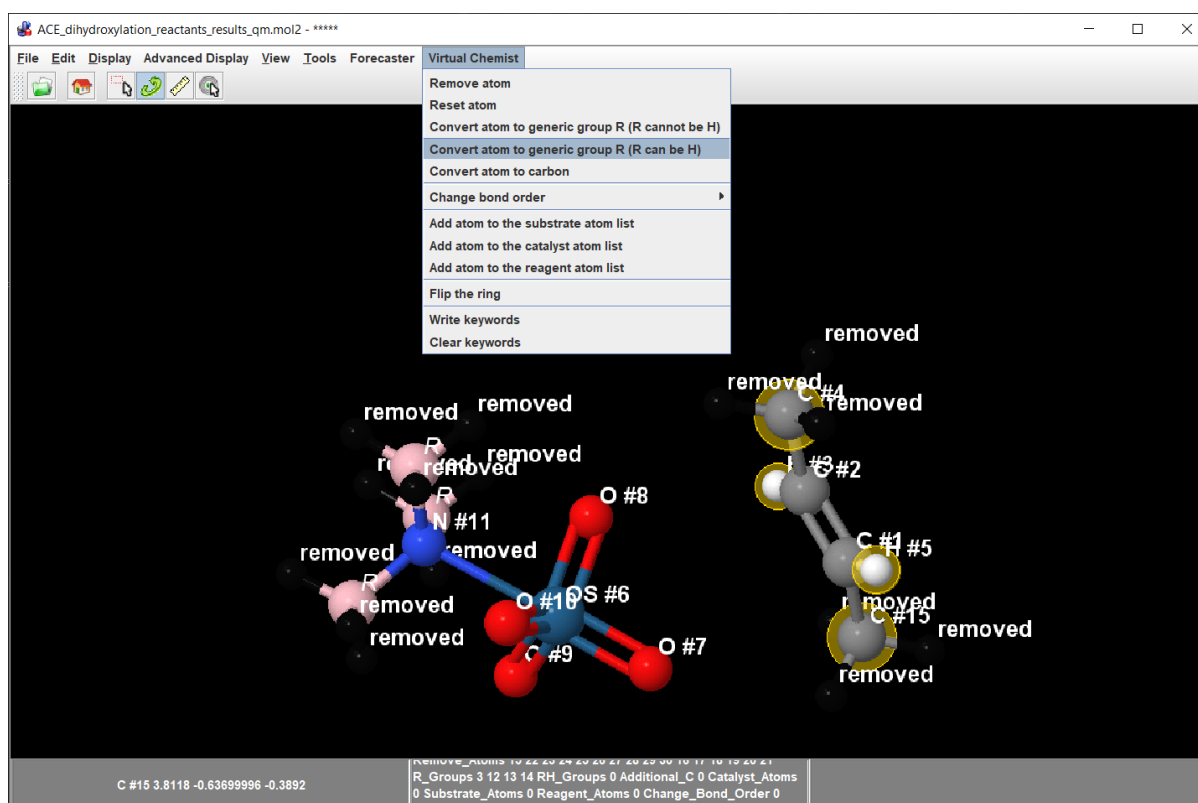
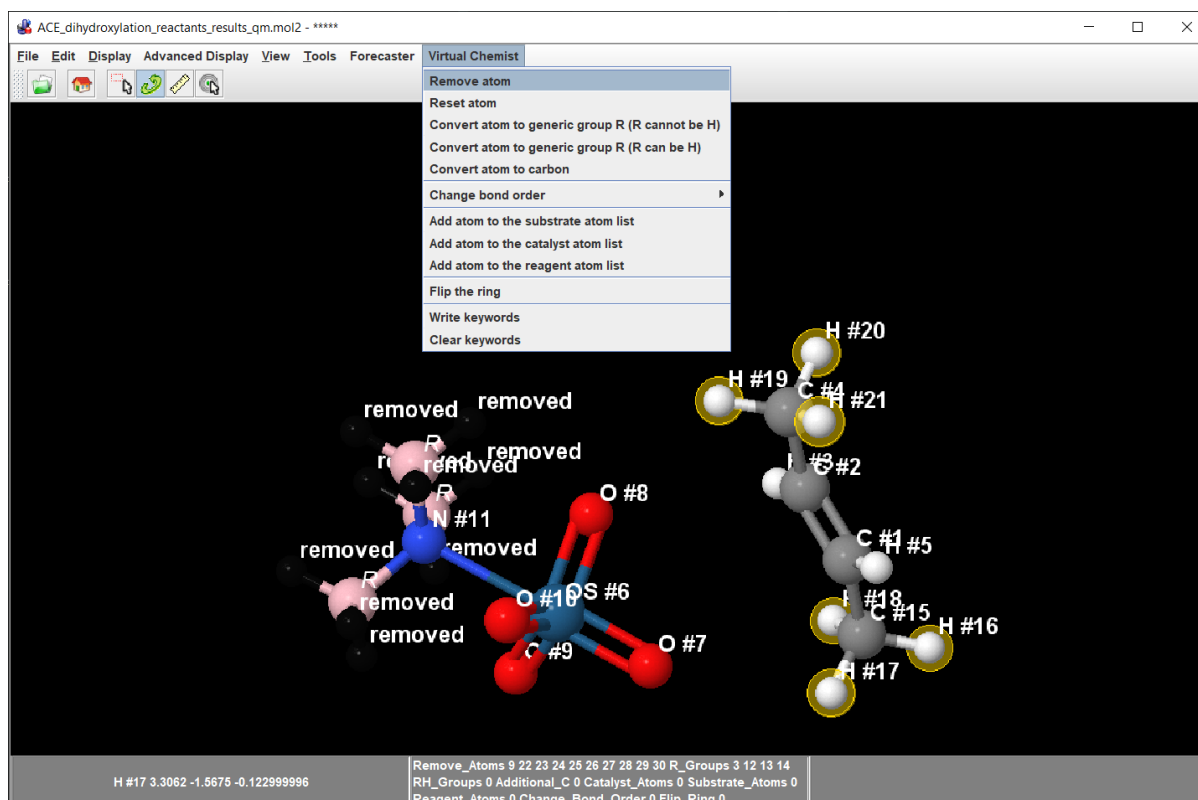
Click on the cog of the CONSTRUCTS box to open the CONSTRUCTS menu.



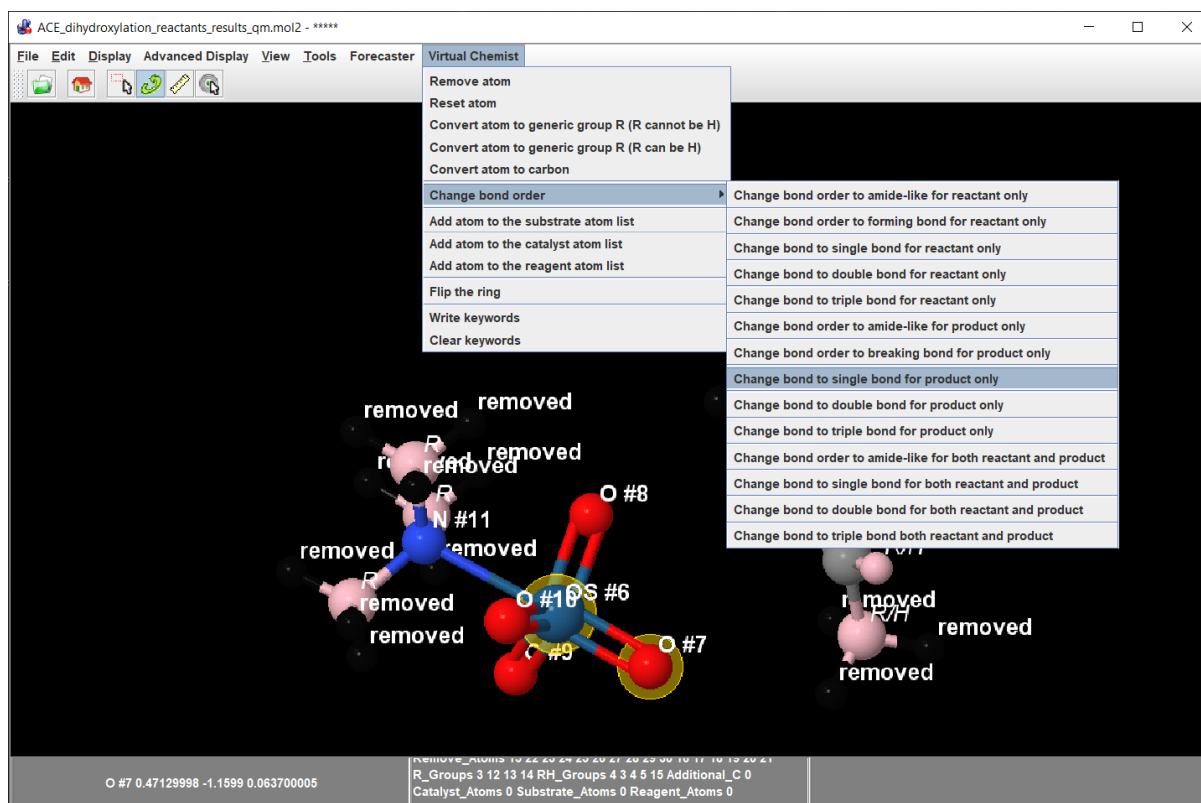
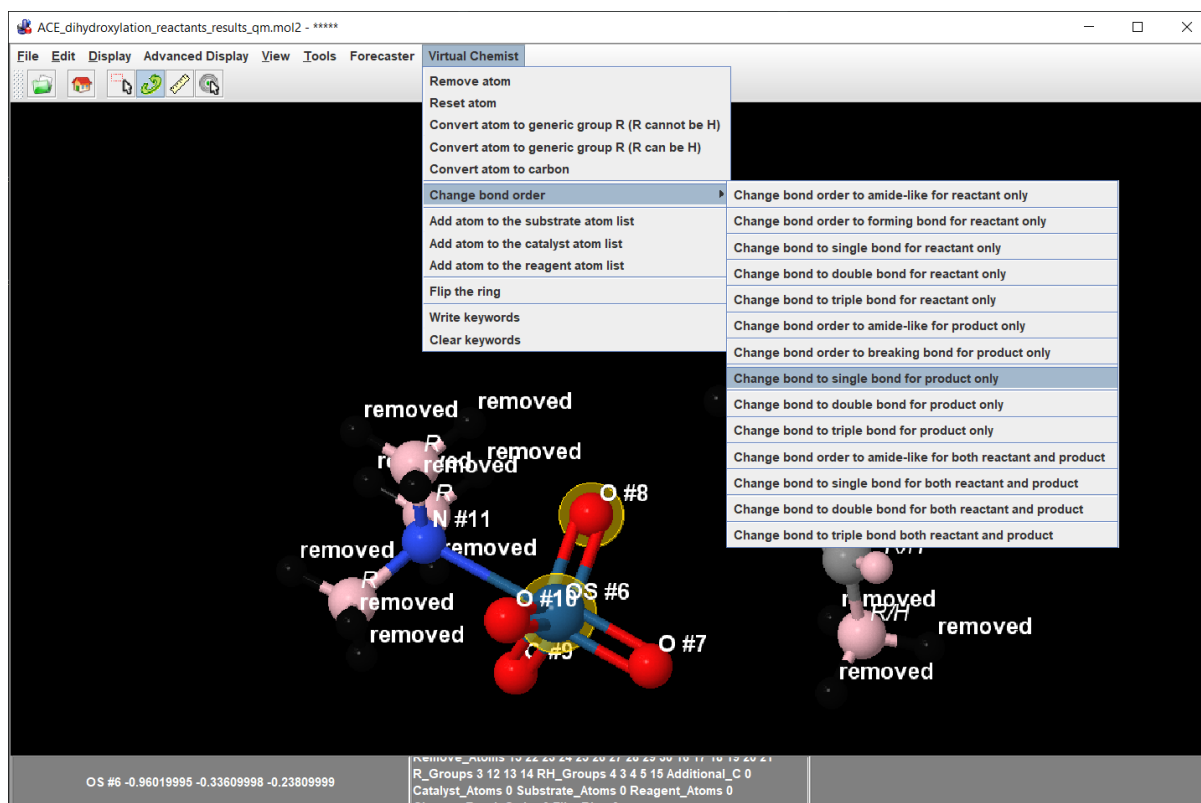
First, we convert NMe_3 used into NR_3 by removing the hydrogens and converting the carbon atoms into generic R groups.



Then the substrate $\text{H}_3\text{C}-\text{CH}=\text{CH}-\text{CH}_3$ is converted into $\text{R}_2-\text{C}=\text{C}=\text{R}_2$.



We next modify bond orders (the ones currently shown are only correct for the reactant). In the product, two of the Os=O bonds will become Os-O.



In the course of the reaction, two O-C will be created, and the double bond of the alkene will be converted into a single bond. We need to instruct CONSTRUCTS of these changes.

The parameters will appear in the “Parameter for template 1” box. As there is no chiral atom in the generic catalyst in this template, no need to look at enantiomers.

Parameters for template 1	
	Remove_Atoms 15 22 23 24 25 26 27 28 29 30 16 17 18 19 20 21 R_Groups 3 12 13 14 RH_Groups 4 3 4 5 15 Additional_C 0 Catalyst_Atoms 4 11 12 13 14 Substrate_Atoms 6 1 2 3 4 5 15 Reagent_Atoms 0 Change_Bond_Order 5 product 1 6 8 p product 1 6 7 product 1 2 8 product 1 1 7 product 1 1 2 Flip_Ring 0
	Only this enantiomer ▼
Catalyst residue name	CAT1
Substrate residue name	REA1
Reagent residue name	REA2
Output File Name	constructs_output
ACE to be ran on multiple cores	Multicore or cluster ▼
Number of cores	200
Number of TS configurations	4

You may now save the parameters.

Running CONSTRUCTS	
Edit Parameter File	
Save	parameters-constructs.txt

The template file should like the one below:

```

#
#   Creating user name:      SMART (Forecaster v. 5620)
#   Creation time:          Fri Apr 12 08:32:37 2019

@<TRIPOS>MOLECULE
template1_1
  15   15 0 0 0
SMALL
USER_CHARGES

@<TRIPOS>ATOM
  1 C      2.6005    0.0595    0.2694 C.3    1 REA1    0.0000 ***** 1
  2 C      2.0615    1.3789    0.1557 C.3    1 REA1    0.0000 ***** 2
  3 R      2.2439    2.0348   -1.2033 R/H    1 REA1    0.0000 ***** 3
  4 R      1.9749    2.2498    1.3984 R/H    1 REA1    0.0000 ***** 4
  5 R      3.3837   -0.1766    1.5505 R/H    1 REA1    0.0000 ***** 5
  6 OS     -1.0346   -0.2557   -0.2151 Os    1 MOL     0.0000 ***** 6
  7 O       0.5280   -1.0824    0.0904 O.3    1 MOL     0.0000 ***** 7
  8 O      -0.2587    1.3336   -0.4991 O.3    1 MOL     0.0000 ***** 8
  9 O      -1.4679   -1.1131   -1.6065 O.3    1 MOL     0.0000 ***** 9
 10 O      -1.7818   -0.4295    1.2939 O.3    1 MOL     0.0000 ***** 10
 11 N      -3.0201    0.8183   -0.6786 N.3    1 CAT1    0.0000 ***** 11
 12 R      -4.1353   -0.2082   -0.5645 R/R    1 CAT1    0.0000 ***** 12
 13 R      -3.2170    1.9424    0.3254 R/R    1 CAT1    0.0000 ***** 13
 14 R      -2.9521    1.3657   -2.0345 R/R    1 CAT1    0.0000 ***** 14
 15 R       2.9638   -0.7085   -0.9910 R/H    1 REA1    0.0000 ***** 15

@<TRIPOS>BOND
  1    11    14    1    1
  2     6    11    1    1
  3     9     6    2    2
  4    12    11    1    1
  5    13    11    1    1
  6     8     6    2    1
  7     7     6    2    1
  8    10     6    2    2
  9     2     8    0    1
 10     3     2    1    1
 11     1     2    2    1
 12    15     1    1    1
 13     4     2    1    1
 14     5     1    1    1
 15     1     7    0    1

@<VC>TS
0
0
1

```

After checking the template, we can now prepare the TS structures.

Settings

Source of catalyst structures	From working directory
Library of Catalysts	Dihydroxylation_catalysts.sdf
Library of Substrates	Dihydroxylation_substrates.sdf
Library of Reagents	none
Preset reactions	No
Transition state template	Build from Reactant and Prod...
Prepare TS templates	Prepare template and TS for A...
Ground states	1 reactant and 1 product
Ground state reactant file #1	ACE_dihydroxylation_product...
Ground state product file #1	ACE_dihydroxylation_product...

Load template structure

Parameters for template 1 Remove_Atoms 15 22 23 24 25 26 27

With these parameters set, we can now save the parameters and run the workflow.

<input checked="" type="checkbox"/> Included	CONSTR...		
<input type="checkbox"/> Excluded	ACE		

Run workflow

VI. REFERENCES

- (1) Seminario, J. M. Calculation of intramolecular force fields from second-derivative tensors. *Int. J. Quant. Chem.* **1996**, *60*, 1271-1277.
- (2) Allen, A. E. A.; Payne, M. C.; Cole, D. J. Harmonic Force Constants for Molecular Mechanics Force Fields via Hessian Matrix Projection. *J. Chem. Theory Comput.* **2018**, *14*, 274-281.
- (3) DelMonte, A. J.; Haller, J.; Houk, K. N.; Sharpless, K. B.; Singleton, D. A.; Strassner, T.; Thomas, A. A. Experimental and Theoretical Kinetic Isotope Effects for Asymmetric Dihydroxylation. Evidence Supporting a Rate-Limiting "(3 + 2)" Cycloaddition. *J. Am. Chem. Soc.* **1997**, *119*, 9907-9908
- (4) Rasmussen, T.; Norrby, P. O. Modeling the stereoselectivity of the beta-amino alcohol-promoted addition of dialkylzinc to aldehydes. *J. Am. Chem. Soc.* **2003**, *125*, 5130-5138