

Study of intermolecular recognition via direct estimative of the partition function

*

*Federal University of São Carlos, Departament of Chemistry, Rodovia Washington Luís km
235, CP 676, CEP 13565-905, São Carlos, SP, Brazil*

E-mail: moura@ufscar.br

THEMIS USER GUIDE

COMMAND LINE OPTIONS

Firstly, to display Themis help, one should use the command below

```
themis@linux:~$ themis --help
-----
                        Program THEMIS version beta
STARTED AT: 26/09/2019 - 09:20:23
COMMAND LINE READ: themis --help
-----
Usage:  themis [RUNTYPE] [GRIDTYPE] [RADIUS|FILENAME]
-----
                        [RUNTYPE] options
--run      Start new calculation.
--rerun    Read energies from previous run. User must
            give an energy.bin file (from Themis) or an
            energy.log file (from external programs).
-----
                        [GRIDTYPE] options
--shell    Translation moves will be performed on a
            spherical shell grid generated on the run.
--user     Translation moves will be performed on an
            user-defined external grid.
-----
                        [RADIUS] value
(real)     Scaling factor for the spherical grid
            radius (in Angstrom)
                        [FILENAME] value
(char)     XYZ file containing the user-defined
            translation grid. It must be aligned
            with molecule 1
-----
                        Other options
--help     Display this help
--version  Display the version
-----
themis@linux:~$
```

Also, to display Themis license information, one should use the following command

```
themis@linux:~$ themis --license
-----
                        Program THEMIS version beta
STARTED AT: 26/09/2019 - 10:29:47
COMMAND LINE READ: themis --license
-----
Copyright (C) 2019 Felipe Mariano Colombari
License GPLv3+:
GNU GPL version 3 or later
see <http://gnu.org/license/gpl.html>
This is a free software
you are free to change it and redistributibe it
There is NO WARRANTY, to the extent permitted by law
Written by Felipe M. Colombari
E-mail: colombarifm@hotmail.com
-----
themis@linux:~$
```

To run a new simulation using a spherical translation grid of radius 4.0 Å, one should enter the following command

```
themis@linux:~$ themis --run --shell 4.0 > resume.log &  
themis@linux:~$
```

On the other hand, the following command

```
themis@linux:~$ themis --run --user vdw.xyz > resume.log &  
themis@linux:~$
```

will run a new simulation using a translation grid corresponding to the VdW surface of molecule 1 saved on “vdw.xyz”.

INPUT FILES

conf1.xyz and **conf2.xyz**: Standard XYZ files containing the coordinates of both structures. For the water dimer mentioned in the previous sections, a dummy site (X) corresponding to water center of mass was used to define an orientation vector along Z-axis, as shown below.

themis@linux:~\$ cat conf1.xyz	themis@linux:~\$ cat conf2.xyz
4	4
blank line	*blank line*
OW 0.00000 0.06682 0.00000	OW 0.00000 0.06682 0.00000
HW -0.76677 -0.53032 0.00000	HW -0.76677 -0.53032 0.00000
HW 0.76677 -0.53032 0.00000	HW 0.76677 -0.53032 0.00000
X 0.00000 0.00000 0.00000	X 0.00000 0.00000 0.00000
themis@linux:~\$	themis@linux:~\$

parameters1 and **parameters2** are plain text files containing potential parameters used for energy calculations. Those files are read differently according to the potential used. For **potential : lj-coul** (eq. 1), TIP3P water parameters are given as follows:

$$U_{\text{ljc}} = U_{\text{lj}} + U_{\text{coul}} = \sum_i \sum_{j < i} 4\epsilon_{ij} \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right) + \frac{1}{4\pi\epsilon_0} \sum_i \sum_{j < i} \frac{q_i q_j}{r_{ij}} \quad (1)$$

where $\epsilon_{ij} = (\epsilon_i \cdot \epsilon_j)^{\frac{1}{2}}$ and $\sigma_{ij} = (\sigma_i \cdot \sigma_j)^{\frac{1}{2}}$

themis@linux:~\$ cat parameters1	themis@linux:~\$ cat parameters2
q sig (Å) eps (kJ/mol)	q sig (Å) eps (kJ/mol)
OW -0.834 3.15061 0.636386	OW -0.834 3.15061 0.636386
HW +0.417 0.00000 0.000000	HW +0.417 0.00000 0.000000
HW +0.417 0.00000 0.000000	HW +0.417 0.00000 0.000000
X 0.000 0.00000 0.000000	X 0.000 0.00000 0.000000
themis@linux:~\$	themis@linux:~\$

For **potential : bh-coul** (eq. 2), MATSUI parameters for a TiO₂ unit must be provided as follows:

$$U_{\text{bhc}} = U_{\text{bh}} + U_{\text{coul}} = \sum_i \sum_{j < i} \left\{ \left(\frac{-C_i C_j}{r_{ij}^6} \right) + f(B_i + B_j) \exp \left[\left(\frac{A_i + A_j - r_{ij}}{B_i + B_j} \right) \right] \right\} + \frac{1}{4\pi\epsilon_0} \sum_i \sum_{j < i} \frac{q_i q_j}{r_{ij}} \quad (2)$$

where the quantity f corresponds to a standard force of 4.184 kJ/mol/Å.

themis@linux:~\$ cat parameters1					themis@linux:~\$ cat parameters2				
	q	A(A)	B (A)	C(A ³ kJ/mol)		q	A(A)	B (A)	C(A ³ kJ/mol)
Ti	2.1960	1.18230	0.07700	22.5000	Ti	2.1960	1.18230	0.07700	22.5000
O	-1.0980	1.63390	0.11700	54.0000	O	-1.0980	1.63390	0.11700	54.0000
O	-1.0980	1.63390	0.11700	54.0000	O	-1.0980	1.63390	0.11700	54.0000
X	0.0000	0.00000	0.00000	0.00000	X	0.0000	0.00000	0.00000	0.00000
themis@linux:~\$					themis@linux:~\$				

NOTE: It is important to highlight that atoms described in both `parameters1` and `parameters2` files MUST BE IN THE SAME ORDER as they appear in both `conf1.xyz` and `conf2.xyz` files. Parameters file MUST contain a header followed by one line for each atom described in structure files.

INPUT file: plain text file containing detailed intructions prior to calculation.

```

themis@linux:~$ cat INPUT

reorientation_factor : 2          # integer
translation_factor : 2          # integer
gyration_factor : 36            # integer
gyration_range : 360.0         # real
temperature : 300.0            # real
potential : lj-coul             # character. valid strings are: none, lj-coul our bh-coul
ref_mol1 : 1                   # integer
rot_ref_mol1 : 4               # integer
ref_mol2 : 1                   # integer
rot_ref_mol2 : 2               # integer
shortest_distance : 0.8        # real
write_xtc : no                 # character. valid strings are: no, F, yes or T
lowest_structures : 20         # integer
write_frames : none            # character. valid strings are: none, MOP or XYZ
mopac_job :                    # character

themis@linux:~$

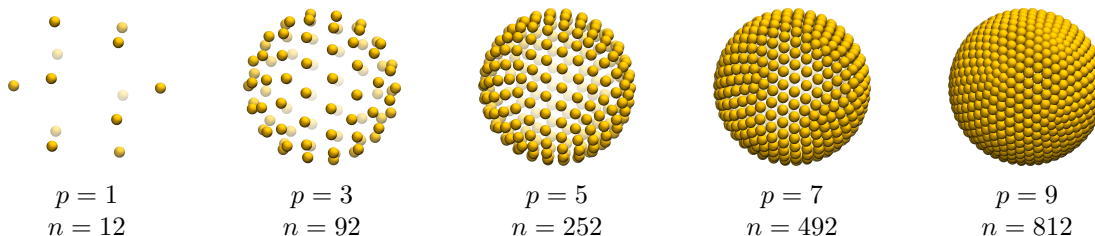
```

reorientation_factor: Parameter (p) used to generate the spherical grid used for reorientation moves.

The number of points (n) obtained along the sphere surface by dodecahedron tessellation is given by

$$n = 12 + 10 \times 3 \times (p - 1) + 10 \times (p - 2) \times (p - 1)$$

If one uses $p = 0$, the reorientation move will correspond to align molecule 2 along Z-axis (1 reorientational move).



translation_factor: Same as **reorientation_grid** if a spherical translation shell is used.

gyration_factor: Corresponds to the number of gyration points around the rotation axis.

gyration_range: Corresponds to the maximum rotation angle (in degrees).

temperature: Absolute temperature (K).

potential: Potential energy function selection. Options are “none”, “lj-coul”, “bh-coul”.

write_frames: Selects the format in which all valid frames will be written: “XYZ”, “MOP” and “none”.

If “MOP” is selected, the optional character variable containing the first line of MOPAC input is read.

ref_mol1: Site of molecule 1 used for centering, according to **conf1.xyz** file.

rot_ref_mol1: Site of molecule 1 that will build its rotation vector, according to **conf1.xyz** file.

ref_mol2: Site of molecule 2 used for centering, according to `conf2.xyz` file.

rot_ref_mol2: Site of molecule 2 that will build its rotation vector, according to `conf2.xyz` file.

shortest_distance: Corresponds to the lowest intermolecular distance to consider the configuration as a valid one. Below such value (in Angstrom), molecular contacts are considered strongly repulsive and an interaction energy value of 10^{10} kJ/mol is attributed to such configuration. This is useful to avoid spending time calculating energies for unphysical configurations since the energy loop is skipped.

write_xtc: Flag to enable the writting of all configurations to a XTC file. WARNING: very large files can be generated ;)

lowest_structures: selects the number of lowest energy/highest probability structures to write after the run.

mopac_job: string containing the header for mopac calculations. enabled when “write_frames : MOP” is selected.

OUTPUT FILES

energy.bin

Binary file containing interaction energy values for all microstates. Since all entries are written in the right loop sequence, they can be read using the `rerun` feature.

energy-sort.log

Contains interaction energy values and probabilities for the N most probable structures. By running Themis with the input files presented below, and considering a spherical grid with radius = 2.8 Å, one obtains

```
themis@linux:~$ cat energy-sort.log
```

inter_energy(g,r,t)	g	r	t	prob.	sum prob.
-2.83500E+001	1	10	3	6.704E-004	6.704E-004
-2.83500E+001	1	4	9	6.704E-004	1.341E-003
-2.83453E+001	2	4	9	6.692E-004	2.010E-003
-2.83453E+001	2	10	3	6.692E-004	2.679E-003
-2.83453E+001	120	4	9	6.692E-004	3.348E-003
-2.83453E+001	120	10	3	6.692E-004	4.017E-003
-2.83312E+001	3	10	3	6.654E-004	4.683E-003
-2.83312E+001	119	10	3	6.654E-004	5.348E-003
-2.83312E+001	3	4	9	6.654E-004	6.014E-003
-2.83312E+001	119	4	9	6.654E-004	6.679E-003
-2.83078E+001	118	10	3	6.592E-004	7.338E-003
-2.83078E+001	118	4	9	6.592E-004	7.997E-003
-2.83078E+001	4	10	3	6.592E-004	8.657E-003
-2.83078E+001	4	4	9	6.592E-004	9.316E-003
-2.82751E+001	117	4	9	6.506E-004	9.966E-003
-2.82751E+001	117	10	3	6.506E-004	1.062E-002
-2.82751E+001	5	10	3	6.506E-004	1.127E-002
-2.82751E+001	5	4	9	6.506E-004	1.192E-002
-2.82333E+001	116	10	3	6.398E-004	1.256E-002
-2.82333E+001	6	10	3	6.398E-004	1.320E-002

```
themis@linux:~$
```

output.log

Contains thermodynamic data for all translation grid points, and also for the overall ensemble. Written in an extended XYZ format containing extra field values for each grid point (probability, free energy, energy and entropic penalty).

```

themis@linux:~$ cat output.log

```

	42								
	X (Å)	Y (Å)	Z (Å)	point	PROB	A (kJ/mol)	-TS (kJ/mol)	E (kJ/mol)	
X	2.38182	1.47205	0.00000	1	2.98845E-003	-1.08134E+001	6.75915E+000	-1.75725E+001	
X	2.38182	-1.47205	0.00000	2	2.98845E-003	-1.08134E+001	6.75915E+000	-1.75725E+001	
X	1.47205	0.00000	2.38182	3	5.85172E-002	-1.82330E+001	6.84167E+000	-2.50746E+001	
X	1.47205	0.00000	-2.38182	4	1.34004E-003	-8.81280E+000	4.19575E+000	-1.30086E+001	
X	0.00000	2.38182	1.47205	5	1.00969E-002	-1.38502E+001	6.78551E+000	-2.06357E+001	
X	0.00000	2.38182	-1.47205	6	1.74725E-001	-2.09615E+001	4.32032E+000	-2.52818E+001	
X	0.00000	-2.38182	1.47205	7	1.00969E-002	-1.38502E+001	6.78551E+000	-2.06357E+001	
X	0.00000	-2.38182	-1.47205	8	1.74725E-001	-2.09615E+001	4.32032E+000	-2.52818E+001	
...									
X	-2.26525	0.86525	-1.40000	40	9.83462E-004	-8.04111E+000	5.57165E+000	-1.36128E+001	
X	-2.26525	-0.86525	-1.40000	41	9.83462E-004	-8.04111E+000	5.57165E+000	-1.36128E+001	
X	-2.80000	0.00000	0.00000	42	2.97201E-003	-1.07996E+001	6.97190E+000	-1.77715E+001	

TOTAL OVER TRANSLATIONAL GRID					1.00000E+000	-1.59900E+001	7.67496E+000	-2.36649E+001	

```

themis@linux:~$

```

output-sort.log

Same as output.log but ordered from most probable point to the least probable point.

```

themis@linux:~$ cat output-sort.log

```

	42								
	X (Å)	Y (Å)	Z (Å)	point	PROB	A (kJ/mol)	-TS (kJ/mol)	E (kJ/mol)	
X	0.00000	-2.38182	-1.47205	8	1.74725E-001	-2.09615E+001	4.32032E+000	-2.52818E+001	
X	0.00000	2.38182	-1.47205	6	1.74725E-001	-2.09615E+001	4.32032E+000	-2.52818E+001	
X	0.00000	0.00000	2.80000	24	8.68520E-002	-1.92179E+001	6.43659E+000	-2.56545E+001	
X	1.47205	0.00000	2.38182	3	5.85172E-002	-1.82330E+001	6.84167E+000	-2.50746E+001	
X	-1.47205	0.00000	2.38182	9	5.85172E-002	-1.82330E+001	6.84167E+000	-2.50746E+001	
X	0.86525	1.40000	2.26525	22	4.04675E-002	-1.73130E+001	7.12969E+000	-2.44427E+001	
X	-0.86525	1.40000	2.26525	29	4.04675E-002	-1.73130E+001	7.12969E+000	-2.44427E+001	
X	0.86525	-1.40000	2.26525	23	4.04675E-002	-1.73130E+001	7.12969E+000	-2.44427E+001	
...									
X	-2.26525	0.86525	-1.40000	40	9.83462E-004	-8.04111E+000	5.57165E+000	-1.36128E+001	
X	2.26525	-0.86525	-1.40000	19	9.83462E-004	-8.04111E+000	5.57165E+000	-1.36128E+001	
X	-2.26525	-0.86525	-1.40000	41	9.83462E-004	-8.04111E+000	5.57165E+000	-1.36128E+001	

TOTAL OVER TRANSLATIONAL GRID					1.00000E+000	-1.59900E+001	7.67496E+000	-2.36649E+001	

```

themis@linux:~$

```

surf_free-energy.vmd, surf_energy.vmd, surf_entropic-penalty.vmd

Contains a VMD script for reading the thermodynamic data along the translation grid from file output.log.

lowest_0001.xyz

XYZ coordinates for the most probable structure from the whole ensemble. The number of lowest structure files is defined by the user in the INPUT file.

```
themis@linux:~$ cat lowest_0001.xyz
```

```

      8
Energy = -2.8350000E+01
O      0.0000    0.0000    0.0000
H     -0.0000   -0.7668   -0.5971
H     -0.0000    0.7668   -0.5971
X     -0.0000   -0.0000   -0.0668
O      1.4720    0.0000    2.3818
H      0.9611    0.0000    1.5551
H      0.7957    0.0000    3.0797
X      1.4056    0.0000    2.3746

```

```
themis@linux:~$
```

grid_log.log

File containing informations of each translation grid point: point number, number of rejected structures, spent time.

```
themis@linux:~$ cat grid_log.log
```

t	point	rejected	structures	time (s)
	1	0 of	5040	0.020
	2	0 of	5040	0.012
	3	0 of	5040	0.012
	4	0 of	5040	0.013
...				
	38	0 of	5040	0.024
	39	0 of	5040	0.013
	40	0 of	5040	0.013
	41	0 of	5040	0.012
	42	0 of	5040	0.014

```
themis@linux:~$
```

full_ensemble.xtc

XTC trajectory file containing the whole ensemble. Written if INPUT option `write_xtc` is enabled.

point_0001_0001_0001.xyz

XYZ file containing the structure of the microstate $t = 1$, $r = 1$, $g = 1$. Files are numbered according to the loop position. Written if INPUT option `write_frames = XYZ` is set. Microstates with intermolecular distances below the one defined by `shortest_distance` are skipped. WARNING: this option will create a very large number of files in the directory. ;)

```

themis@linux:~$ point_0001_0001_0001.xyz

      8
Energy = 0.0000000E+00
O      0.0000    0.0000    0.0000
H     -0.0000   -0.7668   -0.5971
H     -0.0000    0.7668   -0.5971
X     -0.0000   -0.0000   -0.0668
O      2.3818    1.4720    0.0000
H      3.2085    1.9830   -0.0000
H      2.1793    1.3469    0.9423
X      2.4167    1.4936    0.0527

themis@linux:~$

```

point_0001_0001_0001.mop

Same as before, but in MOPAC format, containing the header defined by `mopac_job`. Written if INPUT options `write_frames = MOP` is set. Microstates with intermolecular distances below the one defined by `shortest_distance` are skipped. WARNING: this option will create a very large number of files in the directory. ;)

```

themis@linux:~$ point_0001_0001_0001.mop

PM7 1SCF CHARGE=0 THREADS=1 OUTPUT
*blank line*
*blank line*
O      0.0000  0  0.0000  0  0.0000  0
H     -0.0000  1 -0.7668  1 -0.5971  1
H     -0.0000  1  0.7668  1 -0.5971  1
X     -0.0000  0 -0.0000  0 -0.0668  0
O      2.3818  0  1.4720  0  0.0000  0
H      3.2085  0  1.9830  0 -0.0000  0
H      2.1793  1  1.3469  1  0.9423  1
X      2.4167  1  1.4936  1  0.0527  1

themis@linux:~$

```
