

SMOG-SERVER.ORG

**THE SMOG
USER GUIDE**

EXCERPT ON USING WHAM.jar

Jeff Noel

info@smog-server.org

[HTTP://SMOG-SERVER.ORG/SB_ANALYSIS.HTML](http://smog-server.org/SB_ANALYSIS.HTML)

AUGUST 25, 2012

Contents

1	Weighted Histogram Analysis - WHAM.jar	2
1.1	Introduction	2
1.2	Downloading WHAM.jar	2
1.3	Citing WHAM.jar	3
1.4	Running WHAM.jar	3
1.5	Configuration file	3
1.5.1	Basic configuration file: compute density of states (Ω), i.e. run WHAM	3
1.5.2	Full configuration parameter list	4
1.6	Compute $F(Q), S(Q), E(Q)$ at T from $\Omega(E, Q)$ or $\Omega(E, Q_1, Q_2)$	4
1.7	Compute $C_V(T), E(T), S(T), F(T)$ from $\Omega(E, \dots)$	5
1.8	Compute $\langle Q(T) \rangle$ from $\Omega(E, Q)$	5
1.9	Compute $\langle Q_2(Q_1) \rangle$ from $\Omega(E, Q_1, Q_2)$	6
1.10	Compute $\langle Q_3(Q_1, [Q_2^{\text{low}}, Q_2^{\text{high}}]) \rangle$ from $\Omega(E, Q_1, Q_2, Q_3)$	6
1.11	Reweighting of umbrella sampling	7
1.11.1	Harmonic umbrella	7

1 Weighted Histogram Analysis - WHAM.jar

1.1 Introduction

This section describes a Java application `WHAM.jar` which at its heart uses a well established algorithm described here [1]. `WHAM.jar` is tailored for the sort of analysis that is most often performed on SBMs, including free energy perturbation, thermal averaging of reaction coordinates and free energy as a function of (1 or 2) reaction coordinates. The basic operation involves providing `WHAM.jar` with a set of histograms from constant temperature simulation with various temperatures and umbrella parameters, and a configuration file. Based on the options set in the configuration file (or at the command line) `WHAM.jar` will perform the appropriate analysis. The WHAM algorithm itself [1] provides an optimal density of states (Ω). From the Ω many thermodynamic quantity of interest can be calculated.

1.2 Downloading WHAM.jar

In the near future `WHAM.jar` will be available at `smog-server.org`. There are no plans to make the source code publicly available. Any feature requests or source code requests should be directed to `info@smog-server.org`.

1.3 Citing WHAM.jar

WHAM.jar is part of the SMOG group of tools, so please cite [2].

1.4 Running WHAM.jar

Like any java application, no compilation is necessary, but a virtual machine is required; WHAM.jar requires JRE 1.6.0_29 or greater. At the command line, the basic syntax is

```
user$ java [-Xmx1000m] -jar WHAM.jar --config configurationFile [ argument... ]
```

-Xmx1000m assigns 1000 MB of RAM to the Java virtual machine heap. With large histograms (4 or 5 dimensions) the default heap allocation can run out which gives the following error:

```
java.lang.OutOfMemoryError: Java heap space
```

1.5 Configuration file

1.5.1 Basic configuration file: compute density of states (Ω), i.e. run WHAM

```
numDimensions 2 #number of dimensions in the dos histogram
                #energy and Q, energy is always in the first column, can have up to 3 reaction coordinates

kB 0.008314     #Boltzmann constant

run_wham       #creates optimal density of states, writes to dosFile, comment out to not run

dosFile dos     #density of states output filename for run_wham, input filename for analysis (i.e. run_cv)

threads 1       #on multicore systems you can use up to 8 threads to speed up the calculation

### energy binning ###
numBins 142
start -88
step 2

### reaction coordinate 1 binning ###
numBins 50
start 0
step 2

### reaction coordinate 2 binning, change numDimensions to 3 ###
## numBins 50
## start 0
## step 2

### list of histogram filenames and their temperatures ###
numFiles 1
name data/hist144 temp 144
name data/hist145 temp 145
name data/hist146 temp 146
```

```
name data/hist147 temp 147
```

- The Boltzmann constant (`kb`) that relates the energy units used in the histogram files to the temperature units given with `temp`. The value shown here is appropriate for Gromacs and SMOG, where the temperature is the Gromacs temperature in Kelvin.
- The total considered range for the energy binning is [`start` , `start + numBins*step`]. Adjust `step` to balance detail (finer `step`) with noise reduction (larger `step`). Also, a larger `step` leaves a smaller memory footprint, and runs faster. So large `step` is useful for preliminary tests.
- The histogram files are designated with relative paths with respect to the directory where `WHAM.jar` is launched. Histogram files are space delimited columns with the following format:

```
energy reaction_coord_1 [ more_reaction_coords ] [ umbrella_coords ]
```

```
-23.779612 117
-37.528976 124
-52.246330 123
-43.539364 132
-41.074669 129
-34.133888 130
-25.081692 125
6.063239 117
51.440716 102
21.739162 114
4.184849 112
37.821884 117
-10.340809 121
```

This file (`hist144`) has an energy column and one reaction coordinate column. The configuration file designates it was run at a temperature of 144.

1.5.2 Full configuration parameter list

The following will give a full list of configuration options:

```
user$ java -jar WHAM.jar -help
```

1.6 Compute $F(Q)$, $S(Q)$, $E(Q)$ at T from $\Omega(E, Q)$ or $\Omega(E, Q_1, Q_2)$

Add the following to compute $F(Q)$ from $\Omega(E, Q)$. You can substitute Q for your favorite reaction coordinate, and it is always given in the second column

```

run_free          #comment out to not run, reads dosFile

run_free_out free #prefix for the free energy curves

startTF 135      #first temperature to compute F(Q)

deltaTF 0.1      #step in temperature

ntempsF 200      #total F(Q) to generate

```

- There will be `ntempsF` files with names like `run_free_outtemp`. Output files have four columns in the following format:

```

reaction_coord_value free_energy enthalpy entropy

```

- If two reaction coordinates are given $\Omega(E, Q_1, Q_2)$, `run_free` assumes you want two dimensional analysis $F(Q_1, Q_2)$.

```

reaction_coord_1_value reaction_coord_2_value free_energy enthalpy entropy

```

1.7 Compute $C_V(T)$, $E(T)$, $S(T)$, $F(T)$ from $\Omega(E, \dots)$

Add the following to compute the heat capacity $C_V(T)$ (also the temperature dependence of energy, entropy, or free energy) from $\Omega(E, \dots)$.

```

run_cv           #comment out to not run, reads dosFile

run_cv_out cv    #filename for the temperature curves

startT 135      #starting temperature

deltaT 0.1      #step in temperature

ntemps 200      #total temps to generate

```

- There will be one output file with five columns in the following format:

```

temperature Cv enthalpy free_energy entropy

```

1.8 Compute $\langle Q(T) \rangle$ from $\Omega(E, Q)$

Add the following to compute the thermal expectation of a reaction coordinate Q_1 with respect to reaction coordinate Q_2 .

```

run_coord        #comment out to not run, reads dosFile

run_coord_out coord #filename for the coordinate curve

```

```

startTC 110      #temperature to start at
deltaTC 0.1      #step in temperature
ntempsC 200     #total temps to generate

```

- There will be one output file with two columns in the following format:

```
temperature <Q>
```

1.9 Compute $\langle Q_2(Q_1) \rangle$ from $\Omega(E, Q_1, Q_2)$

Add the following to compute the thermal expectation of a reaction coordinate Q_2 with respect to reaction coordinate Q_1 .

```

run_coord        #comment out to not run, reads dosFile
run_coord_out coord #filename prefix for the coordinate curve
startTC 110      #temperature to use

```

- There will be one output file `run_coord_outstartTC` with two columns in the following format:

```
temperature <Q2(Q1)>
```

1.10 Compute $\langle Q_3(Q_1, [Q_2^{\text{low}}, Q_2^{\text{high}}]) \rangle$ from $\Omega(E, Q_1, Q_2, Q_3)$

Add the following to compute the thermal expectation of a reaction coordinate Q_3 with respect to reaction coordinate Q_1 and averaged over a range of Q_2 .

```

run_coord        #comment out to not run, reads dosFile
run_coord_out coord #filename prefix for the coordinate curve
startTC 110      #temperature to use
lowQ2 38         #start of range for averaging third reaction coord
                  #note that internally it uses the binning, so bins (not values)
                  #greater than lowQ2 are considered
highQ2 42        #end of range for averaging third reaction coord
                  #note that internally it uses the binning, so bins (not values)
                  #less than highQ2 are considered

```

- There will be one output file `run_coord_outstartTC` with two columns in the following format:

```
temperature <Q3(Q1, [Q2])>
```

1.11 Reweighting of umbrella sampling

The umbrella coordinate is always given as the last columns in the input. `numDimensions` should be equal to the number of columns, which is `numUmbrella+1+number_of_reaction_coords`, where the +1 comes from the energy column. The energies (first column) are assumed to be without the umbrella included.

1.11.1 Harmonic umbrella

The potential for is $V = k(Q - Q_0)^2 \rightarrow \Delta E = \text{umbrella_k} * (\text{column_3} - \text{umbrella_0})^2$

```
numDimensions 3      #energy reaction_coord umbrella_coord
numUmbrella 1       #number of umbrella_coord
umbrellaType harmonic #harmonic umbrellas
.
.
.
### list of histogram filenames, temperatures, umbrella stiffness and umbrella centers ###
numFiles 1
name data/hist144.40 temp 144 umbrella_k 0.5 umbrella_0 40
name data/hist144.80 temp 144 umbrella_k 0.5 umbrella_0 80
name data/hist146.40 temp 146 umbrella_k 0.5 umbrella_0 40
name data/hist146.80 temp 146 umbrella_k 0.5 umbrella_0 80
```

- Important! The energy column is the energy without the umbrella perturbation included.

References

- [1] Kumar, S, Rosenberg, J, Bouzida, D, & Swendsen. (1992) The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method. *J. Comput. Chem.* **13**, 1011.
- [2] Noel, J. K, Whitford, P. C, Sanbonmatsu, K. Y, & Onuchic, J. N. (2010) Smog@ctbp: simplified deployment of structure-based models in gromacs. *Nucleic Acids Res.* **38**, W657–61.