# Analyzing Tools

## Analyzing Single-component 'histogram\_complexes\_time.dat' File

**General histogram function:**

hist (FileName, FileNum, InitialTime, FinalTime, SpeciesName, SaveFig)

Description: This function enables users to obtain a histogram during a certain time period of selected species.

Parameters:

* FileName

Acceptable value: String (Must provide)

Description: It is the path to the ‘.dat’ file, which is usually named as ‘histogram\_complexes\_time.dat’, representing the histogram data to be analyzed.

Sample: FileName = ‘/Users/UserName/Documents/histogram\_complexes\_time.dat’

* FileNum

Acceptable value: Int (Must Provide)

Description: It is the number of the total input file. If multiple files are provided, their names should obey the naming rule listed below.

Sample: FileNum = 5

Note (naming rule): If single file is provided, the input file should be named as its original name (‘histogram\_complexes\_time.dat’); if multiple files are provided, the name of input file should also include serial number as ‘histogram\_complexes\_time\_X.dat’ where X = 1,2,3,4,5…

* InitialTime

Acceptable value: Float (Must provide)

Unit: s

Description: It is the initial time that users desire to examine. The acceptable range should not smaller than the starting time or exceed the ending time of simulation.

Sample: InitialTime = 0.5

* FinalTime

Acceptable value: Float (Must provide)

Unit: s

Description: It is the final time that users desire to examine. The acceptable range should not smaller than the value of InitialTime or exceed the ending time of simulation.

Sample: FinalTime = 1.0

* SpeciesName

Acceptable value: String (Must provide)

Description: It is the name of species that users want to examine, which should also be identical with the name written in the input (.inp and .mol) files.

Sample: SpeciesName = ‘clat’

* SaveFig

Acceptable value: Bool (Optional)

Default: False

Description: If True, the plot will be saved as a ‘.png’ file in the current directory; if False, the figure will not be saved.

**Function for plotting maximum number of monomers in single complex over time:**

max\_complex (FileName, FileNum, InitialTime, FinalTime, SpeciesName, SaveFig)

Description: This function enables users to obtain a plot indicating maximum number of monomers in single complex molecule during a certain time period.

Parameters:

* FileName

Acceptable value: String (Must provide)

Description: It is the path to the ‘.dat’ file, which is usually named as ‘histogram\_complexes\_time.dat’, representing the histogram data to be analyzed.

Sample: FileName = ‘/Users/UserName/Documents/histogram\_complexes\_time.dat’

* FileNum

Acceptable value: Int (Must Provide)

Description: It is the number of the total input file. If multiple files are provided, their names should obey the naming rule listed below.

Sample: FileNum = 5

Note (naming rule): If single file is provided, the input file should be named as its original name (‘histogram\_complexes\_time.dat’); if multiple files are provided, the name of input file should also include serial number as ‘histogram\_complexes\_time\_X.dat’ where X = 1,2,3,4,5…

* InitialTime

Acceptable value: Float (Must provide)

Unit: s

Description: It is the initial time that users desire to examine. The acceptable range should not smaller than the starting time or exceed the ending time of simulation.

Sample: InitialTime = 0.5

* FinalTime

Acceptable value: Float (Must provide)

Unit: s

Description: It is the final time that users desire to examine. The acceptable range should not smaller than the value of InitialTime or exceed the ending time of simulation.

Sample: FinalTime = 1.0

* SpeciesName

Acceptable value: String (Must provide)

Description: It is the name of species that users want to examine, which should also be identical with the name written in the input (.inp and .mol) files.

Sample: SpeciesName = ‘clat’

* SaveFig

Acceptable value: Bool (Optional)

Default: False

Description: If True, the plot will be saved as a ‘.png’ file in the current directory; if False, the figure will not be saved.

**Function for plotting mean number of monomers in single complex over time:**

mean\_complex (FileName, FileNum, InitialTime, FinalTime, SpeciesName, ExcludeSize, SaveFig)

Description: This function enables users to obtain a plot indicating mean number of monomers in single complex molecule during a certain time period.

Parameters:

* FileName

Acceptable value: String (Must provide)

Description: It is the path to the ‘.dat’ file, which is usually named as ‘histogram\_complexes\_time.dat’, representing the histogram data to be analyzed.

Sample: FileName = ‘/Users/UserName/Documents/histogram\_complexes\_time.dat’

* FileNum

Acceptable value: Int (Must Provide)

Description: It is the number of the total input file. If multiple files are provided, their names should obey the naming rule listed below.

Sample: FileNum = 5

Note (naming rule): If single file is provided, the input file should be named as its original name (‘histogram\_complexes\_time.dat’); if multiple files are provided, the name of input file should also include serial number as ‘histogram\_complexes\_time\_X.dat’ where X = 1,2,3,4,5…

* InitialTime

Acceptable value: Float (Must provide)

Unit: s

Description: It is the initial time that users desire to examine. The acceptable range should not smaller than the starting time or exceed the ending time of simulation.

Sample: InitialTime = 0.5

* FinalTime

Acceptable value: Float (Must provide)

Unit: s

Description: It is the final time that users desire to examine. The acceptable range should not smaller than the value of InitialTime or exceed the ending time of simulation.

Sample: FinalTime = 1.0

* SpeciesName

Acceptable value: String (Must provide)

Description: It is the name of species that users want to examine, which should also be identical with the name written in the input (.inp and .mol) files.

Sample: SpeciesName = ‘clat’

* ExcludeSize

Acceptable value: Int (Optional)

Default: 0

Description: In the generated plot, the number of monomers in the complex that are no larger than this number will be excluded and will not be considered into the average calculation.

Sample: ExcludeNum = 1 (which means the monomer will be excluded)

* SaveFig

Acceptable value: Bool (Optional)

Default: False

Description: If True, the plot will be saved as a ‘.png’ file in the current directory; if False, the figure will not be saved.

**Function for converting single-species ‘histogram\_complexes\_time.dat’ file to a pandas.DataFrame:**

single\_hist\_to\_df (FileName, SaveCsv)

Description: This function enables users to convert the raw .dat file to a data frame in python pandas package for single-species system. Each column in the data frame includes the simulation time and selected occurrences of species during the simulation; each row is separated by a different simulation time.

Parameters:

* FileName

Acceptable value: String (Must provide)

Description: It is the path to the ‘.dat’ file, which is usually named as ‘histogram\_complexes\_time.dat’, representing the histogram data to be analyzed.

Sample: FileName = ‘/Users/UserName/Documents/histogram\_complexes\_time.dat’

* SaveCsv

Acceptable value: Bool (Optional)

Default: True

Description: If True, the corresponding .csv file will also be saved as ‘histogram.csv’ under the same directory; if False, the generated .csv file will be deleted.

**Function for converting single-species ‘histogram\_complexes\_time.dat’ file to a .csv file:**

single\_hist\_to\_csv (FileName)

Description: This function enables users to convert the raw .dat file to a .csv file for single-species system. Each column in the data frame includes the simulation time and selected occurrences of species during the simulation; each row is separated by a different simulation time.

Parameters:

* FileName

Acceptable value: String (Must provide)

Description: It is the path to the ‘.dat’ file, which is usually named as ‘histogram\_complexes\_time.dat’, representing the histogram data to be analyzed.

Sample: FileName = ‘/Users/UserName/Documents/histogram\_complexes\_time.dat’

**Function for creating 3D histogram (histogram with time axis):**

hist\_3d\_time (FileName, InitialTime, FinalTime, SpeciesName, TimeBins)

Description: This function enables users to generate 3D histogram representing the number of monomers in single complex as simulation time develops. The x-axis is the number of monomers, y-axis is the averaged time and z-axis is the relative occurrence probabilities.

Parameters:

* FileName

Acceptable value: String (Must provide)

Description: It is the path to the ‘.dat’ file, which is usually named as ‘histogram\_complexes\_time.dat’, representing the histogram data to be analyzed.

Sample: FileName = ‘/Users/UserName/Documents/histogram\_complexes\_time.dat’

* InitialTime

Acceptable value: Float (Must provide)

Unit: s

Description: It is the initial time that users desire to examine. The acceptable range should not smaller than the starting time or exceed the ending time of simulation.

Sample: InitialTime = 0.5

* FinalTime

Acceptable value: Float (Must provide)

Unit: s

Description: It is the final time that users desire to examine. The acceptable range should not smaller than the value of InitialTime or exceed the ending time of simulation.

Sample: FinalTime = 1.0

* SpeciesName

Acceptable value: String (Must provide)

Description: It is the name of species that users want to examine, which should also be identical with the name written in the input (.inp and .mol) files.

Sample: SpeciesName = ‘clat’

* TimeBins

Acceptable value: Int (Must provide)

Description: It is the number of bins that users want to divide the selected time period into. The value should be positive integer.

**Function for creating 2D heatmap showing distribution of N-mers vs. time (low contrast option):**

hist\_time\_heatmap (FileName, InitialTime, FinalTime, SpeciesName, TimeBins, ShowNum, SaveFig)

Description: This function enables users to generate 2D histogram of numerical distribution of different N-mers vs. time. The x-axis is the distribution of number of monomers in single complex and y-axis is the time period. The color in each box indicates the number of corresponding N-mers when corresponding time period is reached.

Parameters:

* FileName

Acceptable value: String (Must provide)

Description: It is the path to the ‘.dat’ file, which is usually named as ‘histogram\_complexes\_time.dat’, representing the histogram data to be analyzed.

Sample: FileName = ‘/Users/UserName/Documents/histogram\_complexes\_time.dat’

* InitialTime

Acceptable value: Float (Must provide)

Unit: s

Description: It is the initial time that users desire to examine. The acceptable range should not smaller than the starting time or exceed the ending time of simulation.

Sample: InitialTime = 0.5

* FinalTime

Acceptable value: Float (Must provide)

Unit: s

Description: It is the final time that users desire to examine. The acceptable range should not smaller than the value of InitialTime or exceed the ending time of simulation.

Sample: FinalTime = 1.0

* SpeciesName

Acceptable value: String (Must provide)

Description: It is the name of species that users want to examine, which should also be identical with the name written in the input (.inp and .mol) files.

Sample: SpeciesName = ‘clat’

* TimeBins

Acceptable value: Int (Must provide)

Description: It is the number of bins that users want to divide the selected time period into. The value should be positive integer.

* ShowNum

Acceptable value: Bool (Optional)

Default: True

Description: If True, the corresponding number will be shown in the center of each box; if false, there will be no number shown.

* SaveFig

Acceptable value: Bool (Optional)

Default: False

Description: If True, the plot will be saved as a ‘.png’ file in the current directory; if False, the figure will not be saved.

**Function for creating 2D heatmap showing number of total count of monomers inside N-mers vs. time:**

hist\_time\_heatmap\_mono\_count(FileName, InitialTime, FinalTime, SpeciesName, TimeBins, ShowNum, SaveFig)

Description: This function enables users to generate 2D histogram of total count of monomers in different N-mers vs. time. The x-axis is the number of monomers in single complex and y-axis is the time period. The color in each box indicates the total number of corresponding monomers in N-mers when corresponding time period is reached.

Parameters:

* FileName

Acceptable value: String (Must provide)

Description: It is the path to the ‘.dat’ file, which is usually named as ‘histogram\_complexes\_time.dat’, representing the histogram data to be analyzed.

Sample: FileName = ‘/Users/UserName/Documents/histogram\_complexes\_time.dat’

* InitialTime

Acceptable value: Float (Must provide)

Unit: s

Description: It is the initial time that users desire to examine. The acceptable range should not smaller than the starting time or exceed the ending time of simulation.

Sample: InitialTime = 0.5

* FinalTime

Acceptable value: Float (Must provide)

Unit: s

Description: It is the final time that users desire to examine. The acceptable range should not smaller than the value of InitialTime or exceed the ending time of simulation.

Sample: FinalTime = 1.0

* SpeciesName

Acceptable value: String (Must provide)

Description: It is the name of species that users want to examine, which should also be identical with the name written in the input (.inp and .mol) files.

Sample: SpeciesName = ‘clat’

* TimeBins

Acceptable value: Int (Must provide)

Description: It is the number of bins that users want to divide the selected time period into. The value should be positive integer.

* ShowNum

Acceptable value: Bool (Optional)

Default: True

Description: If True, the corresponding number will be shown in the center of each box; if false, there will be no number shown.

* SaveFig

Acceptable value: Bool (Optional)

Default: False

Description: If True, the plot will be saved as a ‘.png’ file in the current directory; if False, the figure will not be saved.

**Function for creating 2D heatmap showing fractions of monomers forming N-mers vs. time:**

hist\_time\_heatmap\_fraction (FileName, InitialTime, FinalTime, SpeciesName, TimeBins, ShowNum, SaveFig)

Description: This function enables users to generate 2D histogram of fractions of monomers forming different N-mers vs. time. The x-axis is the number of monomers in single complex and y-axis is the time period. The color in each box indicates the fraction of monomers forming corresponding N-mers when corresponding time period is reached.

Parameters:

* FileName

Acceptable value: String (Must provide)

Description: It is the path to the ‘.dat’ file, which is usually named as ‘histogram\_complexes\_time.dat’, representing the histogram data to be analyzed.

Sample: FileName = ‘/Users/UserName/Documents/histogram\_complexes\_time.dat’

* InitialTime

Acceptable value: Float (Must provide)

Unit: s

Description: It is the initial time that users desire to examine. The acceptable range should not smaller than the starting time or exceed the ending time of simulation.

Sample: InitialTime = 0.5

* FinalTime

Acceptable value: Float (Must provide)

Unit: s

Description: It is the final time that users desire to examine. The acceptable range should not smaller than the value of InitialTime or exceed the ending time of simulation.

Sample: FinalTime = 1.0

* SpeciesName

Acceptable value: String (Must provide)

Description: It is the name of species that users want to examine, which should also be identical with the name written in the input (.inp and .mol) files.

Sample: SpeciesName = ‘clat’

* TimeBins

Acceptable value: Int (Must provide)

Description: It is the number of bins that users want to divide the selected time period into. The value should be positive integer.

* ShowNum

Acceptable value: Bool (Optional)

Default: True

Description: If True, the corresponding number will be shown in the center of each box; if false, there will be no number shown.

* SaveFig

Acceptable value: Bool (Optional)

Default: False

Description: If True, the plot will be saved as a ‘.png’ file in the current directory; if False, the figure will not be saved.

## Analyzing 'transition\_matrix\_time.dat' File

**Function for creating line plot of free energy among different size of complexes:**

free\_energy(FileName, FileNum , InitialTime , FinalTime , SpeciesName , SaveFig)

Description: The plot indicates the change in free energy in selected time period among different size of complexes. The x-axis is the size of complex and the y-axis is the free energy calculated as in the unit of , where the refers to the probability of occurrence of the number of times N-mer is counted (including association and dissociation). If multiple input files are given, the output plot will be the average value of all files and an error bar will also be included.

Parameters:

* FileName

Acceptable value: String (Must provide)

Description: It is the path to the ‘.dat’ file, which is usually named as ‘histogram\_complexes\_time.dat’, representing the histogram data to be analyzed.

Sample: FileName = ‘/Users/UserName/Documents/transition\_matrix\_time.dat’

* FileNum

Acceptable value: Int (Must Provide)

Description: It is the number of the total input file. If multiple files are provided, their names should obey the naming rule listed below.

Sample: FileNum = 5

Note (naming rule): If single file is provided, the input file should be named as its original name (‘transition\_matrix\_time.dat’); if multiple file is provided, the name of input file should also include serial number as ‘transition\_matrix\_time\_X.dat’ where X = 1,2,3,4,5…

* InitialTime

Acceptable value: Float (Must provide)

Unit: s

Description: It is the initial time that users desire to examine. The acceptable range should not smaller than the starting time or exceed the ending time of simulation.

Sample: InitialTime = 0.5

* FinalTime

Acceptable value: Float (Must provide)

Unit: s

Description: It is the final time that users desire to examine. The acceptable range should not smaller than the value of InitialTime or exceed the ending time of simulation.

Sample: FinalTime = 1.0

* SpeciesName

Acceptable value: String (Must provide)

Description: It is the name of species that users want to examine, which should also be identical with the name written in the input (.inp and .mol) files.

Sample: SpeciesName = ‘clat’

* SaveFig

Acceptable value: Bool (Optional)

Default: False

Description: If True, the plot will be saved as a ‘.png’ file in the current directory; if False, the figure will not be saved.

**Function for creating line plot of symmetric associate probability:**

associate\_prob\_symmetric(FileName, FileNum, InitialTime, FinalTime, SpeciesName, DivideSize, SaveFig)

Description: This line plot represents the probability of association between complexes of different sizes and other complexes of different sizes. The x-axis is the size of the complex and y-axis is the associate probability. Three lines will exist in the line graph, representing associating to complexes of sizes less than, equal to, or greater than the specified size, respectively. 'Symmetric' in the function name means that for the associate reaction, both sizes of complexes are counted as associating events symmetrically, for example, if an associate event occurs where a trimer associates to a tetramer as a heptamer, then this event is counted twice, which are trimer associates to tetramer and tetramer associates to trimer. If multiple input files are given, the output plot will be the average value of all files and an error bar will also be included.

Parameters:

* FileName

Acceptable value: String (Must provide)

Description: It is the path to the ‘.dat’ file, which is usually named as ‘histogram\_complexes\_time.dat’, representing the histogram data to be analyzed.

Sample: FileName = ‘/Users/UserName/Documents/transition\_matrix\_time.dat’

* FileNum

Acceptable value: Int (Must Provide)

Description: It is the number of the total input file. If multiple files are provided, their names should obey the naming rule listed below.

Sample: FileNum = 5

Note (naming rule): If single file is provided, the input file should be named as its original name (‘transition\_matrix\_time.dat’); if multiple file is provided, the name of input file should also include serial number as ‘transition\_matrix\_time\_X.dat’ where X = 1,2,3,4,5…

* InitialTime

Acceptable value: Float (Must provide)

Unit: s

Description: It is the initial time that users desire to examine. The acceptable range should not smaller than the starting time or exceed the ending time of simulation.

Sample: InitialTime = 0.5

* FinalTime

Acceptable value: Float (Must provide)

Unit: s

Description: It is the final time that users desire to examine. The acceptable range should not smaller than the value of InitialTime or exceed the ending time of simulation.

Sample: FinalTime = 1.0

* SpeciesName

Acceptable value: String (Must provide)

Description: It is the name of species that users want to examine, which should also be identical with the name written in the input (.inp and .mol) files.

Sample: SpeciesName = ‘clat’

* DivideSize

Acceptable Value: int (Optional)

Default: DivideSize = 2

Description: This is the value that distinguishes the size of the associate complex, for example, if DivideSize = 2, that means the associate events are classified as ‘associate size < 2’, ‘associate size = 2’ and ‘associate size > 2’.

* SaveFig

Acceptable value: Bool (Optional)

Default: False

Description: If True, the plot will be saved as a ‘.png’ file in the current directory; if False, the figure will not be saved.

**Function for creating line plot of asymmetric associate probability:**

associate\_prob\_asymmetric(FileName, FileNum, InitialTime, FinalTime, SpeciesName, DivideSize, SaveFig)

Description: This line plot represents the probability of association between complexes of different sizes and other complexes of different sizes. The x-axis is the size of the complex and y-axis is the associate probability. Three lines will exist in the line graph, representing associating to complexes of sizes less than, equal to, or greater than the specified size, respectively. 'Asymmetric' in the function name means that for the associate reaction, only the complexes of smaller size associating to the larger one is counted as associate event asymmetrically, for example, if an associating event occurs where a trimer associates to a tetramer as a heptamer, then this event is counted only once, which is a trimer associates to tetramer. If multiple input files are given, the output plot will be the average value of all files and an error bar will also be included.

Parameters:

* FileName

Acceptable value: String (Must provide)

Description: It is the path to the ‘.dat’ file, which is usually named as ‘histogram\_complexes\_time.dat’, representing the histogram data to be analyzed.

Sample: FileName = ‘/Users/UserName/Documents/transition\_matrix\_time.dat’

* FileNum

Acceptable value: Int (Must Provide)

Description: It is the number of the total input file. If multiple files are provided, their names should obey the naming rule listed below.

Sample: FileNum = 5

Note (naming rule): If single file is provided, the input file should be named as its original name (‘transition\_matrix\_time.dat’); if multiple file is provided, the name of input file should also include serial number as ‘transition\_matrix\_time\_X.dat’ where X = 1,2,3,4,5…

* InitialTime

Acceptable value: Float (Must provide)

Unit: s

Description: It is the initial time that users desire to examine and plot on the plot. The acceptable range should not smaller than the starting time or exceed the ending time of simulation.

Sample: InitialTime = 0.5

* FinalTime

Acceptable value: Float (Must provide)

Unit: s

Description: It is the final time that users desire to examine. The acceptable range should not smaller than the value of InitialTime or exceed the ending time of simulation.

Sample: FinalTime = 1.0

* SpeciesName

Acceptable value: String (Must provide)

Description: It is the name of species that users want to examine, which should also be identical with the name written in the input (.inp and .mol) files.

Sample: SpeciesName = ‘clat’

* DivideSize

Acceptable Value: int (Optional)

Default: DivideSize = 2

Description: This is the value that distinguishes the size of the associate complex, for example, if DivideSize = 2, that means the associate events are classified as ‘associate size < 2’, ‘associate size = 2’ and ‘associate size > 2’.

* SaveFig

Acceptable value: Bool (Optional)

Default: False

Description: If True, the plot will be saved as a ‘.png’ file in the current directory; if False, the figure will not be saved.

**Function for creating line plot of symmetric dissociate probability:**

dissociate\_prob\_symmetric(FileName, FileNum, InitialTime, FinalTime, SpeciesName, DivideSize, SaveFig)

Description: This line plot represents the probability of dissociation of complexes of different sizes into other complexes of different sizes. The x-axis is the size of the complex and y-axis is the dissociate probability. Three lines will exist in the line graph, representing dissociating to complexes of sizes less than, equal to, or greater than the specified size, respectively. 'Symmetric' in the function name means that for the dissociate reaction, both sizes of complexes are counted as dissociating events symmetrically, for example, if an dissociate event occurs where a heptamer dissociates into a tetramer and a trimer, then this event is counted twice, which are heptamer dissociates to tetramer and heptamer dissociates to trimer. If multiple input files are given, the output plot will be the average value of all files and an error bar will also be included.

Parameters:

* FileName

Acceptable value: String (Must provide)

Description: It is the path to the ‘.dat’ file, which is usually named as ‘histogram\_complexes\_time.dat’, representing the histogram data to be analyzed.

Sample: FileName = ‘/Users/UserName/Documents/transition\_matrix\_time.dat’

* FileNum

Acceptable value: Int (Must Provide)

Description: It is the number of the total input file. If multiple files are provided, their names should obey the naming rule listed below.

Sample: FileNum = 5

Note (naming rule): If single file is provided, the input file should be named as its original name (‘transition\_matrix\_time.dat’); if multiple file is provided, the name of input file should also include serial number as ‘transition\_matrix\_time\_X.dat’ where X = 1,2,3,4,5…

* InitialTime

Acceptable value: Float (Must provide)

Unit: s

Description: It is the initial time that users desire to examine. The acceptable range should not smaller than the starting time or exceed the ending time of simulation.

Sample: InitialTime = 0.5

* FinalTime

Acceptable value: Float (Must provide)

Unit: s

Description: It is the final time that users desire to examine. The acceptable range should not smaller than the value of InitialTime or exceed the ending time of simulation.

Sample: FinalTime = 1.0

* SpeciesName

Acceptable value: String (Must provide)

Description: It is the name of species that users want to examine, which should also be identical with the name written in the input (.inp and .mol) files.

Sample: SpeciesName = ‘clat’

* DivideSize

Acceptable Value: int (Optional)

Default: DivideSize = 2

Description: This is the value that distinguishes the size of the dissociate complex, for example, if DivideSize = 2, that means the dissociate events are classified as ‘dissociate size < 2’, ‘dissociate size = 2’ and ‘dissociate size > 2’.

* SaveFig

Acceptable value: Bool (Optional)

Default: False

Description: If True, the plot will be saved as a ‘.png’ file in the current directory; if False, the figure will not be saved.

**Function for creating line plot of asymmetric dissociate probability:**

dissociate\_prob\_asymmetric(FileName, FileNum, InitialTime, FinalTime, SpeciesName, DivideSize, SaveFig)

Description: This line plot represents the probability of dissociation of complexes of different sizes into other complexes of different sizes. The x-axis is the size of the complex and y-axis is the dissociate probability. Three lines will exist in the line graph, representing dissociating to complexes of sizes less than, equal to, or greater than the specified size, respectively. 'Asymmetric' in the function name means that for the dissociate reaction, only the complexes of smaller size dissociating from the original one is counted as dissociate event asymmetrically, for example, if an dissociate event occurs where a heptamer dissociates into a tetramer and a trimer, then this event is counted only once, which is heptamer dissociates to trimer. If multiple input files are given, the output plot will be the average value of all files and an error bar will also be included.

Parameters:

* FileName

Acceptable value: String (Must provide)

Description: It is the path to the ‘.dat’ file, which is usually named as ‘histogram\_complexes\_time.dat’, representing the histogram data to be analyzed.

Sample: FileName = ‘/Users/UserName/Documents/transition\_matrix\_time.dat’

* FileNum

Acceptable value: Int (Must Provide)

Description: It is the number of the total input file. If multiple files are provided, their names should obey the naming rule listed below.

Sample: FileNum = 5

Note (naming rule): If single file is provided, the input file should be named as its original name (‘transition\_matrix\_time.dat’); if multiple file is provided, the name of input file should also include serial number as ‘transition\_matrix\_time\_X.dat’ where X = 1,2,3,4,5…

* InitialTime

Acceptable value: Float (Must provide)

Unit: s

Description: It is the initial time that users desire to examine. The acceptable range should not smaller than the starting time or exceed the ending time of simulation.

Sample: InitialTime = 0.5

* FinalTime

Acceptable value: Float (Must provide)

Unit: s

Description: It is the final time that users desire to examine. The acceptable range should not smaller than the value of InitialTime or exceed the ending time of simulation.

Sample: FinalTime = 1.0

* SpeciesName

Acceptable value: String (Must provide)

Description: It is the name of species that users want to examine, which should also be identical with the name written in the input (.inp and .mol) files.

Sample: SpeciesName = ‘clat’

* DivideSize

Acceptable Value: int (Optional)

Default: DivideSize = 2

Description: This is the value that separates the size of the dissociate complex, for example, if DivideSize = 2, that means the dissociate events are classified as ‘dissociate size < 2’, ‘dissociate size = 2’ and ‘dissociate size > 2’.

* SaveFig

Acceptable value: Bool (Optional)

Default: False

Description: If True, the plot will be saved as a ‘.png’ file in the current directory; if False, the figure will not be saved.

**Function for creating line plot of growth probability:**

growth\_prob(FileName, FileNum, InitialTime, FinalTime, SpeciesName, SaveFig)

Description: This line plot indicates the probability of growth in size for different sizes of complexes. The x-axis is the size of complexes, and the y-axis is the growth probability. If multiple input files are given, the output plot will be the average value of all files and an error bar will also be included.

Parameters:

* FileName

Acceptable value: String (Must provide)

Description: It is the path to the ‘.dat’ file, which is usually named as ‘histogram\_complexes\_time.dat’, representing the histogram data to be analyzed.

Sample: FileName = ‘/Users/UserName/Documents/transition\_matrix\_time.dat’

* FileNum

Acceptable value: Int (Must Provide)

Description: It is the number of the total input file. If multiple files are provided, their names should obey the naming rule listed below.

Sample: FileNum = 5

Note (naming rule): If single file is provided, the input file should be named as its original name (‘transition\_matrix\_time.dat’); if multiple file is provided, the name of input file should also include serial number as ‘transition\_matrix\_time\_X.dat’ where X = 1,2,3,4,5…

* InitialTime

Acceptable value: Float (Must provide)

Unit: s

Description: It is the initial time that users desire to examine. The acceptable range should not smaller than the starting time or exceed the ending time of simulation.

Sample: InitialTime = 0.5

* FinalTime

Acceptable value: Float (Must provide)

Unit: s

Description: It is the final time that users desire to examine. The acceptable range should not smaller than the value of InitialTime or exceed the ending time of simulation.

Sample: FinalTime = 1.0

* SpeciesName

Acceptable value: String (Must provide)

Description: It is the name of species that users want to examine, which should also be identical with the name written in the input (.inp and .mol) files.

Sample: SpeciesName = ‘clat’

* SaveFig

Acceptable value: Bool (Optional)

Default: False

Description: If True, the plot will be saved as a ‘.png’ file in the current directory; if False, the figure will not be saved.

**Function for creating line plot of complexes lifetime:**

complex\_lifetime(FileName, FileNum, InitialTime, FinalTime, SpeciesName, SaveFig)

Description: This line plot indicates the lifetime for different sizes of complexes. The x-axis is the size of complexes, and the y-axis is the corresponding lifetime in unit of second. If multiple input files are given, the output plot will be the average value of all files and an error bar will also be included.

Parameters:

* FileName

Acceptable value: String (Must provide)

Description: It is the path to the ‘.dat’ file, which is usually named as ‘histogram\_complexes\_time.dat’, representing the histogram data to be analyzed.

Sample: FileName = ‘/Users/UserName/Documents/transition\_matrix\_time.dat’

* FileNum

Acceptable value: Int (Must Provide)

Description: It is the number of the total input file. If multiple files are provided, their names should obey the naming rule listed below.

Sample: FileNum = 5

Note (naming rule): If single file is provided, the input file should be named as its original name (‘transition\_matrix\_time.dat’); if multiple file is provided, the name of input file should also include serial number as ‘transition\_matrix\_time\_X.dat’ where X = 1,2,3,4,5…

* InitialTime

Acceptable value: Float (Must provide)

Unit: s

Description: It is the initial time that users desire to examine. The acceptable range should not smaller than the starting time or exceed the ending time of simulation.

Sample: InitialTime = 0.5

* FinalTime

Acceptable value: Float (Must provide)

Unit: s

Description: It is the final time that users desire to examine. The acceptable range should not smaller than the value of InitialTime or exceed the ending time of simulation.

Sample: FinalTime = 1.0

* SpeciesName

Acceptable value: String (Must provide)

Description: It is the name of species that users want to examine, which should also be identical with the name written in the input (.inp and .mol) files.

Sample: SpeciesName = ‘clat’

* SaveFig

Acceptable value: Bool (Optional)

Default: False

Description: If True, the plot will be saved as a ‘.png’ file in the current directory; if False, the figure will not be saved.

## Analyzing Multi-component 'histogram\_complexes\_time.dat' File

**Function for converting multi-species ‘histogram\_complexes\_time.dat’ file to a pandas.DataFrame:**

multi\_hist\_to\_df (FileName, SaveCsv)

Description: This function enables users to convert the raw .dat file to a data frame in python pandas package for multi-species system. Each column in the data frame includes the simulation time and selected occurrences of species during the simulation; each row is separated by a different simulation time.

Parameters:

* FileName

Acceptable value: String (Must provide)

Description: It is the path to the ‘.dat’ file, which is usually named as ‘histogram\_complexes\_time.dat’, representing the histogram data to be analyzed.

Sample: FileName = ‘/Users/UserName/Documents/histogram\_complexes\_time.dat’

* SaveCsv

Acceptable value: Bool (Optional)

Default: True

Description: If True, the corresponding .csv file will also be saved as ‘histogram.csv’ under the same directory; if False, the generated .csv file will be deleted.

**Function for converting multi-species ‘histogram\_complexes\_time.dat’ file to a .csv file:**

multi\_hist\_to\_csv (FileName)

Description: This function enables users to convert the raw .dat file to a .csv file for multi-species system. Each column in the data frame includes the simulation time and selected occurrences of species during the simulation; each row is separated by a different simulation time.

Parameters:

* FileName

Acceptable value: String (Must provide)

Description: It is the path to the ‘.dat’ file, which is usually named as ‘histogram\_complexes\_time.dat’, representing the histogram data to be analyzed.

Sample: FileName = ‘/Users/UserName/Documents/histogram\_complexes\_time.dat’

**Function for generating general multi-species histogram:**

multi\_hist(FileName, FileNum, InitialTime, FinalTime, SpeciesList, xAxis, BarSize, ExcludeSize, SaveFig)

Description: This function enables users to plot general histogram of total size of complex or selected species inside each complex for a multi-species system. It will also analyze multiple input files and show the result along with error bar. The x-axis is the size of selected species or total number of monomers, and the y-axis is the average number of counts for corresponding size.

Parameters:

* FileName:

Acceptable value: String (Must provide)

Description: It is the path to the ‘.dat’ file, which is usually named as ‘histogram\_complexes\_time.dat’, representing the histogram data to be analyzed.

Sample: FileName = ‘/Users/UserName/Documents/histogram\_complexes\_time.dat’

* FileNum

Acceptable value: Int (Must Provide)

Description: It is the number of the total input file. If multiple files are provided, their names should obey the naming rule listed below.

Sample: FileNum = 5

Note (naming rule): If single file is provided, the input file should be named as its original name (‘histogram\_complexes\_time.dat’); if multiple files are provided, the name of input file should also include serial number as ‘histogram\_complexes\_time\_X.dat’ where X = 1,2,3,4,5…

* InitialTime

Acceptable value: Float (Must provide)

Unit: s

Description: It is the initial time that users desire to examine. The acceptable range should not smaller than the starting time or exceed the ending time of simulation.

Sample: InitialTime = 0.0

* FinalTime

Acceptable value: Float (Must provide)

Unit: s

Description: It is the final time that users desire to examine. The acceptable range should not smaller than the value of InitialTime or exceed the ending time of simulation.

Sample: FinalTime = 1.0

* SpeciesList

Acceptable value: List with String Elements (Must provide)

Description: It contains all the name of species inside the simulation system.

Sample: SpeciesList = [‘A’, ‘B’]

* xAxis

Acceptable value: String (Must provide)

Description: It indicates the species shown on the x-axis. If xAxis is included inside SpeciesList, the x-axis will only show the number of selected components. The other input is ‘tot’, which represents the x\_axis will count all species in a single complex.

Sample: xAxis = ‘tot’ or xAxis = ‘A’

* BarSize

Acceptable value: Int (Optional)

Default: 1

Description: It is size of each data bar in x-dimension. The x-axis will be separated evenly according to this number and the count of each size range will be sum up and shown together.

* ExcludeSize

Acceptable value: Int (Optional)

Default: 0

Description: In the generated plot, the number of monomers in the complex that are no larger than this number will be excluded and will not be considered into the average calculation.

* SaveFig

Acceptable value: Bool (Optional)

Default: False

Description: If True, the plot will be saved as a ‘.png’ file in the current directory; if False, the figure will not be saved.

**Function for generating stacked multi-species histogram:**

multi\_hist\_stacked(FileName, FileNum, InitialTime, FinalTime, SpeciesList, xAxis, DivideSpecies, DivideSize, BarSize, ExcludeSize, SaveFig)

Description: This function enables users to plot general histogram of total size of complex or selected species for a multi-species system. Each bar is split by three stacked bars which represent the size distribution of another selected species compared to a desired input. It will also analyze multiple input files and show the result along with error bar. The x-axis is the size of selected species or total number of monomers, and the y-axis is the average number of counts for corresponding size.

Parameters:

* FileName:

Acceptable value: String (Must provide)

Description: It is the path to the ‘.dat’ file, which is usually named as ‘histogram\_complexes\_time.dat’, representing the histogram data to be analyzed.

Sample: FileName = ‘/Users/UserName/Documents/histogram\_complexes\_time.dat’

* FileNum

Acceptable value: Int (Must Provide)

Description: It is the number of the total input file. If multiple files are provided, their names should obey the naming rule listed below.

Sample: FileNum = 5

Note (naming rule): If single file is provided, the input file should be named as its original name (‘histogram\_complexes\_time.dat’); if multiple files are provided, the name of input file should also include serial number as ‘histogram\_complexes\_time\_X.dat’ where X = 1,2,3,4,5…

* InitialTime

Acceptable value: Float (Must provide)

Unit: s

Description: It is the initial time that users desire to examine. The acceptable range should not smaller than the starting time or exceed the ending time of simulation.

Sample: InitialTime = 0.0

* FinalTime

Acceptable value: Float (Must provide)

Unit: s

Description: It is the final time that users desire to examine. The acceptable range should not smaller than the value of InitialTime or exceed the ending time of simulation.

Sample: FinalTime = 1.0

* SpeciesList

Acceptable value: List with String Elements (Must provide)

Description: It contains all the name of species inside the simulation system.

Sample: SpeciesList = [‘A’, ‘B’]

* xAxis

Acceptable value: String (Must provide)

Description: It indicates the species shown on the x-axis. If xAxis is included inside SpeciesList, the x-axis will only show the number of selected components. The other input is ‘tot’, which represents the x\_axis will count all species in a single complex.

Sample: xAxis = ‘tot’ or xAxis = ‘A’

* DivideSpecies

Acceptable value: String (Must provide)

Description: It indicates the name of the species that users want to separate by size.

Sample: DivideSpecies = ‘A’

* DivideSize

Acceptable value: Int (Must provide)

* Description: This is the value that separates the size of the dissociate complex, for example, if DivideSize = 5, that means the dissociate events are classified as ‘DivideSpecies size < 5’, ‘DivideSpecies size = 5’ and ‘DivideSpecies size > 5’.

Sample: DivideSize = 5

* BarSize

Acceptable value: Int (Optional)

Default: 1

Description: It is size of each data bar in x-dimension. The x-axis will be separated evenly according to this number and the count of each size range will be sum up and shown together.

* ExcludeSize

Acceptable value: Int (Optional)

Default: 0

Description: In the generated plot, the number of monomers in the complex that are no larger than this number will be excluded and will not be considered into the average calculation.

* SaveFig

Acceptable value: Bool (Optional)

Default: False

Description: If True, the plot will be saved as a ‘.png’ file in the current directory; if False, the figure will not be saved.

## Locating position for certain size of complexes by PDB/restart file

**By PDB file:**

locate\_position\_PDB(FileNamePdb, NumList, FileNameInp, BufferRatio)

Description: This function enables users to locate specific complexes of certain size from a PDB file after simulation. The result will be output as a separated file named “output\_file.pdb” containing only the desired complex.

Note: Reading only the PDB file slows down the function compared to reading the 'restart.dat' file, because the function needs to calculate the distance between all reactive atoms that can be reacted based on the reaction information to determine whether they are bound or not. Therefore, this function is universal but runs slowly in time.

Parameters:

* FileNamePdb

Acceptable value: String (Must provide)

Description: It is the path to the PDB file, which is usually the last frame of simulation.

Sample: FileNamePdb = ‘/Users/UserName/Documents/999999.pdb’

* Numlist

Acceptable value: List with Int elements (Must provide)

Description: It represents the number of individual monomers in the complex that the user needs to locate, and the order of the monomers in the list is determined by the order in which they appear in the PDB file.

Sample: NumList = [12] (This represents the complexes is consist of only one kind of monomer and the desired complexes is a dodecamer, also known as 12-mer)

* FileNameInp

Acceptable value: String (Must provide)

Description: It is the path to the ‘.inp’ file, which usually stores the reaction information.

Sample: FileNameInp = ‘/Users/UserName/Documents/parms.inp’

* BufferRatio

Acceptable value: Float (Optional)

Default: 0.01

Description: When the sigma value is within the interval of 1.0 plus or minus the ratio, the two reaction interfaces can be considered as bonded.

**By ‘restart.dat’ file:**

locate\_position\_restart(FileNamePdb, NumList, FileNameRestart)

Description: This function enables users to locate specific complexes of certain size from a PDB file along with ‘restart.dat’ file after simulation. The result will be output as a separated file named “output\_file.pdb” containing only the desired complex.

Note: The advantage of reading the 'restart.dat' file is that the file directly stores the binding information of each complex in the system and can be used directly, so the function runs faster; however, the function is not universal, if the 'restart.dat ' file's write logic changes, then this function will no longer work.

Parameters:

* FileNamePdb

Acceptable value: String (Must provide)

Description: It is the path to the PDB file, which is usually the last frame of simulation.

Sample: FileNamePdb = ‘/Users/UserName/Documents/999999.pdb’

* Numlist

Acceptable value: List with Int elements (Must provide)

Description: It represents the number of individual monomers in the complex that the user needs to locate, and the order of the monomers in the list is determined by the order in which they appear in the PDB file.

Sample: NumList = [12] (This represents the complexes is consist of only one kind of monomer and the desired complexes is a dodecamer, also known as 12-mer)

* FileNameRestart

Acceptable value: String (Optional)

Default: 'restart.dat'

Description: It is the path to the ‘restart.dat’ file.

# Platonic Solid Self-assembly Input File Writing

Platonic solid self-assembly include 10 models, so that 10 separate functions are needed, which are shown in the following table:

|  |  |  |
| --- | --- | --- |
| Platonic Solid | Center-of-Mass Position | Name of Function |
| Tetrahedron (4-face) | Each Face | tetr\_face(radius, sigma) |
| Tetrahedron (4-face) | Each Vertex | tetr\_vert(radius, sigma) |
| Cube (6-face) | Each Face | cube\_face(radius, sigma) |
| Cube (6-face) | Each Vertex | cube\_vert(radius, sigma) |
| Octahedron (8-face) | Each Face | octa\_face(radius, sigma) |
| Octahedron (8-face) | Each Vertex | octa\_vert(radius, sigma) |
| Dodecahedron (12-face) | Each Face | dode\_face(radius, sigma) |
| Dodecahedron (12-face) | Each Vertex | dode\_vert(radius, sigma) |
| Icosahedron (20-face) | Each Face | icos\_face(radius, sigma) |
| Icosahedron (20-face) | Each Vertex | icos\_vert(radius, sigma) |

Description: This function enable users to generate NERDSS input files (.inp and .mol files) for Platonic solid self-assembly system.

Parameters:

* radius

Acceptable value: Float (Must provide)

Unit: nm

Description: It is the radius of the Platonic solid, which is defined by the distance from the center of Platonic solid to each vertex.

Sample: radius = 40.0

* sigma

Acceptable value: Float (Must provide)

Unit: nm

Description: It is the distance of each interface when reaction takes place.

Sample: sigma = 1.0