
biometall Documentation

biometall

Mar 13, 2023

Contents:

1	Getting Started	3
2	How to install	5
2.1	Standalone executable	5
2.2	PyPi repository	5
2.3	OS Compatibility	6
3	Learning how a BioMetAll calculation works	7
3.1	Objectives	7
3.2	1. Executing the program with default parameters	7
3.3	2. Executing with custom parameters	8
3.4	3. Searching for a specific motif	10
3.5	4. Proposing mutations to complete a motif	11

BioMetAll is a command line application to allow the identification of metal-binding sites in proteins from backbone preorganization.



CHAPTER 1

Getting Started

This page details how to get started with biometall.

CHAPTER 2

How to install

To install BioMetAll we provide two different options.

- 1 - You can download a standalone executable that you can run in your terminal (Linux/macOS) or command prompt (Win10).
- 2 - You can install it through the PyPi repository.

2.1 Standalone executable

1 - Download the executable for your Operating System (BioMetAll v1.0):

- [Linux](#). Tested on Ubuntu 18.04, 20.04 and 22.04.
- [macOS](#). Tested on macOS 10.15 Catalina.
- [Windows 10](#). Tested on Windows 10.

Tip: These executables have been created with [PyInstaller](#) and have been tested on the indicated operating systems. If the executable doesn't work in your specific distribution, it may be due to the lack of some operating system libraries. If that happens, we recommend installing the BioMetAll package from the PyPi repository (see below).

2.2 PyPi repository

To install BioMetAll from the PyPi repository, you should type in your terminal (Linux/macOS) or command prompt (Win10):

```
pip install biometall
```

Tip: To install packages from the PyPi repository, you must have Python ≥ 3.5 and pip installed in your computer. An expanded explanation of the requirements for installing PyPi packages can be found [here](#).

2.3 OS Compatibility

BioMetAll is compatible with Linux, macOS and Windows 10.

If you find some difficulties when installing it in a concrete distribution, please use the [issues page](#) to report them.

Learning how a BioMetAll calculation works

3.1 Objectives

In this first tutorial, the goal is to understand all the options that BioMetAll has to personalize a calculation.

The only mandatory input for a BioMetAll calculation is a structure of the biological system that will be object of study. It can be provided in a *.pdb* file or by indicating its 4-character PDB code. In the latter case, the structure will be downloaded from the Protein Data Bank and saved in a *xxxx.pdb* file on the working directory.

Tip: *BioMetAll* can handle files with several chains, but if you are interested specifically in one/several of them, removing the non useful chains in your *.pdb* file is recommended.

Along this use case tutorial, the chain A of the PDB code 1dhy (*KKS102 BPHC enzyme*) will be used.

3.2 1. Executing the program with default parameters

To execute the exploration of possible metal-binding zones, you simply should call the executable of *BioMetAll* followed by the *.pdb* file path or the 4-character PDB code:

```
biometall <path/to/the/input_file.pdb>
```

or

```
biometall 1dhy
```

You will obtain a list of solutions (i.e. the possible coordinating environments), followed by the coordinates of the center of the area, its radius and the number of probes which match the coordination criteria. In the case of **1dhy**, 39 solutions are obtained, being the first one an area formed by 44 probes.

- The results of the execution will be shown on the screen and saved in a *results_biometall_1dhy.txt* file in the working directory.

- The different coordinating environments will be ordered by number of probes that match that environment.
- If one probe matches several coordinating environments (e.g. [HIS:145:A HIS:193:A HIS:194:A ASP:243:A] and [HIS:193:A HIS:194:A ASP:243:A]), that probe is accounted in all the environments it matches.
- The center and radius provided for each solution defines the minimum sphere which contains all the probes of that solution.

In addition to the *.txt* results file, you can obtain a *.pdb* file containing the different coordination areas with their corresponding centers. To do so, you should use the *-pdb* parameter:

```
biometall --pdb 1dhy
```

It will output a *probes_1dhy.pdb* that you can open together with your original *.pdb* file in any visualization program:

Every solution is stored as Xeon atoms (probes, depicted in green) and a Helium atom (center of each coordinating environment, depicted in yellow). Each solution is stored as a different residue in the *probes_1dhy.pdb* file, so you can easily see which probes corresponds to each one. For example, for the first coordinating environment (probes in green and center in yellow sphere):

3.3 2. Executing with custom parameters

2.1. Filtering the results by number of probes ('-cutoff')

By default, the program proposes as solutions all the coordinating environments. If you want to obtain solutions only with a high number of probes (i.e. wide areas where the metal could coordinate), you should use the *cutoff* parameter. For example, you can use a *cutoff* of 0.5 to obtain solutions with a number of probes \geq half of the probes of the solution with highest number. In the *1dhy* example, only solutions with number of probes ≥ 22 will be provided:

```
biometall --cutoff 0.5 1dhy
```

#	Coordinating residues	Coordinates of center	Num. probes	Radius search	Proposed mutations
1	HIS:193:A HIS:194:A ASP:243:A	79.055 43.235 18.791	44	5.891	
2	HIS:145:A HIS:193:A HIS:194:A	81.076 46.266 16.771	36	3.643	
3	HIS:208:A HIS:209:A GLU:258:A	91.179 38.184 16.771	35	4.165	
4	HIS:145:A HIS:209:A GLU:260:A	87.137 45.256 16.771	31	3.351	
5	HIS:188:A HIS:193:A HIS:194:A	75.014 45.256 21.822	30	3.031	
6	HIS:209:A GLU:258:A GLU:260:A	89.158 40.204 14.750	29	7.355	
7	HIS:193:A HIS:194:A HIS:240:A	79.055 43.235 16.771	22	3.031	
8	HIS:208:A GLU:258:A GLU:260:A	89.158 41.214 16.771	22	2.858	

- To show solutions with a number of probes higher than 40% of the highest scored:

```
biometall --cutoff 0.4 1dhy
```

#	Coordinating residues	Coordinates of center	Num. probes	Radius search	Proposed mutations
1	HIS:193:A HIS:194:A ASP:243:A	79.055 43.235 18.791	44	5.891	
2	HIS:145:A HIS:193:A HIS:194:A	81.076 46.266 16.771	36	3.643	
3	HIS:208:A HIS:209:A GLU:258:A	91.179 38.184 16.771	35	4.165	
4	HIS:145:A HIS:209:A GLU:260:A	87.137 45.256 16.771	31	3.351	
5	HIS:188:A HIS:193:A HIS:194:A	75.014 45.256 21.822	30	3.031	
6	HIS:209:A GLU:258:A GLU:260:A	89.158 40.204 14.750	29	7.355	
7	HIS:193:A HIS:194:A HIS:240:A	79.055 43.235 16.771	22	3.031	
8	HIS:208:A GLU:258:A GLU:260:A	89.158 41.214 16.771	22	2.858	
9	HIS:145:A HIS:188:A HIS:193:A	80.065 47.276 19.801	20	3.780	
10	ASP:170:A HIS:193:A HIS:194:A	76.024 43.235 20.812	18	3.031	

2.2. Customizing which amino acids are considered as possible coordinators ('-residues')

By default, the list of amino acids that are considered as potential metal coordinators is:

```
:: ASP, HIS, GLU, CYS
```

You can modify this list using the `--residues` parameter. You should indicate the amino acids in their 3-letter code, enclosed between brackets `[]` and separated by commas (without spaces). For example, to search for coordinating environments formed by a (much) more extended set of amino acids, you could type:

```
biometall --pdb --residues [ASN,ASP,CYS,GLN,GLU,HIS,MET,SER,THR,TYR] 1dhy
```

2.3. Obtaining solutions with a minimum of coordinating amino acids ('--min_coordinators')

By default, coordinating environments with a minimum of **three amino acids** are considered as valid solutions. If you want to change this number, it can be done with the `--min_coordinators` parameter.

- To show coordinating environments with a minimum of four amino acids:

```
biometall --pdb --min_coordinators 4 --residues [ASN,ASP,CYS,GLN,GLU,HIS,MET,SER,THR,
↪TYR] 1dhy
```

2.4. Adding solutions that match backbone oxygen atoms ('--backbone')

The usual case when searching metal-coordinating amino acids is to search for atoms of their side chains, and that is the default in *BioMetAll* calculations. However, in some cases could be useful to consider coordinations with backbone oxygens. To do so, you should use the `--backbone` parameter, followed by a list of the amino acids that you consider potential coordinators in their backbone. You should indicate the amino acids in their 3-letter code, enclosed between brackets `[]` and separated by commas (without spaces). Also, you can indicate the word *ALL* if you want to consider coordination with backbone oxygen of all amino acids.

For example, to consider coordinations with backbone oxygens of histidines besides the standard side chains coordinations, you should type:

```
biometall --backbone [HIS] 1dhy
```

To consider coordinations with backbone oxygens of all amino acids and a minimum of four coordinators:

```
biometall --min_coordinators 4 --backbone ALL --residues [ASN,ASP,CYS,GLN,GLU,HIS,MET,
↪SER,THR,TYR] 1dhy
```

Warning: In this latter example, the minimum of coordinators refers to the sum of sidechains and backbone coordinations detected. An amino acid is counted double if it can be coordinated both by a sidechain atom and a backbone atom.

2.5. Defining a more dense (or scattered) grid ('--grid')

BioMetAll defines a grid of points (i.e. probes) containing all the volume of your system. Each point of the grid is tested to see if it can be part of a coordinating environment. By default, points of the grid are generated with a separation of 1.0 Angstroms, but if you want a greater (or lower) detail, you can modify it using the `--grid` parameter. For example, to generate a grid with a separation of 0.5 Angstroms between probes:

```
biometall --grid 0.5 1dhy
```

Warning: *grid* parameter has direct implication in the computational time required to perform the calculation. It affects specially when *-mutations* option is used.

2.6. Defining a search zone ('-center' and '-radius')

By default, the grid of probes is constructed to embed the whole system. If you have an specific zone that you want to search, it can be defined using the *-center* (for the center of coordinates of the zone) and *-radius* parameters. The coordinates should be enclosed between brackets *[]* and separated by commas (without spaces). For example:

```
biometall --center [84.98,42.82,16.04] --radius 10.0 ldhy
```

2.7. Defining how many processors are used for the calculation ('-cores')

By default, *BioMetAll* calculations are run in a parallel mode using all the physical cores available in your computer. If for some reason (e.g. you are running *BioMetAll* in a cluster of computers) you want to change it, you can use the *-cores* parameters to define how many physical cores will be used. For example, to use two cores:

```
biometall --cores 2 ldhy
```

3.4 3. Searching for a specific motif

In addition to searching a list of possible coordinating amino acids with the *-residues* option, there exists also the possibility to search for a specific motif. The syntax offers two options: search for an exact motif and search for a variable motif.

To search for an exact motif, you must enclose the motif in square brackets *[]*. The 3-letter code for the amino acids that make up the motif must be separated with commas (without spaces). For example, to search for motifs composed of two histidines and one aspartate, you would use:

```
biometall --motif [HIS,HIS,ASP] ldhy
```

#	Coordinating residues	Coordinates of center	Num. probes	Radius search	Proposed mutations
1	HIS:193:A HIS:194:A ASP:243:A	79.055 43.235 18.791	44	5.891	
2	ASP:170:A HIS:193:A HIS:194:A	76.024 43.235 20.812	18	3.031	
3	HIS:145:A HIS:193:A ASP:243:A	80.065 46.266 16.771	16	2.858	
4	HIS:145:A HIS:194:A ASP:243:A	81.076 45.256 16.771	14	3.643	
5	HIS:188:A HIS:194:A ASP:243:A	76.024 46.266 18.791	12	2.475	
6	ASP:170:A HIS:188:A HIS:193:A	75.014 44.245 21.822	8	2.259	
7	ASP:170:A HIS:188:A HIS:194:A	75.014 44.245 21.822	7	2.259	
8	HIS:194:A HIS:240:A ASP:243:A	80.065 42.225 17.781	7	2.475	
9	HIS:193:A HIS:240:A ASP:243:A	80.065 42.225 17.781	6	3.780	
10	HIS:188:A HIS:193:A ASP:243:A	77.035 44.245 19.801	4	2.259	
11	HIS:145:A HIS:188:A ASP:243:A	79.055 46.266 18.791	1	0.000	

The second option is to introduce some variability using the forward slash symbol */*. It is used to indicate that a residue of the motif has several alternatives. For example, you could search for motifs composed of two histidines and either an aspartate or a glutamate with the following command:

```
biometall --motif [HIS,HIS,ASP/GLU] ldhy
```

#	Coordinating residues	Coordinates of center	Num. probes	Radius search	Proposed mutations
1	HIS:193:A HIS:194:A ASP:243:A	79.055 43.235 18.791	44	5.891	
2	HIS:208:A HIS:209:A GLU:258:A	91.179 38.184 16.771	35	4.165	
3	HIS:145:A HIS:209:A GLU:260:A	87.137 45.256 16.771	31	3.351	
4	ASP:170:A HIS:193:A HIS:194:A	76.024 43.235 20.812	18	3.031	
5	HIS:145:A HIS:193:A ASP:243:A	80.065 46.266 16.771	16	2.858	
6	HIS:208:A HIS:209:A GLU:260:A	88.148 40.204 15.760	16	3.643	
7	HIS:145:A HIS:194:A ASP:243:A	81.076 45.256 16.771	14	3.643	
8	HIS:188:A HIS:194:A ASP:243:A	76.024 46.266 18.791	12	2.475	
9	HIS:240:A GLU:260:A HIS:285:A	84.107 40.204 11.719	9	2.259	
10	ASP:170:A HIS:188:A HIS:193:A	75.014 44.245 21.822	8	2.259	
11	ASP:170:A HIS:188:A HIS:194:A	75.014 44.245 21.822	7	2.259	
12	HIS:194:A HIS:240:A ASP:243:A	80.065 42.225 17.781	7	2.475	
13	HIS:193:A HIS:240:A ASP:243:A	80.065 42.225 17.781	6	3.780	
14	HIS:188:A HIS:193:A ASP:243:A	77.035 44.245 19.801	4	2.259	
15	HIS:145:A HIS:188:A ASP:243:A	79.055 46.266 18.791	1	0.000	
16	HIS:145:A HIS:194:A GLU:260:A	84.107 45.256 16.771	1	0.000	

Tip: You can use slash notation in more than one position of the motif. Also, more than two options are allowed in each position. For example, [HIS/CYS,HIS,ASP/GLU/TYR] would represent a motif where one position should be either a histidine or a cysteine, another position should be a histidine, and the other position either an aspartate, a glutamate or a tyrosine.

3.5 4. Proposing mutations to complete a motif

BioMetAll can be used as a tool to start the design of a metal binding site by proposing mutations to complete an incipient binding site already present in the protein.

As an illustrative case we can continue with the example of the protein with PDB code 1dhy. It can be seen that, among the binding sites with three coordinating residues, there is none formed by three aspartates:

```
biometall --motif [ASP,ASP,ASP] 1dhy
```

None possible coordinating sites have been found. Try again with other parameters or check/change the input file.

However, there are eight possible binding sites with two coordinating aspartates:

```
biometall --motif [ASP,ASP] 1dhy
```

#	Coordinating residues	Coordinates of center	Num. probes	Radius search	Proposed mutations
1	ASP:67:A ASP:68:A	77.035 67.482 28.894	28	5.977	
2	ASP:225:A ASP:228:A	95.220 41.214 -4.445	15	5.151	
3	ASP:57:A ASP:58:A	102.292 41.214 23.843	15	4.630	
4	ASP:14:A ASP:57:A	105.322 45.256 17.781	12	2.475	
5	ASP:218:A ASP:219:A	84.107 54.348 -0.404	10	3.031	
6	ASP:170:A ASP:243:A	76.024 42.225 18.791	9	2.259	
7	ASP:14:A ASP:19:A	106.333 50.307 12.729	4	1.429	
8	ASP:140:A ASP:67:A	78.045 62.430 22.832	1	0.000	

A design exercise could be to find the options to complete these ASP,ASP motifs with another aspartate to form a triad. This is done with the combination of the `--motif` and `--mutations` options. In the *motif* option we will indicate the incipient motif already present in the protein (in this case ASP,ASP). In the *mutations* option, we will indicate the amino acid/s necessary to complete the desired motif (in this case, another ASP to complete the triad). The command will be the following:

```
biometall --motif [ASP,ASP] --mutations [ASP] 1dhy
```

#	Coordinating residues	Coordinates of center	Num. probes	Radius search	Proposed mutations
1	ASP:14:A ASP:57:A	105.322 45.256 16.771	10	2.259	254:A:[['ASP', 6]] 13:A:[['ASP', 5]] 253:A:[['ASP', 4]] 12:A:[['ASP', 4]]
2	ASP:170:A ASP:243:A	76.024 42.225 18.791	8	2.259	169:A:[['ASP', 6]] 193:A:[['ASP', 6]] 242:A:[['ASP', 4]] 186:A:[['ASP', 4]] 194:A:[['ASP', 2]]
3	ASP:67:A ASP:68:A	77.035 65.461 30.915	6	2.475	71:A:[['ASP', 4]] 66:A:[['ASP', 2]] 70:A:[['ASP', 1]]
4	ASP:218:A ASP:219:A	84.107 54.340 -1.414	6	3.031	134:A:[['ASP', 5]] 216:A:[['ASP', 2]] 215:A:[['ASP', 1]]
5	ASP:57:A ASP:58:A	101.281 41.214 21.822	6	1.429	205:A:[['ASP', 5]] 254:A:[['ASP', 2]] 256:A:[['ASP', 2]] 257:A:[['ASP', 2]]
6	ASP:225:A ASP:228:A	92.189 40.204 -2.425	5	4.049	224:A:[['ASP', 3]] 229:A:[['ASP', 2]]
7	ASP:140:A ASP:67:A	78.045 62.430 22.832	1	0.000	3:A:[['ASP', 1]]

The result shows that seven of the eight ASP,ASP motifs could be completed to achieve the aspartate triad. The different mutation options are listed in the “proposed mutations” column, along with the number of probes of each option. As indicated in the BioMetAll article (benchmark and case study 3), the number of probes can be used to prioritize the solutions, in this case being the most favourable ones:

- i) the mutation of residue 254 to aspartate, forming the triad ASP14,ASP57,ASP254
- ii) the mutation of residue 169 to aspartate, forming the triad ASP170,ASP243,ASP169
- iii) the mutation of residue 193 to aspartate, forming the triad ASP170,ASP243,ASP193.

However, you should note that BioMetAll does not consider energetic terms, so the other solutions cannot be discarded beforehand.

Tip: As with motif searches, you can use forward slash notation to incorporate some variability in your search. The notation can be used in the *motif* or *mutations* options, or in both options at the same time. However, you should be aware that the interpretation of the results will increase in complexity with the variability of the search.