

# Sam User Manual - version 2.0.0

[https://mbi-ds4h.loria.fr/softwaretodownload/sam\\_software/](https://mbi-ds4h.loria.fr/softwaretodownload/sam_software/)



## Table des matières

<b>Installing Sam</b> .....	1
<b>Prerequisites</b> .....	1
<b>The Self-Installer</b> .....	1
<b>Running Sam</b> .....	2
<b>1. Assembling Symmetrical Complexes</b> .....	2
<b>2. Controlling the Output</b> .....	5
<b>3. Selecting the Symmetry Type</b> .....	5
<b>4. Defining a Native Reference Structure</b> .....	5
<b>5. Other Sam Parameters</b> .....	8

## Installing Sam

### Prerequisites

Sam currently only works on Linux and Mac systems. In the future, a Windows version might also become available. In order to run Sam on Linux, you will need a relatively recent Linux distribution such as Linux Mint 17 or Ubuntu 14 . The Sam program compiled on 64-bit versions of Mint 17. Because the program was compiled statically, it should work with other Linux distributions. For the Mac, Sam was built on OSX 10.11 (El Capitan), but it might run on earlier versions of OSX.

### The Self-Installer

The easiest way to install Sam is to download and run the self-installer script. Assuming you have a 64-bit system, open a command terminal and enter something like this:

```
sh sam-2.0.0-x64-mint17.bin
```

The self-installer script will ask some questions about where you wish to install *Sam* and whether you will let the script modify one of your shell start-up scripts in order to run the program from the command line. The installer will normally define an environment variable called `SAM_ROOT`, and it will add `${SAM_ROOT}/bin` to your command path. It will also run a test script located in the `${SAM_ROOT}/test` subdirectory. If you do not trust the self-installer script, you can extract and inspect the installation tar file as follows:

```
sh sam-2.0.0-x64-mint17.bin --noexec cd sam-dist-2.0.0 gunzip sam-2.0.0-x64.tgz tar vtf sam-2.0.0-x64.tar
```

You can then create your own installation directory, define the SAM\_ROOT variable, and install the installation files yourself. Please see the file `${SAM_ROOT}/doc/README` for further details.

## Running *Sam*

1. Assembling Symmetrical Complexes
2. Controlling the Output
3. Selecting the Symmetry Type
4. Defining a Native Reference Structure
5. Other *Sam* Parameters

### 1. Assembling Symmetrical Complexes

To make a complex, it is necessary to provide only the PDB file containing the coordinates of one of the monomers or the structure and to give a command line option describing the desired symmetry. The following example will make a C2 homodimer from the monomer given in the `Sam` test directory:

```
cd $SAM_ROOT/test
sam -c2 1m4g_a.pdb
```

Here is the typical output:

```
SAM_CACHE = /home/ritchied/sam_cache
HEX_CACHE = /home/ritchied/sam_cache
hex_cache_mode = 1
HEX_GTO_SCALE = 40
HEX_ETO_SCALE = 1
Sam 2.0.0 starting at Wed Oct 14 15:38:07 2015 on host hardy.
Using LOG file: sam.log
Creating RESULTS directory: ./sam_results/
File [1] = 1m4g_a.pdb
Assuming 1m4g_a.pdb is a PDB file...

Opened PDB file: 1m4g_a.pdb, ID = 1m4g_a
Loaded PDB file: 1m4g_a.pdb, (182 residues, 1733 atoms, 1 models)
Counted 17 +ve and 20 -ve formal charged residues: Net formal charge: -3
>1m4g_a A
MHTQVHTARLVHTADLDSETRQDIRQMVTGAFAGDFTETDWEHTLGGMHALIWHHGAI IAHAAVIQRRLIYRGNALRC
GYVEGVAVRADWRGQRLVLSALLDAVEQVMRGAYQLGALSSARARRLYASRGWLPWHGPTSVLAPTGPVRTPDDDGTV
FVLPIDISLDTSAELMCDWRAGDVW
Calculating SPF coefficients for 1m4g_a.pdb ...

Contouring surface for molecule 1m4g_a.
Polar probe = 1.40A, Apolar probe = 1.40A
Gaussian sampling over 1412 atoms done in 0.17 seconds.
Contoured 101320 triangles (50662 vertices) in 0.07 seconds.
Surface traversal done in 0.03 seconds - Found 1 surface segments.
Primary surface: Area = 7941.41, Volume = 37974.81.
Culled 0 small segments in 0.04 seconds.
Total contouring time: 0.13 seconds.

Sampling surface and interior volumes for molecule 1m4g_a.
Generated 65151 exterior and 47961 interior skin grid cells.
Exterior skin volume = 14072.62; interior skin volume = 10359.58.
Volume sampling done in 0.14 seconds.

Calculating potential to N = 25 (5525 coefficients) using 24 Tasks ...
Grid: 111x111x111 = 1367631 cells (112918 non-zero) of 0.60 Angstroms.
Done integration over 112918 cells in 0.32s (354103/s).

Found 1 chains in starting monomer 1m4g_a.pdb
```

Starting C2 4D search with R12 guess/range = 20.49 / 50.00  
Cn radial to intermolecular scale factor = 0.50

-----  
Docking will output a maximum of 1000 solutions per pair...

-----  
Docking 1 pair of starting orientations...

Docking receptor: lm4g\_a and ligand: lm4g\_a...

Receptor lm4g\_a: Tag = lm4g\_a  
Ligand lm4g\_a: Tag = lm4g\_a

Setting up shape-only correlation.

Starting SPF search.  
Setting docking\_score threshold = 0.0  
Setting 57 distance samples from 0.00 to 44.80, with steps of 0.80.

Total 6D space: Iterate[57,1,1,812,1] x FFT[64] = 2962176.  
Done dl(180.00) matrix to order L=24 in 0.00s.  
Initial rotational increments (N=16) Receptor: 812 (19Mb), Ligand: 812 (19Mb)  
Applying 812+812 coefficient rotations on 24 CPUs for N=16.  
Done 1624 rotations in a total of 0.82s (1970/s).

Starting 4D FFT search using 24 CPUs and 0 GPUs with N=16, Nalpha=64.  
Estart = 2755.53.  
Translation matrix K[16](1.00) done in 0.38 seconds.  
Translation matrix K[16](3.00) done in 0.37 seconds.  
Translation matrix K[16](5.00) done in 0.23 seconds.  
Translation matrix K[16](6.00) done in 0.22 seconds.  
Translation matrix K[16](7.00) done in 0.21 seconds.  
Translation matrix K[16](2.00) done in 0.22 seconds.  
Done 2962176 orientations in 24.56s (120601/s).  
Found 12532/2962176 within score threshold = 0.0 NOT including start guess.

Solution buffer reached 12532/1600000 = 0.8% occupancy with no culling.  
Starting guess not found in top 12532 solutions.  
Emin = -580.86, Emax = -0.00

Re-sampling top 10000 orientations -> top 10000 retained.  
Surviving rotational steps (N=25) Receptor: 736 (63Mb), Ligand: 736 (63Mb)  
Applying 736+736 coefficient rotations on 24 CPUs for N=25.  
Done 1472 rotations in a total of 0.93s (1579/s).

Starting 4D FFT symmetry refinement using 24 CPUs and 0 GPUs with N=25, n\_omega=1.  
Estart = 2578.09.  
Done 640000 orientations in 4.47s (143074/s).  
Found 9948/640000 within score threshold = 0.0 NOT including start guess.

Solution buffer reached 9948/200000 = 5.0% occupancy with no culling.  
Starting guess not found in top 9948 solutions.  
Emin = -618.02, Emax = -0.12

Docking correlation summary by RMS deviation and steric clashes

-----  
Soln Etotal Eshape Eforce Eair RMS Bumps  
-----

Docked structures lm4g\_a:lm4g\_a in a total of 0 min, 31 sec.

-----  
Saving top 1000 orientations.

Docking done in a total of 0 min, 31 sec.

-----  
Clustering found 711 clusters from 1000 docking solutions in 0.14 seconds.

Soln Clust Sym Score Monomer A Centre Monomer B Centre Dist Angle

-----  
1 1 C2 -618.02 -0.0 12.4 0.0 0.0 -12.4 -0.0 24.8 180.0  
6 1 C2 -557.05 -0.0 12.8 0.0 0.0 -12.8 -0.0 25.6 180.0  
14 1 C2 -515.93 -0.0 12.0 0.0 0.0 -12.0 -0.0 24.0 180.0  
155 1 C2 -385.72 -0.0 13.2 0.0 0.0 -13.2 -0.0 26.4 180.0

```

  2 2 C2 -597.16 -0.0 13.6 0.0 0.0 -13.6 -0.0 27.2 180.0
  3 2 C2 -583.05 -0.0 14.0 0.0 0.0 -14.0 -0.0 28.0 180.0
106 2 C2 -410.54 -0.0 13.2 0.0 0.0 -13.2 -0.0 26.4 180.0
  4 3 C2 -569.53 -0.0 14.0 0.0 0.0 -14.0 -0.0 28.0 180.0
 19 3 C2 -499.43 -0.0 13.6 0.0 0.0 -13.6 -0.0 27.2 180.0
 22 3 C2 -493.42 -0.0 14.4 0.0 0.0 -14.4 -0.0 28.8 180.0
  5 4 C2 -565.24 -0.0 15.6 0.0 0.0 -15.6 -0.0 31.2 180.0
  7 4 C2 -555.98 -0.0 15.2 0.0 0.0 -15.2 -0.0 30.4 180.0
 65 4 C2 -430.41 -0.0 16.0 0.0 0.0 -16.0 -0.0 32.0 180.0
  8 5 C2 -552.20 -0.0 13.2 0.0 0.0 -13.2 -0.0 26.4 180.0
 16 5 C2 -504.90 -0.0 13.6 0.0 0.0 -13.6 -0.0 27.2 180.0
 99 5 C2 -412.86 -0.0 12.8 0.0 0.0 -12.8 -0.0 25.6 180.0
(... output truncated ...)

```

-----

Summary of the top 20 clustered solutions...

-----

Cn-Soln	E-total	RMS-B	RMS-D	RMS-TOT
---------	---------	-------	-------	---------

-----

1	-618.02	-1.00	0.00	-1.00
2	-597.16	-1.00	0.00	-1.00
3	-569.53	-1.00	0.00	-1.00
4	-565.24	-1.00	0.00	-1.00
5	-552.20	-1.00	0.00	-1.00
6	-537.12	-1.00	0.00	-1.00
7	-534.48	-1.00	0.00	-1.00
8	-524.68	-1.00	0.00	-1.00
9	-506.46	-1.00	0.00	-1.00
10	-502.99	-1.00	0.00	-1.00
11	-500.37	-1.00	0.00	-1.00
12	-493.09	-1.00	0.00	-1.00
13	-487.67	-1.00	0.00	-1.00
14	-486.01	-1.00	0.00	-1.00
15	-482.72	-1.00	0.00	-1.00
16	-482.33	-1.00	0.00	-1.00
17	-479.04	-1.00	0.00	-1.00
18	-476.74	-1.00	0.00	-1.00
19	-472.40	-1.00	0.00	-1.00
20	-472.01	-1.00	0.00	-1.00

-----

```

Writing: ./sam_results/lm4g_a_00001_C2.pdb
Writing: ./sam_results/lm4g_a_00002_C2.pdb
Writing: ./sam_results/lm4g_a_00003_C2.pdb
Writing: ./sam_results/lm4g_a_00004_C2.pdb
Writing: ./sam_results/lm4g_a_00005_C2.pdb
Writing: ./sam_results/lm4g_a_00006_C2.pdb
Writing: ./sam_results/lm4g_a_00007_C2.pdb
Writing: ./sam_results/lm4g_a_00008_C2.pdb
Writing: ./sam_results/lm4g_a_00009_C2.pdb
Writing: ./sam_results/lm4g_a_00010_C2.pdb
Writing: ./sam_results/lm4g_a_00011_C2.pdb
Writing: ./sam_results/lm4g_a_00012_C2.pdb
Writing: ./sam_results/lm4g_a_00013_C2.pdb
Writing: ./sam_results/lm4g_a_00014_C2.pdb
Writing: ./sam_results/lm4g_a_00015_C2.pdb
Writing: ./sam_results/lm4g_a_00016_C2.pdb
Writing: ./sam_results/lm4g_a_00017_C2.pdb
Writing: ./sam_results/lm4g_a_00018_C2.pdb
Writing: ./sam_results/lm4g_a_00019_C2.pdb
Writing: ./sam_results/lm4g_a_00020_C2.pdb

```

Peak memory allocation: 217 Mb.  
Total memory on exit: 1 Mb.  
Sam finished in a total of 32.10 seconds (0.54 minutes).

The main values to consider here are the columns labeled "Score" and "E-total". These are the calculated pseudo-energies (in units of KJ/mol), such that large negative values represent favourable docking scores. It can also be seen that the calculated models (the "Solutions" column) are automatically grouped into spatially close clusters (which are identified numerically in the "Clust" column). The best (lowest energy) solution from each cluster is then selected as a representative model to be written out in the final results list (... which are also called "solutions").

## 2. Controlling the Output

The above output shows that *Sam* has created a directory called `$SAM_ROOT/test/sam_results/`, in which it has written its top 20 models, with each model being written to a separate PDB file. Using a separate results directory helps to avoid filling up your current working directory with large numbers of results files. If you really want the output files to appear in your current working directory (not recommended) you can set the `SAM_RESULTS` environment variable to `"."`

It is also possible to write all of the models to a single PDB file using the `"-unified"` option:

```
sam -unified -c2 1m4g_a.pdb
```

## 3. Selecting the Symmetry Type

*Sam* can build complexes for any of usual point group symmetries associated with protein structures, i.e. cyclic ( $C_n$ ), dihedral ( $D_n$ ), tetrahedral (T), octahedral (O), and icosahedral (I). As you might have already guessed from the above example, the symmetry type is selected using a command line option like `"-cN"`, where  $N$  may take any value between 2 and 12. Similarly, the option to request a dihedral symmetry group is `"-dN"`. If you really want a higher order cyclic or dihedral symmetry, you can put `"-c N"` or `"-d N"`, where  $N$  has any value greater or equal to 2. Finally, to request a tetrahedral, octahedral, or icosahedral structure, you should specify `"-t"`, `"-o"`, or `"-i"`, respectively. For example, the command:

```
sam -unified -i 1m4g_a.pdb
```

will generate a single PDB file of multiple icosahedral structures, with each structure being written as a separate "MODEL" in the PDB format.

## 4. Defining a Native Reference Structure

One way to benchmark *Sam* is to compare the calculated models with a given crystallographic, or "native", structure using the `"-native"` option:

```
sam -c2 1m4g_a.pdb -native 1m4g.pdb
```

When given a native reference structure, *Sam* will align and superpose the starting ("A") monomer onto one of the monomers in the reference structure, and, for each generated model, *Sam* will calculate the root-mean-squared deviation (RMSD) between the position of one of the source monomer's symmetry mates (the "B" monomer) and the corresponding monomer in the native structure. The resulting output (somewhat truncated) is shown below:

```
SAM_CACHE = /home/ritchied/sam_cache
HEX_CACHE = /home/ritchied/sam_cache
hex_cache_mode = 1
HEX_GTO_SCALE = 40
HEX_ETO_SCALE = 1
Sam 2.0.0 starting at Wed Oct 14 16:41:09 2015 on host hardy.
Using LOG file: sam.log
Using RESULTS directory: ./sam_results/
File [1] = 1m4g_a.pdb
Assuming 1m4g_a.pdb is a PDB file...

Opened PDB file: 1m4g_a.pdb, ID = 1m4g_a
Loaded PDB file: 1m4g_a.pdb, (182 residues, 1733 atoms, 1 models)
```

Counted 17 +ve and 20 -ve formal charged residues: Net formal charge: -3  
>lm4g\_a A  
MHTQVHTARLVHTADLDSETRQDIRQMVTGAFAGDFTETDWEHTLGGMHALIWHHGAI IAHAAVIQRRLIYRGNALRCG  
GYVEGVAVRADWRGQRLVLSALLDAVEQVMRGAYQLGALSSARARRLYASRGWLPWHGPTSVLAPTGPV RTPDDDGTV  
FVLPIDISLDTS AELMCDWRAGDVW  
Calculating SPF coefficients for lm4g\_a.pdb ...

Contouring surface for molecule lm4g\_a.  
Polar probe = 1.40A, Apolar probe = 1.40A  
Gaussian sampling over 1412 atoms done in 0.17 seconds.  
Contoured 101320 triangles (50662 vertices) in 0.08 seconds.  
Surface traversal done in 0.03 seconds - Found 1 surface segments.  
Primary surface: Area = 7941.41, Volume = 37974.81.  
Culled 0 small segments in 0.04 seconds.  
Total contouring time: 0.16 seconds.

Sampling surface and interior volumes for molecule lm4g\_a.  
Generated 65151 exterior and 47961 interior skin grid cells.  
Exterior skin volume = 14072.62; interior skin volume = 10359.58.  
Volume sampling done in 0.17 seconds.

Calculating potential to N = 25 (5525 coefficients) using 24 Tasks ...  
Grid: 111x111x111 = 1367631 cells (112918 non-zero) of 0.60 Angstroms.  
Done integration over 112918 cells in 0.37s (308967/s).

Found 1 chains in starting monomer lm4g\_a.pdb  
Found 2 chains in native structure lm4g.pdb

```
#####  
Alignment: lm4g + lm4g_a (*=aligned, $=matched, @/#=aligned/matched identity)  
Nquery: 181  
Ntarget: 181  
Ncolumns: 181  
Naligned: 181  
Nmatched: 181  
Nidentity: 181 (100.00%)  
RMSD-aligned: 0.00  
RMSD-matched: 0.00  
RMSD-anchor: 0.00  
MaxDist: 0.00  
Contact: 8.00  
#####
```

```
MHTQVHTARLVHTADLDSETRQDIRQMVTGAFAGDFTETDWEHTLGGMHALIWHHGAI IAHAAVIQRRLIYRGNALRCGY  
MHTQVHTARLVHTADLDSETRQDIRQMVTGAFAGDFTETDWEHTLGGMHALIWHHGAI IAHAAVIQRRLIYRGNALRCGY  
#####  
VEGVAVRADWRGQRLVLSALLDAVEQVMRGAYQLGALSSARARRLYASRGWLPWHGPTSVLAPTGPV RTPDDDGTVFVLP  
VEGVAVRADWRGQRLVLSALLDAVEQVMRGAYQLGALSSARARRLYASRGWLPWHGPTSVLAPTGPV RTPDDDGTVFVLP  
#####  
IDISLDTS AELMCDWRAGDVW  
IDISLDTS AELMCDWRAGDVW  
#####
```

```
#####  
Fitted undocked monomer (181 CA atoms) to native structure with RMSD = 0.00  
Setting up 2 reference poses by fitting 1 reference chains...  
#####  
Alignment: lm4g + lm4g (*=aligned, $=matched, @/#=aligned/matched identity)  
Nquery: 176  
Ntarget: 181  
Ncolumns: 181  
Naligned: 176  
Nmatched: 176  
Nidentity: 176 (100.00%)  
RMSD-aligned: 0.36  
RMSD-matched: 0.36  
RMSD-anchor: 0.36  
MaxDist: 1.83  
Contact: 8.00  
#####
```

```
----HTARLVHTADLDSETRQDIRQMVTGAFAGDFTETDWEHTLGGMHALIWHHGAI IAHAAVIQRRLIYRGNALRCGY  
MHTQVHTARLVHTADLDSETRQDIRQMVTGAFAGDFTETDWEHTLGGMHALIWHHGAI IAHAAVIQRRLIYRGNALRCGY  
#####
```

```
VEGVAVRADWRGQRLVSALLDAVEQVMRGAYQLGALSSARARRLYASRGWLPWHGPTSVLAPTGPVRTPDDDGTVFVLP
VEGVAVRADWRGQRLVSALLDAVEQVMRGAYQLGALSSARARRLYASRGWLPWHGPTSVLAPTGPVRTPDDDGTVFVLP
EEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE
```

```
IDISLDTSAELMCDWRAGDVW
IDISLDTSAELMCDWRAGDVW
EEEEEEEEEEEEEEEEEEEEEEEE
```

```
#=====
```

```
Starting C2 4D search with R12 guess/range = 20.49 / 50.00
Cn radial to intermolecular scale factor = 0.50
```

```
(... output truncated...)
```

```
-----
```

```
Summary of the top 20 clustered solutions...
```

```
-----
```

```
Cn-Soln E-total RMS-B RMS-D RMS-TOT
```

```
-----
```

1	-618.02	2.12	0.00	2.12
2	-597.16	30.03	0.00	30.03
3	-569.53	37.43	0.00	37.43
4	-565.24	42.76	0.00	42.76
5	-552.20	26.63	0.00	26.63
6	-537.12	21.13	0.00	21.13
7	-534.48	33.41	0.00	33.41
8	-524.68	4.21	0.00	4.21
9	-506.46	3.77	0.00	3.77
10	-502.99	27.84	0.00	27.84
11	-500.37	41.27	0.00	41.27
12	-493.09	38.94	0.00	38.94
13	-487.67	36.38	0.00	36.38
14	-486.01	37.36	0.00	37.36
15	-482.72	30.42	0.00	30.42
16	-482.33	41.46	0.00	41.46
17	-479.04	42.61	0.00	42.61
18	-476.74	27.26	0.00	27.26
19	-472.40	33.00	0.00	33.00
20	-472.01	3.15	0.00	3.15

```
-----
Writing top 20 solutions to ./sam_results/lm4g_a_unified_C2.pdb
```

```
Peak memory allocation: 210 Mb.
Total memory on exit: 0 Mb.
Sam finished in a total of 32.01 seconds (0.53 minutes).
```

In this example, the first model produced by Sam contains a docking partner ("B" monomer) which is 2.12 Angstroms RMSD of the correct crystallographic solution. In this particular example, such a good result is not especially surprising because the starting ("A") monomer (1m4g\_a.pdb) was manually extracted from the solution (1m4g.pdb). Thus, there are no conformational differences between the starting structure and the expected solution. On the other hand, re-assembling a known structure in this way serves to verify that the docking calculation has been implemented correctly.

For structures having more symmetry axes than Cn, it is possible to identify two different monomer interfaces with respect to the "A" monomer. For example, a Dn structure may be considered as being made from two Cn ring systems, so that each monomer can be considered to have an in-ring (A-B) interface with one of its Cn neighbours, and also a between-ring (A-D) interface with a monomer in the other ring system. Therefore, for such cases, the "RMS-D" column reports the RMSD of this secondary interaction, and the "RMSD-TOT" columns reports the RMSD of the "B" and "D" monomers taken together (the labels "A", "B", and "D" come from the paper that describes the method in full).

## 5. Other Sam Parameters

Sam has several parameters which control its docking calculations. These mostly correspond to options for the Hex polar Fourier docking code that Sam uses. To see a list and a brief description of all command line options, please use:

```
sam -help
```

## Contact

Sjoerd de Vries  
SISR LORIA  
Campus Scientifique  
BP239  
54500 Vandoeuvre-lès-Nancy  
sjoerd.devries AT loria.fr

