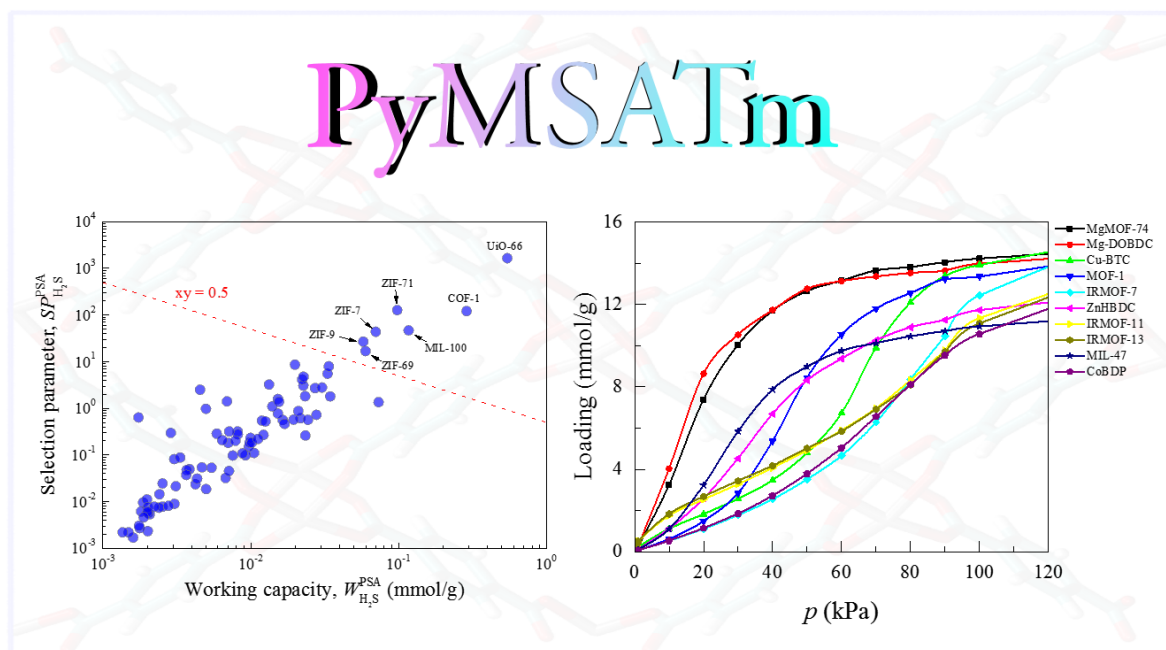


PyMSATm: Python Multipurpose Simulation Analysis Tools Package for MuSiC Software 用户手册



武汉工程大学化工与制药学院

周旭淼 卢嘉峰 杨犁

Email: liyang@wit.edu.cn

中国.湖北.武汉.东湖新技术开发区光谷一路 206 号

2019 年 3 月

目 录

1. 软件简介	4
2. 软件运行要求	5
3. 软件功能	6
3.1 TranslateFormat.py 介绍.....	6
3.2 MuSiCStream.py 介绍	10
3.3 ExtractData.py 介绍.....	14
4. 参数设定	17
4.1 TranslateInfo 格式.....	17
ReadFilePath (绝对路径, 字符串)	17
OutputPath (绝对路径, 字符串)	17
CalculateCharge (yes/no)	18
ImportFormat (cif/pdb)	18
ExportFormat (cif/pdb/mol)	18
ElementSuffix (字符串)	18
4.2 GlueParameters 格式	19
MakeGCMC (open/off)	19
MakePmap (open/off)	20
MakeEmap (open/off)	20
MakeTorque (open/off)	20
UseChargesFromPDBFile (yes/no)	21

ExtractEnergyData (yes/no)	21
InputPath (绝对路径, 字符串)	21
OutputPath (绝对路径, 字符串)	21
AtomParameterPath (绝对路径, 字符串)	21
GasType (列表, 字符串)	21
GasAtomTypeNum (列表, 字符串)	22
GasAtomType (列表, 字符串)	22
Multiple (整数)	22
CutOff (浮点数)	22
GasPartialPressure (列表, 字符串)	22
TemperatureList (列表, 浮点数)	22
PressureList (列表, 浮点数)	23
BalanceStep (整数)	23
CollectionStep (整数)	23
GridSpacingP (浮点数)	23
HighEndPotentialCutoffP (整数)	23
GridSpacingE (浮点数)	23
HighEndPotentialCutoffE (整数)	23
Nodes (字符串) *	24
TaskSuffix (字符串) *	24
TorqueSetting (字符串) *	24
MuSiCSetting (字符串) *	24

4.3 AtomParameter 格式.....	25
5. 扩展程序	26
5.1 dmap2plt.py.....	26
5.2 ZeoReplaceSi2Al.py.....	26
6. 更新与修订	28
PyMSATm 2.0 版本（2019 年 3 月）	28
参考文献.....	29
软件运行总流程图	30

1. 软件简介

PyMSATm 软件是一款预处理和结果提取分析软件，主要针对多目的分子模拟程序 MuSiC 软件编写，目的是在巨正则系综下高通量模拟筛选吸附剂对吸附质微观现象。此程序极大的方便了在进行高通量计算模拟筛选时所消耗的人工操作时间和复杂度。

PyMSATm 软件包的主要基于以下想法：

- 晶体文件的格式转化

晶体的结构模型一般都采用 CIF 格式进行保存，而这个格式不是 MuSiC 软件可以使用的格式。于是对于高通量筛选，把大量晶体文件的格式转换为 MuSiC 软件可读的文件格式是使用 MuSiC 软件进行模拟的第一步。

- 模拟文件的准备

计算机模拟需要输入大量的模拟参数和需求，这些内容通过写入各种形式的模拟准备文件传递给计算机进行读取。对于高通量筛选，其中大量准备文件几乎都是类似的，于是使用一个程序快速的生成这些文件是十分必要的。

- 模拟结果的提取和统计

计算模拟的结果是通过写入文件并返回文件到模拟文件夹的形式提供给人阅读。在巨正则系综下，即使只需要一个晶体的一个温度和一个压力下的吸附结果也会涉及多个文件下多个数据的查找。于是减少重复的结果查找和统计是一个后期极为重要的过程。

对于以上想法的实现，使用 Python 分别编写了三个独立程序以实现上述功能。三个程序分别以 TranslateFormat.py、MuSiCStream.py 和 ExtractData.py 命名，其中 MuSiCStream.py 为主程序，也是最重要的核心程序。

2. 软件运行要求

此 PyMSATm 软件采用 Python3.5 编写和测试的，因此建议使用 Python3.5 或以上版本来运行所有程序。

PyMSATm 软件主要是为了配合 MuSiC 软件而生成 MuSiC 软件所需的必要文件，因此计算服务器上必须装有 MuSiC 软件。关于 MuSiC 软件的安装与使用可以参考 Tina Düren 课题组的官方使用手册：<http://people.bath.ac.uk/td222/research/music/Tutorial/>。

在 TranslateFormat.py 程序中使用了 pybel 库，这个库是使用这个程序的必要函数库。在 MuSiCStream.py 和 ExtractData.py 程序中使用了 openpyxl 库，但这个库并不是必要的。如果没有安装 openpyxl 库，并不会影响这两个程序的运行，但是数据的统计结果不会输出 Excel 文件，关于 openpyxl 库的更多详情请参考：<http://openpyxl.readthedocs.io/en/default/index.html>。

PyMSATm 软件在 Windows 操作系统和 Linux 操作系统下均可运行，并在以上操作系统环境下测试运行成功。但是其中 TranslateFormat.py 程序需要调运特定的 pybel 库^[1]，并且需要计算机上装有 Open Babel 软件才能使用。因为在 Windows 下 Open Babel 的安装和 pybel 库的调用是方便的，所以 TranslateFormat.py 程序的设计初衷是在 Windows 上运行，如果需要移植到 Linux 环境请参考 Open Babel 在 Linux 的安装和使用，官方手册：<http://open-babel.readthedocs.io/en/latest/index.html>。

3. 软件功能

PyMSATm 软件的运行需要如下几个输入文件: TranslateInfo、GlueParameters 和 AtomParameter。这几个文件的名字并没有后缀以防止错误的打开方式, 同时我们强制使用这几个文件名以防止错误的读取信息, 这意味程序只会识别这几个特殊关键词。其中 TranslateInfo 文件是 TranslateFormat.py 程序的信息文件。当运行 TranslateFormat.py 程序时, 该程序只会在当前文件夹路径下搜索 TranslateInfo 文件并读取其中的信息。MuSiCStream.py 和 ExtractData.py 程序则同样通过在当前文件夹下搜索 GlueParameters 文件读取必要信息。此外, MuSiCStream.py 程序还需要 AtomParameter 文件以获取必要的力场参数信息。为了区分不同的立场参数的使用, 把相同名字但不同内容的 AtomParameter 文件放置在不同文件夹中加以区分是十分必要的。因此, MuSiCStream.py 程序会根据 GlueParameters 文件里的参数去指定文件夹中找 AtomParameter 文件。

3.1 TranslateFormat.py 介绍

TranslateFormat.py 程序是整个软件的预备程序, 主要用于大批量材料格式的转化, 但该程序只是一个扩展程序并不是整个软件运行的必要程序, 因此不使用该程序并不会影响后面程序的使用。该程序会在当前路径下寻找名为 TranslateInfo 的文件并从其中读取必要的参数信息, 详细参数内容请看下一章节(参数设定)。

MuSiC 程序中材料原子坐标以一种后缀名为 mol 的非标准格式保存, 这种格式并非标准格式, 仅仅是一种 MuSiC 程序中的内定格式, 如图 1。对于少量材料的模拟, 通过 Materials Studio 等软件手动书写材料的 mol 文件是可以接受的, 但是对于成千上万种材料的计算筛选而言, 这明显是不现实的。此外, 标准格式的转化过程需反复调用 pybel 库, 这很有可能因为缺少库而导致主程序 (MuSiCStream.py) 运行失败。为了极大的增加主程序的通用性, 采用从主程序中分离

TranslateFormat.py 程序的方式来减少主程序对库的依赖性。TranslateFormat.py 程序的主要目的就是方便快捷的预处理标准格式，使之可以被主程序使用。

```
# Basic Molecule Information
# Created by PyMSATm at 2018-06-19
Molecule_name: COF108 CHARGED

Coord_Info: Listed Cartesian Rigid
255
1      19.069    24.27    2.402    O_m -0.521    0    0
2      18.428    25.479    2.922    B_m  1.191    0    0
3      17.24     26.204    2.197    C_m -0.567    0    0
.....
254     22.403    25.891    7.441    H_m  0.066    0    0
255     22.403     2.51    20.96    H_m  0.066    0    0

Fundcell_Info: Listed
28.40100  28.40100  28.40100
90.00000  90.00000  90.00000
0.00000  0.00000  0.00000
28.40100  28.40100  28.40100
```

图 1 MuSiC 软件适用的 mol 格式文件范例

首先，简单介绍一下两种标准格式：CIF 格式和 PDB 格式。国际对于晶体材料的描述上通常采用的是 CIF（Crystallographic Information File）格式，如图 2。详细 CIF 格式用法和书写规则可以参考其他文献，或去官方网站查询使用手册：<http://www.iucr.org/>。这里需要强调的是，CIF 格式所描述的晶体结构大多会使用空间对称的方式来描述晶胞中原子的坐标，以避免大量原子坐标的描述。但是在计算机模拟中，大部分软件必须知道整个晶胞中每一个原子所在的具体位置。这种通过空间对称还原晶胞中每个原子坐标的方式叫做“Make P1”，其中 P1 就是指无空间对称即描述全部原子位置。此外，CIF 格式中原子坐标采用的是分数坐标，简单讲就是原子在晶胞中的相对坐标，因原子不可能出现在晶胞外，所以原子的相对坐标不会大于 1。


```

data_Mg-DOBDC

_audit_creation_method RASPA-1.0
_audit_creation_date 2012-1-5
_audit_author_name 'Ozgur Yazaydin'

_cell_length_a    26.136
_cell_length_b    26.136
_cell_length_c    6.942
_cell_angle_alpha 90
_cell_angle_beta  90
_cell_angle_gamma 120

_symmetry_cell_setting    trigonal
_symmetry_space_group_name_Hall '-R 3'
_symmetry_space_group_name_H-M 'R -3'
_symmetry_Int_Tables_number    148

loop_
_symmetry_equiv_pos_as_xyz
'x,y,z'
.....
'x-y+1/3,x+2/3,-z+2/3'

loop_
_atom_site_label
_atom_site_type_symbol
_atom_site_fract_x
_atom_site_fract_y
_atom_site_fract_z
Mg1      Mg      0.64159      0.68514      0.47189
.....
H1       H       0.16529      0.85803      0.47918

```

图 2 标准晶体学的 CIF 文件范例

PDB (Protein Data Bank) 格式通常用于描述蛋白质结构，是一种生物学中的文件标准格式，

如图 3。PDB 文件中每个原子都有单独的坐标并且采用笛卡尔坐标进行表述。

```

REMARK    PyMSATm PDB file
REMARK    Created: 2018-04-26
CRYST1    28.401    28.401    28.401    90.00    90.00    90.00 P1
ATOM       1 O_m    MOL    2      19.069    24.270    2.402    0.00      O
ATOM       2 B_m    MOL    2      18.428    25.479    2.922    0.00      B
ATOM       3 C_m    MOL    2      17.240    26.204    2.197    0.00      C
ATOM       4 C_m    MOL    2      15.136    27.560    0.841    0.00      C
ATOM       5 C_m    MOL    2      22.065    23.074    4.311    0.00      C
ATOM       6 C_m    MOL    2      16.724    27.399    2.706    0.00      C
.....
ATOM      250 H_m    MOL    2       7.441    22.403    25.891    0.00      H
ATOM      251 H_m    MOL    2      20.960    22.403    2.510    0.00      H
ATOM      252 H_m    MOL    2       5.998    25.891    20.960    0.00      H
ATOM      253 H_m    MOL    2       5.998    2.510    7.441    0.00      H
ATOM      254 H_m    MOL    2      22.403    25.891    7.441    0.00      H
ATOM      255 H_m    MOL    2      22.403    2.510    20.960    0.00      H
TER
END

```

图 3 标准生物学的 PDB 文件范例

本程序通过引入 pybel 库来与 Open Babel 软件中的 Python 接口进行对接以实现 CIF 文件的

“Make P1”过程，这也是此程序的主要目的。CIF 文件中的分数坐标转化为 mol 文件中的笛卡尔坐标的过程也是通过 Open Babel 软件实现。其中的矩阵变化过程对于非正交晶胞来说是十分便利的。此外，本程序还可以通过 Open Babel 软件采用 EQeq 算法进行材料原子的部分电荷分配，关于 EQeq 算法的详细内容请参考相关文献^[2]。如果使用者想采用其它分配电荷的方法也可以选择关闭电荷计算，只进行相关文件的格式转化。总的来说，此程序可以把一般的 CIF 和 PDB 格式转化为已分配电荷（可选）且“Make P1”了的 CIF、PDB 和 mol 格式。其中由于 PDB 格式没有书写原子电荷的位置，程序会默认不给 PDB 分配电荷。其中转化后的 mol 格式可以用于下一步主程序，也可以直接用于 MuSiC 程序。

此外需要强调两点：第一，“Make P1”的过程可能导致错误的晶体结构，如重复的对称计算，所以模拟结果出现异常时需通过可视化软件检查晶体的结构正确性，如对比生成的 CIF 和 PDB 文件。第二，PDB 文件中给予原子电荷描述的字符长度只有两个字符，因此计算出的原子电荷将写在用于描述原子占用大小的地方，即 ATOM 行中第 55-60 的字符位置。这同样意味着，如果使用者想通过其它方式给材料分配电荷，则计算出的电荷同样需要写在那个位置，否则主程序将错误的读取原子电荷。因此我们强烈建议转化成 mol 文件提交给主程序读取。

为了方便读者理解，这里给出 TranslateFormat.py 的运行流程图来加强使用者对程序的理解，如图 4 所示。

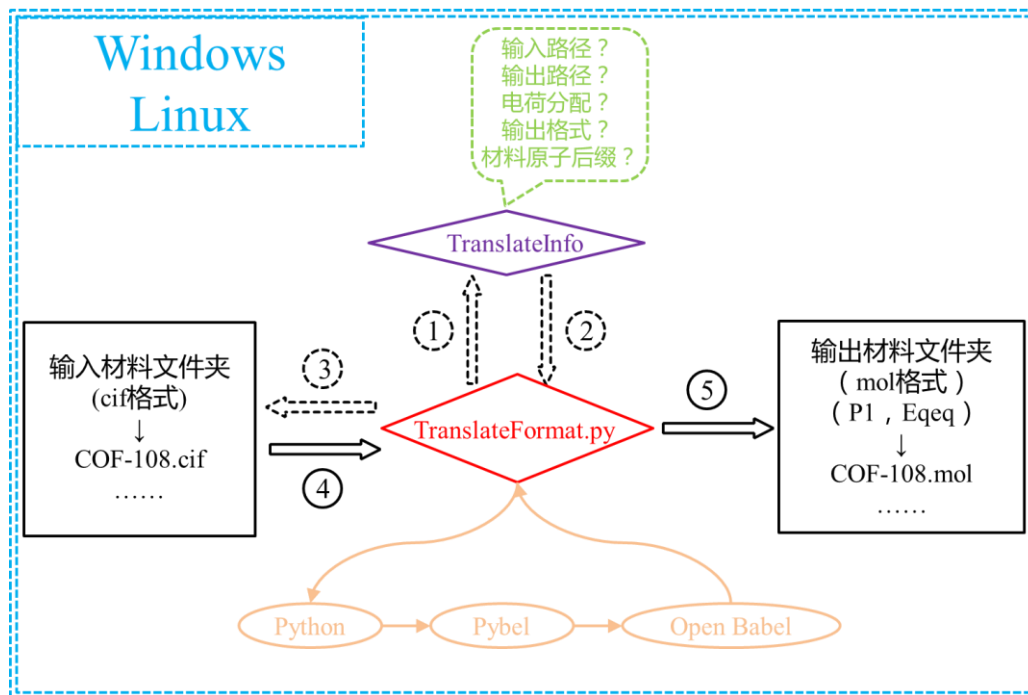


图 4 TranslateFormat.py 运行流程图

3.2 MuSiCStream.py 介绍

MuSiCStream.py 为整个软件的核心程序，主要完成模拟参数设定等一系列计算文件的配置工作，它控制着整个高通量计算筛选的所有过程，包括制作每个晶体 `emap` 和 `pmap` 的过程和 `gcmc` 模拟的过程。此外，因为该程序涉及 MuSiC 程序的环境设置，因此建议该程序在 Linux 系统下进行调用。MuSiCStream.py 程序会读取当前文件夹下名为 `GlueParameters` 的文件，并根据文件中的参数配置整个高通量计算筛选的过程。其次，`GlueParameters` 文件会指出 `AtomParameter` 文件的路径，并根据路径找到 `AtomParameter` 文件并读取其中的原子之间作用力的参数。`GlueParameters` 文件和 `AtomParameter` 文件中的关键字意义与参数设置要求请参考下一章——参数设定。MuSiCStream.py 程序运行后会把生成的文件分类放入自动生成的不同文件夹中，并根据输出路径把这些文件夹返回至同一文件夹中，如图 5 所示：

```
[zxm@gibbs output]$ ls -lthr
total 28K
drwxr-xr-x 2 zxm yangli 4.0K Oct 3 01:56 Statistics
drwxr-xr-x 2 zxm yangli 4.0K Oct 3 01:56 Atoms
drwxr-xr-x 3 zxm yangli 4.0K Oct 3 01:56 MakePmap
drwxr-xr-x 2 zxm yangli 4.0K Oct 3 01:56 Mols
drwxr-xr-x 5 zxm yangli 4.0K Oct 3 01:56 MakeEmap
drwxr-xr-x 3 zxm yangli 4.0K Oct 3 02:32 GCMC
```

图 5 自动生成文件夹类型

首先简短介绍一下什么是 **emap** 和 **pmap** 以及为什么要为晶体制作 **emap** 和 **pmap**。以 **emap** 为例（**pmap** 是一样的原理，只不过是计算范德华作用力），当晶体的原子都固定不动且每个原子携带的电荷是不变时，可以把晶胞在三维方向每隔一定长度设置一个坐标让带一个正电荷的探针原子尝试性地放到晶胞中的每一个坐标上计算此时探针原子受到的静电作用力。于是，可以得到一个记录着晶胞静电作用力的“势能地图”，其势能地图为一个三维的数组。当在进行 **GCMC** 模拟时，程序可以直接从“势能地图”中读取静电作用力而不用花费计算时间在重复计算上。这是一个通过牺牲存储空间来加快计算速度的方式。**MuSiCStream.py** 程序在配置模拟过程时可以选择是否计算 **map** 来加速计算，如果采用 **map** 加速则需要单独在 **MakeEmap** 和 **MakePmap** 中提交计算 **map** 的任务，并在计算完成后把生成的 *****.emap** 和 *****.pmap** 文件复制到 **Map** 文件夹下等待 **GCMC** 模拟的调用（例如使用 **linux** 下指令：**mv MakeEmap/*/*.emap Map**）。提醒一点，由于 **MuSiC** 软件不支持给非正交晶胞制作 **map**，所以程序会跳过给非正交晶胞制作 **map**，在进行 **GCMC** 模拟时所有的非正交晶胞都只能使用非 **map** 模式运行。

Statistics 文件夹用于存放进行高通量计算筛选时晶体信息的统计，如果 **Python** 中装有 **openpyxl** 库，则会输出文本格式和 **Excel** 格式两种统计文件（文件名为：**MaterialsInfo**），否则只会有文本格式的统计文件生成。晶体名可能过长或带有“（）”之类的字符会导致模拟出错或发生一些不可控结果，所以在统计的过程中，程序会自动删去特殊字符并给过长的晶体名分配一个 **ID** 序号作为模拟过程中的晶体名。两种统计文件如图 6：

Data: 2018-06-19
Materials
COF-108
hypotheticalMOF_7003007_i_4_j_29_k_7_m_3
MgMOF-74
NU-2(Pyr_4TT)_Por_4P-ftw
ZIHTEP_clean
TotElements: B,Br,C,Co,H,Li,Mg,N,O,Zr

Mol_ID
COF108
hypotheticalMOF_7003_ID_1
MgMOF74
NU2Pyr_4TT_Por_4Pftw
ZIHTEP_clean

AtomsNumber
255
218
162
286
180

CellAngle
90.0000,90.0000,90.0000
56.0000,87.0300,62.4400
90.0000,90.0000,120.0000
90.0000,90.0000,90.0000
94.6400,97.4400,103.0000

CellLength
28.4010,28.4010,28.4010
31.6400,20.8760,20.8760
26.1360,26.1360,6.9420
19.6220,19.2800,19.3140
8.1180,11.5680,14.2440

SuperCell
1,1,1
1,2,2
2,2,4
2,2,2
4,3,2

Elements
H,C,O,B
N,O,Br,H,C,Zr
H,C,O,Mg
H,C,N,O,Zr
N,O,H,Co,C,Li

	A	B	C	D	E	F	G
1	Data:2018-06-19						
2	Materials	Mol_ID	AtomsNumber	CellAngle	CellLength	SuperCell	Elements
3	COF-108	COF108	255	28.4010 28.4010 28.4010	90.0000 90.0000 90.0000	1 1 1	H C O B
4	hypotheticalMOF_7003007_i_4_j_29_k_7_m_3	hypotheticalMOF_7003_ID_1	218	31.6400 20.8760 20.8760	56.0000 87.0300 62.4400	1 2 2	N O Br H C Zr
5	MgMOF-74	MgMOF74	162	26.1360 26.1360 6.9420	90.0000 90.0000 120.0000	2 2 4	H C O Mg
6	NU-2(Pyr_4TT)_Por_4P-ftw	NU2Pyr_4TT_Por_4Pftw	286	20019.6220 19.2800 19.3140	90.0000 90.0000 90.0000	2 2 2	H C N O Zr
7	ZIHTEP_clean	ZIHTEP_clean	180	8.1180 11.5680 14.2440	94.6400 97.4400 103.0000	4 3 2	N O H Co C Li
8	TotElements: B Br C Co H Li Mg N O Zr						
9							

图 6 晶体命名和参数示意图

Atoms 文件夹用于存放程序生成的原子信息文件，该文件主要给原子提供参考质量。程序会根据统计的晶体和气体元素自动生成 MuSiC 模拟所需的原子信息文件。需要注意的是当要为晶体生成 emap 时，程序会多生成一个 Probe.atm 文件用于 emap 的计算。原子信息文件如下：

```
# Basic Atom Information
Atom_Name:      Cu_m
Atom_Symbol:     Cu
Atom_SS_Charge:  0
Atom_SZ_Charge:  0
Atom_Mass:       63.546
Atom_Valency:    0
```

图 7 MuSiC 定义 atm 文件格式

Mols 文件夹用于存放程序生成的分子信息文件，其中包括晶体信息文件（自动生成），气体信息文件（手动添加）和 pmap 所需的探针信息文件（自动生成）。因为不同气体模型对应的分子信息文件不一样(如硫化氢模型可以描述为三点位或四点位)，所以气体分子的文件需要手动准备，如图 8 为三点位模型的硫化氢分子的范例。如果需要给晶体制作 pmap，本程序还会生成气体原子探针的信息文件，如 S_h2s.mol。如果需要给晶体制作 emap，程序还会生成名为 Probe.mol 的电子探针信息文件。

```
##### Basic Molecule Information
# Created by Yang Li at 05-Apr-2016 21:51:33

Molecule_Name: H2S

Coord_Info: Listed Cartesian Rigid
3 # number of atoms in molecule
1 1.931000 0.005000 0.001000 H_h2s 0.160000 0 0 # x,y,z,Name,Charge,set,type
2 0.967000 0.936000 0.001000 S_h2s -0.320000 0 0 # x,y,z,Name,Charge,set,type
3 0.003000 0.005000 0.001000 H_h2s 0.160000 0 0 # x,y,z,Name,Charge,set,type

Molecule_DOF: 6
```

图 8 MuSiC 定义分子文件格式

介绍一下 GCMC 文件夹里的文件层次。首先，进入 GCMC 文件夹会有一个以模拟气体种类命名的一个文件夹，然后是晶体名的文件夹，接下来在是气体的分压，最后是总的温度和压力，结构如图 9 所示：

```
[zxm@gibbs 100.0kPa]$ pwd
/home/zxm/PyMSATm/example/jobs/GCMC/H2S_N2/C0F108/0.1_0.9/200.0K/100.0kPa
```

图 9 自动生成文件路径范例

对于任务的提交，可以写一个简短的 shell 脚本来完成，如图 10。[这里我们强烈建议在提交大量任务前先随机选择几个任务尝试运行一下，即使在在批量提交时，也要做好任务的记录工作，这样方便出现错误后进行回溯调查错误来源。](#)

```
#!/bin/bash
j=/home/zxm/PyMSATm/example/recording/example_qsubgcmc.txt
touch $j
k=0
echo -e "Time: \c" >> $j
echo `date` >> $j
for i in /home/zxm/PyMSATm/example/jobs/GCMC/H2S_N2/*/*/*
do
let "k=$k+1"
if [ -d $i ]; then
cd $i
echo "$k.-----" >> $j
echo -e "Setting: \c" >> $j
pwd >> $j
pwd
echo -e "Job ID: \c" >> $j
qsub run_gcmc.pbs >> $j
fi
done
```

图 10 提交任务的脚本范例

为了方便读者理解，这里给出 TranslateFormat.py 的运行流程图来加强使用者对程序的理解，如图 4 所示。

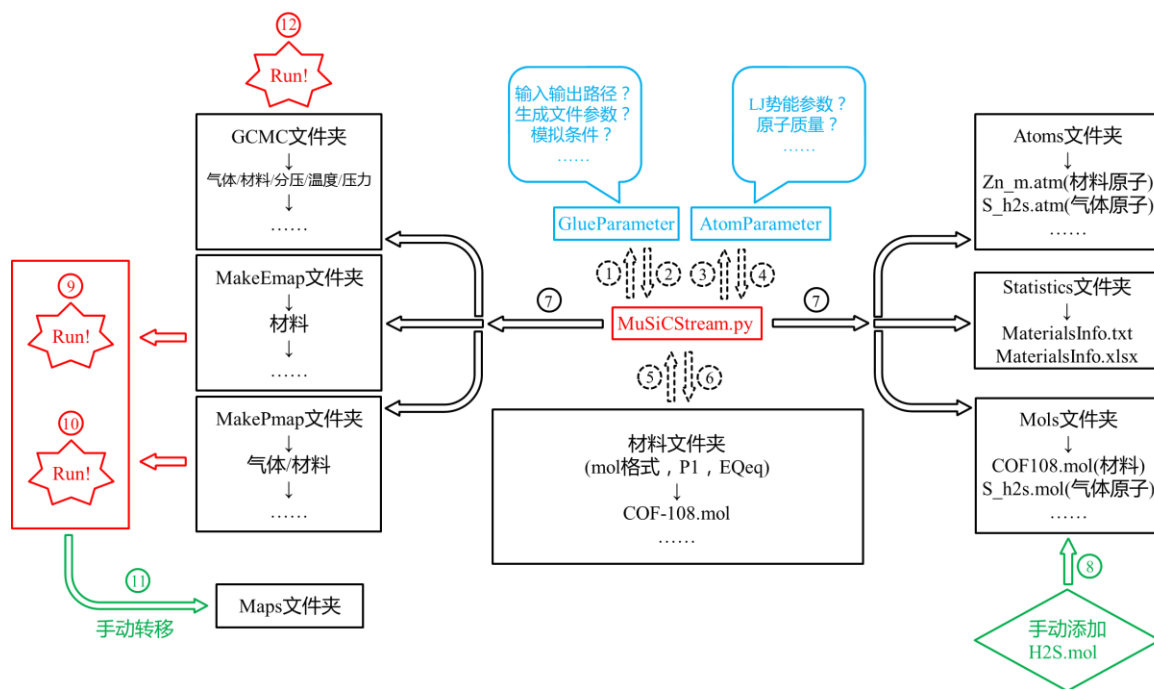


图 11 MuSiCStream.py 运行流程图

3.3 ExtractData.py 介绍

ExtractData.py 程序的功能主要是完成模拟结果的分析和提取。当所有的模拟计算都完成后，ExtractData.py 程序可以提取出的有用的模拟信息。这个程序会根据 GlueParameters 文件中的参数自动搜索并统计完整配置条件下的吸附量。如果使用了 MuSiC 中的 post 功能，该程序还会统计气体在材料中的作用势能，而气体在材料中的作用势能可用于计算气体吸附热。所有的统计结果都会输出到输出文件路径下的 Results 文件夹中。如果 Python 中装有 openpyxl 库，程序则可以输出文本格式和 Excel 格式两种统计文件，否则只会生成文本格式的统计文件。吸附量统计和势能统计文件如下（只展示 Excel 格式文件）：

	A	B	C	D	E
1		100	压力: kPa	200	
2		H2S	N2	H2S	N2
3	COF108	0.02131	1.14406	0.04523	2.17629
4	MgMOF74	15.4954819	3.54042375	16.0335225	4.6434625
5	NU2Pyr_4TT_Por_4Pftw	19.0031913	1.03143625	20.6266025	2.54068375
6	ZIHTEP_clean	0.48577958	0.10197667	0.53921292	0.18303292
7	hypotheticalMOF_7003_ID_1	15.115575	0.8079875	17.3091	1.6506325
8		单位: mol/uc			
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					
22					
23					
24					
25					
26					

图 12 统计吸附量 Excel 输出格式

	A	B	C	D	E
1		100		200	
2		H2S	N2	H2S	N2
3	COF108	0	-0.97233	0	-2.48449
4	MgMOF74	-178.30967	-63.13443	-167.51997	-53.06382
5	NU2Pyr_4TT_Por_4Pftw	-64.78639	-21.07959	-61.67298	-22.0387
6	ZIHTEP_clean	-36.11135	-7.47513	-34.79127	-9.43642
7	hypotheticalMOF_7003_ID_1	-50.42089	-6.73185	-49.58369	-7.17028
8		单位: kJ/mol			
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					
22					
23					
24					
25					
26					
27					
28					
29					

图 13 统计相互作用能 Excel 输出格式

注意：这里的势能是指气体与所有原子（材料+自身+其它气体）之间作用能的加和，如果有其它势能计算需求，请联系我们。硫化氢的势能计算如下图所示：


```

## All the averages below are based on cum avg
## Unstored/not used nrgs might appear as zero
Sorbs:
      Nrg-Total      mol/uc      Nrgs (KJ/mol)
H2S--H2S      Coulombic :      11156.60389      143.56125      19.42830
H2S--H2S      NonCoulom :      365026.17038      143.56125      635.66278
H2S--N2       Coulombic :       86.31156      151.60875       0.14233
H2S--N2       NonCoulom :      7454.24990      151.60875      12.29102
H2S--O2       Coulombic :       0.82768      143.68125       0.00144
H2S--O2       NonCoulom :       23.36876      143.68125       0.04066
H2S--6        Coulombic :     -6556410.75622      143.81125     -11397.59712
H2S--6        NonCoulom :      2215402.29294      143.81125      3851.23259
N2--N2        Coulombic :       1.38113       8.04750       0.04291
N2--N2        NonCoulom :       15.87983       8.04750       0.49332
N2--O2        Coulombic :       0.02485       8.16750       0.00076
N2--O2        NonCoulom :      -0.35907       8.16750      -0.01099
N2--6         Coulombic :     -27140.83643       8.29750     -817.74138
N2--6         NonCoulom :      6247.99312       8.29750      188.24927
O2--O2        Coulombic :       0.00000       0.12000       0.00000
O2--O2        NonCoulom :      -0.00010       0.12000      -0.00020
O2--6         Coulombic :      -73.89198       0.37000     -49.92702
O2--6         NonCoulom :       15.60102       0.37000      10.54123
6--6          Coulombic :       0.00000       0.25000       0.00000
6--6          NonCoulom :       0.00000       0.25000       0.00000

```

图 14 GCMC 模拟相互作用能输出算法

为了方便读者理解，这里给出 ExtractData.py 的运行流程图来加强使用者对程序的理解，流程如图 15 所示。

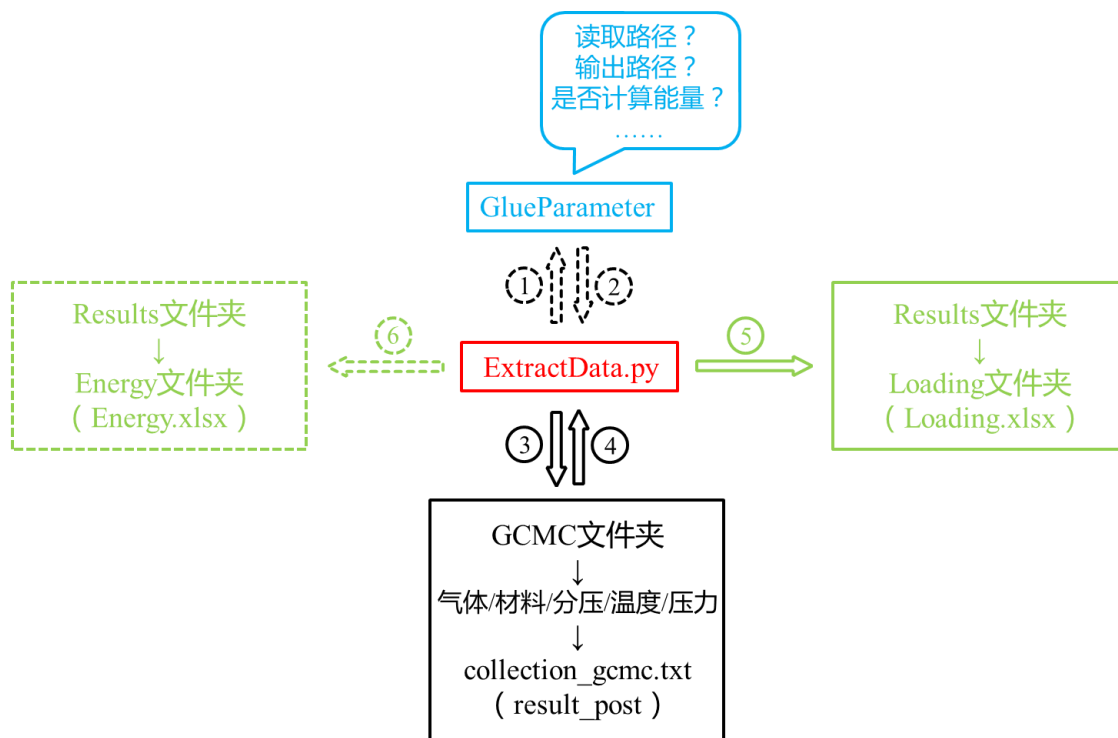


图 15 ExtractData.py 运行流程图

4. 参数设定

在所有输入文件中，被（“#”+空格）标记的行为注释行，（关键词+“:”+空格）后面的内容为输入参数。所有路径参数都最好使用绝对路径。

4.1 TranslateInfo 格式

下图为 TranslateInfo 文件的模板：

```
#####
# PyMSATm -- Python Multipurpose Simulation analysis Tools Package for MuSiC #
# Simulation Software #
# Copyright (C) 2017 Xumiao Zhou, Li Yang #
# #
# 1026715200@qq.com #
# liyang@wit.edu.cn #
# #
# Contact Address : #
# Li Yang #
# School of Chemical Engineering and Pharmacy #
# Wuhan Institute of Technology #
# No.206, Guanggu 1st road #
# Donghu New & High Technology Development Zone #
# Wuhan, Hubei Province, P.R. China #
# Postcode: 430205 #
#####

# This is a information file for TranslateFormat.py
# Please do not modify or delete comments and key words
# Calculate the charge carefully because the pybel calculation results are -
# - inconsistent with the original code calculations

ReadFilePath: D:\Jobs\Code\ZxmTed\ExtraCodes\cif
OutputPath: D:\Jobs\Code\ZxmTed\ExtraCodes\cif_charge
CalculateCharge: yes (yes/no)
# It will be closed, when export format setted as pdb
ImportFormat: cif (cif&pdb)
ExportFormat: mol (cif&pdb&mol)
# The exported files of cif are set to P1
ElementSuffix: _m (Can only input up to two characters due to demand of pdb format,e.g. '_*')
# This function is very useful for distinguishing the same elements in materials and gases
```

图 16 TranslateInfo 文件格式示意图

ReadFilePath（绝对路径，字符串）

用于读取晶体信息文件。参数必须是一个文件夹路径。程序会自动搜索该文件夹下所有指定格式的晶体信息。

OutputPath（绝对路径，字符串）

用于输出晶体信息文件。参数必须是一个文件夹路径。转换好格式的晶体信息文件会全部输

出到指定的文件夹。

CalculateCharge (yes/no)

是否需要给晶体计算电荷。目前只能使用 EQeq 算法来分配电荷。由于 pdb 格式中没有足够的位置书写电荷，因此该功能针对 pdb 格式为默认关闭状态。

ImportFormat (cif/pdb)

读取晶体信息的文件格式。

ExportFormat (cif/pdb/mol)

输出晶体信息的文件格式。

ElementSuffix (字符串)

给每个晶体原子名称后面加上一个后缀。给每个晶体原子名称后面加上一个后缀是十分必要的，可以用于区分同一类元素或原子。例如，气体中的氧和晶体中的氧不能使用同一名称，沸石中的氧（可以使用“_s”）和 MOFs 中的氧（“_m”）不能使用同一名称。由于 PDB 格式需求此后缀的字符长度最好不要超过两个。

4.2 GlueParameters 格式

下图为 GlueParameters 文件的模板（下图存在部分内容的省略）：

```

.....

# This is a information file
# Please do not modify or delete comments and key words
# The symbol of "#" means following contents will ignored
# All parameters cannot be null
# This code will use atomic tags in the material file instead of elements -
# - to distinguish the same atoms species in the different chemistry environment

# ----- Controlled part ----- #
MakeGCMC: open (open/off)
UsePmap: yes (yes/no)
UseEmap: yes (yes/no)
UsePost: yes (yes/no)
# Post file uses to calculate the gas average energy in the ensemble
# such as,in the ensemble of H2S/N2/IRMOF-1
# the H2S energy equal H2S--N2 plus H2S--H2S plus H2S--IRMOF-1
# more details please refer the result_post file
# which was generated after finished calculation

MakePmap: open (open/off)

MakeEmap: open (open/off)

.....

MuSiCSetting: ...
# the following sentences are used to set the MuSiC environment:
# do not miss "..." symbol after "MuSiCSetting:"
# -----

export MUSiC=/program/music/4.0
export PATH=${MUSiC}/bin:${PATH}

# -----

END # the ending symbol of Info file

```

图 16 GlueParameters 文件格式示意图

MakeGCMC（open/off）

是否生成用于 GCMC 模拟的文件。

- UsePmap（yes/no）

是否在 GCMC 模拟中调用 pmap。

- UseEmap（yes/no）

是否在 GCMC 模拟中调用 emap。

- UsePost（yes/no）

是否在 GCMC 模拟中调用 post 功能。这个功能决定是否计算气体在材料中的势能。

MakePmap (open/off)

是否生成用于制作 pmap 的文件。

MakeEmap (open/off)

是否生成用于制作 emap 的文件。

MakeTorque (open/off)

是否给每个工作文件夹生成 Torque 文件。这个文件用于服务器上任务的递交与管理，具体的文件可以参考下图（下图存在部分内容省略，完整内容请参看测试）：

```
#!/bin/bash
#PBS -l nodes=1:ppn=1
#PBS -N MuSiC_gcmc.ZXM
#PBS -o music_gcmc_jobs.out
#PBS -j oe

#
# The number of processors you desire is indicated by replacing
# <nproc> above.
#

#
# GROMACS path and arguments to mdrun :
#
cd $PBS_O_WORKDIR

.....

echo "+++++++ Run MuSiC ++++++"
music_gcmc balance_gcmc.ctr > balance_gcmc.txt
echo `date`
music_gcmc collection_gcmc.ctr > collection_gcmc.txt
echo `date`
music_post post.ctr > post.txt
echo `date`
echo "+++++++ Finish MuSiC ++++++"

cd $PBS_O_WORKDIR
cp -rf ${TEMP_DIR}/* .
rm -rf ${TEMP_DIR}

echo "All files were copied back!"
echo "The work direcotry: " $PBS_O_WORKDIR
echo `date`
echo "=====Finished Job ====="
```

图 17 调度任务文件格式示意图

不同服务器递交与管理任务的方式不同，或者服务器文件夹构架不一致也可能导致不可知的错误，请根据所使用的服务器设定来设定相关功能。关于此部分的详情请参考官方使用手册：

<http://www.adaptivecomputing.com/support/documentation-index/torque-resource-manager-documentation/>。

UseChargesFromPDBFile (yes/no)

是否使用文件中的原子电荷。默认为关闭，此功能是为了避免错误的使用 pdb 文件中的原子电荷，目前该功能还在开发中。其它类型文件中的电荷默认为读取。

ExtractEnergyData (yes/no)

是否提取并统计气体在材料中的势能。此功能服务于 ExtractData.py 程序。

InputPath (绝对路径, 字符串)

晶体信息文件的路径。晶体信息文件的格式必须为后缀为 mol 的 mol 格式，最好使用 TranslateFormat.py 程序生成的文件。

OutputPath (绝对路径, 字符串)

程序输出文件的存放路径。

AtomParameterPath (绝对路径, 字符串)

AtomParameter 文件的存放路径。

GasType (列表, 字符串)

以空格为分隔，列出吸附气体的名称。

GasAtomTypeNum (列表, 字符串)

以空格为分隔, 按照 GasType 列出气体的顺序依次列出每种气体中所包含的原子种类数。注意: 分子中的伪原子也算一种原子种类, 伪原子的命名我们强行要求为 M_加上识别符, 如氮气中的伪原子为 M_n2。

GasAtomType (列表, 字符串)

以空格为分隔, 按照 GasType 列出气体的顺序依次列出每种气体中所包含的原子种类。

Multiple (整数)

这个数值决定最后进行模拟时的晶胞大小。因为晶胞在轴方向的长度必须大于两倍的截断半径, 所以这个值最小为 2。

CutOff (浮点数)

计算静电作用力和范德华作用力时的最远截断半径。

GasPartialPressure (列表, 字符串)

以下划线为分隔, 按照 GasType 列出气体的顺序依次列出每种气体的分压比。支持同时列出多种分压类型, 不同分压类型以空格分开。

TemperatureList (列表, 浮点数)

以空格为分隔, 列出需要模拟的温度。

PressureList (列表, 浮点数)

以空格为分隔, 列出需要模拟的压力。

BalanceStep (整数)

平衡系统的步长数。

CollectionStep (整数)

采样的步长数。

GridSpacingP (浮点数)

制作 pmap 时采样坐标间隔长度, 单位为 Å。间隔长度越小采样结果越精确, 但生成的文件越大。

HighEndPotentialCutoffP (整数)

pamp 中最高作用能的截断值。当计算出的作用能大于该值时则舍弃该值。

GridSpacingE (浮点数)

制作 emap 时采样坐标间隔长度, 单位为 Å。间隔长度越小采样结果越精确, 但生成的文件越大。

HighEndPotentialCutoffE (整数)

emap 中最高作用能的截断值。当计算出的作用能大于该值时则舍弃该值。经研究, 这个值会

强烈影响离子型沸石中 `emap` 制作的准确性。

Nodes（字符串）*

调用服务器计算节点的方式。

TaskSuffix（字符串）*

给递交的任务名加上后缀。这个功能在区分不同人提交的任务时很有用。

TorqueSetting（字符串）*

设置任务管理器的运行环境。

MuSiCSetting（字符串）*

设置 MuSiC 软件的运行环境。

4.3 AtomParameter 格式

设置原子与原子之间的范德华作用力参数。AtomParameter 文件主要分成三个部分，具体的文件格式如下（下图存在部分内容省略，完整内容请参看测试）：

```

.....

# This file records all atom parameters
# Please do not modify or delete comments and key words
# The Pseudo atom uses name by M_, e.g. M_n2
# Lorentz Berthelot mixing rules are used to calculate the cross site-site
# UFF (In silico screening of 4764 computation-ready,
# experimental metal-organic frameworks for CO2 separation)
# The atom mass from Israel Science and Technology Homepage

MaterialAtom:
Number  AtomName  Sigma  Epsilon  FF  Mass
1      Ac_m      3.1    16.6     UFF  227.0
2      Ag_m      2.8    18.11    UFF  107.8682
3      Al_m      4.01   254.09   UFF  26.9815
4      Am_m      3.01   7.04     UFF  243.0
.....
99     Y_m       2.98   36.23    UFF  88.9059
100    Yb_m      114.72  2.99     UFF  173.04
101    Zn_m      2.46   62.39    UFF  65.39
102    Zr_m      2.78   34.72    UFF  91.224

GasAtom:
Number  AtomName  Sigma  Epsilon  Mass
1      H_h2s  0      0         1.0079
2      S_h2s  3.7    275       32.065
3      N_n2   3.31   36        14.0067
4      M_n2   0      0         0
5      O_h2o_4 3.15   78        15.9994
6      H_h2o_4 0      0         1.0079
7      M_h2o_4 0      0         0
8      O_o2   3.05   54.4     15.9994
9      M_o2   0      0         0

SpecialPair:
Number  AtomNameOne  AtomeNameTwo  Sigma  Epsilon
1      S_h2s   C_m          3.5    73

```

图 18 AtomParameter 文件格式示意图

MaterialAtom 和 GasAtom 分别用于描述晶体和气体中原子的参数，不同原子之间的参数计算采用 LB 混合规则：

$$\sigma_{ij} = \frac{\sigma_i + \sigma_j}{2} \quad \epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}$$

SpecialPair 用于描述特定原子对之间的特殊参数。所有的参数排列和内容书写方式必须严格按照给出的参考样板。

5. 扩展程序

5.1 dmap2plt.py

MuSiC 生成的吸附密度文件的格式为 dmap，该程序可以把 dmap 格式转化为 plt 格式。plt 格式可以使用 VMD 软件打开并进行相关的可视化操作，更多关于 VMD 软件的更多详情请参考其官网：<http://www.ks.uiuc.edu/Research/vmd/>。

该程序需要手动修改转化文件的路径与输出路径，如图 19。

```
import struct

with open(r'D:\Jobs\zxm\report\snurr_mofs_separation\New\AdsorptionMap\H2S\UIO-66\H2S_uio_66.poremap.dmap', 'r') as file:
    file_list = file.readlines()
    rank = int(file_list[0].split()[0])
    type = int(file_list[0].split()[1])
    grid_z = int(file_list[1].split()[0])
    grid_y = int(file_list[1].split()[1])
    grid_x = int(file_list[1].split()[2])
    min_z = float(file_list[2].split()[0])
    max_z = float(file_list[2].split()[1])
    min_y = float(file_list[2].split()[2])
    max_y = float(file_list[2].split()[3])
    min_x = float(file_list[2].split()[4])
    max_x = float(file_list[2].split()[5])
    bytes_two = struct.pack('f', float(file_list[3]))
    bytes_one = struct.pack('iiiiifffff', rank, int(3), grid_z, grid_x, grid_y, min_z, max_z, min_y, max_y, min_x, max_x)
    for site in file_list[4:]:
        bytes = struct.pack('f', float(site))
        bytes_two += bytes

with open(r'D:\Jobs\zxm\report\snurr_mofs_separation\New\AdsorptionMap\H2S\UIO-66\H2S_uio_66.plt1', 'wb') as file_out:
    file_out.write(bytes_one)
    file_out.write(bytes_two)
```

输入路径

输出路径

图 19 dmap2plt.py 使用方法

5.2 ZeoReplaceSi2Al.py

该程序主要用于全硅沸石中铝原子的替代并给每种原子分配指定的电荷。沸石在实际生产和应用中往往不是以纯硅的形式存在，一般都含有一定量的铝原子掺杂在沸石的拓扑结构中并通过一些阳离子（钠、钾、钙等）来使材料整体承电中性。该程序以全硅沸石为模板，在符合 Lowenstein 规则的前提下随机换指定数量的硅为铝并为每种原子赋予制定电荷。

该程序需要手动设置参数（246~252 行），如图 20。其中 material_path 为全硅沸石的路径，output_path 为输出文件夹路径，Si_charge、Al_charge、Si_O_Al_charge 和 Si_O_Si_charge 分别为硅、铝、氧（硅铝之间）和氧（硅硅之间）的电荷，output_num 为需要生成的材料个数（由于程序为随机替换，因此可能会出现一样替换位置的材料）。

需要替换的铝原子个数请根据程序提示输入！

```
def main():  
    material_path = 'D:/Jobs/Database/zeolite_95_cif/FAU.cif'  
    output_path = 'D:/Jobs/Database/test'  
    Si_charge = 2.4  
    Al_charge = 1.7  
    Si_O_Al_charge = -1.2  
    Si_O_Si_charge = -1.2  
    output_num=1000
```

图 20 ZeoReplaceSi2Al.py 的参数设置

6. 更新与修订

PyMSATm 2.0 版本（2019 年 3 月）

1. 修改错误。MusicStream 程序生成的材料 mol 文件中的“Rigid”关键字替换为“None”。这个错误会导致模拟过程中晶体原子的整体漂移。
2. 修改关键字。MuSiCStream 程序生成的 GCMC 文件更名为 Equilibrium 和 Production。
3. 添加新功能。MuSiCStream 程序可以在相同气体的相同分压下批量提交任务。
4. 修改读取和输出文件的格式类型。TranslateFormat 程序读取格式为 cif 和 pdb，输出格式为 cif、pdb 和 mol。MuSiCStream 程序的晶体读取格式改为 mol。
5. 新增两个扩展程序。

参考文献

- [1] N. M. O'boyle, C. Morley, and G. R. Hutchison. Pybel: a Python wrapper for the OpenBabel cheminformatics toolkit[J]. Chem. Cent. J, 2008, 2: 5.
- [2] C. E. Wilmer, K. C. Kim, and R. Q. Snurr. An Extended Charge Equilibration Method[J]. J Phys. Chem. Lett., 2012, 3(17): 2506.

软件运行总流程图

