

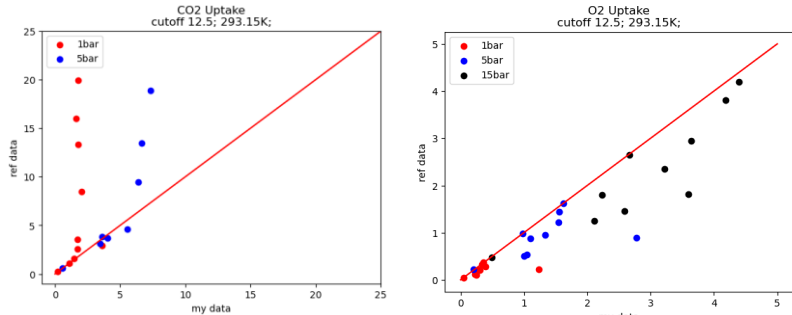
GCMC 시뮬레이션 검증 과정 및 발견된 이슈에 대한 피드백 요청

작성일자 : 2025 02 04

작성자 : 안용상

나현누나 GCMC와 일치하지 않는 이유 찾기

전혀 맞지 않는 결과를 내는
이전에 했던 흡착 시뮬레이션



여러 실험들을 통해 그 문제점을 파악할 수 있었습니다.
결론적으로 문제의 원인은 cif파일 구성에 있었습니다.

먼저 기존에 사용하던 cif파일 구성을 살펴보면 우측과 같습니다.
여러 cif들중 한 개를 가져온 것입니다.

```

1 data_image0
2 _cell_length_a 38.9881
3 _cell_length_b 38.9881
4 _cell_length_c 22.2962
5 _cell_angle_alpha 90
6 _cell_angle_beta 90
7 _cell_angle_gamma 120
8
9 _symmetry_space_group_name_H-M "P 1"
10 _symmetry_int_tables_number 1
11
12 loop_
13 _symmetry_equiv_pos_as_xyz
14 'x, y, z'
15
16 loop
17   _atom_site_label
18   _atom_site_occupancy
19   _atom_site_fract_x
20   _atom_site_fract_y
21   _atom_site_fract_z
22   _atom_site_thermal_displace_type
23   _atom_site_B_iso_or_equiv
24   _atom_site_type_symbol
25 Mn1 1.0000 0.66392 0.73824 0.09236 Biso 1.000 Mn
26 Mn2 1.0000 0.58001 0.65563 0.03378 Biso 1.000 Mn
27 O1 1.0000 0.39461 0.91990 0.99762 Biso 1.000 O
28 O2 1.0000 0.37348 0.89757 0.08970 Biso 1.000 O
29 O3 1.0000 0.68342 0.94699 0.91477 Biso 1.000 O
30 O4 1.0000 0.71682 0.00845 0.94560 Biso 1.000 O
31 O5 1.0000 0.66782 0.04718 0.12943 Biso 1.000 O
32 O6 1.0000 0.61038 0.00212 0.16789 Biso 1.000 O
33 O7 1.0000 0.57184 0.70428 0.01038 Biso 1.000 O
34 O8 1.0000 0.61268 0.74257 0.08365 Biso 1.000 O
35 C1 1.0000 0.58846 0.73516 0.04267 Biso 1.000 C
36 C2 1.0000 0.57754 0.76614 0.02633 Biso 1.000 C
37 C3 1.0000 0.55582 0.76316 0.97364 Biso 1.000 C
38 H1 1.0000 0.54780 0.74190 0.94750 Biso 1.000 H
39 C4 1.0000 0.54700 0.79271 0.96234 Biso 1.000 C
40 H2 1.0000 0.53290 0.79110 0.92780 Biso 1.000 H
41 C5 1.0000 0.55818 0.82428 0.99985 Biso 1.000 C

```

이 cif파일이 담고 있는 구조정보 컬럼입니다.
여기서 **_atom_site_label**이라는 것이 있습니다.
원소 라벨 정보를 나타내는 것으로 보이는데,
제가 다룬바는 cif파일들은 이 **_atom_site_label**이라고 명명
된 컬럼의 데이터들이 원소+번호(몇번째 나오는지)의 규칙으
로 작성되어 있었습니다.

원소+번호(몇번째 나오는지)

Movies	2025-01-23 오전 10:31	파일 폴더
Output	2025-01-23 오전 10:31	파일 폴더
Restart	2025-01-23 오전 10:31	파일 폴더
VTK	2025-01-23 오전 10:31	파일 폴더
System_0	2025-02-04 오후 4:18	파일 폴더
Framework_0_initial_1_1_2.cell	2025-01-23 오전 10:31	한컴오피스 한셀 ... 80KB
Framework_0_initial_1_1_2.txt	2025-01-23 오전 10:31	TXT 파일 70KB
Framework_0_initial_1_1_2.P1.cif	2025-01-23 오전 10:31	CIF 파일 107KB
Framework_0_initial_1_1_2.VASP.cif	2025-01-23 오전 10:31	CIF 파일 107KB

```

1 data_[CoreMOF]ZOFZUR_clean
2
3 _audit_creation_method RASPA-1.0
4 _audit_creation_date 2025-1-23
5 _audit_author_name ''
6
7 _cell_length_a 38.9881
8 _cell_length_b 38.9881
9 _cell_length_c 44.5924
10 _cell_angle_alpha 90
11 _cell_angle_beta 90
12 _cell_angle_gamma 120
13 _cell_volume 58702.4
14
15 _symmetry_cell_setting triclinic
16 _symmetry_space_group_name_Hall 'P 1'
17 _symmetry_space_group_name_H-M 'P 1'
18 _symmetry_int_tables_number 1
19
20 _symmetry_equiv_pos_as_xyz 'x,y,z'
21
22 loop_
23   _atom_site_label
24   _atom_site_type_symbol
25   _atom_site_fract_x
26   _atom_site_fract_y
27   _atom_site_fract_z
28   _atom_site_charge
29 Mn1 Mn 0.663920000000 0.738240000000 0.092360000000 0
30 Mn2 Mn 0.580010000000 0.655630000000 0.033780000000 0
31 O1 O 0.394610000000 0.919900000000 0.997620000000 0
32 O2 O 0.373480000000 0.897570000000 0.089700000000 0
33 O3 O 0.683420000000 0.946990000000 0.914770000000 0
34 O4 O 0.716820000000 0.008450000000 0.945600000000 0
35 O5 O 0.667820000000 0.047180000000 0.129430000000 0
36 O6 O 0.610380000000 0.002120000000 0.167890000000 0
37 O7 O 0.571840000000 0.704280000000 0.010380000000 0
38 O8 O 0.612680000000 0.742570000000 0.083650000000 0
39 C1 C 0.588460000000 0.735160000000 0.042670000000 0
40 C2 C 0.577540000000 0.766140000000 0.026330000000 0
41 C3 C 0.555820000000 0.763160000000 0.973640000000 0
42 H1 H 0.547800000000 0.741900000000 0.947500000000 0
43 C4 C 0.547000000000 0.792710000000 0.962340000000 0
44 H2 H 0.532900000000 0.791100000000 0.927800000000 0
45 C5 C 0.558180000000 0.824280000000 0.999850000000 0

```

_atom_site_label은 RASPA 프로그램이 시뮬레이션을 위해 pseudo atoms 파일의 charge 정보를 참조하여 흡착제 CIF 파일의 원소에 charge를 부여할 때, pseudo atoms 파일의 원소와 짝을 맞추는 기준이 되어주는 정보이다.

즉 **_atom_site_label**의 문자열과 정확히 일치하는 pseudo atoms의 원소를 찾아 charge 정보를 부여해준다는 것인데, 나의 cif의 경우에는 문자열에 원소뿐만 아니라 숫자도 포함되어, pseudo atoms파일에서 적절하게 동일한 원소를 찾아 charge 정보를 추출해낼 수 없었다.

따라서 pseudo_atoms파일에서 누락된 원소라고 RASPA가 판단하고 실제 시뮬레이션 과정에서 charge로 0을 부여하게 되었다.

RASPA 소프트웨어는 사용자가 직접 전달한 cif파일을 받아 RASPA가 읽을 수 있는 cif형태로 한번 변환하는 과정을 거친다.
변환 과정에 pseudo atoms 등의 파일에서 charge와 같은 정보를 가져와 기입하는 과정이 포함된다.

이렇게 변환을 거친 cif 실제로 RASPA가 이용하는 cif로써 쓰이게 되는데 그 정보는 /Movies/System_0/Framework_0_initial_1_1_2.P1.cif 파일에 존재한다.

따라서 이 경우 실제 시뮬레이션에 쓰인 변환된 cif파일을 확인해보니 우측 사진처럼 실제 부여된 **_atom_site_charge**가 모두 0으로 나와있었다.
이런 이유로 나현누나가 계산한 흡착량보다 흡착량이 현저히 적게 나왔었던 것으로 추측된다.

나현누나 GCMC와 일치하지 않는 이유 찾기

따라서 pseudo atoms파일에서 잘 charge 정보를 가져올 수 있도록 내 cif파일의 _atom_site_label 원소 + 번호 문자열의 번호부분을 모두 지워주고 다시 시뮬레이션을 진행하는 실험을 해봤다.

```

1 data_image0
2 _cell_length_a      38.9881
3 _cell_length_b      38.9881
4 _cell_length_c      22.2962
5 _cell_angle_alpha   90
6 _cell_angle_beta    90
7 _cell_angle_gamma   120
8
9 _symmetry_space_group_name_H-M  "P 1"
10 _symmetry_int_tables_number     1
11
12 loop_
13   _symmetry_equiv_pos_as_xyz
14   'x, y, z'
15
16 loop_
17   _atom_site_label
18   _atom_site_occupancy
19   _atom_site_fract_x
20   _atom_site_fract_y
21   _atom_site_fract_z
22   _atom_site_thermal_displace_type
23   _atom_site_B_iso_or_equiv
24   _atom_site_type_symbol
25 Mn      1.0000 0.66392 0.73824 0.09236 Biso 1.000 Mn
26 Mn      1.0000 0.58001 0.65563 0.03378 Biso 1.000 Mn
27 O       1.0000 0.39461 0.91990 0.99762 Biso 1.000 O
28 O       1.0000 0.37348 0.89757 0.08970 Biso 1.000 O
29 O       1.0000 0.68342 0.94699 0.91477 Biso 1.000 O
30 O       1.0000 0.71682 0.00845 0.94560 Biso 1.000 O
31 O       1.0000 0.66782 0.04718 0.12943 Biso 1.000 O
32 O       1.0000 0.61038 0.00212 0.16789 Biso 1.000 O
33 O       1.0000 0.57184 0.70428 0.01038 Biso 1.000 O
34 O       1.0000 0.61268 0.74257 0.08365 Biso 1.000 O
35 C       1.0000 0.58846 0.73516 0.04267 Biso 1.000 C
36 C       1.0000 0.57754 0.76614 0.02633 Biso 1.000 C
37 C       1.0000 0.55582 0.76316 0.97364 Biso 1.000 C
38 H       1.0000 0.54780 0.74190 0.94750 Biso 1.000 H
39 C       1.0000 0.54700 0.79271 0.96234 Biso 1.000 C
40 H       1.0000 0.53290 0.79110 0.92780 Biso 1.000 H
41 C       1.0000 0.55818 0.82428 0.99985 Biso 1.000 C
42 C       1.0000 0.57927 0.82658 0.04937 Biso 1.000 C
43 H       1.0000 0.58700 0.84790 0.07520 Biso 1.000 H
44 C       1.0000 0.58984 0.79837 0.06319 Biso 1.000 C
45 H       1.0000 0.60510 0.80130 0.09700 Biso 1.000 H
46 C       1.0000 0.54660 0.85563 0.98649 Biso 1.000 C
47 H       1.0000 0.56050 0.86980 0.95060 Biso 1.000 H
48 H       1.0000 0.51850 0.84220 0.97770 Biso 1.000 H
49 C       1.0000 0.52654 0.87268 0.07963 Biso 1.000 C
50 H       1.0000 0.54060 0.88550 0.11610 Biso 1.000 H

```

번호 삭제

세 경우를 비교하면 다음과 같았다.

ZOFZUR_clean.cif			
	나현누나 결과	내결과 (atom_site_label이 원소+숫자)	내결과 (atom_site_label에서 숫자 제거)
mol/kg	19.8	1.779	19.8

일치하는 결과를 얻었다

결론적으로 _atom_site_label이 pseudo_atoms의 정보와 일치하지 않아서 charge정보를 끌고오지 못해 제대로된 시뮬레이션이 되지않았던 것으로 볼 수있다.

NET CHARGE 문제

그러나 일치하는 데이터가 나오는 경우에도 한가지 이슈가 있었다.
간혹 몇몇 MOF들에 대한 out파일에 다음과 같은 Warning이 나왔다.

```
2118
2119 WARNING: THE SYSTEM HAS A NET CHARGE
2120
```

NET CHARGE가 있다는 경고였다.

그래서 원인을 파악하고자 CIF파일을 확인해봤다.

RASPA가 charge 정보를 입력하고 변환한 cif를 Movies/System_0/Framework_0_initial_1_1_2_P1.cif 을 열어 확인해보았더니 다음과 같았다.

```
1 data_ZOZUR_clean
2
3 _audit_creation_method RASPA-1.0
4 _audit_creation_date 2025-1-24
5 _audit_author_name ''
6
7 _cell_length_a 38.9881
8 _cell_length_b 38.9881
9 _cell_length_c 44.5924
10 _cell_angle_alpha 90
11 _cell_angle_beta 90
12 _cell_angle_gamma 120
13 _cell_volume 58702.4
14
15 _symmetry_cell_setting triclinic
16 _symmetry_space_group_name_Hall 'P 1'
17 _symmetry_space_group_name_H-M 'P 1'
18 _symmetry_Int_Tables_number 1
19
20 _symmetry_equiv_pos_as_xyz 'x,y,z'
21
22 loop_
23 _atom_site_label
24 _atom_site_type_symbol
25 _atom_site_fract_x
26 _atom_site_fract_y
27 _atom_site_fract_z
28 _atom_site_charge
29 Mn Mn 0.663920000000 0.738240000000 0.046180000000 0
30 Mn Mn 0.580010000000 0.655630000000 0.016890000000 0
31 O O 0.394610000000 0.919900000000 0.498810000000 -1.025
32 O O 0.373480000000 0.897570000000 0.044850000000 -1.025
33 O O 0.683420000000 0.946990000000 0.457385000000 -1.025
34 O O 0.716820000000 0.008450000000 0.472800000000 -1.025
35 O O 0.667820000000 0.047180000000 0.064715000000 -1.025
36 O O 0.610380000000 0.002120000000 0.083945000000 -1.025
37 O O 0.571840000000 0.704280000000 0.005190000000 -1.025
38 O O 0.612680000000 0.742570000000 0.041825000000 -1.025
39 C C 0.588460000000 0.735160000000 0.021335000000 0
40 C C 0.577540000000 0.766140000000 0.013165000000 0
41 C C 0.555820000000 0.763160000000 0.486820000000 0
42 H H 0.547800000000 0.741900000000 0.473750000000 0
43 C C 0.547000000000 0.792710000000 0.481170000000 0
44 H H 0.532900000000 0.791180000000 0.463900000000 0
45 C C 0.558180000000 0.824280000000 0.499925000000 0
46 C C 0.579270000000 0.826580000000 0.024685000000 0
47 H H 0.587000000000 0.847900000000 0.037600000000 0
```

흡착제 내의 모든 원소에 charge가 할당되지 않았고
일부 원소만 charge정보가 있었다.

원인을 파악해보니 pseudo_atoms를 참조하지만 pseudo_atoms파일에 기입되지 않은 원소는 정보를 긁어오지 못해 0으로 charge가 할당되는 것으로 파악되었다.

이렇게 NET Charge가 생기면 전하 균형 문제를 야기하고 전기적 상호작용 왜곡이 생겨 시뮬레이션 신뢰도를 저하시킬 가능성이 있다고 생각했다.

조사해봤을때, 이때 여러가지 방법으로 전하 정보를 부여할 수 있다.

1. DFT 시뮬레이션을 진행해 모든 원자에 직접 정확한 atom_site_charge를 부여한다
2. RASPA에 charge_equilibrium 기능을 이용해 러프하게 계산해 이용한다.

- 1번 방법은 추가 DFT 시뮬레이션 시간이 많이 들 수 있다.
- 2번 방법은 직접 해보니 자꾸 RASPA 소프트웨어에서 메모리 덤프 및 SIGSEGV os segmentation fault 에러가 났고, 라스파 커뮤니티를 찾아보니 다른 사용자에게도 발생하는 소프트웨어 자체 문제인것 같았다.

<https://forums.iraspa.org/index.php?topic=1018.0>










charge를 직접 구하는 것보다 이미 _atom_site_charg가 구해져있는 DB가 있는지 찾아봤다.

그 결과 한 데이터베이스를 찾을 수 있었다.

MOFSimplify 프로젝트의 github를 보니 CoRE2019라는 폴더에서 charg정보를 담은 cif를 제공하는 것을 확인할 수 있었다.

NET CHARGE 문제를 해결할 수 있는 방법

<https://github.com/hjkgrp/MOFSimplify>

 gianmarco-terrones updated citation information	9868303 · 3 months ago	🕒 426 Commits
 CoRE2019	options for user to download information on latent space ne...	4 years ago
 TGA	made all readme files full caps in their name	4 years ago
 VM_info	removed all __pycache__ folders and adjusted README.rtf fil...	4 years ago
 environments	water stability prediction	last year
 images	deleted a DS_store	4 years ago
 libraries	Delete .DS_Store	4 years ago
 list_content	bug fix for CoRE MOF dropdown, regarding refcodes with m...	4 years ago
 model	working C2 predictions	7 months ago

```

1  data_crystal
2
3  _cell_length_a 15.52000
4  _cell_length_b 13.89000
5  _cell_length_c 13.07690
6  _cell_angle_alpha 90.00000
7  _cell_angle_beta  90.00000
8  _cell_angle_gamma 90.00000
9
10 _symmetry_space_group_name_ 'P 1'
11 _symmetry_space_group_name_H-M 'P 1'
12
13 loop_
14 _symmetry_equiv_pos_as_xyz
15   x,y,z
16
17 loop_
18 _atom_site_label
19 _atom_site_type_symbol
20 _atom_site_fract_x
21 _atom_site_fract_y
22 _atom_site_fract_z
23 _atom_site_charge
24 U      U      0.26136 0.98959 0.43733 0.04000
25 U      U      0.27438 0.73750 0.58520 0.04000
26 P      P      0.27580 0.51870 0.17570 0.47800
27 P      P      0.31290 0.73950 0.34590 0.46600
28 O      O      0.30510 0.82500 0.42040 -0.12400
29 O      O      0.23300 0.59360 0.10390 -0.12500
30 O      O      0.27980 0.43140 0.10210 -0.12500
31 O      O      0.28460 0.65370 0.41440 -0.12500
32 O      O      0.21740 0.49460 0.26480 -0.13900
33 O      O      0.24910 0.75210 0.25540 -0.13200
34 O      O      0.40460 0.72760 0.30870 -0.13000
35 O      O      0.36330 0.54970 0.20960 -0.14600
36 U      U      0.76136 0.48959 0.43733 0.04000
37 U      U      0.77438 0.23750 0.58520 0.04000
38 P      P      0.77580 0.01870 0.17570 0.47600











```

charg정보가 기입되어있는것을
확인할 수 있었다.

그러나, 이 charge정보를 어떤 방식
으로 구했는지는
MOFSimplify 논문을 찾아봤지만
알아낼 수 없었다.

NET CHARGE 문제를 해결할 수 있는 방법

<https://github.com/hjkgrp/MOFSimplify>

 gianmarco-terrones updated citation information	9868303 · 3 months ago	 426 Commits
 CoRE2019	options for user to download information on latent space ne...	4 years ago
 TGA	made all readme files full caps in their name	4 years ago
 VM_info	removed all __pycache__ folders and adjusted README.rtf fil...	4 years ago
 environments	water stability prediction	last year
 images	deleted a DS_store	4 years ago
 libraries	Delete_DS_Store	4 years ago
 list_content	bug fix for CoRE MOF dropdown, regarding refcodes with m...	4 years ago
 model	working C2 predictions	7 months ago

- define charges via the CIF-file

If you want a possibly different charge for each atom, then use the option:

```
UseChargesFromCIFFile yes
```

and define the charge using the field 'atom_site.charge' in the CIF-file. Atom-types from the CIF-file that are not defined in the 'pseudo.atom.def' are automatically added, atoms that are already defined as a type in the 'pseudo.atom.def' get the charge from the CIF-file. In the output-file in the list of pseudo-atoms you will see e.g.

```
Charge=0.111115012 (av)
```

which signals that for this atom-type the averages charge is listed (because each atom potentially can have a different value in this case). This is a typical case for simulations based on CHelpG charges from quantum.

- define charges via the 'pseudo.atom.def' file

If you want the same charge for all atoms of the atom-type, then you can list all of these in the 'pseudo.atom.def' file and use

```
UseChargesFromCIFFile no
```

which is the default. Any atoms with a type known in the CIF-file will get a charge given in the 'pseudo.atom.def' file; atoms of unknown type will be added to the pseudo-atoms but with a charge of zero. The latter is probably not what you want, so make sure you have listed all atom type in 'pseudo.atom.def' file.

```

1 data_crystal
2
3 _cell_length_a 15.52000
4 _cell_length_b 13.89000
5 _cell_length_c 13.07690
6 _cell_angle_alpha 90.00000
7 _cell_angle_beta 90.00000
8 _cell_angle_gamma 90.00000
9
10 _symmetry_space_group_name_ 'P 1'
11 _symmetry_space_group_name_H-M 'P 1'
12
13 loop_
14 _symmetry_equiv_pos_as_xyz
15   x,y,z
16
17 loop_
18 _atom_site_label
19 _atom_site_type_symbol
20 _atom_site_fract_x
21 _atom_site_fract_y
22 _atom_site_fract_z
23 _atom_site_charge
24 U U 0.26136 0.98959 0.43733 0.04000
25 U U 0.27438 0.73750 0.58520 0.04000
26 P P 0.27580 0.51870 0.17570 0.47800
27 P P 0.31290 0.73950 0.34590 0.46600
28 O O 0.30510 0.82500 0.42040 -0.12400
29 O O 0.23300 0.59360 0.10390 -0.12500
30 O O 0.27980 0.43140 0.10210 -0.12500
31 O O 0.28460 0.65370 0.41440 -0.12500
32 O O 0.21740 0.49460 0.26480 -0.13900
33 O O 0.24910 0.75210 0.25540 -0.13200
34 O O 0.40460 0.72760 0.30870 -0.13000
35 O O 0.36330 0.54970 0.20960 -0.14600
36 U U 0.76136 0.48959 0.43733 0.04000
37 U U 0.77438 0.23750 0.58520 0.04000
38 P P 0.77580 0.01870 0.17570 0.47600

```

charg정보가 기입되어있는것을 확인할 수 있었다.

그러나, 이 charge정보를 어떤 방식으로 구했는지는 MOFSimplify 논문을 찾아봤지만 알아낼 수 없었다.

또한 이 cif내의 charge 정보를 그대로 이용하기 위해서는 charge정보를 pseudo_atoms가 아닌 cif를 통해 참조하도록 하는 설정을 해야한다. 이때는 .input파일의 UseChargesFromCIFFILE 설정을 yes로 하면 된다. 이 설정은 default로 no로 설정된다고 하여 명시적으로 yes를 하여 cif를 통해 charge를 참조한다.

```

6
7 Forcefield GarciaPerez2006ForceField
8 UseChargesFromCIFFile yes
9

```

ZOFZUR_clean.cif		모든 원소 charge가 0	NET CHARGE 일부 원소만 charge 누락	
	나현누나 결과	내결과 (atom_site_label이 원소+숫자)	내결과 (atom_site_label에서 숫자 제거)	UseChargesFromCIF설정 후
mol/kg	19.8	1.779	19.8	7.340296839

Charge정보를 cif에서 가져오도록 했을 때, NET CHARGE가 발생했을때와 모든 charge를 0으로 할당했을때의 중간정도의 값인 7 mol/kg정도가 나왔다

결론 :

나현누나와 똑같은 시뮬레이션 결과가 나오도록 설정하는 방법을 찾았습니다.

하지만 일부 원소의 pseudo atom charge 누락으로 인한 NETCHARGE문제가 발생하는 MOF들이 간혹 있었고 이 경우에 흡착량이 매우 크게 나오는 현상이 발견되었습니다.

이런 경우에 시뮬레이션의 신뢰도가 떨어지는 것이 맞는지
그렇다면 cif의 charge를 그대로 가져와서 사용하는 것이 올바른 방향이 맞을지 **피드백 받고 싶습니다.**