

Calculate pKas of Lysozyme (MCCE steps 1 - 4)

Welcome to MCCE Tutorial

This tutorial will show you how to set up, run and interpret the results of the PDB file 4LZT (Hen Egg White Lysozyme). MCCE must be installed- [see installation instructions here](#).

The output includes: ALWAYS the protonation states of all Asp, Glu, Arg, Lys, His, Tyr, Cys (and His tautomer) at a pH of interest. The pKa of these residues includes information about why pKas are shifted within the protein from the values of the isolated residues in solution.

`lysozyme` bound or type unknown

Crystal structure of hen egg white lysozyme - PDB ID 4LZT. MCCE starts with atomic coordinates in PDB format.

The experimental pKas of all Lysozyme residues is provided at the bottom of this tutorial. Here we will focus on two residues in the active site that have perturbed pKas. Residue GLU 35 and ASP 52 are in the active sites. The pKas of these two residues play an important role in the enzyme's activity. In order for lysozyme to attack the glucose molecule of the substrate, GLU needs a high pKa and ASP 52 a low pKa. GLU 35 acts as a proton donor, cutting the glucose with protonation of the glycosidic oxygen and a deprotonated ASP 52 stabilizes the highly charged intermediate, making the reaction easier.

[Jens Erik Nielsen and J. Andrew McCammon, Protein Sci. 2003 Sep; 12\(9\): 1894-1901](#)

Prepare the calculation

After the program is installed and the environment variable PATH is configured (see [link](#) for environment configuration details), make a working directory and go to the directory. EACH MCCE RUN IS DESIGNED TO RUN IN ITS OWN DIRECTORY. MCCE will generate intermediate files and result files in the current directory. If you run a new protein, or change conditions, make a new directory.

```
$ mkdir 4lzt  
$ cd 4lzt
```

Then download pdb file 4LZT from rscb.org to the working directory:

```
jmao@jmao-desktop ~/4lzt $ getpdb 4lzt  
Saving as 4LZT.pdb ...  
Download completed.
```

p_info - How does MCCE view the protein before a run?

```
p_info 4lzt.pdb
```

p_info is a tool found in the MCCE_bin. It will tell you:

1. How large the protein is, to estimate computation time.
2. If there are ligands not recognized by MCCE (what do with unknown ligands: [INSERT LINK HERE](#))
3. If MCCE is modifying your protein. For example, it will make the N and C termini ionizable.

Explicit water in the PDB file. We suggest removing explicit waters from the input pdb file. The protonation states and pKas with and without are similar, but without waters the calculation is much faster. **By default, MCCE runs remove most waters from PDB files.** If necessary, waters can be removed manually from 4LZT using the following command:

```
grep -v HOH 4lzt.pdb > 4lzt_noHOH.pdb
```

run_mcce and Key Output Files

```
run_mcce 4lzt.pdb > run.log &
```

run_mcce executes the first four steps of MCCE on a chosen protein. After this command is finished running, you should have a full directory, including key output files **fort.38**, **sum_crg.out**, and **pK.out**.

pK.out has information about how the protein shifts the pK of each residue. To focus on the pKa, use the command

```
cut -c1-48 pK.out
```

In a solution, residues have defined pKs, which are then modified by the protein. All "Arg" residues should have high pKs.

sum_crg.out gives the charge at each pH. It's derived from **fort.38** which gives the % of each conformer at each pH.

Run the four steps of MCCE

MCCE IS MODULAR AND CAN BE STOPPED AND MODIFIED AFTER EACH OF THE FOUR STEPS. Again, all steps should be run in one directory associated with the protein to be run. Step N needs the output of Step N - 1. You can rerun later steps without running from the beginning. E.g. if you have run steps 1-4 you can rerun step 4. This will overwrite the original step 4's output.

The four main steps are:

1. Read in, fix and re-name PDB file
2. Make conformers using the instructions in head1.lst
3. Calculate self and pairwise interactions for all conformers in step2_out.pdb
4. Run Monte Carlo (can control pH or eH limits of titration; can fix individual conformers on or off)

Breakdown of Four MCCE Steps

Step 1. Convert PDB file into MCCE PDB

Step 1 reformats the PDB file. Missing side chain atoms will be added back. Residues that are not in the MCCE lexicon will be flagged. Explicit waters with greater than 5% surface exposure will be stripped off. Chain termini are separated to be NTR and CTR and are treated as protonatable. Other changes in residue naming are made using name.txt.

```
step1.py 4lzt.pdb
```

This command generates step1_out.pdb which is required for step 2.

NOTE: All steps have a "-h" flag to provide additional information. For example, "step1.py -h" will give you all possible options for that step.

Step 2. Make side chain conformers

This step makes alternative side chain locations and ionization states.

```
step2.py
```

This command generates step2_out.pdb which is required of step 3.

Step 3. Make energy table

This step calculates conformer self energy and pairwise interaction table.

```
step3.py
```

This command generates opp files under energies/ folder and file head3.lst which are required of step 4. As part of the energy calculations, the Poisson-Boltzmann equation is solved, and various solvers are available to choose. step3.py takes the most computation time of any MCCE step.

If you want to know the help information and other options of this command:

```
step3.py -h
```

Step 4. Simulate a titration with Monte Carlo sampling

This step simulates a titration and writes out the conformation and ionization states of each side chain at various conditions.

```
step4.py --xts
```

- fort.38. is the primary output file. It gives the average occupancy of each conformer.
- sum_crg.out give the net charge on all residues. Residues where all conformers have a zero charge are not reported in this file.
- pK.out calculates the best single pKa for each residue given the change in charge with pH in sum_crg.out
- The flag "--xts" enables entropy correction

Results

The pKa report is in file pK.out.

```
cat pK.out
```

From the result, GLU 35 (pKa = 5.05) has a higher pKa than ASP 52 (pKa = 3.15).

To analyze the ionization energy of an ionizable residue at the mid point pH=5.05 with pairwise cutoff 0.1, we use another program. mfe.py calculates the mean field energy ionization energy on ionizable residues at a specific pH/eH:

```
$ mfe.py ASP-A0052_ -c 0.1
Residue ASP-A0052_ pKa/Em=3.152
=====
Terms           pH      meV      Kcal
-----
```

```

vdw0      -0.01  -0.85  -0.02
vdw1       0.00   0.23   0.01
tors      -0.10  -5.86  -0.14
ebkb      -1.27 -73.44  -1.73
dsol       1.99 115.38   2.71
offset    -0.62 -36.17  -0.85
pH&pK0     1.60  92.75   2.18
Eh&Em0     0.00   0.00   0.00
- TS       0.00   0.00   0.00
residues  -1.36 -79.01  -1.86
*****
TOTAL      0.22  13.04   0.31  sum_crg
*****
ASNA0044_  -0.46 -26.92  -0.63   0.00
ARGA0045_  -0.11  -6.36  -0.15   1.00
ASNA0046_  -0.19 -11.09  -0.26   0.00
ASPA0048_   0.50  28.96   0.68  -0.96
SERA0050_   0.13   7.28   0.17   0.00
GLNA0057_   0.22  12.78   0.30   0.00
ASNA0059_  -0.99 -57.18  -1.34   0.00
ARGA0061_  -0.26 -15.37  -0.36   1.00
ASPA0066_   0.27  15.72   0.37  -0.94
ARGA0112_  -0.19 -11.11  -0.26   1.00
ARGA0114_  -0.10  -5.86  -0.14   1.00
=====

```

This reports the ionization free energy breakdown. The TOTAL free energy should be close to 0 at the midpoint pH 3.15 (pKa).

To analyze the ionization energy of this residue pH 7 with pairwise cutoff 0.1:

```

$ mfe.py ASP-A0052_ -p 7 -c 0.1
Residue ASP-A0052_ pKa/Em=3.152
=====
Terms          pH      meV    Kcal
-----
vdw0          -0.02   -0.97   -0.02
vdw1          -0.01   -0.52   -0.01
tors          -0.11   -6.31   -0.15
ebkb          -1.26  -73.20  -1.72
dsol           1.97  114.32   2.69

```

```

offset      -0.62  -36.17  -0.85
pH&pK0      -2.25 -130.60  -3.07
Eh&Em0       0.00   0.00   0.00
- TS         0.00   0.00   0.00
residues     -1.89 -109.87  -2.58
*****
TOTAL        -4.19 -243.32  -5.72  sum_crg
*****
GLUA0035_    0.85   49.05   1.15  -0.98
ASNA0044_   -1.00  -57.87  -1.36   0.00
ARGA0045_   -0.10   -6.00  -0.14   1.00
ASNA0046_   -0.86  -49.87  -1.17   0.00
ASPA0048_    0.51   29.70   0.70  -1.00
SERA0050_    0.13    7.70   0.18   0.00
GLNA0057_    0.18   10.59   0.25   0.00
ASNA0059_   -1.20  -69.84  -1.64   0.00
ARGA0061_   -0.27  -15.88  -0.37   1.00
ASPA0066_    0.29   17.04   0.40  -1.00
ARGA0112_   -0.20  -11.53  -0.27   1.00
=====

```

At pH=7, the free energy is -4.19. This means the ionization ASP to ASP- is more favored than at the midpoint pH=3.15.

A more detailed explanation of mfe.py program can be found [here \(To be done\)](#)

Benchmark pKas for Lysozyme

There are 20 experimentally measured pKas in hen white lysozyme.

- Bartik, K., C. Redfield, and C.M. Dobson, Biophys J, 1994. 66(4): p. 1180-4.
- Kuramitsu, S. and K. Hamaguchi, J Biochem (Tokyo), 1980. 87(4): p. 1215-9.
- Takahashi, T., H. Nakamura, and A. Wada, Biopolymers, 1992. 32: p. 897-909.

These pKa values have been used to benchmark MCCE and other programs that calculate pKas. For example:

- Sham, Y. Y., I. Muegge, and A. Warshel. 1999. Simulating proton trans- locations in proteins: probing proton transfer pathways in the Rhodobacter sphaeroides reaction center. Proteins. 36:484-500.
- You, T. J., and D. Bashford. 1995. Conformation and hydrogen ion titration of proteins: a continuum electrostatic model with conformational flexi- bility. Biophys. J. 69:1721-1733.

- Antosiewicz, J., J. A. McCammon, and M. K. Gilson. 1996. The determinants of pKa's in proteins. *Biochemistry*. 35:7819–7833.

Experimentally measured pKas of residues in Lysozyme

Residue	pKa
LYS 1	10.8
GLU 7	2.85
LYS 13	10.5
HIS 15	5.36
ASP 18	2.66
TYR 20	10.3
TYR 23	9.8
LYS 33	10.4
GLU 35	6.2
ASP 48	1.6
ASP 52	3.68
ASP 66	0.9
ASP 87	2.07
LYS 96	10.8
LYS 97	10.3
ASP 101	4.08
LYS 116	10.2
ASP 119	3.2
CTR	2.75

Revision #13

Created 10 March 2023 17:00:55 by Junjun Mao

Updated 12 May 2025 19:48:09 by Jared