How to break down the energy of a residue ionization?

MCCE calculates theoretical pKas as well as explain why these residues have those pKas.

Continuum Electrostatic interpretation of pKa

Free energy and pKa

An amino acid and conjugate base in equilibrium:

$$AH <=> A^- + H^+$$

The free energy of the reaction has relationship with the equilibrium of:

$$\Delta G = \Delta G_0 + 2.303 RT \lg(\frac{[A^-][H^+]}{[HA]})$$

At the equilibrium of midpoint where

$$[A^-] = [HA]$$

therefore residue pKa is linked to the reaction free energy:

$$0 = \Delta G_0 + 2.303RT\lg([H^+])$$

$$pKa = -\lg([H^+]) = \frac{\Delta G_0}{2.303RT}$$

The acids in solution and in protein

We know the ionization energy in solution, which is represented by the solution pKa:

$$AH \xrightarrow{\Delta G1} A^- + H^+ \qquad in\ solution$$
$$\downarrow \Delta G2 \qquad\qquad\qquad \downarrow \Delta G3$$
$$[Protein]|AH \xrightarrow{\Delta G4} [Protein']|A^- + H^+ \qquad in\ protein$$
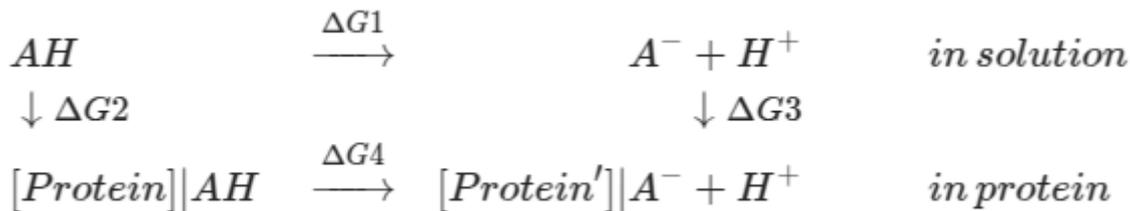
$\Delta G1$ is solution pKa, and $\Delta G4$ is acid pKa in protein, what we are looking for. $\Delta G2$ and $\Delta G3$ involve the free energy of moving acid from solution to protein environment, interaction between the acid and the protein, and the interaction within the protein at AH and A^- states.

The desolvation energy, pairwise interactions are all calculated by continuum electrostatic methods.

MCCE's way to compute pKa

MCCE makes residues as independently ionizable pieces. All residues are given freedom to change conformation and ionization if possible. Residue pairwise interactions are precalculated. The choice of a residue being at which position and ionization state is statistically counted by Monte Carlo sampling.

A pKa calculation is a simulated titration that carried out at multiple pH points. The residue charges change with pH, and their pKas are their midpoints determined by fitting the titration curve.

Therefore the calculated pKa of a residue is a statistical result that came from the simultaneous interactions of many factors. To understand the pKa, we need to decipher the factors with the help of mean field energy analysis.

Understanding the residue pKa and factors that affect pKa

Mean field approximation

The accurate interaction calculation is only possible at microstate level. At the end, what we know is the probability of a residue being at certain conformation and ionization.
The interaction between residues calculated after Monte Carlo sampling os approximated by "mean field".

For example, 4 possible interactions for a pair of acids could be:

$$A_1H \leftrightarrow A_2H \tag{1}$$

$$A_1H \leftrightarrow A_2^- \tag{2}$$

$$A_1^- \leftrightarrow A_2H \tag{3}$$

$$A_1^- \leftrightarrow A_2^- \tag{4}$$
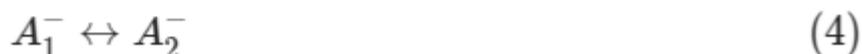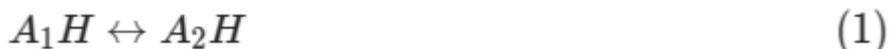
If MCCE sampled acids $A1$ and $A2$ are both at 50% ionization, after MOnte Carlo sampling, from the ionization rate alone, we could not know the interaction is 50% of (2) and 50% of (3), 50% of (1) and 50% of (4), or a mix.

This is the limitation of mean fields energy analysis. Once we know the limitation, we can extract quite useful information of residue ionization.

MFE energy terms

Once step 4 is done, one can run mfe.py on a residue to break down the energy terms of a ionization free energy. The input residue has to be in file pK.out.

For example, in lysozyme pKa calculation, we obtained pKa output file pK.out.

```
$ cat pK.out
  pH              pKa/Em  n(slope) 1000*chi2     vdw0     vdw1     tors     ebkb     dsol
offset  pHpK0   EhEm0    -TS    residues    total
NTR+A0001_       5.095     0.985     0.015    -0.00    -0.01    -0.26     0.52     4.01     -
0.95    -2.90   0.00     0.00    -0.11      0.29
LYS+A0001_       9.593     0.998     0.027    -0.00    -0.00     0.00     0.15     0.46
0.29    -0.81   0.00    -0.00    -0.09      0.00
ARG+A0005_      12.695     0.858     0.051    -0.01    -0.00    -0.00    -0.73     0.89
0.00     0.20   0.00     0.14    -0.35      0.14
GLU-A0007_       3.192     0.944     0.010    -0.01     0.00    -0.19    -1.27     1.17     -
0.22     1.56   0.00     0.30    -1.02      0.31
LYS+A0013_      11.322     0.954     0.019    -0.00    -0.00     0.00     0.19     0.61
```

```
0.29    0.92    0.00    0.00    -2.01    0.00
ARG+A0014_      13.330     0.934     0.004    -0.01   -0.00    0.00    0.07   -0.28
0.00    0.83    0.00    0.43    -0.60     0.44
HIS+A0015_       6.816     0.988     0.065    -0.01   -0.01    0.00    1.44    1.33   -
0.73   -0.16    0.00    0.36    -0.42     1.80
ASP-A0018_       2.013     0.929     0.127    -0.01   -0.00   -0.16   -2.03    0.95   -
0.62    2.74    0.00    0.40    -0.86     0.40
TYR-A0020_      13.498     0.687     0.401     0.00    0.00    0.00    0.24    3.29   -
0.37   -3.30    0.00    0.18     0.04     0.09
ARG+A0021_      13.239     0.796     0.092    -0.01   -0.00    0.00   -0.09    0.11
0.00    0.74    0.00    0.28    -0.74     0.29
TYR-A0023_      10.312     0.863     0.020     0.00    0.00    0.00   -1.17    1.98   -
0.37   -0.11    0.00    0.29    -0.33     0.29
LYS+A0033_      10.581     0.974     0.026    -0.00   -0.00    0.00    0.06    0.61
0.29    0.18    0.00    0.00    -1.09     0.05
GLU-A0035_       5.133     0.880     1.144    -0.04   -0.01   -0.01   -1.26    2.16   -
0.22   -0.38    0.00    0.26    -0.23     0.26
ARG+A0045_      12.919     0.796     1.608    -0.03   -0.00    0.00    0.19   -0.32
0.00    0.42    0.00    0.45    -0.24     0.46
ASP-A0048_       1.528     0.963     0.058    -0.01    0.00   -0.24   -2.52    3.22   -
0.62    3.22    0.00    0.14    -3.01     0.17
ASP-A0052_       3.330     0.859     0.170    -0.01    0.00   -0.10   -1.32    1.99   -
0.62    1.42    0.00    0.46    -1.20     0.61
TYR-A0053_      >14.0                          0.00    0.00    0.00   -1.70    4.45   -
0.37   -3.80    0.00    0.00     8.03     6.61
ARG+A0061_      >14.0                         -0.01   -0.00    0.00    0.69    0.15
0.00    1.50    0.00    0.42    -2.43     0.31
ASP-A0066_       2.066     3.239     0.002    -0.01    0.02   -0.04   -5.30    7.34   -
0.62    2.68    0.00   -0.11    -4.50    -0.55
ARG+A0068_      13.604     0.768     0.028     0.12    1.81    0.16    0.29   -0.80
0.00    1.10    0.00   -0.19    -0.05     2.46
ARG+A0073_      12.750     0.935     0.132    -0.01   -0.00    0.00   -0.12    0.10
0.00    0.25    0.00    0.35    -0.22     0.35
ASP-A0087_       2.073     0.960     0.014    -0.02    0.01   -0.21   -1.88    1.24   -
0.62    2.68    0.00    0.31    -1.18     0.31
LYS+A0096_      10.631     0.831     0.307    -0.00    0.00    0.00   -1.52    1.26
0.29    0.23    0.00    0.00    -0.25     0.01
LYS+A0097_      10.813     0.918     0.072    -0.00   -0.00    0.00   -0.27    0.07
0.29    0.41    0.00    0.00    -0.51    -0.00
ASP-A0101_       3.988     1.009     0.010    -0.00   -0.05   -0.10    0.04    1.15   -
```

```
 0.62    0.76    0.00    0.57   -1.17      0.58
ARG+A0112_        12.956     0.923     0.080    -0.01   -0.00    0.00   -0.27    0.77
 0.00    0.46    0.00    0.55   -0.90      0.59
ARG+A0114_        13.257     0.968     0.050    -0.01   -0.00    0.00   -0.09   -0.29
 0.00    0.76    0.00    0.34   -0.35      0.35
LYS+A0116_         9.554     0.941     0.042    -0.00   -0.00    0.00    0.12    0.43
 0.29   -0.85    0.00    0.00    0.02      0.01
ASP-A0119_         3.596     1.005     0.031    -0.01    0.00   -0.24   -1.08    1.95   -
 0.62    1.15    0.00    0.09   -1.15      0.09
ARG+A0125_        13.083     0.915     0.005    -0.01   -0.00   -0.00    0.36   -0.18
 0.00    0.58    0.00    0.45   -0.75      0.45
ARG+A0128_        13.586     0.939     0.008     0.00   -0.00    0.00    0.08   -0.81
 0.00    1.09    0.00    0.45   -0.36      0.45
CTR-A0129_         2.464     0.905     0.072    -0.01   -0.01   -0.12   -0.06    1.66
 0.00    1.29    0.00    0.49   -2.75      0.49
```

Now we would like to know why ASP52 has a low pKa of 3.33

```
$ mfe.py ASP-A0052_ -c 0.1
Residue ASP-A0052_ pKa/Em=3.33
=================================
Terms         pH     meV    Kcal
---------------------------------
vdw0        -0.01   -0.85   -0.02
vdw1         0.00    0.21    0.00
tors        -0.10   -5.89   -0.14
ebkb        -1.32  -76.87   -1.81
dsol         1.99  115.58    2.72
offset      -0.62  -36.17   -0.85
pH&pK0       1.42   82.42    1.94
Eh&Em0       0.00    0.00    0.00
-TS          0.00    0.00    0.00
residues    -1.20  -69.73   -1.64
*******************************
TOTAL        0.15    8.70    0.20  sum_crg
*******************************
ASNA0044_   -0.44  -25.31   -0.59    0.00
ARGA0045_   -0.12   -6.74   -0.16    1.00
```

```
ASNA0046_   -0.16    -9.25    -0.22     0.00
ASPA0048_    0.51    29.65     0.70    -0.96
SERA0050_    0.12     7.04     0.17     0.00
GLNA0057_    0.22    12.63     0.30     0.00
ASNA0059_   -0.99   -57.31    -1.35     0.00
ARGA0061_   -0.27   -15.65    -0.37     1.00
ASPA0066_    0.39    22.84     0.54    -1.00
ARGA0112_   -0.19   -11.15    -0.26     1.00
ARGA0114_   -0.10    -5.86    -0.14     1.00
================================
```

Program computes the ionization free energy, specifically, the free energy difference between the ionized ASP and the neutral ASP.

If we don't specify pH, mfe.py uses the midpoint pH or Eh.

Energy terms:

- **vdw0:** Van der Waals interaction within the residue. Since the vdw interaction is similar to ionized and neutral ASP, the contribution to ΔG of ionization is small.
- **vdw1:** Van der Waals interaction to the protein backbone. Again the contribution to pKa is usually small.
- **tors:** Side chain torsion energy. The contribution to pKa is usually small.
- **ebkb:** The protein backbone electrostatic interaction. The protein secondary structure especially helices have a dipole that may affect the ionized residue more than neutral form of the residue, therefore it could be a factor of pKa.
- **dsolv:** The desolvation energy. The ionized residue is less stabilized im protein than in solution. This makes the ionization inside protein harder and it links to a positive free energy to the reaction from neutral residue to ionized residue.
- **offset:** The is a system correction of pKa calibrated by benchmark. It captures the terms not counted by MCCE.
- **pH&pK0:** Solution pH effect to the ionization. It is environment pressure to the residue ionization. To an acid, low solution pH makes the ionization (releasing proton) easy so it contributes as a favorable energy. To a base, low pH makes ionization harder. When pH equals residue's solution pKa, the environment pH is at the balance point, where the contribution is 0.
- **Eh&Em0:** Environment Eh effect to redox reaction. This works similarily as pH&pK0.
- **-TS:** Entropy term. The number of rotamers of neutral and ionized residues generated by MCCE may be different. The effect of number of rotamers on two ionization states act like entropy. Since this may not be a desired effect, one can enable entropy correction in step 4, Monte Carlo sampling. If entropy correction is enabled in MC, the entropy effect has been eliminated and entropy should be set to 0 in MFE analysis. Program mfe.py is able to know how MC is done and handle this accordingly, or one can turn on or off in MFE manually.

- **residues:** Total pairwise interaction from other residues. Other residues may shift the ionization free energy depending on their dipole orentation and charge.
- **TOTAL:** Total free enegy of ionization reaction. It the sum of all above terms.
- **Redisue breakdown:** Individual residue contribution. This part is the breakdown of term **residues**.

At the midpoint, $[AH]=[A-]$. This means conformers of AH and conformers of A^- strike a balance, and the transition energy ΔG between them is 0 under this condition. The value ΔG in pH unit also means how far away the pKa is from the solution pH. At the midpoint, $\Delta G=0$ and residue pKa is the same as envrionment pH.

In the above example, ASP52 has low pH because a pretty big stabilization from ASN59.

We can also do mfe anaylysis off the midpoint, for example at physiological pH about 7. The command option "-p 7" sets the mfe point.

```
$ mfe.py ASP-A0052_ -c 0.1 -p 7
Residue ASP-A0052_ pKa/Em=3.33

================================
Terms          pH     meV    Kcal

--------------------------------
vdw0         -0.02   -0.97   -0.02
vdw1         -0.01   -0.52   -0.01
tors         -0.11   -6.30   -0.15
ebkb         -1.32  -76.59   -1.80
dsol          1.97  114.53    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     -1.81 -105.25   -2.47
********************************
TOTAL        -4.17 -241.87   -5.68  sum_crg
********************************
GLUA0035_     0.85   49.08    1.15   -0.98
ASNA0044_    -1.00  -57.84   -1.36    0.00
ARGA0045_    -0.11   -6.28   -0.15    1.00
ASNA0046_    -0.85  -49.43   -1.16    0.00
ASPA0048_     0.52   30.08    0.71   -1.00
SERA0050_     0.13    7.43    0.17    0.00
GLNA0057_     0.18   10.47    0.25    0.00
ASNA0059_    -1.20  -69.75   -1.64    0.00
```

```
ARGA0061_    -0.28  -16.15    -0.38    1.00
ASPA0066_     0.40   22.96     0.54   -1.00
ARGA0112_    -0.20  -11.56    -0.27    1.00
================================
```

We see ASP52 is surrounded by these polar residues.gets extra stabilization from ASN44 and ASN46 besides ASN59. The stabilization is bigger at pH 7. Actually the back calculated pKa $7-4.17=2.83$ is lower than titrated value 3.33, indicating extra stabilization at pH 7.

With MFE analysis, one can quickly determine what residues affect the pKa most therefore identify the influential sites.

Revision #1
Created 3 June 2025 15:37:42 by Junjun Mao
Updated 3 June 2025 16:15:12 by Junjun Mao