Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Supporting Information: Sulfur-Substitution-Induced Base Flipping in DNA duplex

Zhaoxi Sun[1,2*], Xiaohui Wang[1,3], John Z.H.Zhang[1,4,5], and Qiaole He[6,7*]

[1]*State Key Laboratory of Precision Spectroscopy, School of Chemistry and Molecular Engineering, East China Normal University, Shanghai 200062, China*

[2]*Computational Biomedicine (IAS-5/INM-9), Forschungszentrum Jülich, Jülich 52425, Germany*

[3]*Institute of Computational Science, Università della Svizzera italiana (USI), Via Giuseppe Buffi 13, CH-6900, Lugano, Ticino, Switzerland*

[4]*NYU-ECNU Center for Computational Chemistry at NYU Shanghai, Shanghai 200062, China*

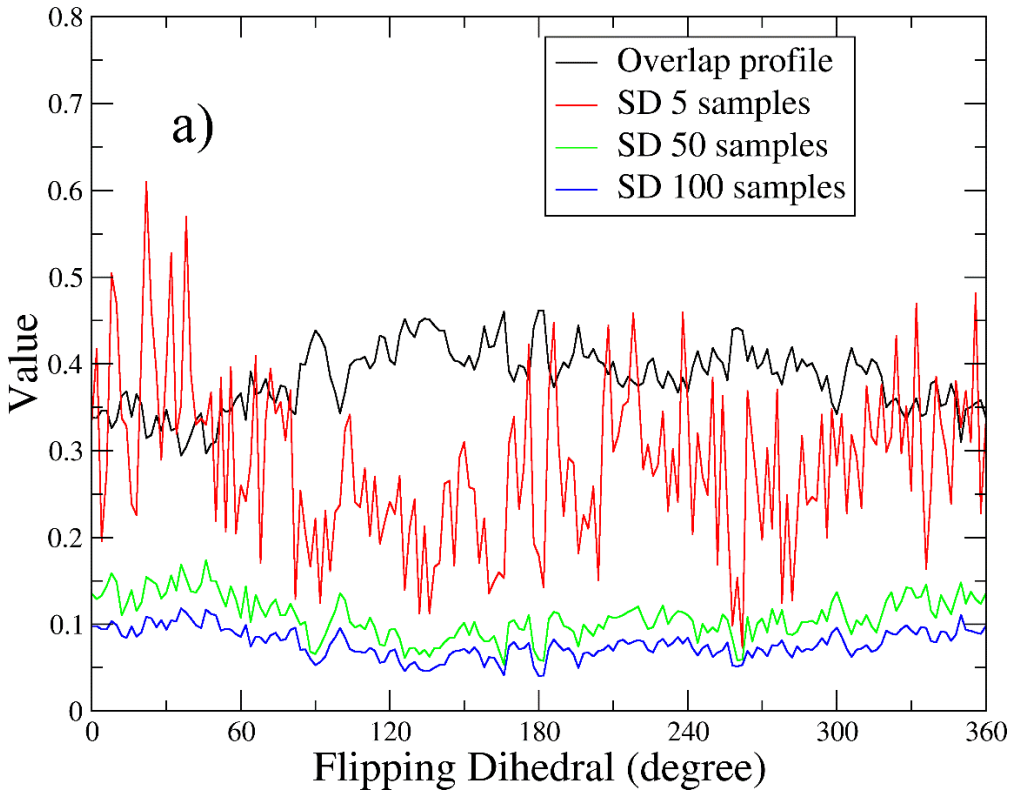
[5]*Department of Chemistry, New York University, NY, NY 10003, USA*

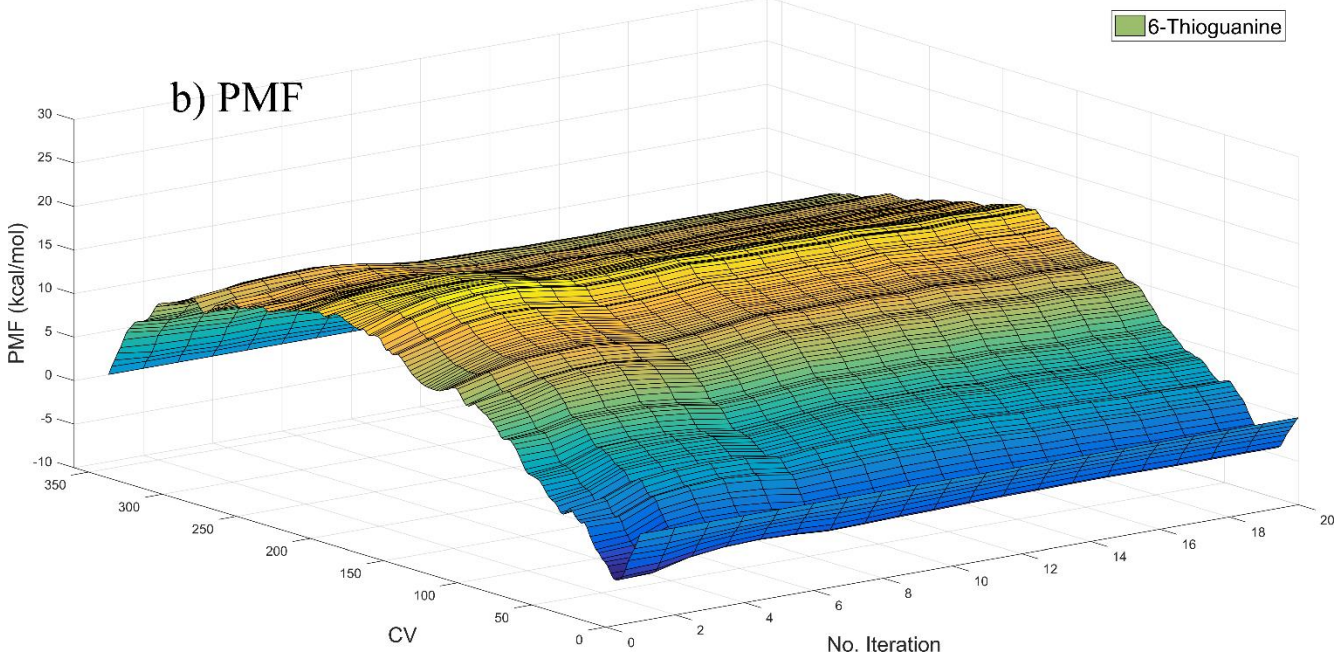
[6]*Forschungszentrum Jülich GmbH, IBG-1: Biotechnology, Wilhelm-Johnen-Str. 1, 52425 Jülich, Germany*

[7]*State Key Laboratory of Bioreactor Engineering, R&D Center of Separation and Extraction Technology in Fermentation Industry, East China University of Science and Technology, Shanghai 200237, China*

*To whom correspondence should be addressed: z.sun@fz-juelich.de, qiaole.he@rwth-aachen.de

Fig. S1. In the sulfur-substitution case described with the OL15 force field, a) the comparison between dimensionless SD profiles and overlap profile in nonequilibrium stratification, b) the convergence behavior of free energy profiles from nonequilibrium stratification on the sample size, and c) the time-evolution of state-specified SD. The standard deviation in the *i*th state is the sum of the components contributed by samples initiated from that state, namely $\sigma_i = \sqrt{\frac{\mathrm{Var}\left(f_{i,i+1}\right)}{n_i^2 f_{i,i+1}^2} + \frac{\mathrm{Var}\left(f_{i-1,i}\right)}{n_i^2 f_{i-1,i}^2}}$. In b), the initial sample size is 5 and in each iteration further 5 samples are added to the dataset. There are small differences between the 5-sample PMF and the later ones. Since the 10th iteration, the fluctuation of the PMF is small. Thus, we define this sample size (50 samples) as the minimum sample size required for convergence.





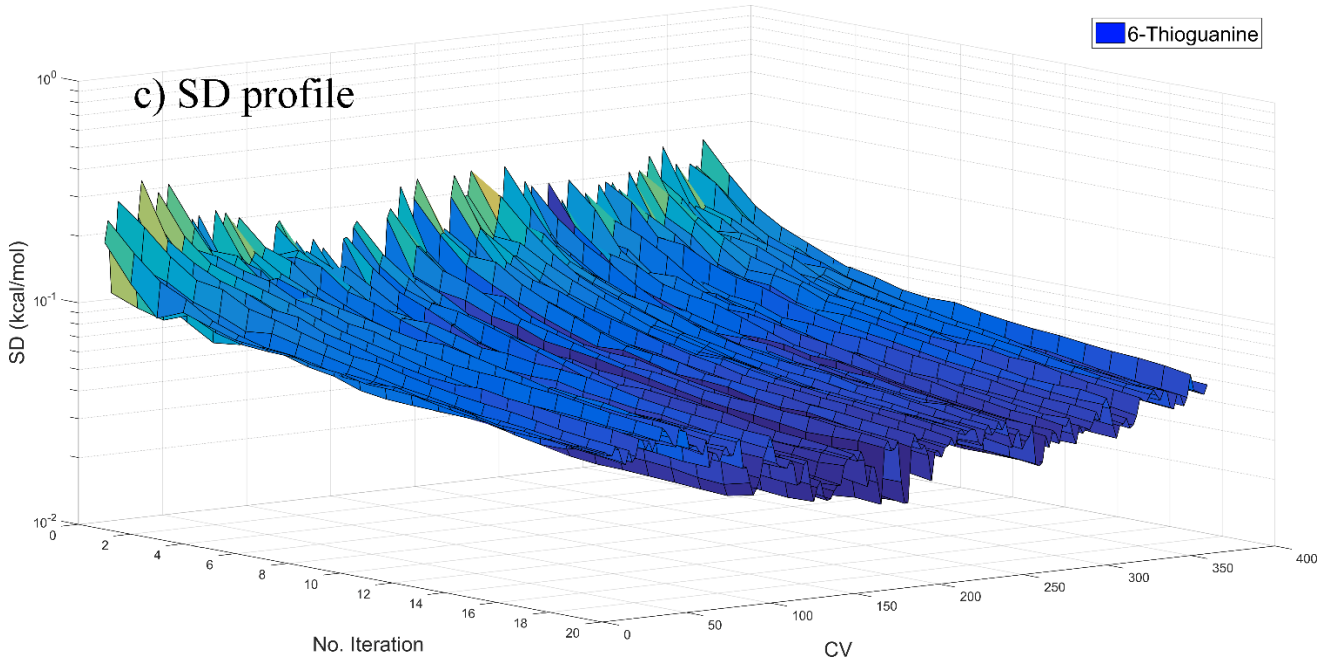

6-Thioguanine
c) SD profile
SD (kcal/mol)
10^{0}
10^{-1}
10^{-2}
0
2
4
6
8
10
12
14
16
18
20
No. Iteration
0
50
100
150
200
250
300
350
400
CV

Fig. S2. Under the force field of bsc1, the convergence behavior of free energy profiles constructed from equilibrium umbrella sampling simulations with vFEP reweighting in a-b) the wild-type GC base-paired case and c-d) the thioguanine substituted case. The lengths of time blocks are 2 ns and 4 ns.

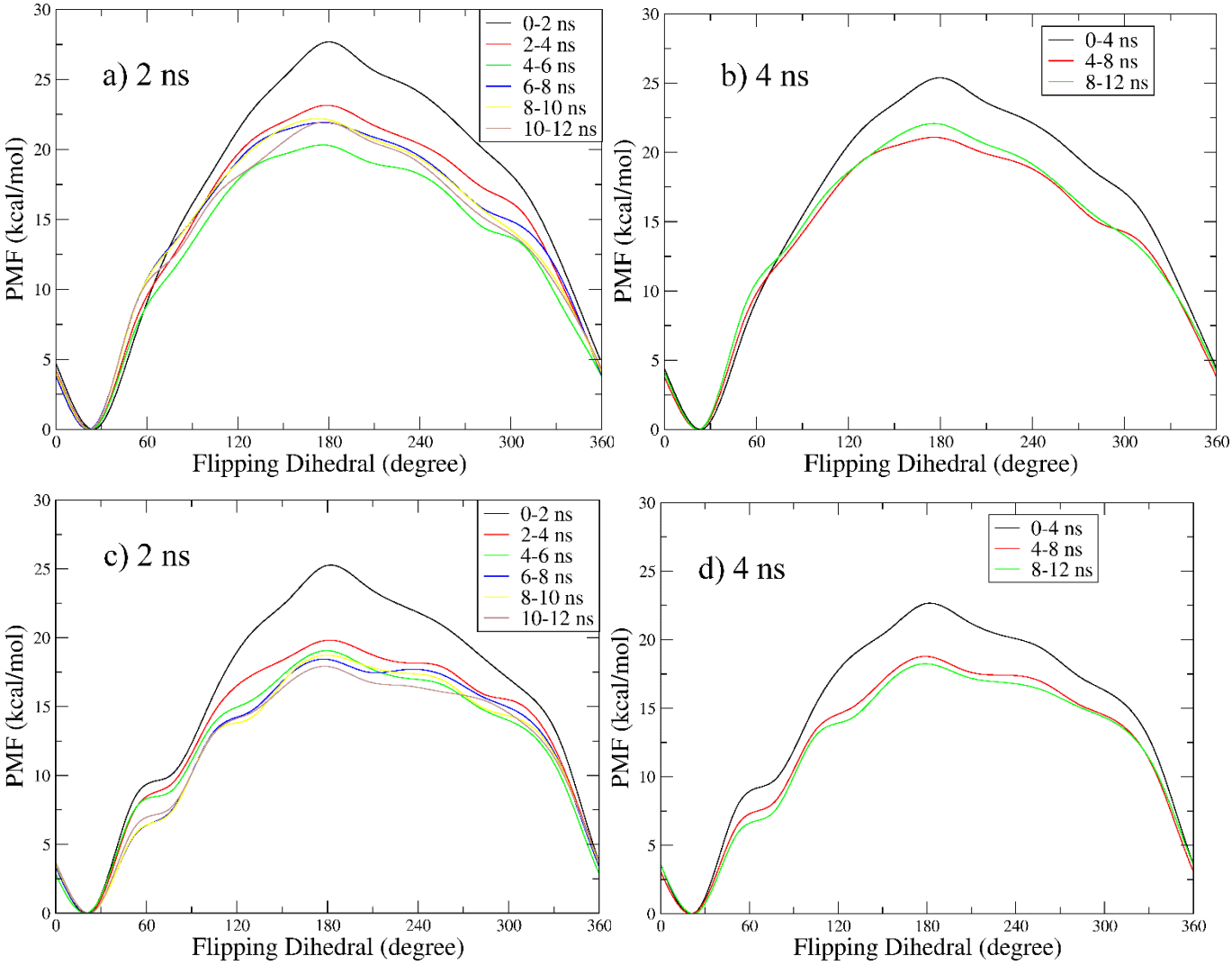
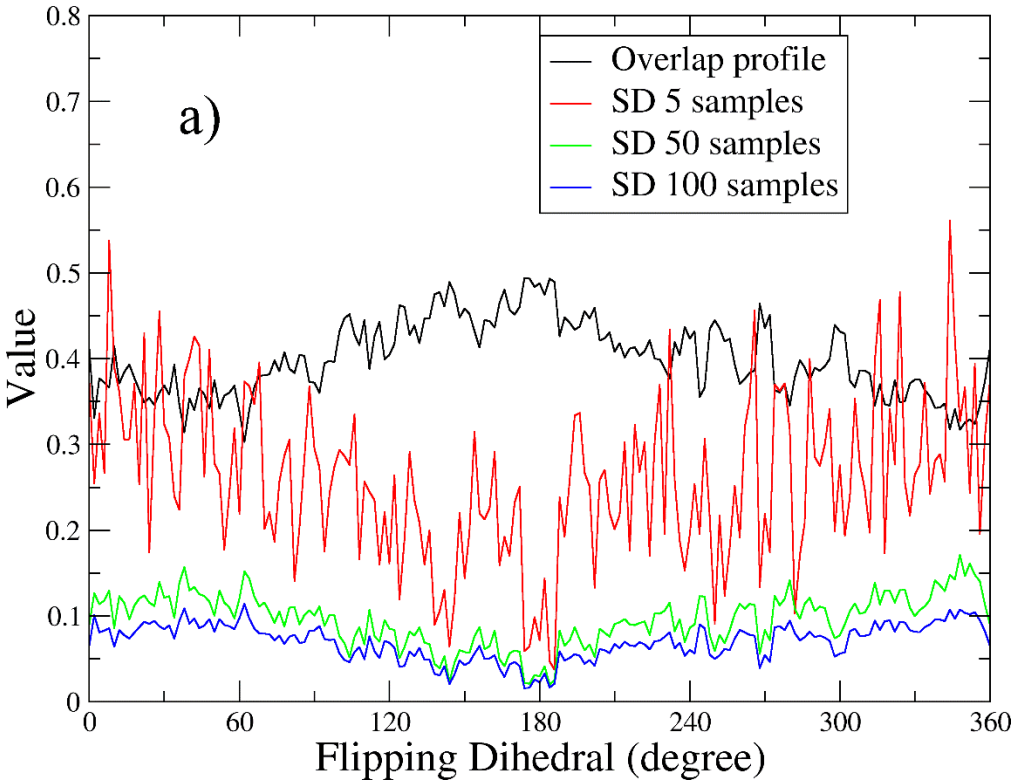
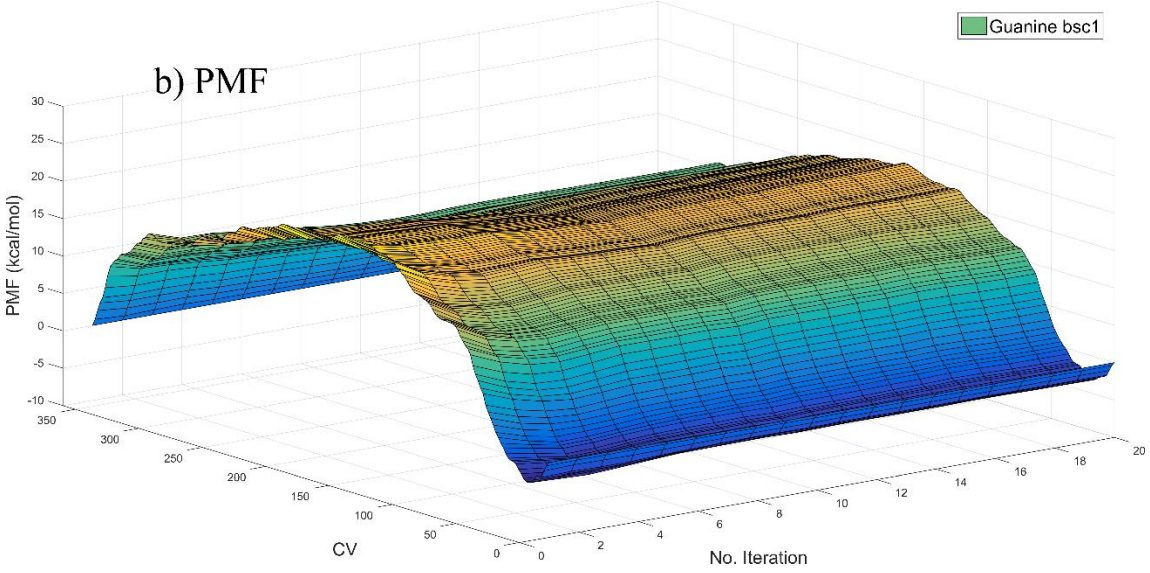


Fig. S3. In the wild-type case described with the bsc1 force field, a) the comparison between dimensionless SD profiles and overlap profile in nonequilibrium stratification, b) the convergence behavior of free energy profiles from nonequilibrium stratification on the sample size, and c) the time-evolution of state-specified SD. The standard deviation in the *i*th state is the sum of the components contributed by samples initiated from that state, namely $\sigma_i = \sqrt{\frac{\mathrm{Var}\left(f_{i,i+1}\right)}{n_i^2 f_{i,i+1}^2} + \frac{\mathrm{Var}\left(f_{i-1,i}\right)}{n_i^2 f_{i-1,i}^2}}$. In b), the initial sample size is 5 and in each iteration further 5 samples are added to the dataset. There are small differences between the 5-sample PMF and the later ones. Since the 10th iteration, the fluctuation of the PMF is small. Thus, we define this sample size (50 samples) as the minimum sample size required for convergence.





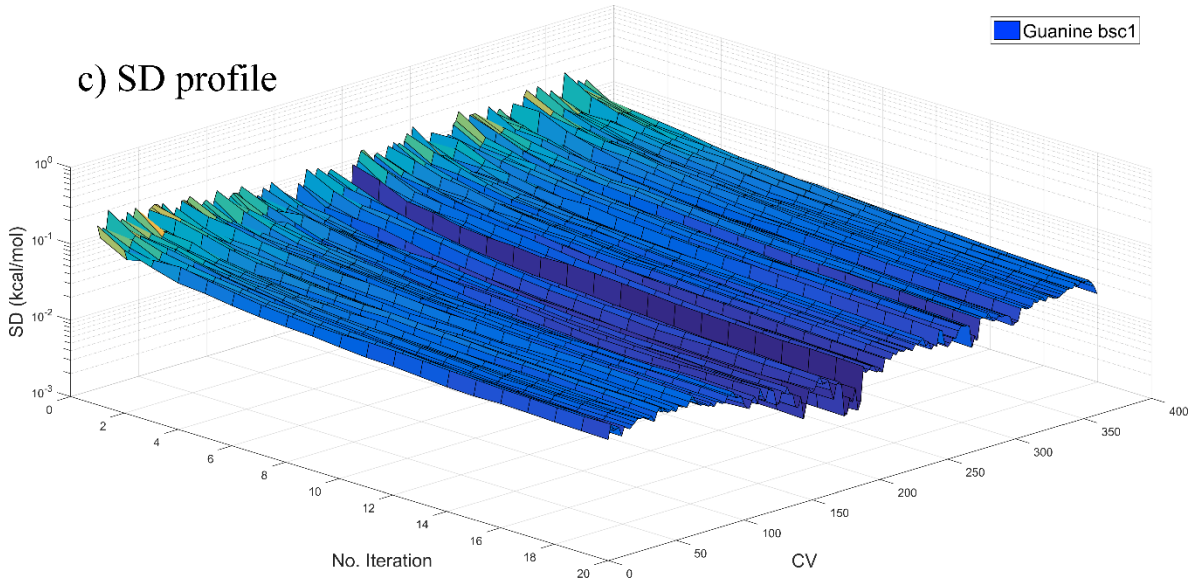
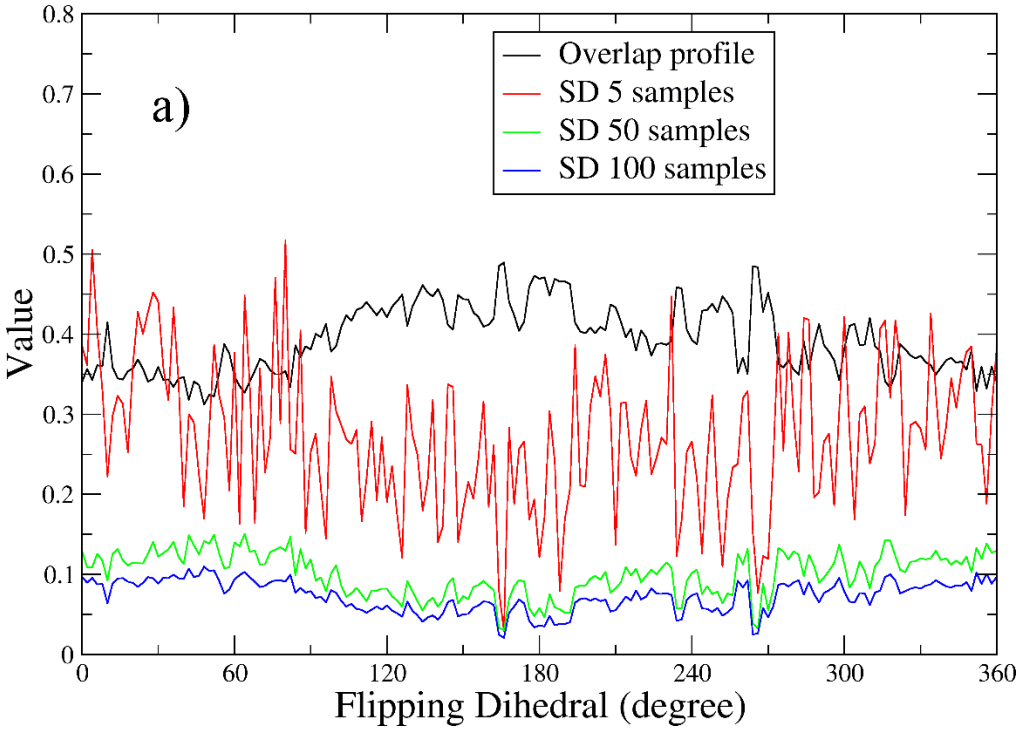
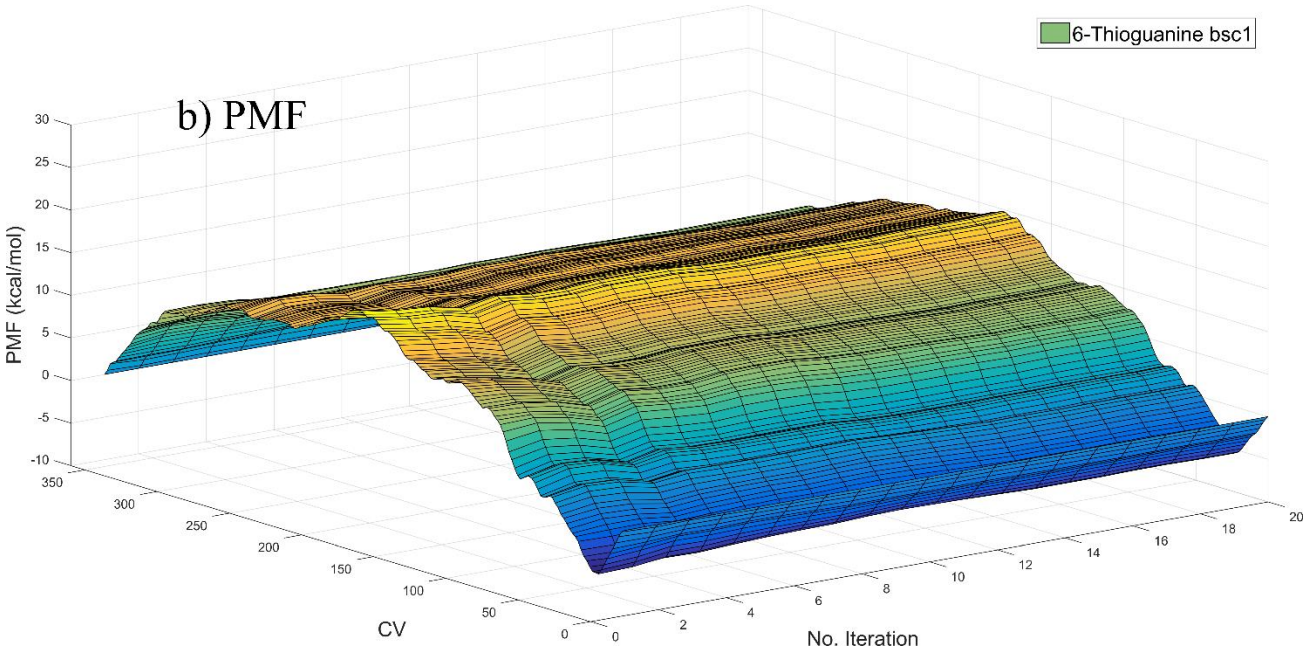

c) SD profile
Guanine bsc1
SD (kcal/mol)
10^{0}
10^{-1}
10^{-2}
10^{-3}
0
2
4
6
8
10
12
14
16
18
20
No. Iteration
0
50
100
150
200
250
300
350
400
CV

Fig. S4. In the sulfur-substitution case described with the bsc1 force field, a) the comparison between SD profiles and overlap profile in nonequilibrium stratification, b) the convergence behavior of free energy profiles from nonequilibrium stratification on the sample size, and c) the time-evolution of state-specified SD. The standard deviation in the *i*th state is the sum of the components contributed by samples initiated from that state, namely $\sigma_i = \sqrt{\frac{\mathrm{Var}\left(f_{i,i+1}\right)}{n_i^2 f_{i,i+1}^2} + \frac{\mathrm{Var}\left(f_{i-1,i}\right)}{n_i^2 f_{i-1,i}^2}}$. In b), the initial sample size is 5 and in each iteration further 5 samples are added to the dataset. There are small differences between the 5-sample PMF and the later ones. Since the 10th iteration, the fluctuation of the PMF is small. Thus, we define this sample size (50 samples) as the minimum sample size required for convergence.





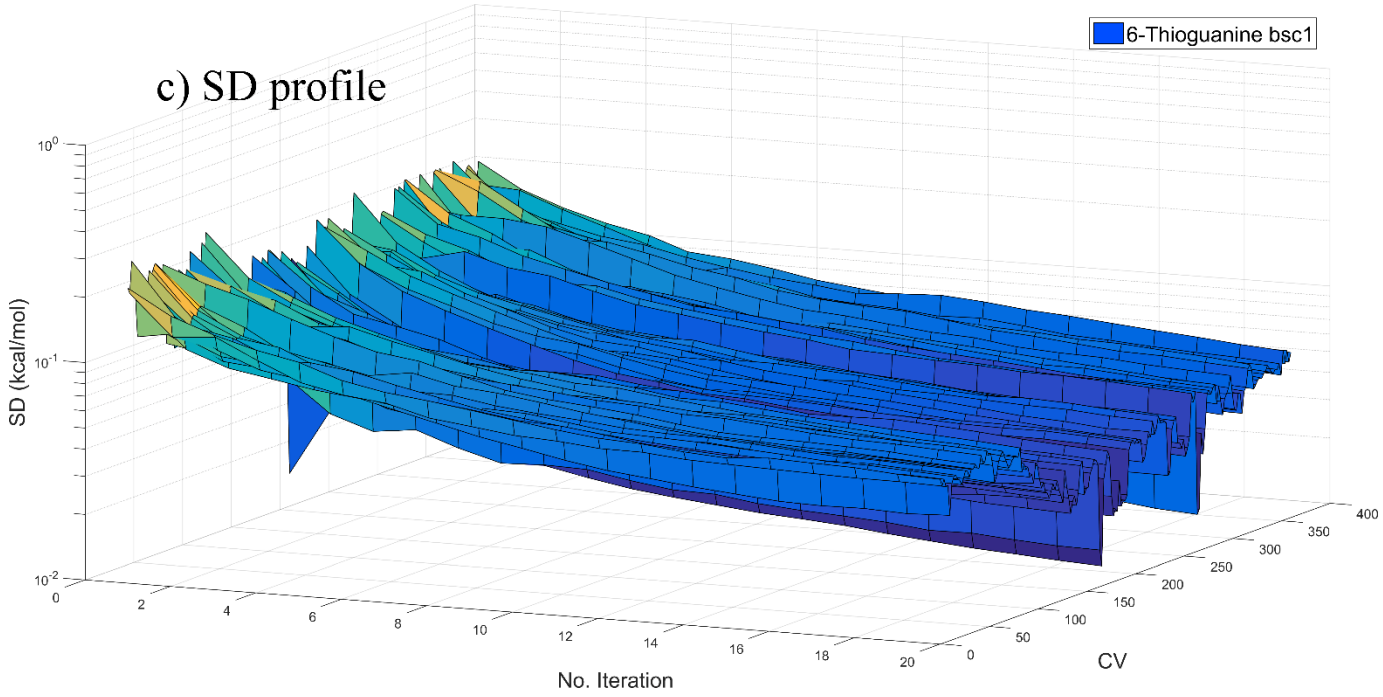

c) SD profile
6-Thioguanine bsc1
SD (kcal/mol)
10^{0}
10^{-1}
10^{-2}
0
2
4
6
8
10
12
14
16
18
20
No. Iteration
0
50
100
150
200
250
300
350
400
CV

Fig. S5. Under the bsc1 force field, comparison between the free energy profiles of a) the wild-type GC base-paired system and b) the sulfur-substituted mutation constructed from nonequilibrium stratification and equilibrium umbrella sampling with vFEP reweighting.

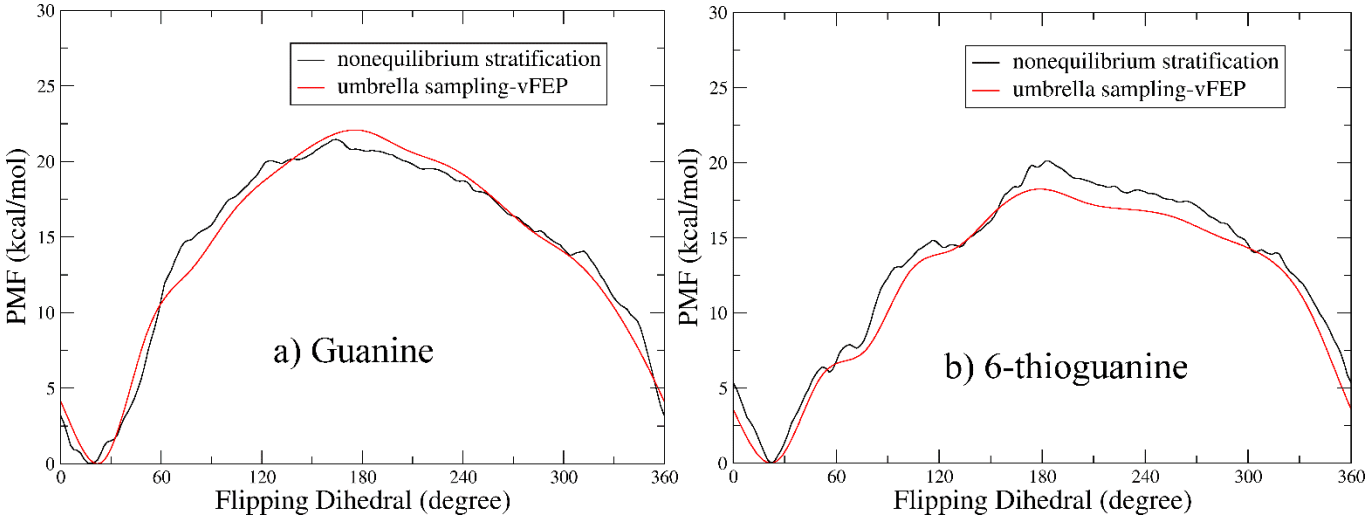


Fig. S6. Under the force field of bsc0, the convergence behavior of free energy profiles constructed from equilibrium umbrella sampling simulations with vFEP reweighting in a) the wild-type GC base-paired case and b) the thioguanine case. The length of time blocks is 4 ns.

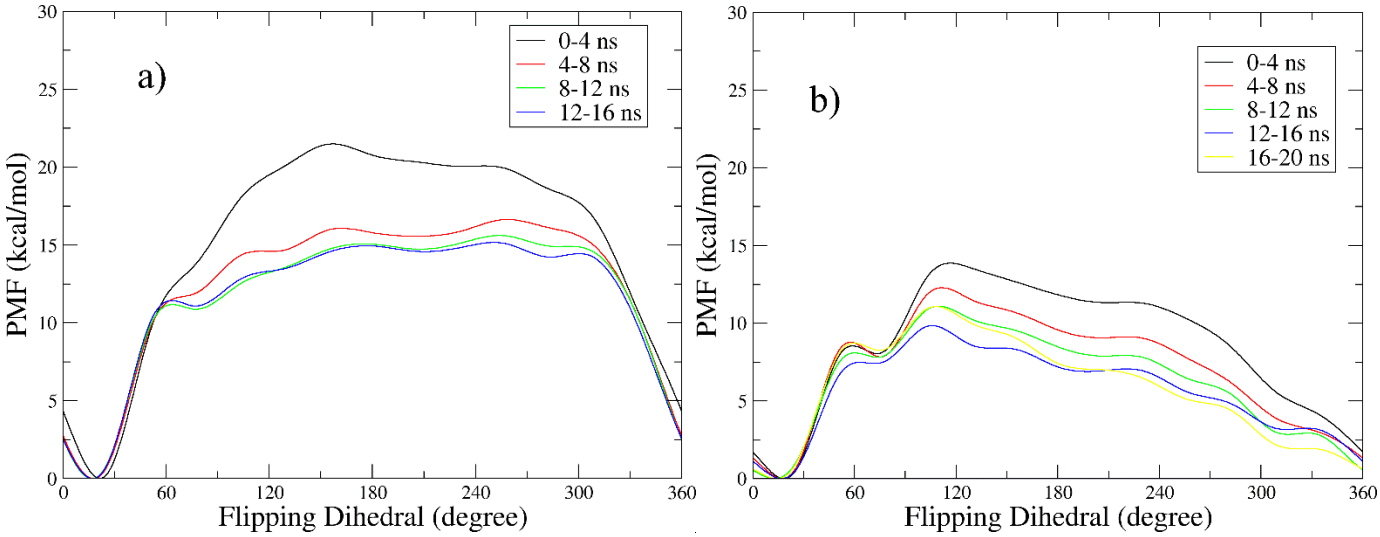


Fig. S7. The unbiased probability distribution of the flipping dihedral a) in the first sulfur-substitution induced base flipping case and b) in the S6G-C and S6G-T case.

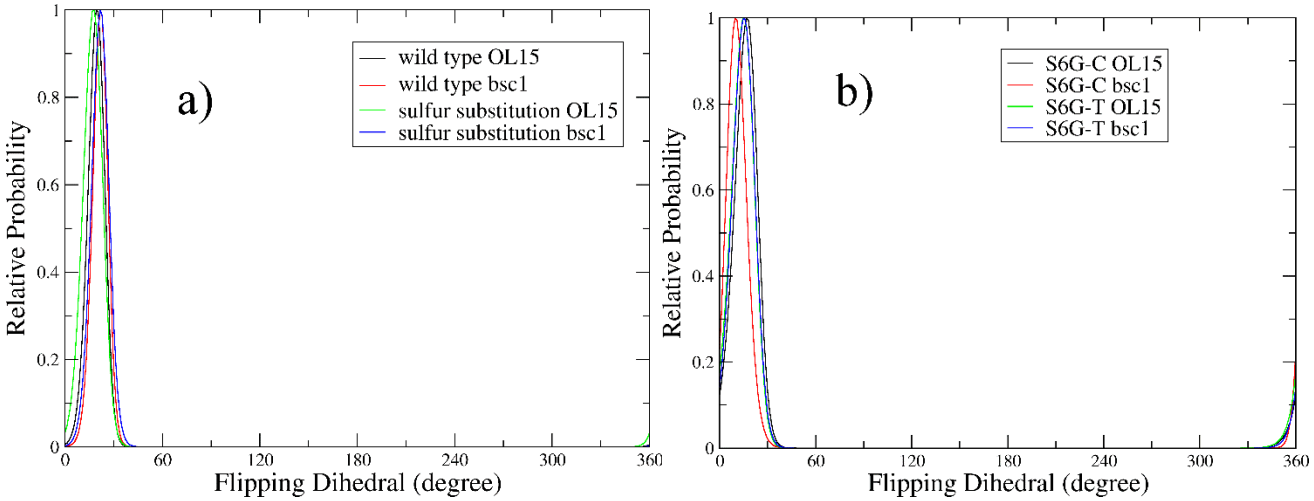
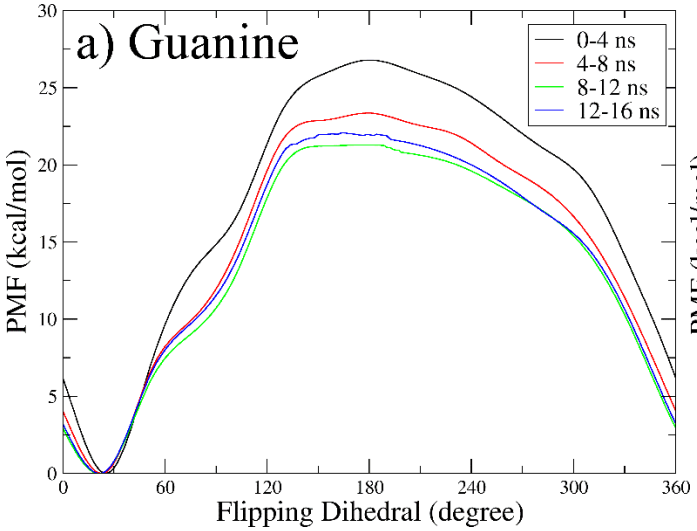
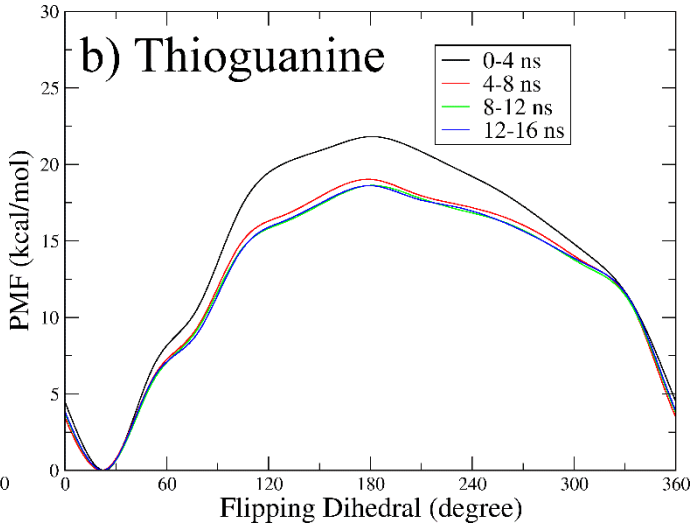


Fig. S8. Under the force field of bsc1 and the ion concentration of 0.1 M, the convergence behavior of free energy profiles constructed from equilibrium umbrella sampling simulations with vFEP reweighting in a) the wild-type GC base-paired case and b) the thioguanine case. The length of time blocks is 4 ns. c) Comparison between free energy profiles under the ion concentration of 0.1 M and those with the zero ion concentration.





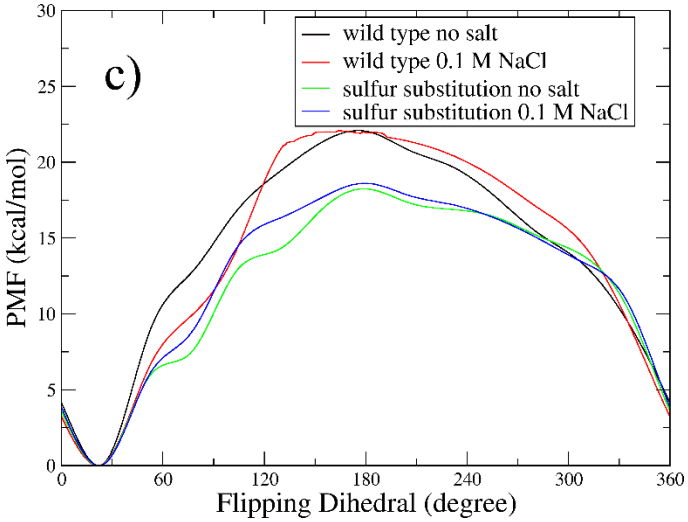


Fig. S9. The number of hydrogen bonds formed between G7-C20 and that in the mutation. The hydrogen bonds are calculated under different cutoff for defining the hydrogen bond. The cutoff is a) 3.1 Å, b) 3.2 Å, c) 3.3 Å and d) 3.4 Å.

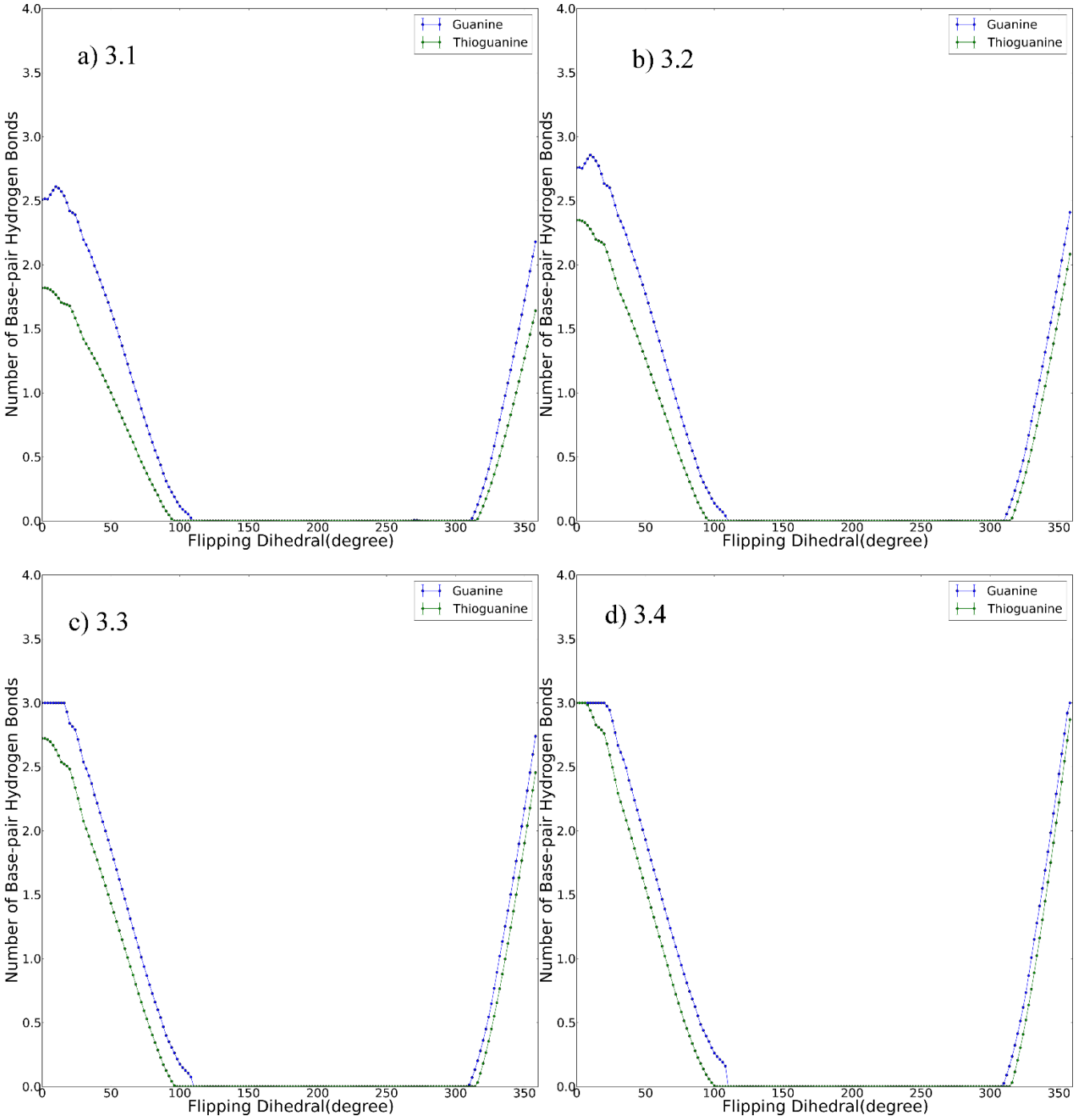


Fig. S10. Under the force field of bsc1, the convergence behavior of free energy profiles constructed from equilibrium umbrella sampling simulations with vFEP reweighting in a) the S6G-C base-paired case and b) the S6G-T mismatched case. The length of time blocks is 4 ns.

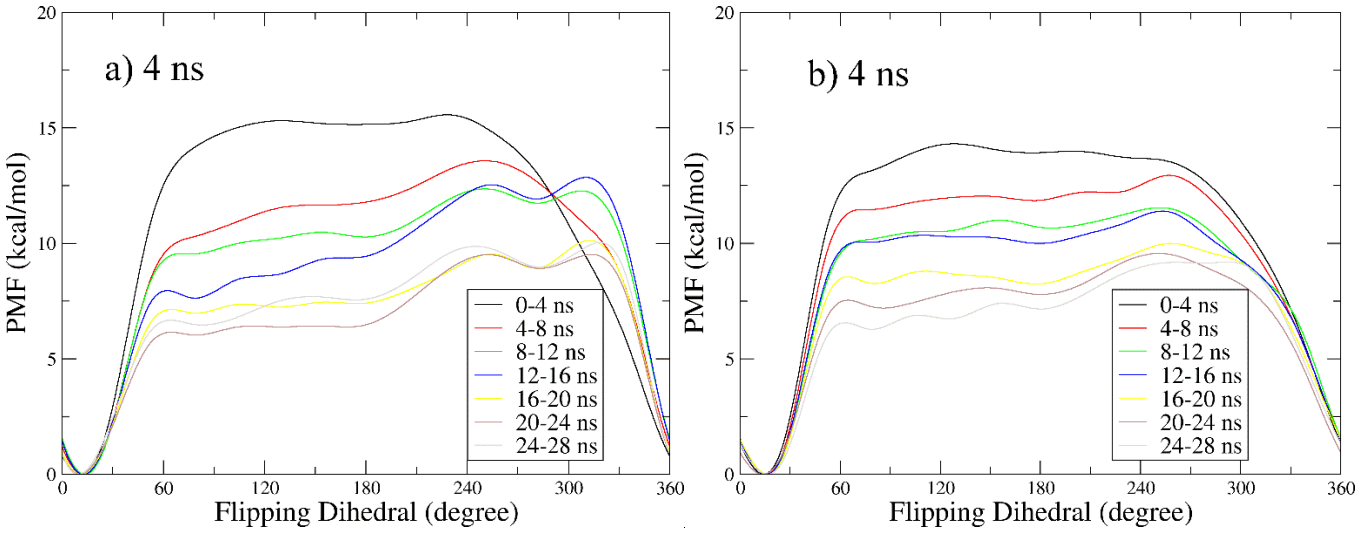


Fig. S11. Under the force field of OL15 and the ion concentration of 0.1 M, the convergence behavior of free energy profiles constructed from equilibrium umbrella sampling simulations with vFEP reweighting in a) the S6G-C base-paired case and b) the S6G-T mismatched case. The length of time blocks is 4 ns. c) Comparison between free energy profiles under the ion concentration of 0.1 M and those with the zero ion concentration.