Supporting Information

A Method for Efficient Calculation of Thermal Stability of Protein upon Point Mutations

Bo Wang[1], Yifei Qi[1,3],Ya Gao[1,2*], John Z. H. Zhang[1,3,4,5*]

[1] *Shanghai Engineering Research Center of Molecular Therapeutics & New Drug Development, Shanghai Key Laboratory of Green Chemistry & Chemical Process, School of Chemistry and Molecular Engineering, East China Normal University, Shanghai 200062, China*

[2] *School of Mathematics, Physics and Statistics, Shanghai University of Engineering Science, Shanghai 201620, China*

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

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

[5]*Collaborative Innovation Center of Extreme Optics, Shanxi University, Taiyuan, Shanxi 030006, China*

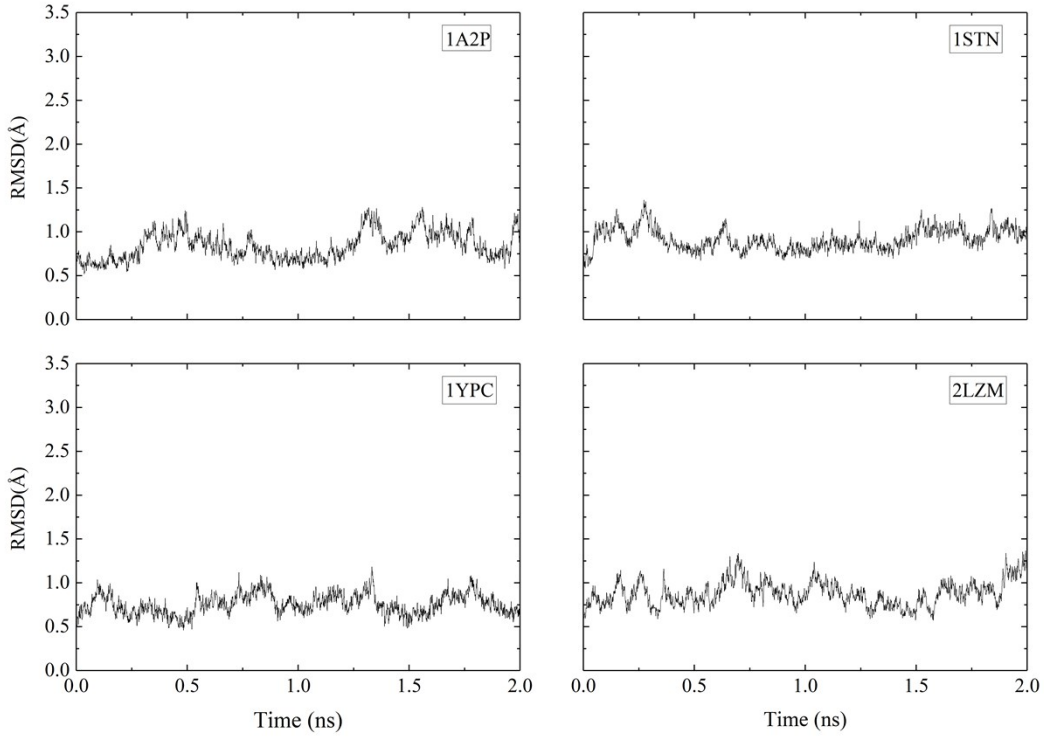

Figure S1. Time evolution of the backbone RMSD from the native structure for four protein systems in data set S712.

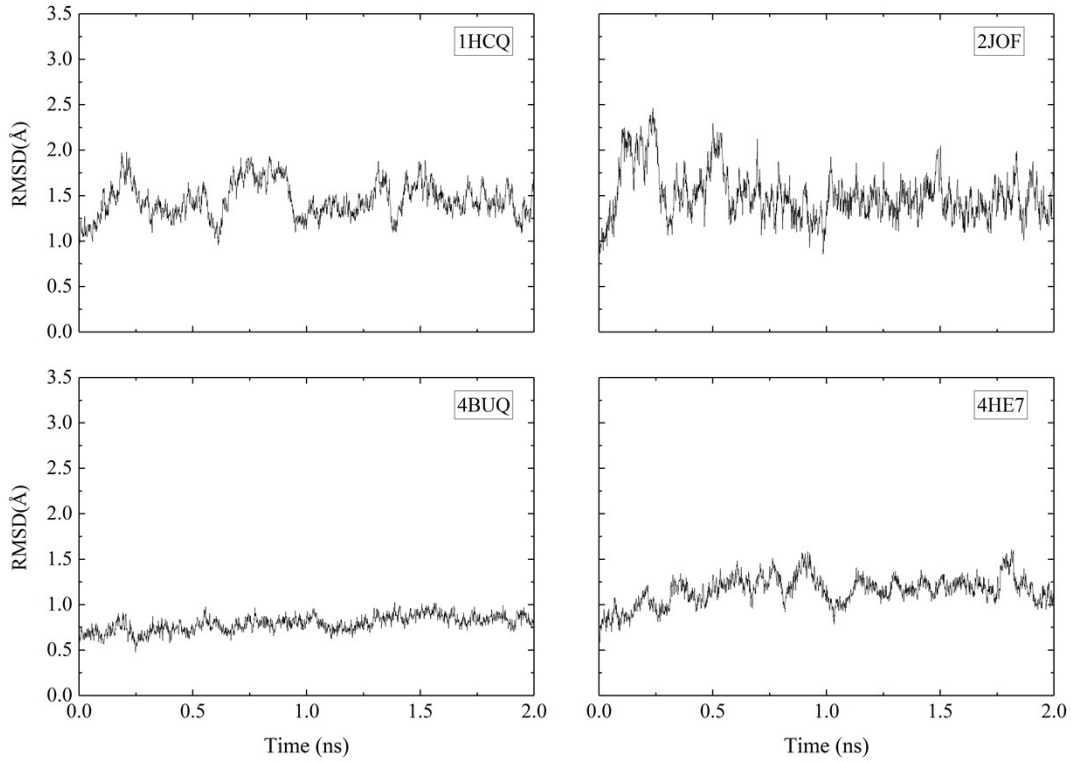

Figure S2. Same as Fig. S1 but for data set S141.

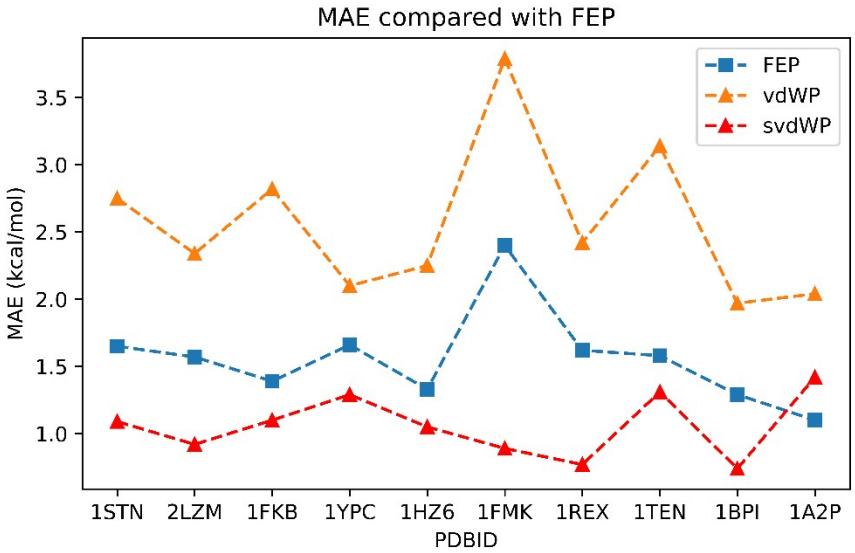

Figure S3. The MAE (mean absolute error) of our calculation result compared with that from FEP (free energy perturbation) calculation for every protein system in data set S712. The svdWP represents the calculated results after scaling.

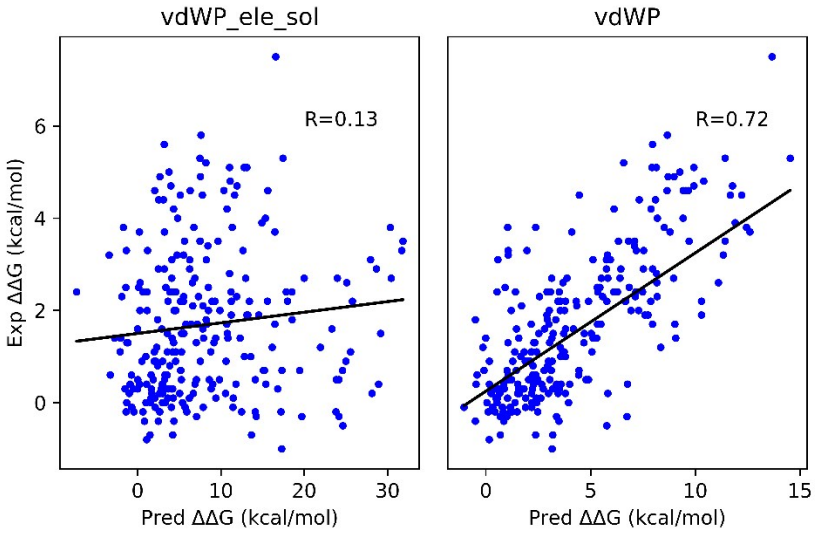

Figure S4. Comparison of the calculated results considering electrostatic interaction and solvation free energy (named vdWP_ele_sol) in the Eq.13 with the vdWP. Taking the protein system 1STN as an example because this system has the most mutation data (255 mutations). The solvation free energies are calculated using MM/GBSA[1] (igb=2) from 100 snapshots. The dielectric constants are set at 1, 3, and 5 for hydrophobic, polar, and charged residues respectively to obtain a minimal mean absolute errors which has been proven in our previous work [2].

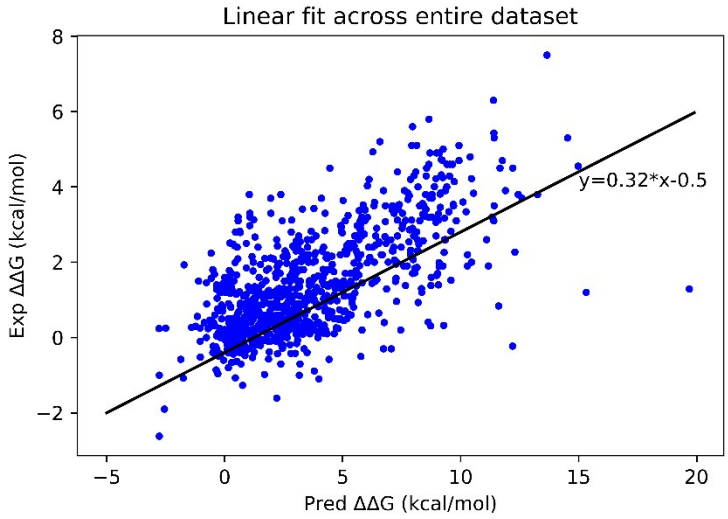

Figure S5. The linear fitting of all the data in this work. The fitting slope and intercept are used for svdWP calculation.

Tabel S1. The influence of the number of mutation snapshots used on the calculation results. Taking the protein system 1STN as an example.

Snapshots	100	200	300	500	1000
R	0.66	0.72	0.71	0.72	0.72

References:

1. A. Onufriev, D. Bashford, D. A. J. P. S. Case, Function, and Bioinformatics, 2004, **55**, 383-394.
2. Y. Yan, M. Yang, C. G. Ji and J. Z. H. Zhang, *J. Chem Inf. Model.*, 2017, **57**, 1112-1122.