Supporting Information for

Serine protease acylation proceeds with a subtle re-orientation of the histidine ring at the tetrahedral intermediate

Yanzi Zhou and Yingkai Zhang*

Department of Chemistry, New York University, New York, NY10003, U.S.A.
Tel: 001-212-998-7882; E-mail: yingkai.zhang@nyu.edu

Computational Details

1. Preparation of initial enzyme reactant systems

The initial structure was obtained from the x-ray crystal structure of trypsin-inhibitor complex (Protein Data Bank identification code: 1MCT).[1] First, the inhibitor was removed. Then the substrate was taken from the inhibitor sequence at the active site (Cys3-Pro4-Arg5-Ile6-Trp7-Met8) and placed at the X-ray coordinates of inhibitor. The scissile bond is between Arg5 and Ile6. Hydrogen atoms were added with all titratable residues in their normal protonation states. Next, the system was neutralized, solved, and equilibrated with a series of minimizations interspersed by short molecular dynamics simulations using Amber10[2] with periodic boundary condition. Then an extensive molecular dynamics simulation of 5ns was carried out and the trajectory is stable. A snapshot at 2 ns was randomly chosen for the subsequent QM/MM simulations.

In all above MD simulations, long-range electrostatic interactions were treated with particle mash Ewald (PME) method[3, 4] and 12 Å cutoff was used for both PME and van deer Waals (vdW) interactions. The pressure was maintained at 1 atm and coupled with isotropic position scaling. Temperature was controlled at 310 K with Berendsen thermostat method.[5] They were performed with Amber10 molecular dynamic package,[2] and amber99SB[6, 7] force fields was used.

2. QM/MM simulations.

With an equilibrated MD snapshot, the QM/MM model was prepared by deleting the ions and waters beyond 27 Å from the reaction center, which was chosen as hydroxyl oxygen atom of Ser195. The resulted system contained about 9,000 atoms. The QM sub-system includes the side chains of catalytic triad (His57, Asp102, and Ser195) and the scissile peptide portion of the substrate as shown in Figure S1. There are 38 QM atoms, and 7 pseudo atoms in total. The QM/MM boundary was described by improved pseudobond approach[8-11]. All other atoms were described by the same molecular mechanical force field used in classical MD simulations. For all QM/MM calculations, the spherical boundary condition was applied, and only the atoms within 22 Å from the reaction center were allowed to move. The 18 and 12 Å cutoffs were employed for electrostatic and van der Waals interactions, respectively. There was no cutoff for electrostatic interactions between QM and MM regions. The prepared QM/MM system was first minimized and then employed to map out a reaction path with B3LYP/6-31+G* QM/MM minimizations. Figure S2 illustrates the reaction coordinate employed for each reaction step. For the initial step of acylation reaction, it is chosen as $d_{\text{OG-HG}} - d_{\text{C-OG}}$, which is the distance between O-H bond of Ser195 and the distance between attacking O of Ser195 and the carbon atom of scissile bond. The reaction coordinate of the second step is $d_{\text{OG-HG}} - d_{\text{N-HG}}$, the distances of two H-bonds formed with the proton. The reaction coordinate of the third step is $d_{\text{C-N}} - d_{\text{N-HG}} + d_{\text{NE2-HG}}$, in which $d_{\text{C-N}}$ means the distance between C and N of the scissile bond, $d_{\text{N-HG}}$ is the distance between N atom of scissile and H, and $d_{\text{NE2-HG}}$ is distance between H and the N atom of His57. For each determined structure along the path, a 500 ps MD simulation with MM force field was performed to equilibrate the MM subsystem, with the QM subsystem being frozen. Finally, the resulting snapshot was used as the starting structure for Born-Oppenheimer B3LYP/6-31+G* QM/MM MD simulation with umbrella sampling. Along the reaction path, 12, 10, 19 umbrella windows [12-14] were chosen for the three steps respectively. Time step of 1 fs was employed, and the Berendsen thermostat method [5] has been used to control the system temperature at 310 K.

Each window was simulated for at least 30 ps. First 10 ps MD simulations were discarded for equilibration, and the next 20ps were used for data analysis. From these biased simulations, the free energy profile for each reaction was obtained with the weighted histogram analysis method (WHAM).[15-17] All ab initio QM/MM calculations were performed with modified Q-Chem[18] and Tinker[19] programs. Born−Oppenheimer MD simulations with ab initio QM/MM potential[8, 20-28] and the umbrella sampling method[12-14] have been successfully applied to study several enzymes reactions in our group.[29-37]

Figure S1. Illustration of the division of QM/MM system for acylation reaction of trypsin. All atoms colored in blue are QM atoms and assigned with 6-31+G* basis set, and the pseudo-atoms (colored in red) are treated with pseudo-bond parameters. All the left atoms are MM atoms that are treated with Amber99SB force field. The atoms connected with pseudo-atoms (colored in green) have no electrostatic interaction with QM atoms.

Figure S2. Illustration of Reaction mechanism of acylation reaction for serine protease and reaction coordinate chosen at each step: a) initial step of acylation, b) second step of acylatoin, c) third step of acylation.

Figure S3. Illustration of atom names at active site.

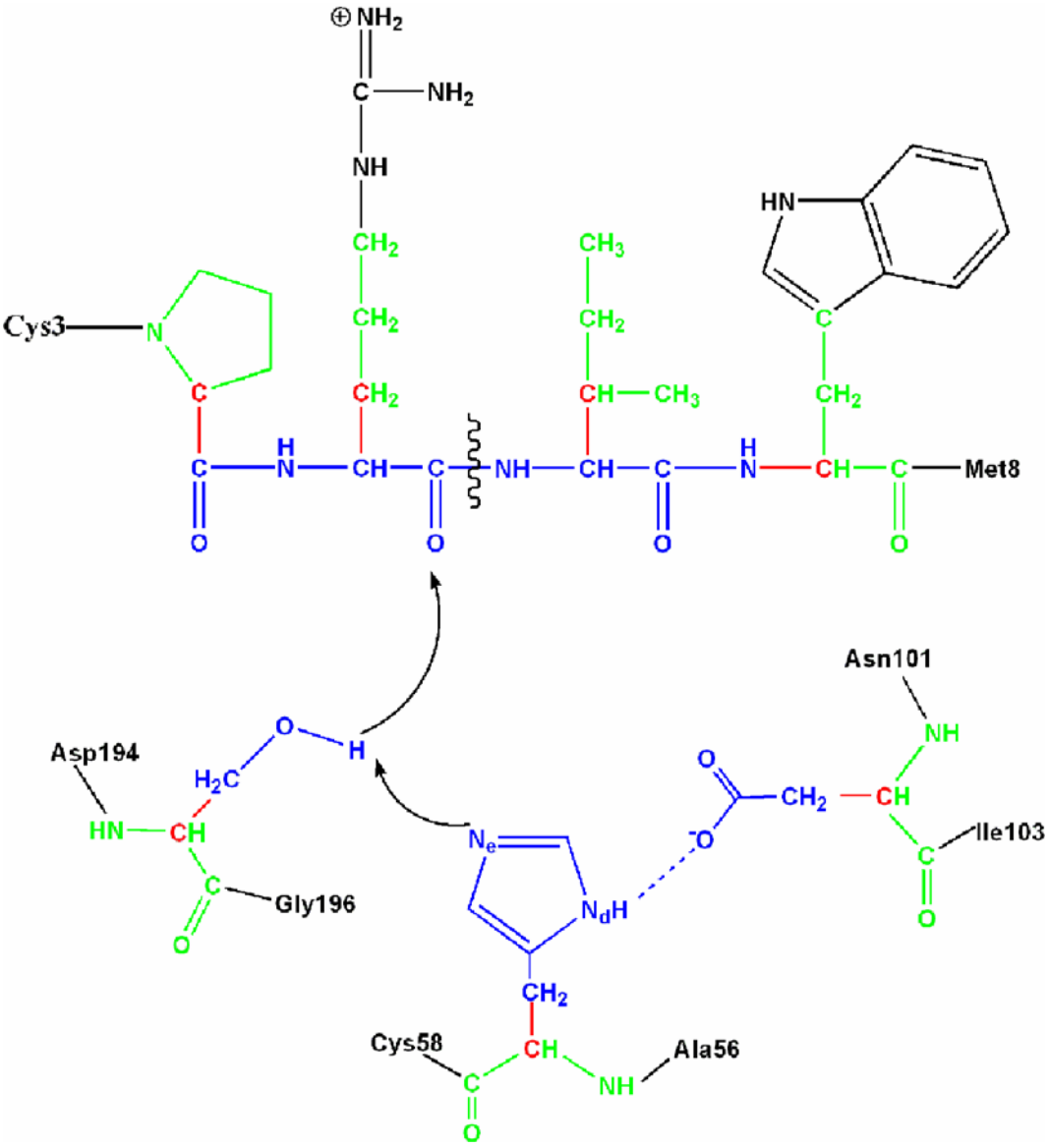

Figure S1.

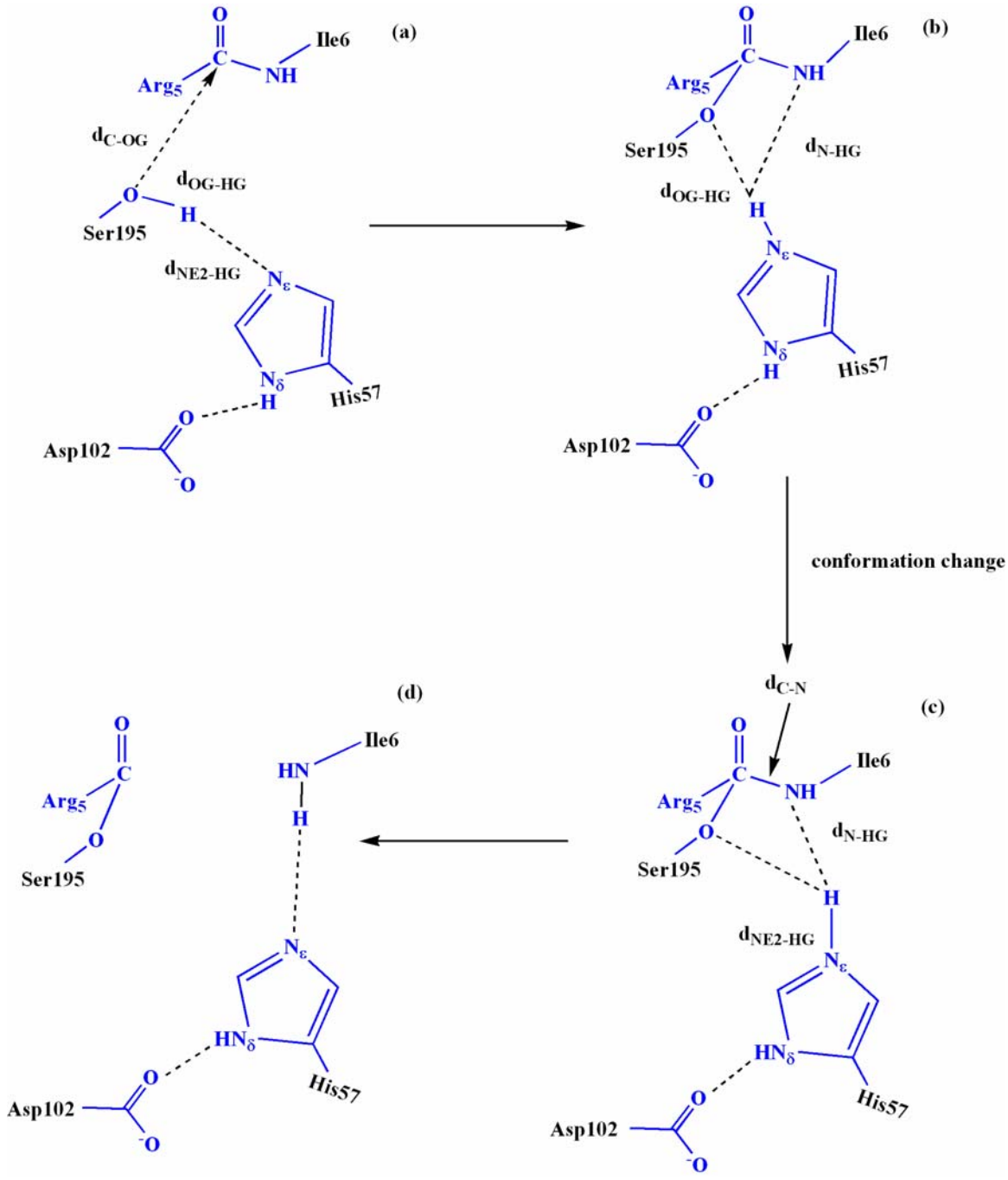
(a)
(b)
(c)
(d)
Ile6
Arg5
Ser195
His57
Asp102
d_{C-OG}
d_{OG-HG}
d_{NE2-HG}
d_{N-HG}
d_{C-N}
conformation change

Figure S2.

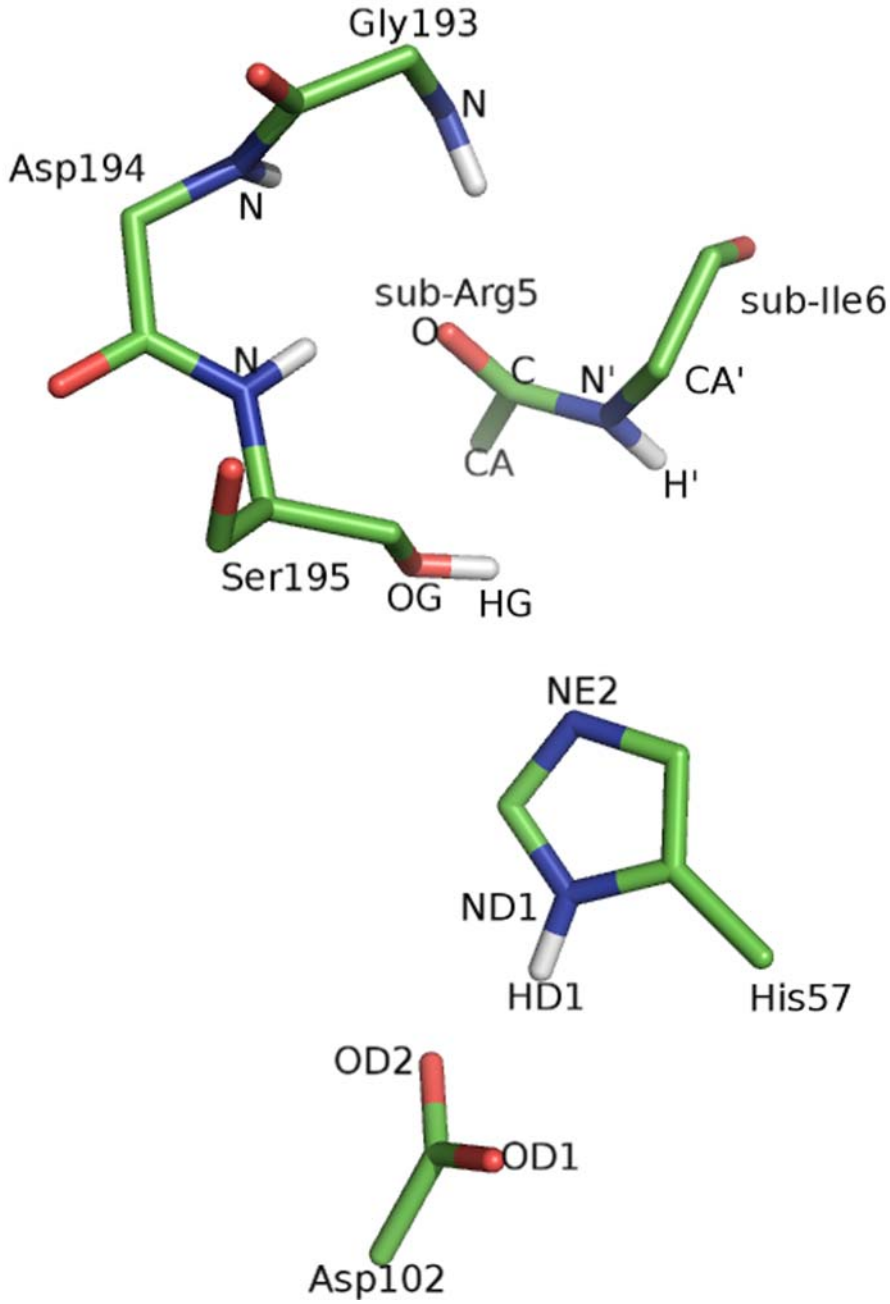
Gly193
N
Asp194
N
sub-Arg5
O
C
N'
CA'
sub-Ile6
CA
H'
N
Ser195
OG
HG
NE2
ND1
HD1
His57
OD2
OD1
Asp102

Figure S3.

Table S1. List of key geometric parameters for the reactant, transition states, intermediates and acyl-enzyme for acylation reaction based on B3LYP/6-31+G* QM/MM MD simulations. Atom names are the same as labeled on Figure S3. Torsion angle ω is the angle for the rotation of NH group around the scissile bond; τ is the improper angle to indicate the inversion of N atom.

Distance (Å)	ES	TS1	TI1	TS2	TI2	TS3	EA1
OG (S195) – C (sub)	2.64 ± 0.04	1.79 ± 0.09	1.56 ± 0.05	1.52 ± 0.05	1.51 ± 0.05	1.46 ± 0.05	1.34 ± 0.03
HG (S195) – OG(S195)	0.99 ± 0.03	1.39 ± 0.09	1.66 ± 0.06	2.05 ± 0.09	2.18 ± 0.16	2.38 ± 0.14	2.90 ± 0.20
HG (S195) – NE2(H57)	1.81 ± 0.13	1.15 ± 0.06	1.06 ± 0.03	1.04 ± 0.03	1.06 ± 0.03	1.14 ± 0.05	2.20 ± 0.13
HG (S195) – N′(sub)	3.37 ± 0.22	2.90 ± 0.19	2.75 ± 0.22	2.10 ± 0.09	1.91 ± 0.06	1.60 ± 0.11	1.02 ± 0.03
C (sub) – N′(sub)	1.37 ± 0.02	1.47 ± 0.04	1.52 ± 0.05	1.58 ± 0.06	1.60 ± 0.06	1.81 ± 0.14	2.97 ± 0.13
ND1 (H57) – OD1(D102)	3.00 ± 0.22	3.23 ± 0.28	3.29 ± 0.28	3.43 ± 0.20	3.46 ± 0.16	3.54 ± 0.18	3.42 ± 0.23
ND1 (H57) – OD2(D102)	3.17 ± 0.30	2.81 ± 0.21	2.77 ± 0.18	2.74 ± 0.12	2.72 ± 0.11	2.79 ± 0.14	2.88 ± 0.15
HD1 (H57) – OD1(D102)	2.02 ± 0.27	2.21 ± 0.37	2.36 ± 0.37	2.59 ± 0.22	2.63 ± 0.17	2.71 ± 0.19	2.57 ± 0.27
HD1 (H57) – OD2(D102)	2.40 ± 0.36	1.89 ± 0.34	1.84 ± 0.32	1.71 ± 0.17	1.68 ± 0.14	1.76 ± 0.17	1.90 ± 0.21
HD1 (H57) – ND1(H57)	1.04 ± 0.03	1.06 ± 0.04	1.07 ± 0.04	1.07 ± 0.04	1.07 ± 0.04	1.06 ± 0.04	1.04 ± 0.03
N (S195) – O(sub)	2.89 ± 0.10	2.86 ± 0.10	2.82 ± 0.10	2.84 ± 0.09	2.87 ± 0.10	2.87 ± 0.10	2.84 ± 0.12
N (G193) – O(sub)	2.79 ± 0.08	2.74 ± 0.08	2.73 ± 0.07	2.74 ± 0.08	2.75 ± 0.08	2.76 ± 0.08	2.80 ± 0.10
N (D194) – O(sub)	3.14 ± 0.15	3.31 ± 0.16	3.25 ± 0.15	3.30 ± 0.14	3.35 ± 0.14	3.36 ± 0.15	3.24 ± 0.18
∠NE2 (H57) – HG (S195) – N′ (sub) degree	119.8 ± 7.8	130.3 ± 6.0	149.6 ± 7.0	159.0 ± 6.1	164.7 ± 7.1	170.8 ± 5.2	166.5 ± 7.6
Torsion ω /degree	173.5 ± 8.3	162.9 ± 7.4	160.5 ± 8.3	147.8 ± 5.9	145.2 ± 6.1	140.0 ± 5.5	143.8 ± 7.8
Torsion τ /degree	-11.2 ± 7.9	-25.7 ± 8.2	-29.9 ± 5.8	-33.5 ± 4.2	-35.0 ± 3.8	-37.4 ± 3.2	-27.1 ± 4.4

1. Q. C. Huang, S. P. Liu and Y. Q. Tang, *J. Mol. Biol.*, 1993, **229**, 1022.
2. D. A. Case, T. A. Darden, T. E. Cheatham, C. L. S. III, J. Wang, R. E. Duke, R. Luo, M. Crowley, R. C. Walker, W. Zhang, K. M. Merz, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K.F.Wong, F. Paesani, J. Vanicek, X.Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D. H. Mathews, M. G. Seetin, C. Sagui, V. Babin and P. A. Kollman, (2008), University of California, San Francisco.
3. T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089.
4. U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**.
5. H. J. C. Berendsen, J. P. M. Postma, W. F. Vangunsteren, A. Dinola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684.
6. W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *JACS*, 1995, **117**, 5179.
7. V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling, *Proteins-Structure Function and Bioinformatics*, 2006, **65**, 712.
8. Y. K. Zhang, *Theor. Chem. Acc.*, 2006, **116**, 43.
9. Y. K. Zhang, T. S. Lee and W. T. Yang, *J. Chem. Phys.*, 1999, **110**, 46.
10. Y. K. Zhang, H. Y. Liu and W. T. Yang, *J. Chem. Phys.*, 2000, **112**, 3483.
11. Y. K. Zhang, *J. Chem. Phys.*, 2005, **122**, 024114.
12. G. N. Patey and J. P. Valleau, *J. Chem. Phys.*, 1975, **63**, 2334.
13. B. Roux, *Comput. Phys. Commun.*, 1995, **91**, 275.
14. E. M. Boczko and C. L. Brooks, *J. Phys. Chem.*, 1993, **97**, 4509.
15. S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman and J. M. Rosenberg, *J. Comput. Chem.*, 1992, **13**, 1011.
16. M. Souaille and B. Roux, *Comput. Phys. Commun.*, 2001, **135**, 40.
17. A. M. Ferrenberg and R. H. Swendsen, *Phys. Rev. Lett.*, 1988, **61**, 2635.
18. Y. Shao, L. F. Molnar, Y. Jung, J. Kussmann, C. Ochsenfeld, S. T. Brown, A. T. B. Gilbert, L. V. Slipchenko, S. V. Levchenko, D. P. O'Neill, R. A. DiStasio, R. C. Lochan, T. Wang, G. J. O. Beran, N. A. Besley, J. M. Herbert, C. Y. Lin, T. Van Voorhis, S. H. Chien, A. Sodt, R. P. Steele, V. A. Rassolov, P. E. Maslen, P. P. Korambath, R. D. Adamson, B. Austin, J. Baker, E. F. C. Byrd, H. Dachsel, R. J. Doerksen, A. Dreuw, B. D. Dunietz, A. D. Dutoi, T. R. Furlani, S. R. Gwaltney, A. Heyden, S. Hirata, C. P. Hsu, G. Kedziora, R. Z. Khalliulin, P. Klunzinger, A. M. Lee, M. S. Lee, W. Liang, I. Lotan, N. Nair, B. Peters, E. I. Proynov, P. A. Pieniazek, Y. M. Rhee, J. Ritchie, E. Rosta, C. D. Sherrill, A. C. Simmonett, J. E. Subotnik, H. L. Woodcock, W. Zhang, A. T. Bell, A. K. Chakraborty, D. M. Chipman, F. J. Keil, A. Warshel, W. J. Hehre, H. F. Schaefer, J. Kong, A. I. Krylov, P. M. W. Gill and M. Head-Gordon, *PCCP*, 2006, **8**, 3172.
19. J. W. Ponder, (2004) software Tools for Molecular Design, Version 4.2, June.
20. A. Warshel and M. Levitt, *J. Mol. Biol.*, 1976, **103**, 227.
21. U. C. Singh and P. A. Kollman, *J. Comput. Chem.*, 1986, **7**, 718.
22. M. J. Field, P. A. Bash and M. Karplus, *J. Comput. Chem.*, 1990, **11**, 700.

23. A. J. Mulholland, *Drug Discovery Today*, 2005, **10**, 1393.
24. H. Hu and W. T. Yang, *Annu. Rev. Phys. Chem.*, 2008, **59**, 573.
25. R. A. Friesner and V. Guallar, *Annu. Rev. Phys. Chem.*, 2005, **56**, 389.
26. J. L. Gao and D. G. Truhlar, *Annu. Rev. Phys. Chem.*, 2002, **53**, 467.
27. H. M. Senn and W. Thiel, *Angewandte Chemie-International Edition*, 2009, **48**, 1198.
28. A. Shurki and A. Warshel, in *Protein Simulations*, 2003, vol. 66, pp. 249.
29. S. L. Wang, P. Hu and Y. K. Zhang, *J. Phys. Chem. B*, 2007, **111**, 3758.
30. P. Hu, S. Wang and Y. Zhang, *JACS*, 2008, **130**, 3806.
31. P. Hu, S. L. Wang and Y. K. Zhang, *JACS*, 2008, **130**, 16721.
32. Z. Y. Lu and Y. K. Zhang, *Journal of Chemical Theory and Computation*, 2008, **4**, 1237.
33. Z. H. Ke, S. L. Wang, D. Q. Xie and Y. K. Zhang, *J. Phys. Chem. B*, 2009, **113**, 16705.
34. Z. H. Ke, Y. Z. Zhou, P. Hu, S. L. Wang, D. Q. Xie and Y. K. Zhang, *J. Phys. Chem. B*, 2009, **113**, 12750.
35. H. Zheng, S. L. Wang and Y. K. Zhang, *J. Comput. Chem.*, 2009, **30**, 2706.
36. R. B. Wu, P. Hu, S. L. Wang, Z. X. Cao and Y. K. Zhang, *Journal of Chemical Theory and Computation*, 2010, **6**, 337.
37. Y. Z. Zhou, S. L. Wang and Y. K. Zhang, *J. Phys. Chem. B*, 2010, **114**, 8817.