ml-2010-00026a

Supporting Information

Computational Analysis of Protein Hotspots

*Chao-Yie Yang, and Shaomeng Wang**

Departments of Internal Medicine, Pharmacology and Medicinal Chemistry, University of Michigan, 1500 E. Medical Center Drive, Ann Arbor, MI 48109, USA

* Corresponding author: Email: shaomeng@umich.edu; Phone:+1 734 6150362; Fax:+1 734 6479647.

Methods and Calculations

Force field parameters generation and preparation of molecular dynamic simulation

Three crystal structures of thermolysin (4TLI[1], 1FJW[2] and 1FJO[2]) from the Protein Databank[3] were used for this study. 4TLI is a structure of thermolysin dispersed in 25% v/v isopropanol, 1FJW is thermolysin in 20 mM phenol and 1FJO is thermolysin in 50% acetone solution. Amber (version 8)[4] was used for the thermodynamic integration calculations and PMEMD from Amber (version 10)[5] was used for molecular dynamics simulations. The Amber 99SB force field parameters[6] were used for the amino acids. In thermolysin, a zinc ion is covalently bonded with His142, His146, and Glu166 in a tetrahedral structure. Because the zinc ion in thermolysin has no catalytic role with the organic solvents, we used the force field parameters derived by Ryde[7] and constrained the three residues in the zinc-coordinated center with modest harmonic forces (1 kcal/mol-Å^2). The force field parameters for isopropanol were taken from the literature.[8] The Antechamber module in Amber was used to derive the force field parameters for the phenol and acetone molecules. The protocol for generating the point charge parameters is as follows: The probe molecule was minimized at the RHF level using a 6-31G** basis set with Gaussian98[9]. The electrostatic field potential calculated from Gaussian98 was used to generate the point charges at each atom site based on the RESP fitting procedure.

To prepare the topology and coordinate files, counter ions were added to neutralize the charges in thermolysin with two probe molecules at sites 1 and 2 before it was placed in a 13 Å octahedron box of water. The TIP3P[10] water model was used. After a 3000-step minimization (steps 1-1000 using conjugated gradient followed by 2000 steps steepest decent), a 0.5 ps simulation was performed to raise the temperature of the system to

150K, followed by another 1 ps of simulation to increase the temperature further to 298K. The system was then equilibrated for 8.5 ps at 298K. The system is then ready for the production run. All the MD simulations were in the isothermal isobaric (NTP, T = 298K and P = 1 atm) ensemble. The SHAKE[11] algorithm was used to fix bonds involving hydrogen. The PME method[12] was used and the non-bonded cutoff distance was set at 10 Å. The time step was 2 fs, and neighboring pairs list was updated every 20 steps.

Double-decoupling method

To estimate the preference of the probe molecule at the crystallographic site over the same molecule in the bulk liquid water, we used the double-decoupling method and follow the notation of Gilson *et. al*.[13] based on the following thermochemical transformation equations,

$$\text{probe(g)} \rightarrow \text{null}, \qquad \Delta G_0^0 \qquad \text{(S1)}$$

$$\text{probe(l)} \rightarrow \text{probe(g)}, \qquad \Delta G_1^0 \qquad \text{(S2)}$$

$$\text{probe(l)-P(l)} \rightarrow \text{probe(g)} + \text{P(l)}, \qquad \Delta G_2^0 \qquad \text{(S3)}$$

where probe-P denotes the protein-probe structure, (l) and (g) refer to liquid and ideal gas phases, null denote a probe molecule without charges and van der Waals radii, and ΔG^0 is the free energy change at the standard state condition. The hydration free energy of the probe molecule is defined as $\Delta G_1^0 - \Delta G_0^0$.[14] Calculation of the standard free energy of decoupling a probe from the protein in the bulk liquid follows the procedure described by Hamelberg and coworkers[15] and our previous work.[16] The probe molecule in this study is equivalent to the constrained water molecule studied previously [15-16]. According to ref. [15],

$$\Delta G_2^0 = \int_0^1 \left\langle \frac{\partial U(\lambda, \vec{r}_P, \vec{r}_{probe}, \zeta_{probe}, \vec{r}_S)}{\partial \lambda} \right\rangle_\lambda d\lambda - RT \ln\left(\frac{\sigma_{P-probe}}{\sigma_P \sigma_{probe}}\right) + RT \ln\left(C^0 V_I\right) + RT \ln\left(\frac{\xi_I}{8\pi^2}\right) + P^0 \Delta\Delta V. \quad \text{(S4)}$$

In eq. (4), σ is the symmetry number, C^0 is the standard state concentration, V_I is the volume of an ideal gas probe molecule constrained in the simulation, ξ_I is the integral of the orientational space of an ideal gas probe molecule sampled (solid angle) in the protein and ΔΔV is the difference of volume changes between the probe-protein and the protein alone in bulk liquid. In our calculations, a 1 ns simulation of the probe molecules and thermolysin was performed to obtain the root-mean-square displacement (RMSD) of the probe molecule at each site. The RMSD values were used to convert into a harmonic force constant based on $k = \frac{3RT}{< \delta r^2 >}$. The harmonic potential was imposed on each probe molecule during thermodynamic integration to determine the integral in eq. (4). Because all three probe molecules have a two-fold symmetry, $\sigma_{probe} = 2$ for all calculations. Free rotation of the probe molecule is assumed because no orientational constraint was imposed during the calculation.

For the thermodynamic integration calculation, two perturbation steps were used. In the first step, charges of the probe molecules were slowly turned off. Five Gaussian quadrature points between 0 and 1 for λ were used. At each λ, 20 ps pre-equilibrium simulations were performed followed by a 100 ps data collection. Gaussian quadrature integration was used to calculate the corresponding integral. In the second step, the van der Waals radii were slowly turned off (the end state of all heavy atoms were assigned radii of 1 Å to avoid singularity of $\frac{\partial U}{\partial \lambda}$) with λ= 0,0.05,0.1,0.15,0.2,0.25,0.3,0.35,0.4,0.45,0.5,0.55,0.6,0.65,0.7,0.75,0.8,0.85,0.9,0.92,0.94

,0.96,0.98. A trapezoid sum was used to calculate the corresponding integral. In this calculation, a 20 ps pre-equlibration was performed followed by a 40 ps data collection. For each calculation, forward and backward paths for the λ values were performed to ensure convergence and calculate the error of the free energy values. The value of $\frac{\partial U}{\partial \lambda}$ changes much slowly when turning off the charges of the probe molecule. No scaling of the potential was used (klambda=1). When calculating the vdW term of $\frac{\partial U}{\partial \lambda}$, a scaling potential was used (klambda=6). To ensure the convergence when calculating vdW term of $\frac{\partial U}{\partial \lambda}$, we compared the $\left\langle \frac{\partial U}{\partial \lambda} \right\rangle$ at each λ value between the forward and backward path. The difference of $\left\langle \frac{\partial U}{\partial \lambda} \right\rangle$ at each λ value was controlled to be at most 3.7 kcal/mol (λ is a unitless parameter) and the total of the errors for all calculations are less than 0.42 kcal/mol (see Table 2). To achieve this convergence, additional simulations at some λ values were needed. In addition, we also found the simulations became less stable when λ is greater than 0.92 in most cases. For these calculations, 1 fs time step was used with the same data collection time as in other λ values. Results of all calculations were listed in Table S1.

Cosolvent mapping simulations and analyses

An equilibrated cosolvent box (20% v/v isopropanol in water) was provided by Dr. Barril. For MD simulations, the structure of thermolysin with two isopropanol molecules at sites 1 and 2 were first neutralized with counter ions and placed in a 13 Å octahedral box of the cosolvent system. After a 3000-step minimization, a series of equilibration protocols were used as follows:

1. 50 ps, NVT, temperature changes from 0 to 298 K, all heavy atoms of proteins were constrained with a 5 kcal/mol/Å^2 harmonic force constant.
2. 50 ps, NPT, temperature changes from 298 to 350 K, all heavy atoms of proteins were constrained with a 1 kcal/mol/Å^2 harmonic force constant.
3. 50 ps, NPT, temperature changes from 400 to 450 K, all heavy atoms of proteins were constrained with a 1 kcal/mol/Å^2 harmonic force constant.
4. 50 ps, NPT, temperature changes from 450 to 500 K, all heavy atoms of proteins were constrained with a 1 kcal/mol/Å^2 harmonic force constant.
5. 50 ps, NPT, temperature changes from 500 to 550 K, all heavy atoms of proteins were constrained with a 1 kcal/mol/Å^2 harmonic force constant.
6. 100 ps, NPT, temperature kept at 550 K, all heavy atoms of proteins were constrained with a 1 kcal/mol/Å^2 harmonic force constant.
7. 50 ps, NPT, temperature changes from 550 to 425 K, all heavy atoms of proteins were constrained with a 1 kcal/mol/Å^2 harmonic force constant.
8. 50 ps, NPT, temperature changes from 425 to 298 K, all heavy atoms of proteins were constrained with a 1 kcal/mol/Å^2 harmonic force constant.
9. 1 ns, NPT, temperature kept at 298K and only three residues coordinated with the zinc ion were constrained with a 1 kcal/mol/Å^2 harmonic force constant.

A production simulation of 16 ns at 298K was performed. Snapshots of the whole system were saved every 2 ps.

The hotspot analyses are similar to those reported by Barril et. al.[8] but with our own implementation. They are as follows. Thermolysin from the 16 ns simulation were aligned first and the cosolvents were imaged into an octahedral box. Evenly spaced 0.5 Å

grids covering the entire thermolysin molecule were created and the counts of the probe atoms in isopropanol (terminal carbon or hydroxyl oxygen atoms) occupying each grid point were calculated. These calculations were performed using ptraj program from Amber suite. After normalizing the counts of the probe atoms occupying each grid point(N_p), an empirical formula:

$$\Delta G^{CM} = - kT \log (N_p/N_0), \quad (S5)$$

where k is the Boltzman constant, T is room temperature and N_0 is the counts of the probe atoms occupying any grid point (expected occupancy) in 20% v/v cosolvent box without a protein. The N_0 values for terminal carbon and hydroxyl oxygen atoms were provided by Dr. Barril. Based on the values of ΔG, only the grids calculated to be lower than -0.83 kcal/mol were kept for the next analysis. A search procedure was used to find the grid point with the lowest ΔG value and grid points within 1.4 Å of this grid point were removed. The same procedure continued until all grid points were visited. Results of this procedure give pseudo atoms in space. A cluster analysis based on a Depth First Search (DFS) algorithm with a criterium of two pseudo atoms being connected within 2.5 Å was implemented in a C++ program to generate chemical graphs. In analogy to the graph theory, pseudo atoms are vertices and bonds are edges. These chemical graphs form the bases of hotspots probes and were reported in Figure 2. The number of pseudo atoms in each graph varies from one to five in this study. For the terminal carbon atom probes of isopropanol (two atoms), a total of 57 chemical graphs were obtained. For the hydroxyl atom probe of isopropanol, a total of 44 chemical graphs were obtained.

Table S1. Standard free energy of the probe molecules decoupled from bulk water and sites in thermolysin calculated by double-decoupling approach using thermodynamic integration method. The unit is kcal/mol.

Ideal Gas	path	elec	vdw	RTln(C0Vi)	-RT*ln($\sigma_{P\text{-}probe}/\sigma_{probe}$)	total	constrain force contants
IPA	forward	22.55	-3.22		0.00	19.33	0
	backward	22.55	-3.29			19.26	
	average					19.29	
IPH	forward	9.64	-3.98		0.00	5.66	0
	backward	9.64	-4.03			5.60	
	average					5.63	
ACN	forward	34.98	4.28		0.00	39.26	0
	backward	34.97	4.46			39.44	
	average					39.35	
Bulk Water							
IPA	forward	26.96	-3.94		0.00	23.02	0
	backward	27.07	-3.30			23.77	
	average					23.40	
IPH	forward	15.04	-3.51		0.00	11.54	0
	backward	15.20	-3.68			11.52	
	average					11.53	
ACN	forward	40.54	3.23		0.00	43.77	0
	backward	40.42	3.31			43.73	
	average					43.75	
Thermolysin							
IPA(site 1)	forward	26.82	0.98	-1.21	0.18	26.77	0.61
	backward	26.98	0.48	-1.21	0.18	26.43	
	average					26.60	
IPA(site 2)	forward	25.44	4.94	-1.94	0.18	28.62	4.01
	backward	24.54	5.03	-1.94	0.18	27.81	
	average					28.22	
IPH(site 1)	forward	13.60	3.04	-1.08	0.18	15.74	0.43
	backward	15.22	1.65	-1.08	0.18	15.96	
	average					15.85	
ACN(site 1)	forward	41.33	6.71	-0.58	0.18	47.64	0.12
	backward	41.75	5.51	-0.58	0.18	46.86	
	average					47.25	

References:

(1) English, A. C.; Done, S. H.; Caves, L. S. D.; Groom, C. R.; Hubbard, R. E. *Proteins: Structure, Function, and Genetics* **1999**, *37*, 628.

(2) English, A. C.; Groom, C. R.; Hubbard, R. E. *Protein Eng.* **2001**, *14*, 47.

(3) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucl. Acids Res.* **2000**, *28*, 235.

(4) Case, D. A.; Darden, T. A.; Cheatham III, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Wang, B.; Pearlman, D. A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J. W.; Ross, W. S.; Kollman, P. A. University of California, San Francisco, 2004.

(5) Case, D. A.; Darden, T. A.; T.E. Cheatham, I.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz, K. M.; B.Wang; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; K.F.Wong; Paesani, F.; Vanicek, J.; X.Wu; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. A.; University of California, San Francisco: 2008.

(6) Wang, J.; Cieplak, P.; Peter A. Kollman *Journal of Computational Chemistry* **2000**, *21*, 1049.

(7) Ryde, U. *Proteins* **1995**, *21*, 40.

(8) Seco, J.; Luque, F. J.; Barril, X. *Journal of Medicinal Chemistry* **2009**, *52*, 2363.

(9) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski Jr., V. G.; Montgomery, J. A.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M.; Johnson, B.; Chen, W.; Wong, M. M.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A.; Gaussian, Inc.: Pittsburgh PA, 1998.

(10) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *The Journal of Chemical Physics* **1983**, *79*, 926.

(11) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *Journal of Computational Physics* **1977**, *23*, 327.

(12) Darden, T. A.; York, D. M.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.

(13) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. *Biophys J* **1997**, *72*, 1047.

(14) Shirts, M. R.; Pitera, J. W.; Swope, W. C.; Pande, V. S. *Journal of Chemical Physics* **2003**, *119*, 5740.

(15) Hamelberg, D.; McCammon, J. A. *J Am Chem Soc* **2004**, *126*, 7683.

(16) Lu, Y.; Yang, C. Y.; Wang, S. *J Am Chem Soc* **2006**, *128*, 11830.