Supporting Information

Ligand Gaussian accelerated molecular dynamics (LiGaMD): Characterization of ligand binding thermodynamics and kinetics

Yinglong Miao*, Apurba Bhattarai, Jinan Wang

Center for Computational Biology and Department of Molecular Biosciences, University of Kansas, Lawrence, KS 66047, USA

* Corresponding email: miao@ku.edu

Supporting Methods

Implementation of ligand Gaussian accelerated molecular dynamics (LiGaMD)

Ligand Gaussian accelerated molecular dynamics (LiGaMD) is currently implemented in the GPU version of AMBER 20[1], but should be transferable to other molecular dynamics programs as well. LiGaMD provides enhanced sampling of protein-ligand binding and unbinding. Following is a list of the input parameters for a LiGaMD simulation:

igamd Flag to apply boost potential
= **0** (default) no boost is applied
= **1** boost on the total potential energy only (GaMD_Tot)
= **2** boost on the dihedral energy only (GaMD_Dih)
= **3** dual boost on both dihedral and total potential energy (GaMD_Dual)
= **4** boost on the non-bonded potential energy only (GaMD_NB)
= **5** dual boost on both dihedral and non-bonded potential energy (GaMD_NB_Dual)
= **10** boost on ligand non-bonded potential energy (LiGaMD)
= **11** dual boost on both non-bonded potential energy of the bound ligand and the remaining potential energy of the entire system (LiGaMD_Dual)

iE Flag to set the threshold energy E for applying all boost potentials
= **1** (default) set the threshold energy to the lower bound $E = V_{max}$
= **2** set the threshold energy to the upper bound $E = V_{min} + (V_{max} - V_{min})/k_0$

iEP Flag to overwrite *iE* and set the threshold energy E for applying the first boost potential in dual-boost schemes
= **1** (default) set the threshold energy to the lower bound $E = V_{max}$
= **2** set the threshold energy to the upper bound $E = V_{min} + (V_{max} - V_{min})/k_0$

iED Flag to overwrite *iE* and set the threshold energy *E* for applying the second boost potential in dual-boost schemes
= 1 (default) set the threshold energy to the lower bound $E = V_{max}$
= 2 set the threshold energy to the upper bound $E = V_{min} + (V_{max} - V_{min})/k_0$

ntcmdprep The number of preparation conventional molecular dynamics steps. This is used for system equilibration and the potential energies are not collected for calculating their statistics. The default is 200,000 for a simulation with 2 *fs* timestep.

ntcmd The number of initial conventional molecular dynamics simulation steps used to calculate the maximum, minimum, average and standard deviation of the system potential energies (i.e., V_{max}, V_{min}, V_{avg}, σ_V). The default is 1,000,000 for a simulation with 2 *fs* timestep.

ntebprep The number of preparation biasing molecular dynamics simulation steps. This is used for system equilibration after adding the boost potential and the potential statistics (i.e., V_{max}, V_{min}, V_{avg}, σ_V) are not updated during these steps. The default is 200,000 for a simulation with 2 fs timestep.

nteb The number of biasing molecular dynamics simulation steps. Potential statistics (V_{max}, V_{min}, V_{avg}, σ_V) are updated between the ***ntebprep*** and ***nteb*** steps and used to calculate the GaMD acceleration parameters, particularly E and k_0. The default is 1,000,000 for a simulation with 2 *fs* timestep. A greater value may be needed to ensure that the potential statistics and GaMD acceleration parameters level off before running production simulation between the ***nteb*** and ***nstlim*** (total simulation length) steps. Moreover, ***nteb*** can be set to ***nstlim***, by which the potential statistics and GaMD acceleration parameters are updated adaptively throughout the simulation. This in some cases provides more appropriate acceleration.

ntave The number of simulation steps used to calculate the average and standard deviation of potential energies. This variable has already been used in AMBER. The default is set to 50,000 for GaMD simulations. It is recommended to be updated as about 4 times of the total number of atoms in the system. Note that ***ntcmd*** and ***nteb*** need to be multiples of ***ntave***.

irest_gamd Flag to restart GaMD simulation
= 0 (default) new simulation. A file "gamd-restart.dat" that stores the maximum, minimum, average and standard deviation of the potential energies needed to calculate the boost potentials (depending on the ***igamd*** flag) will be saved automatically after GaMD equilibration stage.
= 1 restart simulation (***ntcmd*** and ***nteb*** are set to 0 in this case). The "gamd-restart.dat" file will be read for restart.

sigma0P The upper limit of the standard deviation of the first potential boost that allows for accurate reweighting. The default is 6.0 (unit: kcal/mol).

sigma0D The upper limit of the standard deviation of the second potential boost that allows for accurate reweighting in dual-boost simulations (e.g., ***igamd*** = 2, 3, 5 and 11). The default is 6.0 (unit: kcal/mol).

timask1 Specifies atoms of the bound ligand in ambmask format. This variable has already been used in AMBER. The default is an empty string.

scmask1 Specifies atoms of the bound ligand that will be described using soft core in ambmask format. This variable has already been used in AMBER. The default is an empty string.

nlig The total number of ligand molecules in the system. The default is 0.

ibblig The flag to boost the bound ligand selectively with ***nlig*** > 1
= 0 (default) no selective boost
= 1 boost the bound ligand selectively out of ***nlig*** ligand molecules in the system based on the shortest distance to the protein target site
= 2 boost the bound ligand selectively out of ***nlig*** ligand molecules in the system based on the smallest mean square displacement (MSD)

atom_p Serial number of a protein atom (starting from 1 for the first protein atom) used to calculate the ligand distance. It is used only when ***ibblig*** = 1. The default is 0.

atom_l Serial number of a ligand atom (starting from 1 for the first ligand atom) used to calculate the ligand distance to the protein. It is used only when ***ibblig*** = 1 or 2. The default is 0.

ntmsd Number of timesteps corresponding to the lagging time used to calculate the ligand MSD. It is used only when ***ibblig*** = 2. The default is 50,000.

nftau Number of saved simulation frames used to calculate the ligand MSD. MSD is calculated for every time window of *ntwin* = ***ntmsd*** + ***nftau*ntwx*** steps, for which simulation frames are saved every ***ntwx*** steps. It is used only when ***ibblig*** = 2. The default is 10.

dblig **(Optional)** The cutoff distance between atoms ***atom_p*** and ***atom_l*** used to identify the ligand that is bound in the protein when ***ibblig*** = 1 or the cutoff MSD of atom ***atom_l*** used to identify the ligand that is bound in the protein when ***ibblig*** = 2. If ***dblig*** (default 1.0d99 Å) is not set in the input file, the first boost potential will be selectively applied to the ligand with the smallest distance to the protein (***ibblig*** = 1) or the smallest MSD (***ibblig*** = 2) out of ***nlig*** ligand molecules in the system.

Example input parameters used in LiGaMD_Dual simulations of ligand binding to trypsin include the following:

```
igamd = 11, irest_gamd = 0,
ntcmd = 700000, nteb = 27300000, ntave = 140000,
ntcmdprep = 280000, ntebprep = 280000,
sigma0P = 4.0, sigma0D = 6.0, iEP = 2, iED=1,

icfe = 1, ifsc = 1, gti_cpu_output = 0, gti_add_sc = 1,
timask1 = ':225', scmask1 = ':225',
timask2 = '', scmask2 = '',

ibblig = 1, nlig = 10, atom_p = 2472, atom_l = 4
```

The LiGaMD algorithm is summarized as the following:

```
LiGaMD {
   If (irest_gamd == 0) then
      For i = 1, …, ntcmd // run initial conventional molecular dynamics
         If (i >= ntcmdprep) Update Vmax, Vmin
         If (i >= ntcmdprep && i%ntave ==0) Update Vavg, sigmaV
      End
      Save Vmax,Vmin,Vavg,sigmaV to “gamd_restart.dat” file
      Calc_E_k0(iE,sigma0,Vmax,Vmin,Vavg,sigmaV)
```

```
    For i = ntcmd+1, …, ntcmd+nteb // Run biasing molecular dynamics simulation steps
        deltaV = 0.5*k0*(E-V)**2/(Vmax-Vmin)
        V = V + deltaV
        If (i >= ntcmd+ntebprep) Update Vmax, Vmin
        If (i >= ntcmd+ntebprep && i%ntave ==0) Update Vavg, sigmaV
        Calc_E_k0(iE,sigma0,Vmax,Vmin,Vavg,sigmaV)
    End
    Save Vmax,Vmin,Vavg,sigmaV to "gamd_restart.dat" file
  else if (irest_gamd == 1) then
    Read Vmax,Vmin,Vavg, sigmaV from "gamd_restart.dat" file
  End if

  For i = ntcmd+nteb+1, …, nstlim // run production simulation
    deltaV = 0.5*k0*(E-V)**2/(Vmax-Vmin)
    V = V + deltaV
  End

  ntwin = ntmsd+nftau*ntwx
  lig0=1 // ID of the bound ligand

  If (ibblig == 1 && i%ntwx ==0) then // identify the bound ligand according to distance
    For ilig = 1, …, nlig
      dlig = distance(atom_p, atom_l)
      If (dmin <= dlig) blig_min=ilig; dmin=dlig
    End
    If (dmin <= dblig) blig=blig_min
  else if (ibblig == 2 && i%ntwin ==0) then // identify the bound ligand according to MSD
    For ilig = 1, …, nlig
      dlig = msd(atom_l, ntmsd, nftau)
      If (dmin <= dlig) blig_min=ilig; dmin=dlig
    End
    If (dmin <= dblig) blig=blig_min
  End if
  If (blig != lig0) Swap atomic coordinates, forces and velocities of ligand blig with lig0 for selective higher boost
}

Subroutine Calc_E_k0(iE,sigma0,Vmax,Vmin,Vavg,sigmaV) {
if iE = 1 :
      E = Vmax
      k0' = (sigma0/sigmaV) * (Vmax-Vmin)/(Vmax-Vavg)
      k0 = min(1.0, k0')
else if iE = 2 :
      k0" = (1-sigma0/sigmaV) * (Vmax-Vmin)/(Vavg-Vmin)
      if 0 < k0" <= 1 :
            k0 = k0"
            E = Vmin + (Vmax-Vmin)/k0
      else
            E = Vmax
            k0' = (sigma0/sigmaV) * (Vmax-Vmin)/(Vmax-Vavg)
            k0 = min(1.0, k0')
      end
end
}
```

Table S1 The guest binding and unbinding time periods (τ_B and τ_U) recorded from LiGaMD simulations of host-guest binding systems.

Host	Ligand	GaMD (300 ns x 3)	τ_B (ns)	τ_U (ns)
CD: GAFF	Aspirin	LiGaMD	32.0228, 8.5966, 3.1112, 39.9596, 21.2604	9.9081, 194.0138, 86.3999
		LiGaMD_Dual	8.2607, 0.796, 7.9322, 0.5776, 4.5163, 1.758, 5.9326	201.0795, 13.8457, 24.4303, 239.0854, 8.3765
CD: q4MD	Aspirin	LiGaMD	31.9081, 61.633, 43.8801, 8.2827, 16.9167, 20.3661, 14.1932, 19.3123, 27.4969	1.0855, 2.1165, 86.4148, 35.6304, 17.3932, 115.365, 89.763
		LiGaMD_Dual	3.907, 30.0912, 36.3697, 1.4833, 10.1038, 1.7181, 12.8671, 56.2154, 3.9519, 8.4395, 8.9309, 4.4481, 4.2826, 26.9633, 38.3872, 20.3627, 9.6437, 44.3214	8.4001, 29.2999, 6.2442, 42.1422, 12.1334, 39.4287, 54.0979, 62.0065, 15.8736, 45.7133, 6.6287, 29.8733, 13.1733, 60.1833, 65.7046, 9.4815
	1-Butanol	LiGaMD	5.4329, 2.5518, 11.281, 3.0967, 6.9, 2.606, 0.7952, 3.3571, 1.0371, 1.1861, 5.05, 0.9876, 1.275, 0.7967, 0.4231, 1.0814, 1.0247, 0.9549, 0.7596, 4.4845	5.1734, 72.0719, 51.3295, 15.962, 56.6445, 0.9302, 16.088, 12.9821, 19.5307, 84.7936, 62.3141, 27.7493, 39.8894, 61.5153, 32.1647, 24.3296, 24.2369, 38.1307, 27.2135, 79.995
		LiGaMD_Dual	1.4312, 2.3294, 4.6188, 1.5447, 1.2259, 2.1735, 1.6123, 1.8385, 0.8007, 5.3053, 0.6236, 0.9946, 0.924, 0.3333, 4.6931, 1.8042, 0.8413, 2.2369, 1.7726, 2.692, 4.133, 1.7794, 8.4176, 2.262, 1.5251, 6.0076, 1.262, 1.5508, 1.1358, 1.6418	1.8388, 7.8776, 20.9217, 79.5963, 41.6825, 50.1632, 22.8479, 1.5881, 11.4997, 18.477, 65.5491, 16.961, 47.0128, 8.5518, 2.548, 2.8552, 5.14, 16.607, 5.6482, 41.906, 35.6489, 4.0918, 18.9406, 20.069, 14.6097, 8.0379, 85.2561, 4.8654, 24.6768, 45.8237

Table S2 Energy barriers of host-guest dissociation ("off") and binding ("on") calculated from the reweighed ($\Delta \boldsymbol{F}$) and modified (no reweighting, $\Delta \boldsymbol{F^*}$) free energy profiles, curvatures of the reweighed (w) and modified (w^*) free energy profiles near the guest Bound ("B"), Barrier ("Br") and Unbound ("U") states, and the ratio of apparent diffusion coefficients calculated from the LiGaMD and LiGaMD_Dual simulations without reweighting (modified, D^*) and with reweighting (D). Results are listed for the following systems: (A) aspirin binding to CD with the GAFF force field, (B) aspirin binding to CD with the q4MD force field and (C) 1-butanol binding to CD with the q4MD force field.

Sim	$\Delta \boldsymbol{F}$ **(kcal/mol)**		$\Delta \boldsymbol{F^*}$ **(kcal/mol)**		***w***			***w****			***D**/D***	
	Off	**On**	**Off**	**On**	**B**	**Br**	**U**	**B**	**Br**	**U**	**Off**	**On**
LiGaMD	3.09 ± 0.37	0.61 ± 0.37	1.49 ± 0.35	1.87 ± 0.28	0.38 ± 0.21	0.08 ± 0.04	0.11 ± 0.01	0.32 ± 0.08	0.07 ± 0.08	0.15 ± 0.01	1.14	1.53
LiGaMD_Dual	2.06 ± 0.28	0.40 ± 0.34	0.59 ± 0.37	1.83 ± 0.28	0.49 ± 0.08	0.07 ± 0.05	0.12 ± 0.03	0.35 ± 0.03	0.05 ± 0.15	0.13 ± 0.01	1.23	0.85

(A) CD (GAFF) – Aspirin

Sim	$\Delta \boldsymbol{F}$ **(kcal/mol)**		$\Delta \boldsymbol{F^*}$ **(kcal/mol)**		***w***			***w****			***D**/D***	
	Off	**On**	**Off**	**On**	**B**	**Br**	**U**	**B**	**Br**	**U**	**Off**	**On**
LiGaMD	4.00 ± 0.34	0.97 ± 0.35	2.18 ± 0.28	1.95 ± 0.37	0.41 ± 0.08	0.09 ± 0.08	0.13 ± 0.03	0.32 ± 0.05	0.08 ± 0.02	0.13 ± 0.02	0.84	0.12
LiGaMD_Dual	4.46 ± 0.22	0.67 ± 0.39	2.25 ± 0.23	1.67 ± 0.24	0.46 ± 0.03	0.10 ± 0.12	0.18 ± 0.02	0.33 ± 0.02	0.08 ± 0.12	0.15 ± 0.00	0.92	1.80

(B) CD (q4MD) – Aspirin

Sim	$\Delta \boldsymbol{F}$ **(kcal/mol)**		$\Delta \boldsymbol{F^*}$ **(kcal/mol)**		***w***			***w****			***D**/D***	
	Off	**On**	**Off**	**On**	**B**	**Br**	**U**	**B**	**Br**	**U**	**Off**	**On**
LiGaMD	2.65 ± 0.34	1.59 ± 0.42	1.37 ± 0.30	2.09 ± 0.19	2.23 ± 0.02	0.07 ± 0.05	0.12 ± 0.00	2.42 ± 0.02	0.06 ± 0.03	0.14 ± 0.01	0.94	1.28
LiGaMD_Dual	2.83 ± 0.24	1.35 ± 0.25	1.82 ± 0.16	2.21 ± 0.11	2.24 ± 0.00	0.07 ± 0.07	0.03 ± 0.00	2.82 ± 0.03	0.06 ± 0.01	0.12 ± 0.04	0.98	1.11

(C) CD (q4MD) – 1-Butanol

Table S3 The ligand binding and unbinding time periods (τ_B and τ_U) recorded from LiGaMD_Dual simulations of the trypsin-benzamidine binding system.

System	ID	τ_B (ns)	τ_U (ns)
Trypsin - BEN	Sim1	14.90, 8.20, 28.90	165.90, 77.00, 93.40
	Sim2	26.10, 14.00, 27.00, 18.80, 124.10	11.50, 270.70, 70.20, 415.70
	Sim3	47.39, 31.59, 12.63, 18.29, 49.09, 19.84, 41.00, 15.00, 85.13, 17.19, 60.59	145.08, 17.21, 18.71, 28.50, 38.37, 18.52, 32.75, 205.65, 13.37, 32.84
	Sim4	32.10, 70.00, 38.70, 30.30	210.40, 9.40, 21.70, 496.90
	Sim5	7.00, 46.70, 31.30, 39.60	279.80, 179.30, 265.80

Table S4 Energy barriers of trypsin-benzamidine dissociation ("off") and binding ("on") calculated from the reweighed ($\Delta \boldsymbol{F}$) and modified (no reweighting, $\Delta \boldsymbol{F^*}$) free energy profiles, curvatures of the reweighed (w) and modified (w^*) free energy profiles near the guest Bound ("B"), Barrier ("Br") and Unbound ("U") states, and the ratio of apparent diffusion coefficients calculated from the LiGaMD_Dual simulations without reweighting (modified, D^*) and with reweighting (D).

Sim	$\Delta \boldsymbol{F}$ **(kcal/mol)**		$\Delta \boldsymbol{F^*}$ **(kcal/mol)**		***w***			***w****			***D*/D***	
	Off	**On**	**Off**	**On**	**B**	**Br**	**U**	**B**	**Br**	**U**	**Off**	**On**
LiGaMD_Dual	12.17 ± 1.54	3.04 ± 2.04	1.37 ± 0.56	2.40 ± 0.41	2.39 ± 0.21	0.12 ± 0.16	0.06 ± 0.01	0.99 ± 0.05	0.04 ± 0.05	0.06 ± 0.02	1.06	15.07

Figure S1 Time courses of host-guest distances calculated from (A) LiGaMD and (B) LiGaMD_Dual simulations of CD using the GAFF force field with aspirin, (C) LiGaMD and (D) LiGaMD_Dual simulations of CD using the GAFF force field with 1-butanol, (E) LiGaMD and (F) LiGaMD_Dual simulations of CD using the q4MD force field with aspirin, (G) LiGaMD and (H) LiGaMD_Dual simulations of CD using the q4MD force field with 1-butanol.

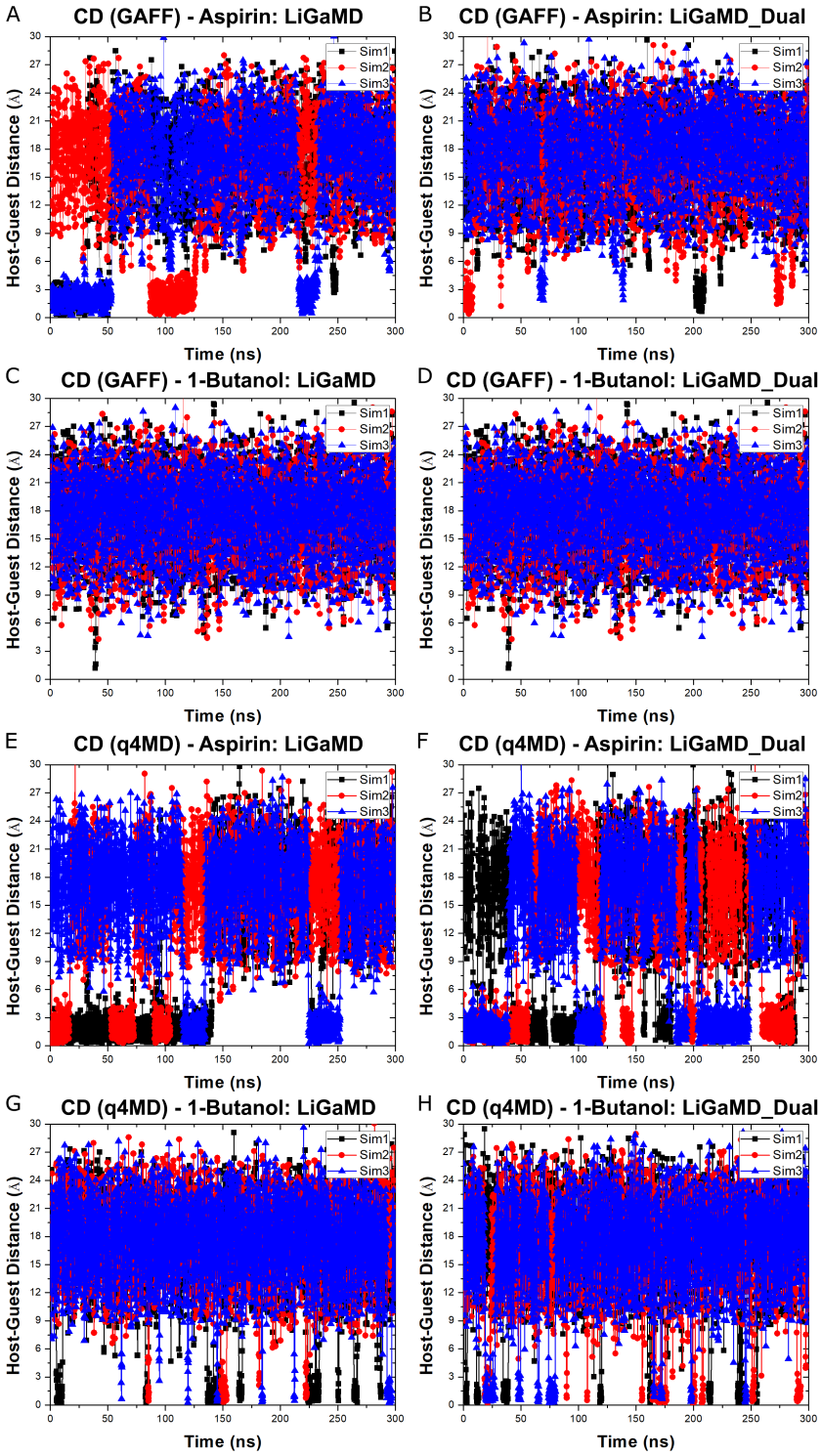


Figure S2 Reweighted and modified PMF profiles of guest 1-butanol binding to the CD host modeled with the GAFF force field: (A) LiGaMD and (B) LiGaMD_Dual simulations.

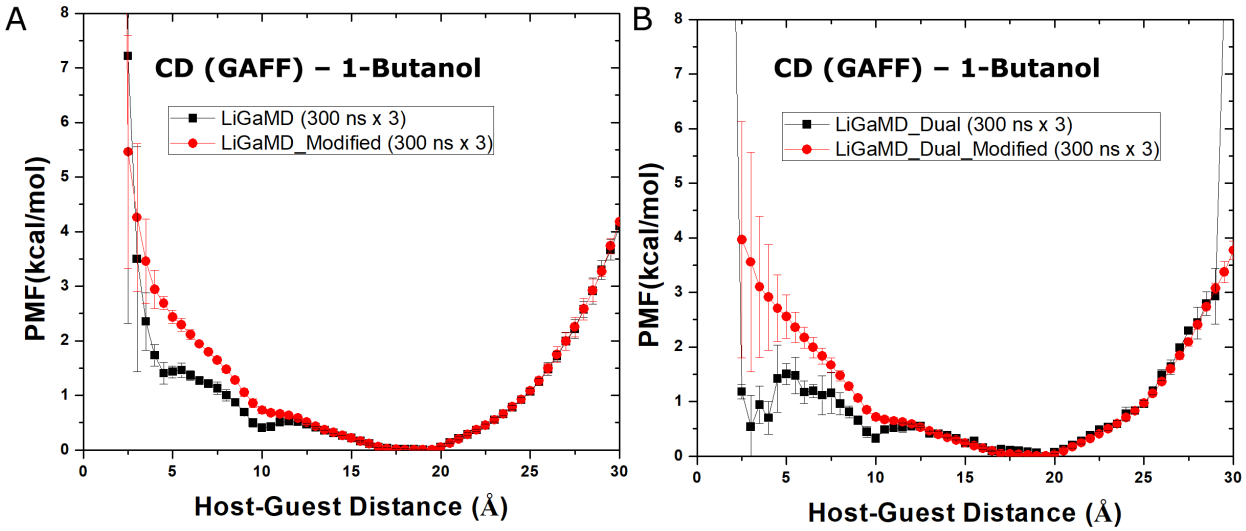


Figure S3 Time courses of the benzamidine ligand RMSD relative to the X-ray conformation obtained from LiGaMD_Dual equilibration simulations of trypsin, where the input parameter σ_{0P} was increased from 1.0 to 6.0 with the threshold energy set to upper bound for applying boost potential to the ligand non-bounded potential energy.

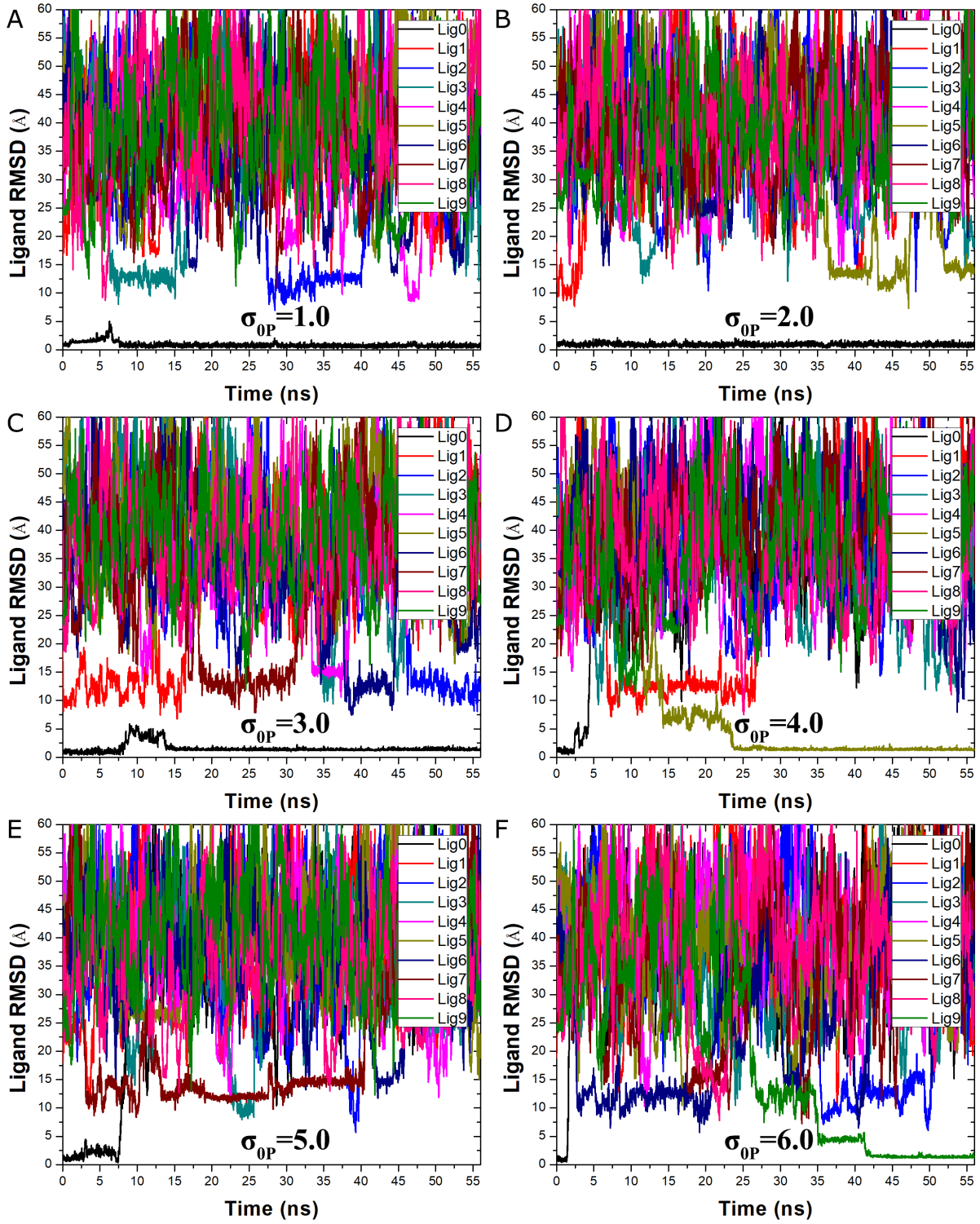


Figure S4 RMSD of the benzamidine ligand relative to the X-ray crystal conformation calculated from LiGaMD_Dual simulations as listed in **Table 4**: (A) Sim1, (B) Sim2, (C) Sim3, (D) Sim4 and (E) Sim5.

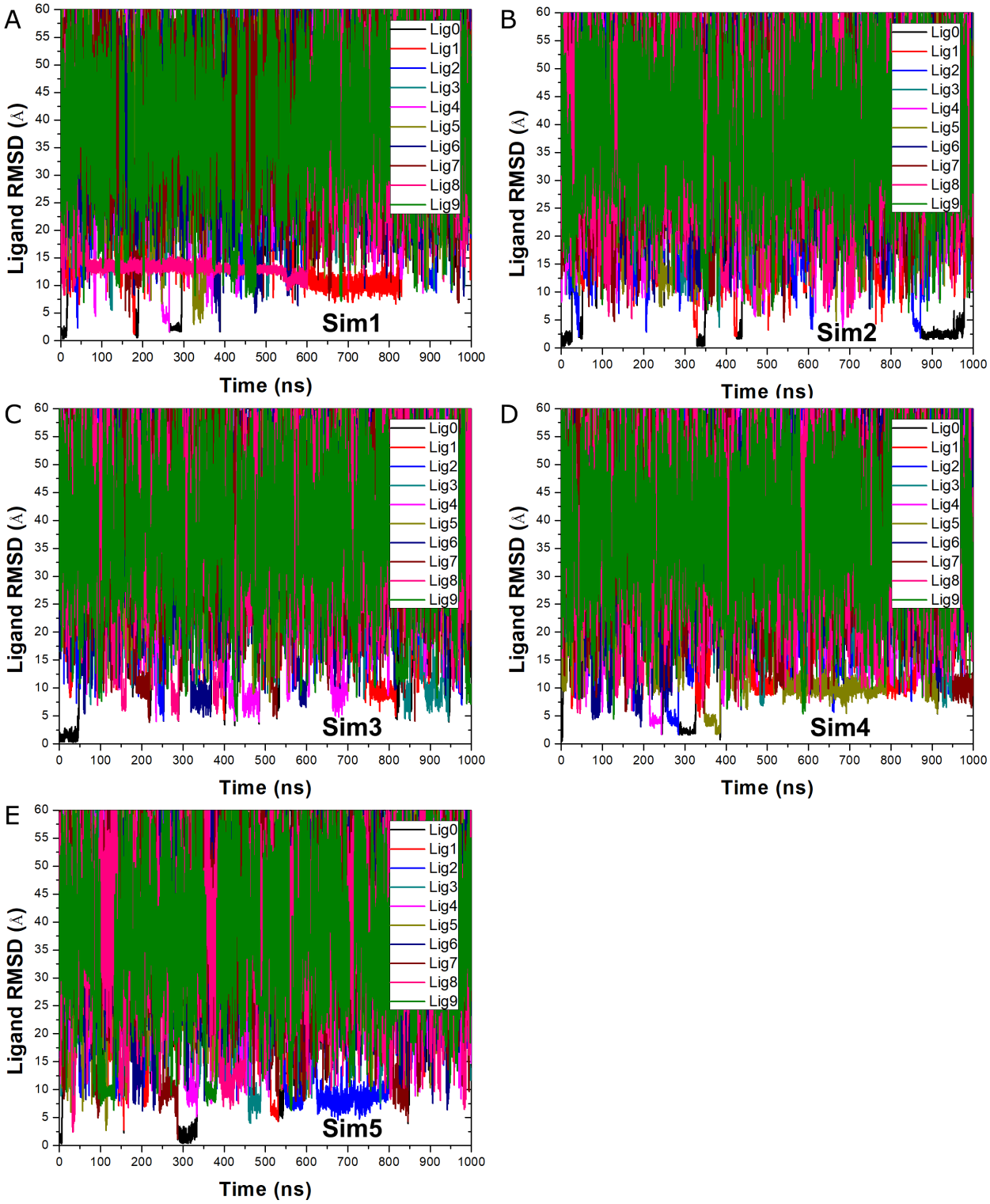


Figure S5 Distances between the charge centers at the amidine C atom in benzamidine and CG atom of Asp189 in trypsin calculated from five LiGaMD_Dual simulations as listed in **Table 4**: (A) Sim1, (B) Sim2, (C) Sim3, (D) Sim4 and (E) Sim5. Distance plots of Sim2 are provided in **Figure 5A**.

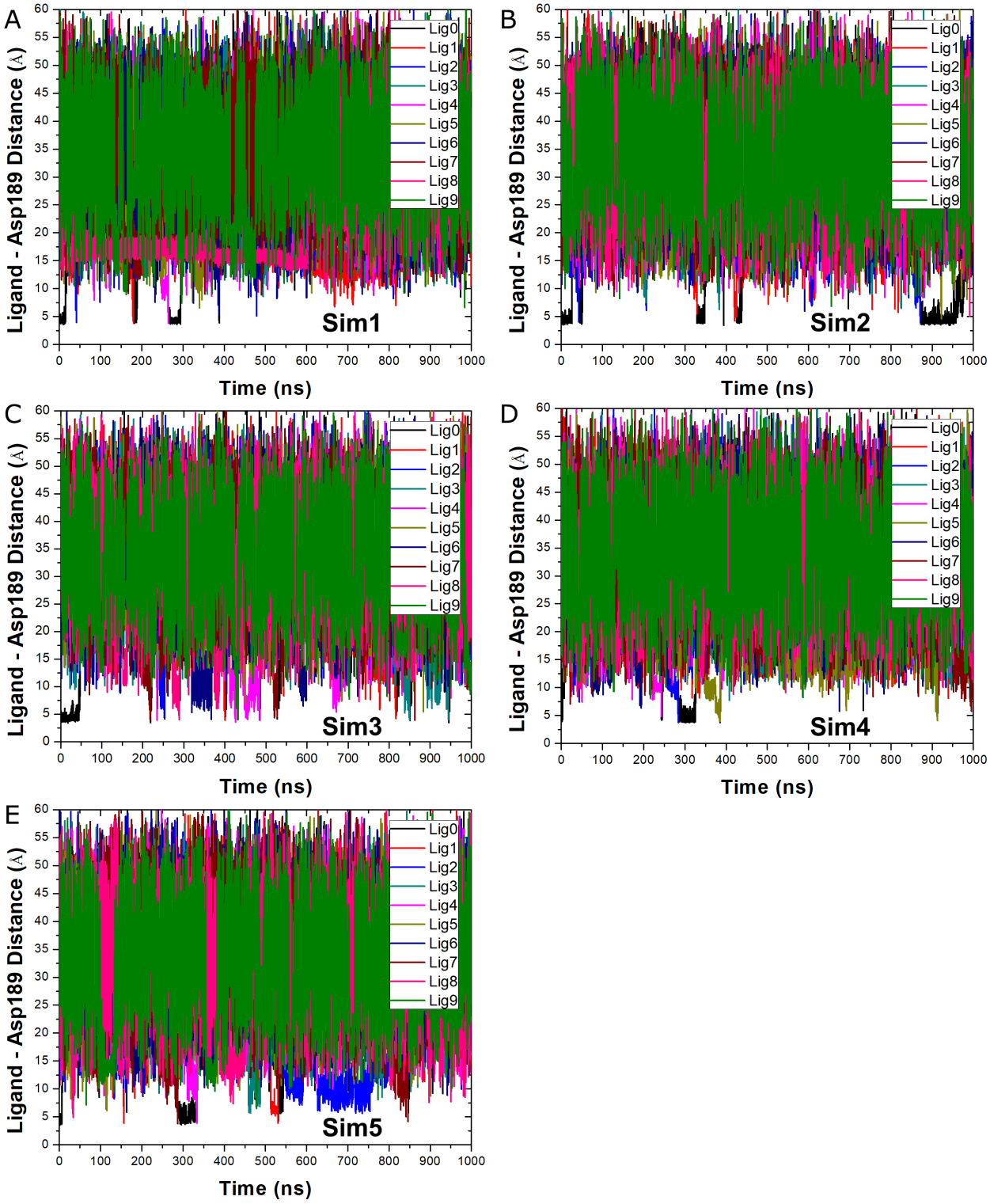


Figure S6 Reweighted 2D PMF profiles of the benzamidine:C – Asp189:CG and Trp215:NE – Asp189:CG atom distances calculated from five individual 1000 ns LiGaMD_Dual simulations of the benzamidine inhibitor binding to trypsin: (A) Sim1, (B) Sim2, (C) Sim3, (D) Sim4 and (E) Sim5.

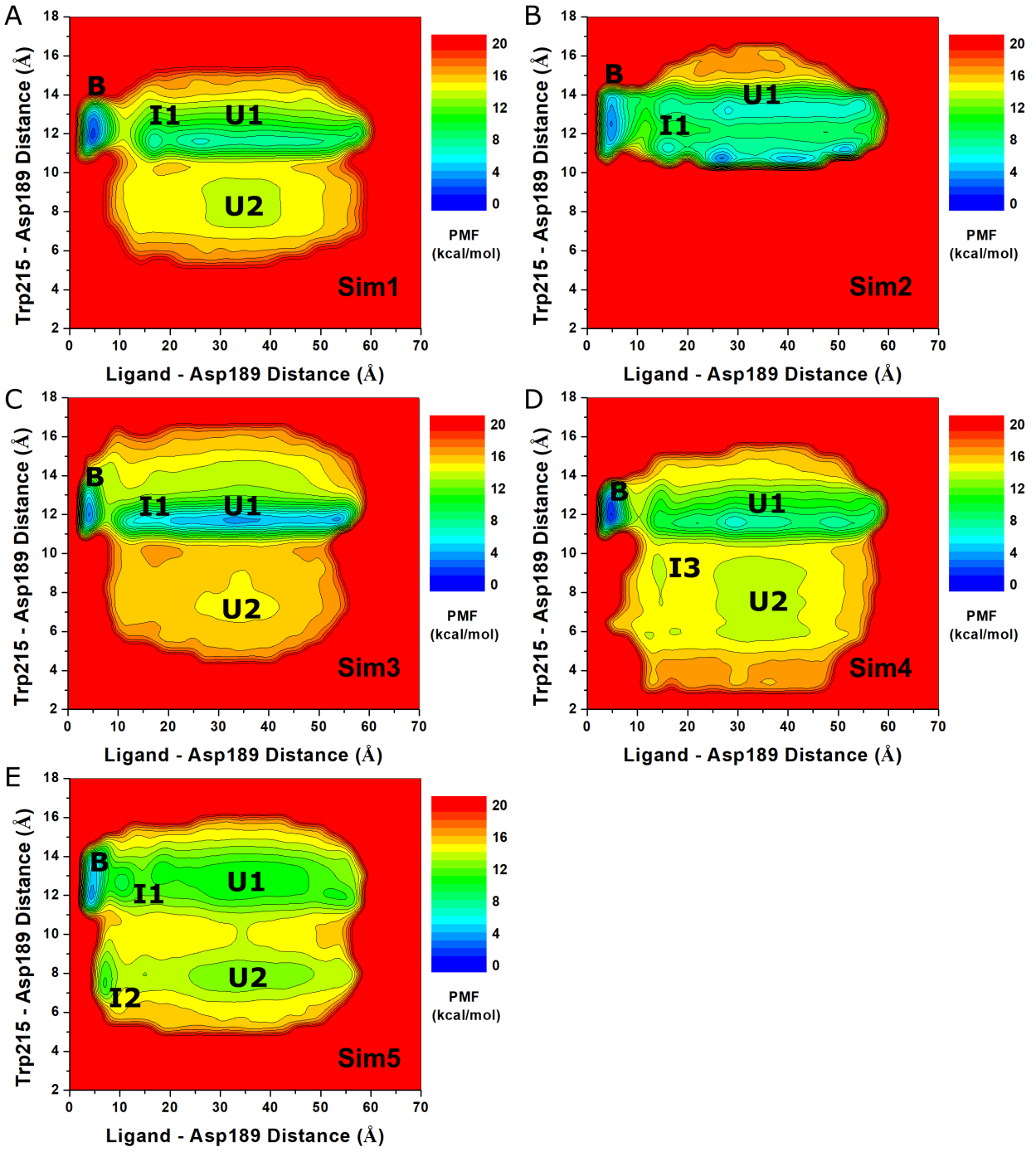
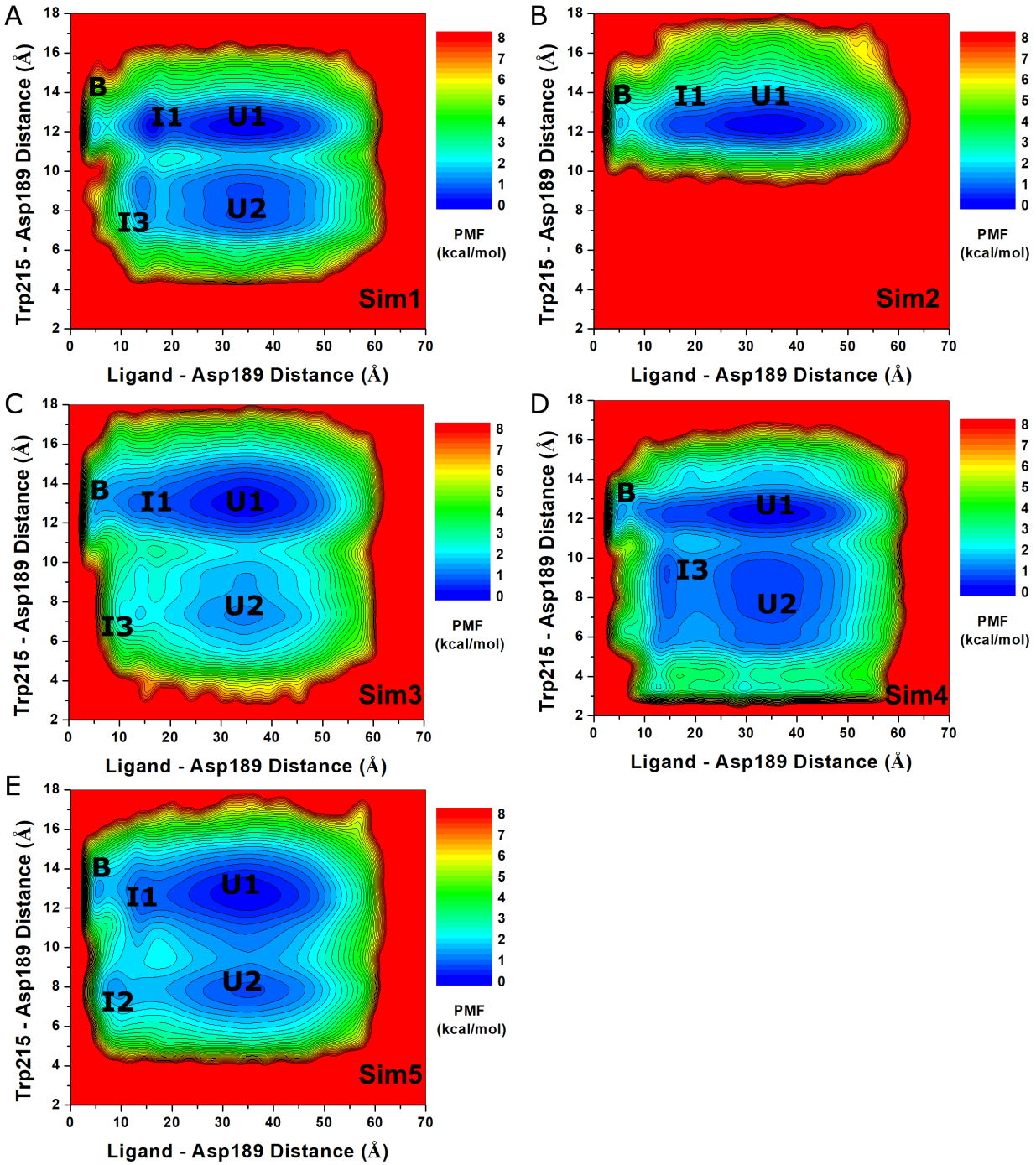


Figure S7 Modified 2D PMF profiles of the benzamidine:C – Asp189:CG and Trp215:NE – Asp189:CG atom distances calculated from five individual 1000 ns LiGaMD_Dual simulations of the benzamidine inhibitor binding to trypsin: (A) Sim1, (B) Sim2, (C) Sim3, (D) Sim4 and (E) Sim5.



References

1. Case, D. A.; Belfon, K.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; T.E. Cheatham, I.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Giambasu, G.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Harris, R.; Izadi, S.; jhala, K. K.-.; Kovalenko, A.; Krasny, R.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Man, V.; Merz, K. M.; Miao, Y.; Mikhailovskii, O.; Monard, G.; Nguyen, H.; Onufriev, A.; Pan, F.; Pantano, S.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Skrynnikov, N.; Smith, J.; Swails, J.; Walker, R. C.; Wang, J.; Wilson, L.; Wolf, R. M.; Wu, X.; York, D. M.; Kollman, P. A., AMBER 20, University of California, San Francisco. **2020**.
2. Tang, Z. Y.; Chang, C. E. A., Binding Thermodynamics and Kinetics Calculations Using Chemical Host and Guest: A Comprehensive Picture of Molecular Recognition. *J Chem Theory Comput* **2018,** *14* (1), 303-318.