

# Supporting Information

## for Gaussian Accelerated Molecular Dynamics in OpenMM

### Implementation algorithm of Gaussian accelerated Molecular Dynamics (GaMD) in OpenMM

```
GaMD {  
    // Stage 1: run short initial conventional molecular dynamics  
    For i = 1, ..., conventional_md:  
        if (i >= conventional_md_prep):  
            n = i - conventional_md_prep  
            Update(V, Vmax, Vmin)  
        if (i >= conventional_md_prep) and (i % averaging_window_interval):  
            Update(i, V, Vavg, sigmaV)  
    End  
  
    if (i == conventional_md):  
        calculate_threshold_energy_with_effective_harmonic_constant(sigma0, sigmaV, Vmax,  
        Vmin, k, E)  
  
    // Equilibrate the system after adding boost potential  
    For i = 1, ..., gamd_equilibration:  
        If (E > V):  
            deltaV = 0.5*k*(E-V)**2  
            V = V + deltaV  
        EndIf  
        Update Vmax, Vmin, Vavg, sigmaV  
        If (i >= gamd_equilibration_prep):  
            calculate_threshold_energy_with_effective_harmonic_constant(sigma0, sigmaV, Vmax,  
            Vmin, k, E)  
        EndIf  
    End  
  
    // run production simulation  
    For i = 1, ..., total_simulation_length:  
        If (E > V):  
            deltaV = 0.5*k*(E-V)**2  
            V = V + deltaV  
        EndIf  
    End  
}  
  
Subroutine UpdateMaxMin(V,Vmax,Vmin):  
    if (V > Vmax) Vmax = V  
    if (V < Vmin) Vmin = V  
}
```

```

Subroutine UpdateAvgSigma(n, V, Vmax, Vmin, Vavg, sigmaV ):
    Vdiff = V - Vavg
    Vavg = Vavg + Vdiff / n
    M2 = M2 + Vdiff * (V - Vavg)
    sigmaV = sqrt(M2 / n)
}

// Lower Bound Integrator
Subroutine calculate_threshold_energy_with_effective_harmonic_constant(sigma0,Vmax,Vmin,
k, E):
    E = Vmax
    k0' = (sigma0/sigmaV) * (Vmax-Vmin)/(Vmax-Vavg)
    k0 = min(1.0, k0')
    k = k0/(Vmax-Vmin)
}
// Upper Bound Integrator
Subroutine calculate_threshold_energy_with_effective_harmonic_constant(sigma0,Vmax,Vmin,
k, E):
    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
    k = k0/(Vmax-Vmin)
}

```

### Example Input XML file for GaMD-OpenMM simulation

```

<?xml version="1.0" ?>
<gamd>
    <temperature>298.15</temperature> <!-- unit.kelvin -->

    <system>
        <nonbonded-method>PME</nonbonded-method>
        <nonbonded-cutoff>1.0</nonbonded-cutoff> <!-- unit.nanometers -->
        <constraints>HBonds</constraints>
    </system>

    <barostat>
        <pressure>1.0</pressure> <!-- unit.bar -->
        <frequency>25</frequency>
    </barostat>

```

```

<run-minimization>True</run-minimization>

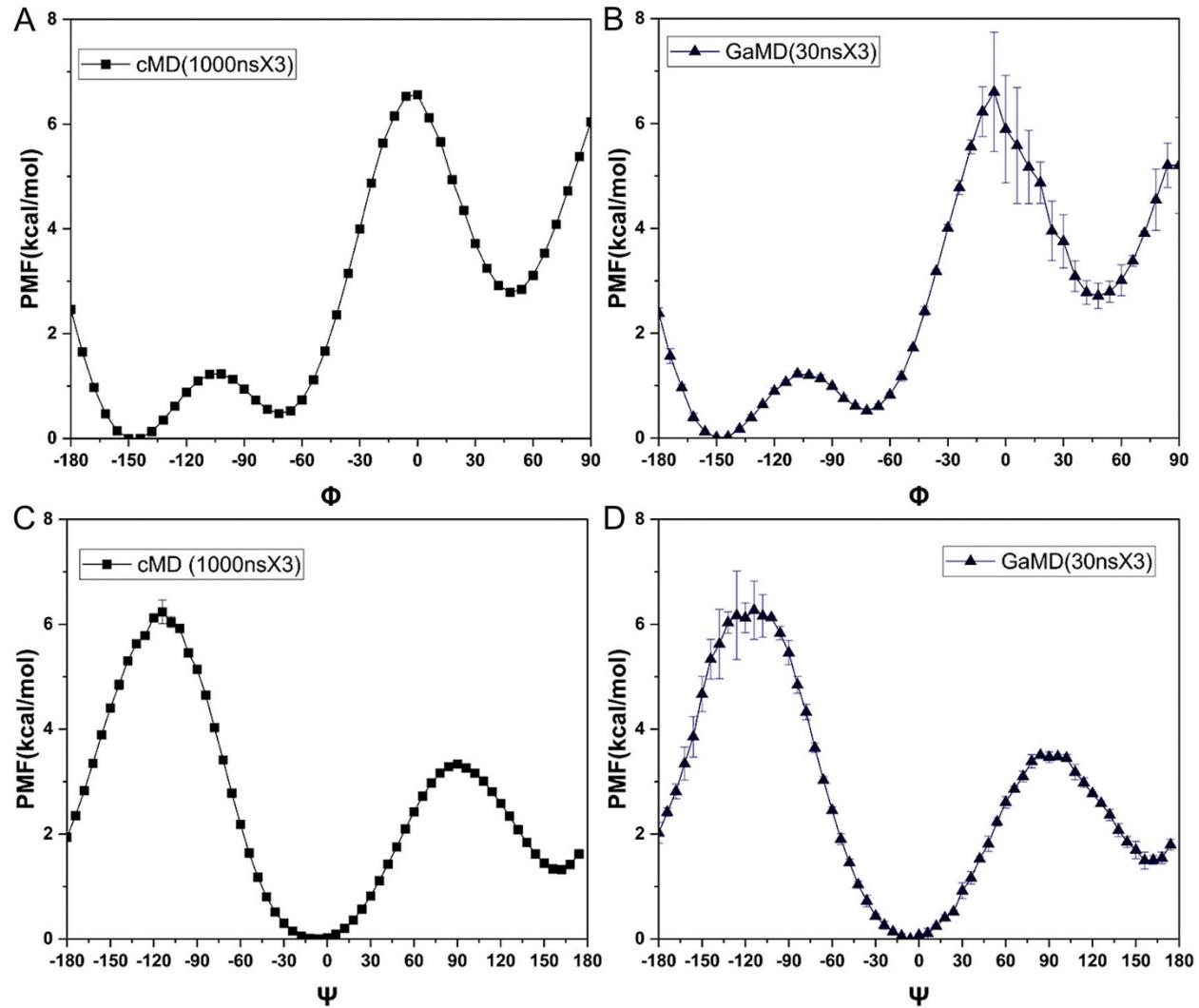
<integrator>
  <algorithm>langevin</algorithm>
  <boost-type>lower-dual</boost-type>
  <sigma0>
    <primary>6.0</primary> <!-- unit.kilocalories_per_mole -->
    <secondary>6.0</secondary> <!-- unit.kilocalories_per_mole -->
  </sigma0>
  <random-seed>0</random-seed>
  <dt>0.002</dt> <!-- unit.picoseconds -->
  <friction-coefficient>1.0</friction-coefficient> <!-- unit.picoseconds**-1 -->
  <number-of-steps>
    <conventional-md-prep>200000</conventional-md-prep>
    <conventional-md>1000000</conventional-md>
    <gamd-equilibration-prep>200000</gamd-equilibration-prep>
    <gamd-equilibration>2000000</gamd-equilibration>
    <gamd-production>3000000</gamd-production>
    <averaging-window-interval>50000</averaging-window-interval>
  </number-of-steps>
</integrator>

<input-files>
  <amber>
    <topology>data/dip.top</topology>
    <coordinates type="rst7">data/md-4ns.rst7</coordinates>
  </amber>
</input-files>

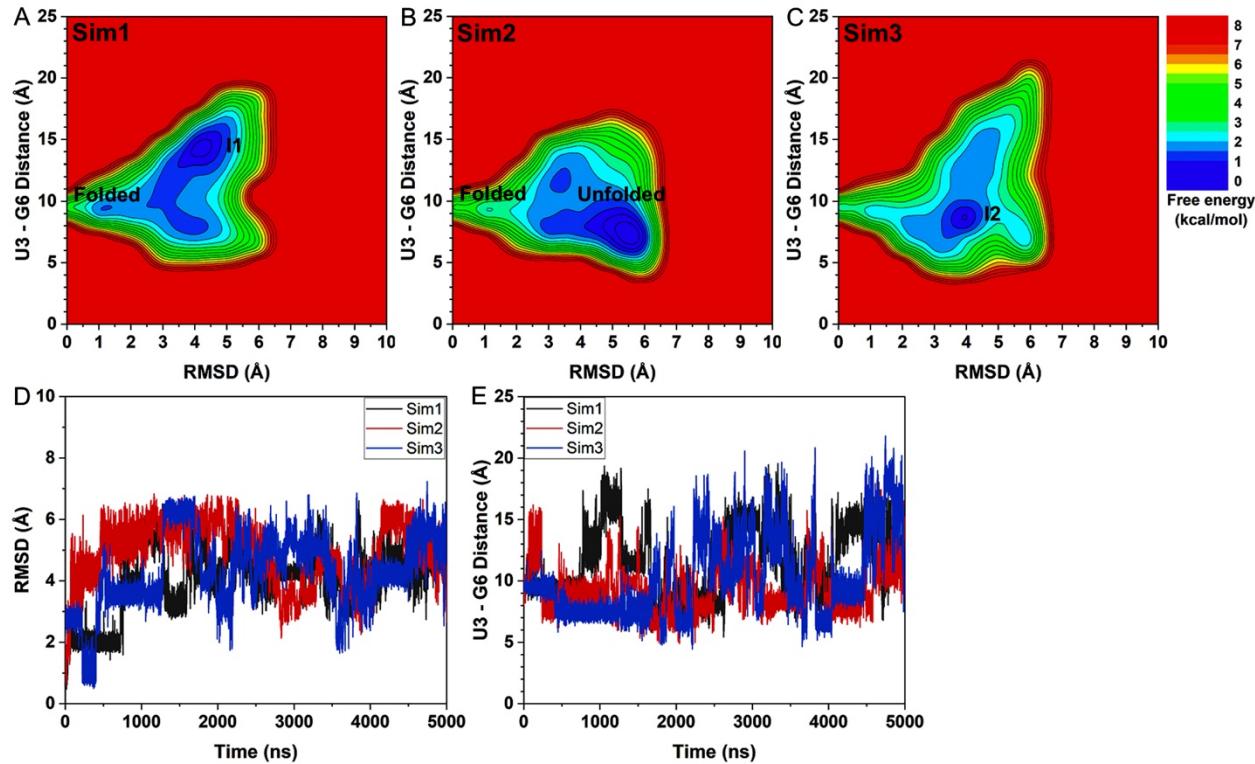
<outputs>
  <directory>output</directory>
  <overwrite-output>True</overwrite-output>
  <reporting>
    <energy>
      <interval>500</interval>
    </energy>
    <coordinates>
      <file-type>DCD</file-type>
    </coordinates>
    <statistics>
      <interval>500</interval>
    </statistics>
  </reporting>
</outputs>
</gamd>

```

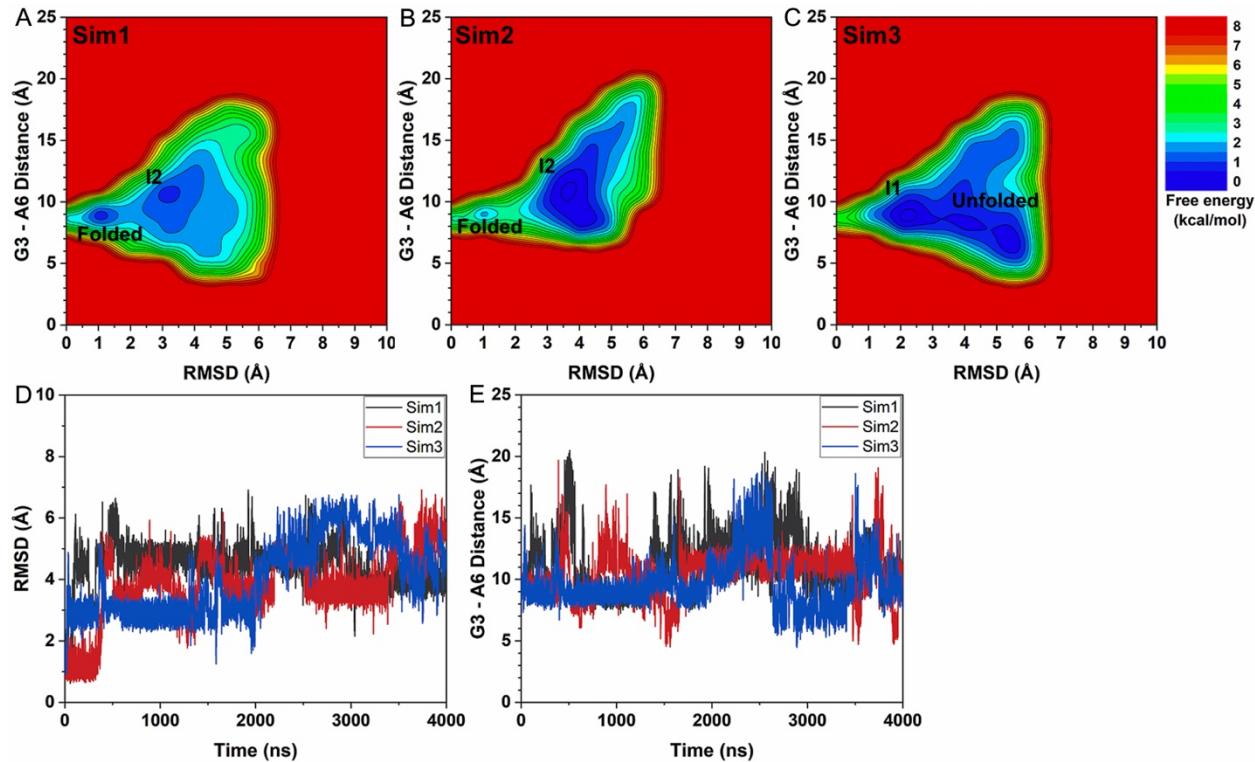
**Figure S1.** Potential of mean force (PMF) profiles of the (A-B)  $\Phi$  and (C-D)  $\Psi$  dihedrals calculated from three 1000ns cMD simulations (A,C) and three 30ns GaMD simulations (B,D).



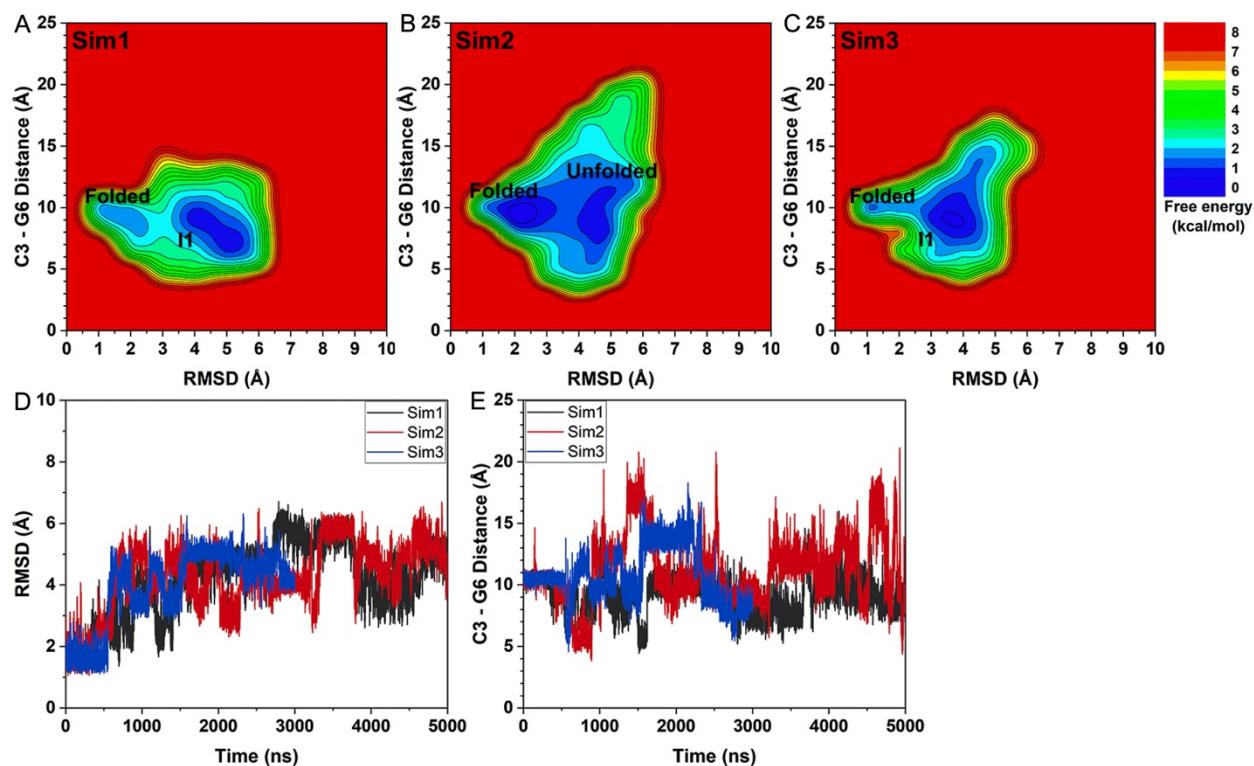
**Figure S2. (A-C)** 2D free energy profiles of the heavy-atom RMSD of UUCG relative to the 1F7Y PDB structure and the COM distance between nucleotides U3 and G6 calculated from three independent 5,000 ns GaMD simulations of the UUCG RNA tetraloop. The low-energy RNA conformational states are labeled “Folded”, “I1”, “I2”, and “Unfolded”. **(D)** Time courses of the heavy-atom RMSD of UUCG relative to the 1F7Y PDB structure calculated from three independent 5,000 ns GaMD simulations of the UUCG RNA tetraloop. **(E)** Time courses of the COM distance between nucleotides U3 and G6 calculated from three independent 5,000 ns GaMD simulations of the UUCG RNA tetraloop.



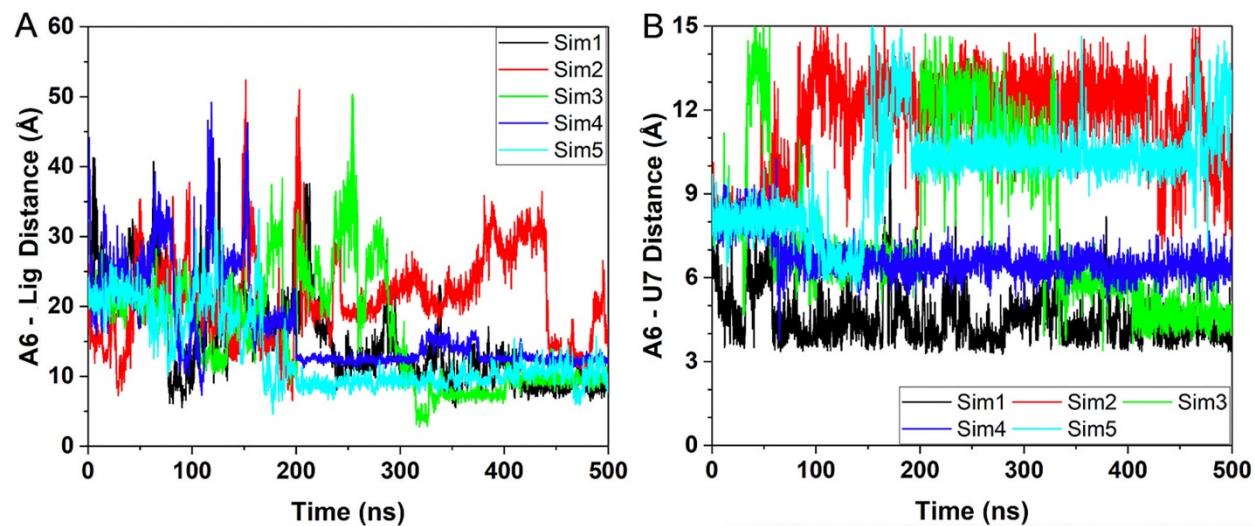
**Figure S3. (A-C)** 2D free energy profiles of the heavy-atom RMSD of GCAA relative to the 1ZIH PDB structure and the COM distance between nucleotides G3 and A6 calculated from three independent 4,000 ns GaMD simulations of the GCAA RNA tetraloop. The low-energy RNA conformational states are labeled “Folded”, “I1”, “I2”, and “Unfolded”. **(D)** Time courses of the heavy-atom RMSD of GCAA relative to the 1ZIH PDB structure calculated from three independent 4,000 ns GaMD simulations of the GCAA RNA tetraloop. **(E)** Time courses of the COM distance between nucleotides G3 and A6 calculated from three independent 4,000 ns GaMD simulations of the GCAA RNA tetraloop.



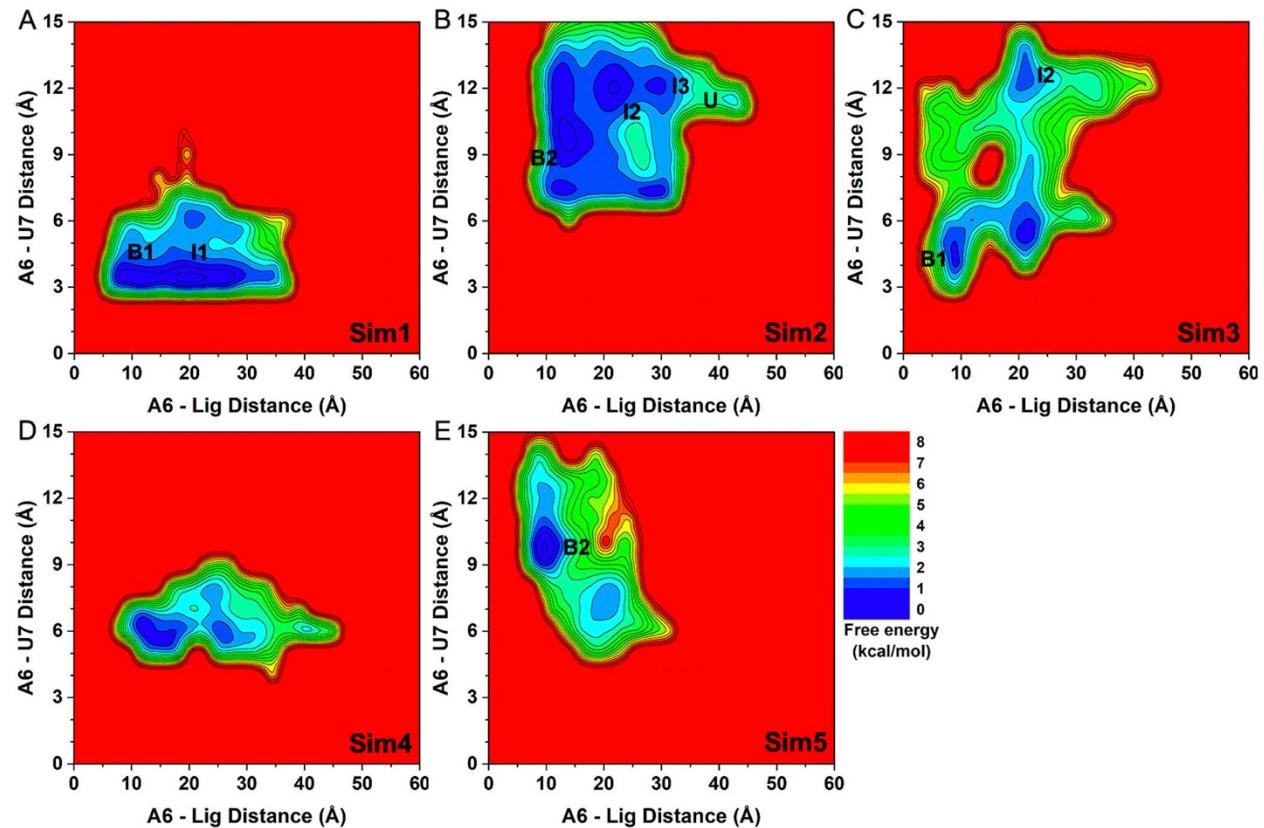
**Figure S4.** (A-C) 2D free energy profiles of the heavy-atom RMSD of CUUG relative to the 1RNG PDB structure and the COM distance between nucleotides C3 and G6 calculated from three independent 3,000-5,000 ns GaMD simulations of the CUUG RNA tetraloop. The low-energy RNA conformational states are labeled “Folded”, “I1”, and “Unfolded”. (D) Time courses of the heavy-atom RMSD of CUUG relative to the 1RNG PDB structure calculated from three independent 3,000-5,000 ns GaMD simulations of the CUUG RNA tetraloop. (E) Time courses of the COM distance between nucleotides C3 and G6 calculated from three independent 3,000-5,000 ns GaMD simulations of the CUUG RNA tetraloop.



**Figure S5.** **(A)** Time courses of the COM distance between the rbt203 ligand and RNA nucleotide A6 calculated from five independent 500 ns GaMD simulations of the rbt203 binding to the HIV-1 Tar RNA. **(B)** Time courses of the COM distance between RNA nucleotides A6 and U7 side chains calculated from five independent 500 ns GaMD simulations of the rbt203 binding to the HIV-1 Tar RNA.



**Figure S6.** 2D free energy profiles of the COM distance between the rbt203 ligand (Lig) and RNA nucleotide A6 and the COM distance between RNA nucleotides A6 and U7 side chains calculated from five independent 500 ns GaMD simulations of the rbt203 binding to the HIV-1 Tar RNA. The low-energy conformational states are labeled “B1”, “B2”, “I1”, “I2”, “I3”, and “U”.



**Figure S7.** Graphical representation of the GaMD algorithm as implemented in OpenMM. The algorithm is divided into the “conventional MD” and “Gaussian accelerated MD” portions, both of which are further divided into a total of five stages.

Stage 1: Conventional MD preparatory stage: no statistics are collected to allow the system to equilibrate.

Stage 2: Conventional MD stage: boost parameters  $V_{\max}$ ,  $V_{\min}$ ,  $V_{\text{avg}}$ , and  $\sigma_v$  are collected.

Stage 3: GaMD pre-equilibration stage: boost potential is applied, but boost parameters are not updated.

Stage 4: GaMD equilibration stage: boost potential is applied, and boost parameters are updated.

Stage 5: GaMD production stage: boost potential is applied, boost parameters are held fixed.

