Supporting Information

Mechanism of RNA recognition by a Musashi RNA-binding protein

Jinan Wang[1,2], Lan Lan[2], Xiaoqing Wu[2], Liang Xu[2,3], Yinglong Miao[1,2,*]

[1]Center for Computational Biology, University of Kansas, Lawrence, KS 66047, USA

[2] Department of Molecular Biosciences, University of Kansas, Lawrence, KS 66047, USA

[3]Department of Radiation Oncology, the University of Kansas Cancer Center, Kansas City, KS 66160, USA

[*]Email: miao@ku.edu

AMBER input file for GaMD equilibration:

****

GaMD equilibration simulation

```
 &cntrl
   imin=0,        ! No minimization
   irest=0,       ! This IS a new MD simulation
   ntx=1,         ! generate new velocity
   ! Temperature control
   ntt=3,         ! Langevin dynamics
   gamma_ln=1.0,  ! Friction coefficient (ps^-1)
   tempi=300.0,   ! Initial temperature
   temp0=300.0,   ! Target temperature
   ig=-1,         ! random seed
   ! Potential energy control
   cut=9.0,       ! nonbonded cutoff, in Angstroms
   ! MD settings
   nstlim=12600000, ! simulation length
   dt=0.002,      ! time step (ps)
   ! SHAKE
   ntc=2,         ! Constrain bonds containing hydrogen
   ntf=2,         ! Do not calculate forces of bonds containing hydrogen
   ! Control how often information is printed
   ntpr=1000,     ! Print energies every 1000 steps
   ntwx=1000,     ! Print coordinates every 1000 steps to the trajectory
   ntwr=5000,    ! Print a restart file every 5K steps
   ! Wrap coordinates when printing them to the same unit cell
   iwrap=1,
   ! Constant pressure control. Note that ntp=3 requires barostat=1
   barostat=1,    ! Berendsen... change to 2 for MC barostat
   ntp=1,         ! 1=isotropic, 2=anisotropic, 3=semi-isotropic w/ surften
```

```
  pres0=1.0,     ! Target external pressure, in bar
  taup=0.5,      ! Berendsen coupling constant (ps)
  ! GaMD parameters
  igamd = 3, iE = 1, irest_gamd = 0,
  ntcmd = 600000, nteb = 12000000, ntave = 120000,
  ntcmdprep = 240000, ntebprep = 240000,
  sigma0P = 6.0, sigma0D = 6.0,
 /
```

****

AMBER input file for GaMD production simulation:

****

```
GaMD restart simulation
 &cntrl
  imin=0,        ! No minimization
  irest=0,       ! This IS a new MD simulation
  ntx=1,         ! generate new velocity
  ! Temperature control
  ntt=3,         ! Langevin dynamics
  gamma_ln=1.0,  ! Friction coefficient (ps^-1)
  tempi=300.0,   ! Initial temperature
  temp0=300.0,   ! Target temperature
  ig=-1,         ! random seed
  ! Potential energy control
  cut=9.0,       ! nonbonded cutoff, in Angstroms
  ! MD settings
  nstlim=600000000, ! simulation length
  dt=0.002,      ! time step (ps)
  ! SHAKE
  ntc=2,         ! Constrain bonds containing hydrogen
```

```
  ntf=2,         ! Do not calculate forces of bonds containing hydrogen
  ! Control how often information is printed
  ntpr=200,     ! Print energies every 200 steps
  ntwx=200,     ! Print coordinates every 200 steps to the trajectory
  ntwr=5000,    ! Print a restart file every 5K steps
  ! Wrap coordinates when printing them to the same unit cell
  iwrap=1,
  ! Constant pressure control. Note that ntp=3 requires barostat=1
  barostat=1,    ! Berendsen... change to 2 for MC barostat
  ntp=1,         ! 1=isotropic, 2=anisotropic, 3=semi-isotropic w/ surften
  pres0=1.0,     ! Target external pressure, in bar
  taup=0.5,      ! Berendsen coupling constant (ps)
  ! GaMD parameters
  igamd = 3, iE = 1, irest_gamd = 1,
  ntcmd = 0, nteb = 0, ntave = 120000,
  ntcmdprep = 0, ntebprep = 0,
  sigma0P = 6.0, sigma0D = 6.0,
 /
********************************************************************************
```

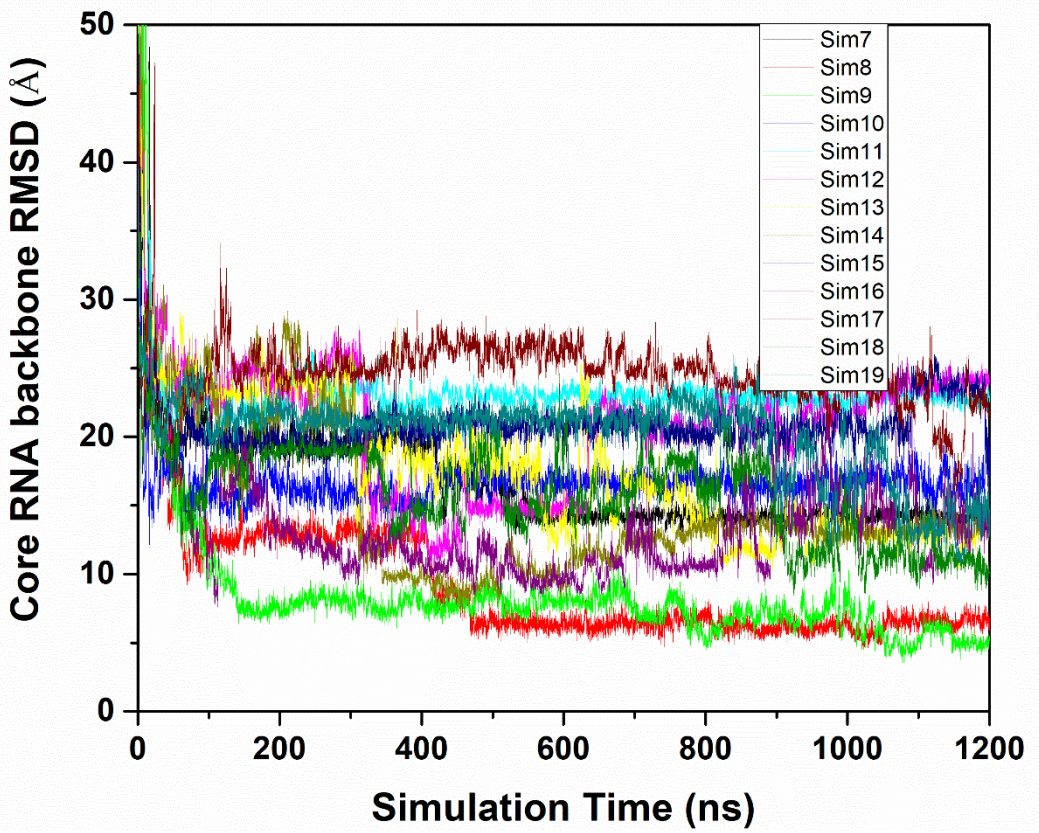

Figure S1. Time courses of the backbone RMSDs of core RNA (central three nucleotides GUA of the Numb RNA) relative to the first NMR conformation (PDB: 2RS2) are calculated from the 13 GaMD simulations (Sim7-Sim19) started from the Unbound state of the MSI1-Numb system.

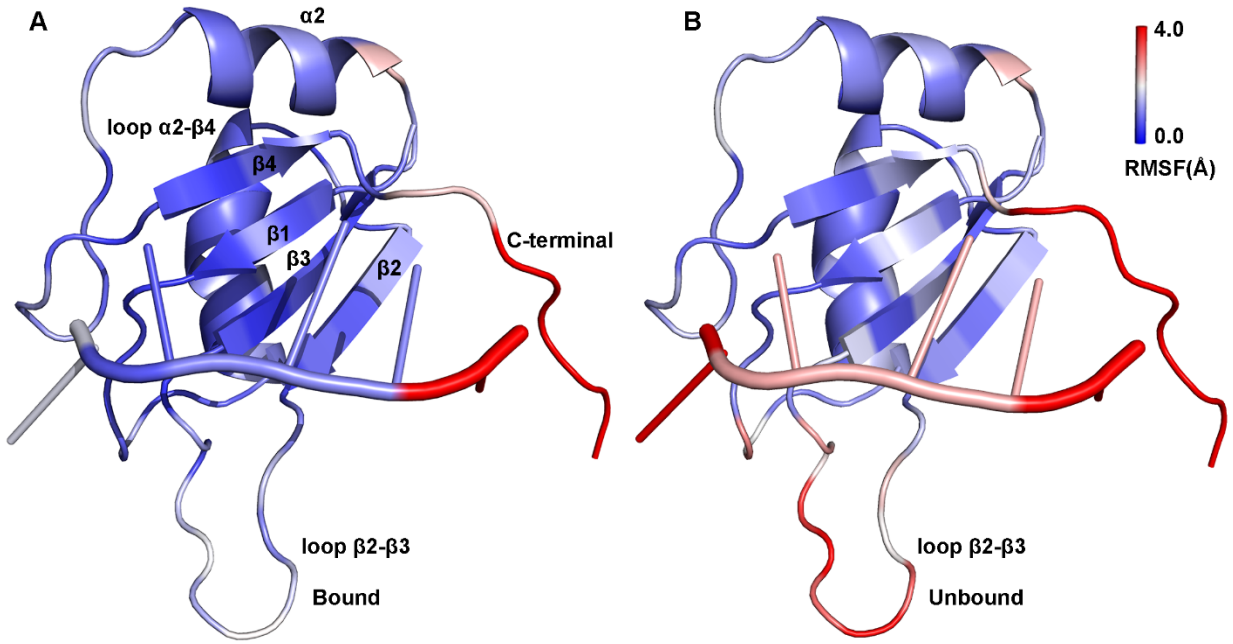

Figure S2. Structural flexibility of MSI1-RNA obtained from GaMD simulations: root-mean-square fluctuations (RMSFs) of the MSI1-RNA complex when the Numb RNA was initially placed in the (**A**) Bound state and (**B**) Unbound state. Only the 6 successful binding trajectories started from the Unbound state were used to calculate the RMSF in **B**. A color scale of 0.0 Å (blue) to 4.0 Å (red) is used for the RMSF.

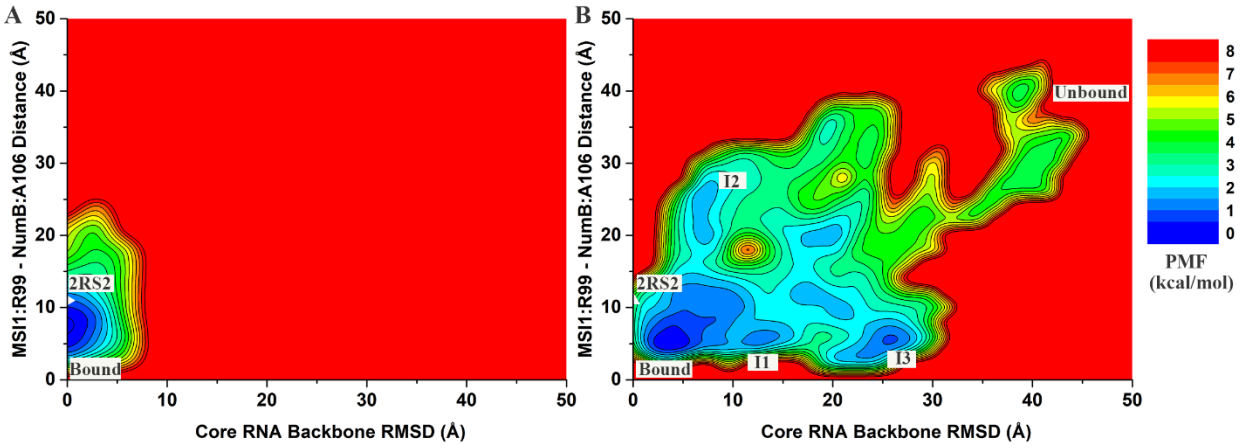

Figure S3. 2D potential of mean force (PMF) free energy profiles of the core RNA backbone RMSD relative to the first NMR conformation of in the PDB structure (ID: 2RS2) with alignment of MSI1 residues 21 to 95 and the distance between MSI1 residue R99 and Numb residue A106 are calculated from GaMD simulations started from the (**A**) Bound and (**B**) Unbound states of the MSI1-Numb system.

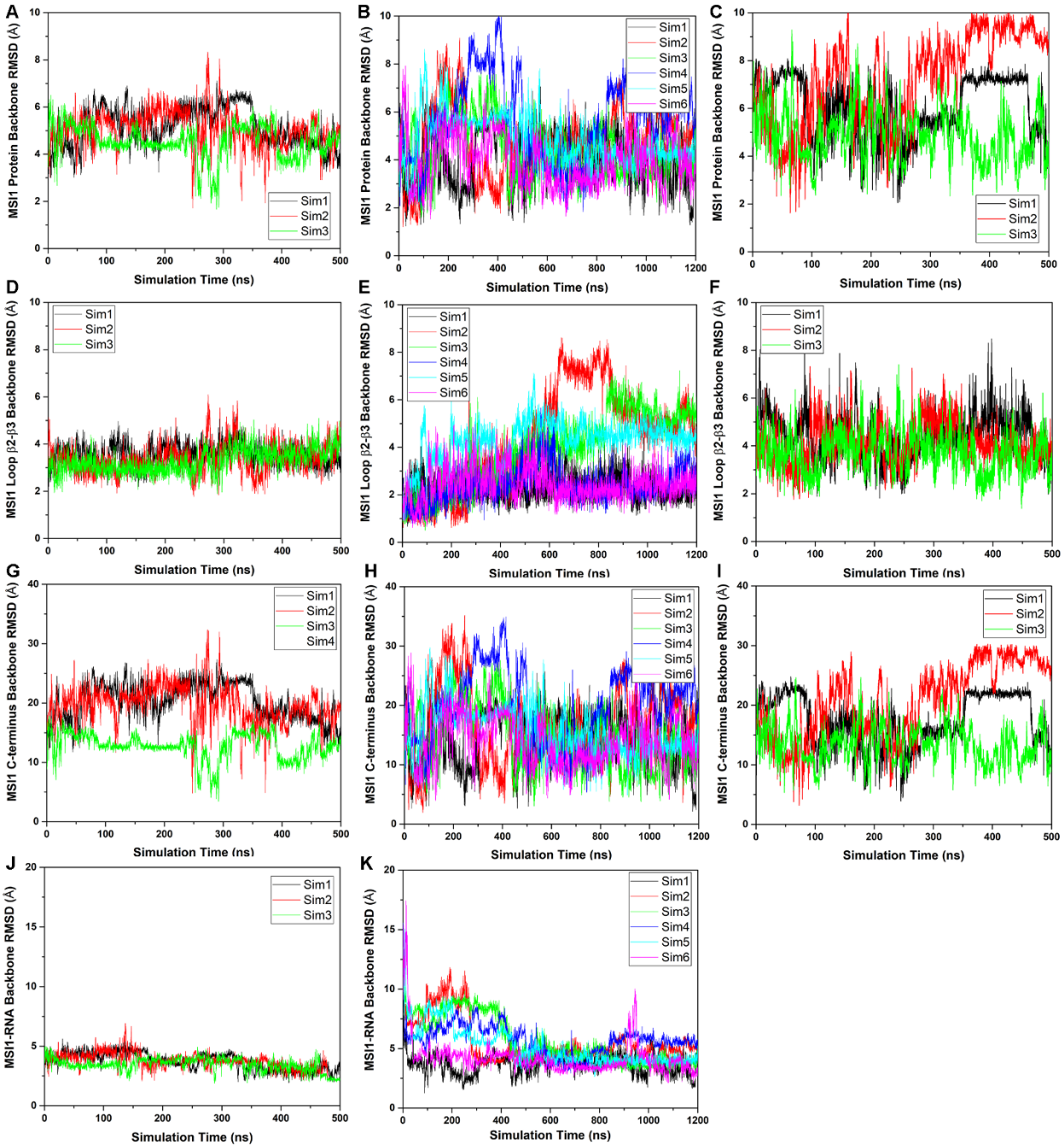

Figure S4. Time courses of the RMSDs of different MSI1 protein regions in GaMD simulations relative to the first NMR conformation in PDB (ID: 2RS2) with the MSI1 protein residues 21 to 95 aligned: **(A-C)** RMSDs of MSI1 backbone in simulations started from the (**A**) Bound, (**B**) Unbound and (**C**) *apo* states. **(D-F)** RMSDs of MSI1 β2-β3 loop backbone in simulations started from the (**D**) Bound, (**E**) Unbound and (**F**) *apo* states. **(G-I)** RMSDs of MSI1 C-terminus (residue 96 to 103) backbone in simulations started from the (**G**) Bound, (**H**) Unbound and (**I**) *apo* states. (**J-K**)RMSDs of MSI1-RNA complex backbone in simulations started from the (**J**) Bound and (**K**) Unbound states. For simulations started from the Unbound state, the time courses are plotted for only the 6 successful binding trajectories here.

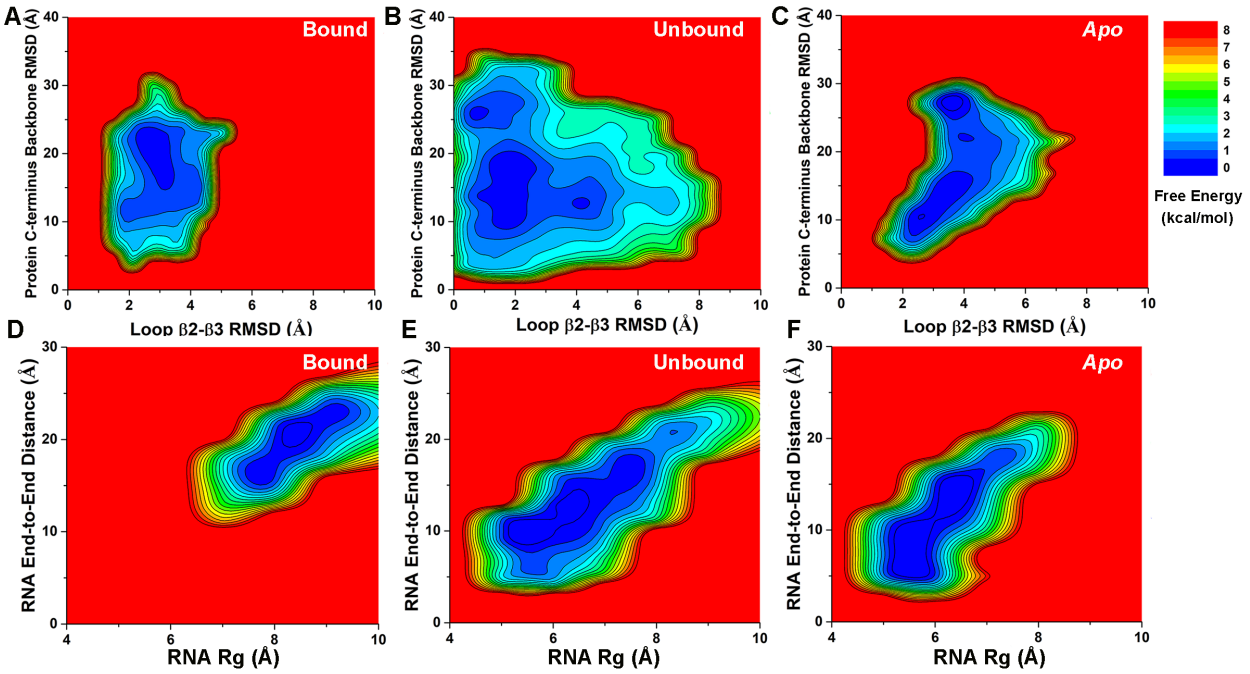

Figure S5. (A-C) 2D PMF free energy profiles of the MSI1 loop β2-β3 loop backbone RMSD and MSI1 C-terminus backbone RMSD relative to the first NMR conformation (PDB: 2RS2) are calculated from GaMD simulations started from the (**A**) Bound and (**B**) Unbound states of MSI1-Numb system and (**C**) *apo* MSI1. (**D-F**) 2D PMF free energy profiles of the radius of gyration (R_g) of and the end-to-end distance of the Numb RNA are calculated from GaMD simulations started from the (**D**) Bound and (**E**) Unbound states of the MSI1-Numb system and (**F**) the *apo* state of Numb RNA system.

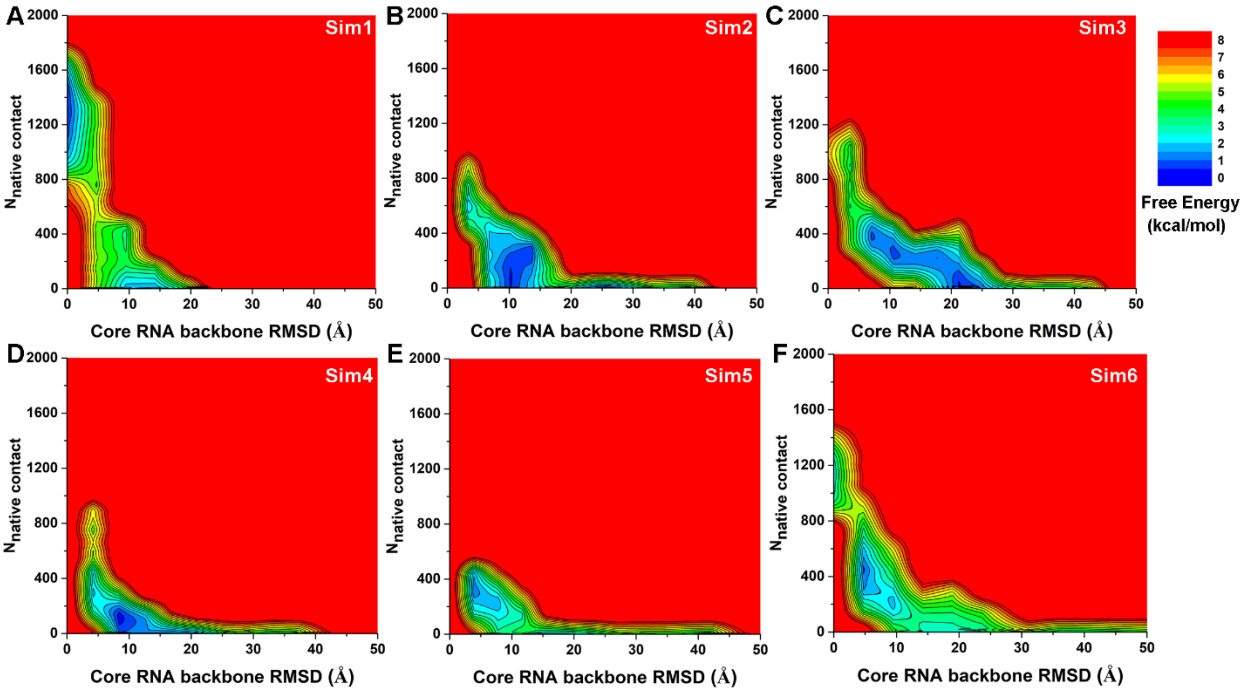

Figure S6. 2D potential of mean force (PMF) free energy profiles of the core RNA backbone RMSD relative to the first NMR conformation (PDB: 2RS2) and number of native contacts between MSI1 and Numb RNA are calculated from individual (**A**) Sim1, (**B**) Sim2, (**C**) Sim3, (**D**) Sim4, (**E**) Sim5 and (**F**) Sim6 GaMD simulations started from the Unbound state of the MSI1-Numb system.