

Mechanism of RNA recognition by a Musashi RNA-binding protein

Jinan Wang^{1,2}, Lan Lan², Xiaoqing Wu², Liang Xu^{2,3}, Yinglong Miao^{1,2,*}

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

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

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

*Email: miao@ku.edu

Abstract

The Musashi RNA-binding proteins (RBPs) regulate translation of target mRNAs and maintenance of cell stemness and tumorigenesis. Musashi-1 (MSI1), long considered as an intestinal and neural stem cell marker, has been more recently found to be overexpressed in many cancers. It has served as an important drug target for treating acute myeloid leukemia and solid tumors such as ovarian, colorectal and bladder cancer. One of the reported binding targets of MSI1 is Numb, a negative regulator of the Notch signaling. However, the dynamic mechanism of Numb RNA binding to MSI1 remains unknown, largely hindering effective drug design targeting this critical interaction. Here, we have performed all-atom simulations using a robust Gaussian accelerated molecular dynamics (GaMD) method, which successfully captured spontaneous and highly accurate binding of the Numb RNA from bulk solvent to the MSI1 protein target site. GaMD simulations revealed that Numb binding to MSI1 involved largely induced fit in both the RNA and protein. The simulations also identified important low-energy intermediate conformational states during RNA binding, in which Numb interacted mainly with the β 2- β 3 loop and C terminus of MSI1. The mechanistic understanding of RNA binding obtained from our GaMD simulations is expected to facilitate rational structure-based drug design targeting MSI1 and other RBPs.

Introduction

Protein-RNA interactions play crucial roles in various cellular activities and their dysfunction leads to a wide range of human diseases¹⁻³. Identification of small molecules that modulate interactions between RNA-binding proteins (RBPs) and RNA is progressing rapidly. It represents a novel strategy for discovery of drugs with new mechanisms¹. The Musashi (MSI) RBPs have been shown to regulate translation of target mRNAs and participate in the maintenance of cell stemness and tumorigenesis. They have been suggested as potential drug targets for treating many types of human cancer, including acute myeloid leukemia, ovarian cancer, colorectal cancer and bladder cancer⁴. The MSI protein family has two members: MSI1 and MSI2. Each MSI protein contains two N-terminal RNA recognition motifs (RRM1 and RRM2) that mediate the binding to their target mRNAs⁵. MSI1 binds to the 3'-untranslated region of Numb mRNA and represses its translation, which confers to the upregulation of Notch signaling. This leads to increased cell proliferation and survival, and decreased apoptosis of cancer cells⁴. Understanding the molecular mechanism of MSI1-Numb RNA interaction is important in both basic biology and applied medical research.

Rational design of small molecules targeting protein-RNA interactions requires structural characterizations of the RBP-RNA complexes. Due to high flexibility of MSI proteins and the lack of potent ligands, only a few MSI structures have been resolved so far, including the *apo* structure of MSI1/2-RRM1⁶⁻¹⁰ and RNA-bound structure of MSI1¹¹⁻¹². These structures have greatly facilitated structure-based modeling and drug design targeting the MSI-RNA interactions¹³⁻¹⁶. For example, we have recently identified one compound Aza-9 by combining fluorescence polarization (FP) assay, surface plasmon resonance (SPR), nuclear magnetic resonance (NMR) spectroscopy and molecular docking¹⁶. However, the experimental structures are rather static

images and the dynamic mechanism of MSI-RNA interactions remains unknown, which has largely hindered the development of potent inhibitors of MSI proteins.

Molecular dynamics (MD) is a powerful technique that enables all-atom simulations of biomolecules. MD simulations are able to fully account for the flexibility of the RBP and RNA during their interactions¹⁷⁻¹⁹. In 2015, Krepl et al.²⁰ provided systematic benchmarking data by simulating six structurally diverse protein/RNA complexes over multiple microsecond timescale MD runs and evaluating the simulations' stability. Their results suggested that the current force fields are able to handle microsecond MD simulations of protein/RNA complexes in many cases. For most systems, MD was possible to achieve a good but imperfect agreement with the experimental structure. However, MD could not maintain the initial experimental structure in one among six cases (3K5Y). The same group further presented a joint MD and NMR study to interpret and expand the available structural data of two RBPs bound with their single-stranded target RNAs²¹. They collected more than 50 μ s simulations and showed that the MD simulation was robust enough to reliably describe structural dynamics of RBP-RNA complexes.²¹ However, due to the slow dynamics and limited simulation timescale, it is rather challenging for conventional MD (cMD) simulations to sufficiently sample RBP-RNA interactions and obtain proper free energy profiles to quantitatively characterize RBP-RNA interactions.

To overcome limitations of cMD, enhanced sampling methods have been developed to improve biomolecular simulations²²⁻²⁸. Enhanced sampling methods have also been applied in studies of RBP-RNA interactions, including the steered MD²⁹, umbrella sampling²⁹ and metadynamics³⁰⁻³². Nevertheless, these enhanced sampling methods require predefined collective variables and may introduce constraints on the conformational space of the proteins. In this context, Gaussian accelerated MD (GaMD) has been developed to allow for unconstrained enhanced

sampling and free energy calculations of large biomolecules.^{28, 33-34} GaMD has been applied to successfully simulate protein folding and conformational changes³³⁻³⁸, ligand binding^{33-35, 39-44}, protein-protein/membrane/nucleic acid interactions⁴⁵⁻⁵⁰. More recently, GaMD simulations have successfully captured spontaneous binding of RNA to the human respiratory syncytial virus M2-1 protein⁵¹.

In this study, we have performed all-atom enhanced sampling simulations using GaMD on MS11-Numb RNA interactions started from the NMR structure in the bound state¹¹, as well as the unbound state with the Numb RNA moved far away from the MS11 protein surface (**Table 1**). While the NMR structure was found to maintain the bound conformation of MS11-Numb during the GaMD simulations, further simulations captured complete binding of the Numb RNA to the MS11 protein. The simulations thus allowed us to characterize structural flexibility and free energy landscapes of the MS11-Numb RNA interactions, which provided important insights into the mechanism of RNA recognition by the MS11 RBP.

Methods

Gaussian accelerated molecular dynamics (GaMD). GaMD enhances the conformational sampling of biomolecules by adding a harmonic boost potential to reduce the system energy barriers.³⁵ When the system potential $V(\vec{r})$ is lower than a reference energy E , the modified potential $V^*(\vec{r})$ of the system is calculated as:

$$V^*(\vec{r}) = V(\vec{r}) + \Delta V(\vec{r})$$
$$\Delta V(\vec{r}) = \begin{cases} \frac{1}{2} k(E - V(\vec{r}))^2, & V(\vec{r}) < E \\ 0, & V(\vec{r}) \geq E, \end{cases} \quad (1)$$

Where k is the harmonic force constant. The two adjustable parameters E and k are automatically determined on three enhanced sampling principles. First, for any two arbitrary potential

values $v_1(\vec{r})$ and $v_2(\vec{r})$ found on the original energy surface, if $V_1(\vec{r}) < V_2(\vec{r})$, ΔV should be a monotonic function that does not change the relative order of the biased potential values; i.e., $V_1^*(\vec{r}) < V_2^*(\vec{r})$. Second, if $V_1(\vec{r}) < V_2(\vec{r})$, the potential difference observed on the smoothed energy surface should be smaller than that of the original; i.e., $V_2^*(\vec{r}) - V_1^*(\vec{r}) < V_2(\vec{r}) - V_1(\vec{r})$. By combining the first two criteria and plugging in the formula of $V^*(\vec{r})$ and ΔV , we obtain

$$V_{max} \leq E \leq V_{min} + \frac{1}{k}, \quad (2)$$

Where V_{min} and V_{max} are the system minimum and maximum potential energies. To ensure that **Eq. 2** is valid, k has to satisfy: $k \leq 1/(V_{max} - V_{min})$. Let us define: $k = k_0 \cdot 1/(V_{max} - V_{min})$, then $0 < k_0 \leq 1$. Third, the standard deviation (SD) of ΔV needs to be small enough (i.e. narrow distribution) to ensure accurate reweighting using cumulant expansion to the second order: $\sigma_{\Delta V} = k(E - V_{avg})\sigma_V \leq \sigma_0$, where V_{avg} and σ_V are the average and SD of ΔV with σ_0 as a user-specified upper limit (e.g., $10k_B T$) for accurate reweighting. When E is set to the lower bound $E = V_{max}$ according to **Eq. 2**, k_0 can be calculated as

$$k_0 = \min(1.0, k'_0) = \min\left(1.0, \frac{\sigma_0}{\sigma_V} \cdot \frac{V_{max} - V_{min}}{V_{max} - V_{avg}}\right), \quad (3)$$

Alternatively, when the threshold energy E is set to its upper bound $E = V_{min} + 1/k$, k_0 is set to:

$$k_0 = k''_0 \equiv \left(1 - \frac{\sigma_0}{\sigma_V}\right) \cdot \frac{V_{max} - V_{min}}{V_{avg} - V_{min}}, \quad (4)$$

If k''_0 is calculated between 0 and 1. Otherwise, k_0 is calculated using **Eq. 3**.

Energetic Reweighting of GaMD Simulations. For energetic reweighting of GaMD simulations to calculate potential of mean force (PMF), the probability distribution along a reaction coordinate is written as $p^*(A)$. Given the boost potential $\Delta V(r)$ of each frame, $p^*(A)$ can be reweighted to recover the canonical ensemble distribution $p(A)$, as:

$$p(A_j) = p^*(A_j) \frac{\langle e^{\beta \Delta V(r)} \rangle_j}{\sum_{i=1}^M \langle p^*(A_i) e^{\beta \Delta V(r)} \rangle_i}, \quad j = 1, \dots, M, \quad (5)$$

where M is the number of bins, $\beta = k_B T$ and $\langle e^{\beta \Delta V(r)} \rangle_j$ is the ensemble-averaged Boltzmann factor of $\Delta V(r)$ for simulation frames found in the j^{th} bin. The ensemble-averaged reweighting factor can be approximated using cumulant expansion:

$$\langle e^{\beta \Delta V(r)} \rangle = \exp \left\{ \sum_{k=1}^{\infty} \frac{\beta^k}{k!} C_k \right\}, \quad (6)$$

where the first two cumulants are given by:

$$\begin{aligned} C_1 &= \langle \Delta V \rangle, \\ C_2 &= \langle \Delta V^2 \rangle - \langle \Delta V \rangle^2 = \sigma_v^2. \end{aligned} \quad (7)$$

The boost potential obtained from GaMD simulations usually follows near-Gaussian distribution³⁴. Cumulant expansion to the second order thus provides a good approximation for computing the reweighting factor^{35,52}. The reweighted free energy $F(A) = -k_B T \ln p(A)$ is calculated as:

$$F(A) = F^*(A) - \sum_{k=1}^2 \frac{\beta^k}{k!} C_k + F_c, \quad (8)$$

where $F^*(A) = -k_B T \ln p^*(A)$ is the modified free energy obtained from GaMD simulation and F_c is a constant.

System setup. Two models were prepared for simulations of MSI1-RNA interactions. One was obtained from the first model of NMR structure of the Numb RNA-bound MSI1 protein (PDB: 2RS2¹¹, denoted as the “Bound” state). Another one was generated with the first model by moving the Numb RNA ~ 30 Å away from its binding site in MSI1 (denoted as the “Unbound” state). The two systems were solvated in explicit water using *tleap* in the AMBER 20 package⁵³. The system charges were then neutralized at 0.15M NaCl using *tleap*. The AMBER ff14SBonlysc force field parameters⁵⁴, RNA.LJbb⁵⁵ and TIP3P model⁵⁶ were applied for the protein, RNA and water molecules, respectively. Each system was minimized using steepest descent for 50,000 steps and

conjugate gradient for another 50,000 steps. After minimization, the system was heated from 0 to 300 K in 1 ns simulation by applying 1 kcal/(mol·Å²) harmonic position restraints to the protein and RNA heavy atoms with a constant number, volume and temperature (NVT) ensemble. Each system was further equilibrated using a constant number, pressure and temperature (NPT) ensemble at 1 atm and 300 K for 1ns with same restraints as in the NVT run. Another 1.2 ns cMD simulations were performed to collect potential energy statistics (including the maximum, minimum, average and standard deviation). Then 24 ns GaMD equilibration after applying the boost potential was performed. Finally, three independent 500 ns and four independent 1,200 ns GaMD production runs with randomized initial atomic velocities were performed on the bound and unbound MSI1-Numb systems, respectively (**Table 1**). Simulation frames were saved every 0.4 ps for analysis.

Simulation Analysis. CPPTRAJ⁵⁷ and VMD⁵⁸ were used to analyze the GaMD simulations. Important reaction coordinates were identified from the simulation trajectories such that they involved dynamic regions (e.g., the β 2- β 3 loop of MSI1) and could be used to differentiate conformational states of the MSI1-Numb system. Therefore, root-mean-square deviations (RMSDs) of the backbone of core RNA (central three nucleotides UAG in Numb) and the β 2- β 3 loop of MSI1 relative to the NMR structure, the number of native contacts between MSI1 and Numb RNA ($N_{contacts}$), the radius of gyration (R_g) and end-to-end distance of the Numb RNA were selected as reaction coordinates. Root-mean-square fluctuations (RMSFs) were calculated for the protein residues and RNA nucleotides, averaged over the GaMD production simulations and color coded for schematic representation of each system. Since only the Sim1 and Sim2 GaMD trajectories successfully captured complete binding of the Numb RNA to MSI1, these trajectories

were used separately for structural clustering to identify the RNA binding pathways using the hierarchical agglomerative algorithm in CPPTRAJ⁵⁷. The RMSD cutoff was set to 3.0 Å for the core RNA backbone to form a cluster.

The PyReweighting⁵² toolkit was applied to reweight GaMD simulations to recover the original free energy or PMF profiles of the two MSI1-Numb systems. All GaMD production simulations from the Unbound state (4,800 ns in total) and Bound state (1,500 ns in total) were combined for calculating the PMF profiles, respectively. A bin size of 1.0 Å was used for the core RNA backbone RMSD, the MSI1 β 2- β 3 loop backbone RMSD, the Numb R_g and the end-to-end distance of Numb. A bin size of 100 was used for $N_{contacts}$. The cutoff was set to 500 frames for 2D PMF calculations.

Results

GaMD simulations captured complete binding of the Numb RNA to the MSI1 protein. Three independent 500 ns and four independent 1,200 ns GaMD simulations were performed on the MSI1-Numb system started from the Bound and Unbound states (**Table 1**). The GaMD simulations started from the Bound state recorded average and SD of the boost potential as 14.08 kcal/mol and 3.79 kcal/mol, respectively (**Table 1**). The GaMD simulations started from the Unbound state showed similar average and SD of boost potential with 11.53 kcal/mol and 3.36 kcal/mol, respectively (**Table 1**). The Bound MSI1-Numb complex was found to maintain the NMR structure with <3 Å RMSD of the core RNA backbone during most of the GaMD simulations (**Fig. 1A**). In the GaMD simulations started from the Unbound state, the core RNA backbone RMSD relative to the NMR structure decreased to minimum values 0.70 Å and 1.78 Å in Sim1 and Sim2, respectively, suggesting that complete binding of the Numb RNA from free diffusion in the solvent to the MSI1 target site was successfully captured (**Fig. 1b**). Spontaneous binding of

RNA was observed in the Sim1 after ~100 ns with the RNA RMSD decreased to ~2.50 Å relative to the NMR structure (**Fig. 1B**). In Sim2, the Numb RNA bound to MSI1 during ~1,010-1,130 ns and then dissociated to the solvent (**Fig. 1B**). The RNA binding events captured in the present GaMD simulations allowed us to characterize the dynamic interactions between the MSI1 protein and Numb RNA.

Variations of structural flexibility upon MSI1-Numb RNA binding. We analyzed structural flexibility of both the MSI1 protein and Numb RNA in the GaMD simulations. During GaMD simulations started from the Bound NMR structure, the MSI1 protein underwent small fluctuations except the loop connecting $\beta 2$ and $\beta 3$ strands (the $\beta 2$ - $\beta 3$ loop) and the C terminus (**Fig. 2A**). The fifth nucleotide in the Numb RNA exhibited significantly higher flexibility than the other nucleotides, especially the central three ones UAG (denoted as the core RNA). This suggested that interactions between the core RNA and the MSI1 were strong. Thus, the core RNA might play an important role in the interactions between the MSI1 protein and Numb RNA. In compared to the GaMD simulations started from the Bound state, the Numb RNA exhibited much higher flexibility in the simulations started from the Unbound state. For the MSI1 protein, both the $\beta 2$ - $\beta 3$ loop and C-terminus exhibited higher flexibility (**Fig. 2B**). These motifs were suggested to be important for binding of the Numb RNA¹¹ and small molecules¹⁴ to the MSI1 protein.

Free energy profiles of RNA binding to the MSI1 protein. Free energy profiles were calculated from the GaMD simulations using the core RNA backbone RMSD relative to the NMR structure and the number of native contacts between MSI1 and Numb RNA ($N_{contacts}$) as reaction coordinates. Only one low-energy minimum of the “Bound” conformation was identified from the

GaMD simulations on the NMR structure, in which the core RNA backbone RMSD and $N_{contacts}$ centered around (1.2 Å, 1600) (**Fig. 3A**). Five low-energy minima were identified from GaMD simulations started from the Unbound state, including the “Bound”, “Intermediate I1”, “Intermediate I2”, “Intermediate I3” and “Unbound” states, in which the core RNA backbone RMSD and $N_{contacts}$ centered around (2.0 Å, 1500), (5.2 Å, 600), (12.5 Å, 200), (25.0 Å, 10) and (40 Å, 0), respectively (**Fig. 3B**). The intermediate I1, I2 and I3 conformational states are shown in **Figs. 4**. Remarkably, positively charged residues (Arg61 and Arg99) in the β 2- β 3 loop and C-terminal region of the MSI1 protein formed favorable salt-bridge and hydrogen bond interactions with the central nucleotide A106 of Numb RNA. In the intermediate I1 state, the Numb RNA formed interactions with both the β 2- β 3 loop and C terminus of MSI1, leading to large conformational changes of these two regions (**Fig. 4A**). Notably, the sidechain of residue Arg99 in the C terminus of the MSI1 protein formed three hydrogen bonds with the sidechain of nucleotide A106 in the Numb RNA (**Fig. 4A**). In the intermediate I2 state, the sidechain of residue Arg61 in MSI1 could flip out to the solvent, forming a hydrogen bond and a salt-bridge with the sidechain and backbone (oxygen atom in the phosphate group) of the nucleotide A106 in Numb RNA, respectively (**Fig. 4B**). In the intermediate I3 state, Arg99 in the C terminus of MSI1 formed a hydrogen bond and a salt-bridge with sidechain and backbone of the nucleotide A106 in Numb RNA, respectively, for which a large conformational change of the protein C terminus was observed (**Fig. 4C**). Thus, these hydrogen bonds and salt-bridge interactions played a significant role in the recognition and binding of the Numb RNA to MSI1 protein. Since the charged residues Arg61 and Arg99 were identified to interact with the Numb RNA via hydrogen bond and long-range salt-bridge interactions in the intermediate states, we assumed that the Numb RNA binding to the MSI1 protein was mainly mediated by electrostatic interactions.

As described above, binding of the Numb RNA induced higher flexibility of the MSI1 β 2- β 3 loop (**Fig. 1B**) and large conformational change of the same region was observed in the intermediate I1 state (**Fig. 4A**). Therefore, the MSI1 β 2- β 3 loop backbone RMSD and core Numb RNA backbone RMSD relative to the experimental structure were used as reaction coordinates to further compute 2D free energy profiles (**Figs. 5**). The MSI1 β 2- β 3 loop was highly flexible, sampling a large conformational space with the backbone RMSD ranging from ~ 0 Å to ~ 8.0 Å (**Fig. 5B**). This loop sampled two distinct low-energy conformations, including the “Closed” (bound) (RMSD < 1 Å) and “Open” (free) states (RMSD ~ 3 -5 Å) (**Fig. 5B**). Compared to the “Open” state, the MSI1 β 2- β 3 loop moved closer to the core domain in the “Closed” state (**Fig. 4B**). Five low-energy states were identified from GaMD simulations starting with the Unbound state, including the “Unbound/Open”, “Intermediate I3/Open”, “Intermediate I2/Open”, “Intermediate I1/Closed” and “Bound/Open” states, in which the MSI1 β 2- β 3 loop backbone RMSD and core RNA backbone RMSD were located around (1.0 Å, 40 Å), (1.2 Å, 25.0 Å), (1.1 Å, 12.5 Å), (4.5 Å, 7.5 Å) and (1.5 Å, 2.0 Å), respectively (**Fig. 5B**). The Numb RNA and MSI1 β 2- β 3 loop accommodated to each other to form the final bound conformation (**Fig. 5B**), suggesting an “induced fit” mechanism in the MSI1-Numb RNA interaction.

Pathways of RNA binding to the MSI1 protein. Next, structural clustering was performed separately on the GaMD trajectories of Sim1 (movie S1 in Supporting Information) and Sim2 (movie S2 in Supporting Information) to identify the representative binding pathways of the Numb RNA to the MSI1 protein. The structural clusters were reweighted to obtain their original free energy values, which ranged from 0.0 kcal/mol to ~ 4.5 kcal/mol. The top reweighted clusters with PMF ≤ 2.0 kcal/mol were selected to represent the pathways of the Numb RNA binding to MSI1

(**Fig. 6**). In Sim1, the Numb RNA bound to MSI1 via interactions with the protein C terminus (**Fig. 6A**). In Sim2, both the β 2- β 3 loop and C terminus of MSI1 contributed important interactions with the Numb RNA during its binding to the protein target site (**Fig. 6B**). These findings again revealed important roles of the β 2- β 3 loop and C terminus of MSI1 in binding of the Numb RNA. This was consistent with the above RMSF analysis that higher flexibilities were observed in these two dynamic regions of MSI1 (**Fig. 2**).

The Numb RNA bound to the MSI1 protein via an induced fit mechanism. In order to further explore the mechanism of RNA binding to the MSI1 protein, we computed free energy profiles to characterize conformational changes of the Numb RNA upon binding to MSI1. In this regard, the radius of gyration (R_g) and the end-to-end distance of Numb were calculated to monitor its possible conformational changes. We used the R_g and end-to-end distance of the Numb RNA and the core RNA backbone RMSDs as reaction coordinates to calculate 2D PMF profiles (**Figs. 7A-7D**). Notably, the Numb RNA sampled a large conformational space during binding to the MSI1 protein in the GaMD simulations started from the Unbound state (**Figs. 7B**). From the reweighted 2D PMF profiles, we identified a similar “Bound” low-energy well in simulations started from both the Bound and Unbound states, for which the Numb RNA R_g and core RNA backbone RMSD centered around (2.0 Å, 8.5 Å) and (2.5 Å, 8.2 Å), respectively (**Fig. 7A-7B**). Another four low-energy states were identified in GaMD simulations started from Unbound conformation, including the “Unbound”, “Intermediate I1”, “Intermediate I2” and “Intermediate I3”, for which the core RNA backbone RMSD and R_g of Numb centered around (40.0 Å, 6.2 Å), (5.7 Å, 6.5 Å), (5.7 Å, 12.5 Å), respectively (**Fig. 7B**).

Furthermore, we calculated 2D PMF profiles regarding the core RNA backbone RMSD and the end-to-end distance of Numb RNA (**Figs. 7C-7D**). Only one low-energy state (“Bound”) was identified in the 2D PMF profile calculated from the GaMD simulations of the bound NMR structure, in which the Numb RNA adopted primarily the “Extended” conformation (**Fig. 7C**). The Numb end-to-end distance and core RNA backbone RMSD centered around (22.5 Å, 1.8 Å) (**Fig. 7C**). In contrast, five distinct low-energy states were identified from the 2D PMF profile in the GaMD simulations started from the Unbound conformation, including the “Unbound”, “Intermediate I1”, “Intermediate I2”, “Intermediate I3” and “Bound”, in which the Numb RNA adopted primarily the “Curled”, “Curled”, “Curled”, “Curled”, and “Extended” conformations, respectively (**Figs. 7D and 4**). The end-to-end distance of Numb RNA and core RNA backbone RMSD centered around (12.5 Å, 40.0 Å) in the “Unbound/Curled” state, (14.0 Å, 25.0 Å) in the “Intermediate I3/Curled” state, (18.5 Å, 10.5 Å) in the “Intermediate I2/Curled” state, (9.5 Å, 7.5 Å) in the “Intermediate I1/Curled” state and finally (20.0 Å, 2.5 Å) in the “Bound/Extended” state (**Fig. 7D**).

In comparison, the Numb RNA sampled a larger conformational space with a wider range of R_g or end-to-end distance in the “Bound” state than in the “Unbound” and even the three Intermediate conformational states (**Figs. 7B and 7D**). Therefore, binding of the Numb RNA to the MSI1 protein involved largely induced fit.

Discussion

In this study, we have applied all-atom GaMD simulations to investigate dynamic interactions between the Numb RNA and MSI1 protein. The GaMD simulations unprecedentedly captured spontaneous and highly accurate binding of the Numb RNA from bulk solvent to the MSI1 protein

with <2 Å RMSD in the core RNA backbone compared with the experimental structure. Proper energetic reweighting of the GaMD simulations allowed us to calculate PMF free energy profiles to characterize the MSI1-Numb binding process.

Important low-energy conformational states were identified from the GaMD derived free energy profiles. Particularly, in the I1, I2 and I3 intermediate states, a significant role of the β 2- β 3 loop and C terminus of the MSI1 protein was revealed in the recognition and binding of the Numb RNA (**Fig. 4**). The charged residues Arg61 and Arg99 of MSI1 that formed hydrogen bond and salt-bridge interactions with the Numb RNA were observed in the GaMD simulations, supporting important roles of electrostatic interactions in forming the “Intermediate” and “Bound” states of the MSI1-Numb RNA complex. The important role of Arg99 was consistent with the result that the mutation of the same residue in its homology MSI2 (Arg100Ala) decreased its binding affinity of Numb¹⁰. The salt-bridge between Arg61 (MSI1 protein) and A106 (Numb RNA) was also observed in the bound NMR structure (**Figs. 4B**). In contrast, our GaMD simulations here captured this interaction during the binding process of the Numb RNA, supporting the important role of electrostatic interactions in forming both the “Intermediate” and “Bound” states. Furthermore, Arg61 was characterized as important residue for the inhibitor binding as the MSI1 Arg61Glu mutant decreased ~ 5 fold for the binding of inhibitor¹⁴. Therefore, GaMD simulations showed that long-range electrostatic interactions played an important role in the Numb RNA binding to the MSI1 protein and identified two critical residues (Arg61 and Arg99) in the binding process. Additionally, RMSF analysis suggested that the binding between the core RNA and the MSI1 protein was highly stable in the GaMD simulations starting from the Bound state, suggesting that the core RNA might play an important role in the interactions between the MSI1 protein and the Numb RNA. This result is consistent with the result obtained by Zearfoss et al.⁵⁹ that the central

core RNA (UAG) forms the core MSI recognition element and makes major contributions to binding affinity.

Conformation selection⁶⁰⁻⁶² and induced fit⁶³⁻⁶⁴ are two common models for describing biomolecular recognition. In this context, our GaMD simulations have revealed that binding of the Numb RNA to the MSI1 protein involved predominantly an induced fit mechanism, in which both the RNA and protein (especially the β 2- β 3 loop) underwent significant conformational changes during binding. This is consistent with previous studies of other protein-RNA interactions⁶⁴⁻⁶⁷, including ribosomal protein S15-rRNA and U1A-RNA complexes. A major conformational change of the rRNA was found upon binding to the S15 protein through comparison of the free and bound structures of S15 and rRNA, suggesting induced fit of the protein and RNA⁶⁴. MD simulations combined with available structure analysis also indicated that binding of the U1A protein and RNA followed an induced fit mechanism⁶⁷. For the U1A protein, MD simulations indicated that induced fit upon binding involved a non-native thermodynamic substate, in which the structure is preorganized for binding. In contrast, induced fit of the RNA involved a distortion of the native structure to an unstable form in solution.

It is worth noting that the presented free energy profiles were not converged since only two binding events were observed in the GaMD simulations. More binding and unbinding events would still need to be simulated in order to quantitatively characterize RNA binding to the MSI1 protein and calculate the binding free energies and kinetic rates. In this regard, our recently developed selective GaMD algorithms⁶⁸⁻⁶⁹ could be useful to address the challenge. In particular, microsecond simulations using the peptide GaMD (Pep-GaMD) method⁶⁹, which works by selectively boosting the peptide potential energy, have been demonstrated to capture repetitive binding and unbinding of highly flexible peptides to the target protein⁶⁹. Apart from enhanced

conformational sampling, accurate force fields are also needed especially for the RNA⁷⁰⁻⁷³ in order to simulate repetitive RNA dissociation and binding to RBPs. Even the force field works well individually for the protein and RNA, combination of protein and RNA force fields in the MD simulations could introduce additional challenges.^{18, 20, 70} Nevertheless, we have observed two complete binding events of the Numb RNA to the MSI1 protein with our current force field settings and GaMD simulations, which will guide our future simulation studies of RBPs-RNA interactions.

Conclusions

In summary, all-atom GaMD simulations with unconstrained enhanced sampling and free energy calculations have provided important insights into the mechanism of the Numb RNA binding to the MSI1 protein. For future studies, the effects of small molecule binding in the MSI1-Numb interactions still need to be determined and our simulation findings await validation in the wet-lab experiments. Further studies are planned to simulate both dissociation and binding of RNA to the RBPs and accurately predict the thermodynamics and kinetics of protein–RNA interactions. These efforts are expected to greatly facilitate rational drug design targeting the MSI1 and other RBPs.

Author Contributions

Y.M. designed research. J.W. performed the research. J.W., L. L, X.W., L. X and Y.M. analyzed the data, and J.W. and Y.M. wrote the paper.

Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgements

This work was supported in part by the National Institutes of Health (R01GM132572) and the startup funding in the College of Liberal Arts and Sciences at the University of Kansas (to Y.M.), NIH R01 CA191785 and KU Cancer Center Pilot Project Award (to L.X.). This work used supercomputing resources with allocation Award No. TG-MCB180049 through the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by the National Science Foundation, Grant No. ACI-1548562, and Project No. M2874 through the National Energy Research Scientific Computing Center (NERSC), which is a U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231, and the Research Computing Cluster at the University of Kansas.

Supplementary Material

The Supporting Information is available free of charge on the ACS Publications website. Movies S1 and S2 showing GaMD trajectories of Sim1 and Sim2 starting from the Unbound state, respectively.

Reference

- (1) Wu, P., Inhibition of RNA-binding proteins with small molecules. *Nat. Rev. Chem.* **2020**, *4* (9), 441-458.
- (2) Hentze, M. W.; Castello, A.; Schwarzl, T.; Preiss, T., A brave new world of RNA-binding proteins. *Nat. Rev. Mol. Cell Biol.* **2018**, *19* (5), 327-341.
- (3) Conlon, E. G.; Manley, J. L., RNA-binding proteins in neurodegeneration: mechanisms in aggregate. *Genes. Dev.* **2017**, *31* (15), 1509-1528.
- (4) Kudinov, A. E.; Karanicolas, J.; Golemis, E. A.; Boumber, Y., Musashi RNA-Binding Proteins as Cancer Drivers and Novel Therapeutic Targets. *Clin. Cancer Res.* **2017**, *23* (9), 2143-2153.
- (5) Sakakibara, S.; Nakamura, Y.; Satoh, H.; Okano, H., Rna-binding protein Musashi2: developmentally regulated expression in neural precursor cells and subpopulations of neurons in mammalian CNS. *J. Neurosci.* **2001**, *21* (20), 8091-107.
- (6) Lan, L.; Xing, M.; Kashipathy, M.; Douglas, J.; Gao, P.; Battaile, K.; Hanzlik, R.; Lovell, S.; Xu, L., Crystal and solution structures of human oncoprotein Musashi-2 N-terminal RNA recognition motif 1. *Proteins* **2020**, *88* (4), 573-583.
- (7) Lan, L.; Xing, M.; Douglas, J. T.; Gao, P.; Hanzlik, R. P.; Xu, L., Human oncoprotein Musashi-2 N-terminal RNA recognition motif backbone assignment and identification of RNA-binding pocket. *Oncotarget* **2017**, *8* (63), 106587.
- (8) Nagata, T.; Kanno, R.; Kurihara, Y.; Uesugi, S.; Imai, T.; Sakakibara, S.-i.; Okano, H.; Katahira, M., Structure, backbone dynamics and interactions with RNA of the C-terminal RNA-binding domain of a mouse neural RNA-binding protein, Musashi1 | Edited by P. E. Wright. *J. Mol. Biol.* **1999**, *287* (2), 315-330.
- (9) Miyanoiri, Y.; Kobayashi, H.; Imai, T.; Watanabe, M.; Nagata, T.; Uesugi, S.; Okano, H.; Katahira, M., Origin of Higher Affinity to RNA of the N-terminal RNA-binding Domain than That of the C-terminal One of a Mouse Neural Protein, Musashi1, as Revealed by Comparison of Their Structures, Modes of Interaction, Surface Electrostatic Potentials, and Backbone Dynamics. *J. Biol. Chem.* **2003**, *278* (42), 41309-41315.
- (10) Minuesa, G.; Albanese, S. K.; Xie, W.; Kazansky, Y.; Worroll, D.; Chow, A.; Schurer, A.; Park, S.-M.; Rotsides, C. Z.; Taggart, J.; Rizzi, A.; Naden, L. N.; Chou, T.; Gourkanti, S.; Cappel, D.; Passarelli, M. C.; Fairchild, L.; Adura, C.; Glickman, J. F.; Schulman, J.; Famulare, C.; Patel, M.; Eibl, J. K.; Ross, G. M.; Bhattacharya, S.; Tan, D. S.; Leslie, C. S.; Beuming, T.; Patel, D. J.; Goldgur, Y.; Chodera, J. D.; Kharas, M. G., Small-molecule targeting of MUSASHI RNA-binding activity in acute myeloid leukemia. *Nat. Commun.* **2019**, *10* (1), 2691.
- (11) Ohyama, T.; Nagata, T.; Tsuda, K.; Kobayashi, N.; Imai, T.; Okano, H.; Yamazaki, T.; Katahira, M., Structure of Musashi1 in a complex with target RNA: the role of aromatic stacking interactions. *Nucleic Acids Res.* **2011**, *40* (7), 3218-3231.
- (12) Iwaoka, R.; Nagata, T.; Tsuda, K.; Imai, T.; Okano, H.; Kobayashi, N.; Katahira, M., Structural Insight into the Recognition of r(UAG) by Musashi-1 RBD2, and Construction of a Model of Musashi-1 RBD1-2 Bound to the Minimum Target RNA. *Molecules* **2017**, *22* (7), 1207.
- (13) Lan, L.; Appelman, C.; Smith, A. R.; Yu, J.; Larsen, S.; Marquez, R. T.; Liu, H.; Wu, X.; Gao, P.; Roy, A., Natural product (-)-gossypol inhibits colon cancer cell growth by targeting RNA-binding protein Musashi-1. *Mol. Oncol.* **2015**, *9* (7), 1406-1420.

- (14) Clingman, C. C.; Deveau, L. M.; Hay, S. A.; Genga, R. M.; Shandilya, S. M. D.; Massi, F.; Ryder, S. P., Allosteric inhibition of a stem cell RNA-binding protein by an intermediary metabolite. *eLife* **2014**, *3*, e02848.
- (15) Lan, L.; Liu, H.; Smith, A. R.; Appelman, C.; Yu, J.; Larsen, S.; Marquez, R. T.; Wu, X.; Liu, F. Y.; Gao, P.; Gowthaman, R.; Karanicolas, J.; De Guzman, R. N.; Rogers, S.; Aubé, J.; Neufeld, K. L.; Xu, L., Natural product derivative Gossypolone inhibits Musashi family of RNA-binding proteins. *BMC Cancer* **2018**, *18* (1), 809.
- (16) Lan, L.; Liu, J.; Xing, M.; Smith, A. R.; Wang, J.; Wu, X.; Appelman, C.; Li, K.; Roy, A.; Gowthaman, R.; Karanicolas, J.; Somoza, A. D.; Wang, C. C. C.; Miao, Y.; De Guzman, R.; Oakley, B. R.; Neufeld, K. L.; Xu, L., Identification and Validation of an *Aspergillus nidulans* Secondary Metabolite Derivative as an Inhibitor of the Musashi-RNA Interaction. *Cancers* **2020**, *12* (8), 2221.
- (17) Cheatham III, T. E.; Case, D. A., Twenty-five years of nucleic acid simulations. *Biopolymers* **2013**, *99* (12), 969-977.
- (18) Šponer, J.; Krepl, M.; Banáš, P.; Kührová, P.; Zgarbová, M.; Jurečka, P.; Havrila, M.; Otyepka, M., How to understand atomistic molecular dynamics simulations of RNA and protein-RNA complexes? *WIREs RNA* **2017**, *8* (3), e1405.
- (19) Borišek, J.; Saltalamacchia, A.; Galli, A.; Palermo, G.; Molteni, E.; Malcovati, L.; Magistrato, A., Disclosing the Impact of Carcinogenic SF3b Mutations on Pre-mRNA Recognition Via All-Atom Simulations. *Biomolecules* **2019**, *9* (10), 633.
- (20) Krepl, M.; Havrila, M.; Stadlbauer, P.; Banas, P.; Otyepka, M.; Pasulka, J.; Stefl, R.; Sponer, J., Can We Execute Stable Microsecond-Scale Atomistic Simulations of Protein-RNA Complexes? *J. Chem. Theory Comput.* **2015**, *11* (3), 1220-1243.
- (21) Krepl, M.; Cléry, A.; Blatter, M.; Allain, F. H. T.; Sponer, J., Synergy between NMR measurements and MD simulations of protein/RNA complexes: application to the RRM, the most common RNA recognition motifs. *Nucleic Acids Res.* **2016**, *44* (13), 6452-6470.
- (22) Mlýnský, V.; Bussi, G., Exploring RNA structure and dynamics through enhanced sampling simulations. *Curr. Opin. Struct. Biol.* **2018**, *49*, 63-71.
- (23) Abrams, C.; Bussi, G., Enhanced Sampling in Molecular Dynamics Using Metadynamics, Replica-Exchange, and Temperature-Acceleration. *Entropy-Switz* **2014**, *16* (1), 163-199.
- (24) Spiwok, V.; Sucer, Z.; Hosek, P., Enhanced sampling techniques in biomolecular simulations. *Biotechnol. Adv.* **2015**, *33* (6 Pt 2), 1130-1140.
- (25) Gao, Y. Q.; Yang, L. J.; Fan, Y. B.; Shao, Q., Thermodynamics and kinetics simulations of multi-time-scale processes for complex systems. *Int Rev Phys Chem* **2008**, *27* (2), 201-227.
- (26) Liwo, A.; Czaplowski, C.; Oldziej, S.; Scheraga, H. A., Computational techniques for efficient conformational sampling of proteins. *Curr. Opin. Struct. Biol.* **2008**, *18* (2), 134-139.
- (27) Christen, M.; van Gunsteren, W. F., On searching in, sampling of, and dynamically moving through conformational space of biomolecular systems: A review. *J. Comput. Chem.* **2008**, *29* (2), 157-166.
- (28) Miao, Y.; McCammon, J. A., Unconstrained Enhanced Sampling for Free Energy Calculations of Biomolecules: A Review. *Mol. Simul.* **2016**, *42* (13), 1046-1055.
- (29) Wu, S.; Wang, J.; Pu, X.; Li, L.; Li, Q., T7 RNA Polymerase Discriminates Correct and Incorrect Nucleoside Triphosphates by Free Energy. *Biophys. J.* **2018**, *114* (8), 1755-1761.

- (30) Laio, A.; Parrinello, M., Escaping free-energy minima. *Proc. Natl. Acad. Sci. USA* **2002**, *99* (20), 12562-12566.
- (31) Alessandro, L.; Francesco, L. G., Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Rep. Prog. Phys.* **2008**, *71* (12), 126601.
- (32) Borkar, A. N.; Bardaro, M. F.; Camilloni, C.; Aprile, F. A.; Varani, G.; Vendruscolo, M., Structure of a low-population binding intermediate in protein-RNA recognition. *Proc. Natl. Acad. Sci. USA* **2016**, *113* (26), 7171-7176.
- (33) Pang, Y. T.; Miao, Y.; Wang, Y.; McCammon, J. A., Gaussian Accelerated Molecular Dynamics in NAMD. *J. Chem. Theory Comput.* **2017**, *13* (1), 9-19.
- (34) Miao, Y.; McCammon, J. A., Gaussian Accelerated Molecular Dynamics: Theory, Implementation, and Applications. *Annu. Rep. Comput. Chem.* **2017**, *13*, 231-278.
- (35) Miao, Y.; Feher, V. A.; McCammon, J. A., Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. *J. Chem. Theory Comput.* **2015**, *11* (8), 3584-3595.
- (36) Peng, Y.; Cao, S.; Kiselar, J.; Xiao, X.; Du, Z.; Hsieh, A.; Ko, S.; Chen, Y.; Agrawal, P.; Zheng, W.; Shi, W.; Jiang, W.; Yang, L.; Chance, M. R.; Surewicz, W. K.; Buck, M.; Yang, S., A Metastable Contact and Structural Disorder in the Estrogen Receptor Transactivation Domain. *Structure* **2019**, *27* (2), 229-240 e4.
- (37) Brown, B. P.; Zhang, Y.-K.; Westover, D.; Yan, Y.; Qiao, H.; Huang, V.; Du, Z.; Smith, J. A.; Ross, J. S.; Miller, V. A.; Ali, S.; Bazhenova, L.; Schrock, A. B.; Meiler, J.; Lovly, C. M., On-target Resistance to the Mutant-Selective EGFR Inhibitor Osimertinib Can Develop in an Allele-Specific Manner Dependent on the Original EGFR-Activating Mutation. *Clin. Cancer Res.* **2019**, *25* (11), 3341-3351.
- (38) Salawu, E. O., The Impairment of TorsinA's Binding to and Interactions With Its Activator: An Atomistic Molecular Dynamics Study of Primary Dystonia. *Front. Mol. Biosci.* **2018**, *5* (64).
- (39) Miao, Y.; McCammon, J. A., Graded activation and free energy landscapes of a muscarinic G-protein-coupled receptor. *Proc. Natl. Acad. Sci. USA* **2016**, *113* (43), 12162-12167.
- (40) Wang, Y.-T.; Chan, Y.-H., Understanding the molecular basis of agonist/antagonist mechanism of human mu opioid receptor through gaussian accelerated molecular dynamics method. *Sci. Rep.* **2017**, *7* (1), 7828.
- (41) Liao, J.-M.; Wang, Y.-T., In silico studies of conformational dynamics of Mu opioid receptor performed using gaussian accelerated molecular dynamics. *J. Biomol. Struct. Dyn.* **2019**, *37* (1), 166-177.
- (42) Miao, Y.; Bhattarai, A.; Nguyen, A. T. N.; Christopoulos, A.; May, L. T., Structural Basis for Binding of Allosteric Drug Leads in the Adenosine A1 Receptor. *Sci. Rep.* **2018**, *8* (1), 16836.
- (43) Pawnikar, S.; Miao, Y., Pathway and mechanism of drug binding to chemokine receptors revealed by accelerated molecular simulations. *Future Med. Chem.* **2020**, *12* (13), 1213-1225.
- (44) Chuang, C. H.; Chiou, S. J.; Cheng, T. L.; Wang, Y. T., A molecular dynamics simulation study decodes the Zika virus NS5 methyltransferase bound to SAH and RNA analogue. *Sci. Rep.* **2018**, *8* (1), 6336.
- (45) Sibener, L. V.; Fernandes, R. A.; Kolawole, E. M.; Carbone, C. B.; Liu, F.; McAfee, D.; Birnbaum, M. E.; Yang, X.; Su, L. F.; Yu, W.; Dong, S.; Gee, M. H.; Jude, K. M.; Davis, M. M.;

- Groves, J. T.; Goddard, W. A., 3rd; Heath, J. R.; Evavold, B. D.; Vale, R. D.; Garcia, K. C., Isolation of a Structural Mechanism for Uncoupling T Cell Receptor Signaling from Peptide-MHC Binding. *Cell* **2018**, *174* (3), 672-687 e27.
- (46) Park, J. B.; Kim, Y. H.; Yoo, Y.; Kim, J.; Jun, S.-H.; Cho, J. W.; El Qaidi, S.; Walpole, S.; Monaco, S.; García-García, A. A.; Wu, M.; Hays, M. P.; Hurtado-Guerrero, R.; Angulo, J.; Hardwidge, P. R.; Shin, J.-S.; Cho, H.-S., Structural basis for arginine glycosylation of host substrates by bacterial effector proteins. *Nat. Commun.* **2018**, *9* (1), 4283.
- (47) Miao, Y.; McCammon, J. A., Mechanism of the G-protein mimetic nanobody binding to a muscarinic G-protein-coupled receptor. *Proc. Natl. Acad. Sci. USA* **2018**, *115* (12), 3036-3041.
- (48) Ricci, C. G.; Chen, J. S.; Miao, Y.; Jinek, M.; Doudna, J. A.; McCammon, J. A.; Palermo, G., Deciphering Off-Target Effects in CRISPR-Cas9 through Accelerated Molecular Dynamics. *ACS Cent. Sci.* **2019**, *5* (4), 651-662.
- (49) Palermo, G.; Miao, Y.; Walker, R. C.; Jinek, M.; McCammon, J. A., CRISPR-Cas9 conformational activation as elucidated from enhanced molecular simulations. *Proc. Natl. Acad. Sci. USA* **2017**, *114* (28), 7260-7265.
- (50) Bhattarai, A.; Wang, J.; Miao, Y., G-Protein-Coupled Receptor-Membrane Interactions Depend on the Receptor Activation State. *J. Comput. Chem.* **2020**, *41* (5), 460-471.
- (51) Gao, Y.; Cao, D.; Pawnikar, S.; John, K. P.; Ahn, H. M.; Hill, S.; Ha, J. M.; Parikh, P.; Ogilvie, C.; Swain, A.; Yang, A.; Bell, A.; Salazar, A.; Miao, Y.; Liang, B., Structure of the Human Respiratory Syncytial Virus M2-1 Protein in Complex with a Short Positive-Sense Gene-End RNA. *Structure* **2020**, *28* (9), 979-990.e4.
- (52) Miao, Y.; Sinko, W.; Pierce, L.; Bucher, D.; Walker, R. C.; McCammon, J. A., Improved Reweighting of Accelerated Molecular Dynamics Simulations for Free Energy Calculation. *J. Chem. Theory Comput.* **2014**, *10* (7), 2677-2689.
- (53) D.A. Case, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, D. Greene, N. Homeyer, S. Izadi, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. Mermelstein, K.M. Merz, G. Monard, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, C.L. Simmerling, W.M. Botello-Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, L. Xiao, D.M. York and P.A. Kollman (2020), AMBER 2020, University of California, San Francisco.
- (54) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696-3713.
- (55) Zgarbová, M.; Otyepka, M.; Šponer, J.; Mládek, A.; Banáš, P.; Cheatham, T. E.; Jurečka, P., Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J. Chem. Theory Comput.* **2011**, *7* (9), 2886-2902.
- (56) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79* (2), 926-935.
- (57) Roe, D. R.; Cheatham, T. E., 3rd, PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9* (7), 3084-95.
- (58) Humphrey, W.; Dalke, A.; Schulten, K., VMD: visual molecular dynamics. *J. Mol. Graph.* **1996**, *14* (1), 33-8, 27-8.

- (59) Zearfoss, N. R.; Deveau, L. M.; Clingman, C. C.; Schmidt, E.; Johnson, E. S.; Massi, F.; Ryder, S. P., A Conserved Three-nucleotide Core Motif Defines Musashi RNA Binding Specificity. *J. Biol. Chem.* **2014**, *289* (51), 35530-35541.
- (60) Tsai, C. J.; Kumar, S.; Ma, B.; Nussinov, R., Folding funnels, binding funnels, and protein function. *Protein Sci.* **1999**, *8* (6), 1181-90.
- (61) Tsai, C. J.; Ma, B.; Nussinov, R., Folding and binding cascades: shifts in energy landscapes. *Proc. Natl. Acad. Sci. USA* **1999**, *96* (18), 9970-2.
- (62) Kumar, S.; Ma, B.; Tsai, C. J.; Sinha, N.; Nussinov, R., Folding and binding cascades: dynamic landscapes and population shifts. *Protein Sci.* **2000**, *9* (1), 10-9.
- (63) Okazaki, K.-i.; Takada, S., Dynamic energy landscape view of coupled binding and protein conformational change: induced-fit versus population-shift mechanisms. *Proc. Natl. Acad. Sci. USA* **2008**, *105* (32), 11182-11187.
- (64) Williamson, J. R., Induced fit in RNA–protein recognition. *Nat. Struct. Biol.* **2000**, *7* (10), 834-837.
- (65) Leulliot, N.; Varani, G., Current Topics in RNA–Protein Recognition: Control of Specificity and Biological Function through Induced Fit and Conformational Capture. *Biochemistry* **2001**, *40* (27), 7947-7956.
- (66) Suryadi, J.; Tran, E. J.; Maxwell, E. S.; Brown, B. A., The Crystal Structure of the Methanocaldococcus jannaschii Multifunctional L7Ae RNA-Binding Protein Reveals an Induced-Fit Interaction with the Box C/D RNAs. *Biochemistry* **2005**, *44* (28), 9657-9672.
- (67) Pitici, F.; Beveridge, D. L.; Baranger, A. M., Molecular dynamics simulation studies of induced fit and conformational capture in U1A–RNA binding: Do molecular substates code for specificity? *Biopolymers* **2002**, *65* (6), 424-435.
- (68) Miao, Y.; Bhattarai, A.; Wang, J., Ligand Gaussian accelerated molecular dynamics (LiGaMD): Characterization of ligand binding thermodynamics and kinetics. *J. Chem. Theory Comput.* **2020**, *16* (9), 5526-5547.
- (69) Wang, J.; Miao, Y., Peptide Gaussian accelerated molecular dynamics (Pep-GaMD): Enhanced sampling and free energy and kinetics calculations of peptide binding. *J. Chem. Phys.* **2020**, *153* (15), 154109.
- (70) Šponer, J.; Bussi, G.; Krepl, M.; Banáš, P.; Bottaro, S.; Cunha, R. A.; Gil-Ley, A.; Pinamonti, G.; Poblete, S.; Jurečka, P.; Walter, N. G.; Otyepka, M., RNA Structural Dynamics As Captured by Molecular Simulations: A Comprehensive Overview. *Chem. Rev.* **2018**, *118* (8), 4177-4338.
- (71) Tan, D.; Piana, S.; Dirks, R. M.; Shaw, D. E., RNA force field with accuracy comparable to state-of-the-art protein force fields. *Proc. Natl. Acad. Sci. USA* **2018**, *115* (7), E1346-E1355.
- (72) Kührová, P.; Mlýnský, V.; Zgarbová, M.; Krepl, M.; Bussi, G.; Best, R. B.; Otyepka, M.; Šponer, J.; Banáš, P., Improving the Performance of the Amber RNA Force Field by Tuning the Hydrogen-Bonding Interactions. *J. Chem. Theory Comput.* **2019**, *15* (5), 3288-3305.
- (73) Cesari, A.; Bottaro, S.; Lindorff-Larsen, K.; Banáš, P.; Šponer, J.; Bussi, G., Fitting Corrections to an RNA Force Field Using Experimental Data. *J. Chem. Theory Comput.* **2019**, *15* (6), 3425-3431.

Table 1. Summary of GaMD simulations performed on the MSI1 protein with RNA Numb started from the Bound and Unbound states.

System	N_{atoms}	Length(ns)	Boost potential (kcal/mol)
MSI1-Numb (Bound)	29,368	500 x 3	14.08 ± 3.79
MSI1-Numb (Unbound)	29,569	1200 x 4	11.53 ± 3.36

Figure captions

Figure 1. Time courses of the backbone RMSDs of core RNA (central three nucleotides GUA of the Numb RNA) relative to the NMR structure (PDB: 2RS2) are calculated from the GaMD simulations started from the (A) Bound and (B) Unbound states of the MSI1-Numb system.

Figure 2. 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 or (B) Unbound state. A color scale of 0.0 Å (blue) to 4.0 Å (red) is used.

Figure 3. 2D potential of mean force (PMF) free energy profiles of the core RNA backbone RMSD relative to the NMR structure (PDB: 2RS2) and number of native contacts between MSI1 and Numb RNA are calculated from GaMD simulations started from the (A) Bound and (B) Unbound states of the MSI1-Numb system.

Figure 4. Low-energy intermediate conformational states (I1, I2 and I3) as identified from the 2D PMF profile of the MSI1-RNA simulation system started from the Unbound state. The MSI1 protein and Numb RNA are shown in green and red, respectively. The NMR structure of the MSI1-Numb complex is shown in blue for comparison. MSI1 protein residues Arg99 and Arg61, and nucleotide A106 of Numb RNA are highlighted in balls and sticks. The hydrogen bonds between the side-chain of residues in the MSI1 protein and Numb RNA are shown in red. The salt-bridge interactions between the side-chain of residues in the MSI1 protein and backbone (oxygen atom) of the Numb RNA are shown in black.

Figure 5. 2D PMF profiles of the MSI1 β 2- β 3 loop backbone RMSD and core RNA backbone RMSD relative to the NMR structure (PDB: 2RS2) are calculated from GaMD simulations started from the (A) Bound and (B) Unbound states of the MSI1-Numb system.

Figure 6. Binding pathways of the Numb RNA to the MSI1 protein revealed from GaMD simulations: **(A)** Starting from diffusion in the solvent, the Numb RNA bound to the target site of the MSI1 via intermediate conformations by interacting with the MSI1 C-terminus in the “Sim1” GaMD trajectory. **(B)** Starting from diffusion in the solvent, the Numb RNA bound to the target site of the MSI1 protein via intermediate conformations by interacting with the MSI1 $\beta 2$ - $\beta 3$ loop and C-terminus in the “Sim2” GaMD trajectory. The MSI1 is shown in blue ribbons. The Numb structural clusters (sticks) are colored by the reweighted PMF free energy values in a green (0.0 kcal/mol)-white-red (2.0 kcal/mol) color scale.

Figure 7. **(A-B)** 2D PMF profiles of the radius of gyration (R_g) of the Numb RNA and core RNA backbone RMSD relative to the NMR structure (PDB: 2RS2) are calculated from GaMD simulations started from the **(A)** Bound and **(B)** Unbound states of the MSI1-Numb system. **(C-D)** 2D PMF profiles of the end-to-end distance of the Numb RNA and core RNA backbone RMSD relative to the NMR structure (PDB: 2RS2) are calculated from GaMD simulations started from the **(C)** Bound and **(D)** Unbound states of the MSI1-Numb system.

Figure 1

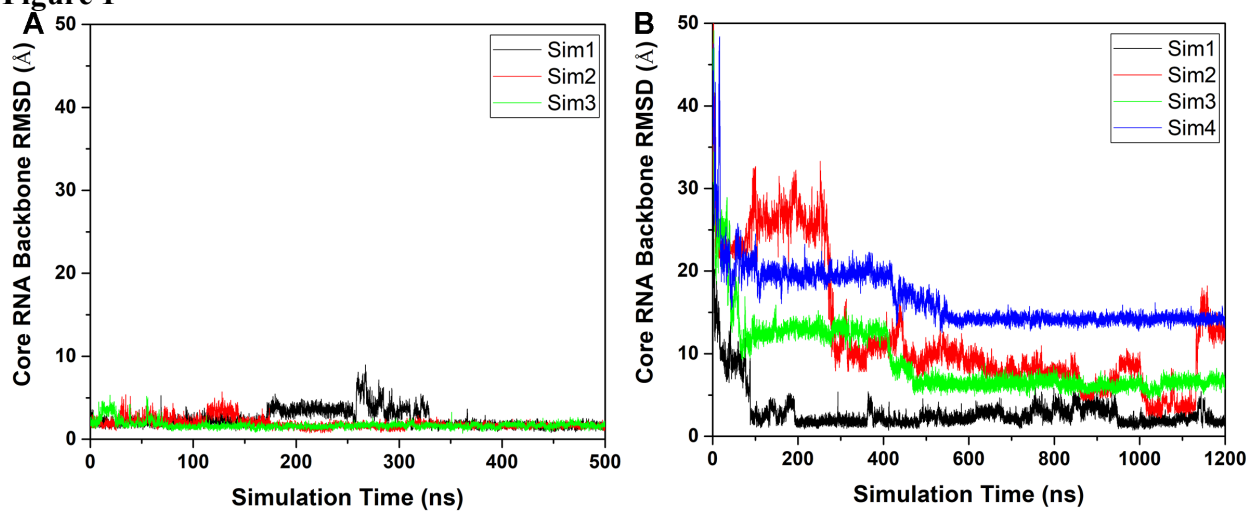


Figure 2

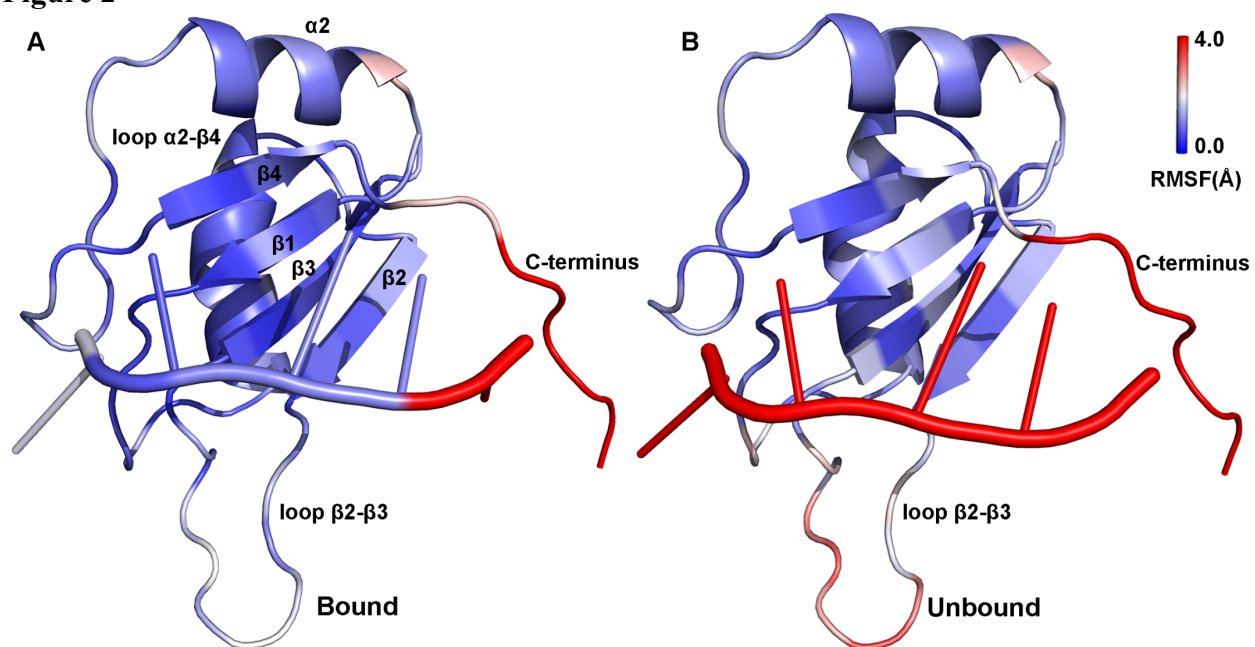


Figure 3

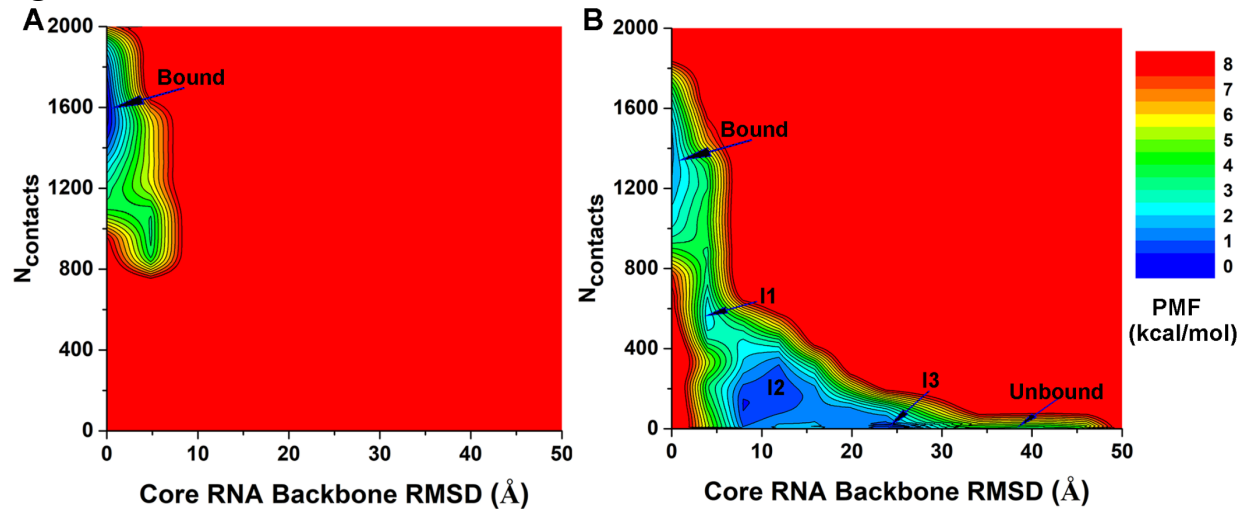


Figure 4

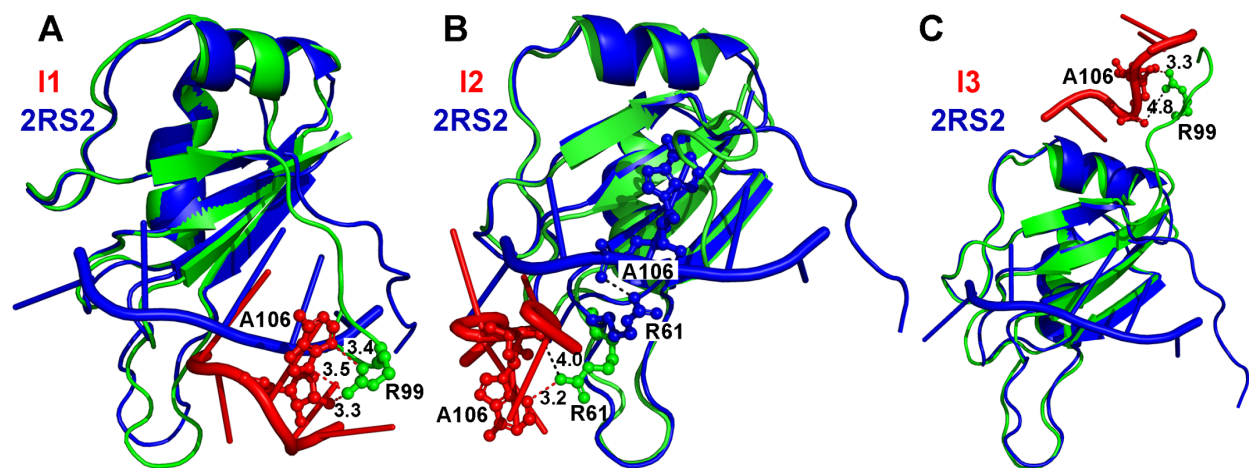


Figure 5

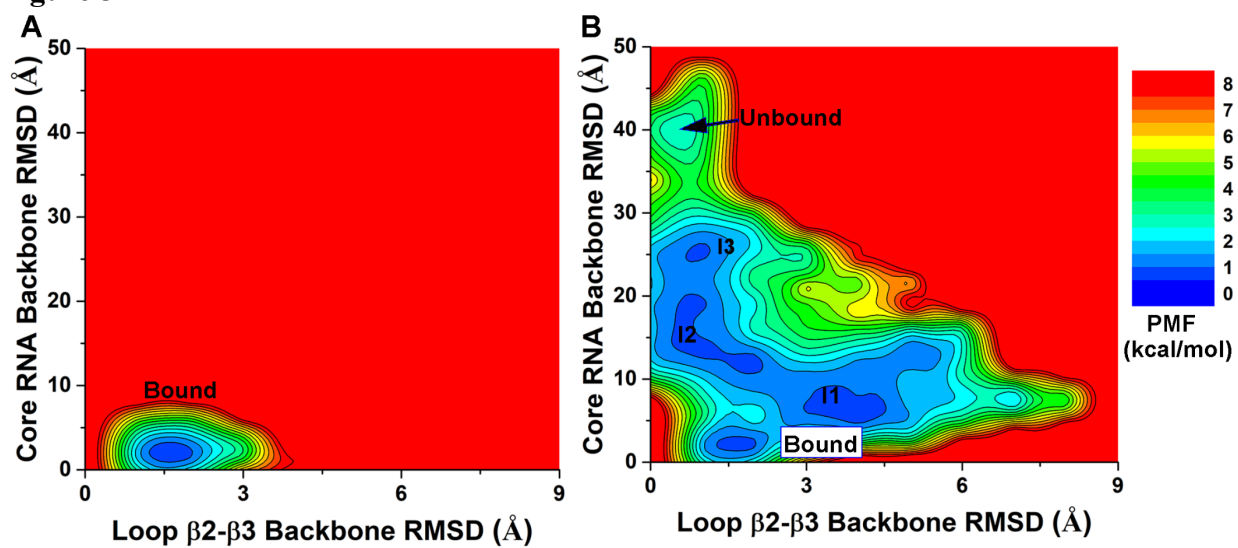


Figure 6

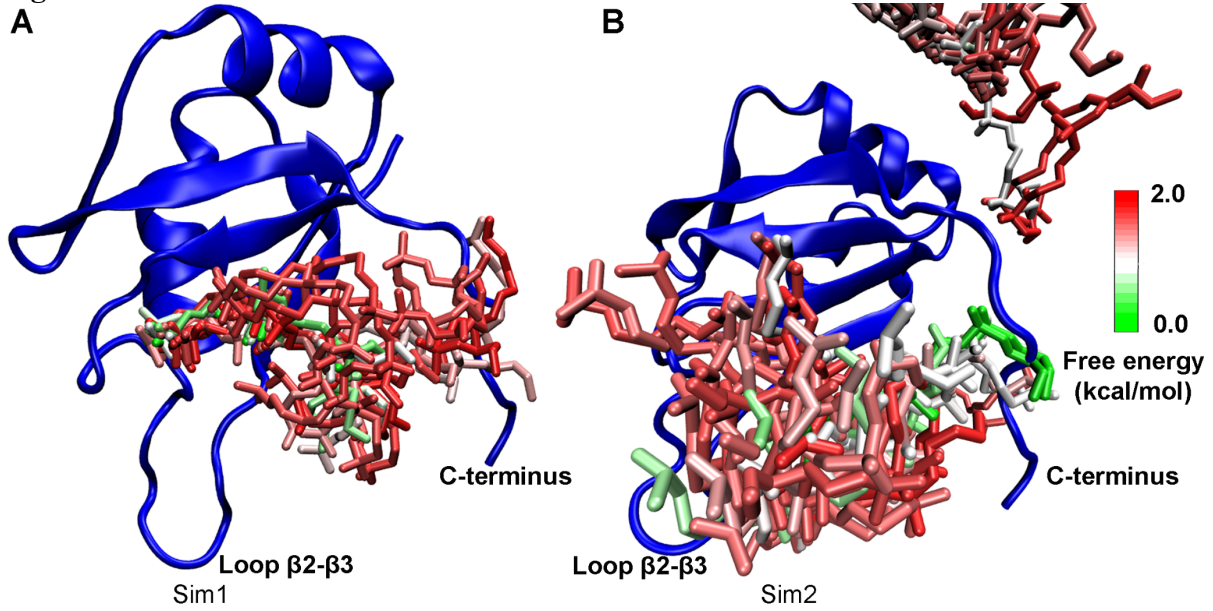


Figure 7

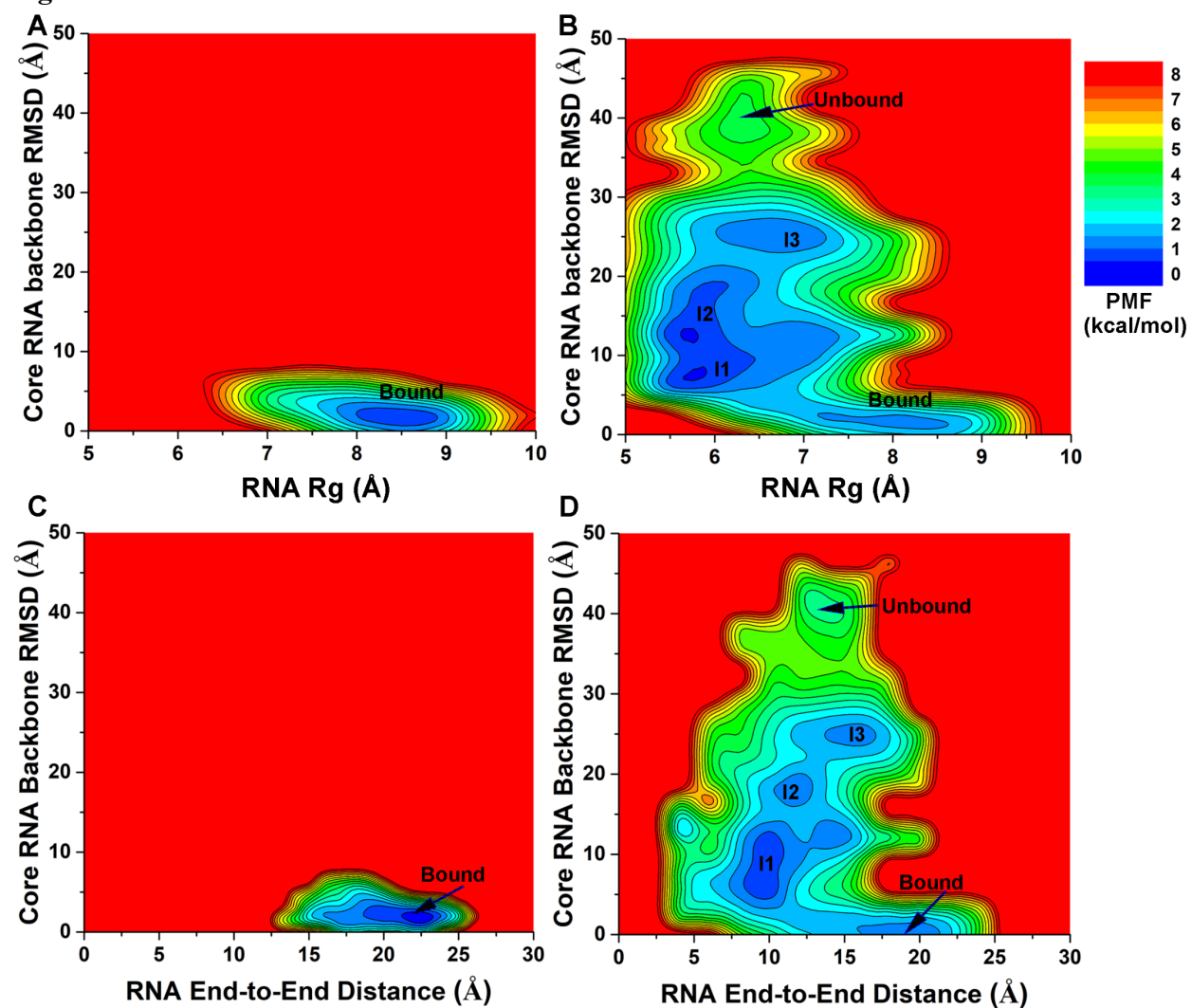
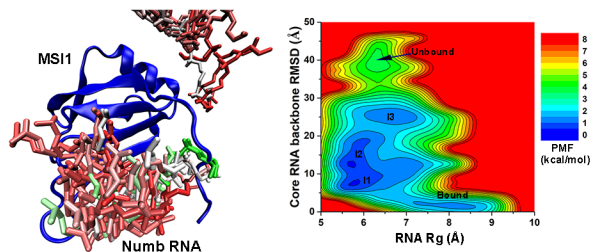


Table of content



Abstract Graphic All-atom simulations using a robust Gaussian accelerated molecular dynamics (GaMD) method successfully captured spontaneous binding of the Numb RNA to the Musashi-1 (MSI1) protein, which involved largely induced fit in both the RNA and protein. GaMD simulations also identified important low-energy intermediate conformational states of RNA binding. The simulation findings are expected to facilitate rational structure-based drug design targeting MSI1 and other RNA-binding proteins.