UC Berkeley
UC Berkeley Previously Published Works

Title

Dynamic docking of small molecules targeting RNA CUG repeats causing myotonic dystrophy type 1.

Permalink

Journal

Authors

Wang, Kye
Riveros, Ivan
DeLoye, James
et al.

Publication Date

DOI

Peer reviewed

Dynamic docking of small molecules targeting RNA CUG repeats causing myotonic dystrophy type 1

Kye Won Wang,[1,2] Ivan Riveros,[1] James DeLoye,[3] and Ilyas Yildirim[1,*]

[1]Department of Chemistry and Biochemistry, Florida Atlantic University, Jupiter, Florida; [2]Departments of Biological Sciences and Chemistry, Lehigh University, Bethlehem, Pennsylvania; and [3]Department of Chemistry, University of California, Berkeley, Berkeley, California

ABSTRACT   Expansion of RNA CUG repeats causes myotonic dystrophy type 1 (DM1). Once transcribed, the expanded CUG repeats strongly attract muscleblind-like 1 (MBNL1) proteins and disturb their functions in cells. Because of its unique structural form, expanded RNA CUG repeats are prospective drug targets, where small molecules can be utilized to target RNA CUG repeats to inhibit MBNL1 binding and ameliorate DM1-associated defects. In this contribution, we developed two physics-based dynamic docking approaches (DynaD and DynaD/Auto) and applied them to nine small molecules known to specifically target RNA CUG repeats. While DynaD uses a distance-based reaction coordinate to study the binding phenomenon, DynaD/Auto combines results of umbrella sampling calculations performed on 1×1 UU internal loops and AutoDock calculations to efficiently sample the energy landscape of binding. Predictions are compared with experimental data, displaying a positive correlation with correlation coefficient (R) values of 0.70 and 0.81 for DynaD and DynaD/Auto, respectively. Furthermore, we found that the best correlation was achieved with MM/3D-RISM calculations, highlighting the importance of solvation in binding calculations. Moreover, we detected that DynaD/Auto performed better than DynaD because of the use of prior knowledge about the binding site arising from umbrella sampling calculations. Finally, we developed dendrograms to present how bound states are connected to each other in a binding process. Results are exciting, as DynaD and DynaD/Auto will allow researchers to utilize two novel physics-based and computer-aided drug-design methodologies to perform in silico calculations on drug-like molecules aiming to target complex RNA loops.

SIGNIFICANCE   The importance of RNA as a new drug target has been recognized because of the discovery of a wide range of RNA-associated diseases. We developed a dynamic docking methodology (DynaD) and a modified version of DynaD (DynaD/Auto) to predict the bound states of small molecules targeting RNA loop motifs. DynaD is designed to be used when the binding site is identified but no other information is available. DynaD/Auto, on the other hand, combines AutoDock calculations with the energy landscape of the binding site predicted by umbrella sampling calculations. Finally, we created dendrograms to interpret large data sets arising from binding calculations and display how bound states are connected to each other.

INTRODUCTION

Myotonic dystrophy type 1 (DM1) is a genetic muscular dystrophy that affects skeletal and smooth muscles as well as many other organs (1,2). It is the most common cause of adult- and late-onset muscular dystrophy that can be characterized by muscle weakness, myotonia, and cardiac conduction abnormality, which cause physical disability and reduction of life expectancy (1,2). DM1 is a repeat expansion disorder due to the repetitive expansion of CTG repeats

in the DNA. Once transcribed, the expanded CUG repeats, located in the three-prime untranslated (3′ UTR) region of messenger RNA (mRNA), with underlined U (U) representing the loop residues, forms a hairpin structure, which attracts muscleblind-like 1 (MBNL1) protein (1,3–8). MBNL1 is a key regulator of alternative splicing of transcripts such as insulin receptor, cardiac troponin T, and muscle-specific chlorine ion channel (4,6,7,9,10). The transcribed CUG repeats have strong binding affinities to MBNL1, which disturbs MBNL1's function via sequestration (6,11,12). Healthy individuals have 5–37 CUG repeats in the DMPK gene, while DM1-affected individuals have 50–6000 CUG copies, where severity of disease increases with repeat size (13,14). Because of its unique structural

characteristics such as the formation of a hairpin structure with continuous 1 × 1 UU internal loops connected with 2 × 2 GC/GC Watson-Crick (WC) basepairs, expanded RNA CUG repeats are prospective drug targets (15,16). Thus, DM1-associated defects can be ameliorated by small molecules specifically targeting RNA CUG repeats to inhibit MBNL1 binding.

To design effective drug-like small molecules targeting RNA, it is necessary to discover the binding mechanism between RNA and small molecules. Computational techniques are promising, as they can virtually screen molecules and visualize bound states for drug developments. There are several examples of in silico molecular docking programs studying ligand-receptor interactions, such as LigandRNA (17), AutoDock Vina (18), Glide (19), AnnapuRNA (20), and RLDOCK (21). The goals of these programs are to predict binding affinity and the corresponding conformation between a small molecule and a macromolecular receptor. LigandRNA is a program designed for specific docking of small molecular ligands to RNA, and uses a grid-based algorithm combined with a knowledge-based potential representing ligand-binding sites. AutoDock is an automated docking software to predict how ligands bind to a binding site such as proteins, DNA, RNA, and other biomolecules. AutoDock uses its own scoring function based on the Amber force field, and estimates the free energy of binding of a ligand to its target. Glide is a ligand docking program developed by Schrödinger that allows small molecules to dock onto all kinds of biomolecular receptors including RNA, where binding affinities are predicted by free energy perturbation or molecular mechanics combined with generalized Born surface area continuum solvation (MM/GBSA) (22). AnnapuRNA is a knowledge-based tool designed to investigate RNA-ligand complexes. The main limitation of AnnapuRNA is that binding poses not in the database are difficult to predict, as the docking process relies on the experimentally determined RNA-ligand structures. RLDOCK is another computational method for predicting binding poses for RNA-ligand complexes with an energy-based scoring function, where RNA structure is kept fixed. The aforementioned methods share the assumption that docking is related to the global minimum conformation of a binding site obtained from experiments. Typically, the structure of the binding site is obtained from existing databases such as Protein Data Bank (PDB). However, since we do not know what the exact mechanism of the binding or docking process is, these methods have the disadvantage of excluding various possibilities derived from the intermediate bindings between small molecules and binding sites. To overcome such disadvantages, Al-Hashimi's group have introduced an ensemble including unique and dominant conformations across the entire RNA structure landscape guided by NMR data, and showed that docking of six small molecules targeting the RNA ensemble in HIV type 1 improves the predictions (23). Another approach to increase the accuracy of the binding pose is to perform a detailed ligand-recognition step. Grubmuller et al. successfully studied the detailed binding pathways of a biotin molecule to a streptavidin by a force-pulling method (24). However, computational binding of small molecules to RNA is more challenging than proteins because RNA molecules are highly dynamic. Guilbert and James presented a novel algorithm, MORDOR (Molecular Recognition with a Driven dynamics OptimizeR), which incorporates flexibility of both ligands and nucleic acid residues (25). However, if the binding site is already known, such as the case in RNA CUG repeats, focus should be given to the binding site to increase the accuracy of predictions. Another method to investigate RNA-ligand binding, supervised molecular dynamics (SuMD) simulations (26), uses short molecular dynamics (MD) simulations to derive binding poses of a ligand to an RNA-binding site. However, the efficiency may not always be high owing to the strong chance of restarting the simulations as a result of ligands drifting away from the binding site during the sampling process. Above all, these two methods require complicated settings to sample the molecular recognition events.

We developed a dynamic docking methodology (DynaD) and successfully applied it to several RNA systems causing myotonic dystrophy (15,16,27–29), amyotrophic lateral sclerosis/frontotemporal dementia (30–32), and breast cancer (33,34). DynaD is designed to be used when the binding site is already identified but no structural information is available about the binding site. The biggest advantage of DynaD is that it can predict the bound states of small molecules to dynamic RNA loops via purely physics-based approaches using simple simulation settings built in simulation packages. DynaD uses a series of simulation techniques to sample the complex energy landscape of a small molecule-RNA interaction and predicts the binding free energies of stable binding modes. This method mimics the natural binding phenomenon using a distance-based reaction coordinate and thus does not use a scoring function as employed by docking programs. Initial bound states are created using this reaction coordinate, which then are utilized in MD simulations to discover the stable binding modes. During MD simulations, ligands will undergo conformational change to optimize binding that is directed by the force field. Even though it is challenging, adequate sampling of the complex energy landscape representing the binding phenomenon will yield stable bound states ranked by their binding free energies calculated with MM combined with Poisson-Boltzmann surface area continuum solvation (MM/PBSA) (35) and with the three-dimensional reference interaction site model (MM/3D-RISM) (36,37) approaches that will include the global minimum. Results of DynaD are crucial, as it has the potential to successfully investigate the binding modes of small molecules targeting dynamic RNA systems that will assist in silico drug-design studies. Furthermore, we developed a modified version of DynaD,

namely DynaD/Auto, which utilizes both the results of umbrella sampling calculations performed on 1×1 UU internal loops and AutoDock calculations to efficiently sample the complex energy landscapes. The strategy of DynaD and DynaD/Auto is in line with the approach performed by Al-Hashimi's group in sampling wide conformational space except that it is purely computational. However, since the number of initial bound states required to discover the correct binding modes in DynaD depends on the properties of the binding site and the ligand targeting the site such as the size of the RNA loop and flexibility of the ligand, undersampling of conformational space can cause challenges while determining the global minimum and crucial bound states in some systems. DynaD/Auto eliminates this uncertainty by exploiting information that characterizes the binding site if available. Umbrella sampling data provide all the global and local minimum conformations for the 1×1 UU internal loops in RNA CUG repeats, which allow construction of an ensemble of structures covering the conformational space of the binding site important in the binding process.

In this contribution, we studied the binding properties of nine small molecules known to target RNA CUG repeats (38) using DynaD and DynaD/Auto. Predictions are compared with experimental data, which displayed a positive correlation with correlation coefficient (R) values of 0.70 and 0.81 for DynaD and DynaD/Auto, respectively. Furthermore, we performed different binding free energy calculations and discovered that the best correlations between predictions and experiments are achieved with MM/3D-RISM calculations, highlighting the importance of solvation in binding calculations. Moreover, we detected that DynaD/Auto performed better than DynaD in determining the correlation between predictions and experiments in the studies of small-molecule RNA CUG binding because of the use of prior knowledge about the binding site, as it allows one to capture challenging binding modes, which might be difficult to observe in conventional sampling methods. Finally, we utilized dendrograms to present simple diagrams displaying how bound states are connected to each other in a binding process. With promising results, DynaD and DynaD/Auto will allow researchers to utilize two novel physics-based and computer-aided drug-design methodologies to investigate drug-like molecules aiming to target complex RNA loops.

MATERIALS AND METHODS

Determination of small molecule/RNA-binding modes

Fig. 1 displays the workflows of DynaD and DynaD/Auto methods in determining the stable binding modes of a small molecule targeting 1×1 UU internal loops in RNA CUG repeats. Both methods follow almost the same procedure except in the preparation stage, which is described as follows.

In DynaD, initial bound states for small molecule/RNA complex are created by moving the small molecule to the binding site repeatedly using a distance-based reaction coordinate, which is defined as the distance between the center of mass (COM) of the heavy atoms of the closing basepairs of the binding site and the COM of the heavy atoms of the small molecule (Figs. 1 and 2 *a*; Video S1). In the "Docking" stage of DynaD, modified generalized Born implicit solvent model (GB^{OBC}) (39) with 0.3 M salt concentration implemented to the Amber18 (40) package was used. During the initial docking stage, the initial distance between the small molecule and the closing base pairs is set to be 40 Å so that they are far away from each other. Small molecule is then gradually moved toward the binding site by 1-Å intervals each within 20 ps using distance restraints imposed on the reaction coordinate until it becomes 0 Å. As an example, a single initial docking to bring the compound from 40 Å to 0 Å will take 20×40 ps $= 800$ ps. During this process, WC base pairing and torsional and chirality restraints are imposed on the RNA residues except the loop residues so that the global RNA structure maintains the A-form orientation. It is worth highlighting the two assumptions we have in the small molecule/RNA-binding process: 1) small molecules target the 1×1 UU internal loops in RNA CUG repeats; and 2) because the ligands are small molecules, they are not going to distort the global RNA structure dramatically. As a result, the set of restraints we use allow the uridine residues to freely move around the system, where both *syn* and *anti* uridine conformations as well as unstacked states are observed without distorting the rest of the RNA (Video S1). The conformation of RNAs in the loop residues and the small molecule reflects the moment when the two molecules meet, which is called an initial bound state. Thereafter, the small molecule is slowly moved away from the binding site randomly by 1-Å intervals until the distance reaches 40 Å again. During this process, WC base pairing and torsional and chirality restraints are applied to all the RNA residues to transform the RNA back to its apo structure. This "forced binding" process is repeated several dozen times sequentially to obtain initial bound states for the small molecule/RNA CUG complex, where the initial bound states correspond to conformations having 0 Å distance between the small molecule and the closing basepairs as determined by the "Docking" step (Fig. 1). The key point of DynaD is to sample different regions of conformational space by starting with different initial bound states, thereby increasing the probability of finding the important binding poses. Thus, increasing the number of initial docked states will increase the accuracy of the predictions. These initial bound states are then utilized in explicit solvent MD simulations where structures are solvated with water and ions to investigate the dynamics (vide infra). No restraints are used in the explicit solvent MD simulations to allow small molecules to reorient themselves freely while sampling the conformational space with respect to the force field. Each MD simulation is run for 120 ns, yielding several-microseconds-long combined MD trajectories for the studied small molecule-RNA interaction. In the "Analysis" step (Fig. 1), cluster analyses are conducted where root-mean-square deviation (RMSD) is utilized to determine the structural similarity throughout the trajectory (vide infra). In the RMSD calculation, the heavy atoms of the target site, which includes the uridine residues, two closing basepairs (Fig. 2 *a*), and the small molecule, are considered. Snapshots with RMSD ≤ 1.25 Å are clustered into the same group, and binding free energies of each cluster having more than 50 snapshots are calculated using the MM/3D-RISM method (36,37) (vide infra).

In DynaD/Auto, an alternative approach was designed to investigate the small molecule/RNA-binding phenomenon. Even though DynaD mimics the natural binding process using a distance-based reaction coordinate while targeting the RNA-binding site, the number of initial bound states required to find the global minimum depends on the nature of the binding site and the properties of the small molecules, which can cause sampling issues. Use of prior knowledge about the binding site, such as the free energy landscape of the 1×1 UU internal loop in RNA CUG predicted by umbrella sampling calculations, can expedite the calculations, as it can provide unique and important conformations for the RNA loop motif not easily
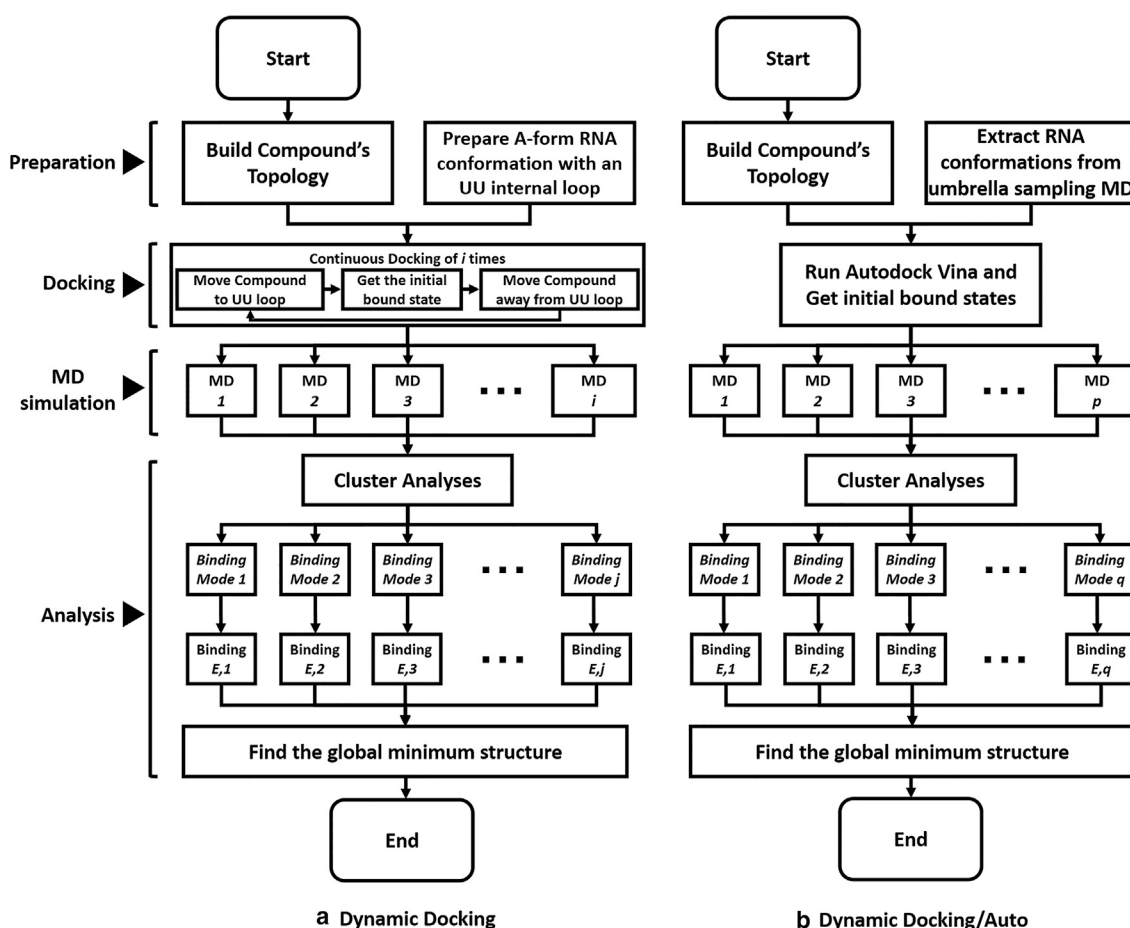
FIGURE 1 Flow charts for dynamic docking (DyanD) (*a*) and dynamic docking/AutoDock (DynaD/Auto) (*b*) approaches to predict binding free energies of small molecules targeting RNA CUG repeats. See Video S1 displaying the initial docking stage of DynaD. The bound state predicted to have the lowest binding energy is the proposed global minimum.

accessible by conventional MD simulations. We therefore created a modified version of DynaD, DynaD/Auto, where umbrella sampling results of 1 × 1 UU internal loop in RNA CUG repeats are utilized to create an ensemble of RNA structures, which are used in AutoDock Vina calculations to create initial bound states (vide infra). Fig. S1 displays the two-dimensional (2D) free energy landscape predicted by umbrella sampling calculations for a single uridine of a 1 × 1 UU internal loop in RNA CUG repeats using two reaction coordinates mimicking two crucial motions in an RNA residue: rotation around χ displaying the base orientation with respect to sugar and rotation around a pseudo-torsion, θ_1, imitating base stacking ↔ unstacking. Umbrella sampling calculations exhibit 11 minima in the 2D free energy landscape for a single uridine (Fig. S1), which we used to build the initial conformations for the model RNA CUG (vide infra). Because of the symmetry in the RNA CUG repeats, a total of 66 unique RNA loop conformations were homology modeled using the umbrella sampling data, as the 2D free energy landscape represents the results for a single uridine residue in a 1 × 1 UU internal loop. AutoDock Vina (18) was then utilized to obtain the best initial bound states for a small molecule targeting each RNA loop conformation. The rest of the process is similar to the DynaD approach, where initial bound states are used as initial conformations in the explicit solvent MD simulations (Fig. 1) (vide infra). All the MD trajectories are combined and used in the cluster analyses, where binding free energies for each cluster are calculated using the MM/3D-RISM method as before, yielding the global minimum bound state for small molecule/RNA interaction (vide infra).

Selection of small molecules targeting 1 × 1 UU internal loops in RNA CUG repeats

Rzuczek et al. performed a set of experiments and investigated 320 drug-like small molecules, which were collected according to their structural diversity and chemical similarity (38). First, they performed a time-resolved fluorescence resonance energy transfer (TR-FRET) assay to discover 28 hit compounds disrupting the r(CUG)$_{12}$/MBNL1 complex in vitro, which could be due to either targeting r(CUG)$_{12}$ or MBNL1. To assess whether the compounds are specifically targeting the RNA CUG repeats or not, they investigated the biological activity of these compounds using a DM1 cellular model, wherein they co-transfected a DM1 mini-gene containing 960 RNA r(CUG) repeats (r(CUG)$_{960}$), which reports on alternative splicing of cardiac troponin T (cTNT) exon 5. While the amount of mature mRNA containing exon 5 in healthy cells (representing absence of r(CUG)$_{960}$) is $\approx 55\%$, in the presence of r(CUG)$_{960}$, $\approx 90\%$ of exon 5 is included in the mature mRNA. As a result, if a compound targets r(CUG)$_{960}$ and rescues MBNL1 in cellulo, a decrease in exon 5 inclusion should be observed, which is what they measured (38). Among the 28 small molecules, 12 small molecules improved cTNT alternative pre-mRNA splicing pattern toward wild type, and only nine small molecules out of 12 showed a similar percentage level observed in healthy cells. Those nine small molecules are considered to have a significant level of binding abilities to RNA CUG repeats to interfere with the formation of r(CUG)$_{960}$/MBNL1 complex. We therefore selected these nine small

a

Compound

Binding site

b

Compound A ← 1

Compound B ← 2

Compound C ← 7

Compound D ← 8

Compound E ← 11

Compound F ← 16

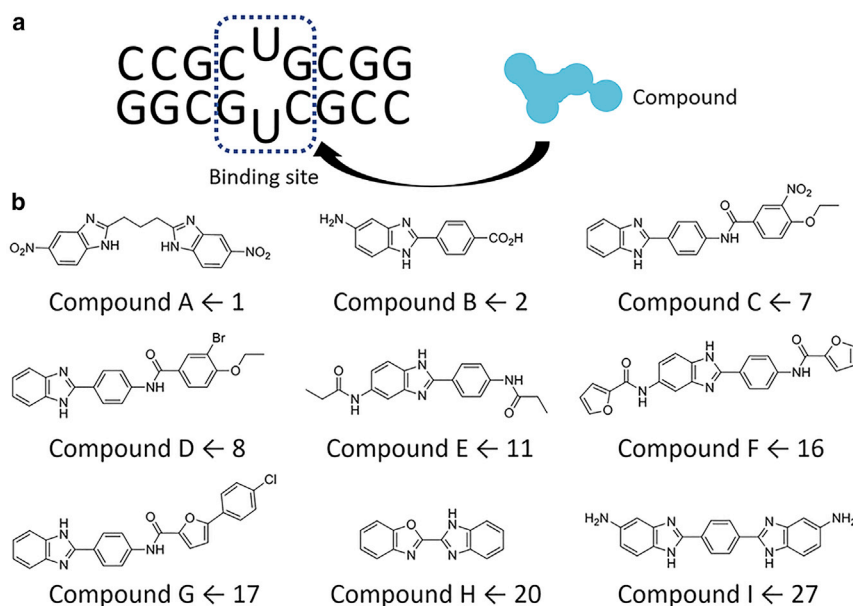Compound G ← 17

Compound H ← 20

Compound I ← 27

FIGURE 2 (*a*) Schematics of the binding process and secondary structure of the model RNA CUG system utilized in this study, where dashed blue rectangle highlights the binding site. (*b*) Nine small molecules targeting RNA CUG repeats inhibiting r(CUG)$_{12}$/MBNL1 complex. Number notation displayed under each molecule is used by Rzuczek et al. (38) while we use the letter notation to differentiate each small molecule. Note that in (*a*), the binding site includes two Watson-Crick GC basepairs defined as "closing basepairs" and a noncanonical 1 × 1 UU pair. To see this figure in color, go online.

molecules to test DynaD and DynaD/Auto on a model RNA CUG sequence in this study (Fig. 2 *b*). Relative experimental binding properties of the small molecules are quantified by their in vitro activity for disruption of the r(CUG)$_{12}$/MBNL1 complex as determined by a TR-FRET assay at 100 μM concentration and used as experimental data to compare predicted binding free energies.

Preparation of model systems and parameterization of small molecules

A model RNA sequence, r(5′-CCG CUG CGG-3′/5′-CCG CUG CGG-3′) (Fig. 2 *a*), which contains a single RNA CUG repeat, was utilized to determine the binding affinities of small molecules to RNA CUG repeats. The Amber99 force field (41) with revised χ (42) and α/γ (43) torsional parameters were chosen to represent the properties of RNA. Nine small molecules (Fig. 2 *b*), which display highest binding affinities among the candidate small molecules in in vitro and in cellulo experiments, were selected for in silico calculations targeting RNA CUG repeats (vide supra). The generalized Amber force field (GAFF) (44) was used to represent the small molecules with restrained electrostatic potential (RESP) charges derived following the RESP charge-fitting protocol (45,46) as described in our previous studies (15,16,27–34), where small molecules were first optimized and then electrostatic potentials as a set of grid points were calculated at the HF level with the 6-31G* basis set using Gaussian09 (Tables S1–S9) (47). Furthermore, we investigated a guanine riboswitch system using DynaD, whereby an X-ray structure (PDB: 2EES) (48) was utilized to compare the structural predictions (see supporting material for details).

Explicit solvent MD simulations

Initial bound states obtained from DynaD and DynaD/Auto were used as initial conformations for the explicit solvent MD simulations. To obtain the initial bound states for a small molecule, DynaD required 3.5 GPU hours using a Quadro P620 GPU whereas DynaD/Auto required 16.5 min using 12 Intel i7-8700K cores. Small molecule/CUG complexes were solvated with TIP3P water (49) molecules in octahedral boxes with a buffer of 8 Å. To neutralize the systems, Na$^+$ ions (50) were included in each system. Each system was minimized in two steps, which were then equili-

brated. In the first step of minimization, the water molecules and ions were minimized while positional restraints were imposed on small molecule/RNA complex. In the second step restraints were removed, and all the residues were included in the minimization process. In each minimization step, steepest descent minimization of 25,000 steps was followed by a conjugate gradient minimization of 25,000 steps. For the equilibration, we used the Langevin thermostat to heat up the system to 300 K in 500,000 steps without including any restraints on the system. Long-range electrostatic interactions were calculated using the particle mesh Ewald method. Production runs were followed after equilibration, where temperature and pressure were maintained throughout the MD simulations at 300 K and 1 bar, respectively, using Langevin dynamics and Berendsen barostat (NPT ensemble) (see Table S10 for sample input files). For each bound state determined by DynaD and DynaD/Auto in the "Docking" stage (Fig. 1), 120-ns-long MD simulations were run with a time step of 2 fs. For 120 ns, 13.3 GPU hours were required using an NVIDIA Tesla V100 GPU card. For the DynaD/Auto calculations, 66 independent MD simulations, each determined after umbrella sampling and AutoDock Vina calculations (vide infra), were conducted for each small molecule/RNA CUG complex. Note that the number of the initial bound states used in DynaD and DynaD/Auto is 66 to have comparable results, where the total combined MD simulation time for each system is 7.92 μs (120 ns × 66 = 7.92 μs).

Extracting significant RNA conformations from umbrella sampling calculations for DynaD/Auto

The structural and thermodynamic properties of 1 × 1 UU internal loops were previously investigated using conventional MD simulations and umbrella sampling MD simulations using a model RNA CUG system, which displays multiple stable conformations (Fig. S1) (51). According to the 2D potential of mean force (PMF) calculations with χ torsion and a pseudo-torsion (θ) mimicking base stacking ↔ unstacking as the reaction coordinates, we previously discovered that the global minimum structure of 1 × 1 UU internal loops in RNA CUG repeats is the stacked state in *anti-anti* UU orientation (Fig. S1) (51). Furthermore, the 2D PMF surface displays several local minima states having *anti* and *syn* base orientations (Fig. S1). Umbrella sampling results provide all the stable and statistically important regions for the 1 × 1 UU internal loops in RNA CUG repeats,

which can be capitalized by small molecules when targeting RNA CUG repeats. Thus, we consider these local minima as well as the global minimum as probable orientations the 1×1 UU internal loops in RNA CUG repeats prefer upon small-molecule binding. Therefore, an ensemble of RNA structures were constructed from umbrella sampling calculations, where 11 conformations representing the stable states observed in 2D (χ,θ) PMF profile (Fig. S1) were utilized to build 66 unique loop conformations for the 1×1 UU internal loop. These 66 structures cover a wide range of conformational space for 1×1 UU internal loops while being targeted by small molecules and have been combined with AutoDock calculations to build the initial bound states described in DynaD/Auto calculations to investigate how small molecules target RNA CUG repeats using an alternative approach (Fig. 1).

AutoDock Vina

DynaD determines initial bound states for a small molecule interacting with 1×1 UU internal loops in RNA CUG repeats by using a series of repetitive implicit solvent MD simulations. DynaD/Auto, however, uses AutoDock Vina to obtain the initial bound states for the 66 1×1 UU internal loop structures extracted from umbrella sampling calculations. RNA and small molecules were prepared using default AutoDock Vina protocols (18,52). The COM of the small molecule was selected as the grid center of docking. The size of the grid box was adjusted manually to contain the internal loop and the closing basepairs completely. The number of binding modes generated was set to 20, and the default values for exhaustiveness of search level and the maximum energy difference were utilized in the calculations (18,52). For each small molecule, virtual screening was performed and only the binding mode with strongest binding affinity was considered for explicit solvent MD simulations.

Cluster analyses and binding free energy calculations

We utilized our in-house code to perform cluster analyses as described previously (15,16,27–34). For each system, all the MD trajectories were first combined and then RMSD was used as the metric to cluster similar structures. Snapshots with RMSD $\leq$ 1.25 Å were clustered into the same group. During this process, symmetry observed in each system was included in the analyses so that we did not overpredict the total number of clusters. Average structures for each cluster were calculated at the end. MM/PBSA (22), MM/GBSA (22), and MM/3D-RISM (36,37) analyses were conducted on each cluster to determine the global minimum bound state for each system, whereby the MMPBSA.py module of Amber18 (40) was utilized in the calculations. In the small molecule/RNA-binding process, water molecules will play crucial roles. Thus, we decided to utilize the RISM of molecular solvation, as it is an inherently microscopic approach calculating the equilibrium distribution of the solvent from which all thermodynamic properties are then determined. In MM/3D-RISM calculations, Kovalenko-Hirata closure (3D-RISM-KH) (37) molecular solvation theory was utilized to calculate solvation properties (see Table S11 for a sample input file).

Dendrogram analyses

Interpreting large sets of data that come with simulations is a daunting task even with analysis tools. Thus, we performed dendrogram analyses to visualize transformation pathways between small molecule/RNA bound states, where dendrogram graphs displaying the "closeness" between a set of clusters as determined by the metric, RMSD, were created for each system (see supporting material for details).

RESULTS AND DISCUSSION

The global minimum predicted by DynaD for a guanine riboswitch overlaps perfectly well with its X-ray structure

To verify the quality of DynaD, we predicted the bound state of a known system (PDB: 2EES), where a small molecule, residue name HPA, interacts with an RNA riboswitch. Comparison of the prediction with its experimental structure (2EES) modeled by X-ray crystallography displayed a perfect overlap with an all-heavy-atom RMSD of 1.2 Å, implying how good DynaD can perform in an RNA riboswitch system (Table S12 and Fig. S2).

The initial bound conformations change in MD simulations because of the force field

The energy landscapes of small molecules targeting RNA CUG repeats are complex because the conformations of both the RNA-binding site and small molecule can change during binding. Therefore, a single RNA conformation representing the binding site will not be enough to mimic the binding process. To effectively scan the conformational space, however, one needs to imitate the binding phenomenon quickly. DynaD tries to do that by following a reaction coordinate that will simulate the physical binding behavior. The advantage of DynaD is that it allows one to study any RNA target without having prior knowledge about the structural details of the binding site. The initial docked state using DynaD is calculated by forcing small molecules to interact with the RNA-binding site. In the modified version, DynaD/Auto, an ensemble of structures important in the binding process is built using prior knowledge about the RNA target site such as the energy landscape predicted by umbrella sampling calculations. These RNA structures are then utilized in AutoDock calculations to come up with initial docked states. These initial states then change during MD simulations, where the force field bears a part in optimizing the binding interactions. Both docking methods can be thought of as two different approaches that place small molecules close to relevant conformational states important in the binding process. Consequently, MD simulation will cause the orientations of both the RNA-binding site and small molecule to change to optimize binding. By running MD simulations on different initial docked states and combining all the trajectories, one can perform cluster analyses to determine the stable bound states. Predictions of relative binding free energies can then describe what the global minimum structure is for small molecule/RNA complex.

As an example, using the DynaD/Auto approach, each small molecule was docked to 66 unique RNA CUG conformations using AutoDock Vina. In each AutoDock Vina calculation, conformations with highest binding affinities

calculated using an empirical scoring function were selected as initial docked states representing small molecule/RNA CUG complexes, which are used as initial structures in MD simulations. A total of 66 independent MD simulations, each 120 ns long, were performed. Each MD simulation was analyzed by calculating the RMSD with respect to initial structure to determine the structural changes observed in MD trajectories. In the RMSD calculations, heavy atoms of the small molecules and loop residues as well as closing basepairs were included. Fig. S3 displays the RMSD variation over simulation time and distance between the COM of small molecule **I** and the RNA-binding site in **I**/RNA CUG complex using DynaD/Auto. For simplicity, only the results of 6 out of 66 MD simulations are displayed and highlighted in unique colors (Fig. S3). Three structural motions representing reorientation of **I** and the RNA-binding site (reorientation), relocation of **I** to the terminal basepairs (relocation), and loss of binding (interruption) are observed in the MD trajectories (Fig. S3). Histogram analyses display that most of the structures have RMSD > 2.0 Å (Fig. S3 *c*), an implication that initial conformations are altered in the MD simulations. RMSD and distance values ≤ 5.0 Å (black, red, green, and blue trajectories in Fig. S3) imply reorientation of both **I** and RNA-binding site, which accounts for 64.8% of structures calculated from histogram analysis (Fig. S3 *c*). When the small molecule did not maintain the initial bound state, both RMSD and distance values were over 5.0 Å, representing relocation and/or interruption as displayed in cyan and brown trajectories, respectively, in Fig. S3, *a* and *b*.

MM/3D-RISM results have a better correlation with the experimental data compared with MM/PBSA and MM/GBSA results

Binding free energy calculations can be performed using MM/GBSA (22), MM/PBSA (22), and MM/3D-RISM (36,37) approaches. In MM/GBSA and MM/PBSA, MM energies are combined with the Poisson-Boltzmann and generalized Born surface area continuum solvation methods, respectively, while in MM/3D-RISM, 3D molecular theory of solvation, also known as the 3D reference interaction site model (3D-RISM), is utilized to calculate the solvation structure and thermodynamics from the first principles of statistical mechanics. An advantage of the use of RISM in binding free energy calculations is that it yields solvent density distribution and, thus, potential water bridges in the binding site, allowing the effect of water molecules in binding. Fig. S4 displays the correlation between the experimental data and the predicted binding free energies for the nine small molecules we investigated using MM/GBSA, MM/PBSA, and MM/3D-RISM approaches. The best correlation was discovered using MM/3D-RISM with a correlation coefficient, *R*, of 0.81, while MM/PBSA and MM/GBSA approaches gave *R* values of 0.44 and 0.30,

respectively. Even though entropy is not included in calculations, water inclusion by RISM improves the correlation dramatically (Fig. S4), highlighting the importance of water molecules in small molecule/RNA binding. The nine small molecules we investigated are structurally comparable. Thus, it is expected that they have similar entropic effects and will not distort the correlation we already observe between the predictions and experimental data (Fig. S4). It is worthy of mention that the conformational entropy due to the structural changes observed in RNA and the small molecules before and after the binding process cannot be captured by simple methods such as normal mode (NMODE) analyses (53). We know that $\Delta G_{\text{binding}} < 0$ kcal/mol because these small molecules inhibit the formation of r(CUG)$_{12}$/MBNL1 complex, implying binding. Nevertheless, predicted binding free energies by MM/3D-RISM are not realistic, as they are around -15 kcal/mol. Inclusion of conformational entropy would have brought the predicted binding free energies to realistic values.

DynaD/Auto combined with 3D-RISM calculations provides the best correlation between predicted binding free energies and experimental data

While investigating the binding properties of each small molecule using the DynaD and DynaD/Auto calculations, 66 initial bound states were used to run 66 MD simulations, in which the 66 trajectories were combined to perform cluster analyses. The combined MD trajectories in DynaD and DynaD/Auto correspond to 7.92 μs MD time. Table 1 displays the total number of clusters determined for each small molecule/RNA system using DynaD and DynaD/Auto. Using the MM/3D-RISM approach, we calculated the binding free energies for each cluster (Tables S13–S30). Bound states with lowest binding free energies are the proposed global minima for each system (Table 1). Predicted binding energies are then compared with experimental data displaying r(CUG)$_{12}$/MBNL1 complex formation rates when small molecules are present (Fig. 3) (38). When there are no small molecules present in the system, r(CUG)$_{12}$/MBNL1 complex will form 100%, which is the reference value used in the analyses (Fig. 3 and Table 1). When small molecules are present, which are known to specifically target RNA CUG repeats to inhibit r(CUG)$_{12}$/MBNL1 complex formation, rates will decrease. The in vitro activity for disruption of the r(CUG)$_{12}$/MBNL1 complex as determined by TR-FRET exhibits complex formation rates between 20% and 70% when one of the nine small molecules is present. It was discovered that the predicted binding free energies are positively correlated with the experimental data (Fig. 3). Namely, the higher the binding affinities predicted, the lower the r(CUG)$_{12}$/MBNL1 complex formed depending on the type of the small molecule targeting RNA CUG repeats. While the correlation coefficient, *R*, for predictions using DynaD is calculated to be 0.70, predictions using

TABLE 1 Results displaying properties of nine small molecules (Fig. 2) targeting RNA CU̲G repeats predicted by DynaD and DynaD/Auto

Small molecule	Experimental r(CU̲G)$_{12}$/ MBNL1 formation rates (%)[a]	Predicted ΔG binding free energies (kcal/mol)			No. of clusters determined[c]		χ Torsions of uridines in global minimum bound states	
		DynaD	DynaD/Auto	$\Delta\Delta G$[b]	DynaD	DynaD/Auto	DynaD	DynaD/Auto
A	24.6964	−16.77	−17.69	−0.92	117 (41)	111 (34)	*anti/anti*	*anti/syn*
B	25.5061	−18.10	−19.37	−1.27	85 (51)	117 (39)	*anti/anti*	*anti/anti*
C	40.0810	−13.63	−15.62	−1.99	91 (32)	54 (23)	*anti/anti*	*anti/anti*
D	40.4858	−12.45	−15.62	−3.17	101 (46)	56 (20)	*anti/anti*	*anti/anti*
E	43.7247	−13.88	−14.02	−0.14	113 (48)	61 (29)	*anti/anti*	*anti/syn*
F	52.6316	−14.34	−15.15	−0.81	117 (38)	60 (10)	*anti/syn*	*anti/anti*
G	52.6316	−13.82	−15.71	−1.89	89 (10)	67 (4)	*anti/anti*	*anti/syn*
H	61.9433	−12.12	−12.76	−0.64	47 (23)	102 (17)	*anti/anti*	*anti/syn*
I	69.2308	−13.89	−14.53	−0.64	103 (49)	88 (41)	*anti/syn*	*anti/syn*

[a]Proportion of the RNA CUG/MBNL1 complex formed relative to control group (100%) (38).
[b]$\Delta\Delta G = \Delta G_{\text{DynaD/Auto}} - \Delta G_{\text{DynaD}}$.
[c]The values in parentheses represent the common structures determined by the two methods. For example, for **A**, cluster analyses found 117 and 111 clusters using DynaD and DynaD/Auto, respectively, where 41 clusters of DynaD are structurally similar to 34 clusters of DynaD/Auto with RMSD ≤ 1 Å.

DynaD/Auto have a higher accuracy with $R = 0.81$ (Fig. 3). The predicted binding free energies by DynaD/Auto were always lower than the predictions of DynaD (Fig. 3). The lowest $\Delta\Delta G = \Delta G_{\text{DynaD/Auto}} - \Delta G_{\text{DynaD}}$ was observed in **E**, with $\Delta\Delta G = -0.14$ kcal/mol while the highest $\Delta\Delta G$ was observed in D, with $\Delta\Delta G = -3.17$ kcal/mol (Table 1). Even though DynaD/Auto outperforms DynaD, increasing the initial docked states in DynaD not only improves the predicted correlation but also the predicted binding free energies (Fig. S5). In DynaD/Auto, 66 unique RNA structures representing the critical RNA orientations important in the small molecule/RNA-binding process were utilized. DynaD, however, does not have such precise initial docked

states as DynaD/Auto while searching for the global minimum bound state. Naturally, increasing the number of initial docked states would be advantageous in DynaD to discover major binding modes, but this will require more sampling of the conformational space and, thus, more computational power. To obtain a sense of how the predictions of DynaD improve with respect to the total number of initial docked states used, we calculated the binding free energies for each small molecule using 10, 20, 30, 40, 50, and 66 initial docked states (Fig. S5). The variation of correlation coefficient with respect to the total number of initial docked states used is shown in Fig. S5 *a*. As expected, the accuracy of predictions determined by correlation coefficient is improved with the increase of the total number of initial docked states from $R \approx 0.5$–0.7 (Fig. S5 *b*). Increasing the total number of initial docked states in DynaD improves the predicted binding free energies and, thus, the predicted trendlines by shifting them downward toward the trendline predicted by DynaD/Auto (Fig. S5 *a*). Nevertheless, DynaD/Auto still has an improved correlation compared with DynaD, as it uses prior knowledge about the RNA-binding site.
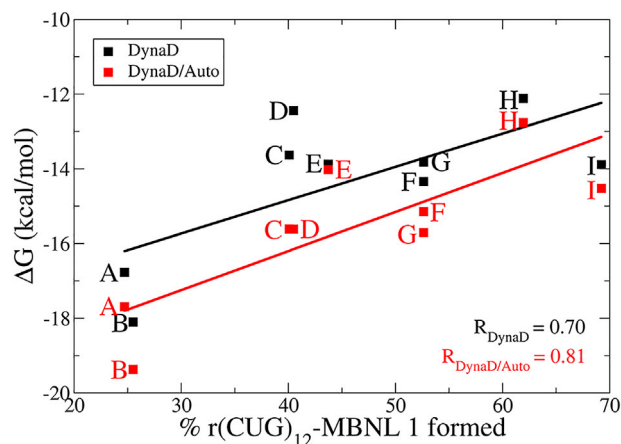

FIGURE 3 Correlation between predicted binding free energies (in kcal/mol) as calculated by MM/3D-RISM approach and in vitro activity of nine small molecules for disruption of r(CU̲G)$_{12}$/MBNL1 complex as determined by TR-FRET assay (38). Color notation is used to display predictions of DynaD (*black*) and DynaD/Auto (*red*). Predicted binding free energies of global minimum states for each small molecule are used in the figure. Note that the lower the percent r(CU̲G)$_{12}$/MBNL1 value, the higher the inhibition of r(CU̲G)$_{12}$/MBNL1 complex due to higher binding affinity of the small molecule to the 1 × 1 UU internal loops in r(CU̲G)$_{12}$. To see this figure in color, go online.

Both DynaD and DynaD/Auto predict almost the same global minimum bound states for E and I with RMSD < 1 Å

As described above, both DynaD and DynaD/Auto sample the complex conformational space using two different approaches, whereby the former utilizes a distance-based reaction coordinate while the latter combines structural data extracted from umbrella sampling calculations with AutoDock calculation to create initial bound states. If enough sampling is performed, both methods should converge to the same global minimum as observed in **E** and **I** (Figs. 2–4). RMSD between the global minimum bound states predicted by DynaD and DynaD/Auto for **E** and **I** are 0.58 Å and 0.17 Å, respectively (Fig. 4).
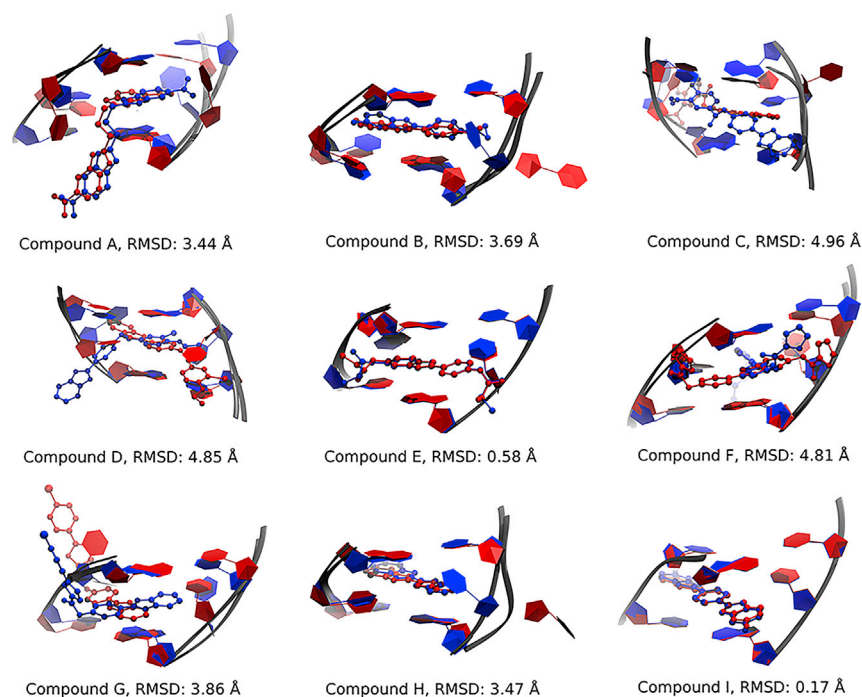
Compound A, RMSD: 3.44 Å

Compound B, RMSD: 3.69 Å

Compound C, RMSD: 4.96 Å

Compound D, RMSD: 4.85 Å

Compound E, RMSD: 0.58 Å

Compound F, RMSD: 4.81 Å

Compound G, RMSD: 3.86 Å

Compound H, RMSD: 3.47 Å

Compound I, RMSD: 0.17 Å

FIGURE 4 The global minimum bound states determined by DynaD (*blue*) and DynaD/Auto (*red*) for the small molecules targeting RNA CUG repeats (Fig. 2). NewRibbons and CPK representations are used to display the RNA and small molecules, respectively. For simplicity, no hydrogen atoms are displayed. RMSD values are displayed under each system to highlight how similar the blue and red structures are (see main text for details). To see this figure in color, go online.

$\Delta\Delta G$, defined as $\Delta G_{\text{DynaD/Auto}} - \Delta G_{\text{DynaD}}$, for **E** and **I** are -0.14 and -0.64 kcal/mol, respectively, implying that both methods predict the same structures with similar binding energies (Fig. 3 and Table 1). The global minimum bound state of small molecule **E**/RNA CUG predicted by DynaD/Auto displays one of the uridines in *syn* state while it is *anti* in DynaD prediction (Table 1). As noted above, DynaD/Auto uses conformations extracted from umbrella sampling calculations, which have *syn* states in it, while DynaD mimics the physical binding phenomenon using a distance-based reaction coordinate while scanning the conformational space, which might require a longer time in finding the global minimum bound states. Indeed, the global minimum bound states of small molecule **I**/RNA CUG predicted by DynaD and DynaD/Auto are identical and display the 1 × 1 UU internal loop in *anti-syn* orientation (Table 1). Furthermore, comparisons of the clusters predicted by DynaD and DynaD/Auto reveal that half of the clusters determined by each method are in common. For example, for the small molecule **I**/RNA CUG interaction, 49 out of 103 predicted by DynaD can be described by 41 out of 88 clusters predicted by DynaD/Auto (Table 1). A similar result is found for small molecule **E**/RNA CUG, where 48 out of 113 clusters predicted by DynaD can be described by 29 out of 61 clusters predicted by DynaD/Auto (Table 1). When a small molecule targets a dynamic system, such as RNA 1 × 1 UU internal loops, both the conformations of the binding site and the small molecule will determine the final bound states. For example, we studied the dynamic behavior of each small molecule in explicit solvent and performed cluster analyses to determine their

conformational flexibility. Results showed that **E** and **I** have nine and four clusters, respectively, implying that they are not as disordered as **C**, **D**, **F**, and **G** (Table 2). As a result, DynaD and DynaD/Auto could predict the same results for these two small molecules, which is partly due to sampling similar conformational space.

The global minimum structures predicted by DynaD and DynaD/Auto for small molecules A, B, and H display RMSD < 3.7 Å

Investigation of conformational flexibility of **A**, **B**, and **H** (Fig. 2 *b*) using explicit solvent MD simulations followed by cluster analyses yielded four clusters for **A**, and one cluster for **B** and **H** implying that these small molecules are entropically ordered (Table 2). We calculated the solvent-accessible surface areas (SASAs) of these molecules using

TABLE 2 Properties of nine small molecules (Fig. 2) in apo state extracted from explicit solvent MD simulations

Small molecule	SASA[a] (Å^2)	No. of clusters determined
A	290.70	4
B	249.53	1
C	386.71	15
D	373.03	17
E	346.13	9
F	382.79	22
G	383.42	13
H	230.21	1
I	334.44	4

Each small molecule was individually studied to determine its flexibility.
[a]SASA, solvent-accessible surface area.

explicit solvent MD simulations in their apo state to display their apparent sizes. SASAs of these small molecules are between 230 and 290 Å^2 (Table 2), where **B** and **H** are structurally the smallest small molecules we investigated (Fig. 2 *b*). Comparison of the global minimum bound states predicted by DynaD and DynaD/Auto display RMSD values <3.7 Å (Fig. 4). Even though these compounds are more ordered than **E** and **I**, DynaD and DynaD/Auto predict different conformations for the global minimum. While the orientations of the compounds in the bound states predicted by DynaD and DynaD/Auto are fairly similar, the orientations of uridine residues in the binding site are very different, which is the reason for the high RMSDs (Fig. 4). As described above, conformations of both the RNA-binding site and small molecules will determine the final bound states. Even though the small molecules of **A**, **B**, and **H** are relatively rigid and ordered, proper sampling of the RNA target site determines the accuracy of the predictions. In the cases of **A** and **H**, the orientations of small molecules are exactly the same except the final orientations of the RNA loop sites, where one of the uridines is in *syn* state and the other one is flipped out from the helical axis, which is why we observe $\Delta\Delta G$ values of -0.92 and -0.64 kcal/mol, respectively (Figs. 2 and 3; Table 1). In the case of **B**, the predicted lowest binding energy states display the benzimidazole ring of **B** in different orientations as well as one of the uridines in the RNA-binding site flipped out, which is likely the reason why we observe a $\Delta\Delta G$ value of -1.27 kcal/mol (Figs. 2 and 4; Table 1). Even though the orientations of small molecules are the same in the predictions of DynaD and DynaD/Auto, DynaD requires more simulation time to sample other RNA loop conformations to predict the global minimum bound state predicted by DynaD/Auto, which utilizes prior knowledge about the RNA-binding site. DynaD predicted 117, 85, and 47 clusters, respectively, for **A**, **B**, and **H**, interacting with RNA C<u>U</u>G repeats, while DynaD/Auto predicted 111, 117, and 88 clusters, respectively (Table 1). Although more than half of the clusters predicted for **B** and **H** by DynaD are observed in the predictions of DynaD/Auto, less than 33% of the clusters predicted by DynaD/Auto are observed in DynaD. For example, only 34, 39, and 17, respectively, of 111, 117, and 102 clusters predicted by DynaD/Auto for **A**, **B**, and **H** are observed in the DynaD approach (Table 1). A wider conformational space is scanned by DynaD/Auto compared with DynaD, which caused differences in predicted binding energies (Table 1).

Lowest energy bound states for small molecules C, D, F, and G are different as predicted by DynaD and DynaD/Auto with RMSD > 3.86 Å

The small molecules **C**, **D**, **F**, and **G** have SASA over 370 Å^2, representing the biggest molecules in the set we studied (Table 2). Furthermore, MD simulations of these small molecules display more than 13 clusters preferred by these small molecules, implying that they are disordered (Table 2). As a result, the lowest energy bound states predicted by DynaD and DynaD/Auto are different with RMSD > 3.86 Å (Fig. 4). DynaD predicted 91, 101, 117, and 89 clusters, respectively, for **C**, **D**, **F**, and **G** interacting with RNA C<u>U</u>G repeats, while DynaD/Auto predicted 54, 56, 60, and 67 clusters, respectively (Table 1). Not many common structures are predicted by both methods. For example, while studying the binding properties of **G**, only 10 out of 89 clusters predicted by DynaD can be described by 4 out of 67 clusters predicted by DynaD/Auto (Table 1). Similar behavior was observed while studying **C**, **D**, and **F**. As a result, $\Delta\Delta G$ for **C**, **D**, **F**, and **G** are -1.99, -3.17, -0.81, and -1.89 kcal/mol, respectively, implying that DynaD requires more sampling compared with DynaD/Auto results (Fig. 3 and Table 1). Prediction of global minimum bound state of a disordered small molecule targeting a dynamic RNA-binding site, such as 1×1 UU internal loops, is a challenging problem. Thus, utilization of prior knowledge about the binding site, such as DynaD/Auto utilizing the energy landscape of 1×1 UU internal loops predicted by umbrella sampling calculations, can expedite the calculations. Nevertheless, $\Delta\Delta G$ values are within error limits, and thus the predicted binding energies can still help in in silico drug-design processes. For example, if a set of small molecules known to target an RNA-binding site were investigated using DynaD, the predicted binding energies could provide enough data to determine the most potent small molecule within that set, which then could be studied in detail.

Undersampling of *syn* orientations in DynaD hinders the predictions

The energy landscape representing the binding of a small molecule to the model RNA C<u>U</u>G is very complex, whereby both methods investigate specific points in the energy landscape that matter in the binding process. One of the differences between DynaD and DynaD/Auto is that the latter investigates a relatively wider conformational space of the bound states compared with former. For example, the 2D (χ,θ) conformational space, where χ and θ represent the base orientation with respect to sugar and a pseudo-torsion mimicking base stacking $\leftrightarrow$ unstacking, respectively, of uridine residues in model RNA C<u>U</u>G targeted by **C** and **I**, display that DynaD/Auto scans 17% and 25% of the available 2D (χ,θ) landscape, respectively, while DynaD only scans 11.5% and 17.2% of it (Fig. 5). A similar trend is observed in the other systems (Fig. 5, *c* and *d*). It is important to note that the 2D (χ,θ) energy landscape predicted for uridine residues in RNA C<u>U</u>G repeats display conformations with high energies, such as the red regions in Fig. S1, which are 16 kcal/mol less stable than the global minimum. Thus, MD simulations will not sample 100% of the conformational space displayed in the 2D (χ,θ)
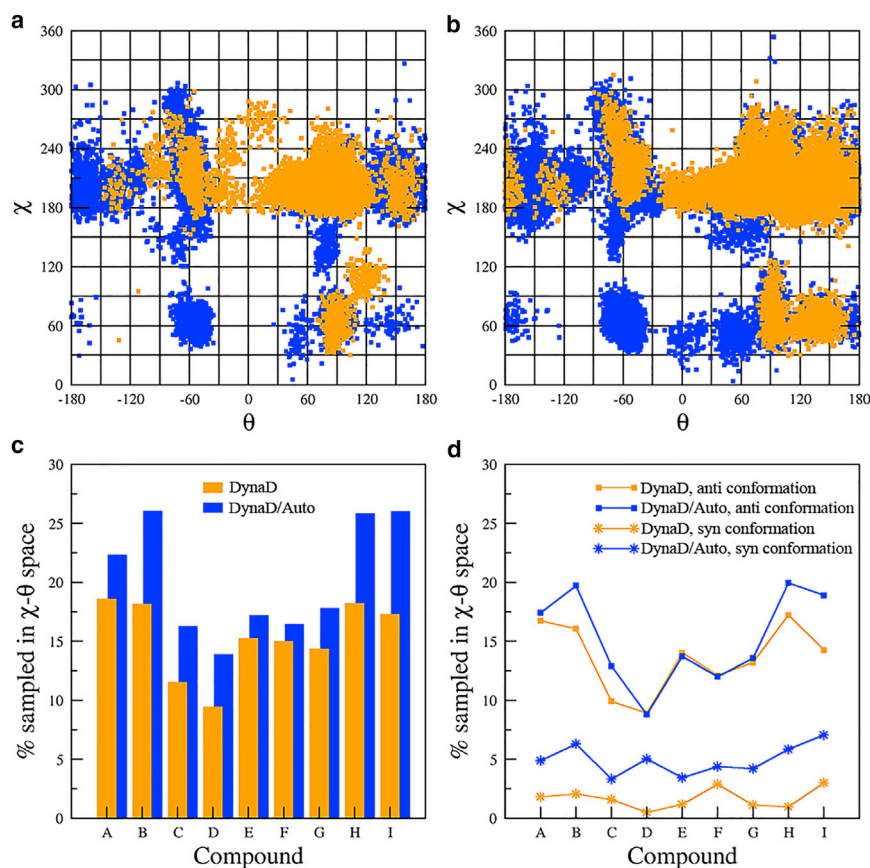
FIGURE 5 Two-dimensional (χ,θ) conformational space sampled by uridine residues in **C**/RNA C**U**G and **I**/RNA C**U**G complexes. χ and θ represent base orientation with respect to sugar and base stacking $\leftrightarrow$ unstacking, respectively. Snapshots observed in the clusters are analyzed to create the results. Small molecules of **C** (*a*) and **I** (*b*) interacting with the model RNA C**U**G display that a wider conformational space is sampled in DynaD/Auto (*blue*) compared with DynaD (*orange*). 2D PMF surface of 1×1 UU internal loops in a model RNA C**U**G predicted by umbrella sampling calculations (Fig. S1) is in line with the sampled regions in **C**/RNA C**U**G and **I**/RNA C**U**G complexes. Percentage of 2D (χ,θ) space sampled by each small molecule (*c*) in DynaD (*orange*) and DynaD/Auto (*blue*) exhibit similar results. In (*c*), grid points are created in 2D (χ,θ) space to determine whether a grid point is sampled, which is then used to calculate the percentages. Percentage of *anti* and *syn* orientations sampled by each small molecule in DynaD and DynaD/Auto (*d*) show that one of the reasons for wider sampling observed in DynaD/Auto is because of broader sampling of *anti* ($\chi \approx 200°$) and, mainly, *syn* ($\chi \approx 60°$) orientations. For example, results of **I** shown in (*d*) display that 14% and 3% of 2D (χ,θ) space representing *anti* and *syn* regions are sampled in DynaD, while 19% and 7% of the corresponding regions are sampled in DynaD/Auto. Note that sum of *anti* and *syn* percentages shown in (*d*) produces results shown in (*c*). To see this figure in color, go online.

landscape. As already described, we performed cluster analyses to determine the stable bound states. The 2D (χ,θ) points observed in these clusters for the cases of **C** and **I** are plotted to explicitly highlight the 2D space sampled in each case, where orange and blue colored points represent the results of DynaD and DynaD/Auto, respectively (Fig. 5, *a* and *b*). Compared with Fig. S1, DynaD/Auto samples almost all the minima discovered for RNA C**U**G repeats while DynaD samples a subset of them (Fig. 5, *a* and *b*). One of the reasons why DynaD/Auto was able to produce better results is that it samples the uridine residues with *syn* orientation much better. Pyrimidine residues such as uridine and cytidine are known to prefer *anti* ($\chi \approx 200°$) over *syn* ($\chi \approx 60°$) (42). As a result, when small molecules are forced to interact with the 1×1 UU internal loops as in DynaD using a distance-based reaction coordinate, uridine residues will tend to stay in the *anti* orientations, causing a limited number of *syn* states being investigated by DynaD compared with DynaD/Auto. To quantify the results, we divided the 2D surface into two categories representing *anti* and *syn* conformations. DynaD/Auto generally sampled a wider *anti* region, except for **A**, **D**, **E**, **F**, and **G**, where the *anti* region was sampled equally by both methods (Fig. 5 *d*). Nevertheless, the *syn* region was sampled much better by DynaD/Auto regardless of the type of small molecule studied, where on average 5% and 2% of regions representing

syn orientations were sampled by DynaD/Auto and DynaD, respectively (Fig. 5 *d*). Analyses of the 1×1 UU internal loop conformations observed in the global minima structures predicted by DynaD/Auto and DynaD, respectively, show five and two systems out of nine having *syn-anti* UU orientations (Table 1), which implies that DynaD cannot capture the *syn* conformations as much as DynaD/Auto. Results suggest that if prior knowledge about an RNA target site is available, DynaD/Auto can yield a better prediction than DynaD.

Conformational flexibility of small molecules can hinder convergence

RNA internal loops are typically dynamic, which is one of the reasons why we utilized 66 different 1×1 UU internal loop conformations in the DynaD/Auto approach to overcome the sampling problem while determining the global minimum bound states. Even though conformational flexibility of an RNA-binding site is an important factor in deciding the efficiency of the binding study, conformational flexibility of small molecules can have a significant impact in efficient sampling of conformational space.

The flexibility of small molecules can be quantified by finding the total number of unique conformations possessed by each small molecule in apo state. As described above, we

studied the dynamics of each small molecule in explicit solvent followed by cluster analyses to determine the total number of unique conformations preferred by each (Table 2). While we found that **A**, **B**, **H**, and **I** have fewer than four unique conformations representing their apo states, **C**, **D**, **F**, and **G** are observed to have over 13 unique conformations implying dynamics. The flexibility of a small molecule is partly related to its molecular size, whereby small-sized molecules usually tend to be stiff due to the reduced rotatable bonds such as the cases in **B**, **H**, and **I** (Fig. 2 b). In the case of **A**, however, even though the molecule has several rotatable bonds connecting two benzimidazole rings, **A** in apo form displays stacked conformations and, therefore, rigid-like behavior. Rigidification of molecular structures of small molecules creates lower SASAs (Table 2), which in the binding process allows the RNA-binding site to sample a wider conformational space while searching for the global minimum bound states in small molecule/RNA CUG complexes (Fig. 5 c). For example, in the binding process of **B**, **H**, and **I**, DynaD/Auto samples over 26% of the available 2D (χ,θ) space, while less than 18% of the 2D (χ,θ) space was sampled in **C**, **D**, **F**, and **G** (Fig. 5 c). Undersampling of the 2D (χ,θ) space can hinder the determination of binding modes, especially the global minimum bound state, of a small molecule/RNA complex. For example, in the case of **D**, DynaD and DynaD/Auto sampled around 9% and 14%, respectively, of the 2D (χ,θ) space (Fig. 5 c), which is the reason why we predict $\Delta\Delta G = -3.17$ kcal/mol (Table 1). Similar results were also observed in **C** (Table 1 and Fig. 5 c). Thus, conformational flexibility of small molecules should be taken into serious consideration while investigating the binding process of a small molecule targeting a dynamic RNA site.

van der Waals interactions are deciding the fate of small molecule/RNA CUG binding

In a binding process, the nonbonded interactions are going to direct what the governing forces are. The binding free energy has four nonbonded terms, which will have crucial roles in binding: van der Waals (vdW), electrostatics, and polar and nonpolar (apolar) solvation free energies. We decomposed the binding free energies (ΔG) into these nonbonded terms for all the clusters predicted for small molecules **A** to **I** (Fig. 2 b). Correlation analyses showed that almost all the small molecules have strong positive correlations between vdW and ΔG (Fig. S6). This result is due to the small molecules preferring intercalated/stacked states while targeting the RNA-binding site and indicates the importance of vdW interactions in the binding process (Fig. S6). Furthermore, relatively weak negative correlations between ΔG_{solv} and ΔG were observed in almost all the cases, where ΔG_{solv} is the sum of polar and apolar solvation free energies ($\Delta G_{polar} + \Delta G_{apolar}$). This result is, again, due to the small molecules preferring intercalated/stacked states, which disfavor polar and apolar interactions.

Dendrogram analysis of small molecule/RNA CUG interaction presents a unique way to visualize bound states predicted by DynaD and DynaD/Auto

Dendrograms (tree representations) can be used as a visualization method whereby both the relationships between the clustered conformations and corresponding binding energies within the data can be seen in a simple way. Both DynaD and DynaD/Auto predict the global minimum conformation as well as many local minimum conformations, which can be connected based on structural similarity to draw a picture of the potential structural transformations between predicted bound states. Fig. 6 a is an example dendrogram representing the results for **D** targeting RNA CUG repeats using DynaD/Auto. Dendrograms of other small molecules are displayed in Figs. S7–S14. The y axis in a dendrogram represents the index numbers of the clusters while the x axis corresponds to half of the average deviation of RMSD calculated between two structures, which can be used as a measure of how different two selected bound states are (see supporting material for details). Depending on the system, if the RMSD value of a structure with respect to another is less than 1.0, we may consider those two bound states structurally similar. Thus, we display the gray dotted line on the dendrogram representing an average RMSD value of $2 \times 0.5 = 1.0$ Å to highlight that clusters connected with RMSD/2 $\leq$ 0.5 Å are structurally alike (Fig. 6 a). Furthermore, relative binding free energies of each bound states with respect to global minimum are displayed on the left-hand side of each dendrogram as yellow bars to characterize each state by their binding free energies and structural similarities. In Fig. 6 a, the largest RMSD deviation from the global minimum is $2 \times 3.06 = 6.12$ Å, representing cluster #158 located at the very top of the dendrogram. As can be easily verified, structural heterogeneity of cluster #158 as well as the binding free energy gap compared with global minimum is very large, implying different binding properties. One of the uses of dendrograms is that they allow one to find potential transformation pathways between two bound states with intermediate states to offer a transformation mechanism. By following the connected structures on the dendrogram and using decrease in energy as a guide, one can easily deduce a possible transformation between any two bound states. To demonstrate this, a path highlighted in red in Fig. 6 a displays a path between cluster #2442 and the global minimum (cluster #2649), where bound states shown in Fig. 6 b corresponding to clusters #2442, #2163, #490, and #2649 highlight the states followed by a transformation from cluster #2442 to cluster #2649. Cluster 2442 has one of its uridines flipped out at the major groove side interacting with the bromobenzene
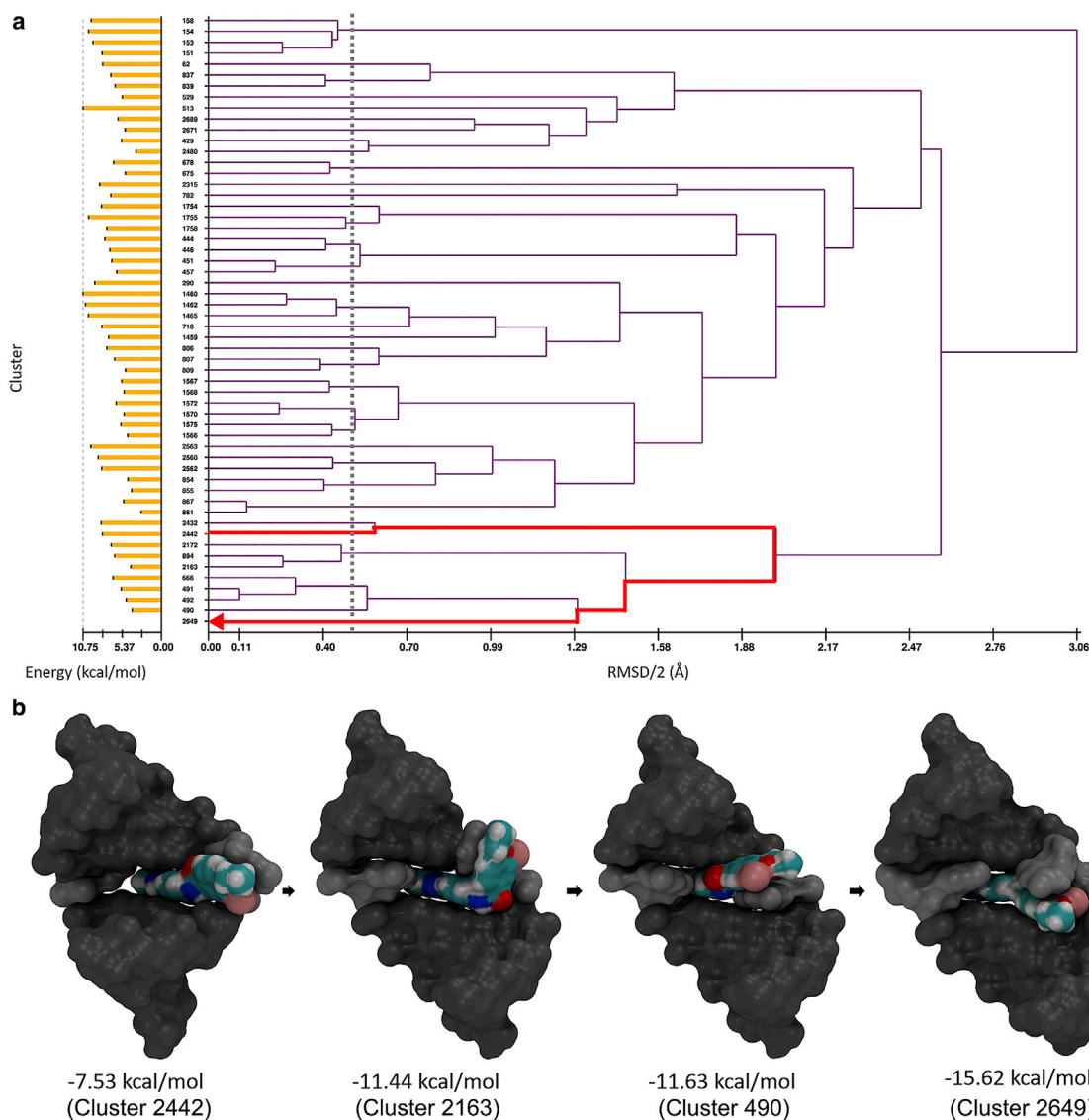
FIGURE 6 Predicted results of **D**/RNA C<u>U</u>G binding using DynaD/Auto. (*a*) Dendrogram analysis displaying how predicted bound states are connected to each other, based on the structural similarity with respect to RMSD. In (*a*), *x* and *y* axes represent RMSD/2 and bound states extracted from cluster analyses, respectively. Yellow bars left of each index represent the relative binding free energies with respect to global minimum, which is the lowermost cluster in the figure (Table S20). The gray vertical dashed line is drawn at $x = 0.5$ to indicate clusters with RMSD $< 2 \times 0.5 = 1.0$ Å, implying relatively similar bound states. (*b*) An example of a structural transformation highlighted in red in (*a*), which displays the use of dendrogram analysis to visualize potential structural transformation observed in the small molecule/RNA-binding phenomenon. In (*b*), bound states of clusters #2442, #2163, #490, and #2649 are displayed from left to right. Molecular surface representation is used to highlight individual regions in **D**/RNA C<u>U</u>G complex, where solid colors are used to represent RNA regions. In (*b*), gray- and silver-colored regions represent the RNA and 1×1 UU internal loops, respectively, while default atom colors are used in **D**. Note that in (*b*), bromine of bromobenzene ring is highlighted in pink. To see this figure in color, go online.

ring of **D**. In clusters #2163 and #490, uridine reorients itself in such a way that **D** moves toward the helical axis to increase intercalation with the RNA C<u>U</u>G binding site. Finally, in cluster #2649, both uridine and **D** optimize their conformation to maximize binding. Using the simplistic display of the dendrograms, one can easily visualize various transformations in the small molecular/RNA-binding process. Furthermore, the structural data representing bound states of a small molecule/RNA binding combined with po-

tential pathways extracted from dendrogram analyses can be used in future studies to investigate the fine details of how small molecules behave near RNA-binding sites.

As described above, the binding phenomenon of a small molecule targeting an RNA loop is very complex, which can dramatically change both the conformation of the small molecule and RNA loop. During the binding process, multiple different conformations representing stable bound states can be observed (Figs. 6 and S7–S14). While it is

determined that the global minimum state has the lowest binding free energy, it is possible that there are local minima with binding free energies close to the global minimum representing alternative binding mechanisms. For example, in the studies of **A**/RNA C$\underline{U}$G binding there are at least two local minima, which are structurally different than global minimum but have binding free energies close to the global minimum (Fig. S7, clusters #143 and #804). Similar observations are also made in the studies of **B**/RNA C$\underline{U}$G (Fig. S8, cluster #1490), **E**/RNA C$\underline{U}$G (Fig. S10, cluster #193), **H**/RNA C$\underline{U}$G (Fig. S13, cluster #93), and **I**/RNA C$\underline{U}$G (Fig. S14, cluster #552) binding. These results can be crucial while optimizing a small molecule to enhance binding affinities.

How can DynaD and DynaD/Auto be extended to study other systems?

Compared with well-structured RNA stem regions, RNA loop motifs are dynamic, and thus can be targeted with small molecules and ligands. To find the binding modes of a small molecule to a dynamic RNA loop motif, one needs to sample the conformational space sufficiently. A single conformational state representing the RNA loop motif will not guarantee the discovery of the global minimum bound state. DynaD and DynaD/Auto try to overcome this issue by creating initial bound states close to the local and global minima states. As described above, the general procedures of DynaD and DynaD/Auto can be described as first finding initial bound states for small molecule/RNA complex and then investigating these states using explicit solvent MD simulations. While DynaD uses a distance-based reaction coordinate to find the initial bound states, DynaD/Auto combines the predicted energy landscape of an RNA-binding site with the AutoDock Vina calculations. The initial bound states can be pictured as structures with small molecules placed nearby an RNA target site, which is reasonable enough to study the binding phenomenon. If there are no energetics data available to the RNA-binding site, one can utilize DynaD to create the initial bound states by repeatedly forcing the small molecule to interact with the RNA-binding site using a distance-based reaction coordinate. If there are energetics data available, one can extract the unique RNA loop conformations and then utilize AutoDock Vina calculations to create the initial bound states. As an example, if one wants to investigate how a small molecule would target a tetraloop RNA hairpin system, one can design a model system for the RNA hairpin structure and utilize the DynaD approach. A forced binding process can be then followed by moving the small molecule to the RNA hairpin site, where everything except the tetraloop RNA residues is restrained to stay in an A-form orientation. Once enough initial bound states are created, one can then run conventional MD simulations to scan the conformational space. Cluster analyses and MM/3D-RISM calculations will then yield the local

and global minimum binding modes. If, however, one has energetics data available for the tetraloop RNA hairpin system such as energy landscapes predicted by umbrella sampling (51,54), replica exchange (55,56), and/or discrete path sampling calculations (43,51,57,58), the DynaD/Auto approach can be followed to investigate the binding phenomenon. Finally, it is worthy of mention that DynaD and DynaD/Auto are not limited to the study of only RNA systems. Because DynaD and DynaD/Auto try to exploit the physical properties of ligands and their targets, these two methods can be utilized to investigate small molecules and ligands interacting with proteins, DNA, carbohydrates, and lipids.

CONCLUSION

As described above, while there are several in silico molecular docking programs available in the literature to study ligand-receptor interactions, they have specific limitations while predicting bound states. Most of these programs rigidify the ligand and/or the receptor, which will not work when dynamic regions such as RNA loop motifs are targeted. Furthermore, most of the docking software utilizes energy-based scoring functions, which can cause challenges while different receptor conformations are targeted with ligands, as the scores are specifically calculated to the receptor conformation. Finally, use of knowledge-based tools performed by some docking programs can miss the binding poses if they are not in the database. In this contribution, we describe two dynamic docking methods, DynaD and DynaD/Auto, and apply them to nine small molecules targeting RNA 1×1 UU internal loops in RNA C$\underline{U}$G repeat expansions causing DM1. While DynaD utilizes a distance-based reaction coordinate to create initial bound states, DynaD/Auto combines the results of 2D (χ,θ) energy landscape predicted by umbrella sampling calculations with AutoDock Vina calculations to create the initial bound states. Both methods then follow the same methodology, whereby explicit solvent MD simulations are performed on each initial bound state followed by cluster analyses performed on combined MD trajectories, and finally MM/3D-RISM calculations to determine the global minimum bound states. Predictions are then compared with experimental data, where we observe a positive correlation between the predictions and experiments with correlation coefficient (R) values of 0.70 and 0.81 for DynaD and DynaD/Auto, respectively. Comparison of different binding free energy calculations show that the best correlation between predictions and experiments is achieved with MM/3D-RISM compared with MM/GBSA and MM/PBSA, owing to the inclusion of solvation in MM/3D-RISM calculations.

If no prior knowledge about the structure is available for an RNA-binding site, forced targeting can be achieved by DynaD, which is particularly useful in determining the

bound states even in bigger systems. If the binding site is known, DynaD can be utilized to target the site with a small molecule continuously to create initial bound states, which then can be utilized in explicit solvent MD simulations. For example, DynaD can be utilized to study how lead compounds can target different sites in RNA riboswitches, which then can be used to optimize the lead compounds. If, however, prior information is available about the RNA-binding site, such as free energy landscape data, accurate calculations are possible as displayed by DynaD/Auto calculations. We show that by following the DynaD/Auto approach, we can improve the correlation between the predictions and experimental results in the studies of small molecule/RNA CUG binding. The use of prior knowledge about the RNA CUG binding site overcomes the sampling problem generally observed in challenging systems, such as 1×1 UU internal loops in RNA CUG repeats, and thus helps determination of the global minimum bound state. The challenge is partly due to the dynamics of the RNA-binding site and the flexibility of small molecules. Thus, both DynaD and DynaD/Auto will converge to the same result if conformational space is sampled sufficiently, as observed in several systems we studied. Furthermore, the binding studies produce more than a dozen bound states, for which the structural connectivity between each other is difficult to present in traditional ways. Therefore, we designed a simple diagram using dendrograms, which are often used in gene network analysis. These dendrograms display all the bound states and how they are connected to each other using RMSD as the metric, which helps one to see the bigger picture of a small molecule when targeting an RNA site. The two dynamic docking methods we describe in this article, DynaD and DynaD/Auto, will allow researchers to perform in silico calculations using two physics-based approaches for small molecules targeting dynamic RNA loops, and will allow us to rethink the traditional assumption that docking is related to the global minimum conformation of a binding site obtained from experiments.

Data and software availability

The data set representing the predicted structures using DynaD and DynaD/Auto can be found at https://cescos. fau.edu/~iyildirim/rna_cug_compounds.tar.gz.

SUPPORTING MATERIAL

AUTHOR CONTRIBUTIONS

I.Y. designed the project. K.W.W. carried out the MD simulations and performed the analyses. I.R. and J.D. performed the dendrogram analysis. K.W.W., I.R., J.D., and I.Y. wrote the manuscript.

ACKNOWLEDGMENTS

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Ashley, C. T., Jr., and S. T. Warren. 1995. Trinucleotide repeat expansion and human disease. *Annu. Rev. Genet.* 29:703–728. https://doi.org/10.1146/annurev.ge.29.120195.003415.

2. Emery, A. E. H. 2002. The muscular dystrophies. *Lancet.* 359:687–695. https://doi.org/10.1016/S0140-6736(02)07815-7.

3. Kino, Y., D. Mori, …, S. Ishiura. 2004. Muscleblind protein, MBNL1/EXP, binds specifically to CHHG repeats. *Hum. Mol. Genet.* 13:495–507. https://doi.org/10.1093/hmg/ddh056.

4. Kino, Y., C. Washizu, …, S. Ishiura. 2009. MBNL and CELF proteins regulate alternative splicing of the skeletal muscle chloride channel CLCN1. *Nucleic Acids Res.* 37:6477–6490. https://doi.org/10.1093/nar/gkp681.

5. Miller, J. W., C. R. Urbinati, …, M. S. Swanson. 2000. Recruitment of human muscleblind proteins to $(CUG)_n$ expansions associated with myotonic dystrophy. *EMBO J.* 19:4439–4448. https://doi.org/10.1093/emboj/19.17.4439.

6. Warf, M. B., and J. A. Berglund. 2007. MBNL binds similar RNA structures in the CUG repeats of myotonic dystrophy and its pre-mRNA substrate cardiac troponin T. *RNA.* 13:2238–2251. https://doi.org/10.1261/rna.610607.

7. Warf, M. B., J. V. Diegel, …, J. A. Berglund. 2009. The protein factors MBNL1 and U2AF65 bind alternative RNA structures to regulate splicing. *Proc. Natl. Acad. Sci. USA.* 106:9203–9208. https://doi.org/10.1073/pnas.0900342106.

8. Yuan, Y., S. A. Compton, …, M. S. Swanson. 2007. Muscleblind-like 1 interacts with RNA hairpins in splicing target and pathogenic RNAs. *Nucleic Acids Res.* 35:5474–5486. https://doi.org/10.1093/nar/gkm601.

9. Itskovich, S. S., A. Gurunathan, …, L. H. Lee. 2020. MBNL1 regulates essential alternative RNA splicing patterns in MLL-rearranged leukemia. *Nat. Commun.* 11:2369. https://doi.org/10.1038/s41467-020-15733-8.

10. Kanadia, R. N., K. A. Johnstone, …, M. S. Swanson. 2003. A muscle-blind knockout model for myotonic dystrophy. *Science.* 302:1978–1980. https://doi.org/10.1126/science.1088583.

11. Konieczny, P., E. Stepniak-Konieczna, and K. Sobczak. 2014. MBNL proteins and their target RNAs, interaction and splicing regulation. *Nucleic Acids Res.* 42:10873–10887. https://doi.org/10.1093/nar/gku767.

12. Liquori, C. L., K. Ricker, …, L. P. W. Ranum. 2001. Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. *Science.* 293:864–867. https://doi.org/10.1126/science.1062125.

13. Prior, T. W.; On behalf of the American College of Medical Genetics Laboratory Quality Assurance C. 2009. Technical standards and guidelines for myotonic dystrophy type 1 testing. *Genet. Med.* 11:552–555. https://doi.org/10.1097/GIM.0b013e3181abce0f.

14. Yum, K., E. T. Wang, and A. Kalsotra. 2017. Myotonic dystrophy: disease repeat range, penetrance, age of onset, and relationship between repeat size and phenotypes. *Curr. Opin. Genet. Dev.* 44:30–37. https://doi.org/10.1016/j.gde.2017.01.007.

15. Childs-Disney, J. L., E. Stepniak-Konieczna, …, M. D. Disney. 2013. Induction and reversal of myotonic dystrophy type 1 pre-mRNA splicing defects by small molecules. *Nat. Commun.* 4. https://doi.org/10.1038/ncomms3044.

16. Childs-Disney, J. L., I. Yildirim, …, M. D. Disney. 2014. Structure of the myotonic dystrophy type 2 RNA and designed small molecules that reduce toxicity. *ACS Chem. Biol.* 9:538–550. https://doi.org/10.1021/cb4007387.

17. Philips, A., K. Milanowska, …, J. M. Bujnicki. 2013. LigandRNA: computational predictor of RNA-ligand interactions. *RNA.* 19:1605–1616. https://doi.org/10.1261/rna.039834.113.

18. Trott, O., and A. J. Olson. 2010. Software news and update AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* 31:455–461. https://doi.org/10.1002/jcc.21334.

19. Friesner, R. A., J. L. Banks, …, P. S. Shenkin. 2004. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* 47:1739–1749. https://doi.org/10.1021/jm0306430.

20. Stefaniak, F., and J. M. Bujnicki. 2021. AnnapuRNA: a scoring function for predicting RNA-small molecule binding poses. *PLoS Comput. Biol.* 17, e1008309. https://doi.org/10.1371/journal.pcbi.1008309.

21. Sun, L.-Z., Y. Jiang, …, S.-J. Chen. 2020. RLDOCK: a new method for predicting RNA–ligand interactions. *J. Chem. Theor. Comput.* 16:7173–7183. https://doi.org/10.1021/acs.jctc.0c00798.

22. Genheden, S., and U. Ryde. 2015. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expet Opin. Drug Discov.* 10:449–461. https://doi.org/10.1517/17460441.2015.1032936.

23. Stelzer, A. C., A. T. Frank, …, H. M. Al-Hashimi. 2011. Discovery of selective bioactive small molecules by targeting an RNA dynamic ensemble. *Nat. Chem. Biol.* 7:553–559. https://doi.org/10.1038/nchembio.596.

24. Grubmüller, H., B. Heymann, and P. Tavan. 1996. Ligand binding: molecular mechanics calculation of the streptavidin-biotin rupture force. *Science.* 271:997–999. https://doi.org/10.1126/science.271.5251.997.

25. Guilbert, C., and T. L. James. 2008. Docking to RNA via root-mean-square-deviation-driven energy minimization with flexible ligands and flexible targets. *J. Chem. Inf. Model.* 48:1257–1268. https://doi.org/10.1021/ci8000327.

26. Bissaro, M., M. Sturlese, and S. Moro. 2020. Exploring the RNA-recognition mechanism using supervised molecular dynamics (SuMD) simulations: toward a rational design for ribonucleic-targeting molecules? *Front. Chem.* 8. https://doi.org/10.3389/fchem.2020.00107.

27. Benhamou, R. I., S. Vezina-Dawod, …, M. D. Disney. 2020. Macrocyclization of a ligand targeting a toxic RNA dramatically improves potency. *Chembiochem.* 21:3229–3233. https://doi.org/10.1002/cbic.202000445.

28. Angelbello, A. J., R. I. Benhamou, …, M. D. Disney. 2021. A small molecule that binds an RNA repeat expansion stimulates its decay via the exosome complex. *Cell Chem. Biol.* 28:34–45.e36. https://doi.org/10.1016/j.chembiol.2020.10.007.

29. Vezina-Dawod, S., A. J. Angelbello, …, M. D. Disney. 2021. Massively parallel optimization of the linker domain in small molecule dimers targeting a toxic r(CUG) repeat expansion. *ACS Med. Chem. Lett.* 12:907–914. https://doi.org/10.1021/acsmedchemlett.1c00027.

30. Wang, Z.-F., A. Ursu, …, M. D. Disney. 2019. The hairpin form of r(G(4)C(2))(exp) in c9ALS/FTD is repeat-associated non-ATG translated and a target for bioactive small molecules. *Cell Chem. Biol.* 26:179. https://doi.org/10.1016/j.chembiol.2018.10.018.

31. Ursu, A., K. W. Wang, …, M. D. Disney. 2020. Structural features of small molecules targeting the RNA repeat expansion that causes genetically defined ALS/FTD. *ACS Chem. Biol.* 15:3112–3123. https://doi.org/10.1021/acschembio.0c00049.

32. Bush, J. A., H. Aikawa, …, M. D. Disney. 2021. Ribonuclease recruitment using a small molecule reduced c9ALS/FTD r(G4C2) repeat expansion in vitro and in vivo ALS models. *Sci. Transl. Med.* 13, eabd5991. https://doi.org/10.1126/scitranslmed.abd5991.

33. Costales, M. G., H. Aikawa, …, M. D. Disney. 2020. Small-molecule targeted recruitment of a nuclease to cleave an oncogenic RNA in a mouse model of metastatic cancer. *Proc. Natl. Acad. Sci. USA.* 117:2406–2411. https://doi.org/10.1073/pnas.1914286117.

34. Suresh, B. M., W. C. Li, …, M. D. Disney. 2020. A general fragment-based approach to identify and optimize bioactive ligands targeting RNA. *Proc. Natl. Acad. Sci. USA.* 117:33197–33203. https://doi.org/10.1073/pnas.2012217117.

35. Miller, B. R., T. D. McGee, …, A. E. Roitberg. 2012. MMPBSA.py: an efficient program for end-state free energy calculations. *J. Chem. Theor. Comput.* 8:3314–3321. https://doi.org/10.1021/ct300418h.

36. Luchko, T., S. Gusarov, …, A. Kovalenko. 2010. Three-dimensional molecular theory of solvation coupled with molecular dynamics in amber. *J. Chem. Theor. Comput.* 6:607–624. https://doi.org/10.1021/ct900460m.

37. Genheden, S., T. Luchko, …, U. Ryde. 2010. An MM/3D-RISM approach for ligand binding affinities. *J. Phys. Chem. B.* 114:8505–8516. https://doi.org/10.1021/jp101461s.

38. Rzuczek, S. G., M. R. Southern, and M. D. Disney. 2015. Studying a drug-like, RNA-focused small molecule library identifies compounds that inhibit RNA toxicity in myotonic dystrophy. *ACS Chem. Biol.* 10:2706–2715. https://doi.org/10.1021/acschembio.5b00430.

39. Onufriev, A., D. Bashford, and D. A. Case. 2004. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins.* 55:383–394. https://doi.org/10.1002/prot.20033.

40. Case, D. A., I. Y. Ben-Shalom, …, P. A. Kollman. 2018. AMBER 18. University of California, San Francisco, CA.

41. Cornell, W. D., P. Cieplak, …, P. A. Kollman. 1995. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* 117:5179–5197. https://doi.org/10.1021/ja00124a002.

42. Yildirim, I., H. A. Stern, …, D. H. Turner. 2010. Reparameterization of RNA χ torsion parameters for the AMBER force field and comparison to NMR spectra for cytidine and uridine. *J. Chem. Theor. Comput.* 6:1520–1531. https://doi.org/10.1021/ct900604a.

43. Wales, D. J., and I. Yildirim. 2017. Improving computational predictions of single-stranded RNA tetramers with revised α/γ torsional parameters for the amber force field. *J. Phys. Chem. B.* 121:2989–2999. https://doi.org/10.1021/acs.jpcb.7b00819.

44. Wang, J. M., R. M. Wolf, …, D. A. Case. 2005. Development and testing of a general amber force field (vol 25, pg 1157, 2004). *J. Comput. Chem.* 26:114. https://doi.org/10.1002/jcc.20145.

45. Bayly, C. I., P. Cieplak, …, P. A. Kollman. 1993. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges - the Resp Model. *J. Phys. Chem.* 97:10269–10280. https://doi.org/10.1021/j100142a004.

46. Cornell, W. D., P. Cieplak, …, P. A. Kollman. 1993. Application of RESP charges to calculate conformational energies, hydrogen-bond energies, and free-energies of solvation. *J. Am. Chem. Soc.* 115:9620–9631. https://doi.org/10.1021/ja00074a030.

47. Frisch, M. J., G. W. Trucks, …, D. J. Fox. 2016. Gaussian 09. Revision A, 02 edn. Gaussian, Inc., Wallingford, CT.

48. Gilbert, S. D., C. E. Love, …, R. T. Batey. 2007. Mutational analysis of the purine riboswitch aptamer domain. *Biochemistry.* 46:13297–13309. https://doi.org/10.1021/bi700410g.

49. Jorgensen, W. L., J. Chandrasekhar, …, M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935. https://doi.org/10.1063/1.445869.

50. Joung, I. S., and T. E. Cheatham. 2008. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B.* 112:9020–9041. https://doi.org/10.1021/jp8001614.

51. Yildirim, I., D. Chakraborty, …, G. C. Schatz. 2015. Computational investigation of RNA CUG repeats responsible for myotonic dystrophy 1. *J. Chem. Theor. Comput.* 11:4943–4958. https://doi.org/10.1021/acs.jctc.5b00728.

52. Morris, G. M., R. Huey, …, A. J. Olson. 2009. AutoDock4 and Auto-DockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.* 30:2785–2791. https://doi.org/10.1002/jcc.21256.

53. Genheden, S., O. Kuhn, …, U. Ryde. 2012. The normal-mode entropy in the MM/GBSA method: effect of system truncation, buffer region, and dielectric constant. *J. Chem. Inf. Model.* 52:2079–2088. https://doi.org/10.1021/ci3001919.

54. Yildirim, I., H. Park, …, G. C. Schatz. 2013. A dynamic structural model of expanded RNA CAG repeats: a refined X-ray structure and computational investigations using Molecular Dynamics and Umbrella Sampling simulations. *J. Am. Chem. Soc.* 135:3528–3538. https://doi.org/10.1021/ja3108627.

55. Nymeyer, H., S. Gnanakaran, and A. E. García. 2004. Atomic simulations of protein folding, using the replica exchange algorithm. *In* Methods Enzymol., *vol 383* Academic Press, pp. 119–149. https://doi.org/10.1016/S0076-6879(04)83006-4.

56. Mitsutake, A., Y. Sugita, and Y. Okamoto. 2001. Generalized-ensemble algorithms for molecular simulations of biopolymers. *Peptide Science.* 60 (2):96–123. https://doi.org/10.1002/1097-0282(2001)60.

57. Taghavi, A., I. Riveros, …, I. Yildirim. 2022. Evaluating geometric definitions of stacking for RNA dinucleoside monophosphates using molecular mechanics calculations. *J. Chem. Theor. Comput.* 18:3637–3653. https://doi.org/10.1021/acs.jctc.2c00178.

58. Wales, D. J., M. D. Disney, and I. Yildirim. 2019. Computational investigation of RNA A-bulges related to the microtubule-associated protein tau causing frontotemporal dementia and parkinsonism. *J. Phys. Chem. B.* 123:57–65. https://doi.org/10.1021/acs.jpcb.8b09139.