Investigating RNA–protein recognition mechanisms through supervised molecular dynamics (SuMD) simulations

Matteo Pavan, Davide Bassani, Mattia Sturlese and Stefano Moro

Molecular Modeling Section (MMS), Department of Pharmaceutical and Pharmacological Sciences University of Padova, via Marzolo 5, 35131 Padova, Italy

Received May 23, 2022; Revised October 20, 2022; Editorial Decision November 01, 2022; Accepted November 09, 2022

ABSTRACT

Ribonucleic acid (RNA) plays a key regulatory role within the cell, cooperating with proteins to control the genome expression and several biological processes. Due to its characteristic structural features, this polymer can mold itself into different three-dimensional structures able to recognize target biomolecules with high affinity and specificity, thereby attracting the interest of drug developers and medicinal chemists. One successful example of the exploitation of RNA’s structural and functional peculiarities is represented by aptamers, a class of therapeutic and diagnostic tools that can recognize and tightly bind several pharmaceutically relevant targets, ranging from small molecules to proteins, making use of the available structural and conformational freedom to maximize the complementarity with their interacting counterparts. In this scientific work, we present the first application of Supervised Molecular Dynamics (SuMD), an enhanced sampling Molecular Dynamics-based method for the study of receptor–ligand association processes in the nanoseconds timescale, to the study of recognition pathways between RNA aptamers and proteins, elucidating the main advantages and limitations of the technique while discussing its possible role in the rational design of RNA-based therapeutics.
few thousand nucleotides, as in long non-coding sequences (9). Compared to DNA, the higher conformational freedom of ribonucleic acids implies that they can assume a wide variety of three-dimensional structures in solution (10,11), organizing themselves in functional domains specifically designed to recognize other nucleic acids (12), proteins (13), glycated derivatives (14) or small organic molecules (15).

From a functional perspective, the successful exploitation of RNA’s ability to mold itself into different three-dimensional structures able to recognize target biomolecules with high affinity and specificity is represented by aptamers (16). This class of single-stranded oligonucleotides that fold into defined and complex architectures including stems, loops, bulges, hairpins, pseudoknots, triplets or quadruplex, can bind several molecular targets, such as proteins, small organic molecules, and ions, thus classifying as a useful tool both for a therapeutic and diagnostic purpose (17,18).

As is the case for other nucleic acids, interactions between aptamers and proteins are characterized by a complex network of van der Waals, hydrogen bond, stacking, and general non-polar interactions that define the complementarity of shape and electrostatic properties at the surface between the two interactors and determine the specificity of binding to a certain target (19,20). To the present date, the rising amount of experimentally solved three-dimensional RNA structures, especially concerning their complexes with both macro and small molecules, has led to an increased interest by the scientific community in the investigation of RNA structures at an atomistic level of detail, to apply structure-based drug design (SBDD) strategies for the rational design of novel therapeutic entities (21,22).

Among the tools that are routinely used, both in the academic and industrial environment, for investigating the structural determinants of biological entities’ recognition, molecular docking is by far the most widely and successfully adopted (23). Originally developed for predicting the interaction between small organic molecules and proteins (24), throughout the years this computational technique has been also applied to the investigation of protein–protein (25), protein–peptides (26), and antigen–antibodies (27) complexes with various degrees of success. Contrary to proteins, the application of molecular docking to the study of complexes involving nucleic acids has been so far very limited, and mainly applied to the prediction of the binding mode of small molecules (28). Compared to the aforementioned established docking protocols, a smaller number of methods are available for the investigation of nucleic acids–protein complexes, due to some intrinsic structural peculiarities of nucleic acids, particularly in the case of RNA (29). The first limitation is represented by the distinctive charge distribution that characterizes the RNA surface compared to that of proteins (30), the second one is portrayed by the neglected treatment of the role of solvent (31) and the third one is the lack or limited consideration of the structural flexibility and dynamicity of ribonucleic acids (32).

One possible approach to overcome the limitations of molecular docking is represented by molecular dynamics (MD) simulations. Despite the enhanced description of the binding event derived from the explicit treatment of solvent molecules and the consideration of both receptor and ligand flexibility, MD simulations are rarely carried out to investigate the whole binding event due to the long simulation times and computational effort required to sample these infrequent events and are therefore mostly exploited for the refinement of docking results (33).

To mitigate the time constraints of classic molecular dynamics, one possible strategy is the exploitation of enhanced sampling techniques that allow increasing the frequency of observing desired events, such as the receptor–ligand association (34). Among the plethora of enhanced sampling protocols that have been developed throughout the years, supervised molecular dynamics (SuMD) has proven to be particularly successful in investigating ligand-receptor recognition pathways at an atomistic level of detail without applying any energetic bias to the system while contemporarily reducing by orders of magnitude the simulation time compared to classic MD simulations (35). Particularly, from a ligand perspective, SuMD simulations have proven useful to work with a variety of molecular entities, ranging from small organic molecules (both fragment (36–41) and mature, lead-like, compounds (42–44)), to more complex chemical species such as macrocycles (45) and peptide ligands (46). From a receptor point of view, instead, SuMD was successfully applied to the study of both soluble (47,48) and membrane (49–51) systems, including both protein (52,53) and nucleic (54) targets.

In this scientific work, SuMD simulations were applied for the first time to the study of the recognition process between RNA macromolecules and proteins, to extend the applicability domain of the methodology. Particularly, we decided to focus our attention on RNA aptamers, due to their relevance as diagnostic and therapeutic tools and due to the variety of their structural landscape.

Briefly, we present four different applications of the SuMD methodology to RNA aptamer–protein complexes. Three test cases, involving systems in which the three-dimensional structure of the complex is known and deposited in the Protein Data Bank (55), are used to validate the ability of the SuMD protocol to correctly reproduce the experimental data while giving additional useful information that goes beyond the final state of the recognition process. The fourth and final case, concerning instead a complex whose structure has not yet been experimentally determined, is presented to show a possible prospective application of the SuMD protocol, discussing at the same time the advantages and limitations of the technique other than its possible role in a typical drug discovery pipeline.

**MATERIALS AND METHODS**

**Hardware overview**

General molecular modeling operations, such as the preparation of RNA aptamer–protein complex structures, the system setup for molecular dynamics simulations, and trajectory analysis were performed on an 8 CPU Linux workstation equipped with an Intel Xeon E5-1620 3.50 GHz processor. All molecular dynamics simulations were performed on a GPU cluster composed of 20 NVIDIA devices ranging from GTX980 to Titan V.
Structures preparation

The three-dimensional coordinates of the three RNA aptamer–protein complex used as control cases in this study were retrieved from the Protein Data Bank (PDB ID: 3DD2, 4PDB, 5VOE) and prepared for subsequent simulations exploiting several modules from the Molecular Operating Environment (MOE) 2019.01 suite (57). At first, structures were pre-processed through the ‘Structure Preparations’ tool, to assign each residue with alternate conformations to the one with the highest occupancy, build missing loops through homology modeling, and correct inconsistencies between the primary sequence and the tertiary structure. For compatibility with each piece of software used in this work and for consistency with previous studies involving the SuMD approach, structures were manually edited to mutate each non-natural nucleic residue (e.g. fluorinated nucleotides) to the corresponding natural alternative. Afterward, titratable residues were assigned to the most probable protonation state at pH = 7.40 exploiting the ‘Protonate3D’ tool. Finally, each non-protein or non-nucleic residue was removed, and the nucleic ligand was moved away from the binding site at a distance of at least 30 Å from the nearest receptor atom, to explore the conformational degree of freedom of the ligand throughout the recognition process.

For the investigation of the recognition process between the SARS-CoV-2 Spike RNA aptamer, the structure of the Spike Receptor Binding Domain (RBD) was retrieved from the Protein Data Bank (accession code: 6M0J (58)). Concerning the RNA aptamer, the experimental structure was not available, therefore the primary sequence was obtained from the Supplementary Material of the original work from Valero et al. (59) and sequentially submitted to the NUPACK (60) and 3DRNA (61) web servers to retrieve the predicted secondary and tertiary structure respectively. In both cases, default parameters were chosen, and the lowest energy structure coming from the 3DRNA webserver was used for both docking calculations and MD simulations. Every other preparation passage is identical to the ones executed for the three control cases.

System setup and equilibration protocol

Nucleic-protein systems coming from the preliminary preparation stage were then further processed making use of both visual molecular dynamics (VMD) 1.9.2 (62) and several tools from the Ambertools14 suite (63). Each protein or nucleic atom was parameterized according to the ff14SB force field with χ modification tuned for RNA (XOL3) (64–66). At first, each system was solvated in a cubic box of TIP3P (67) water molecules with a padding of 35 Å. Afterward, each solvated system was neutralized through the addition of an appropriate number of Na+ and Cl− counterions until a salt concentration of 0.154 M was reached. Before molecular dynamics (MD) simulations, each system was subjected to a 1500-step energy minimization phase with the conjugate-gradient algorithm.

Each minimized system underwent then a two stages equilibration protocol. The first stage consisted of 1 ns of simulation in the canonical ensemble (NVT), applying a 5 kcal mol−1 Å−2 harmonic positional restraint on each protein and nucleic atom. The second stage consisted instead of a 2 ns simulation in the isothermal-isobaric ensemble (NPT), with the same restraints applied only to the backbone atoms of both the protein and the nucleic acid. For each MD simulation performed in this scientific work, an integration step of 2 fs was used, the temperature was kept at a constant value of 310 K through a Langevin thermostat (68), and the M-SHAKE algorithm (69) was used to constrain the length of bonds involving hydrogen atoms, the particle-mesh Ewald (PME) (70) method was exploited to compute electrostatic interactions using cubic spline interpolation and a 1 Å grid spacing, while a 9.0 Å cutoff was set for calculation of Lennard–Jones interactions. For simulations in the NPT ensemble, the pressure was kept at a constant value of 1 atm through a Monte Carlo barostat (71). Finally, all MD simulations were run through the ACEMD 3 (72) engine, which is based on OpenMM 7 (73).

Supervised molecular dynamics (SuMD) simulations

SuMD is a well-established enhanced-sampling molecular dynamics approach that has been successfully applied to the study of the recognition process between various molecular entities at an atomic level of details on the nanosecond timescale (35).

The main advantage of the SuMD approach compared to traditional molecular dynamics simulations is the improved ability to sample infrequent events such as molecular association processes, thus reducing the timescale of the simulation that is required to spontaneously observe a binding event from the microseconds range to a few nanoseconds.

In detail, this task is accomplished by performing a sequence of short, unbiased, MD simulations followed by an evaluation of the simulation progress by a tabu-like algorithm. In this case, each of these MD simulations, defined as the ‘SuMD-step’, is run in the canonical ensemble at a constant temperature of 310 K for 300 ps, but the length of the ‘SuMD-step’ can vary and is chosen according to the system that is studied. At the end of each ‘SuMD-step’, the distance between the center of mass of the ligand and one of the user-defined binding sites is computed at each step of the simulation, and this data is then fitted into a linear function: if the slope of the resulting straight-line is negative, indicating that the ligand is approaching the binding site, the ‘SuMD-step’ is considered productive and retained for the generation of the final trajectory, while the final state of the simulation is used as the starting point for the successive step. On the contrary, if the slope is positive, thereby indicating that the ligand is not approaching the binding site, the ‘SuMD-step’ is considered not productive and therefore discarded: in this case, the step is repeated by randomly reassigning the particle velocities through the Langevin thermostat and retaining the final coordinates from the end of the previous ‘SuMD-step’. The supervision algorithm is switched off when the distance between the two centers of mass falls below a threshold value (10 Å, in this case): from that point on, the simulation proceeds for the other 10 ns of classic molecular dynamics, allowing the system to relax and reach the final state of the simulation without any external geometric biased imposed by the supervision. For each control case study, 10 SuMD simulations were collected: while every single one was visually inspected, only the best one according
to the geometrical agreement with the reference (based on the RMSD between the ligand coordinates in the final step of the simulation and the ligand coordinate in the reference experimental structure after optimal superposition of the protein backbone) was thoroughly analyzed and discussed in the manuscript. In the case where the structure of the complex was not available, the best replica was chosen based on the MMGBSA interaction energy instead. In the current implementation, the SuMD code is written in Python and exploits the Numpy and ProDy (74) modules to perform the aforementioned geometrical supervision throughout the simulation. A list of residues utilized to define the ligand and protein binding site for each case is provided in Supplementary Table S4 (Supplementary Materials).

**Trajectory analysis**

The SuMD trajectories were analyzed by making use of an in-house tool written in Python 3 which represents an evolution and customization of the original one which is described in the work of Salmaso et al. (46).

Initially, trajectories representing each single ‘SuMD-step’ were merged into a single collective trajectory. Then, obtained trajectories were pre-processed by applying a stride and retaining one frame every 20 ps, superposing and aligning each frame on the protein backbone atoms of the first frame, and wrapping it into an image of the system simulated under periodic boundary conditions (PBC). Both geometric and energetic analyses were performed on the so-obtained SuMD trajectories.

Concerning the geometric properties of the system, regarding both the nucleic ligand and the protein receptor, the time-dependent evolution of both backbone RMSD and radius of gyration, a global and a time-dependent per-residue decomposition of the backbone RMSF were collected and reported in an aggregated panel. Furthermore, the geometric performance of the SuMD protocol in reproducing the experimental bound conformation of the ligand was evaluated by computing the ligand backbone RMSD compared to the experimental reference throughout the entire simulation. All these geometric analyses were performed making use of the appropriate functions of the MDAnalysis (75,76) Python library and plotted through the Matplotlib module.

Regarding the energetic analysis, an estimation of the ligand-receptor interaction energy alongside the SuMD trajectory was obtained both through the MMGBSA protocol, as implemented in AMBER 14, and through the ‘NAMD Energy’ plugin for VMD, which exploits the NAMD (78) package to retrieve an estimate of the interaction energy for both the receptor and the ligand and a bidimensional interaction energy matrix in which interacting residues on the ligand side are correlated with the corresponding interacting residues on the receptor side. For all these per-residue analyses, the 25 most frequently contacted residues throughout the trajectory are considered (25 for the receptor, as well as for the ligand), defining contacting residues as the ones that are at a maximum distance of 4.5 Å from the nearest atom of the counterpart, either the ligand or the receptor.

A movie representation of the trajectory alongside the dynamic evaluation of its geometric and energetic features is also provided by the same analysis tool, which exploits VMD for the visual rendering of the simulated system. For uniformity reasons, in each plot and video, residue numbering is related to the fasta sequence for the wild-type receptor, as retrieved from the UniProt database.

**Docking**

To evaluate SuMD’s ability to reproduce the native conformation of the RNA aptamer–protein complex, we decided to compare its performance with the one of molecular docking. The program chosen to accomplish this task was HADDOCK (25) (‘High Ambiguity Driven protein-protein DOCKing’, version 2.4) since it has already been extensively used for dealing with protein-nucleic acid complexes (79) and it uses a priori information to steer the docking calculation in a similar way to how SuMD works.

Each one of the crystal structures used as control (the ones coming from PDB codes 3DD2, 4PDB, and 5VOE) was subjected to a docking run. For all these cases, the nucleic acid was treated as the ligand, while the protein was considered as the receptor. The binding site was defined based on residues at the contact surface in the crystallographic structures, both on the protein and nucleic acid sides.

Concerning the SARS-CoV-2 Spike RBD RNA aptamer, the selected protein residues were chosen instead based on the contact surface with human ACE2 in the structure 6M0J, while for the aptamer the residues were selected based on the information coming from the original paper by Valero et al. The list of protein and aptamer residues used as input for each docking calculation is reported in Supplementary Table S3 (Supplementary Material). All user-definable parameters for molecular docking were kept as default.

HADDOCK starts with a randomization stage, in which the docking partners are placed far in space from one another (about 150 Å) and randomly rotated around their centers of mass. The following step consists of a rigid-body energy minimization, which is followed by the rigid-body docking of the ligand and the receptor, allowing to obtain of 1000 complexes. The 200 best solutions in terms of intermolecular energies obtained at this stage are subjected to simulated annealing refinements. Both the intra- and intermolecular energies are evaluated by HADDOCK using full electrostatic and van der Waals energy terms with an 8.5 Å distance cutoff using the OPLS (80) nonbonded parameters. The final complexes are then clustered based on the Fraction of Common Contacts (81) (FCCs) with a 0.6 cut-
off similarity for clustering, and the clusters are then ranked for energetics.

**RNA aptamer–protein complexes molecular dynamics simulations**

To evaluate the dynamic behavior of aptamer-bound protein complexes, we performed several classic molecular dynamics simulations. At first, each system was subjected to a preparation step exploiting both AmberTools14 and VMD 1.9.2, as previously mentioned in the preparation stage for SuMD simulations. Specifically, each protein-nucleic acid complex was singularly solvated in an explicit TIP3P water box with a 40 Å padding. Each of these simulation boxes was then neutralized using Na⁺/Cl⁻ ions until reaching a physiological salt concentration of 0.154 M. The preparation phase was followed by a two-step equilibration protocol. The first equilibration was carried out in the canonical ensemble (NVT) and was composed of 1500 steps of energy minimization with a conjugate-gradient algorithm followed by a 1 ns MD simulation. In this first passage, harmonic positional restraints of 5 kcal mol⁻¹ Å⁻² were applied on both the protein and the nucleic acid, while the temperature was kept at the constant value of 310 K exploiting a Langevin thermostat (with friction coefficient set to 0.1 ps⁻¹). The second equilibration was performed in the isothermal-isobaric ensemble (NPT), also this time for 1 ns of MD simulation in which the harmonic positional restraints of 5 kcal mol⁻¹ Å⁻² were applied just on the protein and nucleic acid backbones. The pressure was kept constant at the value of 1.0 atm using a Monte Carlo barostat. In each of the equilibration steps, a 2 ps integration step was adopted, the bonds involving the hydrogen atoms were constrained through the M-SHAKE algorithm, and a 9.0 Å cutoff was used for the calculation of the Lennard-Jones interaction. For the electrostatic interaction, a particle-mesh Ewald method (PME) was used. After this preparation phase, three different 50 ns replicates of classic MD simulation in the NVT ensemble at 310 K were executed.

**Free RNA–aptamer molecular dynamics simulations**

To complement the investigation of the structural dynam-icity of investigated RNA aptamer–protein complexes, we also performed a classic MD simulation of the free RNA aptamer. To accomplish this task, we retrieved the coordinates for each nucleic acid molecule from the previously mentioned complexes. Each of these aptamers was then subjected to the same protocol described before for RNA aptamer–protein complexes, except for the parts related to the protein which, in this case, was not part of the system.

**RESULTS**

To assess the applicability domain and accuracy of Supervised Molecular Dynamics simulation in the context of the nucleic acids-protein recognition processes, we opted for a retrospective validation approach, evaluating the ability of the protocol to correctly reproduce the binding mode of nucleic ligands found in experimentally solved complex structures, focusing both on the sampling and ranking capabilities of the protocol. Particularly, we decided to focus our attention on the class of RNA aptamers, both for their therapeutical relevance and for their challenging nature due to their peculiar structural features such as intrinsically higher flexibility and density of negative charge compared to ligands considered in past applications of the SuMD protocol. In the following paragraphs, we present the application of SuMD to three different case studies for which the experimental structure of the RNA aptamer–protein complex is available on the Protein Data Bank, focusing on the geometrical accuracy of the technique in reproducing the experimentally determined binding mode and monitoring both the geometric and energetic features of the recognition process, stretching beyond the final state of the simulation. The three test cases are reported in chronological order, starting from the oldest structure to the most recent one. Furthermore, we also present a prospective application of the SuMD protocol to the investigation of a complex whose structure has not yet been experimentally determined, to present and discuss the role, advantages, and limitations of implementing the SuMD protocol in a pipeline for the rational design of RNA-based therapeutics. Information about each SuMD simulation reported in the manuscript are encompassed in Supplementary Table S2 (Supplementary Material).

**RNA aptamer bound to human thrombin (PDB ID: 3DD2)**

Due to its ability to process several proteins that are part of the coagulation cascade, including the cleavage of soluble fibrinogen into fibrin, which is responsible for the formation of clots, human thrombin is a serine protease that exerts a pivotal role in blood coagulation and is, therefore, a target of interest for anticoagulation therapy (82–84). Two surface regions of thrombin (exosite-1 and exosite-2), which are located on opposite sides of the molecule and away from the catalytic site, are responsible for its ability to interact with various macromolecular substrates. Particularly, exosite-2 is responsible for the binding of thrombin to heparin, a clinically used oligosaccharide that mediates its anticoagulant effect by facilitating the interaction of thrombin with its endogenous inhibitor antithrombin (85).

In 2001, White et al. reported the discovery of Toggle-25, an RNA aptamer that was developed to bind with a high affinity to both human and porcine thrombin, leading to the inhibition of both plasma clot formation and platelet activation (86). In 2008, Long et al. were able to solve the crystal structure of Toggle-25 at, a 25 nucleotide truncated version of Toggle-25, bound to the exosite-2 of human thrombin at a resolution of 1.90 Å (PDB ID: 3DD2 (87)), thereby allowing a structural characterization of the complex which nicely complements previous biochemical analysis (87). As reported in the original publication, the aptamer recognizes the thrombin exosite-2 in its native state, since no conformational changes can be observed between the unbound and bound state. Despite a relatively simple secondary structure, defined by a stem-loop with an internal bulge, Toggle-25 can achieve a selective and high-affinity binding to human thrombin ($K_d = 0.54 \pm 0.1$ nM) thanks to the good complementarity of shape and electrostatic properties be-
tween the negatively charged aptamer and the basic protein region responsible for its recognition. The absence of significant structural alteration of the protein upon binding, the therapeutic relevance of the target, and the relatively modest size of the aptamer (25 residues), make this complex an ideal target for the application of the SuMD protocol to the study of RNA aptamer–protein interactions. As previously introduced, two simplifications were introduced in the system investigated through SuMD: firstly, each 2′ fluororo substituted pyrimidine residue was retro-mutated to the correspondent naturally occurring nucleotide, and secondly, the divalent Mg2+ ion was not included in the system. The choice of retro-mutating the fluorine-containing nucleotides was done for compatibility reasons since some of the software used in this scientific work (e.g. HADDOCK) could not work with non-natural nucleic residues. According to the original publication by Long et al. (87), the introduction of fluorine mainly impacted the aptamer resistance to ribonuclease rather than the binding affinity, as also underlined by the fact that only one single 2′ fluororo group is in direct contact with the protein, specifically at the level of the U17-Arg126 interaction. Concerning the presence of Mg2+ ions, despite their undisputable importance in the field of RNA folding, we opted not to include them at all in the simulations due to some intrinsic molecular mechanics limitations that hamper the possibility of fully and accurately describing their interaction with RNA (88), other than the difficulties in accurately predicting their locations without hints from experimental data (89).

As can be seen in Video 1 (Supplementary Material), about 15 ns of simulation time was sufficient to sample a putative molecular recognition event between the Toggle-25t RNA aptamer and the human thrombin exosite-2. This is a quite remarkable result, considering the usual hundreds of nanoseconds that are necessary to spontaneously sample a binding event with classic MD simulations. As illustrated by Video 1 and summarized in Figure 1, the final state of the SuMD simulation converged quite well both from a geometrical and an interactive point of view toward the crystal reference.

Concerning the geometric accuracy, the RMSD between the ligand backbone conformation in the final state of the simulation and the native ligand backbone conformation observed in the crystal structure is 6.41 Å, which is quite impressive considering the intrinsic structural flexibility of these objects, which is also confirmed by the structural deviation than can be observed in classic MD simulations of both the crystal complex (average RMSD of the nucleic acid backbone in the final step of the simulation across three MD replicates: 2.87 Å) and the free aptamer (average RMSD of the nucleic acid backbone in the final step of the simulation across three MD replicates: 3.85 Å). It is not surprising, therefore, that SuMD performs worse compared to molecular docking from a geometrical point of view (RMSD between the best docking pose and the crystal binding mode of the aptamer: 2.51 Å).

Despite a lower geometrical accuracy of the method compared to the one of docking, which derives mainly from the high intrinsic flexibility of the nucleic ligand during the simulation (see also Supplementary Figure S4, Supplementary Material, which reports the pairwise RMSD matrix of the free RNA aptamer during classic MD simulations), SuMD can correctly pose the negatively charged RNA–aptamer in a native-like conformation that maximizes the complementarity of shape and electrostatic features with the electronegative concave surface of the thrombin exosite-2, as highlighted in Figure 1 (panel A).

Concerning the capability of the protocol to correctly recapture the pivotal binding features despite a suboptimal geometric accuracy, SuMD can accurately describe the main interaction determinants, as illustrated by a comparison between the per-residue energy decomposition of the first 300 ps of the crystal complex classic MD simulation and the last 300 ps of the SuMD simulation (Supplementary Figure S5, Supplementary Material). Particularly, as can be seen in Figure 1 (panels C and D), SuMD correctly captures the pivotal role of both Arg 101 (461) and Arg 233 (608) in driving the recognition mechanism, serving as electrostatic recruiters in the initial phases of the process and acting as anchors to stabilize the bound state in the final part of the simulation, in agreement with mutagenesis data that assess how mutation of each of these two residues completely abrogates any aptamer’s effect (91). Particularly, Arg 233 (608) along with Arg 165 (533) are responsible for the formation of a stacked interaction domain motif defined as an ‘A-Arg zipper’ which involves five unpaired adenine residues on the ligand side, A4, A5, A7, A15, and A18 respectively. As can be seen in Figure 1 panel D, SuMD individually captures all these five adenine residues as key interaction determinants, in agreement with the experimental data. Finally, SuMD also discriminates the non-relevance of Gln 239 (614), which is not reported in the analysis as it is not one of the 25 most contacted residues during the trajectory, coherently with mutagenesis data that shows how a mutation of this residue does not affect the aptamer binding (91).

On the ligand side, SuMD once again correctly encompasses the different roles portrayed by different nucleotides. As can be seen in Figure 1 (panel D), the interaction between U17 and Arg 126 (486) is retrieved by the SuMD simulation: mutagenesis data shows how the mutation of U17, one of the flipped-out nucleotides, with adenine has essentially no effect on the aptamer’s affinity for the target. Looking at our ligand-based interaction map it can be noticed how this interaction is the only one in which this nucleotide is involved, other than being less intense compared to more prominent interaction such as the aforementioned ‘A-Arg zipper’, which suggest a non-pivotal contribution to the binding affinity. This can be related to the fact the U17 interacts with Arg126 through the backbone and not through its sidechain so that, as pointed out by the work of Jeter et al., this interaction would be maintained even when substituting the base (91). Substitution of U12 with adenine results in a nearly three-orders-of-magnitude diminished binding affinity: this is due to the stabilizing role that U12 plays towards A15 through a non-Watson-Crick base pairing, one of the bases involved in the formation of the ‘A-Arg zipper’(91). Once again, as depicted in Figure 1 panel D, SuMD correctly recognizes the pivotal role portrayed by A15, while contemporarily elucidating the indirect role of U12, which is not involved in any major interactions with
Figure 1. This panel encompasses the putative recognition pathway between the RNA aptamer Toggle-25t and the exosite-2 of human thrombin described by the best trajectory obtained through the SuMD protocol (the one with the lowest RMSD to the crystal reference). (A) Visual representation of the Toggle-25t conformation sampled in the last frame of the SuMD trajectory (green) superposed with the native Toggle-25t conformation (yellow), as found in the crystal structure deposited in the Protein Data Bank with accession code 3DD2 (RMSD SuMD-Crystal: 6.41 Å). The aptamer is represented as a ribbon, while the protein is represented as a Connolly surface colored according to the electrostatic potential as calculated with the APBS software (90), where red indicates a negatively charged area while blue indicates a positively charged one. (B) Profile of the ligand-receptor interaction energy (defined as the sum of the electrostatic and van der Waals contribution) throughout the recognition process as a function of both the simulation time and the RMSD between the ligand position during the trajectory and the ligand position in the crystal. (C) Receptor per-residue decomposition of the receptor–ligand interaction energy throughout the SuMD trajectory as a function of the simulation time: the 25 most-contacted residues are reported in the plot. (D) Per-residue interaction energy matrix: the 25 most-contacted residues for both the receptor and the ligand are considered, while each square composing the heatmap represents the average value of the interaction energy between the two paired residues alongside the trajectory.

To assess the predictive power of the method, we retrospectively analyzed all trajectories using two different metrics, i.e. the electrostatic interaction energy and the MMGBSA interaction energy of the final state of the simulation. The idea to use these two metrics stems from the consideration that RNA binding to proteins usually requires a good level of complementarity of steric and electrostatic properties at the binding interface. As reported in Supplementary Figure S19, both metrics can successfully distinguish the native and native-like poses, i.e. the ones with a superimposable interaction pattern with the reference (measured through the mean signed error and the root-mean squared error of the per-residue interaction energy decomposition, panels C and D respectively), from the decoys. This observation suggests that both metrics could be utilized in a prospective application of SuMD to rank poses coming from different simulations prioritizing the ones that are most similar to the native binding mode.

RNA aptamer bound to bacillus anthracis ribosomal protein S8 (PDB ID: 4PDB)

Thanks to its central function in the vital cycle of bacteria, the bacterial ribosome, a complex machinery responsible for protein synthesis in prokaryotic organisms, has been extensively studied, both from a structural and a functional perspective, and has been validated as a target for multiple antibiotic drugs (92). Protein-RNA interactions play a key role in the assembly, maturation, and function of the bacterial ribosome (93). Among these, association processes involving ribosomal protein S8 are particularly relevant since it not only participates in the 30s subunit assembly by binding to 16S rRNA but, additionally, it also serves as a translational repressor of the spc operon mRNA, which encodes for 11 ribosomal proteins including S8 itself (94).
Due to its relevance, the complex formed between bacterial ribosomal protein S8 and 16S rRNA has been thoroughly characterized through different techniques, allowing us to establish that most of the protein–RNA contacts involve helices 21 and 25 and that a small RNA portion located in helix 21 is sufficient to confer specificity and high affinity to the S8–RNA interaction (95). Furthermore, the interaction determinants between S8 and its RNA targets are largely conserved, and the same degree of conservation applies also to the overall fold of various S8 proteins (96–98). Finally, the complementarity of shape and electrostatic properties that is required for the binding entails a high level of nucleotide sequence and secondary structure conservation. This is proportional to the squared distance between the two interactors, even small differences in the relative position of interacting residues can alter the quantitative estimation of the interaction energy. Coherently with this interpretation of the data coming from the simulation analysis, our SuMD protocol can qualitatively describe the vast majority of the key interaction determinants.

In the crystal structure, the interaction between the S8 protein and the RNA aptamer involves a strip of electrostatic charge (as is also visible in Video 2 and Figure 2, panel A) along which the phosphate backbone of the aptamer traverses from residues A4–C7 and U27–A29: this behavior is correctly captured by our simulation, as depicted in Figure 2, panel D. Moreover, as can be noticed in Figure 2, panels C and D, SuMD correctly interpolates the polar interactions that form between the backbone of aptamer residues C16, C17, A24, U25, U27 and A26 and their counterparts on the protein side such as Glu 126, Ser 107, Gly124, Lys 110, Ser 109, Ala 91 and Thr123. Additionally, as depicted in Figure 2, panel D, the SuMD protocol is also able to spot some water-mediated interactions such as the contact between U27 to Glu 126. Finally, as can be depicted in Figure 2, panel D, SuMD can retrieve the stacked interaction between the peptide bond of highly conserved residues Ser 107–Thr 108–Ser 109 and the purine ring of A26: an analogous stacking interaction, involving A642, is present also in the complexes between the S8 protein and its natural RNA interactors, where it represents the only base-specific contacts of the complex (99). All other geometric and energetic analyses performed on the trajectory are summarized in Supplementary Figures S6–S8 (Supplementary Materials).

As for the previous case, we once again retrospectively analyzed all trajectories using the same metrics utilized before (electrostatic interaction energy and MMGBSA interaction energy), to assess the predictive power of the method. As reported in Supplementary Figure S20, also in this case both metrics can successfully distinguish the native and native-like poses from the decoys. This observation further supports the idea that either of the two metrics could be utilized in a prospective application of SuMD to rank poses coming from different simulations to prioritize the most similar to the native binding mode.

**RNA aptamer bound to human factor xa (PDB ID: 5VOE)**

One of the key events of the coagulation cascade is represented by the formation of the prothrombinase complex, macromolecular machinery formed by the serine protease
factor Xa (FXa) and its cofactor factor Va (FVa) (100,101): this membrane-mediated interaction enhances the catalytic activity of the FXa, leading to increased conversion of prothrombin into thrombin by a factor of $\sim 10^5$ and is, thereby, an interesting target for anticoagulation therapy (102,103). Despite a great effort devoted to the development of both small-molecule inhibitors of the FXa catalytic site and peptide inhibitors directed at epitopes on the binding interface with FVa, both these approaches have led to disappointing therapeutic results since the interaction surface is large and involves multiple hotspots (104,105) and the inhibition at the catalytic site-level interferes with natural regulatory agents such as antithrombin III (106).

To avoid the limitations of traditional small molecule ligands, in 2010 Buddai et al. developed RNA$_{11F7t}$, an RNA aptamer that exerts a potent anticoagulant effect by binding to FXa with high affinity ($K_d$ 1.1 ± 0.2 nM) and selectivity ($\sim 3000$ fold over other coagulation proteases) and inhibiting its interaction with FVa (107).

In 2018, Gunaratne et al. were able to solve the crystal structure of 11F7t bound to FXa, allowing us to better comprehend the key structural features that characterize this interaction (108). Specifically, in agreement with previous biochemical data, the analysis of the structure confirmed that the interaction occurs at a protein site which is implicated in the binding of both anticoagulant drug heparin and coagulation factor FactorVa (FVa) (109), with the interaction surface involving a central aptamer loop formed by residues C8, A10, A21 and C28–C30 and a protease exosite formed by Leu 59, Arg 64, Val 88, Ile 89, Asn 92, Arg 93, Lys 236 and Arg 240 (108). The absence of notable structural alterations of the protease upon aptamer binding, the therapeutic relevance of the target, and the relatively contained size of the aptamer (36 residues) induced us to consider it to validate the SuMD protocol. As in the case of structure 3DDD2, 2’ fluoro-modified nucleotides were retro-mutated to the corresponding natural alternatives, and the presence of the two Mg$^{2+}$ ions was not considered in the simulations.
As can be deemed by Video 3 (Supplementary Materials), in this case about 20 ns of simulation time were sufficient to sample a presumptive association pathway between the RNA aptamer 11F7t and human coagulation factor Xa. As can be noticed in Figure 3 (panel A), in this case, the geometric accuracy of the SuMD protocol in reproducing the crystal complex was worse compared to the first two cases, as also denoted by the 9.12 A RMSD value between the final state of the simulation and the crystal reference. A first explanation of the lower geometric accuracy of the method can be found in the analysis of the classic MD simulation performed on both the crystal complex and the free aptamer: as highlighted by Supplementary Figure S14 (Supplementary Materials) and by the RMSD of the nucleic backbone in the crystal complex MD (5.98 A) and the free aptamer MD (4.93 A), this aptamer has a significantly higher degree of conformational freedom compared to the previous two cases, which increase the difficulty of sampling the binding event due to the reduced amount of time that the aptamer spends in the binding competent conformation. Intriguingly, in this case, also, molecular docking compared worse than in the previous one, with the most correct pose having an RMSD of 3.17 A to the crystal reference, indicating that the structural flexibility of the object (which is not considered by molecular docking) is not the only determinant of the performance of the protocol.

Contrary to the expectations, SuMD converged quite well from an interactive point of view with the experimental data: as can be seen in Supplementary Figure S15 (Supplementary Materials), SuMD can qualitatively retrieve most of the native crystal interactions even if, as in the case of complex 4PDB, the estimation of the relative interaction strength is not always congruent.

Specifically, as can be noticed in Figure 3, panel D, the SuMD protocol can correctly retrieve the pivotal part played by residues A10, A21 and C29-C30, while slightly missing out on the importance of contacts with residues C8 and C28. On the protein side, as illustrated by Figure 3, panel C, SuMD precisely describes the central role portrayed by residues Arg 64 (283), Val 88 (308), Ile 89 (309), Asn 92 (312), Arg 93 (313), Lys 236 (460) and Arg 240 (464), while only passing up on the interaction with Leu 59 (278). Interestingly, Arg 240 (464) and Lys 236 (460) are key residues for the binding of heparin, according to mutagenesis studies (110). Finally, SuMD analysis highlights how Arg 165 (387) and Lys 169 (391), two critical residues in the recognition of factor FXa by either factor Va and/or prothrombin, are not contacted during the trajectory, in agreement with both the crystal structure and previous observations which pointed out to the possibility that the abrogation of factor Va binding happened through an indirect effect rather than through occlusion of the interaction surface (107,108). All other geometric and energetic analyses performed on the trajectory are summarized in Supplementary Figures S11–S13 (Supplementary Materials).

Finally, encouraged by the promising insights provided by the first two cases, we retrospectively analyzed all trajectories using the same scoring metrics defined before, to establish if they were once again able to distinguish the native and native-like poses from the decoys. As can be noticed in Supplementary Figure S21, disappointingly both metrics fail to prioritize the most geometrically accurate solution, preferring instead the second-best one. Curiously, the reference crystal is also scored poorly by both metrics, indicating even if the SuMD protocol would have been able to sample it we would not have been able to prioritize it. As can be noticed in Supplementary Table S1, this case also HADDOCK fails to rank the most geometrically accurate pose as the top solution, attributing to it a lower rank than a completely incorrect pose (RMSD: 19.79 A), which ranks as the second best one. The observation that HADDOCK, despite incorporating information about native contacts in its scoring protocol, has trouble in correctly ranking poses for this case, combined with the intrinsic instability of the crystal complex, as indicated by the high RMSD value (5.98 A) in the classic MD simulations and the low interaction energy values attributed by both metrics, indicate how this case might be an outlier, thus justifying the hypothesis of using the previously proposed metrics in prospective applications of the SuMD protocol.

**RNA aptamer bound to SARS-CoV-2 spike glycoprotein receptor binding domain (RBD)**

The outbreak of the COVID-19 pandemic in December 2019 caused an unprecedented worldwide public health crisis, leading to the death of more than six million people all over the world (111,112). This illness is caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a betacoronavirus able to infect human cells by expressing a surface glycoprotein known as spike (S) glycoprotein, which interacts through its receptor-binding domain (RBD) with the human angiotensin-converting enzyme 2 (hACE2) that mediates the viral uptake process in conjunction with the associated transmembrane serine protease 2 (TMPRSS2) (113). Due to the central role that this interaction plays in SARS-CoV-2 infectivity, the vast majority of therapeutic and prophylactic efforts in contrasting the COVID-19 pandemic have therefore been directed towards the inhibition of the spike-hACE2 interaction, either through vaccination or administration of monoclonal antibodies (114,115).

To overcome the main disadvantages of monoclonal antibodies, such as their high production costs, poor room temperature stability, and immunogenicity, in 2021 Valero et al. performed a SELEX experiment for the identification of a serum-stable RNA aptamer that could tightly bind the RBD of SARS-CoV-2 spike protein preventing the interaction with hACE2 thereby neutralizing viral infectivity (59). This experiment led to the identification of RBD–PB6, an elongated stem-loop RNA aptamer that can selectively interact with the RBD with low nanomolar affinity $K_D \approx 18$ nM, inhibiting the binding of RBD to hACE2 in a concentration-dependent manner (59).

Considering the encouraging results shown by the three applications of the SuMD protocol to the study of recognition processes between RNA aptamers and proteins and to illustrate a possible application of the protocol in a prospective scenario, we used SuMD simulations to shed light on the possible association pathway between the RBD–PB6 aptamer and the SARS-CoV-2 spike RBD. A model of the RNA aptamer structure was obtained through the 3dRNA webserver, based on the input primary sequence retrieved...
Figure 3. This panel encompasses the putative recognition pathway between the 11F7t RNA aptamer and human factor Xa described by the best trajectory obtained through the SuMD protocol (the one with the lowest RMSD to the crystal reference). (A) Visual representation of the RNA aptamer conformation sampled in the last frame of the SuMD trajectory (green) superposed with the native RNA aptamer conformation (yellow), as found in the crystal structure deposited in the Protein Data Bank with accession code 5VOE (RMSD_{SuMD-Crystal} = 9.12 Å). The aptamer is represented as a ribbon, while the protein is represented as a Connolly surface colored according to the electrostatic potential as calculated with the APBS software (90), where red indicates a negatively charged area while blue indicates a positively charged one. (B) Profile of the ligand-receptor interaction energy (defined as the sum of the electrostatic and van der Waals contribution) throughout the recognition process as a function of both the simulation time and the RMSD between the ligand position during the trajectory and the ligand position in the crystal. (C) Receptor per-residue decomposition of the receptor–ligand interaction energy throughout the SuMD trajectory as a function of the simulation time: the 25 most-contacted residues are reported in the plot. (D) Per-residue interaction energy matrix: the 25 most-contacted residues for both the receptor and the ligand are considered, while each square composing the heatmap represents the average value of the interaction energy between the two paired residues alongside the trajectory.

from the original publication and on the secondary structure prediction by the NUPACK webserver, while the structure of the SARS-CoV-2 spike RBD was retrieved from the crystal structure of the complex between the RBD and hACE2, deposited in the PDB with accession code 6M0J. The RBD structure that is present in this crystal comprises protein residues ranging from Thr 333 to Gly 526, slightly shorter than the construct used in the SELEX experiment which included residues from Arg 319 to Asn 532. However, residues that are not experimentally solved in the crystal structure are on the opposite side relative to the hACE2 interaction interface, so their absence should not impact the validity of the simulation. In agreement with experimental data that indicate how the RBD–PB6 RNA aptamer and hACE2 compete for the same binding site on the RBD surface, SuMD simulations were carried out to sample a putative recognition mechanism between RBD–PB6 and RBD surface that is responsible for interaction with hACE2.

As can be observed in Video 4 (Supplementary Materials), about 20 ns of simulation time was enough to sample a putative recognition mechanism between the RBD–PB6 aptamer and the SARS-CoV-2 spike RBD. Interestingly, as can be noticed in Figure 4, panel A, there is a discrete level of convergence between the final state of the simulation and the docking-predicted binding mode, especially in the region ranging from A10 to A61, which is the minimal portion of the full-length aptamer which fully retains its binding capabilities to the RBD (RMSD\textsubscript{backbone\_SuMD\_docking} = 11.69 Å). Regarding the recognition mechanism proposed by the SuMD protocol, an analysis of the interaction pattern reveals that the first contacts on the aptamer side involve residues A10–A14 and G48–G52, all of which fall into the conserved aptamer moiety that is required for binding to RBD. After these initial contacts take place and steer the binding event, other ancillary stabilizing interactions occur, such as the one with residues U24–U26 and the one involving residues U62–U66. On the protein side, instead, the main contacts involve polar and charged residues, such as Thr 376, Lys 378, Arg 403, Arg 408, Tyr 449, Gln 498, Thr 500, Asn 501 and Tyr 505. Interestingly, the most contacted residues during the trajectory do not include either Lys 417 or Glu 484, which are involved in the mutations K417N
and E484K that characterize most viral variants with augmented infectivity compared to the wild-type virus. This evidence is in agreement with experimental data showing the RBD–PB6 affinity for the spike protein is practically unaffected by these mutations (59). On the contrary, SuMD simulations indicate Asn 501 (involved in the N501Y mutation) as one of the most important residues in the recognition process: contrary to other previously mentioned residues, Asn 501 is surrounded by other interacting residues that can be found both in the complex between hACE2 and RBD and in the final state of the SuMD simulation, such as Gln 498, Thr 500 and Tyr 505, which could justify the contained impact of this mutation on the binding affinity of the aptamer compared to the other two. Interestingly, the most recent viral variants of concern include mutations such as Q498R, Y505H, and D405N that increase the positive charge on the spike surface: based on the analysis of the interaction pattern predicted by SuMD (Figure 4, panel C), all these mutations should increase the affinity with the negatively charged RBD–PB6 RNA aptamer, not only justifying its affinity towards the alpha and beta variant of SARS-CoV-2 but also towards the one from omicron on (116,117). All other geometric and energetic analyses performed on the trajectory are summarized in Supplementary Figures S16–S18 (Supplementary Materials).

DISCUSSION

In this scientific work, we presented the first-ever application of Supervised Molecular Dynamics (SuMD) to the study of recognition processes between RNA macromolecules and proteins. Specifically, we concentrated our efforts on the aptamer class, due to their relevance as both therapeutic and diagnostic tools.

Threedifferentretrospectivecasestudies,wherethestructureoftheRNAaptamer–proteincomplexwasavailable, were presented to validate the ability of the SuMD protocol to reproduce the experimental data. In all three cases,
Despite the intrinsic limitations derived from the relatively high conformational freedom of the ribonucleic ligands, SuMD was able to converge quite well, both from a geometric and interactive point of view, with the experimentally solved complex structures. The lower geometric accuracy did not impair the ability to retrieve most if not all the binding determinants of the native complex, with the increased RMSD compared to docking being related to the portion of the ligand not directly involved in the binding interface. Despite the increased complexity of the considered system compared to the usually investigated protein-small molecule complexes, the simulation times required to sample a putative recognition mechanism between the RNA aptamers and their protein target were comparable: in all presented cases, indeed, 10–20 ns of simulation time were sufficient to capture the entire association pathway, from the unbound state to the final complex. The reduced computational effort that the SuMD platform provides compared to classic, unsupervised, molecular dynamics simulation, makes it more suitable for its implementation in a drug discovery pipeline, flanking and complementing the role of already established approaches such as molecular docking. Due to the limited sampling capability of molecular dynamics-based methods compared to molecular docking, now the optimal strategy would be to combine the two techniques rather than using them in a mutually exclusive fashion: the rapidity of molecular docking could be useful to generate a series of a reasonable binding hypothesis that could be then more thoroughly investigated through SuMD simulations. The rapid increase in computational power available to scientists will hopefully make it possible to solve the sampling issue of MD-based methods, allowing them to fully replace more physically approximate methods such as molecular docking.

Concerning the applicability of the SuMD protocol in a prospective scenario, we also presented a case study where the experimental structure of the RNA aptamer–protein complex was not experimentally solved. Particularly, due to its therapeutic relevance, we decided to investigate the recognition process between the RBD–PB6 RNA aptamer, developed by Valero et al. (59), and the SARS-CoV-2 spike protein receptor-binding domain RBD. We showed how, even in the absence of an experimentally solved structure of the aptamer, the SuMD protocol can be coupled with structure prediction tools to give a structural prediction congruent with experimental evidence.

Regarding the ability of the SuMD protocol to be applied in a prospective scenario, one crucial point regards the capability to discriminate between the native-like binding mode and decoys. In the three presented cases, the simulation that was carefully analyzed and discussed in the manuscript was the one that presented the best geometric agreement with the crystal reference but such a metric could not be applied to a prospective investigation of complexes whose structures have not already been solved. By retroactively analyzing the geometric and energetic profile of the generated trajectories, we noticed that the electrostatic component of the interaction energy plays a fundamental role in steering the association process. Particularly, as can be noticed in Supplementary Figures S19–S21 (Supplementary Materials), both the MMGBSA interaction energy and the electrostatic component on its own can discriminate and prioritize the native complex and native-like poses from decoys in two out of three case studies. The only exception to this rule is represented by the complex between the RNA aptamer 11F7t and human factor Xa (PDB ID: 5VOE), for which both the electrostatic and MMGBSA scoring metrics indicate the second most geometrically accurate pose as the one with the most favorable energetic profile. Interestingly, in this case, molecular docking also fails to prioritize the most geometrically accurate pose, ranking it as the third-best one (Supplementary Table S1, Supplementary Materials). The indication that docking and SuMD, despite using different scoring metrics, failed to prioritize the native-like conformation suggests that there is still room for improvement regarding the scoring of complexes involving nucleic acids. However, despite this, the use of MMGBSA and or electrostatic interaction energy as scoring metrics can still be relatively accurate in suggesting reasonable binding mode hypotheses that are congruent with experimental evidence, as also previously pointed out by a benchmark study by Chen et al. from 2018 (118). Due to these considerations, we opted for using MMGBSA as a scoring metric for our prospective study of the interaction between the RNA aptamer RBD–PB6 and the SARS-CoV-2 spike RBD.

The last aspect that is worth addressing is related to the choice of residues to consider for the supervision of the association process throughout the SuMD trajectory. As is the case for molecular docking, where the binding site is defined by the user through a sphere or a box wrapped around the area of interest, SuMD also requires the user to specify a residue selection on both the receptor and ligand sides that are used to compute the distance between the center of mass of the binding site and of the ligand that is fed to the supervision algorithm. The choice of residues is usually based on prior knowledge of the interaction site derived from experimental evidence, but there could be some cases where this choice is not obvious. The first possible solution to this problem is represented by the analysis of the electrostatic potential of the receptor surface: as can be seen in Video 1–4, the recognition between the RNA aptamers and their protein targets usually involves a high level of complementarity of electrostatic properties, with the negatively charged ribonucleic surface being nicely harbored by positively charged patches on the protein side. A second possible solution is to perform a docking calculation to have a first indication of the preferable binding mode of the object, followed by more extensive characterization of the binding mode through SuMD simulations. This solution was used in the context of this article for the study of the interaction between the RBD–PB6 RNA aptamer and the SARS-CoV-2 spike RBD.

The possibility to investigate different binding sites and binding mode hypotheses can also be viewed as a strong point of the SuMD technique: for example, in the case of complex ribosomal protein S8, two different RNA recognition sites are available for the aptamer, specifically the site involved in the interaction with helix 21 and the site that mediates interaction with helix 25 (99). SuMD simulations would allow investigating of both possibilities at the same time, elucidating the mechanistic details that determine the preferential recognition of the primary binding
site thereby helping the rational development of selective binders.

Furthermore, concerning the exploration of different binding hypotheses, SuMD would allow deciphering the possibility of alternative stoichiometries. For example, in the case of aptamer 11F7t, Buddai et al. noticed a peculiar and difficult to rationalize binding stoichiometry, other than a strong Ca\(^{2+}\) dependence of the interaction (107). The authors discussed various possibilities, including a possible effect on the protein and/or aptamer structure, but also a possible calcium-induced aptamer dimerization (107). In this case, the exploitation of the SuMD technique would have allowed the exploration of all these different hypotheses, which could not be investigated through static, time-independent techniques such as molecular docking.

DATA AVAILABILITY

The code to perform SuMD simulations is available free of charge at https://github.com/molecularmodelingsession/SuMD and at https://doi.org/10.5281/zenodo.7289442. The script utilized to perform analysis on SuMD trajectories is also available at https://github.com/molecularmodelingsession/SuMD-analyzer and at https://doi.org/10.5281/zenodo.728944. All trajectories presented in the article can be found at https://doi.org/10.5281/zenodo.6973437.

SUPPLEMENTARY DATA

Supplementary Data are available at NARGAB Online.

ACKNOWLEDGEMENTS

MMS lab is very grateful to Chemical Computing Group, OpenEye, and Acellera for their scientific and technical partnership. MMS lab gratefully acknowledges the support of NVIDIA Corporation with the donation of the Titan V GPU, used for this research. This work has been supported by “PNRR M4C2-Investimento 1.4- CN00000041” financed by – NextGenerationEU.

FUNDING

None declared.
Conflict of interest statement. None declared.

REFERENCES

1. Gilbert, W. (1986) Origin of life: the RNA world. Nature, 319, 618–618.
2. Breaker, R. R. and Joyce, G. F. (2014) The expanding view of RNA and DNA function. Chem. Biol., 21, 1059–1065.
3. Hangauer, M. J., Vaughn, I. W. and McManus, M. T. (2013) Pervasive transcription of the human genome produces thousands of previously unidentified long intergenic noncoding RNAs. PLoS Genet., 9, e1003569.
4. Cheetham, S. W., Gruhl, F., Mattick, J. S. and Dinger, M. E. (2013) Long noncoding RNAs and the genetics of cancer. Br. J. Cancer, 108, 2419–2425.
5. Esteller, M. (2011) Non-coding RNAs in human disease. Nat. Rev. Genet., 12, 861–874.
6. Morris, K. V. and Mattick, J. S. (2014) The rise of regulatory RNA. Nat. Rev. Genet., 15, 423–437.
7. Luo, M., Grozal, E., De Jonghe, S., Snoeck, R., Andrei, G. and Herdewijn, P. (2018) Amidate produgs of cyclic 9-(3-Hydroxy-2-(phosphonomethoxy)propyl)adenine with potent anti-herpesvirus activity. ACS Med. Chem. Lett., 9, 381–385.
8. Bartel, D. P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. Cell, 116, 281–297.
9. Ponting, C. P., Oliver, P. L. and Reik, W. (2009) Evolution and functions of long noncoding RNAs. Cell, 136, 629–641.
10. Salmon, L., Yang, S. and Al-Hashimi, H. M. (2014) Advances in the determination of nucleic acid conformational ensembles. Annu. Rev. Phys. Chem., 65, 293–316.
11. Cruz, J. A. and Westhof, E. (2009) The dynamic landscapes of RNA architecture. Cell, 136, 604–609.
12. Bartel, D. P. (2009) MicroRNAs: target recognition and regulatory functions. Cell, 136, 215–233.
13. Draper, D. E. (1995) Protein-RNA recognition. Annu. Rev. Biochem., 64, 593–620.
14. Lorger, M., Engstler, M., Homann, M. and Göringer, H. U. (2003) Targeting the variable surface of African trypanosomes with variant surface glycoprotein-specific, serum-stable RNA aptamers. Eur. J. Cell Biol., 82, 84–94.
15. Disney, M. D. (2019) Targeting RNA with small molecules to capture opportunities at the intersection of chemistry, biology, and medicine. J. Am. Chem. Soc., 141, 6776–6790.
16. Stoltenburg, R., Reinemann, C. and Strehlitz, B. (2007) SELEX—a (re)volutionary method to generate high-affinity small nucleic acid ligands. Biol. Chem., 234, 381–403.
17. Keefe, A. D., Pai, S. and Ellington, A. (2010) Aptamers as therapeutics. Nat. Rev. Drug Discov., 9, 537–550.
18. Jayasena, S. D. (1999) Aptamers: an emerging class of molecules that rival antibodies in diagnostics. Clin. Chem., 45, 1628–1650.
19. Jones, S. D., Daley, D. T. A., Luscombe, N. M., Berman, H. M. and Thornton, J. M. (2001) Protein–RNA interactions: a structural analysis. Nucleic Acids Res., 29, 943–954.
20. Morozova, N., Myers, J. and Shamoy, Y. (2006) Protein–RNA interactions: exploring binding patterns with a three-dimensional superposition analysis of high resolution structures. Bioinformatics, 22, 2746–2752.
21. Reynolds, A., Leake, D., Boese, Q., Scaringe, S., Marshall, W. S. and Khvorova, A. (2004) Rational siRNA design for RNA interference. Nat. Biotechnol., 22, 326–330.
22. Boese, Q., Leake, D., Reynolds, A., Read, S., Scaringe, S. A., Marshall, W. S. and Khvorova, A. (2005) Mechanistic insights aid computational short interfering RNA design. Methods Enzymol., 392, 73–96.
23. Meng, X. Y., Zhang, H. X., Mezei, M. and Cui, M. (2012) Molecular docking: a powerful approach for structure-based drug discovery. Curr. Comput. Aided-Drug Des., 7, 146–157.
24. Kuntz, J. D., Blaney, J. M., Otley, S. J., Langridge, R. and Ferrin, T. E. (1992) A geometric approach to macromolecule-ligand interactions. J. Mol. Biol., 161, 269–288.
25. Dominguez, C., Boelens, R. and Bonvin, A. (2003) HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. J. Am. Chem. Soc., 125, 1731–1737.
26. Cieplak, M., Kurcinski, M., Kamel, K., Kolinski, A., Alam, N., Schuler-Furman, O. and Kmiecik, S. (2018) Protein–peptide docking: opportunities and challenges. Drug Discovery Today, 23, 1530–1537.
27. Pedotti, M., Simonelli, L., Livoti, E. and Varani, L. (2011) Computational docking of antibody-antigen complexes, opportunities and pitfalls illustrated by influenza hemagglutinin. Int. J. Mol. Sci., 12, 226.
28. David Morley, S. and Ashfar, M. (2004) Validation of an empirical RNA-ligand scoring function for fast flexible docking using ribodock. J. Comput. Aided Mol. Des., 18, 189–208.
29. Nithin, C., Ghosh, P. and Bujnicki, J. M. (2018) Bioinformatics tools and benchmarks for computational docking and 3D structure prediction of RNA-Protein complexes. Genes, 9, 452.
30. Disney, M. D. (2019) Targeting RNA with small molecules to capture opportunities at the intersection of chemistry, biology, and medicine. J. Am. Chem. Soc., 141, 6776.
31. Fulze, S. and Gohlke, H. (2010) Molecular recognition of RNA: challenges for modelling interactions and plasticity. J. Mol. Recognit., 23, 220–231.
32. Hermann, T. (2002) Rational ligand design for RNA: the role of static structure and conformational flexibility in target recognition. *Biochimie*, **84**, 869–875.

33. De Vivo, M., Masetti, M., Bottegoni, G. and Cavalli, A. (2016) Role of molecular dynamics and related methods in drug discovery. *J. Med. Chem.*, **59**, 4035–4061.

34. Bernardi, R.C., Melo, M.C.R. and Schulten, K. (2015) Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochim. Biophys. Acta Gen. Subj.*, **1850**, 872–877.

35. Sabbadin, D. and Moro, S. (2014) Supervised molecular dynamics (SuMD) as a helpful tool to depict GPCR-ligand recognition pathway in a nanoscale time scale. *J. Chem. Inf. Model.*, **54**, 372–376.

36. Deganutti, G., Moro, S., Ciruela, F. and Sotelo, E. (2017) Exploring the complexity of ligand-protein recognition pathways using supervised molecular dynamics (SuMD) simulations. *J. Chem. Inf. Model.*, **56**, 687–705.

37. Cuzzolino, A., Sturlese, M., Deganutti, G., Salsamo, V., Sabatini, S., Sturlese, M., Cecchetti, V., Barreca, M.L. and Moro, S. (2019) Deciphering the molecular recognition mechanism of multidrug resistance staphylococcus aureus NorA efflux pump using a supervised molecular dynamics approach. *Int. J. Mol. Sci.*, **20**, 4041.

38. Ferrari, F., Bissaro, M., Fabbian, S., De Almeida, J., Mammi, S., Bissaro, M., Bolcato, G., Pavan, M., Bassani, D., Sturlese, M. and Moro, S. (2020) Exploring the RNA-Ribonucleic Targeting Molecules? *Front. Chem.*, **8**, 107.

39. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) The protein data bank. *Nucleic Acids Res.*, **28**, 235.

40. Berman, H.M. (2000) The protein data bank. *Nucleic Acids Res.*, **28**, 235–242.

41. Molecular Operating Environment (2021) *Chemical Computing Group ULC, 1010 Sherbrooke St. West, Suite 8910, Montreal, QC, Canada, H3A 2R7*. https://www.chemcomp.com/Research-Citing

42. Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., Zhang, Q., Shi, X., Wang, Q., Zhang, L. et al. (2020) Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*, **581**, 215–220.

43. Valero, J., Civit, L., Dupont, D.M., Selnhin, D., Reinert, L.S., Idorn, M., Israels, B.A., Bednarz, A.M., Bus, C., Asbach, B. et al. (2021) A serum-stable RNA aptamer specific for SARS-CoV-2 neutralizes viral entry. *Proc. Natl. Acad. Sci. U.S.A.*, **118**, e2112942118.

44. Zadeh, J.N., Stoneberg, C.D., Bois, J.S., Wolfe, B.R., Pierce, N.A. (2011) NUPACK: analysis and design of nucleic acid systems. *J. Comput. Chem.*, **32**, 170–173.

45. Wang, J., Wang, J., Huang, Y. and Xiao, Y. (2019) 3dRNA v2.0: an updated web server for RNA 3D structure prediction. *Int. J. Mol. Sci.*, **20**, 4116.

46. Humphrey, W., Dalke, A. and Schulten, K. (1996) VMD: visual molecular dynamics. *J. Mol. Graph.*, **14**, 33–38.

47. Case, D.A., Cheatham, T.E. 3rd, Darden, T., Gohlke, H., Luo, R., Merz, K.M. Jr, Onufriev, A., Simmerling, C., Wang, B. and Woods, R.J. (2005) The amber biomolecular simulation programs. *J. Comput. Chem.*, **26**, 1668–1688.

48. Palloleta, S., Sabbadin, D., von Kägelten, I., Hinz, S., Katrich, V., Hoffmann, K., Abdelrahman, A., Straßburger, J., Baci, Y., Zhao, Q. et al. (2015) Modeling ligand recognition at the P2Y12 receptor in light of X-ray structural information. *J. Comput. Aided Mol. Des.*, **29**, 737–756.

49. Palazzotti, D., Bissaro, M., Bolcato, G., Astolfi, A., Felicetti, T., Sabatini, S., Sturlese, M., Cecchetti, V., Barreca, M.L. and Moro, S. (2019) Deciphering the molecular recognition mechanism of multidrug resistance staphylococcus aureus NorA efflux pump using a supervised molecular dynamics approach. *Int. J. Mol. Sci.*, **20**, 4041.

50. Deganutti, G., Moro, S. and Reynolds, C.A. (2020) Exploring the RNA-Ribonucleic Targeting Molecules? *Front. Chem.*, **8**, 107.

51. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) The protein data bank. *Nucleic Acids Res.*, **28**, 235.

52. Berman, H.M. (2000) The protein data bank. *Nucleic Acids Res.*, **28**, 235–242.

53. Molecular Operating Environment (2021) *Chemical Computing Group ULC, 1010 Sherbrooke St. West, Suite 8910, Montreal, QC, Canada, H3A 2R7*. https://www.chemcomp.com/Research-Citing

54. Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., Zhang, Q., Shi, X., Wang, Q., Zhang, L. et al. (2020) Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*, **581**, 215–220.

55. Valero, J., Civit, L., Dupont, D.M., Selnhin, D., Reinert, L.S., Idorn, M., Israels, B.A., Bednarz, A.M., Bus, C., Asbach, B.et al. (2021) A serum-stable RNA aptamer specific for SARS-CoV-2 neutralizes viral entry. *Proc. Natl. Acad. Sci. U.S.A.*, **118**, e2112942118.

56. Zadeh, J.N., Stoneberg, C.D., Bois, J.S., Wolfe, B.R., Pierce, N.A. (2011) NUPACK: analysis and design of nucleic acid systems. *J. Comput. Chem.*, **32**, 170–173.

57. Wang, J., Wang, J., Huang, Y. and Xiao, Y. (2019) 3dRNA v2.0: an updated web server for RNA 3D structure prediction. *Int. J. Mol. Sci.*, **20**, 4116.

58. Humphrey, W., Dalke, A. and Schulten, K. (1996) VMD: visual molecular dynamics. *J. Mol. Graph.*, **14**, 33–38.

59. Case, D.A., Cheatham, T.E. 3rd, Darden, T., Gohlke, H., Luo, R., Merz, K.M. Jr, Onufriev, A., Simmerling, C., Wang, B. and Woods, R.J. (2005) The amber biomolecular simulation programs. *J. Comput. Chem.*, **26**, 1668–1688.

60. Maier, J.A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K.E. and Simmerling, C. (2015) fI4SB: improving the accuracy of protein side chain and backbone parameters from NMR. *J. Chem. Theory Comput.*, **11**, 3696–3713.

61. Perez, A., Marchan, I., Swozil, D., Sponer, J., Cheatham, T.E. 3rd, Laughton, C.A. and Orozco, M. (2007) Refinement of the AMBER force field for nucleic acids: improving the description of α/γ conformers. *Biophys. J.*, **92**, 3817–3829.

62. Zgarbovska, M., Otyepka, M., Sponer, J., Malek, A., Bana, P., Cheatham, T.E. 3rd and Jurcika, P. (2011) Nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *J. Chem. Theory Comput.*, **7**, 2886–2902.

63. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W. and Klein, M.L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926–935.

64. David, R.L., Handel, R. and Tretjakov, M.V. (2009) Langevin thermostat for rigid body dynamics. *J. Chem. Phys.*, **130**, 234101.

65. Kräutler, V., Gunsteren, W.F. and Hünenberger, P.H. (2001) A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. *J. Comput. Chem.*, **22**, 501–508.
114. Thanh Le, T., Andreadakis, Z., Kumar, A., Gómez Román, R., Tollefsen, S., Saville, M. and Mayhew, S. (2020) The COVID-19 vaccine development landscape. *Nat Rev Drug Discov.*, *19*, 305–306.

115. Taylor, P.C., Adams, A.C., Hufford, M.M., de la Torre, I., Winthrop, K. and Gottlieb, R.L. (2021) Neutralizing monoclonal antibodies for treatment of COVID-19. *Nat. Rev. Immunol.*, *21*, 382–393.

116. Sartore, G., Bassani, D., Ragazzi, E., Traldi, P., Lapolla, A. and Moro, S. (2021) In silico evaluation of the interaction between ACE2 and SARS-CoV-2 spike protein in a hyperglycemic environment. *Sci. Rep.*, *11*, 22860.

117. Bassani, D., Ragazzi, E., Lapolla, A., Sartore, G. and Moro, S. (2022) Omicron variant of SARS-CoV-2 virus: in silico evaluation of the possible impact on people affected by diabetes mellitus. *Front. Endocrinol. (Lausanne)*, *13*, 284.

118. Chen, F., Sun, H., Wang, J., Zhu, F., Liu, H., Wang, Z., Lei, T., Li, Y. and Hou, T. (2018) Assessing the performance of MM/PBSA and MM/GBSA methods. 8. Predicting binding free energies and poses of protein–RNA complexes. *RNA*, *24*, 1183–1194.