

RESEARCH ARTICLE

Modelling aptamers with nucleic acid mimics (NAM): From sequence to three-dimensional docking

Ricardo Oliveira^{1,2}, Eva Pinho^{1,2}, Ana Luísa Sousa¹, Óscar Dias^{3,4}, Nuno Filipe Azevedo², Carina Almeida^{1,2,3*}

1 INIAV - National Institute for Agrarian and Veterinarian Research, Rua dos Lagidos, Vairão, Vila do Conde, Portugal, **2** LEPABE - Laboratory for Process Engineering, Environment, Biotechnology and Energy, Faculty of Engineering, University of Porto, Rua Dr. Roberto Frias, Porto, Portugal, **3** Centre of Biological Engineering (CEB), University of Minho, Campus de Gualtar, Braga, Portugal, **4** LABELS – Associate Laboratory, Braga, Guimarães, Portugal

☯ These authors contributed equally to this work.

* carina.almeida@iniav.pt



OPEN ACCESS

Citation: Oliveira R, Pinho E, Sousa AL, Dias Ó, Azevedo NF, Almeida C (2022) Modelling aptamers with nucleic acid mimics (NAM): From sequence to three-dimensional docking. PLoS ONE 17(3): e0264701. <https://doi.org/10.1371/journal.pone.0264701>

Editor: Prabodhika Mallikaratchy, Lehman College, UNITED STATES

Received: October 22, 2021

Accepted: February 15, 2022

Published: March 23, 2022

Copyright: © 2022 Oliveira et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its [Supporting information files](#).

Funding: This work was financially supported by: Project POCI-01-0145-FEDER-028659, funded by FEDER funds through COMPETE2020 – Programa Operacional Competitividade e Internacionalização (POCI) and by national funds (PIDDAC) through FCT/MCTES; Base Funding - UIDB/00511/2020 of the Laboratory for Process Engineering, Environment, Biotechnology and Energy – LEPABE

Abstract

Aptamers are single-stranded oligonucleotides, formerly evolved by Systematic Evolution of Ligands by EXponential enrichment (SELEX), that fold into functional three-dimensional structures. Such conformation is crucial for aptamers' ability to bind to a target with high affinity and specificity. Unnatural nucleotides have been used to develop nucleic acid mimic (NAM) aptamers with increased performance, such as biological stability. Prior knowledge of aptamer-target interactions is critical for applying post-SELEX modifications with unnatural nucleotides since it can affect aptamers' structure and performance. Here, we describe an easy-to-apply *in silico* workflow using free available software / web servers to predict the tertiary conformation of NAM, DNA and RNA aptamers, as well as the docking with the target molecule. Representative 2'-O-methyl (2'OMe), locked nucleic acid (LNA), DNA and RNA aptamers, with experimental data deposited in Protein Data Bank, were selected to validate the workflow. All aptamers' tertiary structure and docking models were successfully predicted with good structural similarity to the experimental data. Thus, this workflow will boost the development of aptamers, particularly NAM aptamers, by assisting in the rational modification of specific nucleotides and avoiding trial-and-error approaches.

Introduction

Aptamers are single-stranded oligonucleotides generated by Systematic Evolution of Ligands by EXponential Enrichment (SELEX), whose functionality is strictly dependent on their tertiary structure [1]. Traditional DNA and RNA aptamers have low chemical and biological stability; but unnatural nucleotides can be used to improve their performance. Unnatural nucleotides comprise chemical modifications on the heterocyclic base or sugar-phosphate backbone of native nucleotides [2]. They can be inserted by either *de novo* SELEX (a selection

- funded by national funds through the FCT/MCTES (PIDDAC). RO thanks FCT for the PhD Fellowship SFRH/BD/138883/2018. OD acknowledges FCT for the Assistant Research contract obtained under CEEC Individual 2018.

Competing interests: The authors have declared that no competing interests exist.

process starting with a NAM (nucleic acid mimic)-based library) or post-SELEX modifications (nucleotide substitutions by unnatural nucleotides at specific positions) [3]. Although post-SELEX approaches have been widely used to enhance previously selected aptamers due to their ease of application, they are often based on trial-and-error methodologies [4, 5]. In addition, the applied modifications often alter aptamers' tertiary structure and, consequently, affect aptamer-target binding interaction [3, 6, 7]. Thus, knowledge of the tertiary structure of aptamers and their interaction with the target molecule is crucial to successfully perform post-SELEX modifications [8]. Such information can be obtained by experimental techniques such as magnetic resonance spectroscopy (NMR) and X-ray crystallography [8, 9]. However, NMR methods are usually limited to relatively small molecules (<40 kDa) due to the complexity of the data analysis, and X-ray crystallography requires crystals to provide adequate quality diffraction data and crystal structures of aptamer-target complexes have proven to be challenging to obtain. Moreover, both techniques require specialized equipment and technicians not available in most research laboratories [10].

Alternatively, *in silico* approaches have been proposed to assist post-SELEX modifications [8], including computational tools to predict aptamer structures and thermodynamic properties, as well as aptamer-target docking models (identification of key interaction residues, structural motifs, docking structures, etc.) [11]. Thus, they might allow a tailored selection of nucleotides to be modified to improve aptamer performance, saving experimental time in the conventional trial-and-error approach [8, 11]. Some workflows have been recently published to predict aptamers' tertiary structure, thermodynamic properties or docking models, starting from the aptamer primary structure (linear sequence), using bioinformatics tools designed for modelling nucleic acids [8, 11–17]. At the moment, there is no complete *in silico* workflow to help determine the three-dimensional conformation of aptamers and the binding to their target [17]. Software and web servers are dispersed by the internet, requiring logical assemble and careful choice of the most suitable tools.

Moreover, most computational tools have been designed for RNA; thus, having to be adapted to predict the tertiary structure of single-stranded DNA. Additionally, they are either unable to address NAM and/or require specialized computational and molecular modelling skills to use them [13, 18, 19].

Hence, to assist NAM aptamers development and boost their applicability, we propose a workflow that allows predicting the tertiary structure and interaction model with the target molecule of aptamers containing unnatural nucleotides. Moreover, the developed workflow is based on freely available software and web server without the need of specialized skills to use such tools.

Methods

Software workflow

A workflow, based on freely available bioinformatics tools, was developed to predict the tertiary structure of NAM, RNA and DNA aptamers (using the aptamers sequence as a starting point) and the docking model (including the identification of residues interacting with the target molecule). Several software/ web servers were assessed to establish the workflow. For the secondary structure assemble, Kinefold [20], Mfold [21], MPGAfold [22], NUPACK [23], RNAFold [24], RNAstructure [25] and RNA2D3D [26] were evaluated. For tertiary structure prediction, 3D-DART [27], 3dRNA [28], ASSEMBLE2 [29], NAMD [30], RNAComposer [31], Rosetta [32], YASARA [33] and Vfold3D [34] were tested. For inserting chemically modified nucleotides, compatible with the selected unnatural nucleotides, ASSEMBLE2 [29], BIOVIA Discovery Studio, PyMOL, and YASARA [33] were assessed. Finally for the docking model,

AutoDock [35], GOLD [36], GRAMM-X [37], HDOCK [38], NPDock [39] and Patchdock [40] were used. Suitability, ease of handling, and cost were used as criteria for selecting the software/web servers to assemble the workflow.

Briefly, the workflow comprises seven main steps (Fig 1): (1) predict the secondary structure from the nucleotide sequence using the Mfold web server (<http://unafold.rna.albany.edu/?q=mfold>) [21], (2) assemble the tertiary structure of the corresponding RNA model from the secondary structure on the 3dRNA v2.0 web server (<http://biophy.hust.edu.cn/new/3dRNA>) [41], (3) transform the tertiary RNA structure into a DNA or NAM structure through BIOVIA Discovery Studio software (v 20.1.0.19295), (4) add hydrogen atoms using the PyMOL software (windows version 2.4.0) since these atoms (often omitted in chemical structures) play a major role on the stabilization of aptamers tertiary structure and interaction with the target [42], (5) refine the final tertiary structure on QRNAS software (Ubuntu version 0.3—Quick Refinement of Nucleic Acids 0.3) [42], (6) simulate the docking models through the HDOCK web server (<http://hdock.phys.hust.edu.cn/>) [38] and (7) identify the interaction residues using the Protein-Ligand Interaction Profiler (PLIP) web server (<https://projects.biotec.tu-dresden.de/plip-web/plip/index>) [43].

Step 1: Aptamers' secondary structure. The nucleotide sequences were used as input to predict aptamers' secondary structure based on a free energy minimization model applying the Mfold web server [21]. Regarding DNA-like sequences, the folding temperature and ionic buffer conditions (Na^+ and Mg^{2+} concentrations) were adjusted at the "DNA Folding Form" according to the experimental selection conditions of each aptamer, while the remaining parameters were used with the default values (<http://www.unafold.org/mfold/applications/dna-folding-form.php>). For RNA-like sequences, the secondary structure of the aptamers was predicted with the RNA Folding Form V2.3 (<http://www.unafold.org/mfold/applications/rna-folding-form-v2.php>), using the folding temperature and ionic buffer conditions fixed at 1 M. For each aptamer, the most thermodynamically stable structure (lowest Gibbs free energy (ΔG) value) was selected, and the corresponding Vienna output file (dot-bracket notation -. ct file) saved.

Step 2: Tertiary structures of the equivalent RNA aptamers. Aptamers tertiary structure was assembled using the fully automatic 3dRNA v2.0 web server [28]. The nucleotide sequence and the respective dot-bracket notation (Vienna file), obtained in step 1, were used as input. As the 3dRNA v2.0 was developed for RNA structures, the nucleotide thymine (T) was replaced by uracil (U) for DNA-like aptamers. The *Procedure Optimize*, *5 predictions*, *3dRNA-Lib2* and *Minimization* were used as advanced options. The tertiary structures with the lowest score (better statistical model) were saved as a Protein Data Bank (pdb) file [41, 44].

Step 3: Mutation of RNA structures to DNA or NAM structures. The Biovia Discovery Studio Visualizer software was used convert RNA tertiary structures to DNA or NAM structures (when applicable). Conversion to DNA or NAM was performed by changing the nitrogenous bases from uracil to thymine, the pentoses from ribose to deoxyribose, or inserting the corresponding unnatural nucleotide.

Steps 4 and 5: Refine the final tertiary structures of aptamers. The tertiary structures were imported into PyMOL to add hydrogen atoms, often omitted from the simulated structures but with an important rule on the tertiary and molecular docking. The structures were then saved as pdb files and, finally, optimized by the steepest descent energy minimization method (100 000 steps) using the automated QRNAS software. This tool performs a fine-grained refinement of nucleic acid tertiary structures generated in previous steps by simplified methods that use pre-existing structures in databases as templates to develop a comparative model with low- to medium-resolution. QRNAS is an improved version of the AMBER

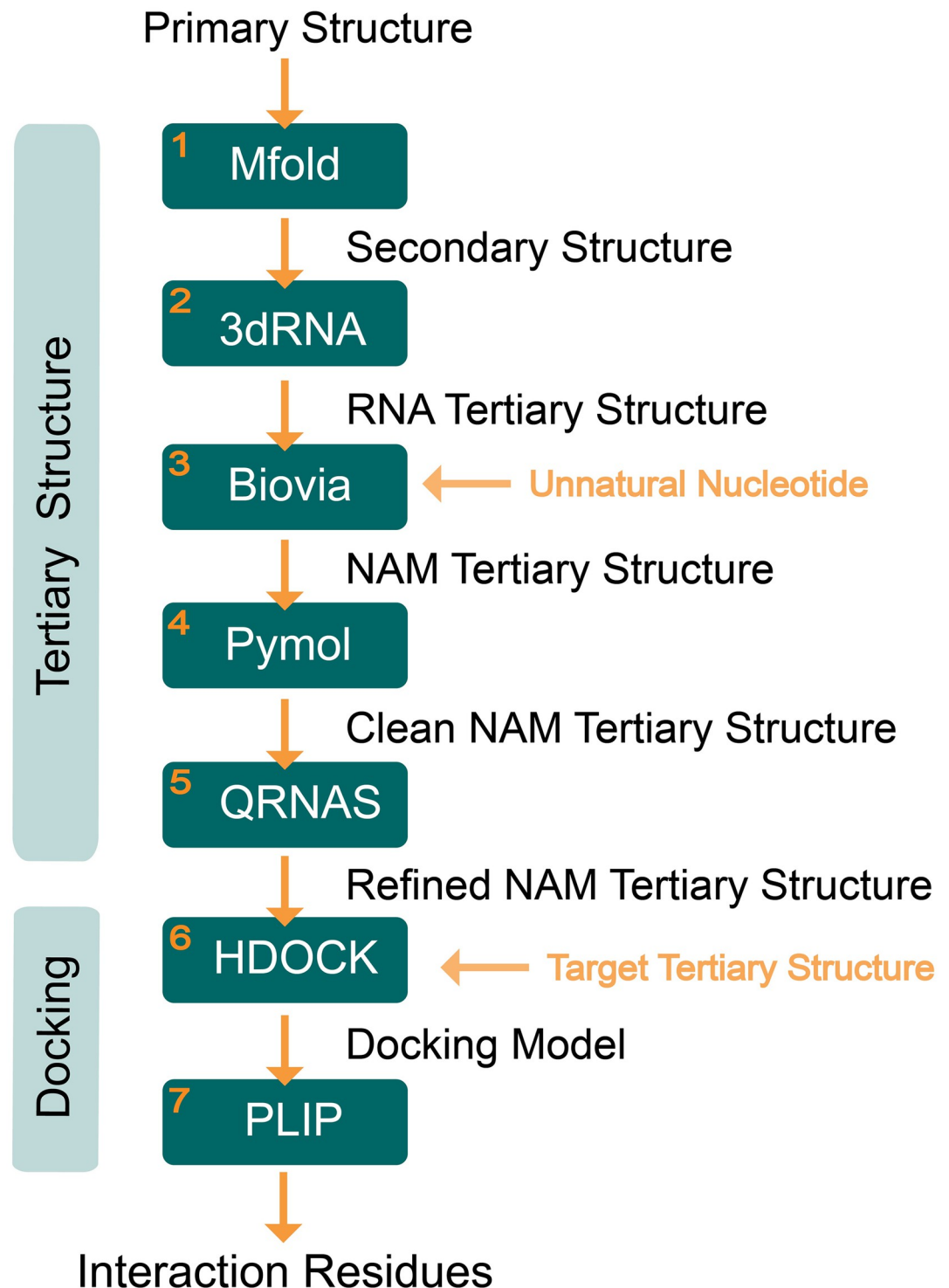


Fig 1. Workflow to predict the tertiary structure of NAM aptamers (using the aptamers sequence as a starting point) and the docking model (including the identification of nucleotides that interact with the target molecule). The yellow horizontal arrows are additional inputs necessary to run the workflow.

<https://doi.org/10.1371/journal.pone.0264701.g001>

simulation method that includes energy minimization of hydrogen bonds, base pairs, backbone irregularities, and custom restraints [42].

Step 6: Simulation of the interaction between aptamers and their targets. The target molecule was isolated by removing the experimentally resolved aptamer structure, ions or water molecules and adding the omitted hydrogens atoms using PyMOL software (clean structure) from the file downloaded from the database *Protein Data Bank in Europe* (PDBe, <https://www.ebi.ac.uk/pdbe/>). This step aims to normalize the effect of the solvent and ensure proper hydration of the functional groups [45]. The docking simulation was performed on HDock web server [38], using the tertiary structure of each aptamer (Step 5 pdb file) and target molecule (clean structure) as input.

Step 7: Identification of interactions between the aptamer and the target. The best docking model (lowest docking energy score) was selected and compared with the experimentally resolved aptamer-protein structures using PyMOL. The interacting residues between the aptamers and the targets were identified with the PLIP web server, using the docking model obtained on step 6 as input.

Workflow validation

A search for experimentally resolved structures of NAM, RNA and DNA, aptamers complexed with a protein molecule, was performed in the PDBe (<https://www.ebi.ac.uk/pdbe/>). From the three independent searches—using the terms "nucleic acid mimic aptamer", "DNA aptamer", and "RNA aptamer"—925 structures were identified. The structures were then individually reviewed, and the following exclusion criteria were applied: 1) lack of experimental resolved aptamer-protein complex; 2) inclusion of triplexes, duplexes, G-quadruplexes or kissing complexes (due to software/web server limitations); and 3) inclusion of other types of unnatural nucleotides besides LNA (locked nucleic acid) or 2'-O-methyl (2'OMe) nucleotides. The version of the software (BIOVIA) used to include unnatural nucleotides in the aptamer sequence only accepts 2'OMe and LNA.

After the first five steps of the workflow, the predicted tertiary structure of the aptamers was compared with experimental data from PDBe using RMSD (root mean standard deviation) as a quantitative parameter. For this, the files were open on the PyMOL software, and the RMSD was obtained with the align plugin, assuming five outlier rejection cycles and a cut-off of 2.0. The *in silico* docking models were evaluated based on the superposition with experimental data by RMSD and the number of shared contact nucleotides.

Results and discussion

The aptamers' performance is strictly linked to their conformational structure since aptamers' functionality relies on the physical fitting between the oligonucleotide and the target [46]. In contrast, their structure is determined by the nucleotide sequence and how that specific combination of nucleotides assembles into a three-dimensional conformation. Hence, the prediction of the tertiary structure is vital to identify residues responsible for the aptamer conformation and aptamer-target interaction.

Computational tools are an asset for developing and optimizing aptamers by avoiding time-consuming experimental procedures, especially when aiming post-SELEX modifications. Although there are several software and web servers for application of natural nucleic acids, no works have been published using such tools with NAMs. Thus, the authors compiled a workflow with free software and web servers (Fig 1) able to build the tertiary structure of NAM aptamers, besides RNA and DNA aptamers, using the nucleotide sequences as a starting point. Also, the workflow can simulate the molecular docking between the aptamers and their

Table 1. Aptamer models selected to validate the assembled workflow. The position of the modified residues is identified in the sequence with "m" preceding the respective nucleotide. The SELEX conditions were used as input on step 1 of the workflow.

PDBe ID	nt	Sequence	Aptamer type	Target	SELEX Conditions	Ref
5D3G	38	TAATACmCCmCCCCTTCGGTGCCTTG CACCGAAGGGGGGG	2'OMe	HIV-1 Reverse Transcriptase	25°C	[47]
					0 mM Na ⁺	
					2 mM Mg ²⁺	
5HRT	34	CCTGGAmCGmGAACmAmGAATmAm CTTTGGTCTCCmAmGmG	2'OMe	Autotaxin	25°C	[48]
					145 mM Na ⁺	
					0.8 mM Mg ²⁺	
2PN9	16	CACGGUCCmCmAGACGUG	LNA	TAR RNA element of HIV-1	23°C	[49]
					0 mM Na ⁺	
					0 mM Mg ²⁺	
6U82	38	GCTAATCTAATCAACCGCAGGT TGATTAGCCCATAGC	DNA	Double homeobox protein 4	20°C	[50]
					150 mM Na ⁺	
					5 mM Mg ²⁺	
4PDB	38	GGGAUGCUCAGUGAUCCUUCGG GAUAUCAGGGCAUCCC	RNA	<i>B. anthracis</i> S8 protein	4°C	[51]
					100 mM Na ⁺	
					5 mM Mg ²⁺	

<https://doi.org/10.1371/journal.pone.0264701.t001>

protein targets and identifying the physical-chemical interactions. The combination of these functionalities allows to predict the structural effect of modifications with unnatural nucleotides on the aptamers tertiary structure and docking model.

Selection of aptamer-target models

After an extensive analysis, two NAM aptamers containing 2'OMe nucleotides (PDB ID: 5D3G and 5HRT), a DNA (PDB ID: 6U82) and an RNA (PDB ID: 4PDB), were selected as study models to validate the workflow (Table 1). No experimentally resolved structure of LNA aptamers complexed to a protein-ligand was found in PDBe. Thus, an exception was made for an LNA aptamer (PDB ID: 2PN9) whose target is a nucleic acid (kissing complex). The LNA aptamer was used only as a model for the first phase of the workflow to predict the tertiary structure and compared with the experimental structure deposited on PDBe.

Prediction of aptamers' tertiary structure

The secondary and tertiary structures of NAM, RNA and DNA aptamers were successfully built (Fig 2) by applying the assembled workflow, using the nucleotide sequences as a starting point. Regarding the secondary structure, the ΔG of the selected structures ranged between $-33.50 \text{ kcal.mol}^{-1}$ (4PDB) and $4.58 \text{ kcal.mol}^{-1}$ (2PN9) (Fig 2A). This parameter is influenced not only by the nucleotide composition, but also by its folding pattern and the conditions in which it was determined [21]. Based on Fig 2A, the five aptamers showed similar secondary structures with stems, internal loops and hairpin loops. Accordingly, the tertiary structures were also quite identical (Fig 2B). Each predicted tertiary structure was aligned with the experimental structure deposited in the PDBe database. The RMSD value, a measure of the average distance between the atoms of superimposed macromolecules [52], was calculated using the PyMOL alignment plugin to measure the degree of similarity between predicted and experimentally resolved tertiary structures.

Aptamers structures obtained *in silico* were structurally identical to the experimental data (Fig 2(B)) with RMSD values ranging from 1.285 Å (2PN9) to 6.110 Å (6U82). The LNA

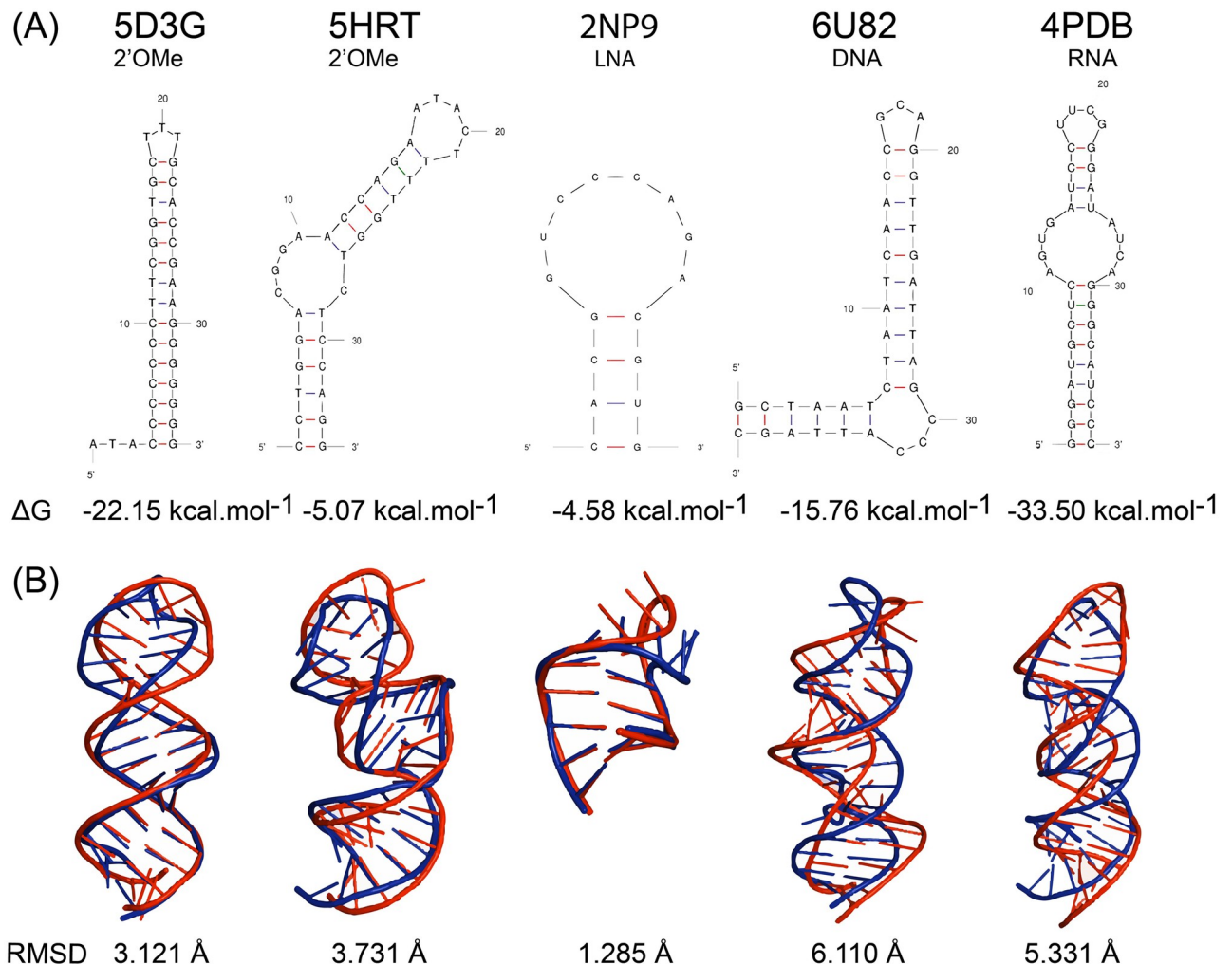


Fig 2. Aptamers' secondary (A) and tertiary (B) structure obtained after executing the first five steps of the workflow (Mfold, 3dRNA, Biovia, PyMOL and QRNAS). (A) The secondary structures obtained by Mfold and the Gibbs free energy (ΔG). (B) The overlap of the predicted tertiary structures (in red) and the corresponding experimentally resolved structures downloaded from the PDBe (in blue), and the RMSD values.

<https://doi.org/10.1371/journal.pone.0264701.g002>

aptamer (2PN9) had a simpler three-dimensional structure mainly due to its small size (16 nucleotides), probably justifying its lower RMSD. Similar RMSD values were reported for DNA aptamers and their experimentally deposited structures by other *in silico* workflows [13, 19].

It is important to highlight that the aptamers' tertiary structures were predicted without considering the unnatural nucleotides since the web server are only able to use natural nucleotides. For the secondary structure prediction (Step 1), no significant interference of unnatural nucleotides with modifications on the sugar molecule, such as 2'OMe or LNA, is expected since it is determined by intramolecular Watson-Crick complementary base pairing [12]. However, the tertiary structure is dependent on geometrical and steric constraints imposed by the nucleotides that compose aptamers [53]. Thus, the fine-grained refinement of the aptamers (Step 4 and 5) was vital to accommodate the modifications inserted after the tertiary structure prediction (Step 3) to minimize the imperfections in the prediction of NAM and DNA structures in the form of RNA (Step 2).

In the present work, only aptamers with 2'OMe or LNA modifications were used since the bioinformatic tool applied to change the tertiary structure from RNA to NAM was only capable of working with these two type of unnatural nucleotides. Other tools, such as Assemble or RNA2D3D, may allow the use of other unnatural nucleotides. Although, they require manual molecular interactions inputs that are usually unknown due to the lack of experimentally resolved structures of NAM aptamers. In fact, one major limitation when working with NAM aptamers is the absence of experimental data to feed the databases used by the bioinformatics tools. Moreover, from the authors' experience, most of the available tools that could work with unnatural nucleotides involve, besides advanced computational skills, access to online databases (not always available) and are often limited to small size oligonucleotides [9, 13, 17, 54].

Simulation of the interaction between aptamers and their targets

The molecular docking of the aptamers with the target molecules was simulated on the HDOCK web server [38] using the tertiary structures obtained after step 6 as input. *In silico* docking resorts to a bioinformatic process that sample all possible binding positions of the aptamer to the target. Then, the assembled docking models are sorted by a scoring function. In the case of HDOCK, the score classification takes into account the shape complementarity as well as the potential intermolecular interactions involved in the nucleic acid-target stabilization [38, 55]. Although several possible docking models were obtained, only the model with the best docking score for each aptamer was used for further analysis (Fig 3). As mentioned before, the LNA aptamer was not included in this step since its target is a nucleic acid (kissing complex), and the applied bioinformatics tool can only predict nucleic acid-protein target docking models [38].

Comparing the *in silico* molecular docking and the experimentally deposited complexes, a similarity of the aptamer-target interacting regions was verified for the four aptamer models used. The docking model obtained for 5D3G was the most identical to the experimental data, with both aptamers (*in silico* and experimental) showing the same orientation within the target (Fig 3). Similar behavior was observed for the RNA aptamer (4PDB), where both tertiary structures interacted with the target with the same orientation. For the aptamers 5HRT and 6U82,

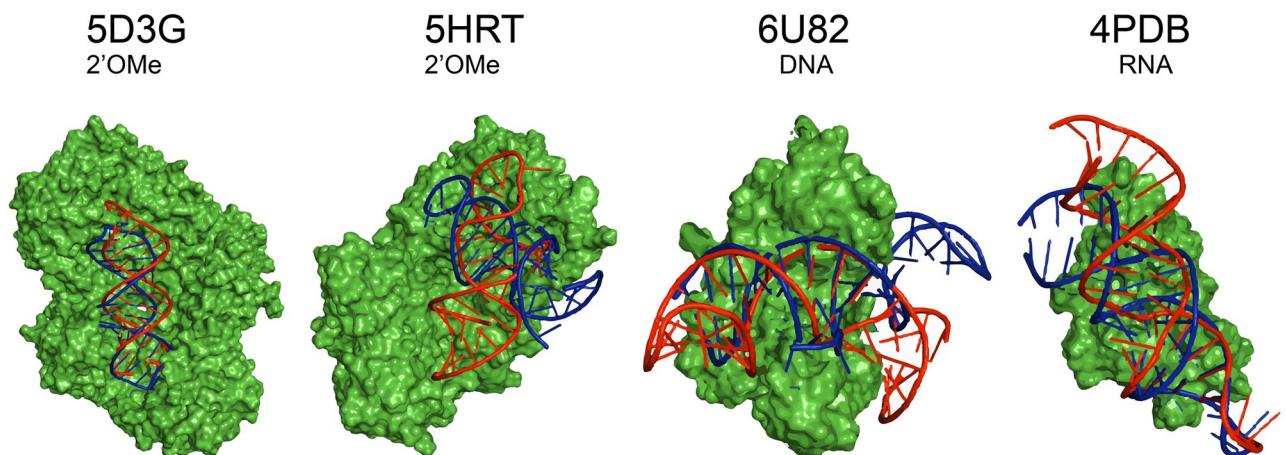


Fig 3. Comparison of the docking models deposited experimentally (in blue) in the PDBe database and the *in silico* docking models predicted through the described workflow (in red) for the 2'OMe (5D3G and 5HRT), DNA (5HRU) and RNA (6SY4) aptamers. The target molecules (in green) were isolated from the aptamer-target complexes determined experimentally using the PyMOL software and used as a receptor in the docking prediction.

<https://doi.org/10.1371/journal.pone.0264701.g003>

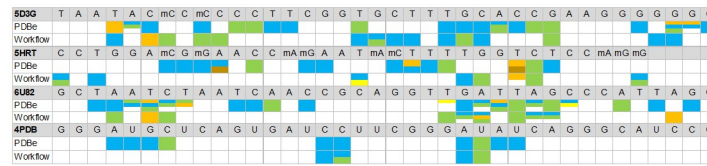


Fig 4. Identification of aptamer-target contact nucleotides and type of non-covalent interaction involved in the stabilization of the complexes for *in silico* (workflow) and experimental data (PDBe) docking. Hydrophobic Interactions (orange), Hydrogen Bonds (blue), Salt Bridges (green), π -Stacking (brown) and π -Cation Interactions (yellow).

<https://doi.org/10.1371/journal.pone.0264701.g004>

although the docking occurs in the same target and aptamer region, *in silico* aptamers orientation differed from the experimental data.

Based on the best molecular docking models, the non-covalent interactions between the target molecules and aptamers were also determined (Fig 4) using the PLIP web server [43]. This computational tool allowed identifying residues and the type of interactions between the aptamer and the target protein. The interaction report identifies the chemical group involved in the interaction, improving the knowledge on the mechanisms involved in stabilizing the aptamer-target interaction. Assuming the interacting residues of the experimentally resolved aptamer-target complexes as a reference, the 2'OMe aptamers (5D3G and 5HRT) present 9/20 (3 residues matched the type of non-covalent interactions) and 2/14 (1 residue correspond to the type of non-covalent interactions) coincident interacting residues, respectively. DNA aptamers have 9/21 coincident interacting residues (2 residues matching the type of non-covalent interactions), and the RNA aptamer has 4/10 (3 residues are full matched with the type of non-covalent interactions).

Comparison between the experimentally deposited data and the *in silico* simulations reveals a low correspondence of interacting residues. This observation is, in part, expected since, despite the structural similarity of tertiary structures, small differences can have a great influence on the shape-based docking and, consequently, on the intermolecular interactions between the aptamer and the target molecule. Even so, a greater number of corresponding interacting residues was observed for the 5D3G model, which presents better predictions of tertiary structure and docking region.

Moreover, the aptamer's (deoxy)ribose-phosphodiester backbone has high flexibility with a total of six torsional angles/flexible bonds, thus enabling a wide variety of tertiary structures with distinct degrees of energetic stability [55]. Hence, different conformations of the same aptamer sequence can coexist in solution, with various degrees of affinity (or, even, no affinity) to the target. The fact is that the conformation with a higher affinity to the target is not necessarily the most stable/energetic favorable. Thus, conformations with lower stability but higher affinity for the target can be experimentally favorable at the expense of the most stable structures *in silico*. The compiled workflow uses the predicted structures (secondary, tertiary and molecular docking models) ranked by the bioinformatics tools as the most likely statistically, i.e., more energetic favorable, to simplify analysis. Also, ions and temperature play a key role in experimental conformation, and some conditions might be challenging to predict by computational tools [56]. In this workflow, environmental conditions (i.e., temperature and ions concentration) are only controllable to a certain degree in the secondary structure prediction (Step 1). This may justify the differences obtained from the *in silico* structures and the experimental data. The selected *in silico* structures might not correspond to the ones with higher affinity under experimental conditions. A deeper analysis of all possible structures on each step may increase the accuracy of the *in silico* prediction, despite being more time

consuming and laborious. As an example, a supplementary analysis was carried to the 2'OMe aptamers (PDB ID: 5D3G and 5HRT) by performing variations on the simulation conditions: (i) environmental conditions (i.e., $\text{Na}^+/\text{Mg}^{2+}$ conditions and temperature, Step 1); (ii) using a tertiary structure with a lower score (Step 2), and (iii) not carrying out the mutation of the structures with the unnatural nucleotides (Step 3) (S1, S2 and S3 Figs, respectively). It was observed that the RMSD values of the predicted tertiary structures were higher for all alternative conditions and the docking with the target protein deviated from the experimentally determined data. These structural differences were reflected in the lower correspondence of the interacting residues between the simulated models and the experimental data deposited in the PDB database (S4 Fig).

Despite the above-mentioned limitations, the developed workflow proved to be an important tool for predicting aptamers tertiary structures (including NAM aptamers), as well as aptamers-protein target docking. It also serves as a tool for an initial analysis of the possible effect of post-SELEX modifications in aptamers conformation and interactions with the target. With this valuable information, it is possible to identify residues in different motifs of the aptamer for further substitutions with unnatural nucleotides, such as 2'OMe and LNA, which do not compromise the aptamer-target docking model. This allows a rational substitution of nucleotides and a suitable *in silico* platform to guide further benchwork.

Conclusions

Applying unnatural nucleotides is an essential evolution of aptamers, considering the limitations in the chemical and biological stability of traditional DNA and RNA aptamers. The development of this type of nucleic acid lacks intuitive computational tools to assist decisions involving structural predictions. Moreover, from the authors' knowledge, no workflow to predict NAM aptamer's tertiary structure and their binding to a target molecule has been published.

Here we present a simple and easy-to-apply workflow for researchers without deep computer experience to predict the tertiary structure and the molecular docking with the target of aptamers containing unnatural nucleotides.

Although the *in silico* data generated by the workflow correlated well to the experimental data, the proposed workflow has limitations on the type of unnatural nucleotides. However, with the increasing applicability of this type of computational approach, an improvement of these tools regarding the use of more types of unnatural nucleotides is expected. Also, a broad adoption of these tools for aptamer-target interactions, associated with the experimental validation of the *in silico* data, will serve to feed and perfect the algorithms of these bioinformatics tools.

Supporting information

S1 Fig. Effect of changing the environmental conditions (Step 1) on 2'OMe (PDB ID: 5D3G and 5HRT) aptamers' secondary (A), tertiary (B) structure and molecular docking model (C) simulation. For aptamer 5D3G, the secondary structure was predicted at 37°C, 100 mM Na^+ and 1 mM Mg^{2+} . Regarding, 5HRT the following conditions were used: 37°C, 50 mM Na^+ and 2 mM Mg^{2+} . (A) The secondary structures obtained by Mfold and the Gibbs free energy (ΔG). (B) The overlap of the predicted tertiary structures (in red) and the corresponding experimentally resolved structures downloaded from the PDB (in blue), and the RMSD values. (C) Molecular docking models deposited experimentally (in blue) in the PDB database and the *in silico* docking models predicted through the described workflow (in red). The target molecules (in green) were isolated from the aptamer-target complexes determined

experimentally using the PyMOL software and used as a receptor in the molecular docking prediction.

(TIF)

S2 Fig. Prediction of 2'OMe (PDB ID: 5D3G and 5HRT) aptamers' secondary (A), tertiary (B) structure and docking model (C) by selecting a tertiary structure with a lower score (Step 2). (A) The secondary structures obtained by Mfold and the Gibbs free energy (ΔG) using the corresponding environmental conditions. (B) The overlap of the predicted tertiary structures (in red) and the corresponding experimentally resolved structures downloaded from the PDBe (in blue), and the RMSD values. (C) Docking models deposited experimentally (in blue) in the PDBe database and the *in silico* docking models predicted through the described workflow (in red). The target molecules (in green) were isolated from the aptamer-target complexes determined experimentally using the PyMOL software and used as a receptor in the docking prediction.

(TIF)

S3 Fig. Prediction of 2'OMe (PDB ID: 5D3G and 5HRT) aptamers' secondary (A), tertiary (B) structure and docking model (C) without mutation of the unnatural nucleotides (Step 3). (A) The secondary structures obtained by Mfold and the Gibbs free energy (ΔG) using the corresponding environmental conditions. (B) The overlap of the predicted tertiary structures (in red) and the corresponding experimentally resolved structures downloaded from the PDBe (in blue), and the RMSD values. (C) Docking models deposited experimentally (in blue) in the PDBe database and the *in silico* docking models predicted through the described workflow (in red). The target molecules (in green) were isolated from the aptamer-target complexes determined experimentally using the PyMOL software and used as a receptor in the docking prediction.

(TIF)

S4 Fig. Identification of aptamer-target contact nucleotides and type of non-covalent interaction in the *in silico* simulations performed with alternative conditions for the 2'OMe aptamers. Hydrophobic Interactions (orange), Hydrogen Bonds (blue), Salt Bridges (green), π -Stacking (purple) and π -Cation Interactions (yellow).

(TIF)

S1 File. Input files of all steps of the proposed workflow for all tested aptamer models.

(RAR)

Author Contributions

Conceptualization: Ricardo Oliveira, Eva Pinho, Ana Luísa Sousa, Nuno Filipe Azevedo, Carina Almeida.

Formal analysis: Eva Pinho.

Funding acquisition: Carina Almeida.

Investigation: Ricardo Oliveira, Eva Pinho, Ana Luísa Sousa.

Methodology: Ricardo Oliveira, Eva Pinho.

Project administration: Nuno Filipe Azevedo, Carina Almeida.

Resources: Eva Pinho, Ana Luísa Sousa.

Supervision: Óscar Dias, Nuno Filipe Azevedo, Carina Almeida.

Validation: Ricardo Oliveira, Eva Pinho.

Writing – original draft: Ricardo Oliveira, Eva Pinho.

Writing – review & editing: Óscar Dias, Nuno Filipe Azevedo, Carina Almeida.

References

1. Stoltenburg R, Reinemann C, Strehlitz B. SELEX-A (r)evolutionary method to generate high-affinity nucleic acid ligands. *Biomolecular Engineering*. 2007. pp. 381–403. <https://doi.org/10.1016/j.bioeng.2007.06.001> PMID: 17627883
2. Morihiro K, Kasahara Y, Obika S. Biological applications of xeno nucleic acids. *Molecular BioSystems*. 2017. pp. 235–245. <https://doi.org/10.1039/c6mb00538a> PMID: 27827481
3. Wang T, Chen C, Larcher LM, Barrero RA, Veedu RN. Three decades of nucleic acid aptamer technologies: Lessons learned, progress and opportunities on aptamer development. *Biotechnology Advances*. Elsevier; 2019. pp. 28–50. <https://doi.org/10.1016/j.biotechadv.2018.11.001> PMID: 30408510
4. Vorobyeva MA, Davydova AS, Vorobjev PE, Pysnyi D V., Venyaminova AG. Key aspects of nucleic acid library design for in vitro selection. *International Journal of Molecular Sciences*. 2018. p. 470. <https://doi.org/10.3390/ijms19020470> PMID: 29401748
5. Hernandez L, Machado I, Schafer T, Hernandez F. Aptamers Overview: Selection, Features and Applications. *Curr Top Med Chem*. 2015; 15: 1066–1081. <https://doi.org/10.2174/1568026615666150413153717> PMID: 25866270
6. Gao S, Zheng X, Jiao B, Wang L. Post-SELEX optimization of aptamers. *Analytical and Bioanalytical Chemistry Anal Bioanal Chem*; Jul 1, 2016 pp. 4567–4573. <https://doi.org/10.1007/s00216-016-9556-2> PMID: 27173394
7. Lapa SA, Chudinov A V., Timofeev EN. The Toolbox for Modified Aptamers. *Molecular Biotechnology*. 2016. pp. 79–92. <https://doi.org/10.1007/s12033-015-9907-9> PMID: 26607475
8. Cai S, Yan J, Xiong H, Liu Y, Peng D, Liu Z. Investigations on the interface of nucleic acid aptamers and binding targets. *Analyst. The Royal Society of Chemistry*; 2018. pp. 5317–5338. <https://doi.org/10.1039/c8an01467a> PMID: 30357118
9. Buglak AA, Samokhvalov A V., Zherdev A V., Dzantiev BB. Methods and applications of in silico aptamer design and modeling. *International Journal of Molecular Sciences*. MDPI AG; 2020. pp. 1–25. <https://doi.org/10.3390/ijms21228420> PMID: 33182550
10. Ruigrok VJB, Levisson M, Hekelaar J, Smidt H, Dijkstra BW, van der Oost J. Characterization of aptamer-protein complexes by x-ray crystallography and alternative approaches. *International Journal of Molecular Sciences*. MDPI AG; 2012. pp. 10537–10552. <https://doi.org/10.3390/ijms130810537> PMID: 22949878
11. Sharma TK, Bruno JG, Dhiman A. ABCs of DNA aptamer and related assay development. *Biotechnol Adv*. 2017; 35: 275–301. <https://doi.org/10.1016/j.biotechadv.2017.01.003> PMID: 28108354
12. Krüger A, Zimbres FM, Kronenberger T, Wrenger C. Molecular modeling applied to nucleic acid-based molecule development. *Biomolecules*. Multidisciplinary Digital Publishing Institute (MDPI); 2018. <https://doi.org/10.3390/biom8030083> PMID: 30150587
13. Jeddi I, Saiz L. Three-dimensional modeling of single stranded DNA hairpins for aptamer-based biosensors. *Sci Rep*. 2017; 7: 1178. <https://doi.org/10.1038/s41598-017-01348-5> PMID: 28446765
14. Darmostuk M, Rimpelova S, Gbelcova H, Ruml T. Current approaches in SELEX: An update to aptamer selection technology. *Biotechnol Adv*. 2014; 33: 1141–1161. <https://doi.org/10.1016/j.biotechadv.2015.02.008> PMID: 25708387
15. Savory N, Abe K, Sode K, Ikebukuro K. Selection of DNA aptamer against prostate specific antigen using a genetic algorithm and application to sensing. *Biosens Bioelectron*. 2010; 26: 1386–1391. <https://doi.org/10.1016/j.bios.2010.07.057> PMID: 20692149
16. Zhou Q, Xia X, Luo Z, Liang H, Shakhnovich E. Searching the Sequence Space for Potent Aptamers Using SELEX in Silico. *J Chem Theory Comput*. 2015; 11: 5939–5946. <https://doi.org/10.1021/acs.jctc.5b00707> PMID: 26642994
17. Bruno JG. Do it yourself 3-dimensional aptamer-ligand molecular modeling. *J Bionanoscience*. 2017; 11: 183–188. <https://doi.org/10.1166/jbns.2017.1437>
18. Andersen ES. Prediction and design of DNA and RNA structures. *New Biotechnology Elsevier*; Jul 31, 2010 pp. 184–193. <https://doi.org/10.1016/j.nbt.2010.02.012> PMID: 20193785

19. Sabri MZ, Hamid AAA, Hitam SMS, Rahim MZA. The assessment of three dimensional modelling design for single strand DNA aptamers for computational chemistry application. *Biophys Chem.* 2020; 267. <https://doi.org/10.1016/j.bpc.2020.106492> PMID: 33035750
20. Xayaphoummine A, Bucher T, Isambert H. Kinefold web server for RNA/DNA folding path and structure prediction including pseudoknots and knots. *Nucleic Acids Res.* 2005; 33: 605–610.
21. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 2003; 31: 3406–3415. <https://doi.org/10.1093/nar/gkg595> PMID: 12824337
22. Shapiro BA, Kasprzak W, Grunewald C, Aman J. Graphical exploratory data analysis of RNA secondary structure dynamics predicted by the massively parallel genetic algorithm. *J Mol Graph Model.* 2006; 25: 514–531. <https://doi.org/10.1016/j.jmgm.2006.04.004> PMID: 16725358
23. Allouche A. Software News and Updates Gabedit—A Graphical User Interface for Computational Chemistry Softwares. *J Comput Chem.* 2012; 32: 174–182.
24. Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL. The Vienna RNA websuite. *Nucleic Acids Res.* 2008; 36: 70–74. <https://doi.org/10.1093/nar/gkn188> PMID: 18424795
25. Reuter JS, Mathews DH. RNAstructure: Software for RNA secondary structure prediction and analysis. *BMC Bioinformatics.* 2010; 11: 129. <https://doi.org/10.1186/1471-2105-11-129> PMID: 20230624
26. Martinez HM, Maizel J V., Shapiro BA. Rna2d3d: A program for generating, viewing, and comparing 3-dimensional models of rna. *J Biomol Struct Dyn.* 2008; 25: 669–683. <https://doi.org/10.1080/07391102.2008.10531240> PMID: 18399701
27. van Dijk M, Bonvin AMJJ. 3D-DART: A DNA structure modelling server. *Nucleic Acids Res.* 2009; 37: 235–239. <https://doi.org/10.1093/nar/gkp287> PMID: 19417072
28. Zhang Y, Wang J, Xiao Y. 3dRNA: Building RNA 3D structure with improved template library. *Comput Struct Biotechnol J.* 2020; 18: 2416–2423. <https://doi.org/10.1016/j.csbj.2020.08.017> PMID: 33005304
29. Jossinet F. Assemble2: An interactive graphical environment dedicated to the study and construction of RNA architectures. 2015 IEEE 1st Int Work Virtual Augment Real Mol Sci VARMS@IEEEVR 2015. 2015; 37–38.
30. Phillips JC, Hardy DJ, Maia JDC, Stone JE, Ribeiro J V., Bernardi RC, et al. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J Chem Phys.* 2020; 153: 1–33. <https://doi.org/10.1063/5.0014475> PMID: 32752662
31. Popenda M, Szachniuk M, Antczak M, Purzycka KJ, Lukasiak P, Bartol N, et al. Automated 3D structure composition for large RNAs. *Nucleic Acids Res.* 2012; 40: 1–12.
32. Cheng CY, Chou FC, Das R. Modeling complex RNA tertiary folds with Rosetta. 1st ed. *Methods in Enzymology.* Elsevier Inc.; 2015.
33. Krieger E, Vriend G. New ways to boost molecular dynamics simulations. *J Comput Chem.* 2015; 36: 996–1007. <https://doi.org/10.1002/jcc.23899> PMID: 25824339
34. Zhao C, Xu X, Chen S-J. Predicting RNA Structure with Vfold. *Physiology & behavior.* 2017. pp. 3–15. https://doi.org/10.1007/978-1-4939-7231-9_1 PMID: 28986779
35. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem.* 2009; 30: 2785–2791. <https://doi.org/10.1002/jcc.21256> PMID: 19399780
36. Jones G, Willett P, Glen RC, Leach AR, Taylor R. Development and validation of a genetic algorithm for flexible docking 1 Edited by F. E. Cohen. *J Mol Biol.* 1997; 267: 727–748. <https://doi.org/10.1006/jmbi.1996.0897> PMID: 9126849
37. Tovchigrechko A, Vakser IA. GRAMM-X public web server for protein-protein docking. *Nucleic Acids Res.* 2006; 34: 310–314. <https://doi.org/10.1093/nar/gkl206> PMID: 16845016
38. Yan Y, Zhang D, Zhou P, Li B, Huang SY. HDOCK: A web server for protein-protein and protein-DNA/RNA docking based on a hybrid strategy. *Nucleic Acids Res.* 2017; 45: W365–W373. <https://doi.org/10.1093/nar/gkx407> PMID: 28521030
39. Tuszynska I, Magnus M, Jonak K, Dawson W, Bujnicki JM. NPDock: A web server for protein-nucleic acid docking. *Nucleic Acids Res.* 2015; 43: W425–W430. <https://doi.org/10.1093/nar/gkv493> PMID: 25977296
40. Duhovny D, Nussinov R, Wolfson HJ. Efficient unbound docking of rigid molecules. *Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics).* 2002; 2452: 185–200.
41. Wang J, Wang J, Huang Y, Xiao Y. 3dRNA v2.0: An updated web server for RNA 3D structure prediction. *Int J Mol Sci.* 2019; 20. <https://doi.org/10.3390/ijms20174116> PMID: 31450739
42. Stasiewicz J, Mukherjee S, Nithin C, Bujnicki JM. QRNAS: Software tool for refinement of nucleic acid structures. *BMC Struct Biol.* 2019; 19: 5. <https://doi.org/10.1186/s12900-019-0103-1> PMID: 30898165

43. Salentin S, Schreiber S, Haupt VJ, Adasme MF, Schroeder M. PLIP: Fully automated protein-ligand interaction profiler. *Nucleic Acids Res.* 2015; 43: W443–W447. <https://doi.org/10.1093/nar/gkv315> PMID: 25873628
44. Wang J, Zhao Y, Zhu C, Xiao Y. 3dRNAscore: A distance and torsion angle dependent evaluation function of 3D RNA structures. *Nucleic Acids Res.* 2015; 43. <https://doi.org/10.1093/nar/gkv141> PMID: 25712091
45. Dans PD, Gallego D, Balaceanu A, Darré L, Gómez H, Orozco M. Modeling, Simulations, and Bioinformatics at the Service of RNA Structure. *Chem.* 2019; 5: 51–73. <https://doi.org/10.1016/j.chempr.2018.09.015>
46. Kim YS, Gu MB. Advances in aptamer screening and small molecule aptasensors. *Adv Biochem Eng Biotechnol.* 2014; 140: 29–67. https://doi.org/10.1007/10_2013_225 PMID: 23851587
47. Miller MT, Tuske S, Das K, DeStefano JJ, Arnold E. Structure of HIV-1 reverse transcriptase bound to a novel 38-mer hairpin template-primer DNA aptamer. *Protein Sci.* 2016; 25: 46–55. <https://doi.org/10.1002/pro.2776> PMID: 26296781
48. Kato K, Ikeda H, Miyakawa S, Futakawa S, Nonaka Y, Fujiwara M, et al. Structural basis for specific inhibition of Autotaxin by a DNA aptamer. *Nat Struct Mol Biol.* 2016; 23: 395–401. <https://doi.org/10.1038/nsmb.3200> PMID: 27043297
49. Lebars I, Richard T, Di primo C, Toulmé JJ. NMR structure of a kissing complex formed between the TAR RNA element of HIV-1 and a LNA-modified aptamer. *Nucleic Acids Res.* 2007; 35: 6103–6114. <https://doi.org/10.1093/nar/gkm655> PMID: 17768146
50. Klingler C, Ashley J, Shi K, Stiefvater A, Kyba M, Sinnreich M, et al. DNA aptamers against the DUX4 protein reveal novel therapeutic implications for FSHD. *FASEB J.* 2020; 34: 4573–4590. <https://doi.org/10.1096/fj.201902696> PMID: 32020675
51. Davlieva M, Donarski J, Wang J, Shamoo Y, Nikonowicz EP. Structure analysis of free and bound states of an RNA aptamer against ribosomal protein S8 from *Bacillus anthracis*. *Nucleic Acids Res.* 2014; 42: 10795–10808. <https://doi.org/10.1093/nar/gku743> PMID: 25140011
52. Parisien M, Cruz JA, Westhof É, Major F. New metrics for comparing and assessing discrepancies between RNA 3D structures and models. *Rna.* 2009; 15: 1875–1885. <https://doi.org/10.1261/rna.1700409> PMID: 19710185
53. Hopfinger MC, Kirkpatrick CC, Znosko BM. Predictions and analyses of RNA nearest neighbor parameters for modified nucleotides. *Nucleic Acids Res.* 2020; 48: 8901–8913. <https://doi.org/10.1093/nar/gkaa654> PMID: 32810273
54. Heiat M, Najafi A, Ranjbar R, Latifi AM, Rasaee MJ. Computational approach to analyze isolated ssDNA aptamers against angiotensin II. *J Biotechnol.* 2016; 230: 34–39. <https://doi.org/10.1016/j.jbiotec.2016.05.021> PMID: 27188956
55. Ahmad NA, Mohamed Zulkifli R, Hussin H, Nadri MH. In silico approach for Post-SELEX DNA aptamers: A mini-review. *J Mol Graph Model.* 2021; 105: 107872. <https://doi.org/10.1016/j.jmglm.2021.107872> PMID: 33765525
56. Tinoco I, Bustamante C. How RNA folds. *J Mol Biol.* 1999; 293: 271–281. <https://doi.org/10.1006/jmbi.1999.3001> PMID: 10550208