PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

fingeRNAt—A novel tool for high-throughput analysis of nucleic acid-ligand interactions

Natalia A. Szulc*¤a, Zuzanna Mackiewicz¤b, Janusz M. Bujnicki*, Filip Stefaniak*

Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell Biology in Warsaw, Warsaw, Poland

¤a Current address: Laboratory of Protein Metabolism, International Institute of Molecular and Cell Biology in Warsaw, Warsaw, Poland
¤b Current address: Laboratory of RNA Biology—ERA Chairs Group, International Institute of Molecular and Cell Biology in Warsaw, Warsaw, Poland
* nszulc@iimcb.gov.pl (NAS); janusz@iimcb.gov.pl (JMB); fstefaniak@iimcb.gov.pl (FS)

Citation: Szulc NA, Mackiewicz Z, Bujnicki JM, Stefaniak F (2022) fingeRNAt—A novel tool for high-throughput analysis of nucleic acid-ligand interactions. PLoS Comput Biol 18(6): e1009783. https://doi.org/10.1371/journal.pcbi.1009783

Editor: Shi-Jie Chen, University of Missouri, UNITED STATES

Received: December 23, 2021

Accepted: May 6, 2022

Published: June 2, 2022

Data Availability Statement: The fingeRNAt program is freely available and distributed under the open-source GPL-3.0 License. It can be downloaded, along with a manual, collection of helper utilities, and sample data from https://github.com/n-szulc/fingeRNAt. The program was extensively tested on Python 3.6, 3.7, 3.8, and 3.9 under Ubuntu Linux (18.04, 20.04, and 21.10) and macOS (macOS Catalina 10.15). The supporting data presented in the manuscript along with the code used for the analysis can be found at https://github.com/n-szulc/fingeRNAt-supplementary.

Abstract

Computational methods play a pivotal role in drug discovery and are widely applied in virtual screening, structure optimization, and compound activity profiling. Over the last decades, almost all the attention in medicinal chemistry has been directed to protein-ligand binding, and computational tools have been created with this target in mind. With novel discoveries of functional RNAs and their possible applications, RNAs have gained considerable attention as potential drug targets. However, the availability of bioinformatics tools for nucleic acids is limited. Here, we introduce fingeRNAt—a software tool for detecting non-covalent interactions formed in complexes of nucleic acids with ligands. The program detects nine types of interactions: (i) hydrogen and (ii) halogen bonds, (iii) cation-anion, (iv) pi-cation, (v) pi-anion, (vi) pi-stacking, (vii) inorganic ion-mediated, (viii) water-mediated, and (ix) lipophilic interactions. However, the scope of detected interactions can be easily expanded using a simple plugin system. In addition, detected interactions can be visualized using the associated PyMOL plugin, which facilitates the analysis of medium-throughput molecular complexes. Interactions are also encoded and stored as a bioinformatics-friendly Structural Interaction Fingerprint (SIFt)—a binary string where the respective bit in the fingerprint is set to 1 if a particular interaction is present and to 0 otherwise. This output format, in turn, enables high-throughput analysis of interaction data using data analysis techniques. We present applications of fingeRNAt-generated interaction fingerprints for visual and computational analysis of RNA-ligand complexes, including analysis of interactions formed in experimentally determined RNA-small molecule ligand complexes deposited in the Protein Data Bank. We propose interaction fingerprint-based similarity as an alternative measure to RMSD to recapitulate complexes with similar interactions but different folding. We present an application of interaction fingerprints for the clustering of molecular complexes. This approach can be used to group ligands that form similar binding networks and thus have similar biological properties. The fingeRNAt software is freely available at https://github.com/n-szulc/fingeRNAt.

Funding: This research was supported by the Foundation for Polish Science and the EU European Regional Development Fund https://www.fnp.org.pl/ (grant number POIR.04.04.00-00-3CF0/16 to J. M.B.) and National Science Centre, Poland https://www.ncn.gov.pl/ (grant number 2020/39/B/NZ2/03127 to F.S.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

Author summary

We present a novel bioinformatics tool, fingeRNAt, aiming to support scientists in the analysis of complexes of nucleic acids with various types of ligands. The software automatically detects non-covalent interactions and presents them in a form that is understandable to both humans and computers. Such data can help decipher the nature of interactions between nucleic acids and ligands and determine the main factors responsible for forming such complexes in nature. fingeRNAt finds application in multiple studies, both structure- and drug discovery-oriented. Here, we analyzed the experimentally solved structures of RNA complexes with small molecules to determine which binding features are most prevalent, i.e., most common interactions or their hot spots. The results of this analysis may help elucidate the mechanisms of binding and design new active molecules. Moreover, we propose to use the data generated by our software as a new metric for the quantitative pairwise comparison of molecular complexes. We have shown that it is more reliable than the currently used methods in certain "difficult" cases. We have shown that the results of our program can be used for high-throughput analysis of molecular complexes and the search for active molecules. We are confident that fingeRNAt will be a valuable tool for exploring the complex world of interactions of nucleic acids with ligands.

Introduction

Nucleic acids are essential bioorganic molecules present in every living organism. Although deoxyribonucleic acid (DNA) is traditionally viewed as a mere genetic information carrier and ribonucleic acid (RNA) as a scaffold in the protein synthesis process, their functions go far beyond that [1–3]. Both DNAs and RNAs regulate diverse biological pathways and thus have a central role in cellular metabolism. Non-coding DNAs constitute the majority of the human genome and regulate protein-coding sequences by acting as a binding site for other transcriptional regulatory factors, an origin of replication site, a centromere, or a telomere [4,5]. Moreover, some non-coding DNAs can be transcribed into non-coding RNAs, which play a fundamental role in the cell, as they build large macromolecular machines, deliver amino acids to ribosomes, or regulate different molecular processes, e.g., by silencing genes or driving catalytic reactions.

Nucleic acids possess the ability to adopt tertiary structures and have grooves acting as binding sites for other factors. They are capable of forming complexes with other nucleic acids, proteins [6], ions [7–10], and naturally occurring small molecules, such as metabolites [11]. These interactions are essential for cell functioning as they may modulate transcription and translation processes, DNA repair, splicing, apoptosis, or stress responses. Nucleic acids are also targets for synthetic small molecule drugs. DNAs remain a primary target for several anticancer chemotherapeutics [12] and potential antimicrobial compounds [13]. RNA molecules, such as the bacterial ribosomes or human pre-mRNA of survival of motor neuron 2 (SMN2) protein, are also known targets for a number of drugs, e.g., bacterial ribosome-targeting antibiotics [14,15], or risdiplam [16,17], respectively. Other RNAs, such as mRNAs [18] and regulatory RNAs in humans [19], riboswitches in bacteria [20], and conserved non-coding RNAs in viruses [21], are considered as promising targets for new therapeutics (for review, see [4,22,23]).

Weak non-covalent bonds are crucial in the molecular recognition process. Their identification and characterization help elucidate the basics of intermolecular binding and support the rational design of bioactive compounds [24,25]. Typically, this process requires a laborious

visual inspection of three-dimensional (3D) models of complexes by structural biologists or medicinal chemists [26]. With the advent of computational methods, this procedure may be supported with programs aiming at detecting and characterizing non-covalent interactions. As most of the available programs were designed to analyze protein complexes with small molecule ligands [27–31], the number of tools focusing on nucleic acids is currently very limited. LigPlot and LigPlot+ may be used to visualize nucleic acid-ligand complexes, but they display only hydrogen bonds and lipophilic interactions [32,33]. Arpeggio is a web server dedicated to detecting and visualizing interatomic interactions in protein structures; however it may be applied to DNA macromolecules [34]. The recent update of the PLIP program and its web server introduced support for DNA and RNA receptors, enabling the detection and visualization of several types of non-covalent bonds [35]. ProLIF is a Python library developed to generate interaction fingerprints for protein, DNA, or RNA complexes from molecular dynamics simulations, experimentally determined structures, and molecular docking [36].

To facilitate high-throughput analysis of intermolecular interactions, detected non-covalent bonds might be encoded in the form of Structural Interaction Fingerprint (SIFt), which describes the existence of specific molecular interactions between all structure's residues and a ligand. SIFt, firstly published by Deng *et al*. as a method to study protein-ligand binding, translates information about 3D interactions within the complex into a 1D binary string (bit vector) [37]. SIFt calculation consists of two main steps. First, the presence of interaction of the specified type for each residue-ligand pair is checked, and an appropriate binary value is assigned (1 if the interaction occurs and 0 otherwise). Subsequently, all calculated binary substrings are merged into one long string—SIFt, preserving the structure's residue order. Typical SIFt applications include post-docking analyses, such as clustering molecule's poses from molecular docking and comparing them with reference structures or scoring functions [38]. It is also frequently applied in interpreting activity landscapes, supporting structural databases, and analyzing protein-ligand complexes to search for similarities, e.g., by calculating the Tanimoto coefficient of bit vectors. In rational drug discovery, SIFt supports processing virtual screening results [39,31,40] or developing new scoring functions [41,42]. With the growing importance of artificial intelligence methods in drug discovery, new applications of interaction fingerprints emerged. SIFt can be associated with the information about ligand's biological activity, thus becoming an excellent input to the machine learning algorithms. This approach was already used, e.g., in training models to predict ligands' activity towards protein targets [43–47].

Here we present the fingeRNAt—a Python 3 program that detects and visualizes nucleic acid-ligand interactions. As an input, it takes a 3D structure of a nucleic acid (RNA or DNA) and a file containing ligands that form complexes with this macromolecule (e.g., the output from molecular docking with an external program). fingeRNAt accepts nucleic acids, small molecules, proteins, and metal cations as ligands. The output is a fingerprint—a bit vector containing information on interactions detected between interacting partners and optionally a human-readable file containing detailed information on detected interactions. By default, it detects nine interactions (see Implementation section in Materials and methods) but can be easily extended to detect virtually any type of interaction using a simple plugin system; the provided sample plugin file enables detection of eight additional interactions. Moreover, accompanying programs allow for convenient post-processing and visualization of detected interactions, calculation of Receptor Preferences (aka Receptor's Interactions Hot Spots), which represent the spatial occurrence frequency of a given interaction type in receptor atoms, and Ligand Preferences (aka Ligand's Interactions Hot Spots), which represent the spatial occurrence frequency of a given interaction type in ligand binding site.

To the best of our knowledge, there are no nucleic acid-dedicated tools for detection and classification of interactions that encode them in both machine- and human-readable formats,

are highly customizable, and allow for exhaustive post-processing such as calculation of similarity/distance metrics and interactive visualization. A detailed comparison of the fingeRNAt features (this work) and similar software tools (Arpeggio [34], PLIP2021 [35], and ProLIF [36]) can be found in S1 Table.

fingeRNAt is freely available to download from github.com/n-szulc/fingeRNAt. Program installation guide, together with an extensive manual, multiple usage examples, and Sphinx documentation, are also accessible from the repository. fingeRNAt can be used as a standalone command-line tool, but it also has an intuitive graphical user interface with the same functionalities.

Results and discussion

fingeRNAt is a program for detecting and classifying non-covalent interactions between a nucleic acid (RNA or DNA; called a *receptor*) and *ligands* (metal cations, small molecules, nucleic acids, or proteins). These data are encoded in the form of Structural Interaction Fingerprints (SIFts)—a 1D bit vector indicating the presence or absence of a given type of interaction, as well as in the form of a detailed listing of all detected interactions, spatial coordinates of the interacting partners, and distances between interacting atoms or aromatic rings.

Here we present three analyses performed for RNA-ligand complexes. In all the cases, the fingeRNAt played a pivotal role in data gathering and analysis.

Classification of interactions in experimentally solved RNA-ligand structures

Experimentally solved structures of macromolecules and their complexes are an invaluable source of knowledge on intermolecular interactions. At the time this publication was written, 1570 structures of RNA had been deposited in the Nucleic Acid Database, with 946 structures of RNA complexes with small molecule ligands (as of 16-Dec-2021, [48]). The information on statistics of interactions in RNA-ligand complexes derived from the solved structures can be used to develop bioinformatics methods to predict the structure of such complexes. Methods that enable an analysis of RNA-ligand interactions include docking programs (such as rDock [49,50]) or scoring functions (such as DrugScoreRNA [51], RNAPosers [52], and developed in our laboratory LigandRNA and AnnapuRNA [53,54]). As an output, the aforementioned methods return the proposed binding pose with numerical score(s). Although these data offer great help in compound prioritization processes or virtual screening, they do not explain the nature of the binding phenomena nor give insights into the main driving forces of the investigated interaction. To shed light on the landscape of interactions with small molecule ligands, we analyzed a diversified dataset of experimentally solved RNA structures deposited in the PDB. We determined the nature of formed non-covalent interactions, including frequency and distance distribution for each investigated contact type.

The performed analysis reveals that the most frequently occurring interactions are hydrogen bonds (5026) and lipophilic interactions (3582; see Fig 1A and S2 Table). Next, with an order of magnitude lower number, are cation-anion bonds (899), water-mediated interactions (151), and Pi-stacking interactions (146). The number of the remaining interactions is two orders of magnitude lower than the number of detected hydrogen bonds, with halogen bonds being the least frequent detected interaction (6). Three metal cations that were present only in one complex each (namely: Pb, Mn, and Sr) were removed from plots for clarity.

Statistics indicating the absolute number of interactions, although informative, may be somehow misleading. Seven types of interactions (namely hydrogen bonds, lipophilic, cation-anion, halogen bonds, and three types of mediated interactions) are detected between single

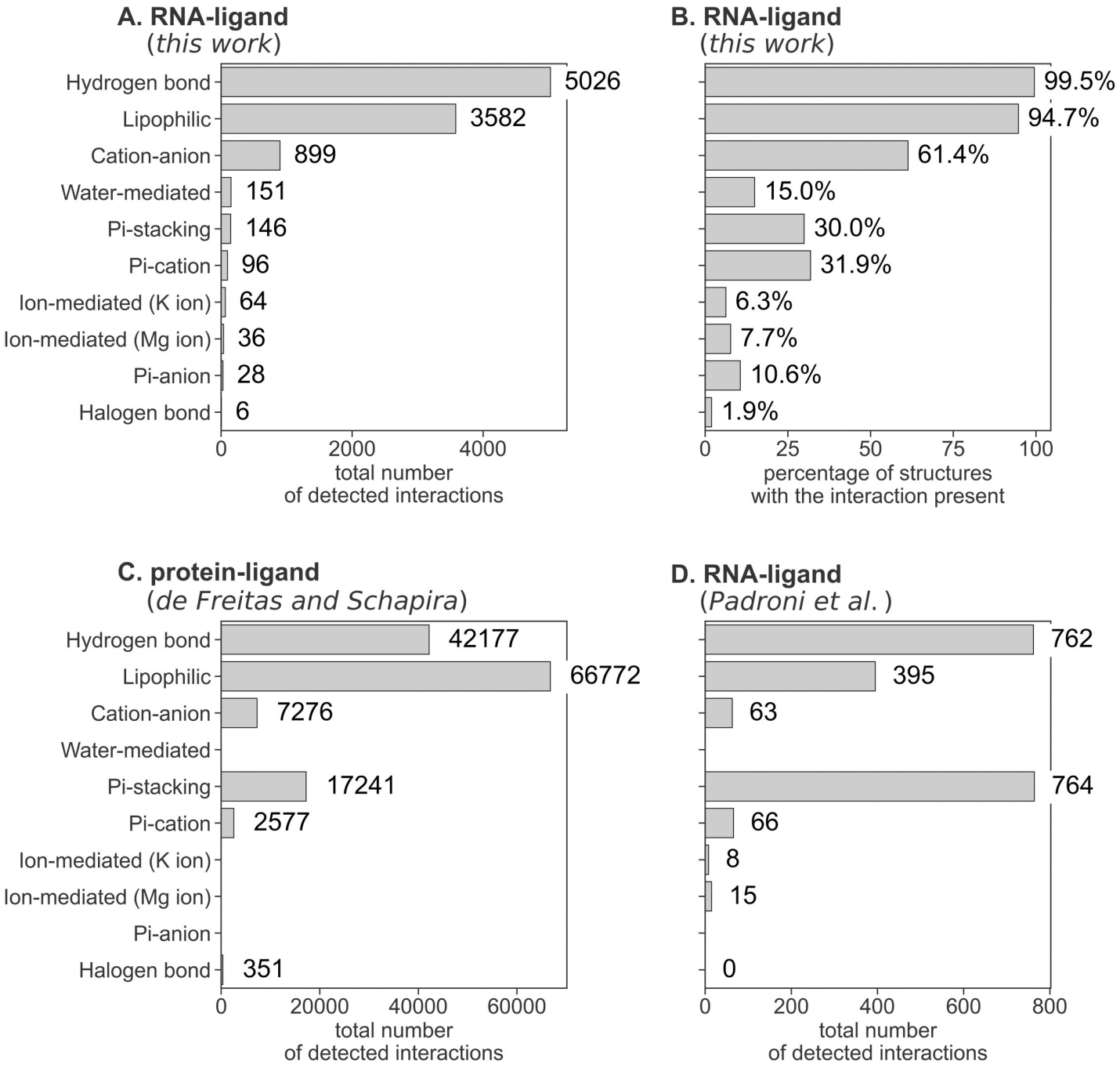

Fig 1. Statistics of interactions formed in macromolecular complexes. (**A**) Total number of interactions detected for RNA-ligand complexes (this work); (**B**) the percentage of RNA-ligand complexes with at least one occurrence of a given interaction (this work); (**C**) number of interactions for protein-ligand complexes presented by de Freitas and Schapira; (**D**) number of interactions for RNA-ligand complexes presented by Padroni *et al.*

https://doi.org/10.1371/journal.pcbi.1009783.g001

atoms/groups of RNA and ligands. Thus, a single interaction point may form multiple bonds with the partner (e.g., a single lipophilic atom of RNA may form bonds with multiple atoms of the ligand and *vice versa*). Interactions involving Pi-systems (Pi-stacking and Pi-ions interactions) are formed between at least one cyclic system, which due to geometrical restraints may form only a limited number of interactions.

To circumvent this bias, we calculated the number and the percentage of complexes with at least one of the given interactions detected (see Fig 1B). This analysis revealed that the hydrogen bonds, lipophilic, and cation-anion interactions are still the most abundant, present in over 99%, 94%, and 61% of structures, respectively. In this analysis, the role of the interactions formed by Pi-systems is more pronounced—the Pi-stacking and Pi-cation interactions are formed in over 30% of structures, and Pi-anion interactions are detected in 11% of structures. The percentage of structures forming mediated interactions is lower than suggested by the number of interactions formed (15%, 8%, and 6% for structures with water-, Mg^{2+}-, and K^{+}-mediated interactions, respectively). We also counted the number of interactions formed by unique residues of RNA (i.e., only one interaction of a given type with a given residue is counted). Results confirm the order of the frequency of interactions revealed by our first analysis (Fig 1A), but with a slightly decreased role of water-mediated interactions (S1 Fig).

In the dataset analyzed in this work, 141 structures (68%) contain at least a single water molecule. Within this group, we observed the formation of a water-mediated interaction only in 22% of structures (15% of the whole dataset). On the other hand, in a subset of eight high-resolution structures (resolution $\leq$ 1.5 Å), all complexes have water coordinates, and seven (88%) of structures form water bridges between RNA and ligand (see the discussion on the structure resolution below). This value is very close to the data presented by Lu *et al*., who calculated that over 85% of the protein-ligand complex structures have at least one bridging water molecule present at the interface of the protein and the ligand (data from 392 high-resolution crystal structures) [55]. This suggests that the water-mediated interactions may play a crucial role in the molecular recognition process for RNA and small molecule ligands. However, the lack of high-quality experimentally solved structures prevents an accurate assessment of the scale of this phenomenon.

In fingeRNAt we implemented two methods for the detection of hydrogen bonds. The default algorithm takes into account the distance between the heavy atom of the hydrogen bond donor and the acceptor atom (D···A, [56]). An alternative, more selective algorithm takes into account not only the distance but also the angle between the heavy atom of the hydrogen bond donor, the hydrogen atom, and the acceptor atom (D-H···A, [35]). This method depends on the position of hydrogen atoms, and as most experimental structures do not have coordinates of hydrogens, these must be added computationally before calculations. The resulting structures may vary depending on the hydrogen-adding algorithm used (S2 Fig).

We compared the interaction statistics obtained when using the default method of hydrogen bond detection (based on the distance of D···A) and when using the distance and angle variant (D-H···A) with four algorithms for placing hydrogen atoms: the fingeRNAt's built-in algorithms utilizing the OpenBabel and RDKit libraries, and the external software: PyMOL and Chimera (Fig 2 and S3 Table) [57–60].

The number of detected interactions for methods that consider the position of hydrogen is almost three times lower (ranging from 1735 to 1822) than for the default method (5026). As expected, for the distance and angle variant (D-H···A), the number of detected interactions depends on the hydrogen-adding algorithm and is caused by the aforementioned variability of the location of some hydrogen atoms. A slightly higher number of detected interactions for Chimera-added protons could be explained by the fact that—unlike for other methods—hydrogens were added for ligands in complex with RNA, which may favor such positions of hydrogen atoms which contribute most to the intermolecular hydrogen-bond network. This method, however, is designed for protein-ligand complexes and may give suboptimal results for nucleic acids. To the best of our knowledge, a method designed for adding polar hydrogen atoms to RNA-ligand complexes currently does not exist. Applying such a tool in our analysis could provide the most reliable results for the number of hydrogen bonds formed within RNA-ligand complexes.

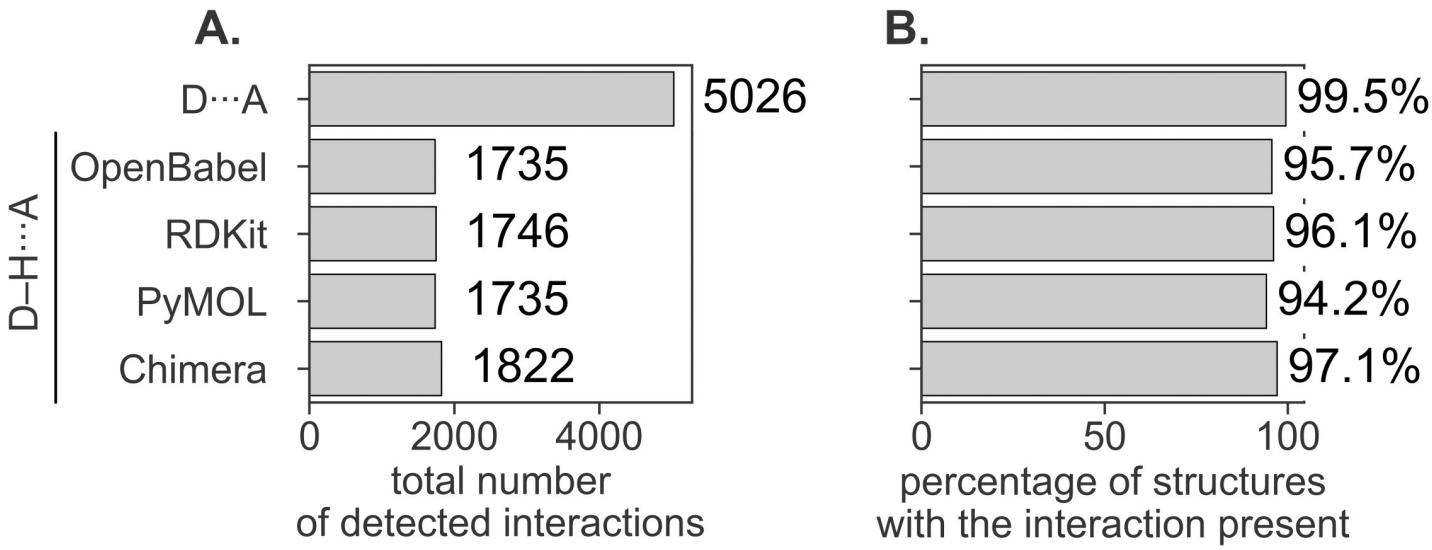

Fig 2. Statistics of hydrogen bonds formed in macromolecular complexes for the detection method without taking into account the position of a hydrogen atom (the default method, based on the distance of D···A) and when using this information (the alternative method invoked when -dha flag is passed to the program, based on the distance and angle D-H···A), calculated for four methods of adding hydrogens (OpenBabel, RDKit, PyMOL, and Chimera). (**A**) Total number of hydrogen bonds detected for RNA-ligand complexes; (**B**) the percentage of RNA-ligand complexes with at least one occurrence of a hydrogen bond.

https://doi.org/10.1371/journal.pcbi.1009783.g002

Comparison of the RNA-ligand interactions statistics to the data derived from protein-ligand complexes published by de Freitas and Schapira reveals that the two most frequent interactions are the same (hydrogen bonds and lipophilic interactions), however in the reversed order (Fig 1C) [61]. To our surprise, the Pi-stacking interactions are more pronounced for protein complexes than for RNA. This observation may be explained by the fact that only 38% of ligands in the RNA-ligand dataset contain an aromatic ring, which is a prerequisite for forming the Pi-stacking interaction with RNA (S3 Fig; data for protein complexes are not published), and almost 80% of ligands with an aromatic ring is forming a Pi-stacking interaction with RNA (see S4 Table for detailed statistics of Pi-stacking interactions, S5 Table for hydrogen bonds, and S6 Table for cation-anion interactions).

We also compared statistics of the interactions derived using the fingeRNAt to data published recently by Padroni *et al.* [62]. The authors used a proprietary software ICM to extract contact information from a set of 37 experimentally solved structures, which covered 14 unique RNAs (data summarized in Fig 1D). They found the Pi-stacking interaction and hydrogen bonds as the most common type of interaction in RNA-small molecule ligand complexes, with an almost equal number of detected interactions (764 and 762, respectively). Our analysis also ranks the hydrogen bonds as the most frequently occurring interaction, but as discussed above, the Pi-stacking interaction is less commonly discovered (the fifth most frequent interaction detected by our method). The third most frequent contact found by Padroni *et al.* is the lipophilic interaction (the second most frequent interaction detected by our method). The substantial difference in contact count values is, however, in the number of cation-anion interactions. In our analysis, this is the third most frequent interaction, while Padroni *et al.* rank it in the fifth position. We also detected four cases of halogen bonds, while Padroni *et al.* did not observe this kind of interaction. Observed discrepancies in absolute values and frequency ranks are especially pronounced in the high number of the detected Pi-stacking interactions (Padroni *et al.* reported 764 Pi-stacking interactions detected in 37 complexes). We analyzed the dataset of Padroni *et al.* using our software fingeRNAt (S4 Fig). As expected, the absolute numbers of interactions are different from those presented in our work; however, the ranking

of interactions (in terms of the total number of detected interactions) is the same as calculated for our dataset, with only minor differences in ion- and water-mediated interactions. This means that these two algorithms use different methods to detect interactions, especially for Pi-stacking interactions.

Observed differences could also result from the fact that we used protonated ligand molecules as an input which more realistically reflects the ionization state of the binding partners and their interactions, thus explaining the differences in the charge-involving interactions (cation-anion, Pi-cation interactions). The protonation model used in this analysis (the RNA and ligands have the protonation state assigned in isolation) is, however, only the approximation of the dynamic phenomena of protonation observed in nature. It was shown earlier that pK_a of RNA nucleotides is highly dependent on the structural environment [63,64] and can be affected by binding small molecule ligands [65,66]. Although there are methods for predicting protonation states for protein-ligand complexes [67,68], such a method dedicated to RNA-ligand complexes, to the best of our knowledge, currently does not exist.

The number of detected interactions, however, must be treated as an approximation only. As recently shown by Xu *et al*. for protein-ligand and protein-protein complexes, insufficient resolution of the data deposited in the PDB may result in overlooking a significant number of non-covalent interactions [69]. Most probably, this problem also exists for RNA-ligand complexes. The diversified set of 207 structures analyzed in our work consists of structures determined by NMR and X-ray crystallography (27 and 180 structures, respectively), and the latter are determined at various resolutions (ranging from 0.61 Å to 4.50 Å, see S5 Fig). To estimate the role of the resolution on the type and number of detected interactions, we repeated the analysis of interaction occurrence frequency for the seven subsets of structures: X-ray determined structures with a resolution equal to or below 1.5, 2.0, 2.5, 3.0, 3.5, and 4.0 Å, and the structures determined by the NMR (See S6 Fig for the complete data and Fig 3 for selected examples). In all groups of X-ray structures, the ranking of the frequency of detected interactions is the same as in the complete dataset, with a single exception of the highest quality structures (resolution $\leq$ 1.5 Å). In this group of eight complexes, the number of the water-mediated

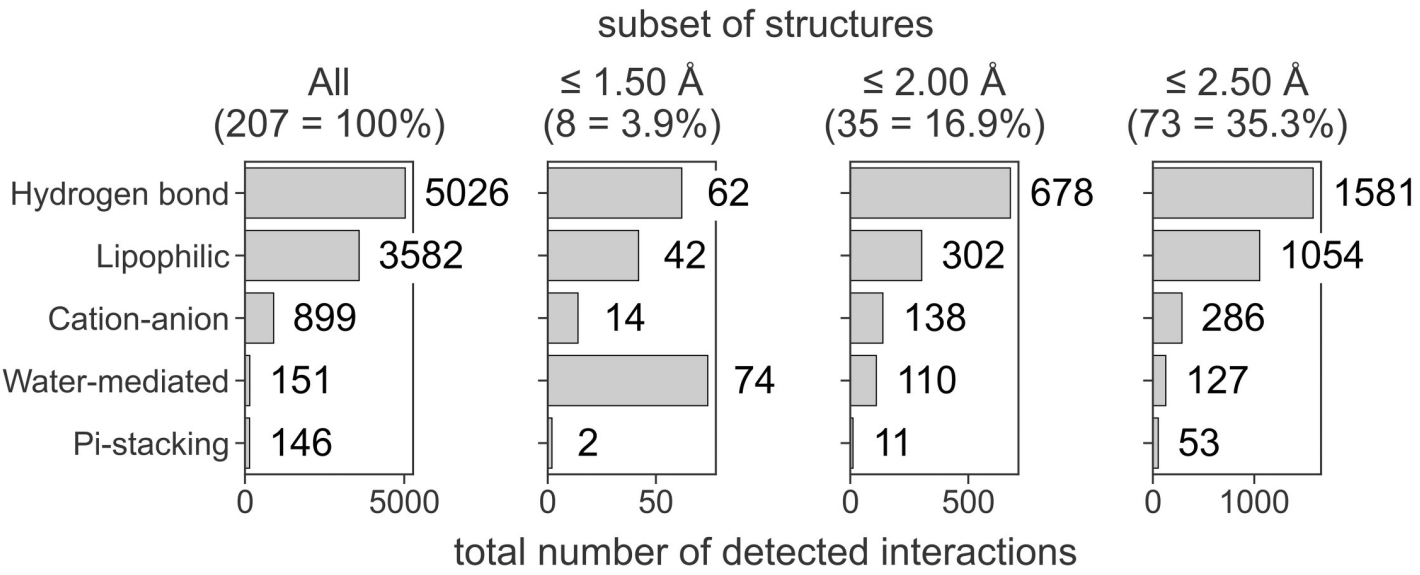

Fig 3. Total number of interactions detected for RNA-ligand complexes with a given resolution. The five most frequent interactions are shown for clarity. The number of structures in a given set is shown in parentheses.

https://doi.org/10.1371/journal.pcbi.1009783.g003

interactions is higher than the observed number of hydrogen bonds and lipophilic interactions (74, 62, and 42, respectively), which highly pronounces the role of water bridges in RNA-ligand complexes. This was expected since all these high-resolution structures contain water coordinates, and 88% of structures in this group form water bridges between RNA and ligand (compared to 15% in the complete dataset). In the group of 27 NMR-determined structures, the list of four most frequently detected interactions is also the same as in the complete dataset (apart from the water-mediated interactions, which cannot be determined due to the lack of water coordinates in the NMR structures). We may conclude that although the number and quality of experimentally determined structures are still limited, the presented statistics are a good estimation of the relative abundance of the non-covalent bonds in RNA-ligand structures.

As noted earlier, the interactions formed between ligands and RNA depend on the chemical structure of the ligand. We grouped all ligands into seven clusters (alcohols and polyols, nucleosides and nucleoside derivatives, amino acids, amino sugars and aminoglycosides, aliphatic amines, heterocycles, and others) and calculated interaction statistics in each group (S1 Text). In all groups but "others", the most frequent interaction formed is hydrogen bond, as it was in the case of the general population of ligands. For amino acids, nucleosides, and heterocycles, the next most frequent contact is lipophilic interaction. For positively charged molecules containing amino groups (amino sugars and aliphatic amines), the second most abundant interaction is cation-anion, formed between negatively charged OP1/OP2 atoms of RNA and protonated amine cations in ligands. This is in line with the observation made by Padroni *et al*. that the frequency of interactions formed by aminoglycosides is different from the one observed in the entire analyzed dataset, with a pronounced role of hydrogen bonding, lipophilic interactions, and charge-involving interactions.

We also used the data generated by the fingeRNAt to investigate the preferred distances for detected interactions. The distribution is multimodal for hydrogen bonds and cation-anion interactions, with two clearly distinguished peaks (Fig 4, S7 Fig, and S7 Table). Such distribution was observed earlier for strong hydrogen bonds (for data derived from structures deposited in the Cambridge Structural Database (CSD) [70,71] and protein-ligand complexes [72]), metal cations with oxygen anion [73], salt bridges in proteins [74], and Pi-anion interactions [75]. For hydrogen bonds length distribution, the first peak is observed for the distance 2.7 Å, which is very close to the median length observed for hydrogen bonds in data derived from the PDB and CSD (2.75 Å and 2.9 Å for interactions of amide C = O with OH and NH, respectively, [76]). We observed the second-main peak at the length 3.7 Å, which may result from water-mediated contacts via water molecules not visible in the experimentally determined structures.

Using the interaction statistics gathered with the fingeRNAt software, we calculated the frequency of interactions formed by the individual RNA atoms. We also mapped the data into the nucleotide structures to visualize the Interactions' Hot Spots. For hydrogen bonds, most of the hydrogen bonds are formed by nucleobases (61%) and phosphate group atoms (23%), while ribose oxygen atoms are responsible only for a small fraction of the hydrogen bonds (15%, see Fig 5 and S8 Table). This observation is in agreement with results obtained by Kligun and Mandel-Gutfreund, indicating that nucleobases form 65%, while interactions with backbone atoms form 35% of RNA-ligand hydrogen bonds [77].

As expected, most hydrogen bonds are formed with a Watson-Crick and Hoogsteen face of the nucleobase (see S9 Table). The most frequent interactions for adenine, cytosine, and uracil are using a Watson–Crick face (31.2%, 63.0%, and 68.8% of all hydrogen bonds formed by the given nucleotide), while guanine tends to use a Hoogsteen face (41.3%). Hydrogen bonds formed using a sugar face of the nucleobase are the least frequent (ranging from 8.9% for

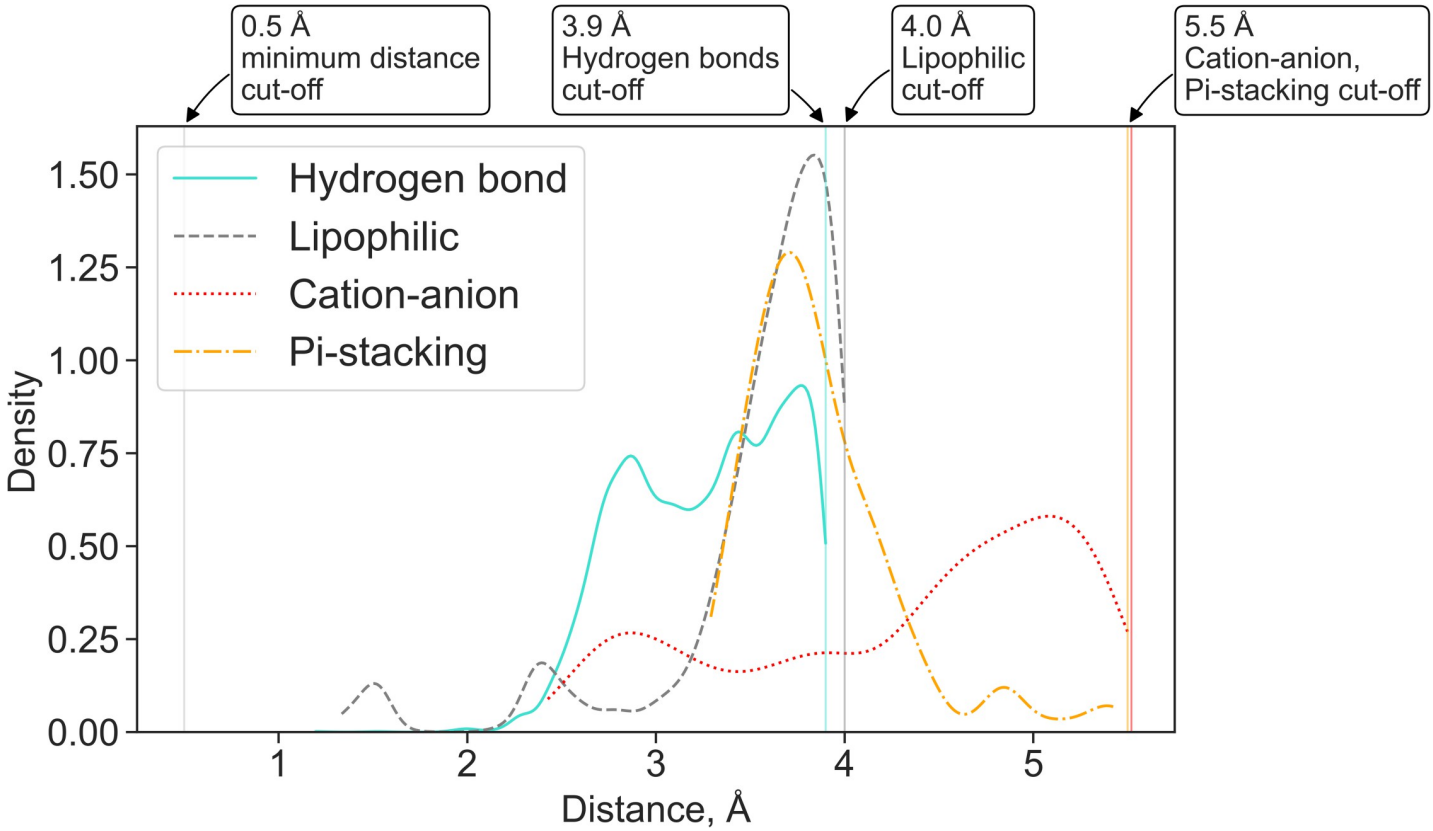

Fig 4. A kernel density estimate plot of the distribution of lengths for four most frequently observed non-covalent interactions in a dataset of experimentally solved RNA-ligand structures. For hydrogen bonds, the distance between the non-hydrogen donor atom and the acceptor is reported. The minimum and maximum cut-off values for each interaction are marked with vertical lines.

https://doi.org/10.1371/journal.pcbi.1009783.g004

adenine to 19.3% for cytosine). Kondo and Westhof, who analyzed 231 RNA-ligand structures, also noted a frequent formation of hydrogen bonds involving the Watson-Crick face of RNA and suggested that the formation of such interactions may be a key to ligand selectivity [78]. The Watson-Crick is the face most frequently forming interactions also in nucleotide-protein complexes, while the sugar face is rarely recognized by either the side-chain or peptide backbone of amino acid residues [79].

Analysis performed for lipophilic interactions indicates that most contacts between RNA and ligands are also made with nucleobase atoms (74%, Fig 6 and S10 Table). For Pi-anion interactions, 82% of bonds are formed with nucleobases, with RNA acting as an anion acceptor (in the remaining 18% of Pi-anion interactions, the phosphate group of RNA acts as an anion donor; see S11 Table). Interestingly, we observed the formation of a halogen bond mostly with the ribose atoms (5 cases, 83%) with a single interaction detected with a nucleobase atom (see S12 Table). However, the overall number of these interactions in our dataset is low (6 interactions), therefore the observed trend may not be reliable. Due to the molecular features of the RNA, all Pi-stacking and Pi-cation bonds are formed exclusively with nucleobases. Taken together, most of the observed interactions of all kinds are formed with the nucleobases (60.53%, while 21.63% and 17.84% with phosphate and ribose fragments, respectively), which highlights a vital role of this region of RNA in structure recognition (see S13 Table).

Presented data can pave the path toward a better understanding of the nature of such interactions and define the main driving forces responsible for forming these complexes. Medicinal

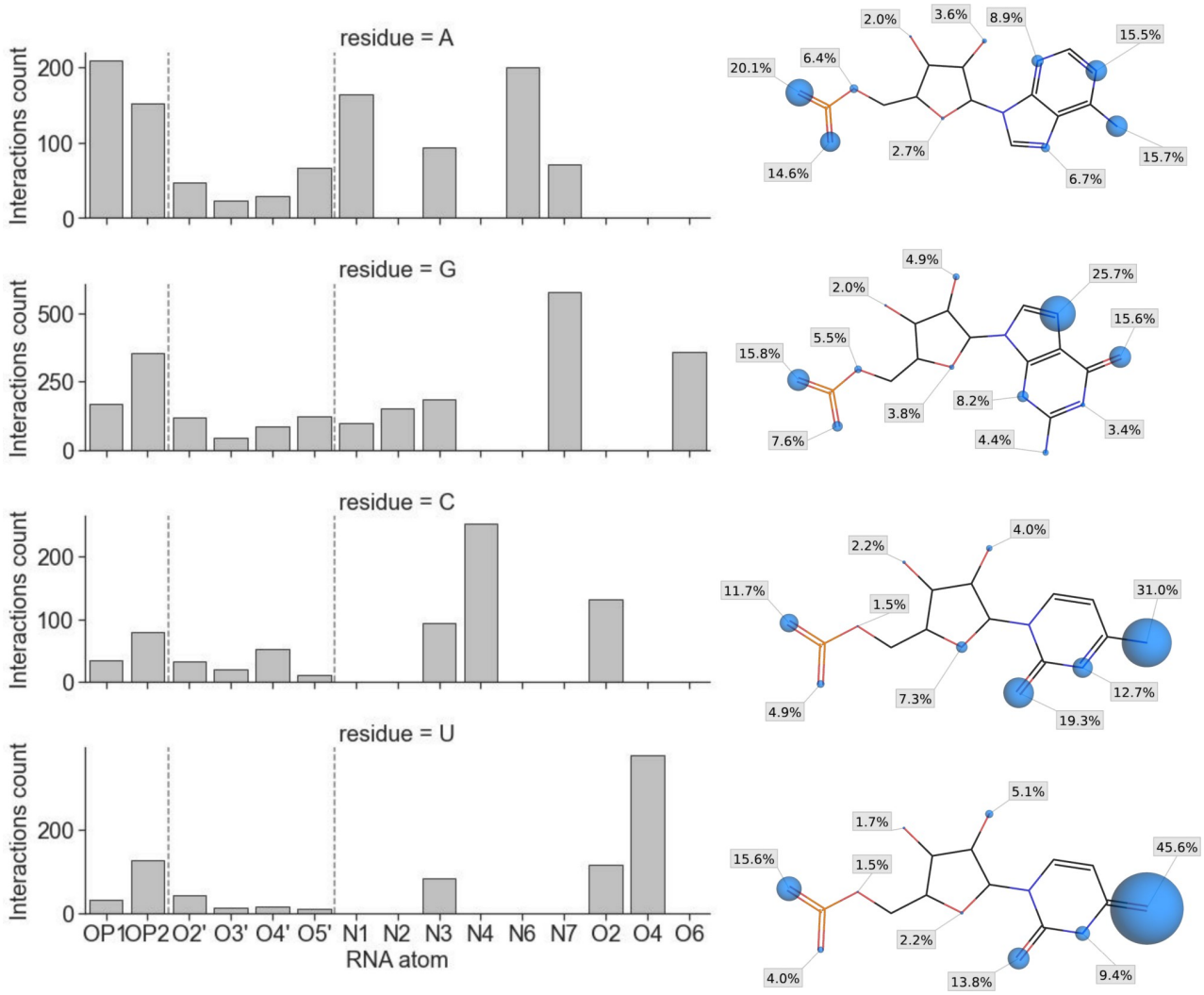

Fig 5. The number of hydrogen bonds formed with ligand molecules grouped by the RNA atoms (bar plots, left; vertical dashed lines separate atoms of a phosphate group, ribose, and nucleobase) and interaction sites statistics for each atom with a percentage of all hydrogen bonds formed by this residue (right column; the radii of spheres are proportional to the percentage values).

https://doi.org/10.1371/journal.pcbi.1009783.g005

chemists can directly use provided information on the preferred interaction distances and interaction sites to support the rational design or modification of existing small molecule ligands to improve their binding affinity or selectivity toward RNA molecules.

SIFt-based structure similarity assessment

The most widely used criterion of the accuracy of macromolecule-ligand modeling tools is the ability to reproduce the binding mode of ligands. This is usually measured by calculating the root-mean-square deviation (RMSD) between the non-hydrogen atoms of the ligand in the experimentally determined structure and the corresponding atoms in the modeled pose. Although widely used for the assessment of molecular docking programs [80], where the macromolecule structure is usually kept rigid, RMSD has several shortcomings, mostly seen in simulations involving the flexibility of both interacting partners. It may happen that although the structure of the predicted complex is very similar to the reference structure and the binding

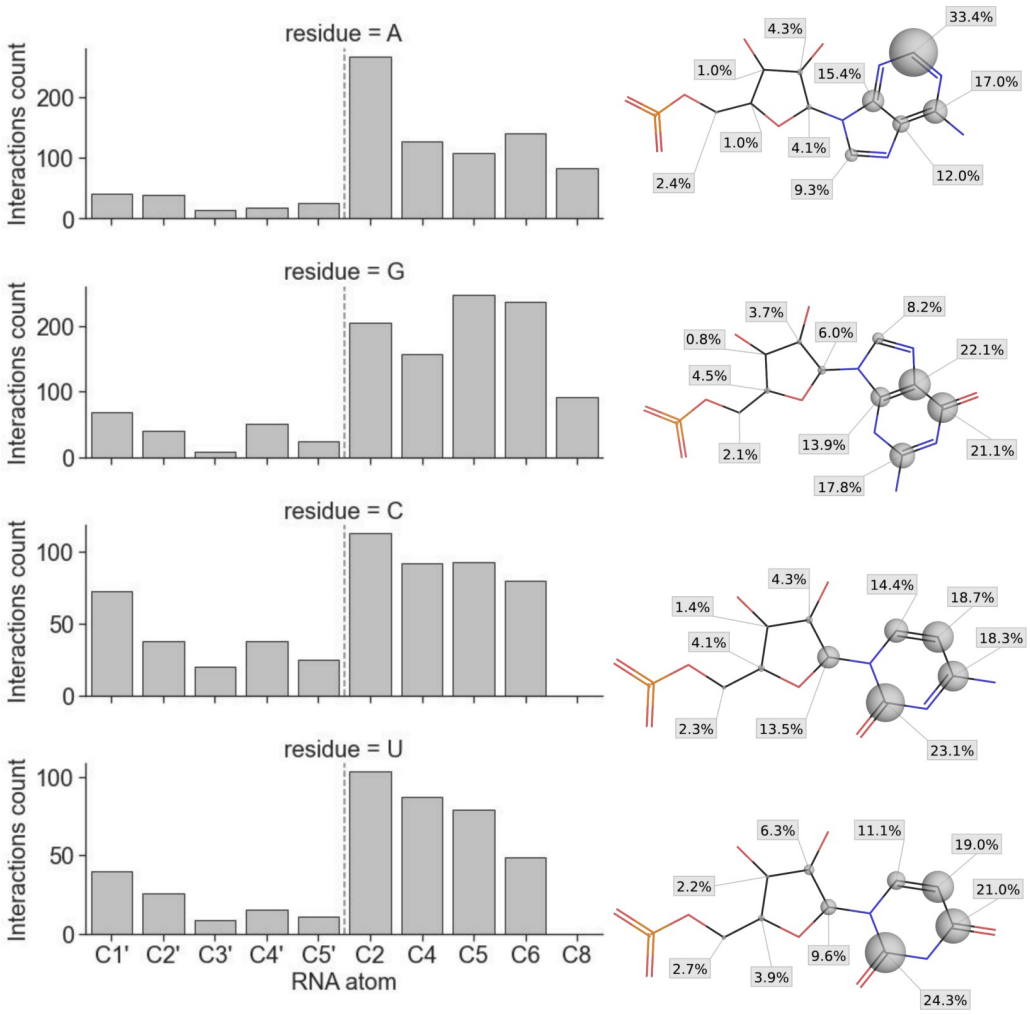

Fig 6. The number of lipophilic interactions formed with ligand molecules grouped by the RNA atoms (bar plots, left; vertical dashed lines separate atoms of ribose and nucleobase) and interaction sites statistics for each atom with a percentage of all hydrogen bonds formed by this residue (right column; the radii of spheres are proportional to the percentage values).

https://doi.org/10.1371/journal.pcbi.1009783.g006

mode of the ligand is perfectly recapitulated, the RMSD value for the whole complex is very high when the model is compared with the reference structure. For example, this can be caused by the fluctuations of the receptor structure in a region distant from the binding site, which does not influence the predicted binding mode of the ligand nor the shape of the binding pocket but has a negative impact on the calculated RMSD value. Another example is when the fragment of a ligand, which is not involved in the binding process (e.g., a solvent-exposed group), deviates from the reference ligand structure and thus results in a high RMSD value (see [81] for examples). In addition, the RMSD is ligand-size dependent, tangling direct comparison of RMSD values obtained for molecules of different sizes as well as using a fixed RMSD threshold as a criterion for successful molecular docking.

To circumvent the above mentioned drawbacks of RMSD, several alternatives have been proposed, however, they were designed and tested, to the best of our knowledge, exclusively for protein complexes. The list includes methods such as RSR (Real Space R-factor, which measures how well a group of ligand atoms fits the experimental electron density, [82]),

GARD (Generally Applicable Replacement for rmsD, which takes into account relative importance to binding of atoms, [81]), TFD (Torsion Fingerprint Deviation, which compares conformations of molecules [83]), or SuCOS (for assessing shape complementarity and overlapping of chemical features [84]). Also, several metrics utilizing a comparison of contacts between proteins and ligands have been proposed. Ding *et al*. developed the Contact Mode Score (CMS), a metric to assess the conformational similarity based on intermolecular protein-ligand contacts [85]. CMS is expressed as the Matthews correlation coefficient (MCC) between contact matrices generated for the reference structure and the investigated complex. It was shown to be a valuable metric to evaluate results of flexible docking, which at the same time considers the changes upon ligand binding. In the IBAC approach (Interactions-Based Accuracy Classification, [86]) proposed by Kroemer *et al*., the scoring is derived from the comparison of (manually defined) key interactions of the reference protein-ligand complex and the docked pose. Although defining the key interactions may be perceived as subjective, the IBAC method was proved to be a more meaningful measure of docking accuracy for the examined test set than RMSD. Balius *et al*. proposed an FPS score (footprint similarity, [87]) which is derived from the comparison of electrostatic, steric, and hydrogen bonding energy profiles for protein-ligand complexes using the Pearson correlation coefficient.

The similarity of interaction fingerprints for protein-ligand complexes was also explored as a measure for binding mode similarity. Drwal *et al*. analyzed four protein targets to compare binding modes of fragments, crystallization additives, and drug-like molecules [88]. For a given target, they calculated a consensus fingerprint containing the relative frequency of each interaction type with each residue. The similarity between a docking pose fingerprint and a consensus fingerprint was calculated using the Tanimoto metric for continuous variables (and thus, not limited to values "0" and "1"). Leung *et al*. benchmarked the PLIF similarity (Protein-Ligand Interaction Fingerprint) as a metric for evaluating docking of the ligands to proteins [84]. They concluded that this metric, contrary to ligand-centric ones (such as the RMSD and SuCOS), was able to capture information about interactions across multiple crystal structures of ligands bound to the same protein, making it a handy feature for experiments where multiple protein conformations are used.

Inspired by structure-centric methods developed for assessing protein-ligand complexes, we examined the applicability of the interaction fingerprints generated by the fingeRNAt as a measure for RNA-ligand complexes' similarity. As input structural data, we used predictions submitted by participants of the RNA-Puzzles (https://www.rnapuzzles.org, [89])—a collective experiment for blind RNA structure prediction. In the RNA-Puzzles round 23, the task was to predict structures of a Mango-III aptamer in complex with biotinylated TO1 dye, based on the sequence of RNA and structure of the ligand. The main criterion of the evaluation of the prediction's quality was the RMSD, deformation index (DI), and interaction network fidelity (INF, [90]) of RNA. Seven groups submitted their models of the target complex (RNA with ligand)—Bujnicki, Chen, Das, Ding, Dokholyan, Adamiak (code-named RNAComposer), and Xiao. The structures of ligands provided by the last group contained structural errors, which made it impossible to calculate RMSD values for the ligand; however, the calculation of interaction fingerprints was possible due to the structure-agnostic (structure-independent) nature of the algorithm proposed in this work.

First, we compared the correlation of the RMSD and INF of RNA and RMSD of the ligand with interaction fingerprint similarity. The similarity of interaction fingerprints (and, in general, of bit vectors) can be expressed in a number of ways. Currently, the fingeRNAt package offers eight methods for calculating the similarity or distance of bit vectors. Here, we used a widely used metric for the comparison of interaction fingerprints—the Tversky similarity [84,91]. It focuses on the recapitulation of the true interactions (formed in the reference

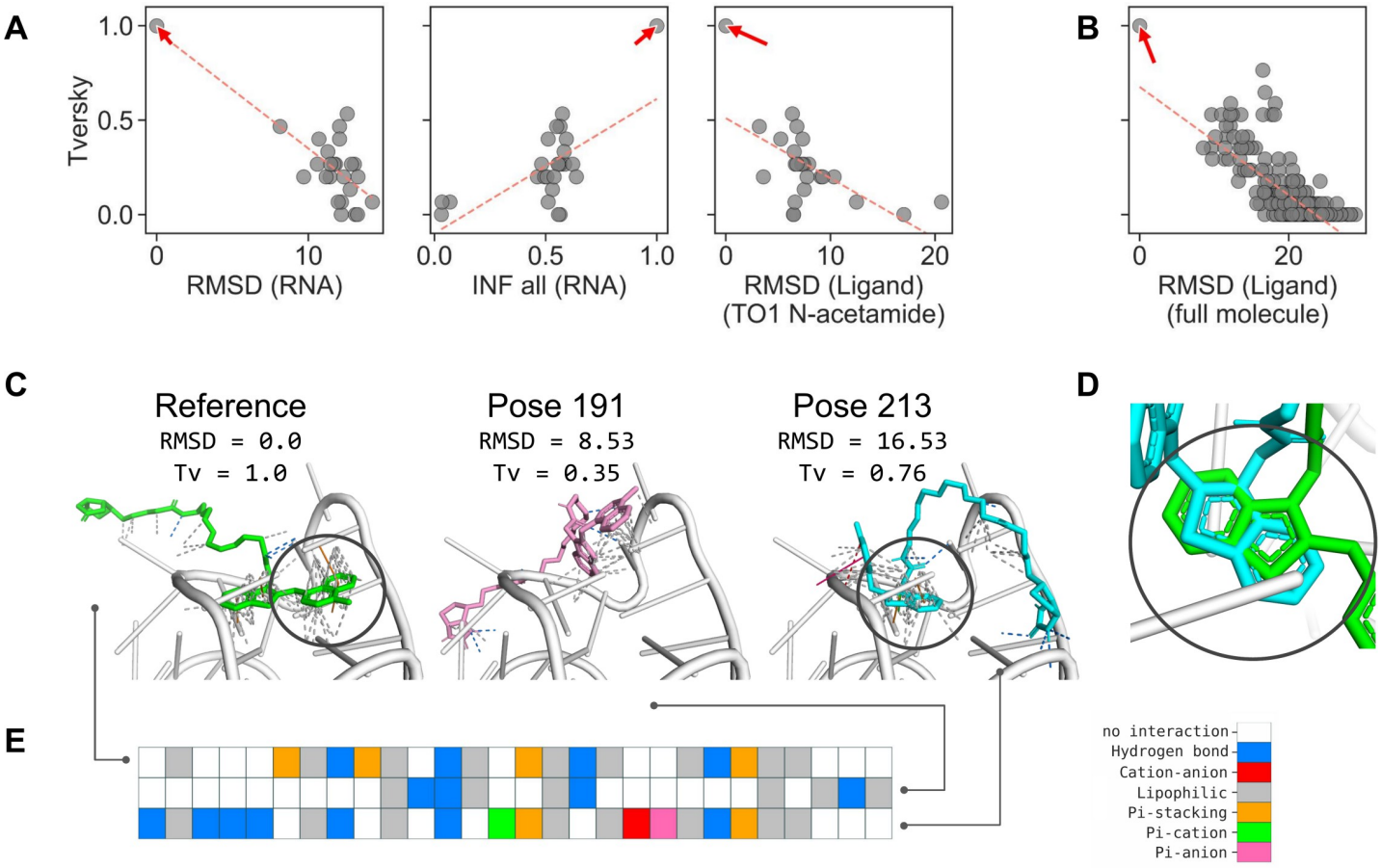

Fig 7. Relationship between structure-centric similarity metrics (RMSD and INF) and fingeRNAt interaction fingerprints similarity (Tversky). (**A**) Relationship between RMSD of RNA, INF all of RNA, RMSD of ligands, and Tversky similarity of interaction fingerprints for models submitted for the RNA-Puzzles round 23 and (**B**) for the redocking experiment. The reference structure is marked with an arrow, and the linear least-squares regression is marked with a red dashed line. (**C**) Selected poses from the docking experiment with ligand RMSD and Tversky similarity of fingerprints, (**D**) the overlay of the benzothiazole fragment of the ligand of the reference structure, and pose 213, and (**E**) the corresponding color-coded interaction fingerprint. Benzothiazole moiety was marked with a gray circle in (C) and (D).

https://doi.org/10.1371/journal.pcbi.1009783.g007

complex) in the predicted model of a complex and ranges from 0 (investigated model has no interactions which are present in the reference complex) to 1 (model has all interactions the reference has).

We observed only a weak correlation between RMSD (of RNA or ligand), INF, and Tversky similarity (R^2 for linear least-squares regression ranges from 0.31 to 0.61; see Fig 7A and S14 Table). The detailed analysis of the rank list shows that most of the best models in terms of RMSD, DI, or INF do not recapitulate the interaction network formed between RNA and ligand in the experimentally solved structure (S15 Table). Only the model submitted by the Das group (Das_07) had relatively good values of RMSD (for both RNA and ligand), DI, and interaction fingerprint similarities (S8 Fig). Also, the mutual similarity of interaction networks in the submitted models is relatively low (with average Tversky similarity equal 0.24 and median similarity 0.18), indicating that in most complexes, the proposed binding mode is unique (S9 Fig).

As a complementary experiment, we performed molecular redocking of the ligand to the reference structure of the Mango-III aptamer. Again, we observed only a weak correlation between RMSD of ligand and Tversky similarity of interaction fingerprints (R^2 for linear least-

squares regression equal 0.60; see Fig 7B and S14 Table). For the detailed analysis, we selected two poses proposed by the docking program—one with relatively low (good) RMSD of ligand but low (unfavorable) fingerprint similarity (pose 191) and one with relatively high (unfavorable) RMSD value but high (good) fingerprint similarity (pose 213) (Fig 7C). Comparison of the two selected poses with the reference complex confirms that pose 191, despite its better RMSD value, has a very different binding mode than the reference ligand. In this case, only six out of 17 interactions were correctly predicted. Conversely, pose 213 with a very high RMSD value recapitulates 13 out of 17 interactions formed by the reference structure. This is especially pronounced by the correct prediction of the position of the central benzothiazole ring, although in a different orientation (Fig 7D).

Moreover, for the redocking data, we did not observe a strong correlation between RMSD of the ligand and any of the eight metrics expressing similarity/distance of interaction fingerprints (for the mutual relationship plots of RMSD and bit vector similarity/distance values, see S10 Fig), meaning that the assessment method proposed by us is distinct from other metrics and could be easily employed as another mean of structure's prediction scoring. Also, lower resolution fingerprints (SIMPLE, PBS) can be used for the rough estimation of the binding mode—the similarity of these fingerprints does not correlate with the RMSD of the ligand as well (S11 Fig).

This observation is also confirmed by the analysis of a larger dataset. For the redocking experiment of 144 RNA-ligand complexes (100 docked poses + one reference pose for each complex), we observed, in most cases, only a weak correlation between RMSD of a ligand and interaction fingerprint similarity (see S12 Fig and S16 Table). For example, the average R^2 value for Tanimoto and Tversky similarity was as low as 0.417 and 0.380, respectively, and the average Spearman rank correlation coefficient was -0.523 and -0.491, respectively. A weak correlation was also observed for other similarity metrics and rank correlation measures.

In summary, the similarity of the interaction fingerprints can be used as an alternative to the RMSD measure for comparing structures of complexes. The proposed approach focuses on the interactions formed between interacting partners, while the actual location of particular atoms and groups of ligands is not considered. By using different fingerprint similarity measures, various aspects of interaction similarity can be emphasized. While the Tversky similarity expresses the number of correctly predicted interactions, the Tanimoto similarity expresses the overall mutual similarity of the interactions in compared complexes.

These results are in line with data obtained earlier for protein-ligand complexes. It was shown that the interaction data could improve docking accuracy by selecting the binding pose closer to the reference structure and recapitulate models poorly assessed by other methods [92,93].

Fingerprints for clustering interactions and detecting preferred patterns

As stated in the previous section, RMSD is the most widely used measure for assessing the similarity of ligands in complex with macromolecules. It may be treated as a distant proxy for comparing interactions formed in two investigated complexes, assuming that two molecules that are close in space will form the same (or similar) interaction network. As we showed, this is not always true, as two ligands with low values of RMSD may form very different interactions with the receptor and *vice versa*. Another limitation of the RMSD used for comparing the binding modes in two complexes is the restriction that both ligands must be the same. Although some recently published methods partially circumvent this constraint by enabling the calculation of the RMSD of molecules sharing some structural features (LigRMSD, [94]), application of RMSD is still limited to structurally similar compounds.

In this experiment, we tested the applicability of SIFts for grouping small molecule compounds of different chemical structures but with akin binding patterns. We hypothesized that molecules forming similar types of interactions with the molecular target would have similar binding properties, such as biological activity. We composed a dataset of diversified small molecule ligands with known or putative activity toward the HIV-1 trans-activation response (TAR) element. This is a medically important and relatively well-explored RNA target with a substantial amount of available experimental data. This includes solved three-dimensional RNA structures and experimentally validated ligands. Also, a structure-based virtual screening with a subsequent ligand activity validation was performed, resulting in new HIV-1 TAR binding ligands [95,96] (for a review, see [97]). The library consisted of 30 active and 1478 inactive molecules (see Materials and methods section Datasets for the detailed description of the library preparation). To avoid artificial enrichments observed when multiple active ligands have a very similar chemical structure and thus possibly have an analogous binding mode (so-called "analog bias"), we ensured that these compounds are dissimilar to each other and belong to different chemical classes (see S13 Fig). Using molecular docking, we predicted the structure of these compounds in complexes with HIV-1 TAR RNA and calculated SIFts for each RNA-ligand complex.

In a pool of 1508 docked compounds, there were 1149 unique fingerprints, and 998 compounds (66.2%) had a unique fingerprint (i.e., the fingerprint that is unique for this molecule only). All 30 active compounds had a unique fingerprint. We used PCA for dimensionality reduction and *k*-means clustering for grouping compounds with similar fingerprints, i.e., forming a similar interaction network with the target RNA. This method offers a reasonable separation of clusters (average silhouette score of 0.516 for 15 clusters, Fig 8A and 8B; for the performance of clustering for various numbers of clusters, see S17 Table). In five clusters, the ratio of active compounds was significantly different than in the input dataset (p-value ≤ 0.05, clusters 8–12), although it did not include the cluster with the highest ratio of active compounds (cluster 7, 6.67% of active compounds, p-value = 0.29, Fig 8B; for clusters composition and statistical significance analysis, see S18 Table). In clusters 8, 11, and 12, the ratio of active compounds is significantly higher than in the input population (with the percentage of active compounds 2.38%, 4.51%, and 4.00%, respectively, p-value ≤ 0.05). Thus, we conclude that the interaction pattern present in the latter three clusters may be "favorable" for active compounds. Conversely, two clusters (9 and 10) do not contain any active compounds (and this number is significantly lower than in the input population), which could indicate that interaction networks formed by members of this group are "unfavorable" for HIV-1 TAR binding ligands.

We performed the same type of analysis for the dataset with lipophilic interactions removed. These interactions are most common among all complexes, and we hypothesized that they might introduce noise into the fingerprint data. If the lipophilic interactions are removed, the separation of the clusters is even better (average silhouette score of 0.72 for nine clusters, Fig 8C and 8D). We detected two clusters in which the ratio of active compounds was significantly lower than in the input dataset (p-value ≤ 0.05, clusters 1 and 6) and which did not contain any active ligand. The ratio of active compounds in cluster 0 was 4.97%, which is 2.5 times higher than in the input dataset (however, this result is not statistically significant, p-value = 0.08). This analysis enabled us to define nine groups of interaction patterns formed by the ligands in complex with HIV-1 TAR, some of which are more "favorable" or "unfavorable" for active ligands. Comparing the results of this analysis with the previous one, where all interaction types were present, we conclude that lipophilic interactions introduced noise to the fingerprint data, making generated clusters fuzzier. On the other hand, it enabled a better separation (more distinctive clusters) of active and inactive compounds. The moderate

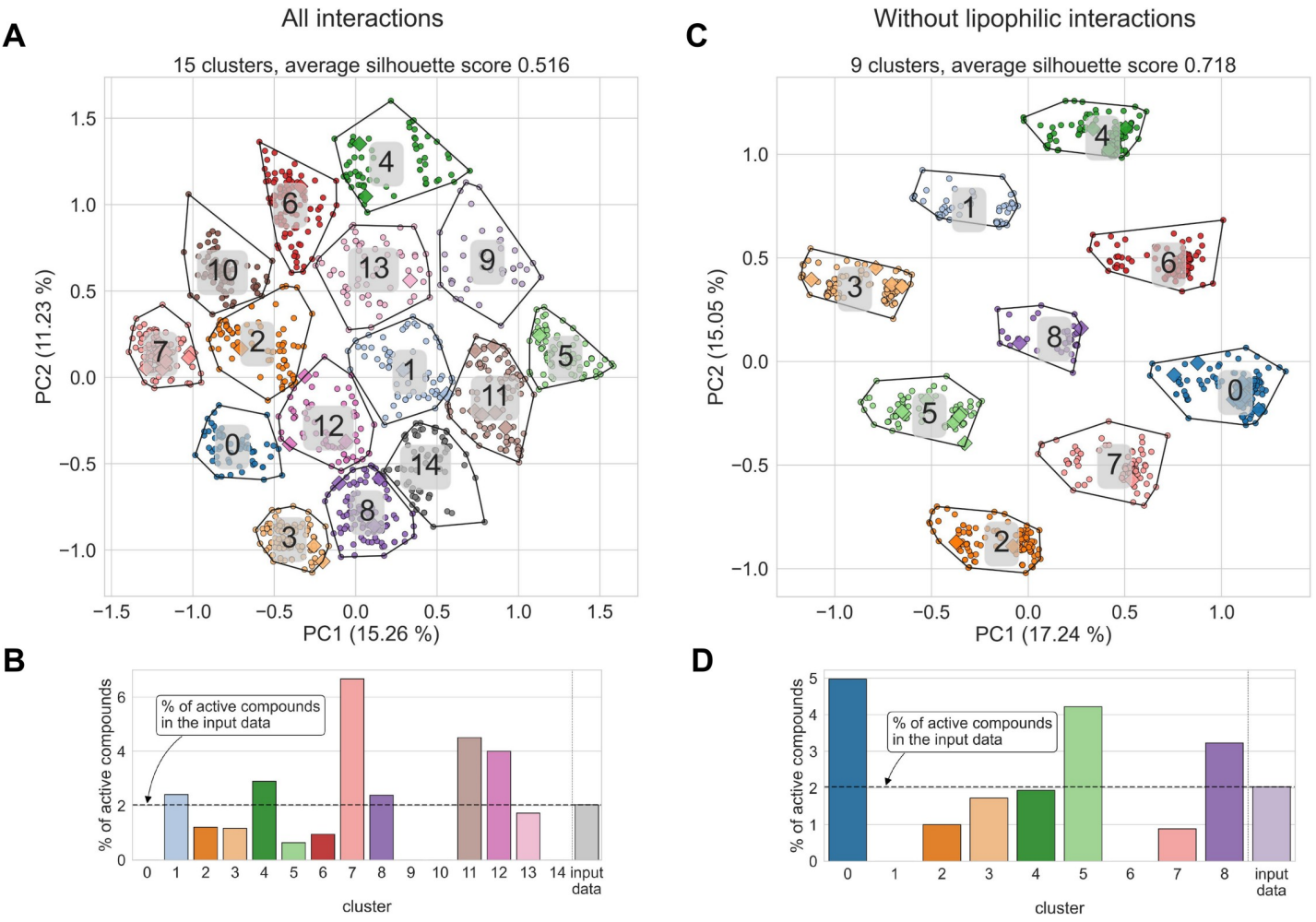

Fig 8. Interaction fingerprints used for visualization and clustering of predicted interaction patterns calculated for data from molecular docking of a set of active and inactive compounds to HIV-1 TAR structure (PDB entry 1UTS). (**A**) Interaction fingerprints mapped on the 2D space using PCA with color-coded clusters for the all-interactions dataset and (**C**) with lipophilic interactions removed. (**B**) Percentage of active compounds in each detected cluster compared to the value obtained from the input dataset, calculated for all-interactions dataset and (**D**) obtained from the dataset with lipophilic interactions removed.

https://doi.org/10.1371/journal.pcbi.1009783.g008

enrichment observed in the active-containing clusters could be a consequence of the docking program's inability to recapitulate the correct binding pose of some active ligands. This shortcoming could be at least partially circumvented by using an optimized combination of docking program and scoring function to capture the more realistic structure of the RNA-ligand complex. Unfortunately, the availability of RNA-specific docking programs is currently very limited.

Taken together, dimensionality reduction and clustering of fingerprints offer a high-throughput method of analysis of structural data of complexes. As shown, it may be used to define groups of ligands interacting with a receptor in a similar way and thus sharing similar properties (such as biological activity or lack of thereof). This kind of analysis may also be used in high-throughput virtual screening to generate a diversified set of compounds (in terms of forming diverse interactions with the receptor of interest) for further testing for their biological activity. Clustering fingerprint data may help select groups of compounds forming similar interactions to the active compounds, thus having a higher probability of being biologically active. It could be directly applied as a post-processing step of virtual screening with molecular

docking. Decomposition methods other than PCA may be used, leading to potentially meaningful results (such as obtaining clusters enriched with molecules with a given property; see S14 Fig). Contrary to the ligand-based metrics (such as the RMSD), the presented approach is not limited to the same or structurally similar ligands.

Summary

Interactions between nucleic acids and ligands play a pivotal role in many biological processes. Characterization of these interactions may elucidate our understanding of those phenomena and help to explain the nature of molecular recognition. This knowledge may also be utilized to modulate the binding process in the desired way, for example, by small molecule inhibitors binding to RNA and preventing its interactions with a molecular partner. The presented software tool, fingeRNAt, detects and characterizes interactions in nucleic acid-ligand complexes. We showed its applications in different bioinformatics problems to help answer structural biology and drug development questions. These included analyzing experimentally solved RNA-small molecule ligand complexes deposited in the PDB database to determine the statistics of non-covalent bond types and their features. We also proposed SIFts' similarity as an alternative measure to RMSD. Contrary to the RMSD, SIFt-based metrics do not depend on the receptor conformation nor the ligand structure and focus on how well the interaction network is recapitulated in the model compared to the reference complex. Besides, we presented an application of molecular fingerprints for the clustering of complexes. Fingerprint data, processed with multidimensional scaling and clustering, yields groups of complexes with similar binding patterns. We demonstrated that these clusters might be enriched with compounds with desired properties, such as biological activity, facilitating a high-throughput analysis of the structure-activity relationship and visual analyses of multiple complexes. The accompanying PyMOL plugin enables visualization of the detected interactions, an inspection of the results, and preparation of publication-quality images.

fingeRNAt is relatively fast. Calculation of fingerprint type FULL for the redocking of guanidine ligand to guanidine III riboswitch (RNA with 39 residues, ligand with 10 atoms, 100 ligand poses) took less than 7 seconds, while calculating the Tanimoto similarity matrix took under 2 seconds (calculated on Ubuntu Linux 20.04 with Intel(R) Core(TM) i5-8400 CPU and 32 GB RAM; see S15 Fig and S19 Table for the detailed benchmark).

Applications of the fingeRNAt-generated SIFts significantly go beyond the ones described in this manuscript. The program may be used to generate interaction profiles of nucleic acid with ions, which may help understand ion binding preferences and enable comparing ion profiles between multiple structures. Moreover, fingeRNAt ideally fits the pipeline for analysis of molecular dynamics trajectories, indicating forming and breaking non-covalent bonds during a simulation. SIFts, paired with the bioactivity data for small molecules, may also be used to develop predictive models for the molecular target of interest.

The roadmap of the software development includes detection of less frequently observed types of non-covalent bonds, such as halogen-Pi [98], "mixed" type bonds, such as cation-anion hydrogen bonds [99], or a separate class of anions binding to amino, imino, and hydroxyl groups of RNA [100]. Recent publications suggest that these interactions may greatly contribute to the molecular recognition process (for a recent review on "unusual" interactions, see [101]). New levels of the resolution of the fingerprint will include differentiation of strong and weak interactions (e.g., for hydrogen bonds). We also plan to define geometrical rules for ions and water molecules to more reliably detect such interactions.

We are confident that fingeRNAt will be highly useful for the bioinformatics community and will facilitate research on nucleic acid interactions.

Materials and methods

The fingeRNAt method

fingeRNAt is a set of Python 3 programs for detection, classification, and analysis of interactions formed within nucleic acid-ligand interactions. It consists of three main tools, each serving a different purpose.

fingeRNAt.py

fingeRNAt.py is a program for the detection and classification of non-covalent nucleic acid-ligand interactions. It can be run from the command line or via the graphical user interface. As an input, it takes a 3D structure of a receptor (RNA or DNA) and a file containing ligands, which form a complex with this macromolecule. fingeRNAt.py accepts six types of ligand molecules: small molecules, proteins, metal cations, DNAs, RNAs, and LNAs (locked nucleic acids) (Fig 9). The output is a fingerprint—a bit vector containing information on the declared interactions detected between the receptor and the ligand.

Input. For all ligand types but metal cations, the program requires two input files: (i) a receptor file, which is an RNA or DNA structure in the pdb format (one model per file), with explicit hydrogens added, and (ii) a ligand file in sdf format, which may contain multiple structures. It is possible to calculate profiles of inorganic ions' interactions with the receptor; in such a case, only one input file containing an RNA or DNA structure in the pdb format should be passed, and inorganic ions should be present within the same file. fingeRNAt will then treat all inorganic ions as ligands and calculate SIFt for each residue—ion pair.

The receptor's structure may contain water and metal cations, but any ligands or buffer molecules must be removed prior to the analysis. Input ligand molecules must have assigned desired protonation states and formal charges. Formal charges on the phosphate groups do not need to be explicitly indicated in the receptor molecule, as fingeRNAt.py always treats OP1 and OP2 atoms as negatively charged anions.

Output. The output is a SIFt calculated for each nucleic acid-ligand pose, saved in a tab-separated (tsv) file, with separate columns for each residue and interaction type. Optionally, the human-readable file (also in tsv format) with detailed information on detected interactions can also be created (-detail option), which includes a listing of all detected interactions, spatial coordinates of the interacting partners, and distances between interacting atoms or aromatic rings.

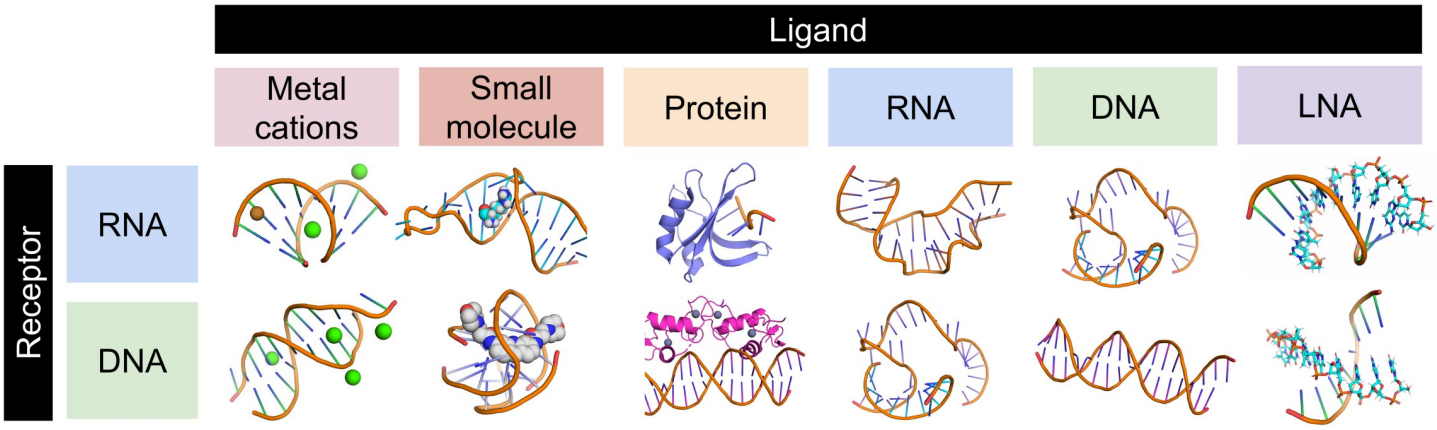

Fig 9. Combinations of receptor and ligand types accepted by the fingeRNAt.py.

https://doi.org/10.1371/journal.pcbi.1009783.g009

Implementation. For each ligand in the input file, the program iterates over all nucleic acid's residues and detects interactions of a given type. If the interaction is detected, the respective bit in the fingerprint is set to "1" and to "0" otherwise. Interactions can be detected at three resolutions: (i) low-resolution SIMPLE variant detects contacts between any atom of each of nucleic acid residue and ligand; (ii) medium-resolution PBS variant detects interactions between atoms of Phosphate, Base, and Sugar fragments of a nucleic acid residue and a ligand; (iii) high-resolution FULL variant detects and classifies the type of non-covalent interactions between a nucleic acid residue and a ligand; in this variant, fingeRNAt detects nine hard-coded interaction types: (i) hydrogen bonds, (ii) halogen bonds, (iii) cation-anion interactions, (iv) Pi-cation interactions, (v) Pi-anion interactions, (vi) Pi-stacking interactions, (vii) metal cation-mediated: magnesium, potassium, sodium, and other metal cation-mediated, (viii) water-mediated interactions, and (ix) lipophilic interactions. We distinguished magnesium, potassium, and sodium cations as those are the most prevalent metal cations in nucleic acid complexes [102,103]; other metal cation-mediated interactions refer to interactions mediated by not the aforementioned ions.

Additional interaction types can be defined by plugins encoded in a human-readable yaml file. Interacting atoms or groups are defined by SMARTS patterns, separately for the receptor and ligand. Interaction criteria are defined based on the distance, distance and angle, or distance and dihedral angles between atoms/groups. The sample plugin file contains a definition of five interactions: (x) any interaction (any contact between nucleic acid and ligand), (xi) polar interactions, i.e., hydrogen bonds without angle restraints, (xii) weak polar interactions, i.e., weak hydrogen bonds without angle restraints, (xiii) $n \rightarrow \pi^*$ interactions, (xiv) weak hydrogen bonds, and (xv) halogen multipolar interactions. Taken together, with the default configuration of the fingeRNAt.py, it is possible to detect 15 different interaction types.

For the summary of fingerprints' variants and features, see Fig 10. Geometrical criteria for interactions were taken from the literature: ([56,104] for hydrogen bonds, [105] for halogen bonds, [106] for cation-anion interactions, [107,108] for interactions with p-orbitals: Pi-stacking and Pi-ion interactions, [109] for ion-mediated interactions, [110] for water-mediated interactions, [62] for lipophilic interactions), [111] for $n \rightarrow \pi^*$ interactions, and the PLIP algorithm (protein-ligand interaction profiler, [30]). See S20 Table for the geometrical criteria summary.

Calculated fingerprints can be further post-processed by three wrappers: (i) ACUG for reporting interactions for four nucleotide types of the nucleic acid (A, G, C, and U or T), (ii) PuPy for reporting interactions for purines and pyrimidines, and (iii) Counter for reporting the total number of occurrence of a given interaction type. The detailed description of parameters accepted by the fingeRNAt is available in S21 Table.

fingeRNAt.py uses the OpenBabel Python module to parse input files and perform most cheminformatics calculations (such as hydrogen bonds acceptors/donors detection) and the RDKit Python module to detect aromatic rings and lipophilic atoms in ligands [57,58,112]. A detailed description of the algorithm of fingerprints' calculations is available in the program manual in the code repository. The summary of detected molecular features together with methods used can be found in S22 Table. If desired, the interaction definitions (such as distance threshold can be easily modified in a configuration file (for default values used in the program, see S20 Table).

fingerDISt.py

In addition to the main program, we provide an auxiliary tool, fingerDISt.py, to calculate different types of distances between fingerprints. It supports eight metrics: Tanimoto coefficient, Cosine, Manhattan, Euclidean, Square Euclidean, Half Square Euclidean, Soergel [113], and

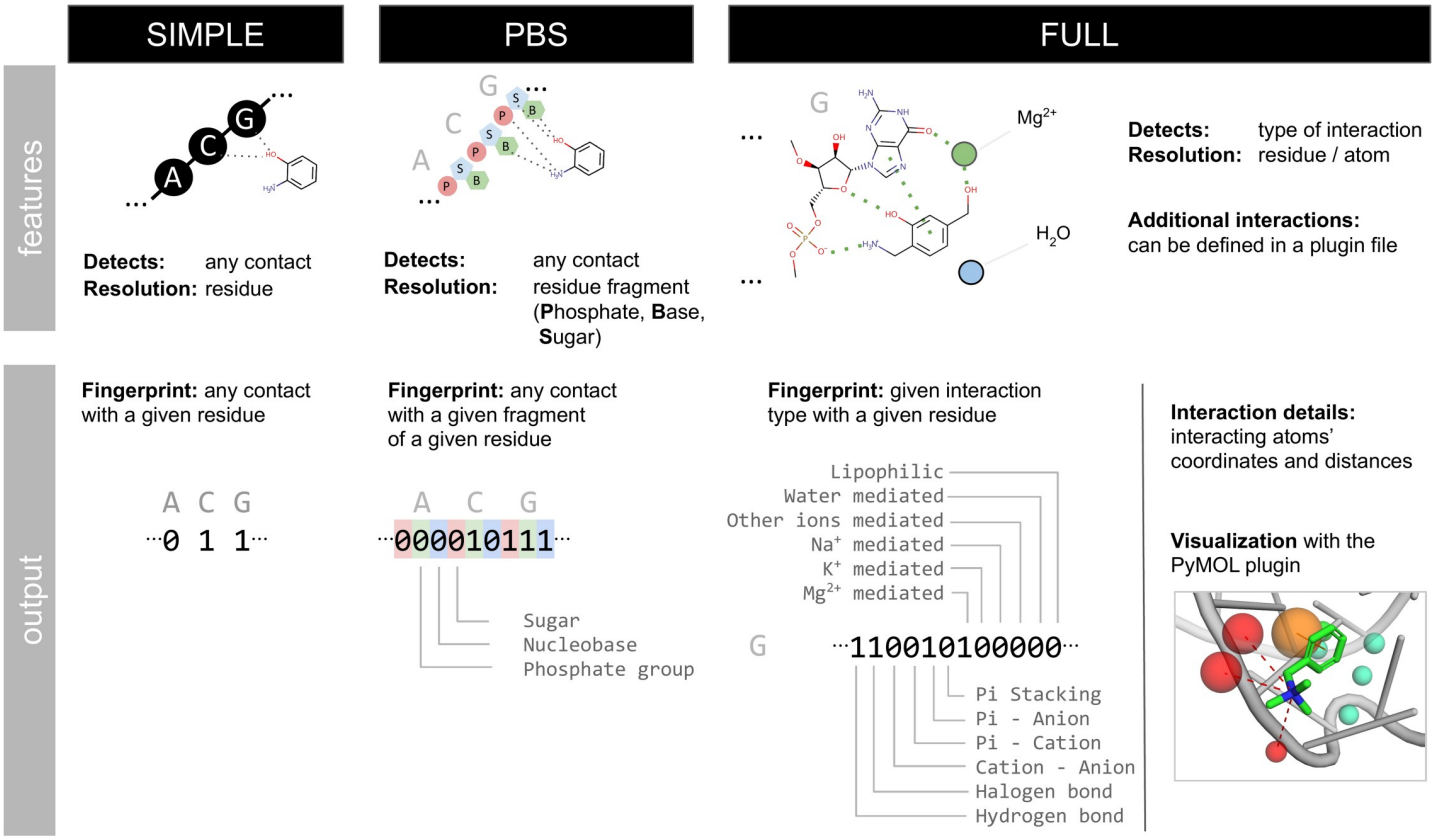

Fig 10. Fingerprints' variants available in the fingeRNAt.py and their output.

https://doi.org/10.1371/journal.pcbi.1009783.g010

Tversky distances [91]. fingerDISt.py accepts SIFts' tsv files generated by the fingeRNAt.py and returns a tsv file with a distance matrix for the fingerprints (all vs. all).

PyMOL plugin

The PyMOL plugin offers a convenient method of visualization of interactions detected and classified by the fingeRNAt program. After loading and processing the interaction file, the plugin generates three groups of objects: (i) Interactions, with objects representing detected interactions (Fig 11A); (ii) Receptor Preferences (aka Receptor's Interactions Hot Spots), with objects representing the spatial occurrence frequency of a given interaction type in receptor atoms (Fig 11B); (iii) Ligand Preferences (aka Ligand's Interactions Hot Spots), with objects representing the spatial occurrence frequency of a given interaction type in ligand binding site (Fig 11C). Occurrence frequency is represented by spheres with centers at the interacting atoms, with the radius proportional to the ratio of interactions of a given type formed by the given interaction site to the total number of all interactions. Additionally, an auxiliary object representing only interacting residues of nucleic acid is created as well as a legend describing the color code and line style used for visualization of different types of interactions and Interactions Hot Spots (Fig 11D).

Known limitations

The analysis of the X-ray structures of protein-ligand complexes shows that both median, minimum, and maximum lengths of the hydrogen bond depend on the interacting partners and

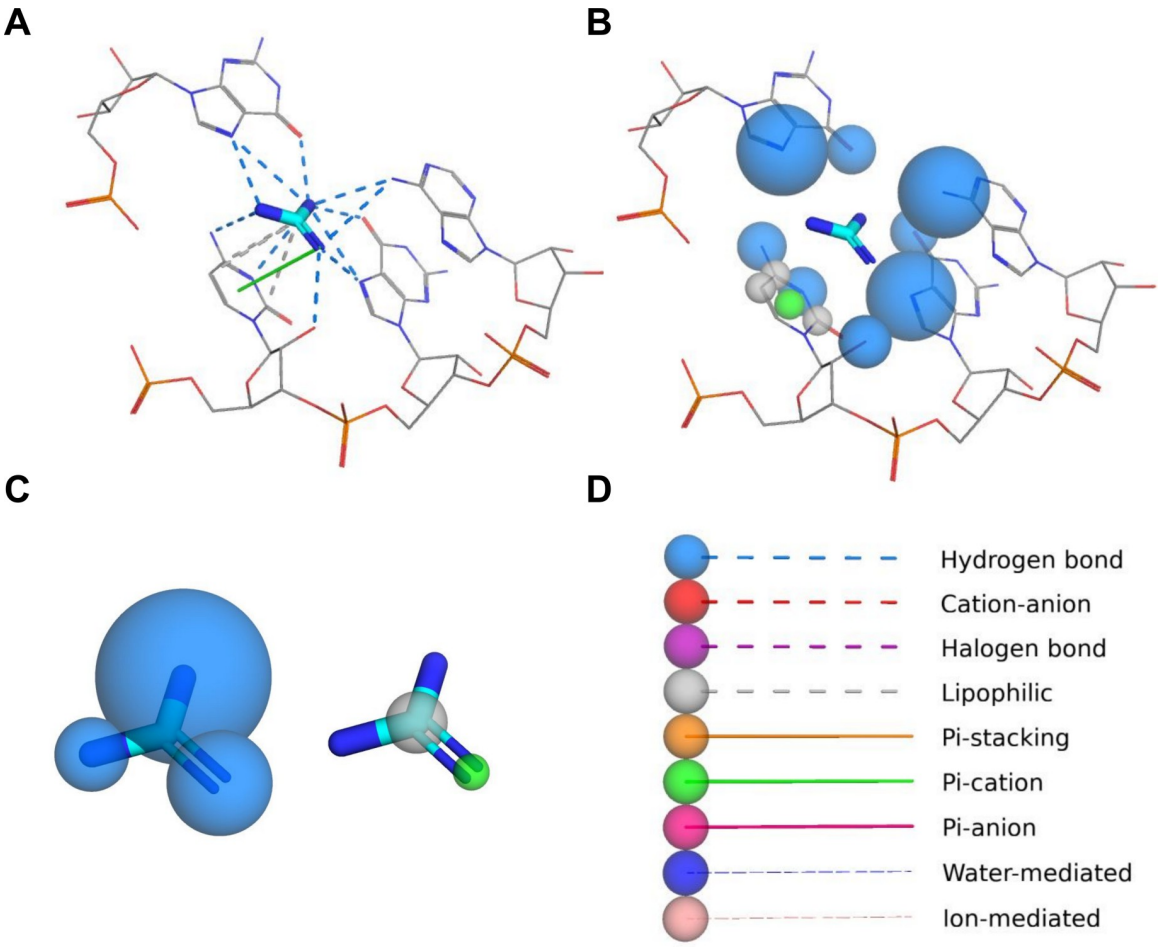

Fig 11. Three types of visualization generated by the PyMOL plugin. (**A**) Detected non-covalent interactions between nucleic acid and ligand; (**B**) Receptor Preferences (aka Receptor's Interactions Hot Spots), with objects representing the spatial occurrence frequency of a given interaction type in receptor atom; (**C**) Ligand Preferences (aka Ligand's Interactions Hot Spots), with objects representing the spatial occurrence frequency of a given interaction type in ligand atoms; (**D**) a legend describing the color code and line styles used for visualization of different types of interactions and Interactions Hot Spots.

https://doi.org/10.1371/journal.pcbi.1009783.g011

are correlated to the hydrogen donor strength. In the fingeRNAt program, we used the single, fixed distance cut-off values for a given interaction type. For example, the default cut-off value for all hydrogen bonds is 3.9 Å, and this distance does not depend on the interacting partners. However, thanks to the plugin-ready architecture of the fingeRNAt, this limitation can be easily circumvented by defining custom hydrogen bond types for atoms/groups of interest and with adjusted distance criteria.

The numbering of residues in the receptor structure must be unique (i.e., there should be no residues with the same number); however, multiple nucleic acid chains are supported. Also, fingeRNAt currently operates on strictly defined file formats (pdb for nucleic acid receptors and sdf for ligands).

Datasets

Non-redundant RNA-small molecule ligand dataset. The non-redundant dataset of complexes of RNA with small molecule ligands was based on the diversified dataset prepared by Philips *et al.* [53]. Briefly, for RNAs with sequence identity >90%, complexed with the same

small molecule ligand, only the structure with the highest resolution was used. Buffer components and inorganic ions other than metal cations were removed. All complexes were inspected visually, removing or correcting ligands with erroneous chemical structure, missing atoms, or ambiguous stereochemistry. Ionization of the ligands was corrected according to the literature data and was supported by pK_a calculations with the Chemicalize platform [114]. The final dataset contains 207 RNA-ligand structures (see S23 Table).

Hydrogens were added to ligands using built-in algorithms (with the OpenBabel and RDKit libraries) and the external software (PyMOL 2.6 h_add command and Chimera 1.14 addh command, while ligand was in complex with the RNA molecule [60]).

HIV-1 trans-activation response element activity dataset. To construct a dataset containing information on active and inactive ligands towards human immunodeficiency virus type 1 (HIV-1) trans-activation response element (TAR), an extensive literature search was performed. Information about the chemical structures of molecules and their binding affinities was collected and tabulated together [115–127]. Ligands, whose binding affinity was not described precisely, or was calculated only on the grounds of cell-based assays, were not included in the database. If the HIV-1 TAR sequence, for which the ligand's activity is described, differed from the sequence of experimentally solved structure 1UTS (taken arbitrarily) within the binding site or its proximity up to 8Å, the ligand was rejected. Structures of the ligands were normalized, and a set of filters was applied to include only drug-like molecules (molecular weight from 90 to 900 daltons, an octanol-water partition coefficient—SlogP from -7 to 9, number of hydrogen bond acceptors up to 18, number of hydrogen bond donors up to 18, number of rotatable bonds up to 18). Next, the PAINS (Pan-assay interference compounds, [128]) filter was applied to exclude ligands for which observed activity was potentially a result of interference with assays and not the interactions with the RNA. Finally, ligands were assigned into one of two groups—active or inactive. The criterion was the activity (expressed as K_D, IC_{50}, MIC, or K_i), and ligands with an activity parameter below 300 μM were classified as active. Ligands with the activity parameter higher than 300 μM, or described by the authors as "inactive", were classified as inactive. This allowed us to ignore the methodological differences associated with various activity parameters and use a binary class to compare ligands indirectly. Afterward, to remove structurally similar compounds, both active and inactive ligands were clustered using the *k*-medoids algorithm, and for each cluster, a representative ligand was selected. Data processing and analysis were performed using the KNIME 4.0.1 analytics platform [129]. As expected, the number of inactive compounds in the dataset was low (10 compounds). To simulate the content of the real chemical library, where the percentage of active molecules is low, additional putative inactive ligands were generated using the DUD-E methodology utilizing the dedicated web server (http://dude.docking.org/generate) [130]. This well-established procedure enabled us to generate decoys with physicochemical properties similar to the active compounds but with a different topology. First, for each active ligand, a pool of decoys is generated, having a similar molecular weight, calculated *log*P, net charge, number of rotatable bonds, and hydrogen bond donors and acceptors as the active molecule. Next, a set of up to 50 most dissimilar molecules is selected from this pool (using ECFP4 fingerprint Tanimoto similarity to the active molecule). This methodology, although widely adopted, has its limitations: the possible presence of active compounds in a computationally generated set of decoys may cause an artificial underestimation of the enrichments of active compounds in clusters [131]. The final dataset consisted of 30 active and 1478 inactive compounds (for the plots showing a diversity of the generated dataset, see S13 Fig).

SIFt-based structure similarity assessment. The goal of the RNA-Puzzles round 23 was to predict the structure of the Mango-III (A10U) aptamer bound to TO1-Biotin (PDB code: 6E8U, ligand structure: S16A Fig). RMSD values for submitted ligands were calculated for

aligned RNA (PyMOL 2.5.0, function align with cycles = 0 parameter) using the LigRMSD web server with FlexibleMatch option (allowing the matching of a pair of different atom and bond types when the structure of the submitted ligand was slightly different from the reference ligand; https://ligrmsd.appsbio.utalca.cl/, [94]). As some teams submitted only the fragment of the ligand, for calculation of RMSD, we used the maximum common scaffold present in all submissions—Thiazole Orange *N*-acetamide (SMILES: C[N+]1 = CC = C(CC2 = [N+](CC(N) = O)C3 = CC = CC = C3S2)C2 = CC = CC = C12, S16B Fig). RMSD and INF (interaction network fidelity) values for the RNA models were provided by the RNA-Puzzles' organizers.

Molecular docking was performed with the rDock (version 2013.1; 400 poses were generated); molecular fingerprints were calculated for docked poses with RDKit hydrogens added (-addH rdkit option of the fingeRNAt). RMSD was calculated for the complete ligand structure.

The similarity of fingerprints was expressed as the Tanimoto coefficient (expressing the fraction of interactions common for the reference structure and a model) and the Tversky distance (with commonly used values of $\alpha = 1$ and $\beta = 0$ and thus representing the fraction of the interactions in the reference structure which are recapitulated in a model) [91].

Computational protocols

Calculations of fingerprints. In all case studies presented in this work, the FULL variant of interaction fingerprint was used with the -detail parameter, and its outputs were used for further calculations.

Molecular docking. The rDock docking program (version 2013.1) was used with dock_solv desolvation potential and a docking radius set to 10 Å [50]. For docking of small molecule ligands to HIV-1 TAR RNA, the 1UTS structure was used [117]; macromolecules were preprocessed with the Chimera dockprep pipeline [60,132]. Fingerprint type FULL was calculated for the best-scored pose of small molecule ligands docked to HIV-1 TAR RNA. Redocking data for 144 non-redundant RNA-ligand complexes were taken from the supplementary materials of [54] https://github.com/filipsPL/annapurna-additional/tree/master/docking, docking with the rDock, starting from the reference ligand structure, with 100 generated poses per complex).

Analysis of fingerprints. Fingerprint analyses and visualization were performed in the jupyter notebook with the Python 3 kernel. Principal Component Analysis (PCA) calculation and *k*-means clustering were performed using the scikit-learn Python module [133]. Optimal clustering parameters were determined by probing a series of cluster numbers and evaluating the quality of clustering with scores: Silhouette score, Calinski-Harabasz score, and Davies-Bouldin score (see S17 Table). Statistical tests were performed in SciPy using the *t*-test for the means of two independent samples of scores, and the two-tailed *p*-values were reported [134].

Clustering of chemical structures was performed in the KNIME environment, using RDKit fingerprint with *k*-medoids algorithm (10 clusters), as described in [54]. Clusters containing the same chemical classes of molecules (i.e., clusters containing amino acids, amino sugars, and other molecules) were merged into groups (named"amino acids", "amino sugars", and "other", respectively).

Software availability. The fingeRNAt program is freely available and distributed under the open-source GPL-3.0 License. It can be downloaded, along with a manual, collection of helper utilities, and sample data from https://github.com/n-szulc/fingeRNAt. The program was extensively tested on Python 3.6, 3.7, 3.8, and 3.9 under Ubuntu Linux (18.04, 20.04, and 21.10) and macOS (macOS Catalina 10.15). The supporting data presented in the manuscript along with the code used for the analysis can be found at https://github.com/n-szulc/fingeRNAt-supplementary.

Supporting information

S1 Text. Interaction statistics for seven chemical groups of ligands: alcohols and polyols, nucleosides and nucleoside derivatives, amino acids, amino sugars and aminoglycosides, aliphatic amines, heterocycles, and others.
(PDF)

S1 Fig. The number of distinct RNA residues forming a given interaction. Counts for Pb, Mn, and Sr cations, which were present only in a single complex each, were removed for clarity.
(PNG)

S2 Fig. Ligand clindamycin (PDB ID: 4V7V) with hydrogens added with the fingeRNAt internal function using RDKit library (orange spheres), OpenBabel library (magenta), and the external software: PyMOL (cyan) and Chimera (light green).
(PNG)

S3 Fig. Percentage of structures with ligands having at least one of given molecular features. HBA—hydrogen bond acceptor; HBD—hydrogen bond donor.
(PNG)

S4 Fig. Statistics of interactions formed in macromolecular complexes in the dataset by Padroni *et al*. (**A**) Total number of interactions detected for RNA-ligand complexes (analyzed with our software fingeRNAt); (**B**) the percentage of RNA-ligand complexes with at least one occurrence of a given interaction (analyzed with our software fingeRNAt); (**C**) the number of interactions for RNA-ligand complexes presented by Padroni *et al*. in their original manuscript.
(PNG)

S5 Fig. Statistics of the structures in the analyzed dataset: (A) structures count depending on the experimental method used; (B) resolution histogram for the structures determined by the X-ray diffraction (0.25 Å bin size); (C) number of structures with resolution below the given threshold.
(PNG)

S6 Fig. Statistics of interactions formed in macromolecular complexes for structures determined with a given resolution and the structures determined by NMR. (A) Total number of interactions detected for RNA-ligand complexes; (B) the percentage of RNA-ligand complexes with at least one occurrence of a given interaction.
(PNG)

S7 Fig. Histograms of bond lengths for non-covalent interactions in a dataset of experimentally solved RNA-ligand structures. The minimum and maximum cut-off values for each interaction are marked with gray dashed lines. Histogram bin width is set to 0.2 Å.
(PNG)

S8 Fig. The RNA-Puzzles round 23 solution and the best scored model. (A) The RNA-Puzzles round 23 solution and **(B)** the model submitted by the Das group (Das_7); **(C)** both complexes overlayed and **(D)** ligands.
(PNG)

S9 Fig. Tanimoto and Tversky similarity of interaction fingerprints calculated for complexes submitted to the RNA-Puzzles round 23.
(PNG)

S10 Fig. Relationship between RMSD of ligand and various measures of the interaction fingerprint similarity/distance, calculated for the data from the redocking experiment.
(PNG)

S11 Fig. Similarity vs. RMSD of ligands for structures from the RNA-Puzzles collective experiment round 23, calculated for three different resolutions of fingerprints (SIMPLE, PBS, and FULL). Ligand RMSD was calculated for the TO1 *N*-acetamide substructure common in all submitted models.
(PNG)

S12 Fig. Distribution of R^2 values (left) and Spearman rank correlation values (right) between RMSD and various metrics of SIFts similarities (y axis), calculated for a redocking experiment of 144 RNA-ligand complexes.
(PNG)

S13 Fig. The heatmaps (upper row) and the histogram (lower row) of similarity of all ligand pairs in the HIV-1 dataset expressed as Tanimoto coefficient (ranging from 0 to 1, the higher the value, the more similar are the ligands) (A) for the subset of active compounds and (B) the complete dataset.
(PNG)

S14 Fig. Interaction fingerprints calculated for data from molecular docking of a set of active and inactive compounds to HIV-1 TAR RNA structure 1UTS mapped on the two-dimensional space, with lipophilic interactions removed. Data was mapped using **(A)** TruncatedSVD, **(B)** FastICA, **(C)** KernelPCA, and **(D)** MDS.
(PNG)

S15 Fig. FULL fingerprint calculation times for docking poses of small molecule ligands of various sizes (guanidine, ibuprofen, and sildenafil) to guanidine III riboswitch (RNA with 39 residues). The fingeRNAt executed as a python script and as a singularity image. The number of poses analyzed was 10, 100, 500, 1000, and 5000 poses. The benchmark was performed on Ubuntu Linux 20.04 with Intel(R) Core(TM) i5-8400 CPU and 32 GB RAM.
(PNG)

S16 Fig. Structures of small molecule ligands used in modeling. (**A**) Structure of TO1-Biotin solved in complex with Mango-III (A10U) aptamer (6E8U), and (**B**) structure of TO1 *N*-acetamide used for RMSD calculation for submitted models.
(PNG)

S1 Table. Comparison of the features of the fingeRNAt software (this manuscript) and similar programs (Arpeggio, PLIP2021, and ProLIF).
(PDF)

S2 Table. The total number of interactions detected for RNA-ligand complexes and the number of RNA-ligand complexes with at least one occurrence of a given interaction.
(PDF)

S3 Table. Statistics of hydrogen bonds detected in macromolecular complexes using various methods—without (the default method) and with taking the position of hydrogen atoms into account (hydrogens added with OpenBabel, RDKit, or PyMOL). (A) Total number of interactions detected for RNA-ligand complexes; (B) the number and (C) the percentage of RNA-ligand complexes with at least one occurrence of a given interaction.
(PDF)

S4 Table. Statistics of complexes and detected Pi-stacking interactions in the RNA-ligand dataset for ligands with or without at least one aromatic ring.
(PDF)

S5 Table. Statistics of complexes and detected hydrogen bonds in the RNA-ligand dataset for ligands with or without at least one hydrogen bond donor and acceptor.
(PDF)

S6 Table. Statistics of complexes and detected cation-anion interactions in the RNA-ligand dataset for ligands with or without at least one charged atom.
(PDF)

S7 Table. Statistics of the lengths for observed non-covalent interactions in a dataset of experimentally solved RNA-ligand structures.
(PDF)

S8 Table. Statistics of hydrogen bonds formed by different RNA atoms.
(PDF)

S9 Table. Total number of hydrogen bonds formed by the nucleotides and the percentage of all hydrogen bonds formed by the given nucleotide using a given face of the nucleobase (H = Hoogsteen, WC = Watson-Crick, S = sugar).
(PDF)

S10 Table. Statistics of lipophilic interactions formed by different RNA atoms.
(PDF)

S11 Table. Statistics of Pi-anion interactions formed by different RNA groups (when RNA is an anion acceptor) and atoms (where RNA is an anion donor).
(PDF)

S12 Table. Statistics of halogen bonds formed by different RNA atoms.
(PDF)

S13 Table. Statistics of all types of interactions formed by different RNA atoms and groups.
(PDF)

S14 Table. Linear least-squares regression R^2 values calculated for parameters of structures from the RNA-Puzzles collective experiment and redocking experiment.
(PDF)

S15 Table. Quality of models submitted to the RNA-Puzzles competition round 23. The three best values in a given category are highlighted in green. Due to errors in the structure of the ligand, it was not possible to calculate ligand RMSD for models submitted by the Xiao group.
(PDF)

S16 Table. Statistics of relationship between RMSD and SIFts similarity—(A) Tanimoto coefficient and (B) Tversky distance, calculated for a redocking experiment of 144 RNA-ligand complexes.
(PDF)

S17 Table. Silhouette score, Calinski-Harabasz score, and Davies-Bouldin score calculated for a various number of clusters, for all-interactions dataset and the one with lipophilic

interactions removed.
(PDF)

S18 Table. Composition of the input dataset and clusters (number of active and inactive compounds, and percentage of active compounds in a given cluster), and *p*-values for comparing the ratio of active compounds in the input dataset and in a given cluster. Two-tailed *p*-values obtained in a *t*-test for the means of two independent samples of scores are reported; *p*-values ≤ 0.05 are bolded.
(PDF)

S19 Table. Mean FULL fingerprint calculation times (in seconds, the average from 10 experiments) of docking poses of small molecule ligands of various sizes (guanidine, ibuprofen, and sildenafil) to guanidine III riboswitch (RNA with 39 residues). The fingeRNAt executed as a python script or as a singularity image. The number of poses analyzed was 10, 100, 500, 1000, and 5000 poses. The benchmark was performed on Ubuntu Linux 20.04 with Intel(R) Core(TM) i5-8400 CPU and 32 GB RAM.
(PDF)

S20 Table. Definitions of the criteria of nine non-covalent interactions calculated by the fingeRNAt.py.
(PDF)

S21 Table. Parameters accepted by the fingeRNAt.py.
(PDF)

S22 Table. Summary of detected chemical properties and external modules used in the fingeRNAt.py.
(PDF)

S23 Table. List of structures used to derive statistics on RNA-small molecule ligand interactions.
(PDF)

Acknowledgments

We thank Dr. Marcin Magnus and Dr. Zhichao Miao for providing RNA-Puzzles' datasets. We thank Dr. Eugene Baulin for his constructive comments on the manuscript. This research was carried out in part with the support of the Interdisciplinary Centre for Mathematical and Computational Modelling (ICM), University of Warsaw under computational allocation no GB76-20 to F.S.

Author Contributions

Conceptualization: Natalia A. Szulc, Janusz M. Bujnicki, Filip Stefaniak.

Data curation: Natalia A. Szulc, Zuzanna Mackiewicz.

Funding acquisition: Janusz M. Bujnicki, Filip Stefaniak.

Investigation: Natalia A. Szulc, Filip Stefaniak.

Project administration: Janusz M. Bujnicki.

Resources: Janusz M. Bujnicki.

Software: Natalia A. Szulc, Filip Stefaniak.

Supervision: Janusz M. Bujnicki, Filip Stefaniak.

Visualization: Natalia A. Szulc, Filip Stefaniak.

Writing – original draft: Natalia A. Szulc, Zuzanna Mackiewicz, Janusz M. Bujnicki, Filip Stefaniak.

Writing – review & editing: Natalia A. Szulc, Zuzanna Mackiewicz, Janusz M. Bujnicki, Filip Stefaniak.

References

1. Travers A, Muskhelishvili G. DNA structure and function. FEBS J. 2015; 282(12):2279–95. https://doi.org/10.1111/febs.13307 PMID: 25903461
2. Sharp PA. The Centrality of RNA. Cell. 2009 Feb 20; 136(4):577–80. https://doi.org/10.1016/j.cell.2009.02.007 PMID: 19239877
3. Breaker RR, Joyce GF. The Expanding View of RNA and DNA Function. Chem Biol. 2014 Sep 18; 21(9):1059–65. https://doi.org/10.1016/j.chembiol.2014.07.008 PMID: 25237854
4. Warner KD, Hajdin CE, Weeks KM. Principles for targeting RNA with drug-like small molecules. Nat Rev Drug Discov. 2018 Aug; 17(8):547–58. https://doi.org/10.1038/nrd.2018.93 PMID: 29977051
5. Palazzo AF, Gregory TR. The Case for Junk DNA. PLoS Genet. 2014 May 8; 10(5):e1004351. https://doi.org/10.1371/journal.pgen.1004351 PMID: 24809441
6. Hudson WH, Ortlund EA. The structure, function and evolution of proteins that bind DNA and RNA. Nat Rev Mol Cell Biol. 2014 Nov; 15(11):749–60. https://doi.org/10.1038/nrm3884 PMID: 25269475
7. Pyle A. Metal ions in the structure and function of RNA. JBIC J Biol Inorg Chem. 2002 Sep 1; 7(7):679–90. https://doi.org/10.1007/s00775-002-0387-6 PMID: 12203005
8. Draper DE. A guide to ions and RNA structure. RNA. 2004 Jan 3; 10(3):335–43. https://doi.org/10.1261/rna.5205404 PMID: 14970378
9. Lipfert J, Doniach S, Das R, Herschlag D. Understanding Nucleic Acid–Ion Interactions. Annu Rev Biochem. 2014; 83:813–41. https://doi.org/10.1146/annurev-biochem-060409-092720 PMID: 24606136
10. Morris DL. DNA-bound metal ions: recent developments. Biomol Concepts. 2014 Oct; 5(5):397–407. https://doi.org/10.1515/bmc-2014-0021 PMID: 25367620
11. Sudarsan N, Barrick JE, Breaker RR. Metabolite-binding RNA domains are present in the genes of eukaryotes. RNA. 2003 Jan 6; 9(6):644–7. https://doi.org/10.1261/rna.5090103 PMID: 12756322
12. Sheng J, Gan JH, Huang Z. Structure-Based DNA-Targeting Strategies with Small Molecule Ligands for Drug Discovery. Med Res Rev. 2013 Sep; 33(5):1119–73. https://doi.org/10.1002/med.21278 PMID: 23633219
13. Godzieba M, Ciesielski S. Natural DNA Intercalators as Promising Therapeutics for Cancer and Infectious Diseases. Curr Cancer Drug Targets. 2020 Jan 1; 20(1):19–32. https://doi.org/10.2174/1568009619666191007112516 PMID: 31589125
14. Poehlsgaard J, Douthwaite S. The bacterial ribosome as a target for antibiotics. Nat Rev Microbiol. 2005 Nov; 3(11):870–81. https://doi.org/10.1038/nrmicro1265 PMID: 16261170
15. Maguire BA. Inhibition of Bacterial Ribosome Assembly: a Suitable Drug Target? Microbiol Mol Biol Rev. 2009 Mar; 73(1):22–35. https://doi.org/10.1128/MMBR.00030-08 PMID: 19258531
16. Mullard A. FDA approves RNA-targeting small molecule. Nat Rev Drug Discov. 2020 Sep 3; 19(10):659–659. https://doi.org/10.1038/d41573-020-00158-1 PMID: 32884109
17. Gandhi G, Abdullah S, Foead AI, Yeo WWY. The potential role of miRNA therapies in spinal muscle atrophy. J Neurol Sci. 2021 Aug 15; 427:117485. https://doi.org/10.1016/j.jns.2021.117485 PMID: 34015517
18. Chen JL, Zhang P, Abe M, Aikawa H, Zhang L, Frank AJ, et al. Design, Optimization, and Study of Small Molecules That Target Tau Pre-mRNA and Affect Splicing. J Am Chem Soc. 2020 May 13; 142(19):8706–27. https://doi.org/10.1021/jacs.0c00768 PMID: 32364710
19. Matsui M, Corey DR. Non-coding RNAs as drug targets. Nat Rev Drug Discov. 2017 Mar; 16(3):167–79. https://doi.org/10.1038/nrd.2016.117 PMID: 27444227
20. Deigan KE, Ferre-D'Amare AR. Riboswitches: Discovery of Drugs That Target Bacterial Gene-Regulatory RNAs. Acc Chem Res. 2011 Dec; 44(12):1329–38. https://doi.org/10.1021/ar200039b PMID: 21615107

21. Hermann T. Small molecules targeting viral RNA. Wiley Interdiscip Rev-Rna. 2016 Nov; 7(6):726–43. https://doi.org/10.1002/wrna.1373 PMID: 27307213

22. Yu AM, Choi YH, Tu MJ. RNA Drugs and RNA Targets for Small Molecules: Principles, Progress, and Challenges. Pharmacol Rev. 2020 Oct 1; 72(4):862–98. https://doi.org/10.1124/pr.120.019554 PMID: 32929000

23. Winkle M, El-Daly SM, Fabbri M, Calin GA. Noncoding RNA therapeutics—challenges and potential solutions. Nat Rev Drug Discov. 2021 Jun 18;1–23.

24. Zhou P, Huang J, Tian F. Specific Noncovalent Interactions at Protein-Ligand Interface: Implications for Rational Drug Design. Curr Med Chem. 2012 Jan 1; 19(2):226–38. https://doi.org/10.2174/092986712803414150 PMID: 22320300

25. Persch E, Dumele O, Diederich F. Molecular Recognition in Chemical and Biological Systems. Angew Chem Int Ed. 2015; 54(11):3290–327. https://doi.org/10.1002/anie.201408487 PMID: 25630692

26. Fischer A, Smieško M, Sellner M, Lill MA. Decision Making in Structure-Based Drug Discovery: Visual Inspection of Docking Results. J Med Chem. 2021 Mar 11; 64(5):2489–500. https://doi.org/10.1021/acs.jmedchem.0c02227 PMID: 33617246

27. Durrant JD, McCammon JA. BINANA: A Novel Algorithm for Ligand-Binding Characterization. J Mol Graph Model. 2011 Apr; 29(6):888–93. https://doi.org/10.1016/j.jmgm.2011.01.004 PMID: 21310640

28. Weisel M, Bitter HM, Diederich F, So WV, Kondru R. PROLIX: rapid mining of protein-ligand interactions in large crystal structure databases. J Chem Inf Model. 2012 Jun 25; 52(6):1450–61. https://doi.org/10.1021/ci300034x PMID: 22582806

29. Radifar M, Yuniarti N, Istyastono EP. PyPLIF: Python-based Protein-Ligand Interaction Fingerprinting. Bioinformation. 2013; 9(6):325–8. https://doi.org/10.6026/97320630009325 PMID: 23559752

30. Salentin S, Schreiber S, Haupt VJ, Adasme MF, Schroeder M. PLIP: fully automated protein–ligand interaction profiler. Nucleic Acids Res. 2015 Jul 1; 43(W1):W443–7. https://doi.org/10.1093/nar/gkv315 PMID: 25873628

31. Da C, Kireev D. Structural Protein–Ligand Interaction Fingerprints (SPLIF) for Structure-Based Virtual Screening: Method and Benchmark Study. J Chem Inf Model. 2014 Sep 22; 54(9):2555–61. https://doi.org/10.1021/ci500319f PMID: 25116840

32. Wallace AC, Laskowski RA, Thornton JM. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. Protein Eng Des Sel. 1995 Feb 1; 8(2):127–34. https://doi.org/10.1093/protein/8.2.127 PMID: 7630882

33. Laskowski RA, Swindells MB. LigPlot+: Multiple Ligand–Protein Interaction Diagrams for Drug Discovery. J Chem Inf Model. 2011 Oct 24; 51(10):2778–86. https://doi.org/10.1021/ci200227u PMID: 21919503

34. Jubb HC, Higueruelo AP, Ochoa-Montaño B, Pitt WR, Ascher DB, Blundell TL. Arpeggio: A Web Server for Calculating and Visualising Interatomic Interactions in Protein Structures. J Mol Biol. 2017 Feb 3; 429(3):365–71. https://doi.org/10.1016/j.jmb.2016.12.004 PMID: 27964945

35. Adasme MF, Linnemann KL, Bolz SN, Kaiser F, Salentin S, Haupt VJ, et al. PLIP 2021: expanding the scope of the protein–ligand interaction profiler to DNA and RNA. Nucleic Acids Res. 2021 May 5; 49(W1):W530–4. https://doi.org/10.1093/nar/gkab294 PMID: 33950214

36. Bouysset C, Fiorucci S. ProLIF: a library to encode molecular interactions as fingerprints. J Cheminformatics. 2021 Sep 25; 13(1):72. https://doi.org/10.1186/s13321-021-00548-6 PMID: 34563256

37. Deng Z, Chuaqui C, Singh J. Structural interaction fingerprint (SIFt): a novel method for analyzing three-dimensional protein-ligand binding interactions. J Med Chem. 2004 Jan 15; 47(2):337–44. https://doi.org/10.1021/jm030331x PMID: 14711306

38. Rácz A, Bajusz D, Héberger K. Life beyond the Tanimoto coefficient: similarity measures for interaction fingerprints. J Cheminformatics. 2018 Oct 4; 10(1):48. https://doi.org/10.1186/s13321-018-0302-y PMID: 30288626

39. Pérez-Nueno VI, Rabal O, Borrell JI, Teixidó J. APIF: A New Interaction Fingerprint Based on Atom Pairs and Its Application to Virtual Screening. J Chem Inf Model. 2009 May 22; 49(5):1245–60. https://doi.org/10.1021/ci900043r PMID: 19364101

40. Istyastono EP, Radifar M, Yuniarti N, Prasasty VD, Mungkasi S. PyPLIF HIPPOS: A Molecular Interaction Fingerprinting Tool for Docking Results of AutoDock Vina and PLANTS. J Chem Inf Model. 2020 Jul 20; https://doi.org/10.1021/acs.jcim.0c00305 PMID: 32687350

41. Jasper JB, Humbeck L, Brinkjost T, Koch O. A novel interaction fingerprint derived from per atom score contributions: exhaustive evaluation of interaction fingerprint performance in docking based virtual screening. J Cheminformatics. 2018 Mar 16; 10(1):15. https://doi.org/10.1186/s13321-018-0264-0 PMID: 29549526

42. Xiong G, Shen C, Yang Z, Jiang D, Liu S, Lu A, et al. Featurization strategies for protein–ligand interactions and their applications in scoring function development. WIREs Comput Mol Sci. n/a(n/a): e1567.
43. Witek J, Rataj K, Mordalski S, Smusz S, Kosciolek T, Bojarski AJ. Application of Structural Interaction Fingerpints (SIFts) into post-docking analysis—insight into activity and selectivity. J Cheminformatics. 2013 Mar 22; 5(1):P28.
44. Wójcikowski M, Kukiełka M, Stepniewska-Dziubinska MM, Siedlecki P. Development of a protein-ligand extended connectivity (PLEC) fingerprint and its application for binding affinity predictions. Bioinforma Oxf Engl. 2019 Apr 15; 35(8):1334–41. https://doi.org/10.1093/bioinformatics/bty757 PMID: 30202917
45. Meng Z, Xia K. Persistent spectral-based machine learning (PerSpect ML) for protein-ligand binding affinity prediction. Sci Adv. 2021 May; 7(19):eabc5329. https://doi.org/10.1126/sciadv.abc5329 PMID: 33962954
46. Kumar S, Kim MH. SMPLIP-Score: predicting ligand binding affinity from simple and interpretable on-the-fly interaction fingerprint pattern descriptors. J Cheminformatics. 2021 Mar 25; 13(1):28. https://doi.org/10.1186/s13321-021-00507-1 PMID: 33766140
47. Gainza P, Sverrisson F, Monti F, Rodolà E, Boscaini D, Bronstein MM, et al. Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning. Nat Methods. 2020 Feb; 17(2):184–92. https://doi.org/10.1038/s41592-019-0666-6 PMID: 31819266
48. Coimbatore Narayanan B, Westbrook J, Ghosh S, Petrov AI, Sweeney B, Zirbel CL, et al. The Nucleic Acid Database: new features and capabilities. Nucleic Acids Res. 2014 Jan 1; 42(D1):D114–22. https://doi.org/10.1093/nar/gkt980 PMID: 24185695
49. Morley SD, David Morley S, Afshar M. Validation of an empirical RNA-ligand scoring function for fast flexible docking using RiboDock. J Comput Aided Mol Des. 2004; 18(3):189–208. https://doi.org/10.1023/b:jcam.0000035199.48747.1e PMID: 15368919
50. Ruiz-Carmona S, Alvarez-Garcia D, Foloppe N, Garmendia-Doval AB, Juhos S, Schmidtke P, et al. rDock: a fast, versatile and open source program for docking ligands to proteins and nucleic acids. PLoS Comput Biol. 2014 Apr; 10(4):e1003571. https://doi.org/10.1371/journal.pcbi.1003571 PMID: 24722481
51. Pfeffer P, Gohlke H. DrugScoreRNA—knowledge-based scoring function to predict RNA-ligand interactions. J Chem Inf Model. 2007 Sep; 47(5):1868–76. https://doi.org/10.1021/ci700134p PMID: 17705464
52. Chhabra S, Xie J, Frank AT. RNAPosers: Machine Learning Classifiers for Ribonucleic Acid–Ligand Poses. J Phys Chem B. 2020; 124(22):4436–45. https://doi.org/10.1021/acs.jpcb.0c02322 PMID: 32427491
53. Philips A, Milanowska K, Lach G, Bujnicki JM. LigandRNA: computational predictor of RNA-ligand interactions. RNA. 2013 Dec; 19(12):1605–16. https://doi.org/10.1261/rna.039834.113 PMID: 24145824
54. Stefaniak F, Bujnicki JM. AnnapuRNA: A scoring function for predicting RNA-small molecule binding poses. PLOS Comput Biol. 2021 Feb 1; 17(2):e1008309. https://doi.org/10.1371/journal.pcbi.1008309 PMID: 33524009
55. Lu Y, Wang R, Yang CY, Wang S. Analysis of Ligand-Bound Water Molecules in High-Resolution Crystal Structures of Protein−Ligand Complexes. J Chem Inf Model. 2007 Mar 1; 47(2):668–75. https://doi.org/10.1021/ci6003527 PMID: 17266298
56. Torshin IY, Weber IT, Harrison RW. Geometric criteria of hydrogen bonds in proteins and identification of 'bifurcated' hydrogen bonds. Protein Eng Des Sel. 2002 May 1; 15(5):359–63. https://doi.org/10.1093/protein/15.5.359 PMID: 12034855
57. O'Boyle NM, Morley C, Hutchison GR. Pybel: a Python wrapper for the OpenBabel cheminformatics toolkit. Chem Cent J. 2008 Mar 9; 2:5. https://doi.org/10.1186/1752-153X-2-5 PMID: 18328109
58. RDKit, Open-Source Cheminformatics. http://www.rdkit.org.
59. The PyMOL Molecular Graphics System, Version 2.6 Schrödinger, LLC.
60. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem. 2004 Oct; 25 (13):1605–12. https://doi.org/10.1002/jcc.20084 PMID: 15264254
61. Ferreira de Freitas R, Schapira M. A systematic analysis of atomic protein-ligand interactions in the PDB. MedChemComm. 2017 Oct 1; 8(10):1970–81. https://doi.org/10.1039/c7md00381a PMID: 29308120

62. Padroni G, N. Patwardhan N, Schapira M, E. Hargrove A. Systematic analysis of the interactions driving small molecule–RNA recognition. RSC Med Chem. 2020; 11(7):802–13. https://doi.org/10.1039/d0md00167h PMID: 33479676

63. Tang CL, Alexov E, Pyle AM, Honig B. Calculation of pKas in RNA: On the Structural Origins and Functional Roles of Protonated Nucleotides. J Mol Biol. 2007 Mar 9; 366(5):1475–96. https://doi.org/10.1016/j.jmb.2006.12.001 PMID: 17223134

64. Wolter AC, Weickhmann AK, Nasiri AH, Hantke K, Ohlenschläger O, Wunderlich CH, et al. A Stably Protonated Adenine Nucleotide with a Highly Shifted pKa Value Stabilizes the Tertiary Structure of a GTP-Binding RNA Aptamer. Angew Chem Int Ed. 2017; 56(1):401–4. https://doi.org/10.1002/anie.201609184 PMID: 27885761

65. Kaul M, Barbieri CM, Kerrigan JE, Pilch DS. Coupling of Drug Protonation to the Specific Binding of Aminoglycosides to the A Site of 16S rRNA: Elucidation of the Number of Drug Amino Groups Involved and their Identities. J Mol Biol. 2003 Mar 7; 326(5):1373–87. https://doi.org/10.1016/s0022-2836(02)01452-3 PMID: 12595251

66. Barbieri CM, Pilch DS. Complete Thermodynamic Characterization of the Multiple Protonation Equilibria of the Aminoglycoside Antibiotic Paromomycin: A Calorimetric and Natural Abundance 15N NMR Study. Biophys J. 2006 Feb 15; 90(4):1338–49. https://doi.org/10.1529/biophysj.105.075028 PMID: 16326918

67. Bas DC, Rogers DM, Jensen JH. Very fast prediction and rationalization of pKa values for protein–ligand complexes. Proteins Struct Funct Bioinforma. 2008; 73(3):765–83. https://doi.org/10.1002/prot.22102 PMID: 18498103

68. Bietz S, Urbaczek S, Schulz B, Rarey M. Protoss: a holistic approach to predict tautomers and protonation states in protein-ligand complexes. J Cheminformatics. 2014 Apr 3; 6(1):12. https://doi.org/10.1186/1758-2946-6-12 PMID: 24694216

69. Xu Z, Zhang Q, Shi J, Zhu W. Underestimated Noncovalent Interactions in Protein Data Bank. J Chem Inf Model. 2019 Aug 26; 59(8):3389–99. https://doi.org/10.1021/acs.jcim.9b00258 PMID: 31294978

70. Lakshmi B, G. Samuelson A, Jose KVJ, R. Gadre S, Arunan E. Is there a hydrogen bond radius? Evidence from microwave spectroscopy, neutron scattering and X-ray diffraction results. New J Chem. 2005; 29(2):371–7.

71. Kaźmierczak M, Katrusiak A. Bimodal Distribution of the Shortest Intermolecular Contacts in Crystals of Organic Compounds. Cryst Growth Des. 2014 May 7; 14(5):2223–9.

72. Panigrahi SK, Desiraju GR. Strong and weak hydrogen bonds in the protein–ligand interface. Proteins Struct Funct Bioinforma. 2007 Apr 1; 67(1):128–41. https://doi.org/10.1002/prot.21253 PMID: 17206656

73. Gagné OC, Hawthorne FC. Bond-length distributions for ions bonded to oxygen: alkali and alkaline-earth metals. Acta Crystallogr Sect B Struct Sci Cryst Eng Mater. 2016 Aug 1; 72(Pt 4):602–25. https://doi.org/10.1107/S2052520616008507 PMID: 27484381

74. Piovesan D, Minervini G, Tosatto SCE. The RING 2.0 web server for high quality residue interaction networks. Nucleic Acids Res. 2016 Jul 8; 44(Web Server issue):W367–74. https://doi.org/10.1093/nar/gkw315 PMID: 27198219

75. Borozan SZ, Zlatović MV, Stojanović S. Anion–π interactions in complexes of proteins and halogen-containing amino acids. JBIC J Biol Inorg Chem. 2016 Jun 1; 21(3):357–68. https://doi.org/10.1007/s00775-016-1346-y PMID: 26910415

76. Bissantz C, Kuhn B, Stahl M. A Medicinal Chemist's Guide to Molecular Interactions. J Med Chem. 2010 Jul 22; 53(14):5061–84. https://doi.org/10.1021/jm100112j PMID: 20345171

77. Kligun E, Mandel-Gutfreund Y. Conformational readout of RNA by small ligands. RNA Biol. 2013 Jun 1; 10(6):981–9. https://doi.org/10.4161/rna.24682 PMID: 23618839

78. Kondo J, Westhof E. Base pairs and pseudo pairs observed in RNA–ligand complexes. J Mol Recognit. 2010; 23(2):241–52. https://doi.org/10.1002/jmr.978 PMID: 19701919

79. Kondo J, Westhof E. Classification of pseudo pairs between nucleotide bases and amino acids by analysis of nucleotide–protein complexes. Nucleic Acids Res. 2011 Oct 1; 39(19):8628–37. https://doi.org/10.1093/nar/gkr452 PMID: 21737431

80. Wagner JR, Churas CP, Liu S, Swift RV, Chiu M, Shao C, et al. Continuous Evaluation of Ligand Protein Predictions: A Weekly Community Challenge for Drug Docking. Structure. 2019 Aug 6; 27(8):1326–1335.e4. https://doi.org/10.1016/j.str.2019.05.012 PMID: 31257108

81. Baber JC, Thompson DC, Cross JB, Humblet C. GARD: A Generally Applicable Replacement for RMSD. J Chem Inf Model. 2009 Aug 24; 49(8):1889–900. https://doi.org/10.1021/ci9001074 PMID: 19618919

82. Yusuf D, Davis AM, Kleywegt GJ, Schmitt S. An Alternative Method for the Evaluation of Docking Performance: RSR vs RMSD. J Chem Inf Model. 2008 Jul 1; 48(7):1411–22. https://doi.org/10.1021/ci800084x PMID: 18598022

83. Schulz-Gasch T, Schärfer C, Guba W, Rarey M. TFD: Torsion Fingerprints As a New Measure To Compare Small Molecule Conformations. J Chem Inf Model. 2012 Jun 25; 52(6):1499–512. https://doi.org/10.1021/ci2002318 PMID: 22670896

84. Leung S, Bodkin M, von Delft F, Brennan P, Morris G. SuCOS is Better than RMSD for Evaluating Fragment Elaboration and Docking Poses. 2019 May 10;

85. Ding Y, Fang Y, Moreno J, Ramanujam J, Jarrell M, Brylinski M. Assessing the similarity of ligand binding conformations with the Contact Mode Score. Comput Biol Chem. 2016 Oct 1; 64:403–13. https://doi.org/10.1016/j.compbiolchem.2016.08.007 PMID: 27620381

86. Kroemer RT, Vulpetti A, McDonald JJ, Rohrer DC, Trosset JY, Giordanetto F, et al. Assessment of Docking Poses: Interactions-Based Accuracy Classification (IBAC) versus Crystal Structure Deviations. J Chem Inf Comput Sci. 2004 May 1; 44(3):871–81. https://doi.org/10.1021/ci049970m PMID: 15154752

87. Balius TE, Mukherjee S, Rizzo RC. Implementation and Evaluation of a Docking-Rescoring Method using Molecular Footprint Comparisons. J Comput Chem. 2011 Jul 30; 32(10):2273–89. https://doi.org/10.1002/jcc.21814 PMID: 21541962

88. Drwal MN, Jacquemard C, Perez C, Desaphy J, Kellenberger E. Do Fragments and Crystallization Additives Bind Similarly to Drug-like Ligands? J Chem Inf Model. 2017 May 22; 57(5):1197–209. https://doi.org/10.1021/acs.jcim.6b00769 PMID: 28414463

89. Miao Z, Adamiak RW, Antczak M, Boniecki MJ, Bujnicki J, Chen SJ, et al. RNA-Puzzles Round IV: 3D structure predictions of four ribozymes and two aptamers. RNA. 2020 Jan 8; 26(8):982–95. https://doi.org/10.1261/rna.075341.120 PMID: 32371455

90. Parisien M, Cruz JA, Westhof É, Major F. New metrics for comparing and assessing discrepancies between RNA 3D structures and models. RNA. 2009 Oct; 15(10):1875–85. https://doi.org/10.1261/rna.1700409 PMID: 19710185

91. Leach AR, Gillet VJ. An Introduction to Chemoinformatics. Springer; 2007. 260 p.

92. Marcou G, Rognan D. Optimizing Fragment and Scaffold Docking by Use of Molecular Interaction Fingerprints. J Chem Inf Model. 2007 Jan 1; 47(1):195–207. https://doi.org/10.1021/ci600342e PMID: 17238265

93. Desaphy J, Raimbaud E, Ducrot P, Rognan D. Encoding Protein–Ligand Interaction Patterns in Fingerprints and Graphs. J Chem Inf Model. 2013 Mar 25; 53(3):623–37. https://doi.org/10.1021/ci300566n PMID: 23432543

94. Velázquez-Libera JL, Durán-Verdugo F, Valdés-Jiménez A, Núñez-Vivanco G, Caballero J. LigRMSD: a web server for automatic structure matching and RMSD calculations among identical and similar compounds in protein-ligand docking. Bioinformatics. 2020 May 1; 36(9):2912–4. https://doi.org/10.1093/bioinformatics/btaa018 PMID: 31926012

95. Filikov AV, Mohan V, Vickers TA, Griffey RH, Cook PD, Abagyan RA, et al. Identification of ligands for RNA targets via structure-based virtual screening: HIV-1 TAR. J Comput Aided Mol Des. 2000 Aug; 14(6):593–610. https://doi.org/10.1023/a:1008121029716 PMID: 10921774

96. Ganser LR, Lee J, Rangadurai A, Merriman DK, Kelly ML, Kansal AD, et al. High-performance virtual screening by targeting a high-resolution RNA dynamic ensemble. Nat Struct Mol Biol. 2018 May; 25(5):425–34. https://doi.org/10.1038/s41594-018-0062-4 PMID: 29728655

97. Abulwerdi FA, Le Grice SFJ. Recent Advances in Targeting the HIV-1 Tat/TAR Complex. Curr Pharm Des. 2017 Aug 1; 23(28):4112–21. https://doi.org/10.2174/1381612823666170616081736 PMID: 28625133

98. Imai YN, Inoue Y, Nakanishi I, Kitaura K. Cl–π interactions in protein–ligand complexes. Protein Sci Publ Protein Soc. 2008 Jul; 17(7):1129–37. https://doi.org/10.1110/ps.033910.107 PMID: 18434503

99. D'Oria E, Novoa JJ. Cation-anion hydrogen bonds: a new class of hydrogen bonds that extends their strength beyond the covalent limit. A theoretical characterization. J Phys Chem A. 2011 Nov 17; 115(45):13114–23. https://doi.org/10.1021/jp205176e PMID: 21942671

100. Auffinger P, Bielecki L, Westhof E. Anion Binding to Nucleic Acids. Structure. 2004 Mar 1; 12(3):379–88. https://doi.org/10.1016/j.str.2004.02.015 PMID: 15016354

101. Kuhn B, Gilberg E, Taylor R, Cole J, Korb O. How Significant Are Unusual Protein–Ligand Interactions? Insights from Database Mining. J Med Chem. 2019 Nov 27; 62(22):10441–55. https://doi.org/10.1021/acs.jmedchem.9b01545 PMID: 31730345

102. Stefan LR, Zhang R, Levitan AG, Hendrix DK, Brenner SE, Holbrook SR. MeRNA: a database of metal ion binding sites in RNA structures. Nucleic Acids Res. 2006 Jan 1; 34(suppl_1):D131–4. https://doi.org/10.1093/nar/gkj058 PMID: 16381830

103. Müller J. Functional metal ions in nucleic acids. Metallomics. 2010 May 1; 2(5):318–27. https://doi.org/10.1039/c000429d PMID: 21072378

104. Jeffrey GA. An Introduction to Hydrogen Bonding. Oxford University Press; 1997. 366 p.

105. Auffinger P, Hays FA, Westhof E, Ho PS. Halogen bonds in biological molecules. Proc Natl Acad Sci. 2004 Nov 30; 101(48):16789–94. https://doi.org/10.1073/pnas.0407607101 PMID: 15557000

106. Barlow DJ, Thornton JM. Ion-pairs in proteins. J Mol Biol. 1983 Aug 25; 168(4):867–85. https://doi.org/10.1016/s0022-2836(83)80079-5 PMID: 6887253

107. Gallivan JP, Dougherty DA. Cation-pi interactions in structural biology. Proc Natl Acad Sci U S A. 1999 Aug 17; 96(17):9459–64. https://doi.org/10.1073/pnas.96.17.9459 PMID: 10449714

108. McGaughey GB, Gagné M, Rappé AK. π-stacking interactions: alive and well in proteins. J Biol Chem. 1998 Jun 19; 273(25):15458–63. https://doi.org/10.1074/jbc.273.25.15458 PMID: 9624131

109. Zheng H, Chruszcz M, Lasota P, Lebioda L, Minor W. Data mining of metal ion environments present in protein structures. J Inorg Biochem. 2008 Sep 1; 102(9):1765–76. https://doi.org/10.1016/j.jinorgbio.2008.05.006 PMID: 18614239

110. Poornima CS, Dean PM. Hydration in drug design. 1. Multiple hydrogen-bonding features of water molecules in mediating protein-ligand interactions. J Comput Aided Mol Des. 1995 Dec; 9(6):500–12. https://doi.org/10.1007/BF00124321 PMID: 8789192

111. Newberry RW, Raines RT. The n→π* Interaction. Acc Chem Res. 2017 Aug 15; 50(8):1838–46. https://doi.org/10.1021/acs.accounts.7b00121 PMID: 28735540

112. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: An open chemical toolbox. J Cheminform. 2011 Oct 7; 3:33. https://doi.org/10.1186/1758-2946-3-33 PMID: 21982300

113. Bajusz D, Rácz A, Héberger K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? J Cheminformatics. 2015 May 20; 7(1):20. https://doi.org/10.1186/s13321-015-0069-3 PMID: 26052348

114. https://chemicalize.com/ developed by ChemAxon.

115. Du Z, Lind KE, James TL. Structure of TAR RNA complexed with a Tat-TAR interaction nanomolar inhibitor that was identified by computational screening. Chem Biol. 2002 Jun; 9(6):707–12. https://doi.org/10.1016/s1074-5521(02)00151-5 PMID: 12079782

116. Davis B, Afshar M, Varani G, Murchie AIH, Karn J, Lentzen G, et al. Rational design of inhibitors of HIV-1 TAR RNA through the stabilisation of electrostatic "hot spots." J Mol Biol. 2004 Feb 13; 336(2):343–56. https://doi.org/10.1016/j.jmb.2003.12.046 PMID: 14757049

117. Murchie AIH, Davis B, Isel C, Afshar M, Drysdale MJ, Bower J, et al. Structure-based Drug Design Targeting an Inactive RNA Conformation: Exploiting the Flexibility of HIV-1 TAR RNA. J Mol Biol. 2004 Feb 20; 336(3):625–38. https://doi.org/10.1016/j.jmb.2003.12.028 PMID: 15095977

118. Kumar S, Arya DP. Recognition of HIV TAR RNA by triazole linked neomycin dimers. Bioorg Med Chem Lett. 2011 Aug 15; 21(16):4788–92. https://doi.org/10.1016/j.bmcl.2011.06.058 PMID: 21757341

119. Stelzer AC, Frank AT, Kratz JD, Swanson MD, Gonzalez-Hernandez MJ, Lee J, et al. Discovery of Selective Bioactive Small Molecules by Targeting an RNA Dynamic Ensemble. Nat Chem Biol. 2011 Jun 26; 7(8):553–9. https://doi.org/10.1038/nchembio.596 PMID: 21706033

120. Davidson A, Begley DW, Lau C, Varani G. A small-molecule probe induces a conformation in HIV TAR RNA capable of binding drug-like fragments. J Mol Biol. 2011 Jul 29; 410(5):984–96. https://doi.org/10.1016/j.jmb.2011.03.039 PMID: 21763501

121. Ranjan N, Kumar S, Watkins D, Wang D, Appella DH, Arya DP. Recognition of HIV-TAR RNA using Neomycin-Benzimidazole Conjugates. Bioorg Med Chem Lett. 2013 Oct 15; 23(20):5689–93. https://doi.org/10.1016/j.bmcl.2013.08.014 PMID: 24012122

122. Sztuba-Solinska J, Shenoy SR, Gareiss P, Krumpe LRH, Le Grice SFJ, O'Keefe BR, et al. Identification of Biologically Active, HIV TAR RNA-Binding Small Molecules Using Small Molecule Microarrays. J Am Chem Soc. 2014 Jun 11; 136(23):8402–10. https://doi.org/10.1021/ja502754f PMID: 24820959

123. Joly JP, Mata G, Eldin P, Briant L, Fontaine-Vive F, Duca M, et al. Artificial nucleobase-amino acid conjugates: a new class of TAR RNA binding agents. Chem Weinh Bergstr Ger. 2014 Feb 10; 20(7):2071–9. https://doi.org/10.1002/chem.201303664 PMID: 24431237

124. Zeiger M, Stark S, Kalden E, Ackermann B, Ferner J, Scheffer U, et al. Fragment based search for small molecule inhibitors of HIV-1 Tat-TAR. Bioorg Med Chem Lett. 2014 Dec 15; 24(24):5576–80. https://doi.org/10.1016/j.bmcl.2014.11.004 PMID: 25466178

125. Kumar S, Ranjan N, Kellish P, Gong C, Watkins D, Arya DP. Multivalency in Recognition and Antagonism of HIV TAR RNA–TAT Assembly using an Aminoglycoside Benzimidazole Scaffold. Org Biomol Chem. 2016 Feb 2; 14(6):2052–6. https://doi.org/10.1039/c5ob02016f PMID: 26765486

126. Patwardhan NN, Ganser LR, Kapral GJ, Eubanks CS, Lee J, Sathyamoorthy B, et al. Amiloride as a new RNA-binding scaffold with activity against HIV-1 TAR. MedChemComm. 2017 Mar 15; 8 (5):1022–36. https://doi.org/10.1039/C6MD00729E PMID: 28798862

127. Desantis J, Massari S, Sosic A, Manfroni G, Cannalire R, Felicetti T, et al. Design and Synthesis of WM5 Analogues as HIV-1 TAR RNA Binders. Open Med Chem J. 2019 Feb 28; 13(1):16–28.

128. Baell JB, Holloway GA. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. J Med Chem. 2010; https://doi.org/10.1021/jm901137j PMID: 20131845

129. Berthold MR, Cebron N, Dill F, Gabriel TR, Kötter T, Meinl T, et al. KNIME: The Konstanz Information Miner. Data Anal Mach Learn Appl. 2008;319–26.

130. Mysinger MM, Carchia M, Irwin JJ, Shoichet BK. Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. J Med Chem. 2012; 55(14):6582–94. https://doi.org/10.1021/jm300687e PMID: 22716043

131. Cleves AE, Jain AN. Structure- and Ligand-Based Virtual Screening on DUD-E+: Performance Dependence on Approximations to the Binding Pocket. J Chem Inf Model. 2020 Sep 28; 60(9):4296–310. https://doi.org/10.1021/acs.jcim.0c00115 PMID: 32271577

132. Lang PT, Brozell SR, Mukherjee S, Pettersen EF, Meng EC, Thomas V, et al. DOCK 6: combining techniques to model RNA-small molecule complexes. RNA. 2009 Jun; 15(6):1219–30. https://doi.org/10.1261/rna.1563609 PMID: 19369428

133. Pedregosa F and V G and Gramfort A and Michel V and Thirion Band Grisel Oand Blondel Mand Prettenhofer Pand Weiss R and Dubourg V and Vanderplas J and Passos A and Cournapeau D and Brucher M and Perrot M and Duchesnay E. Scikit-learn: Machine Learning in Python. J Mach Learn Res. 2011; 12:2825–30.

134. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nat Methods. 2020; 17:261–72. https://doi.org/10.1038/s41592-019-0686-2 PMID: 32015543