Article

RPpocket: An RNA–Protein Intuitive Database with RNA Pocket Topology Resources

Rui Yang [†]**, Haoquan Liu** [†]**, Liu Yang, Ting Zhou, Xinyao Li and Yunjie Zhao** *

Department of Physics, Institute of Biophysics, Central China Normal University, Wuhan 430079, China;
ryang@mails.ccnu.edu.cn (R.Y.); liuhaoquan@mails.ccnu.edu.cn (H.L.); YL2019212120@mails.ccnu.edu.cn (L.Y.);
tzhou@mails.ccnu.edu.cn (T.Z.); lixinyao@mails.ccnu.edu.cn (X.L.)
* Correspondence: yjzhaowh@mail.ccnu.edu.cn
† These authors contributed equally to this work.

Abstract: RNA–protein complexes regulate a variety of biological functions. Thus, it is essential to explore and visualize RNA–protein structural interaction features, especially pocket interactions. In this work, we develop an easy-to-use bioinformatics resource: RPpocket. This database provides RNA–protein complex interactions based on sequence, secondary structure, and pocket topology analysis. We extracted 793 pockets from 74 non-redundant RNA–protein structures. Then, we calculated the binding- and non-binding pocket topological properties and analyzed the binding mechanism of the RNA–protein complex. The results showed that the binding pockets were more extended than the non-binding pockets. We also found that long-range forces were the main interaction for RNA–protein recognition, while short-range forces strengthened and optimized the binding. RPpocket could facilitate RNA–protein engineering for biological or medical applications.

Keywords: RNA–protein interaction; pocket database; interaction mechanism

1. Introduction

RNA–protein complexes play irreplaceable roles in biological processes, including gene regulation, protein synthesis, and virus replication [1–3]. For example, the ribosome is a large RNA–protein complex for protein synthesis [4]. RBPs bind the target RNA to form ribonucleoprotein complexes and perform biological functions [5–7]. For example, the SARS-CoV-2 S protein stabilizes the virus RNA and enhances virus translation by hijacking the host factor IGF2BP1 [8]. Thus, the RNA–protein structural pocket information helps understand biological processes and drug design [9–11].

Currently, the available RNA–protein databases can be divided into three categories. (1) The structural RNA–protein database. The Nucleic Acid Database (NDB) and Protein Data Bank (PDB) provide experimentally determined RNA–protein complexes with sequences and structures [12,13]. In addition, the 3D shapes of the complex assemblies help researchers to understand structure and function principles. (2) The comprehensive RNA–protein database. The RISE, RNAInter, DBBP, and NPInter integrate multiple resources and provide the sequences, secondary structures, and annotations on the interaction interface [14–17]. (3) The RNA–protein docking database. The DM-RPIs, RPI-Pred, NPDOCK, HDOCK, and RNAct provide RNA–protein structures with weakly homologous complexes [18–22]. However, none of the available RNA–protein databases provide detailed analyzed interaction features. Especially, pocket-based information is needed for RNA–protein interaction mechanism understanding at a molecular level.

We propose a pocket-based interaction database, RPpocket, for RNA–protein complexes. We first identify the RNA–protein interaction interface and binding sites. Then, we systematically analyze the secondary structures and topological features for the interface interaction. The results show that the protein loop prefers to bind in the RNA stem region. The electrostatic interactions are the main force for long-range recognition,

while short-range forces optimize and strengthen the interactions. RPpocket provides a straightforward framework to reveal RNA–protein recognition principles for RNA-related drug development or medical applications.

2. Results

2.1. Overview of the RPpocket Database

There are 74 non-redundant RNA–protein structures and 793 pockets in the RPpocket database. The RNA length is 20~266 nucleotides, and the protein length is 22~477 amino acids. According to the different functions, we divided these 74 structures into the following categories (Figure 1 and Table S1): tRNA (14), rRNA (9), mRNA (5), riboswitch (2), microRNA (3), aptamer (6), dsRNA (5), snRNA (1), ribozyme (2), and other (27). Unlike most databases that only contain sequence and structure information, RPpocket provides more detailed interaction information with binding and topological characteristics.
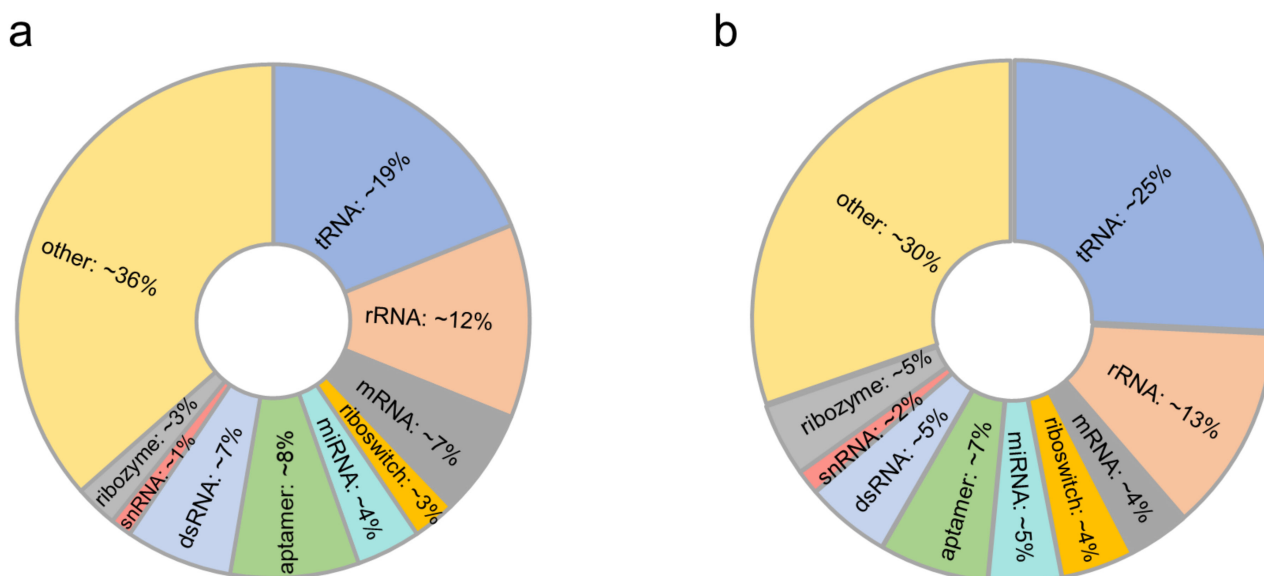

Figure 1. (**a**,**b**) the proportion of various types of RNA and the pocket distribution of each RNA type in our dataset.

2.2. Characteristics of Binding Motifs

The sequence features of binding sites can help us understand the RNA–protein interaction mechanism. We focused on the RNA–protein interactions formed by single RNAs and single proteins to avoid the induced structural changes of multi-body interactions. Here, we retained the binding fragments with length $\geq$ 2 nt/aa, which led to 149 RNA- and 285 protein-binding sites. In addition, we screened the binding fragments, which appeared at least twice. Thus, 11 RNA- and 44 protein-binding fragments are frequently involved in RNA–protein interactions (Figure 2a,b). It is noted that 'CG' and 'GG' were the most favorable binding fragments in RNA while 'KR' was the most favorable binding fragment in protein. We further investigated the nucleotide distributions (Figure 2c,d). The distributions of G (31.14%) were significantly higher than C (26.41%), A (21.42%), and U (21.03%). G formed interactions with amino acids more easily due to its double-ring side-chain chemical structure.
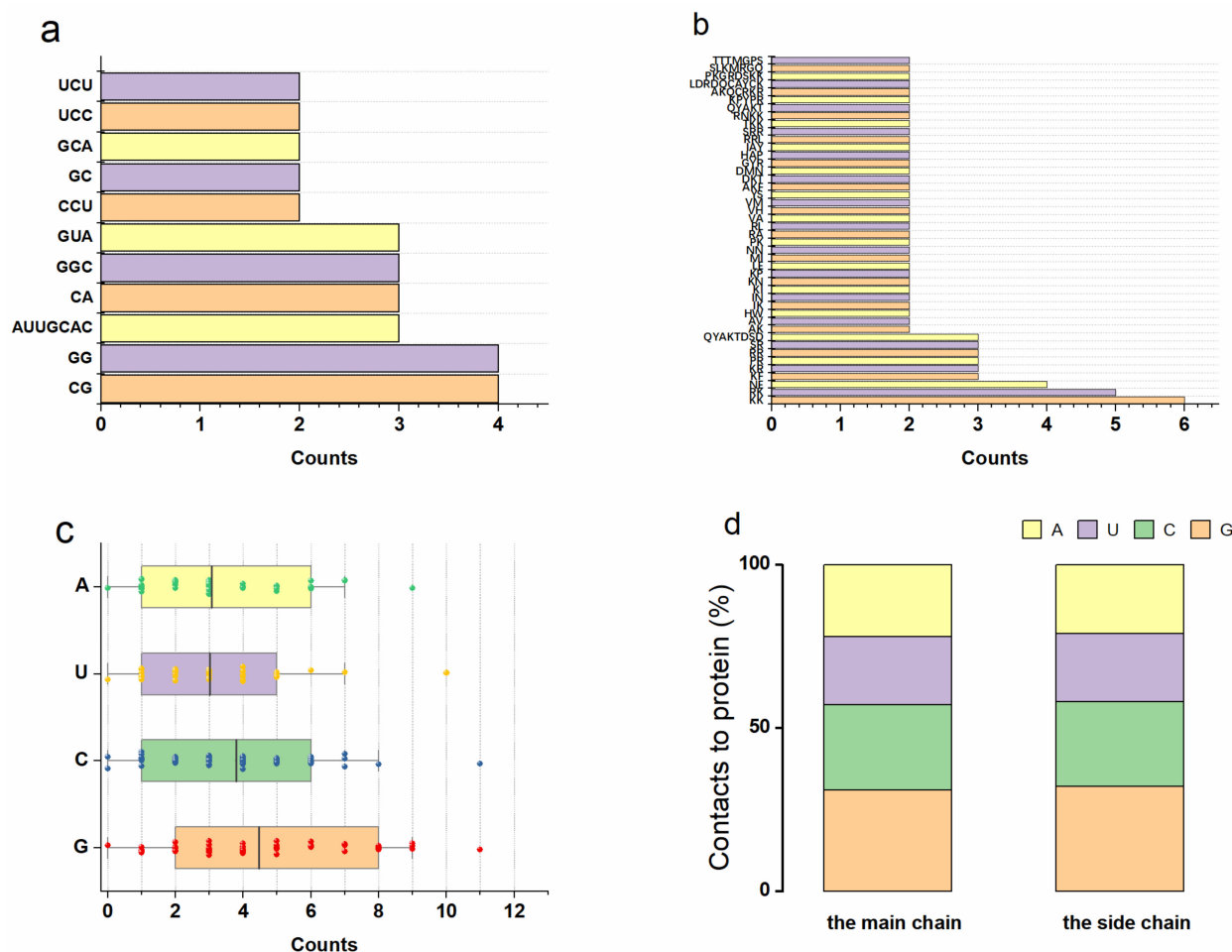
Figure 2. (**a**,**b**) The statistical results of the RNA/protein-binding fragments. (**c**) The RNA nucleotide distributions located at the RNA–protein interface. The distributions of G (31.3%) and C (26.4%) are significantly higher than A (21.5%) and U (20.8%). (**d**) The number of the RNA nucleotide backbone (phosphate and sugar) and side-chain (base) interactions with proteins.

In the previous research, Wang et al. analyzed two RNA–ligand datasets for binding mechanisms. One RNA–ligand dataset was constructed by Wang et al. [23], and the other RNA–ligand dataset was built by Philips et al. [24]. The results showed that 98% of nucleotides were located at or near the loop regions, suggesting that secondary structures can identify binding features [23]. Therefore, we identified and analyzed 137 RNA secondary structure fragments and 312 protein secondary structure fragments involved in RNA–protein interactions. The sequence and secondary structure information for RNAs are presented in Table S2. Table S3 shows the statistical secondary structure analysis for RNA–protein interactions. The stem region is the favorite secondary structure motif for RNA–protein interactions in rRNA, tRNA, aptamer, and others. However, the hairpin loop and internal loop are the favorite secondary structure motifs for RNA–protein interactions in riboswitch and mRNA. The loop region is the favorite secondary structure motif for RNA–protein interactions for proteins. Overall, the stem region is the most favorite motif in RNA secondary structures, followed by the hairpin loop, internal loop, single-stranded, bulge, and multiple loops. At the protein interaction interface, the secondary structure motifs are most commonly in the loop region, followed by the helix region, and the β-sheet region (Figure 3a). The structural views show that the protein loops prefer to interact with the major groove in the RNA stem (Figure 3b).
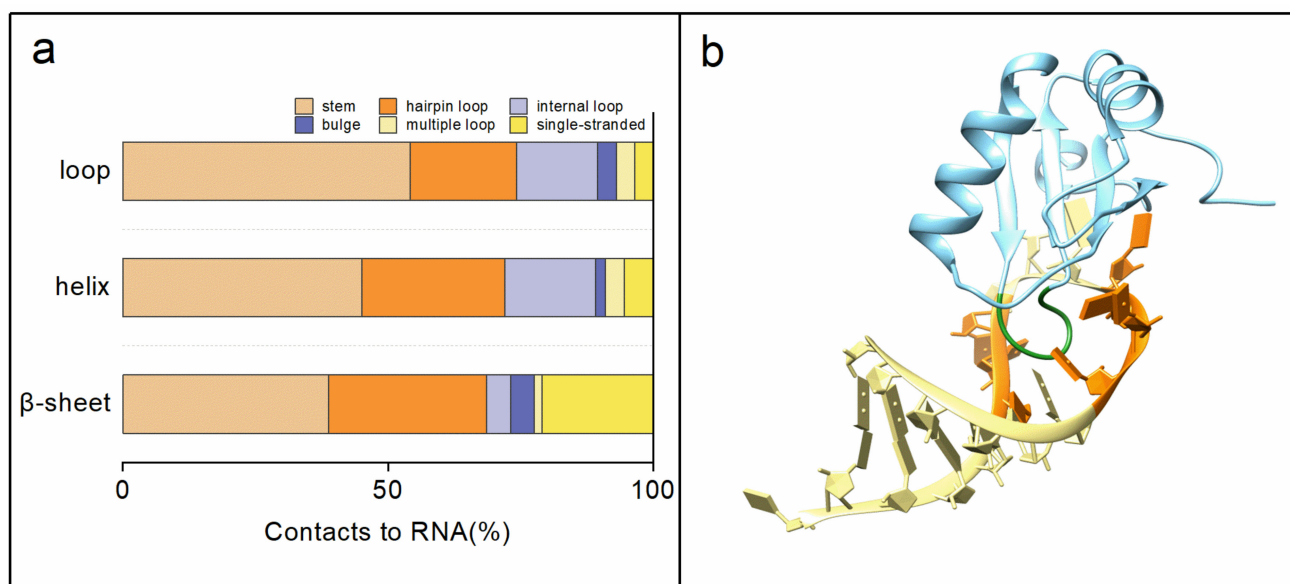
Figure 3. (**a**) The secondary structure distributions at the RNA–protein interaction interface. The protein loops prefer to bind to the RNA stem region. (**b**) An example of a protein loop binding to an RNA stem region (PDB code: 1URN). Green and orange represent the binding interface between protein and RNA.

The pocket geometrical volume and surface area information are some of the most critical characteristics of RNA–protein interactions. We analyzed the topological features of the RNA pockets, including 60 binding and 32 non-binding pockets. The volume, surface area, effective radius, sphericity, and pocket centroid were calculated and recorded in our database. We performed a statistical pocket geometrical analysis for different RNA categories (Table S4). The results showed that the binding pockets' volume and surface area were larger than the non-binding pockets in rRNA, riboswitch, tRNA, and microRNA. However, the volume and surface area of the binding pockets were smaller than the non-binding pockets in aptamer. Figure 4 indicates that the binding pockets (volume of 1670.03 $Å^3$; surface area of 923.77 $Å^2$) were more extended than the non-binding pockets (volume of 1122.91 $Å^3$; surface area of 678.94 $Å^2$).
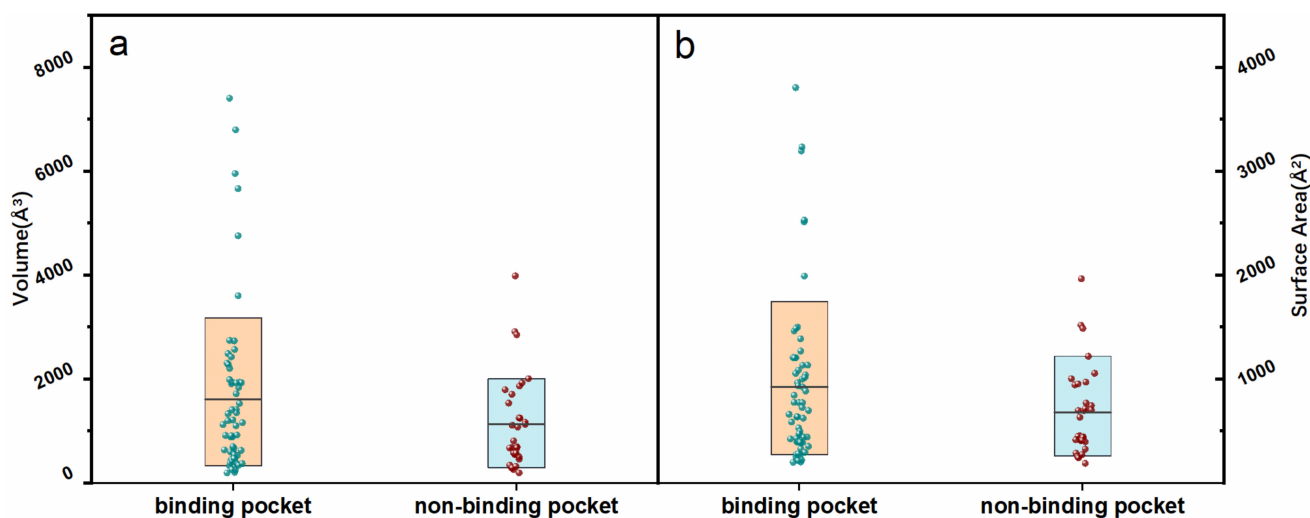

Figure 4. The geometric information distribution of surface area (**a**) and volume (**b**) for RNA-binding pockets and -non-binding pockets. Boxes in (**a**,**b**) represent the 80% core distributions (ranging from 10th to 90th percentile). The black lines on the box represent the average value.

2.3. The RNA–Protein Interaction Mechanism

To understand RNA–protein interaction patterns, we calculated the binding frequency between the nucleotides and amino acids of the representative structures in the RPpocket database. There were more positive amino acids (histidine, lysine, and arginine) involved in the interactions (31.48%) than negative amino acids (aspartic and glutamic, 7.57%). Moreover, the proportions of hydrophobic amino acids (alanine, valine, leucine, isoleucine, phenylalanine, tyrosine, tryptophan, and methionine), polar amino acids (cysteine, serine, threonine, asparagine, and glutamine), and other amino acids (glycine and proline) were 26.94%, 23.21%, and 10.80%, respectively (Figure 5a). We further divided the nucleotides into the backbone (phosphate and sugar) and side chain to explore the interaction characteristics. Figure 5b,c show that the positive amino acids have the highest probability in all interactions. The phosphate group also showed a higher probability than the sugar and side chain. Finally, we analyzed the hydrogen bonds and hydrophobic interactions using Ligpolt+ [25,26]. Figure 6a,b show that HIV-1 TAR RNA binds to the RNA recognition motif (RRM) (PDB ID: 6CMN). Fourteen hydrogen bonds and two hydrophobic contacts were involved in this RNA–protein interaction. Moreover, there were also 14 electrostatic interactions between HIV-1 TAR RNA binding with RRM. It is noted that the β2-β3 loop of the RRM was inserted into the pocket of the RNA stem region. In another example (Figure 6c,d, PDB ID:1A1T), the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA and showed similar characteristics. The results indicate that long-range electrostatic interactions bring RNA and protein together. Then, the short-range interactions optimize the complex.
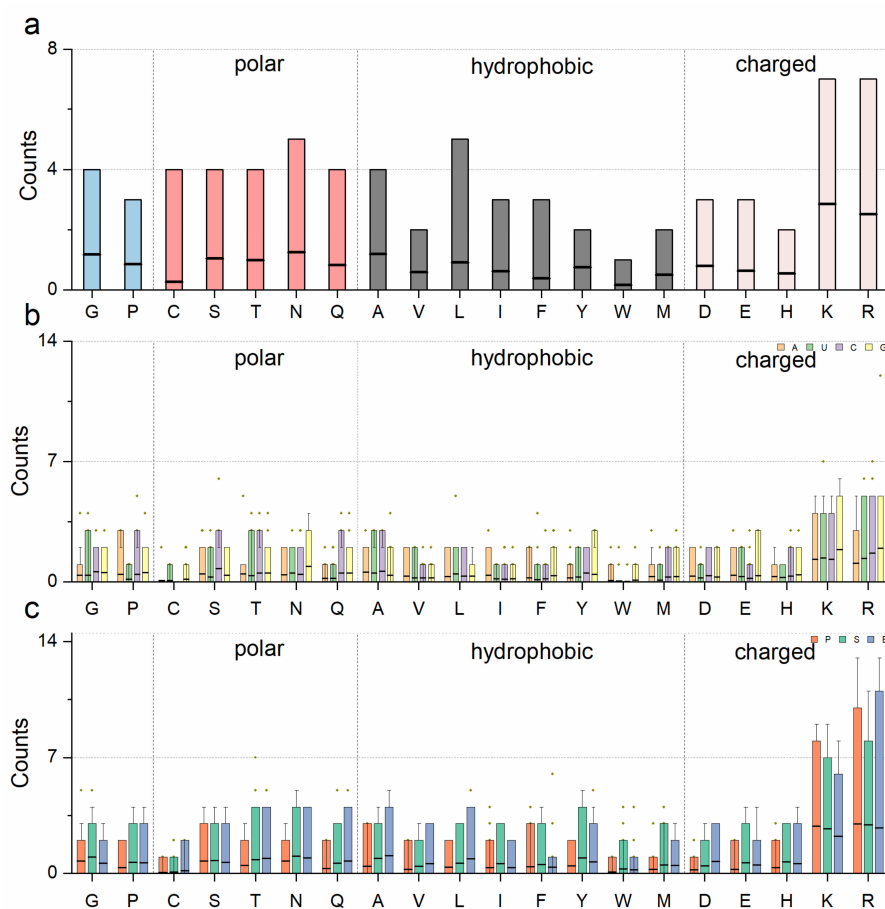

Figure 5. (**a**) The average frequency of the amino acids located at the RNA–protein interface. (**b**) The average contribution of the amino acids interacting with nucleotides. (**c**) The average contribution of the amino acids interacting with RNA backbone (phosphate and sugar) and bases. Boxes in (**a**–**c**) represent the 90% core distribution (ranging from the 5th to 95th percentile). The black lines on the box represent the average values.
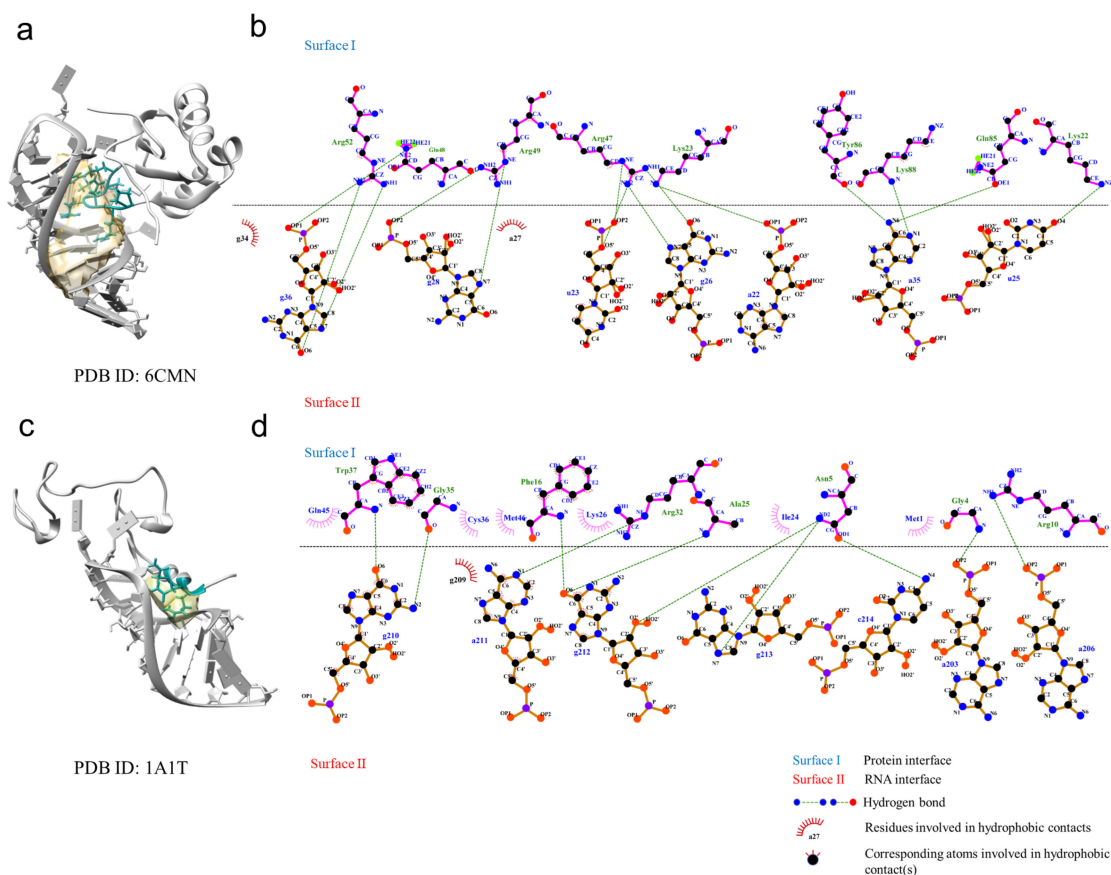
Figure 6. Examples of RNA-binding pocket proteins. (**a**) The structure of HIV-1 TAR RNA binds to the RNA recognition motif (PDB ID: 6CMN). (**b**) Hydrogen bond and hydrophobic contacts for the RNA–protein interface (PDB ID: 6CMN). (**c**) The structure of HIV-1 nucleocapsid protein binds to the SL3 psi-RNA (PDB ID: 1A1T). (**d**) Hydrogen bond and hydrophobic contacts for the RNA–protein interface (PDB ID: 1A1T).

2.4. The Advantages of the RPpocket Database

In general, RPpocket has the following characteristics and advantages: (1) containing 132 RNA pockets and 661 protein pockets extracted from non-redundant 74 RNA–protein complexes; (2) providing the sequence and tertiary structure of all complexes in the dataset; (3) showing the interaction patterns of binding sites and secondary structures between RNA and protein; (4) analyzing the topological pocket features (surface area, volume, effective radius, spherical similarity score, etc.); (5) allowing the users to rotate and translate the structure in the visualization module; (6) supplying tools for RNA complex prediction.

RPpocket has eight modules: Home, Search, Visualization, Download, Links, Tutorial, Statistics, and Contacts. The Home module mainly introduces the RPpcoket database and the navigation to other modules (Figure S1). The search module consists of four parts (Figure 7): a drop-down selection box, a table of RNA descriptions (Figure S2), a summary table of RNA–protein complexes, and a sequence preview module (Figure S3). Users can select the pocket by the RNA class drop-down selection box, choose a complex from the entry representative, and click the submit button to obtain the relevant information about the complex. In the visualization module, users can upload and investigate the structure. The structure will be visualized in four representations: "spacefill", "wire", "ball&stick", and "cartoon". Users can zoom and rotate the structure, generate and save pictures. The download module provides pocket-based interactions for each RNA–protein complex. Users can download the information in xlsx format and the structures in MRC or PDB formats. The link module provides resources for RNA-related simulation, prediction, and

databases. The Tutorial module offers the RPpocket introduction and the abbreviation for the RPpocket database. Data analysis results are provided in the Statistics module. The Contacts module provides emails for users to comment or ask questions.

Figure 7. The search module of RPpocket. The user interface provides the RNA cluster, RNA–protein pocket topology, and sequence motif patterns. The sequence colored in green shadow represents the RNA-protein interaction or binding sites on the pocket.

3. Discussion

The NDB (Nucleic Acid Database) and PDB (Protein Data Bank) are widely used databases for biologists [12,13]. The NDB provides information about experimentally determined nucleic acids and complex structures. The PDB includes archive information with the tertiary shapes of proteins, nucleic acids, and complex structures. The NDB and PDB provide a structural view of biology. However, both databases do not provide detailed interaction features between RNA and protein.

Currently, several databases focus on RNA–protein complexes (Table S5). NPInter is a database that collects RNA-related structural interactions by integrating published literature and processing high-throughput sequencing. It provides detailed annotations and predicted interaction scores for complexes [17]. RNAInter integrates experimental validation and computational predictions of RNA interaction data from the literature and 35 other resources [27]. In addition, it provides interaction annotations, including RNA modification sites and RNA subcellular localization [15]. Unfortunately, these databases do not perform detailed calculations of the RNA pockets. For example, HDOCK used three published protein–RNA benchmarks as testing sets. The protein–RNA docking benchmark 1.0 by the Zou group [28], the protein–RNA docking benchmark v1.1 from the Fernandez-Recio group, and the protein–RNA docking benchmark version 2 from the Bahadur group [29,30]. The NPDOCK used 12 protein–RNA complexes from the Varani and Fernandez benchmarks [31,32].

Several existing protein-related databases also provide sequence, structure, or interaction information (Table S5). For example, (1) HIPPIE, Tissue Net, and MyProtein Net

provide tertiary structure information [33–35], and (2) the BioGRID provides interaction motif information [36]. However, well-analyzed pocket and binding sites are still needed, especially when targeting the pockets in RNA for drug development. We plan to update the database every two years.

4. Materials and Methods

4.1. Construction of RNA–Protein Dataset

To construct the RNA–protein dataset (Figure S4), we used the search options "Entry Polymer Types" and "Protein/NA" in the structure attribute of the advanced search page of the Protein Data Bank before 22 March 2021, then added the additional condition of "Structure Keywords" and "RNA", which means that our dataset includes complexes containing only RNA and protein. We extracted the RNA–protein structures with a single RNA chain binding to proteins with chains ranging from 1 to 10. Moreover, the RNA–protein complex structures were composed of 20~500 nucleotides or amino acids. The first structural model was selected if the RNA had several NMR structures. Then, we used the CD-hit server to discard all the RNA–protein complexes with RNA sequence identities > 95% [37,38]. Thus, 74 RNA–protein complexes were left to construct the dataset. Among them, 50 structures were obtained by X-ray diffraction, 23 structures by NMR, and 1 by electron microscopy.

4.2. Pocket and Binding Site Identification

An RNA–protein interface was defined if the heavy atom distance between nucleotide and amino acid was less than 4 Å. We extracted the secondary structure in dots and brackets format from the RNA FRABASE 2.0 [39]. The secondary structure and tertiary structure visualizations are shown by TBI-forna and PyMOL [40,41].

The RNA pockets were detected using a rolling probe method with the 3V program [42–45]. This extracts the volume quickly by taking the difference between the surface of the solvent with two rolling probes. We used the default outer (10 Å) and inner (3 Å) probe radius for volume detection. The sphericity ψ represents the similarity between the pocket and the standard sphere of the pocket, ranging from 0 to 1. A higher value indicates a more standard sphere. The following formula defines the sphericity:

$$\psi = \frac{\pi^{1/3}(6V)^{2/3}}{A}$$

where *V* and *A* represent the volume and area of the pocket [45,46].

The protein pockets were detected by DoGSiteScorer, which is widely used for detecting protein surface pockets and subpockets [47–50]. The DoGSiteScorer uses heavy atom coordinates to detect the pockets on the protein surface, provide the pocket mesh shape, and mark the grid points. Then, DoGSiteScorer calculates the volume and surface area values by multiplying the number of mesh points.

5. Conclusions

We statistically analyzed RNA–protein complex sequences, secondary structures, and topological pocket features in this work. The results show that protein loops prefer to bind to the RNA stem region. Long-range electrostatic interactions bring RNA and protein together. Short-range interactions optimize the complex. We also developed one easy-to-use database, RPpocket, to provide pocket-based topological interaction information. We hope that RPpocket will facilitate RNA-related study and inhibitor design.

Author Contributions: R.Y. and H.L. collected and calculated the data; L.Y. and X.L. collated and checked the data table; T.Z. wrote the calculation program; Y.Z. supervised the overall study, analyzed

the data, and wrote the paper. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: RPpocket is available at http://zhaoserver.com.cn/RPpocket/RPpocket.html, accessed on 23 May 2022.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Butler, J.E.F.; Kadonaga, J.T. The RNA polymerase II core promoter: A key component in the regulation of gene expression. *Genes Dev.* **2002**, *16*, 2583–2592. [CrossRef] [PubMed]
2. Khalil, A.M.; Rinn, J.L. RNA–protein interactions in human health and disease. *Semin. Cell Dev. Biol.* **2011**, *22*, 359–365. [CrossRef] [PubMed]
3. Li, Z.; Nagy, P.D. Diverse roles of host RNA binding proteins in RNA virus replication. *RNA Biol.* **2011**, *8*, 305–315. [CrossRef] [PubMed]
4. Wool, I.G. The Structure and Function of Eukaryotic Ribosomes. *Annu. Rev. Biochem.* **1979**, *48*, 719–754. [CrossRef] [PubMed]
5. Gehring, N.H.; Wahle, E.; Fischer, U. Deciphering the mRNP Code: RNA-Bound Determinants of Post-Transcriptional Gene Regulation. *Trends Biochem. Sci.* **2017**, *42*, 369–382. [CrossRef]
6. Müller-McNicoll, M.; Neugebauer, K.M. How cells get the message: Dynamic assembly and function of mRNA-protein complexes. *Nat. Rev. Genet.* **2013**, *14*, 275–287. [CrossRef]
7. Rissland, O.S. The organization and regulation of mRNA—protein complexes. *Wiley Interdiscip. Rev. RNA* **2016**, *8*, e1369. [CrossRef]
8. Zhang, S.; Huang, W.; Ren, L.; Ju, X.; Gong, M.; Rao, J.; Sun, L.; Li, P.; Ding, Q.; Wang, J.; et al. Comparison of viral RNA–host protein interactomes across pathogenic RNA viruses informs rapid antiviral drug discovery for SARS-CoV-2. *Cell Res.* **2021**, *1*, 9–23. [CrossRef]
9. Jian, Y.; Wang, X.; Qiu, J.; Wang, H.; Liu, Z.; Zhao, Y.; Zeng, C. DIRECT: RNA contact predictions by integrating structural patterns. *BMC Bioinform.* **2019**, *20*, 497. [CrossRef]
10. Wang, H.; Guan, Z.; Qiu, J.; Jia, Y.; Zeng, C.; Zhao, Y. Novel method to identify group-specific non-catalytic pockets of human kinome for drug design. *RSC Adv.* **2020**, *10*, 2004–2015. [CrossRef]
11. Wang, H.; Qiu, J.; Liu, H.; Xu, Y.; Jia, Y.; Zhao, Y. HKPocket: Human kinase pocket database for drug design. *BMC Bioinform.* **2019**, *20*, 617. [CrossRef]
12. Coimbatore Narayanan, B.; Westbrook, J.; Ghosh, S.; Petrov, A.I.; Sweeney, B.; Zirbel, C.L.; Leontis, N.B.; Berman, H.M. The Nucleic Acid Database: New features and capabilities. *Nucleic Acids Res.* **2014**, *42*, D114–D122. [CrossRef]
13. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242. [CrossRef]
14. Gong, J.; Shao, D.; Xu, K.; Lu, Z.; Lu, Z.J.; Yang, Y.; Zhang, Q.C. RISE: A database of RNA interactome from sequencing experiments. *Nucleic Acids Res.* **2017**, *46*, D194–D201. [CrossRef]
15. Lin, Y.; Liu, T.; Cui, T.; Wang, Z.; Zhang, Y.; Tan, P.; Huang, Y.; Yu, J.; Wang, D. RNAInter in 2020: RNA interactome repository with increased coverage and annotation. *Nucleic Acids Res.* **2019**, *48*, D189–D197. [CrossRef]
16. Park, B.; Kim, H.; Han, K. DBBP: Database of binding pairs in protein-nucleic acid interactions. *BMC Bioinform.* **2014**, *15*, S5. [CrossRef]
17. Teng, X.; Chen, X.; Xue, H.; Tang, Y.; Zhang, P.; Kang, Q.; Hao, Y.; Chen, R.; Zhao, Y.; He, S. NPInter v4.0: An integrated database of ncRNA interactions. *Nucleic Acids Res.* **2019**, *48*, D160–D165. [CrossRef]
18. Cheng, S.; Zhang, L.; Tan, J.; Gong, W.; Li, C.; Zhang, X. DM-RPIs: Predicting ncRNA-protein interactions using stacked ensembling strategy. *Comput. Biol. Chem.* **2019**, *83*, 107088. [CrossRef]
19. Suresh, V.; Liu, L.; Adjeroh, D.; Zhou, X. RPI-Pred: Predicting ncRNA-protein interaction using sequence and structural information. *Nucleic Acids Res.* **2015**, *43*, 1370–1379. [CrossRef]
20. Tuszynska, I.; Magnus, M.; Jonak, K.; Dawson, W.; Bujnicki, J.M. NPDock: A web server for protein–nucleic acid docking. *Nucleic Acids Res.* **2015**, *43*, W425–W430. [CrossRef]

21. Yan, Y.; Zhang, D.; Zhou, P.; Li, B.; Huang, S.-Y. HDOCK: A web server for protein–protein and protein–DNA/RNA docking based on a hybrid strategy. *Nucleic Acids Res.* **2017**, *45*, W365–W373. [CrossRef] [PubMed]

22. Lang, B.; Armaos, A.; Tartaglia, G.G. RNAct: Protein–RNA interaction predictions for model organisms with supporting experimental data. *Nucleic Acids Res.* **2018**, *47*, D601–D606. [CrossRef] [PubMed]

23. Wang, K.; Jian, Y.; Wang, H.; Zeng, C.; Zhao, Y. RBind: Computational network method to predict RNA binding sites. *Bioinformatics* **2018**, *34*, 3131–3136. [CrossRef] [PubMed]

24. Philips, A.; Milanowska, K.; Łach, G.; Bujnicki, J.M. LigandRNA: Computational predictor of RNA–ligand interactions. *RNA* **2013**, *19*, 1605–1616. [CrossRef]

25. Laskowski, R.A.; Swindells, M.B. LigPlot+: Multiple ligand–protein interaction diagrams for drug discovery. *J. Chem. Inf. Model.* **2011**, *51*, 2778–2786. [CrossRef]

26. Wallace, A.C.; Laskowski, R.A.; Thornton, J.M. LIGPLOT: A program to generate schematic diagrams of protein-ligand interactions. *Protein Eng. Des. Sel.* **1995**, *8*, 127–134. [CrossRef]

27. Brown, G.R.; Hem, V.; Katz, K.S.; Ovetsky, M.; Wallin, C.; Ermolaeva, O.; Tolstoy, I.; Tatusova, T.; Pruitt, K.; Maglott, D.R.; et al. Gene: A gene-centered information resource at NCBI. *Nucleic Acids Res.* **2014**, *43*, D36–D42. [CrossRef]

28. Huang, S.-Y.; Zou, X. A nonredundant structure dataset for benchmarking protein-RNA computational docking. *J. Comput. Chem.* **2012**, *34*, 311–318. [CrossRef]

29. Nithin, C.; Mukherjee, S.; Bahadur, R.P. A non-redundant protein-RNA docking benchmark version 2.0. *Proteins Struct. Funct. Bioinform.* **2016**, *85*, 256–267. [CrossRef]

30. Pérez-Cano, L.; Jiménez-García, B.; Fernandez-Recio, J. A protein-RNA docking benchmark (II): Extended set from experimental and homology modeling data. *Proteins Struct. Funct. Bioinform.* **2012**, *80*, 1872–1882. [CrossRef]

31. Zheng, S.; Robertson, T.A.; Varani, G. A knowledge-based potential function predicts the specificity and relative binding energy of RNA-binding proteins. *FEBS J.* **2007**, *274*, 6378–6391. [CrossRef]

32. Pérez-Cano, L.; Solernou, A.; Pons, C.; Fernández-Recio, J. Structural Prediction of Protein-RNA Interaction by Computational Docking with Propensity-Based Statistical Potentials. *Biocomputing* **2009**, *2010*, 293–301. [CrossRef]

33. Alanis-Lobato, G.; Andrade-Navarro, M.A.; Schaefer, M.H. HIPPIE v2.0: Enhancing meaningfulness and reliability of protein–protein interaction networks. *Nucleic Acids Res.* **2016**, *45*, D408–D414. [CrossRef]

34. Barshir, R.; Basha, O.; Eluk, A.; Smoly, I.Y.; Lan, A.; Yeger-Lotem, E. The TissueNet database of human tissue protein–protein interactions. *Nucleic Acids Res.* **2012**, *41*, D841–D844. [CrossRef]

35. Basha, O.; Flom, D.; Barshir, R.; Smoly, I.; Tirman, S.; Yeger-Lotem, E. MyProteinNet: Build up-to-date protein interaction networks for organisms, tissues and user-defined contexts. *Nucleic Acids Res.* **2015**, *43*, W258–W263. [CrossRef]

36. Oughtred, R.; Stark, C.; Breitkreutz, B.-J.; Rust, J.; Boucher, L.; Chang, C.; Kolas, N.; O'Donnell, L.; Leung, G.; McAdam, R.; et al. The BioGRID interaction database: 2019 update. *Nucleic Acids Res.* **2019**, *47*, D529–D541. [CrossRef]

37. Fu, L.; Niu, B.; Zhu, Z.; Wu, S.; Li, W. CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* **2012**, *28*, 3150–3152. [CrossRef]

38. Li, W.; Godzik, A. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **2006**, *22*, 1658–1659. [CrossRef]

39. Popenda, M.; Szachniuk, M.; Blazewicz, M.; Wasik, S.; Burke, E.K.; Blazewicz, J.; Adamiak, R.W. RNA FRABASE 2.0: An advanced web-accessible database with the capacity to search the three-dimensional fragments within RNA structures. *BMC Bioinform.* **2010**, *11*, 231. [CrossRef]

40. Kerpedjiev, P.; Hammer, S.; Hofacker, I. Forna (force-directed RNA): Simple and effective online RNA secondary structure diagrams. *Bioinformatics* **2015**, *31*, 3377–3379. [CrossRef]

41. Delano, W.L. The PyMol Molecular Graphics System. *Proteins Struct. Funct. Bioinform.* **2002**, *30*, 442–454.

42. Connolly, M.L. Analytical molecular surface calculation. *J. Appl. Crystallogr.* **1983**, *16*, 548–558. [CrossRef]

43. Richards, F.M. Areas, volumes, packing and protein structure. *Annu. Rev. Biophys. Bioeng.* **1977**, *6*, 151–176. [CrossRef]

44. Lee, B.; Richards, F. The interpretation of protein structures: Estimation of static accessibility. *J. Mol. Biol.* **1971**, *55*, 379–400. [CrossRef]

45. Voss, N.R.; Gerstein, M. 3V: Cavity, channel and cleft volume calculator and extractor. *Nucleic Acids Res.* **2010**, *38*, W555–W562. [CrossRef]

46. Zhou, T.; Wang, H.; Zeng, C.; Zhao, Y. Additional file 1 of RPocket: An intuitive database of RNA pocket topology information with RNA-ligand data resources. *BMC Bioinform.* **2021**, *22*, 428. [CrossRef]

47. Fährrolfes, R.; Bietz, S.; Flachsenberg, F.; Meyder, A.; Nittinger, E.; Otto, T.; Volkamer, A.; Rarey, M. ProteinsPlus: A web portal for structure analysis of macromolecules. *Nucleic Acids Res.* **2017**, *45*, W337–W343. [CrossRef]

48. Petrova, N.V.; Wu, C.H. Prediction of catalytic residues using Support Vector Machine with selected protein sequence and structural properties. *BMC Bioinform.* **2006**, *7*, 312. [CrossRef]

49. Volkamer, A.; Kuhn, D.; Rippmann, F.; Rarey, M. DoGSiteScorer: A web server for automatic binding site prediction, analysis and druggability assessment. *Bioinformatics* **2012**, *28*, 2074–2075. [CrossRef]

50. Noble, W.S. What is a support vector machine? *Nat. Biotechnol.* **2006**, *24*, 1565–1567. [CrossRef]