

# Analysis of Protein-RNA Interactions at Atomic Level\*

Hyunwoo Kim<sup>1</sup>

hwkim@wilab.inha.ac.kr

Seong-Wook Lee<sup>2</sup>

swl0208@unitel.co.kr

Euna Jeong<sup>1</sup>

eajeong@munhak.inha.ac.kr

Kyungsook Han<sup>1</sup>

khan@inha.ac.kr

<sup>1</sup> School of Computer Science and Engineering, Inha University, Incheon 402-751, Korea

<sup>2</sup> Department of Molecular Biology, Dankook University, Seoul 140-714, Korea

**Keywords:** protein-RNA binding, hydrogen bond, interaction propensity

## 1 Introduction

Many studies of protein-RNA complexes focus on protein instead of RNA or both, and often ignore water-mediated protein-RNA complexes despite their abundant existence. The primary focus of our study is to find interaction propensities not only for proteins but also for RNAs. This paper presents the analysis of 952 hydrogen bonds in the most representative set of protein-RNA complexes, including both direct hydrogen bonds and water-mediated hydrogen bonds.

## 2 Materials and Methods

Protein-RNA complex structures were obtained from the NDB (<http://ndbserver.rutgers.edu/>) and PDB (<http://www.rcsb.org/pdb/>). The complexes solved by X-ray crystallography with a resolution higher than 3.2Å were selected. In order to avoid identical protein-RNA interactions, we ran PSI-BLAST [1] with the  $E$  value of 0.001 and the *identities* value of 80% or below. 29 protein-RNA complexes were finally left as the representative, non-homologous interactions. These include 1A34, 1AQ4, 1ASZ, 1AV6, 1B23, 1B2M, 1B7F, 1C0A, 1CWP, 1DDL, 1DFU, 1DI2, 1DK1, 1DUL, 1E7K, 1E8O, 1EC6, 1F7V, 1FEU, 1FXL, 1H4S, 1I6U, 1JID, 1K8W, 1MMS, 1QU3, 1SER, 2A8V, and 2FMT.

The number of hydrogen bonds between amino acids and nucleotides was calculated using HBPLUS [2]. We follow the definition of a hydrogen bond used in HBPLUS; contacts with a maximum  $d(\text{onor})-a(\text{cceptor})$  distance of 3.9Å, maximum  $h(\text{ydrogen})-a$  distance of 2.5Å, and the minimum  $d-h-a$  angle set to 90° are considered to form hydrogen bonds. In order to know the properties of water-mediated bonds, we have done separate experiments and analyzed the results for two classes of protein-RNA complexes: (1) those with directed bonds and (2) those with water-mediated bonds.

The interaction propensity of Moodie *et al.* [3] was modified as follows:  $P_{ab} = (N_{ab} / \sum N_{ij}) / (N_a \cdot N_b / \sum N_i \cdot \sum N_j)$  where  $N_{ab}$  is the number of amino acid  $a$  in contact with nucleotide  $b$ ,  $\sum N_{ij}$  is the total number of amino acids in contact with any nucleotide,  $N_a$  and  $N_b$  are the total number of  $a$  and  $b$ , respectively, in the dataset ( $D$ ),  $\sum N_i$  is the total number of amino acids in  $D$ , and  $\sum N_j$  is the total number of nucleotides in  $D$ . The propensity value  $P_{ab}$  can reveal the frequency of co-occurrences of amino acid  $a$  and nucleotide  $b$  in the protein-RNA complexes.

---

\*This study was supported by a grant of the IMT-2000 R&D Project, Ministry of Information & Communication, Republic of Korea.

### 3 Experimental Results

The 29 protein-RNA complexes had 952 hydrogen bond contacts between amino acids and nucleotides. Arginine showed the highest propensity ( $P = 3.59$ ), followed by asparagine (2.78), lysine (1.99), and serine (1.77). Common feature of these amino acids is that they are hydrophilic and contain highly electronegative atoms located in the outermost region of their side chains. On the other hand, tryptophan showed the lowest propensity to interact with RNA ( $P = 0.11$ ), followed by valine (0.12) and phenylalanine (0.14). These amino acids are hydrophobic and do not have any highly electronegative atoms. In the case of nucleotides, uracil was the most favored ( $P = 1.89$ ), followed by cytosine (1.12), guanine (0.84) and adenine (0.61). Amino acids showed more diverse interaction propensities ( $P$  in the range [0.11-3.59]) than nucleotides ([0.61-1.89]). This indicates that amino acids are more distinguishable than nucleotides by their propensity values.

While hydrogen bonds were more frequently observed in the side chain (65%) of protein than in the main chain (35%) of protein, there were more hydrogen bonds (57%) observed in the backbone part of RNA than in the base part of RNA (43%). Amino acids, in which side chain contacts are dominant, naturally revealed more diverse interaction propensities than nucleotides. One of the reasons that backbone contacts are more frequent than base contacts in RNA may be that the backbone part has more atoms and highly electronegative atoms than the base part, which makes the backbone more favorable to form hydrogen bonds than the base.

It is essential to analyze water-mediated bonds since they are frequently found in the protein-RNA complexes. A water-mediated bond means that one water molecule forms a hydrogen bond with an amino acid on one side and a nucleotide on the other side. In our analysis, histidine strongly preferred direct bonds to water-mediated bonds, while aspartic acid and cysteine preferred water-mediated bonds to direct bonds. The rest amino acids and nucleotides showed no preference of the two.

Highly electronegative atoms such as nitrogen and oxygen are very important in hydrogen bonding. Amino acids that contain polar atoms (e.g., arginine and lysine) had a strong tendency to form hydrogen bonds. Amino acids that contain an internal ring (e.g., tryptophan and histidine) had a weak tendency to form hydrogen bonds since the ring restricts the movement of atoms. One exception to this is tyrosine; tyrosine indeed has a ring, but it showed a moderate tendency to form hydrogen bonds because of the oxygen at the end of its side chain.

### 4 Conclusions

We have analyzed 952 hydrogen bonds in the most representative set of protein-RNA complexes, including both direct hydrogen bonds and water-mediated hydrogen bonds. The interaction propensity function we developed for this analysis indicates the frequency of co-occurrences of amino acids and nucleotides in the protein-RNA complexes for every combination of amino acids and nucleotides.

Amino acids that are hydrophilic and contain highly electronegative atoms in their side chains showed high propensity. In contrast, hydrophobic amino acids with no highly electronegative atoms had a low propensity to interact with RNA. Amino acids that contain an internal ring had a weak tendency to form hydrogen bonds, except tyrosine. Our long-term goal of this study is to predict the structure of RNA binding protein, and we plan to extend this study to high-level structural analysis of protein-RNA complexes.

### References

- [1] Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J., Gapped BLAST and PSI-BLAST: A new generation of protein database search programs, *Nucleic Acids Res.*, 25:3389–3402, 1997.
- [2] McDonald, I.K. and Thornton, J.M., Satisfying hydrogen bonding potential in proteins, *J. Mol. Biol.*, 238:777–793, 1994.
- [3] Moodie, S.L., Mitchell, J., and Thornton, J., Protein recognition of adenylate: An example of a fuzzy recognition template, *J. Mol. Biol.*, 263:468–500, 1996.