

A Volumetric Survey of Cavities and Electrostatic Patterns in Protein-RNA Binding Sites

Devan Bicher and Brian Y. Chen*

dlb213@lehigh.edu, chen@cse.lehigh.edu

Department of Computer Science and Engineering, Lehigh University
Bethlehem, PA

ABSTRACT

Surveys of cavities and electrostatics in protein structures have yielded significant insights into the role of shape and charge at binding sites on proteins and DNA. Due to the irregularities in the RNA backbone, similar studies have not been performed in RNA, to our knowledge. In particular, non-helical regions of RNA structure lack the frames of reference often used in the analysis of DNA, where the helical secondary structure is more common. To enable an exhaustive analysis of all regions of RNA structure, this paper observes that volumetric methods developed originally for the analysis of protein structures to provide the first survey of binding cavity geometry in every region of RNA structure. Volumetric methods also create an opportunity to examine electrostatic fields in the same binding regions. On a nonredundant representative subset of available protein-RNA structures, this paper reports the relationship between cavities and focused electrostatic fields in RNA structures, as well as their relationship to bound proteins, clarifying our understanding of protein-RNA recognition mechanisms.

KEYWORDS

Protein-RNA interactions, Volumetric Analysis, Electrostatic Focusing

ACM Reference Format:

Devan Bicher and Brian Y. Chen. 2019. A Volumetric Survey of Cavities and Electrostatic Patterns in Protein-RNA Binding Sites. In *10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics (ACM-BCB '19)*, September 7–10, 2019, Niagara Falls, NY, USA. ACM, New York, NY, USA, 7 pages. <https://doi.org/10.1145/3307339.3343869>

1 INTRODUCTION

Protein-RNA interactions play critical roles in transcription, translation and regulation, performing important functions in disease, gene regulation, and the development of the cell. In all molecular interactions, molecular shape and electric fields play crucial roles in protein-RNA interactions. Cavities in protein shape, for example, have been exhaustively surveyed in regard to their roles in binding ligands [2, 5, 8, 10], and DNA [16]. Cavities create chemical microenvironments and steric complementarity that can permit

catalysis, support binding affinity, and control binding specificity. Likewise, the importance of electric fields in proteins have also been carefully examined [6, 13, 15, 17], confirming the ubiquitous role of complementing and noncomplementing charges in binding specificity. However, while much effort has been spent discovering trends in the effect of shape and charge at protein binding sites, no comparable survey of cavity size and electrostatics has been performed on RNA. This paper examines this challenge and proposes a new way to survey these properties in RNA structures.

To perform such a study, a survey of cavities in RNA structure requires software for identifying and computing cavities as they exist on RNA. This capability has been partially possible for some time, because analysis tools like X3DNA [12] and DSSR [11] enable the identification of grooves in the double stranded “stem” regions of RNA structure. Outside of stem regions, however, the generality of RNA tertiary structure often violates the helical frame of reference used by many methods, so an exhaustive survey of RNA cavities is prevented. This study proposes first to apply a structurally general approach, based only on a representation of the molecular surface as a geometric solid and avoiding dependencies on stem regions to identify all clefts in RNA structure. While the method itself has been applied for the study of protein structures, it is the first general survey of cavity geometry in RNA structures to our knowledge.

A second advantage of geometric solids is that they can also be used represent electrostatic isopotentials and regions of electrostatic focusing, making it possible to directly analyze the relationship between electrostatic and structure features. The role of electric fields has long been established in molecular recognition (e.g. [7, 16]). We used solids to understand how frequently electrostatic interactions between proteins and RNA could be found within RNA cavities. We also used solid representations to describe regions of electrostatic focusing, where narrow cavities exclude the polar solvent and decrease the local dielectric within the cavity. As a result, our methods enable a first survey of the electrostatic nature of RNA cavities outside of canonical helical regions.

Together, the capacity to analyze molecular shape and charge with geometric solids without a reliance on helical frame of reference enables us to examine the relationship between cavities and focusing regions and other characteristics of RNA structure. We examine here the relationship between regions of electrostatic focusing in RNA and the positions of amino acids in bound proteins, the relationship between electrostatic focusing and RNA secondary structure, and the relationship between cavities in RNA structure and secondary structure. Together, these findings paint a general

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
ACM-BCB '19, September 7–10, 2019, Niagara Falls, NY, USA

© 2019 Association for Computing Machinery.

ACM ISBN 978-1-4503-6666-3/19/09...\$15.00

<https://doi.org/10.1145/3307339.3343869>

picture of protein-RNA interactions based on a large set of structures and they highlight similarities and notable differences with protein-DNA interactions.

2 METHODS

In order to survey the spatial relationships between cavities, regions of electrostatic isopotential, RNA secondary structure, and bound proteins, we use a novel combination of existing methods to measure the relationships between these phenomena. We briefly paraphrase these methods for completeness, explaining the novelty of their application here.

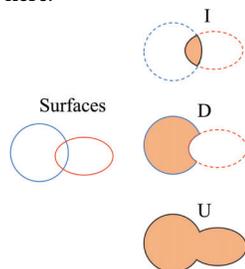


Figure 1: VASP CSG Operations 3D Input surfaces (left, blue and red). CSG operations: (I) the Intersection between the input surfaces, (D) the Difference of the input surfaces, and (U) the Union of the two surfaces.

2.1 Representing Molecular Structure and Electrostatic Fields as Geometric Solids

This study represents molecular structure and electrostatic isopotentials as geometric solids. Each solid is a region of three dimensional space defined by boundary surfaces, and might not be contiguous. The benefit of using geometric solids arises from the fact that they can be analyzed using operations in Constructive Solid Geometry (CSG). These operations, which treat geometric solids as subsets of three dimensional space, include the union, intersection and difference operations. We compute CSG unions, intersections and differences using VASP [4], which was developed originally for the analysis of protein-ligand binding cavities. We are first to repurpose VASP for the analysis of RNA in this study.

The significance of union, intersection, and difference operations is that they can be used to mimic the visual reasoning that is frequently applied by structural biologists. For example, given two aligned protein structures with similar binding sites A and B , it is common to consider the possibility that the region within A that is not within B might sterically hinder ligands that bind in A but not in B . Regions like these are often considered by structural biologists, and they can be found using the CSG difference.

2.2 Solid representations of RNA Cavities

We use CSG operations to define cavities in RNA structure using a method developed originally for VASP [4]. First, we begin with an RNA structure supplied from a PDB file. Next, we compute a molecular surface using the trolbase library [14]. We use the molecular surface to define a geometric solid representing the region occupied by the atoms of the RNA. This surface is the standard solvent accessible surface developed originally by Lee and Richards [9], constructed with a rolling probe sphere that has a radius of 1.4 Å.

We also use a second surface, called the *envelope* surface that is generated in the same manner with a 5.0 Å probe sphere. Using these two surfaces to define geometric solids, we compute the CSG difference of the region within the envelope surface minus the region in the molecular surface. The remaining difference is then separated into individual contiguous regions, and each contiguous region defines a potential cavity on the RNA structure.

2.3 Solid Representations of Electrostatic Isopotentials

To represent electrostatic isopotentials as geometric solids, we first generate electrostatic fields using GRASP[14]. Following the method, developed first for VASP-E [3], the field is queried for regions with electrostatic potential equal to or larger in positive or negative charge than a given threshold. These regions represent the volumes within the electrostatic isopotentials at the given threshold, defining a solid for analysis with CSG. Solid representations of electrostatic isopotentials are used analytically by examining how they overlap with oppositely charged isopotentials, generated with opposite thresholds. The overlap is assessed using a CSG intersection operation, and measuring the volume of the resulting intersection. Larger intersections are said to represent greater electrostatic complementarity, while smaller intersections represent lesser complementarity.

A second way to use geometric solids to examine electrostatic fields is to create solid representations of regions of electrostatic focusing. Electrostatic focusing occurs where the narrowness of molecular cavities exclude solvent, reducing the local dielectric within the cavity and enhancing the field potentials. Focusing has been observed in binding sites in proteins [] and DNA [] because the structure of the molecule enhances the field, enabling interactions that would not occur if the dielectric was treated uniformly, as in the simple coulombic inverse-square calculation. To our knowledge, our characterization of focusing in RNA is the first of its kind.

We generate this representation by using GRASP to compute the electrostatic potential field both with and without a nonuniform dielectric. At every point, the difference in potentials between the two fields represents the degree of enhancement that occurs because of the change in dielectric. Using a technique we developed for VASP-E [], we query this difference in fields for regions where the difference in potentials is larger than a given threshold. We refer to this threshold as the *focusing threshold*. This region is defined as an electrostatic focusing region (FR).

2.4 Large focusing regions

Since all molecules displace the solvent around them, regions of electrostatic focusing can be frequently observed as a thin boundary effect just outside the molecular surface. Inside narrow cavities, however, the focusing effect is enhanced and it can occupy the interiors of entire cavities. For this reason, it is important to disconnect the regions where it occurs as a simple boundary effect from the regions where it is substantial enough to play a role in binding. We refer to these filtered regions as large focusing regions (LFRs).

We developed a method for separating LFRs from FRs in earlier work [], where we demonstrated that LFRs could be used to detect electrostatically active sites in protein and DNA structure. LFRs are

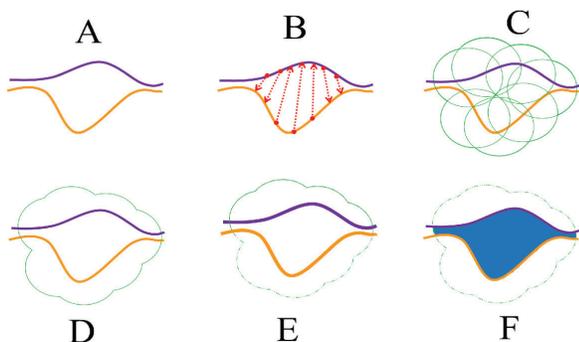


Figure 2: Steps for isolating a large focusing region. Given focusing surface, purple (A), and molecular surface, orange (A), the closest points from one surface to another, red lines in (B), determine if spheres are to be placed on the respective surfaces (green circles in (C)). The union of the spheres (D) minus the molecular surface (E) intersected with the focusing surface creates the large focusing region (F, blue).

generated through a series of CSG operations (Figure 2), beginning with the FR generated earlier, F , and the molecular surface of the same protein, M . For each point p on F , we find the closest distance to any point on M . If the distance is greater than 2.0 \AA , we generate a sphere centered on p with radius 2.0 \AA . We repeat this process on M , generating a series of spheres centered on points of M of the same size. These spheres overlap densely in regions where the focusing surface is distant from the molecular surface. In regions where electrostatic focusing is merely a boundary effect, the distance between the FR and the molecular surface is far less than 2 \AA , so the spheres do not aggregate.

Next, we generate the CSG union of all the spheres that we just generated. We then compute the CSG difference between the focusing region and the molecular surface, and finally the CSG intersection between that difference and the sphere union. This approach identifies the regions where the focusing surface is distant from the molecular surface. Every contiguous region in the resulting solid is separated into an individual LFR. This method was evaluated in earlier work with a parameter sweep of sphere radii [] and shown to accurately identify electrostatically active ligand binding sites in proteins and DNA with the distance threshold at 2.0 \AA . We will apply it here to find potential binding sites in RNA.

2.5 Dataset Construction

We began this survey of RNA protein complexes by running some trial experiments on a small set of RNA structures with the pattern recognition motif and several other representative protein RNA pairs. The initial trials helped guide and refine the experimental methodology in our approach to better understand the complexity of volumetrically based protein RNA electrostatic interactions.

We constructed our dataset to be a nonredundant representation of the protein-RNA complexes available for use in the Protein Data Bank [1]. To construct our dataset we started with any PDB file that explicitly contains both RNA and protein. Any structure that contained the RNA recognition motif were also removed. We parsed the files and removed any that also contained DNA. We also removed structures that contained any RNA or protein component

of the ribosome because many ribosome RNA's are too large for electrostatic field representations with GRASP.

Next, we removed all mutants by performing a text search in the PDB files metadata for the term *mutant* and related word fragments. In doing this we removed structures that were highly similar to one another, but varied in that they were often single mutations of nucleotides or amino acids. This selection reduces output biases that could arise from the many similar inputs.

Finally, we removed members with protein sequence similarity above 95%. We found this similarity through use of the sequence alignment and clustering software package clustalw [7, 6]. ClustalW took the proteins sequences of our dataset as input and clustered everything together that had 95% sequence similarity or greater. With all of the structures clustered together we chose one representative structure from each cluster arbitrarily.

Here we reintroduced the original dataset from the first round of experiments, which was about 150 in total. The final dataset contained 355 structures. Some additional structures were removed out of necessity for different reasons: RNA shorter than 25 nucleotides were discarded. Very large RNA chains were also occasionally reduced to chains long enough to completely describe their interface with a given protein, with generous margins. Structures were removed when our analysis software could not process their expansive size, or because the RNA and protein did not make any contact.

3 RESULTS

3.1 Example Inputs and Outputs

Figure 4 illustrates several example structures and volumes from our study at a focusing level of $0.65(\text{kt}/e)$. In 4(A) the original RNA is in red and protein in blue. In 4(B) and (D) there is a moderately sized focusing chunk, of volume 918.5 \AA^3 , inside a stem secondary structure of the RNA. In 4(C) we can see this focusing region passes through a hole in the stem of the RNA. In 4(E) and (F) we show two of several amino acids, from the blue protein in (A), that intersect this particular structure: arginine and alanine with intersection volumes of 125.68 \AA^3 and 23.83 \AA^3 respectively. In 4(E) the amino acid intersections and distinctions are clearer through the now translucent focusing region; arginine is the larger blue piece in front-left of the smaller alanine in the back-right.

In figure 4 the RNA conforms to the familiar double helix shape and is presented as a more classical RNA example from our data for visual clarity. In reality, the structures in our dataset have a diverse array of shape, focusing region definition, secondary structure conformations and amino acid intersections.

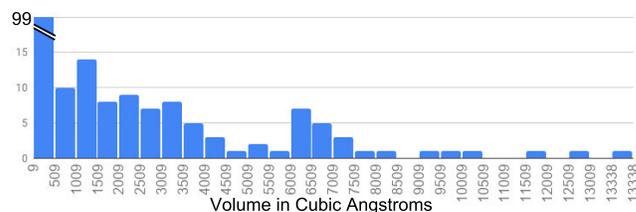


Figure 3: Cavities detected in RNA structures.

A histogram of the volumes of the largest cavity detected in the RNA structures of the dataset.

3.2 A Survey of Cavity Volumes in RNA

To survey the volume of cavities on RNA structure, we separated all RNA chains in our dataset from all other molecules and computed the volume of all cavities on each RNA. On the 355 structures, 1899 cavities were detected in total.

Figure 3 illustrates the distribution of cavity volumes observed in our dataset. Like proteins, RNAs in complex with proteins exhibit many small cavities. Of the 1899 cavities, 380 had volume smaller than 13.26 \AA^3 . Larger cavities exist in lower frequencies with the very largest having volumes in the tens of thousands of cubic angstroms. If we examine only the largest cavity in each RNA, 99 have volume smaller than 500 \AA^3 , but the average volume is 138.9 \AA^3 .

Upon visual examination, the largest cavities by volume are long grooves in RNA where it assumes the double helical conformation. In multiple examples, smaller medium sized cavities also formed in regions where the secondary structure of the RNA was not as regular. Figure 4 illustrates one such example. These findings are consistent with the hypothesis that binding sites can frequently exist outside of groove regions.

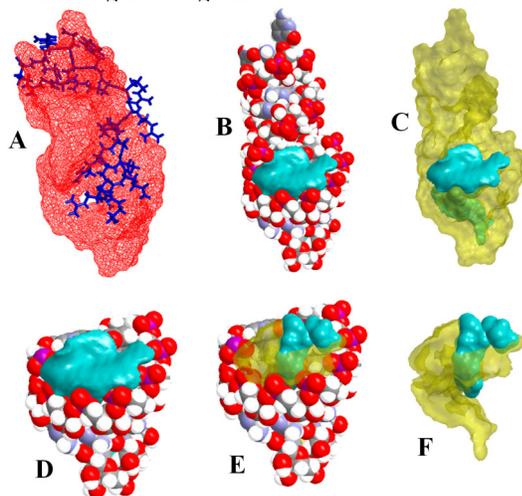


Figure 4: Volumetric Examples of Large Focusing Regions, Secondary Structure, and Amino Acid. (A) RNA in red wire-frame, protein in blue. (B) RNA with a large focusing region, cyan. (C) RNA, yellow transparent & the same large focusing region, cyan. (D) A stem-type RNA secondary structure surrounding the focusing region, cyan. (E) Arginine & alanine in cyan intersecting the same focusing region in yellow surrounded by a stem. (F) The large focusing region, yellow, & two amino acids that intersect it. For clarity, D, E, & F are magnified slightly from A, B, & C but rendered from the same perspective.

3.3 Amino Acids Bound to Large Focusing Regions

The negative charge of the RNA backbone creates a substantial electrostatic field that is attractive to positive amino acids, such as arginine and lysine. Where this field is enhanced by electrostatic focusing, it can be particularly attractive, as has been seen in protein [7] and DNA structures [1]. In such cases, the electrostatic

interaction can stabilize the partner molecule within the focusing region. To examine protein-RNA interactions with this mechanism, we examined how frequently amino acids from bound proteins penetrated the large focusing regions of RNA electrostatic fields.

Using CSG, we measured the average volume of intersection between the molecular surfaces of each of the 20 canonical amino acids and large focusing regions in dataset RNA structures (Figure 5). Tryptophans, arginines, glutamates, leucines and lysines had the highest average volumes of intersection. Since amino acids vary greatly in volume, we also counted the number of amino acids intersecting large focusing regions at more than 1 \AA^3 (Figure 6). Large electrostatic focusing regions were computed at thresholds of 0.65 kt/e, 1.25 kt/e, and 2.5 kt/e.

While several amino acids had large average volumes of intersection, it is clear that arginine and lysine interact with large focusing regions much more frequently than the other amino acids. This is to be expected, since their positive charge complements the negative charge of the RNA backbone. The other amino acids with large average volumes of intersection, glutamic acid, tryptophan, and leucine, do not intersect focusing regions as frequently as arginine and lysine, so while they have a few large intersection volumes they do not display this behavior as consistently as arginine and lysine.

In figure 5 Cysteine, methionine, and tryptophan intersect the least frequently. A few of the amino acids have unusually large average volumes of intersection for focusing level .65, isoleucine, leucine, and methionine, but these averages generally conform with the rest of the amino acids at smaller focusing thresholds. Furthermore, for ILE, LEU, and MET the number of intersections is generally small.

Another anomaly that deserves attention is glutamic acid; it has a rather high intersection average across all three focusing levels. This is particularly unusual considering glutamic acid has a negatively charged side chain. Its number of intersections is not particularly high, so we know the large intersections do not happen often, but the large average is still unusual.

3.4 Relationship between RNA Secondary Structure and Large Focusing Regions

DNA achieves electrostatic focusing in regions where the minor groove of the canonical double helix have a narrow conformation than normal [16]. Similar properties could be expected of RNA when its secondary structure falls into double helical stem regions. Outside these regions, electrostatic fields have not been exhaustively surveyed. For this reason, we counted how frequently large focusing regions come into contact with secondary structures of each type identified by DSSR [11].

Figure 7 illustrates the number and type of secondary structure elements that come into contact with large focusing regions of different volumes. It is clear that on average large focusing regions are frequently in contact with stems, but also frequently with hairpins and i-loops. DSSR defines hairpins as the nucleotides through which RNA transitions from one backbone of a double helical stem to the other. i-loops are small regions where the canonical nucleotide pairings of the stem break down, creating flexible regions in the RNA.

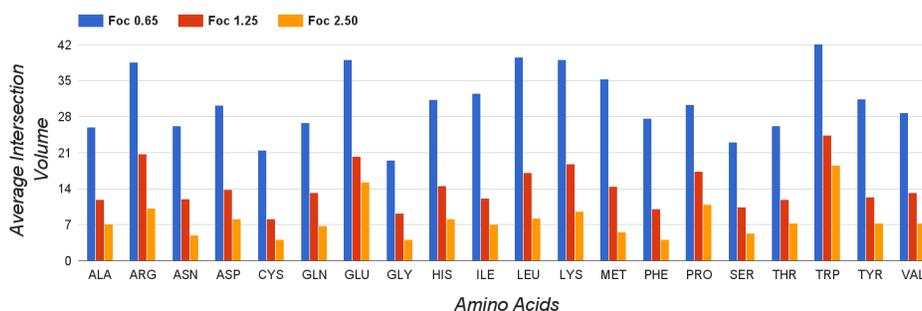


Figure 5: Average intersection volume between amino acids and large focusing regions (\AA^3) at 0.65 kt/e, 1.25 kt/e, and 2.5 kt/e.

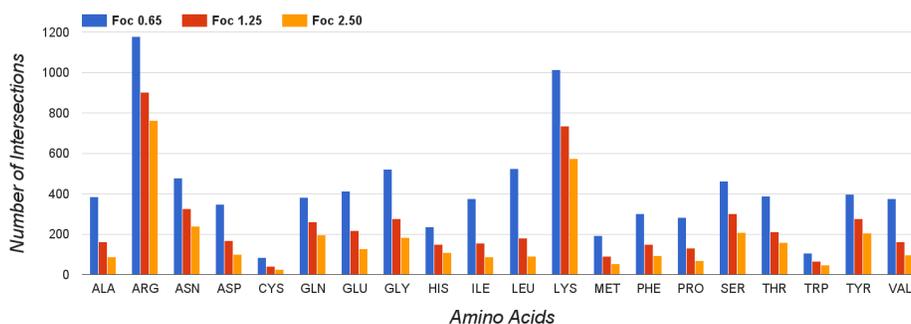


Figure 6: Intersections counted between amino acids and large focusing regions at thresholds 0.65 kt/e, 1.25 kt/e, and 2.5 kt/e.

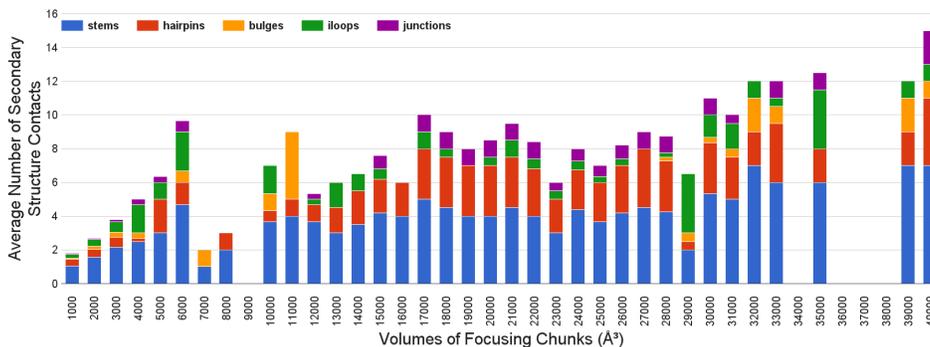


Figure 7: Average RNA secondary structure contacts per volume of large focusing regions at 0.65 kt/e

Since many large focusing regions have volumes less than 1000\AA^3 , we expanded on this region in Figure 8. Smaller LFRs also tend to contact stems approximately half the time, even though the average number of contacts with secondary structures averaged between two and three. These findings indicate that while focusing regions are often in contact with stems, they are also frequently in contact with other secondary structures, notably hairpins, indicating that the double helical structure is not the only region where focusing is occurring. Thus, the relationship between electrostatic focusing and secondary structure takes on a character that is different from DNA, which inhabits the double helical structure with far greater frequency.

These data represent the secondary structure contacts for the 205 RNA structures that had secondary structure. Not all of the RNA molecules had secondary structures identified by DSSR. In

order to be classified as having secondary structure, RNA molecules were required to have at least two backbone strands in a double helical conformation. This requirement is a consequence of the way that DSSR identifies secondary structures. The overall prevalence of secondary structure elements is plotted in Figure 9.

As LFRs increase in volume, it is logical to assume that they might propagate along helical regions of the DNA, thereby maintaining more contacts with stem regions as volume increases. However, if we normalize the heights of the stacked columns of Figure 7, we can see that the proportion of secondary structures in contact with LFRs remains relatively constant (Figure 10). Indeed, similar proportions are maintained even among the contacts of LFRs smaller than 1000\AA^3 , as seen in Figure 11.

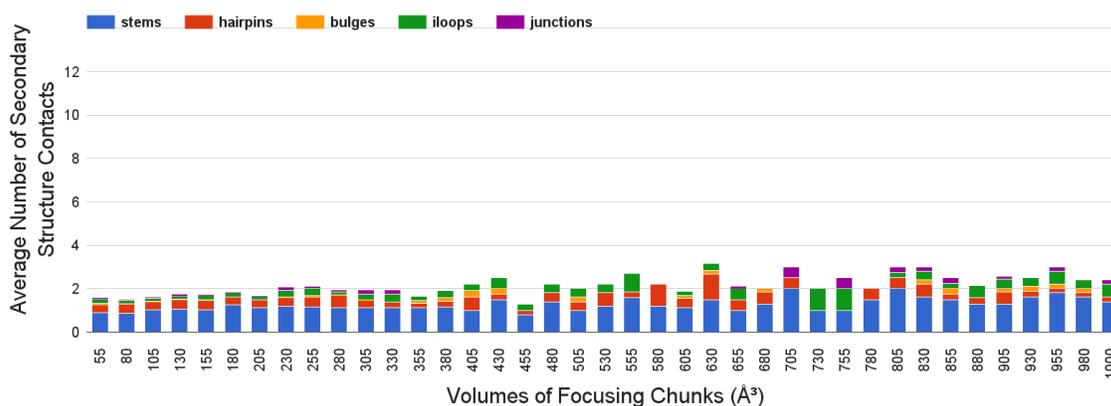


Figure 8: Average RNA secondary structure contacts per volume in large focusing regions smaller than 1000\AA^3 , at 0.65 kT/e .

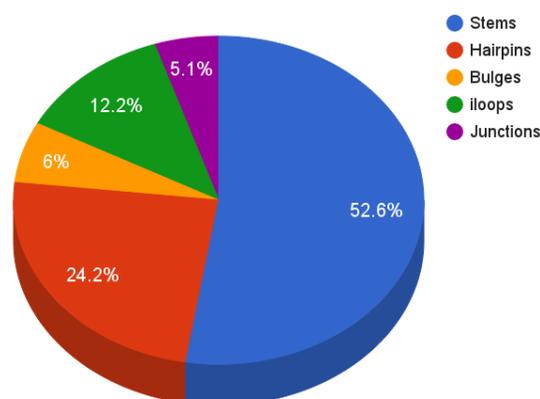


Figure 9: Proportion of RNA secondary structure elements identified by DSSR.

It is not a surprising result that the stems are omnipresent, since stems are required by DSSR to predict secondary structure. However, it is interesting to note that consistently a quarter of the time where focusing occurs, the LFR makes contact with a hairpin.

Discussion We have applied software designed originally for the analysis of protein structure and electrostatics to volumetrically survey RNA structure and electrostatic focusing in protein-RNA complexes. This volumetric approach is distinct from previous methods which are typically based on double helical frames of reference and thus limited to stem regions in RNA. As a result, this study is able to offer a first glimpse into the nature of shape and charge as it exists in RNA-protein interactions outside of stem regions.

Our findings illustrate connections between large focusing regions and secondary structure elements in RNA and between large focusing regions and bound amino acids. Knowing the importance of focusing in protein DNA binding our findings suggest that focusing plays a role in protein-RNA interactions that is cognate to that of protein-DNA interactions. As observed by other studies of protein-DNA interactions (e.g. [16]), we observed the enrichment of positively charged amino acids bound to negatively charged regions of RNA.

In our analysis of amino acid intersections with the large focusing region of RNA, we showed that arginine and lysine not only have the most frequent contact with the focusing region around RNA, figure 6, but that they have the most consistently large volumes of intersection, figure 5. These data suggest that across all the RNA protein binding pairs that we studied, arginine and lysine have the strongest connection with the focusing region.

This finding is consistent with what we know about the charges of both RNA and these amino acids. RNA is strongly negatively charged and lysine and arginine have the strongest positive charge of all amino acids. This result gives support to our theory that focusing enhances interactions with proteins since the electrostatically enhanced areas of focusing around the negatively charged RNA are most frequently interacting with the strongest oppositely charged amino acids of protein.

Whereas we observed the above similarities to the protein-DNA recognition system, we also observed some notable variations. For example, non-stem secondary structures in RNA are almost always in contact with large focusing regions regardless of how large the LFRs are. Whereas electrostatic focusing is a product of variations in groove width in DNA, it is clear that non-stem secondary structures in RNA play a role in focusing, which may be a novel property of RNA.

Overall, these findings illustrate how the generality of a volumetric tool for the examination molecular shape and charge can yield new insights into the mechanisms supporting protein-RNA interactions.

REFERENCES

- [1] Seth Blumenthal, Yisheng Tang, Wenjie Yang, and Brian Chen. 2013. Isolating influential regions of electrostatic focusing in protein and DNA structure. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)* 10, 5 (2013), 1188–1198.
- [2] John A Capra, Roman A Laskowski, Janet M Thornton, Mona Singh, and Thomas A Funkhouser. 2009. Predicting protein ligand binding sites by combining evolutionary sequence conservation and 3D structure. *PLoS computational biology* 5, 12 (2009), e1000585.
- [3] Brian Y Chen. 2014. VASP-E: Specificity Annotation with a Volumetric Analysis of Electrostatic Isopotentials. *PLoS Comput Biol* (2014).
- [4] Brian Y Chen and Barry Honig. 2010. VASP: a volumetric analysis of surface properties yields insights into protein-ligand binding specificity. *PLoS Comput Biol* 6, 8 (2010), 11.

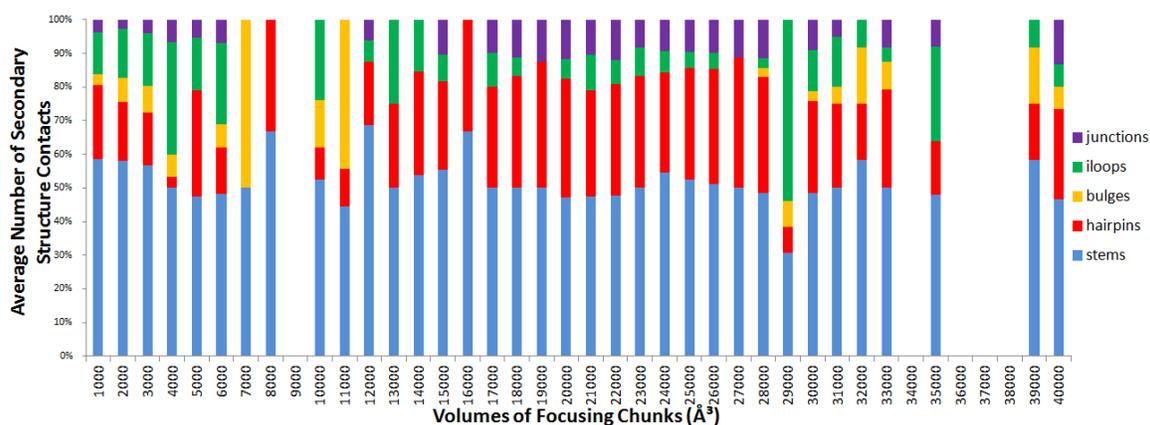


Figure 10: Vertically Normalized Average of RNA Secondary Structure Contacts Per Volume of Large Focusing Regions. Evenly distributed average of times each RNA secondary structure contacts focusing chunks, at level $0.65(kT/e)$, for all volumes.

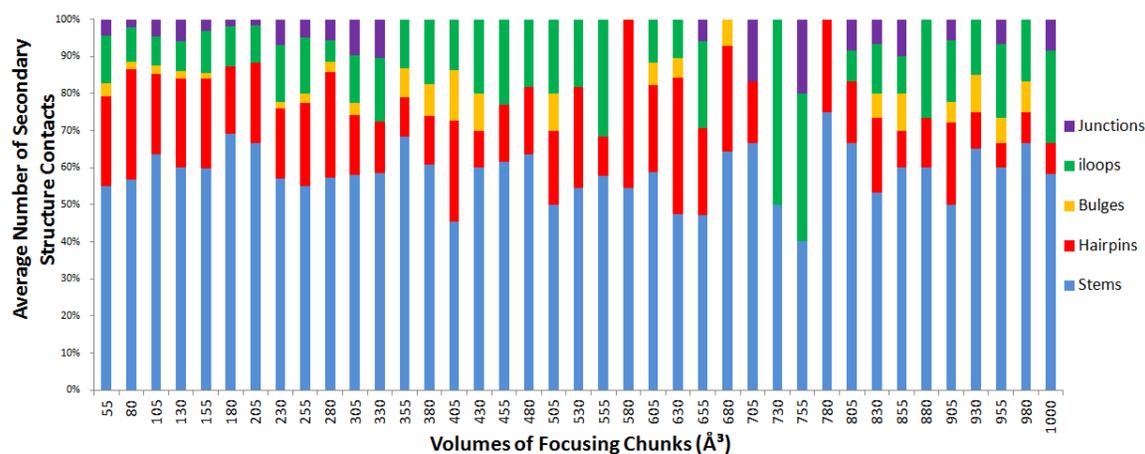


Figure 11: Vertically Normalized Average of RNA Secondary Structure Contacts Per Volume of Large Focusing Regions Less than 1000\AA^3 . Evenly distributed average of times each RNA secondary structure contacts focusing chunks, at level $0.65(kT/e)$, for volumes no greater than 1000\AA^3 .

- [5] Mu Gao and Jeffrey Skolnick. 2013. A comprehensive survey of small-molecule binding pockets in proteins. *PLoS computational biology* 9, 10 (2013), e1003302.
- [6] Barry H Honig, Wayne L. Hubbell, and Ross F. Flewelling. 1986. Electrostatic Interactions in Membranes and Proteins. *Annual Review of Biophysics and Biophysical Chemistry* 15 (1986), 163–193.
- [7] Isaac Klapper, Ray Hagstrom, Richard Fine, Kim Sharp, and Barry Honig. 1986. Focusing of electric fields in the active site of Cu-Zn superoxide dismutase: Effects of ionic strength and amino-acid modification. *Proteins: Structure, Function, and Bioinformatics* 1, 1 (1986), 47–59.
- [8] Roman A Laskowski. 1995. SURFNET: a program for visualizing molecular surfaces, cavities, and intermolecular interactions. *Journal of molecular graphics* 13, 5 (1995), 323–330.
- [9] Byungkook Lee and Frederic M Richards. 1971. The interpretation of protein structures: estimation of static accessibility. *Journal of molecular biology* 55, 3 (1971), 379–IN4.
- [10] Jie Liang, Clare Woodward, and Herbert Edelsbrunner. 1998. Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. *Protein science* 7, 9 (1998), 1884–1897.
- [11] Xiang-Jun Lu, Harmen J Bussemaker, and Wilma K Olson. 2015. DSSR: an integrated software tool for dissecting the spatial structure of RNA. *Nucleic acids research* (2015), gkv716.
- [12] Xiang-Jun Lu and Wilma K Olson. 2003. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic acids research* 31, 17 (2003), 5108–5121.
- [13] H Nakamura. 1996. Roles of electrostatic interaction in proteins. *Quarterly Reviews of Biophysics* 29 (1996), 1–90.
- [14] Donald Petrey and Barry Honig. 2003. GRASP2: visualization, surface properties, and electrostatics of macromolecular structures and sequences. In *Methods in enzymology*. Vol. 374. Elsevier, 492–509.
- [15] M T Record, C F Anderson, and T M Lohman. 1978. Thermodynamic analysis of ion effects on the binding and conformational equilibria of proteins and nucleic acids: the roles of ion association or release, screening, and ion effects on water activity. *Quarterly Reviews of Biophysics* 11, 2 (1978), 103–178. <http://journals.cambridge.org/action/displayAbstract?fromPage=online&aid=4220232>
- [16] Remo Rohs, Sean M West, Alona Sosinsky, Peng Liu, Richard S Mann, and Barry Honig. 2009. The role of DNA shape in protein–DNA recognition. *Nature* 461, 7268 (2009), 1248–1253.
- [17] A Warshel and S T Russell. 1984. Calculations of electrostatic interactions in biological systems and in solutions. *Quarterly Reviews of Biophysics* 17, 3 (1984), 283–422. <http://www.ncbi.nlm.nih.gov/pubmed/6098916>