

Pittsburg State University

Pittsburg State University Digital Commons

Electronic Theses & Dissertations

Fall 12-14-2018

IN SILICO DEVELOPMENT OF AN RNA APTAMER LIBRARY TO BE USE FOR THE SELECTION OF RNA APTAMER THAT TARGET BIOMOLECULES

Nehad Nawfawi

Pittsburg State University, nnawfawi@gus.pittstate.edu

Follow this and additional works at: <https://digitalcommons.pittstate.edu/etd>



Part of the [Biochemistry Commons](#)

Recommended Citation

Nawfawi, Nehad, "IN SILICO DEVELOPMENT OF AN RNA APTAMER LIBRARY TO BE USE FOR THE SELECTION OF RNA APTAMER THAT TARGET BIOMOLECULES" (2018). *Electronic Theses & Dissertations*. 410.

<https://digitalcommons.pittstate.edu/etd/410>

This Thesis is brought to you for free and open access by Pittsburg State University Digital Commons. It has been accepted for inclusion in Electronic Theses & Dissertations by an authorized administrator of Pittsburg State University Digital Commons. For more information, please contact digitalcommons@pittstate.edu.

IN SILICO DEVELOPMENT OF AN RNA APTAMERS LIBRARY TO BE USED FOR THE
SELECTION OF RNA APTAMERS THAT TARGET BIOMOLECULES

A Thesis Submitted to the Graduate School
In Partial Fulfillment of the Requirements
For the Degree of Master of Science

Nehad Omar Nawfawi

Pittsburg State University

Pittsburg, Kansas

November 2018

IN SILICO DEVELOPMENT OF AN RNA APTAMERS LIBRARY TO BE USED FOR THE
SELECTION OF RNA APTAMERS THAT TARGET BIOMOLECULES

Nehad Omar Nawfawi

APPROVED:

Thesis Advisor _____
Dr. Irene Zegar, Department of Chemistry

Committee Member _____
Dr. James McAfee, Department of Chemistry

Committee Member _____
Dr. Khamis Siam, Department of Chemistry

Committee Member _____
Dr. Christine Brodsky, Department of Biology

ACKNOWLEDGEMENTS

I give thanks to God for his protection and for giving me the ability to do my work. I am so grateful to SACM scholarship scheme and the Faculty of Chemistry at Pittsburg state university for making it possible for me to study here. I would like to thank my advisor Dr. Irene Zegar, you have been a great advisor for me. I have been extremely lucky to have an advisor who cared so much about my work. I would also like to thank my committee members, Dr. James Mcafee, Dr. Khamis Siam, and Dr. Christine Brodsky for serving as my committee members. Dr. Siam, you have been a good person and you were as a father for me throughout my Master degree. Furthermore, a special thanks to my family and my friends. Words cannot express how grateful I am to my mother, my sibling, and my friends Balgees and Taghreed for their continued support and encouragement throughout my master's degree. Your prayer for me was what I have been up to now. I would also like to thank you for allowing my defence to be an enjoyable moment, for your wonderful comments and suggestions, thank you.

IN SILICO DEVELOPMENT OF AN RNA APTAMERS LIBRARY TO BE USED FOR THE
SELECTION OF RNA APTAMERS THAT TARGET BIOMOLECULES

An Abstract of the Thesis by
Nehad Omar Nawfawi

The systematic evolution of ligands by exponential enrichment (SELEX) is a powerful method for the development of high affinity RNA ligands toward and infinite array of target molecules. SELEX based upon the generation of a randomized population of RNA or DNA molecules followed by a target molecule that selects high affinity ligands from the randomized population followed by the subsequent amplification of the selected molecules. The procedure of selection and amplification typically carried out through multiple cycles to insure that the identified ligands exhibits the highest affinity toward the target. The procedure is very time- consuming often- taking months to complete, as well as being costly. Another drawback is that every surface that the randomized RNA molecules come in contact with becomes a potential target. Therefore, numerous artefactual ligands are often the product of the procedure. To enhance the efficiency and to reduce the problems associated with controlling surface selection problems we are investigating the efficacy of using a computer generated SELEX procedure. To accomplish this we have generated a library of 5,000 randomized RNA molecules and used comparative modeling to determine the structure of about 1000 of these sequences. SELEX used to identify high affinity RNA ligands for the RNA binding domain of the heterogeneous nuclear ribonucleoprotein C (hnRNP C). To test our system, we included one of these RNA aptamers identified from this study, in our library of randomized molecules, and demonstrated that our computational method could select the high affinity ligand from the random population.

TABLE OF CONTENTS

| CHAPTER | PAGE |
|---|------|
| I. INTRODUCTION..... | 1 |
| II. Materials and Methods..... | 4 |
| i. The Protein Receptor Structure..... | 5 |
| ii. Random RNA Sequence Library Generation..... | 5 |
| iii. Determining the RNA Molecules Secondary Structures..... | 5 |
| iv. Prediction of the Three-Dimensional Structures of the Generated Random RNA Sequences..... | 6 |
| v. Energy Minimization of Lowest-Energy Generated Three-Dimensional RNA Structures..... | 6 |
| vi. Virtual Screening of RNA-Binding Site in HnRNP C RRM Domain...6 | |
| vii. Design the 3D structures by using AutoDock..... | 8 |
| III. Results and Discussion..... | 9 |
| i. Efficacy of Auto Dock in docking RNA molecules to RRM domain of hnRNPC protein | 9 |
| IV. Conclusion..... | 25 |
| V. REFERENCES..... | 27 |

LIST OF TABLES

| Tables | PAGE |
|---|------|
| Table 1: H-Bonding interactions between amino acids in the RRM domain and the nucleotides of the RNA sequence, AUUUUUC found in the co-crystal structure..... | 11 |
| Table 2: H-Bonding interactions between amino acids in the RRM domain and the nucleotides of the RNA sequence, AUUUUUC found in the Docked structure. | 12 |
| Table 3: H-Bonding interactions between amino acids in the RRM domain and the nucleotides of the RNA sequence, AUUUUUC found in the co-crystal structure..... | 12 |
| Table 4: The RNA sequences that resulted in the top 10 RRM-binding affinities..... | 14 |
| Table 5: The Hydrogen Bonds Observed Between RNA 9 Nucleotides and the RRM Domain of HnRNPC. Also Included are Nucleotide-Nucleotide Hydrogen Bonds..... | 15 |
| Table 6: The Hydrogen Bonds Observed Between RNA 16 Nucleotides and the RRM Domain of HnRNPC. Also Included are Nucleotide-Nucleotide Hydrogen Bonds..... | 16 |
| Table 7: The Hydrogen Bonds Observed Between RNA 94 Nucleotides and the RRM Domain of HnRNPC. Also Included are Nucleotide-Nucleotide Hydrogen Bonds | 17 |
| Table 8: The Hydrogen Bonds Observed Between RNA 102 Nucleotides and the RRM Domain of HnRNPC. Also Included are Nucleotide-Nucleotide Hydrogen Bonds..... | 18 |
| Table 9: The Hydrogen Bonds Observed Between RNA 104 Nucleotides and the RRM Domain of HnRNPC. Also Included are Nucleotide-Nucleotide Hydrogen Bonds..... | 19 |
| Table10: The Hydrogen Bonds Observed Between RNA 105 Nucleotides and the RRM Domain of HnRNPC. Also Included are Nucleotide-Nucleotide Hydrogen Bond..... | 20 |
| Table 11: The Hydrogen Bonds Observed Between RNA 106 Nucleotides and the RRM Domain of HnRNPC. Also Included are Nucleotide-Nucleotide Hydrogen Bonds..... | 21 |
| Table 12: The Hydrogen Bonds Observed Between RNA 125 Nucleotides and the RRM Domain of HnRNPC. Also Included are Nucleotide-Nucleotide Hydrogen Bonds..... | 22 |
| Table 13: The Hydrogen Bonds Observed Between RNA 137 Nucleotides and the RRM Domain of HnRNPC. Also Included are Nucleotide-Nucleotide Hydrogen Bonds..... | 23 |
| Table 14: The Hydrogen Bonds Observed Between RNA 144 Nucleotides and the RRM Domain of HnRNPC. Also Included are Nucleotide-Nucleotide Hydrogen Bonds..... | 24 |

LIST OF FIGURES

| Figures | PAGE |
|---|------|
| Figure 1: The structures of RRM Bound to the SELEX Generated RNA Sequence, AUUUUUC..... | 10 |
| Figure 2: The Molecular Docking Structure of the RNA 9 sequence | 15 |
| Figure 3: The Molecular Docking Structure of the RNA 16 sequence..... | 16 |
| Figure 4: The Molecular Docking Structure of the RNA 94 sequence | 17 |
| Figure 5: The Molecular Docking Structure of the RNA 102 sequence..... | 18 |
| Figure 6: The Molecular Docking Structure of the RNA104 sequence | 19 |
| Figure 7: The Molecular Docking Structure of the RNA 105 sequence | 20 |
| Figure 8: The Molecular Docking Structure of the RNA 106 sequence..... | 21 |
| Figure 9: The Molecular Docking Structure of the RNA125 sequence | 22 |
| Figure 10: The Molecular Docking Structure of the RNA 137 sequence..... | 23 |
| Figure 11: The Molecular Docking Structure of the RNA 144 sequence | 24 |

CHAPTER I

Introduction

Aptamers are oligonucleotide or peptide molecules that bind to a target molecule with high affinity. The manner of binding can be likened to that of antigens and antibodies. The word ‘aptamer’ was derived from a Greek word called ‘aptus’ that means ‘to fit’ ^[1].

RNA aptamers are used for biochemical applications because they work in a similar way as antibodies and they have little or no immunogenic effect. They are also easy to synthesize in large quantities and they are considered to be stable thermodynamically compared to antibodies and peptides. They can also be modified to make them more stable in the blood and prevent RNase degradation. They are also single stranded and this allows the binding to be more specific. They also have a smaller size, which enables them to easily penetrate inside the cell. The molecules can also be easily conjugated with ligands, which enables them to be used as agents of drug delivery inside the cell ^[2].

Another useful application of RNA aptamers in diagnosis and treatment of diseases, as they have the ability to bind to target protein sequences of effected cells. For example, the RNA aptamer sequence (5'-CUGCGAUCAGGGGUAAAUUUCCGCGCAGGCUCCACGCCGC-3') that helps in the recognition of 4, 4' methylenedianiline (MDA) in the body has been developed ^[4]. This aptamer binds to the cytotoxic T-lymphocyte-(cytotoxic T cell) associated protein 4 (CTLA-4), and inhibits its activity hence slowing down cancer progression ^[5]. Furthermore, the aptamers of TCF 1 and tenascin –C have also been developed and used in the diagnosis and treatment of colon and breast cancers, respectively ^[2].

Other examples of RNA aptamers as anti-cancer agents was demonstrated by Santulli-Marotto et al who developed the RNA sequence, (5'-GGGAGAGAGGAAGAGGGAUGGGCCGACGUGCCGCAACUUCAACCCUGCACAACCAAUCCGCCCAUAACCCAGAGGUCGAUAGUACUGGAUCCCCC-3') that was found to bind and inhibit the function of CTLA-4 in cancer cells ^[5]. When upregulated in cancer cells, CTLA-4 acts to inhibit the immune system, which results in the progression of cancer cells. RNA aptamers are also found to function as therapeutic and diagnostic vehicles against a number of other diseases ^[2]. In age-related macular degeneration, Pegaptanib, a 28-nucleotide polyethylene glycol conjugated to a 28-nucleotide RNA aptamer is used as therapeutic agent against the disease by binding to Vascular Endothelial Growth Factor (VEGF) that cause vision damage and visual shortage ^[7]. Furthermore RNA aptamers are used for Diabetic Macular Edema (DME) that affects millions of people worldwide and other angiogenesis-related diseases occurring through VEGF^[8].

The use of RNA Aptamers as delivery tools for targeted therapy has gained a lot of attention due to the fact that it is fairly easy to develop specific cell-type functionalized RNA that can deliver toxic drugs to specific cancer cells efficiently while avoiding the cytotoxic effects of the anti-cancer drugs. Moreover, they are biocompatible and they have low or no immunogenicity ^{[2] [6]}.

Traditionally, the generation of RNA aptamers is conducted using systematic evolution of ligands by exponential enrichment (SELEX) to create irregular libraries ^[3]. The SELEX system has been used to generate RNA aptamers for a wide range of targets, including purified proteins, small molecules, live cells, tissues, and microbes ^[9]. The process is composed of three steps which are selection, partitioning and amplification^[10]. A pool of RNA oligonucleotides having fixed flanking regions and randomly arranged center is used to create the RNA aptamers. The pool is mixed with the desired target under suitable binding conditions. A part of the oligonucleotide pool will bind to the target while the rest will remain unbound. By using negative selection, the oligonucleotides that bind partially or incompletely to the target are removed by

binding them to similar targets with unidentical structure. The ones that bind undergo reverse transcription followed by PCR to form a new pool of amplified RNA aptamers. The steps are repeated at least 5 times to get the aptamers with best binding properties ^[21].

The SELEX method takes almost a month to complete and the procedures used for reverse transcription can become dull and tiring. Reverse transcription can become expensive if done on a regular basis. However, we did this project because the traditional SELEX method is labor intensive and pricy, we wanted to use an *in silico* method of developing RNA aptamer library of a large number random sequences that can be used to bind any cellular target.

Once developed it would revolutionize the cancer treatment field because of its cost effectiveness, speed and ability to generate RNA aptamers for any disease or biological process of interest. We found one study that conducted a similar work to what we proposed to do but they have not produced aptamer libraries, this would be the first such work aimed at generating a large library of aptamers that can be used to find specific RNA to bind any biological target ^[22].

CHAPTER II

Materials and Methods:

The computer programs that I used for this work included, RanDNA (Random sequence generator) is a free software used to generate random RNA sequences. This is done by setting the desired percent of each nucleotide, the desired length of the random sequences and the desired number of sequences generated. This software can produce up to 4×10^9 sequences^[11]. The software can be downloaded from the website: www.introni.it/en/software/. RNAfold 2.4.0 is a module included in the Vienna RNA software package. It is used to predict RNA secondary structures. This is done by determining the base-pairing probabilities and structures with the lowest free energy^[12]. However, the Full Atom Refinement protocol (FARFAR) of the Rosetta structural prediction software package was used to predict the three-dimensional structures of the RNA molecules generated by RNAfold. This was done remotely using the webserver, (http://rosie.rosettacommons.org/rna_denovo/submit)^[13]. In addition, UCSF Chimera is a free software that is used to perform structural editing of the three-dimensional RNA structures generated by FARFAR. The structural editing includes adding missing hydrogens, charges and other forcefield parameters needed for energy minimization. Chimera was downloaded from using the URL: <http://www.rbvi.ucsf.edu/chimera>^[14]. YASARA (Yet Another Scientific Artificial Reality Application) View is a molecular graphics program that is used in this thesis to perform energy minimization of three-dimensional RNA structures that have been edited by UCSF Chimera. The program can be downloaded from the URL: <http://www.yasara.org/viewdl.htm>^[15]. Moreover, AutoDock Vina was used to simulate the

interactions between the generated RNA structures (ligands) and the RRM domain of Hnrnp C protein^[16]. The software Vina was used within the virtual screening program PyRx, which was used to perform the Vina calculations on a library of generated RNA structures^[17]. Autodock is software for dock a ligand into protein. It uses to discover many drugs such as HIV1 integrase inhibitors^[18]. It also uses to do structures drug design and view molecules in grid3D^[19]. It shows how to build and edit the hydrogen bonds. In 1990, it developed in Olson's laboratory by Dr. David S. Goodsel to perform rotations and the version is 1.5.6^[20].

Methods:

The Protein Receptor Structure:

The solution structure of HnRNP C RRM domain complexed with the RNA sequence AUUUUUC sequene (PDB ID: 2MXY) was downloaded from the RCSB protein data bank using the URL: www.rcsb.org^[21]. The protein and RNA structures were separated and individually energy minimized using "Repair Object" from the Foldx plugin found under the Analyze Pulldown Menu. The receptor and ligand structures were saved sparately for later use.

Random RNA Sequence Library Generation:

A total of 5000 Random sequences of 7-nucleotide RNA molecules were generated using RanDNA. The sequences were generated by first setting the desired percentage of nucleotides, A, C, G, and U to 25% each. A datafile that will contain the generated sequences was then created by the Save option from the File Pulldown menu followed by the Start Option.

Determining the RNA Molecules Secondary Structures:

The secondary structures of the randomly generated RNA molecules were determined using RNAfold. This was conducted using the following Linux command:

```
RNAfold --constraint=constraints.txt --batch rna1.txt --outfile=rna-out
```

Where input.txt is the file containing the randomly generated RNA sequences, and rna-out is the file containing the secondary structures of each of the random sequences as predicted by RNAfold.

Prediction of the Three-Dimensional Structures of the Generated Random RNA Sequences:

The three-dimensional structures of the generated Random RNA Sequences were calculated using the FARFAR Module of the Rosetta Molecular Simulation Package. Essentially the RNA structures were calculated using the fragment assembly approach which uses an energy function that takes into account backbone conformations and side-chain interactions of experimentally determined RNA structures. The calculations submitted at the Rosetta Online server which can be accessed using the following URL:

rosie.rosettacommons.org/rna_denovo/submit

For successful submission, the RNA sequence must be inputted along with the secondary structure predicted by RNAfold. A total of 1070 structures of each RNA molecule were generated from which the structure with the lowest-energy (most stable structure) which was chosen to be used in subsequent docking calculations.

Energy Minimization of Lowest-Energy Generated Three Dimensional RNA Structures:

Before using the generated three-dimensional RNA structure in subsequent HnRNP C docking simulations, energy minimizations were conducted using UCSF Chimera using 100 steepest descent steps with a steepest descent step size of 0.02 Å followed by 10 conjugated gradient steps with a conjugated gradient step size of 0.02 Å. The energy minimization was first preceded by AddH and Add Charge from the Structure Editing Module that can be accessed from the Tools Pulldown menu of Chimera. The energy minimization module can also be accessed from Chimera's Pull-down menu. The energy minimized RNA structures were saved using the PDB format for subsequent docking calculations.

Virtual Screening of RNA-Binding Site in HnRNP C RRM Domain:

Virtual screening of RNA-HnRNP C RRM binding conformation and binding affinity was conducted using Autodock Vina which is used to conduct virtual screening of large libraries of ligands. Vina uses a docking program that uses the Lamarckian genetic algorithm search

method, which performs a combination of global and local searches in optimizing its docking results. Docking results at every stage are evaluated by determining the free energy of binding of the ligand to the receptor which take into the equation:

$$\Delta G_{\text{binding}} = \Delta G_{\text{vdW}} + \Delta G_{\text{elec}} + \Delta G_{\text{Hbond}} + \Delta G_{\text{desolv}} + \Delta G_{\text{tor}}$$

(ΔG_{vdW}) van der Waals interaction free energy, ($\Delta G_{\text{H-bond}}$) hydrogen-bonding free energy, (ΔG_{elec}) electrostatic interaction free energy, (ΔG_{desolv}) desolvation of ligand in the receptor environment, and (ΔG_{tor}) torsional free energy which represents the change in free energy as the ligand goes from unbound to bound state. In conducting the virtual docking of the RNA molecules developed in this work to the RRM domain of HnRNP C, the receptor was kept as a rigid while the ligand was allowed to be flexible around the rotatable bond (torsions). In setting up the docking protocol a grid box was determined by loosely assigning XYZ coordinates that encompass the binding site of the RRM domain of HnRNP C as previously determined experimentally using NMR structural studies. Furthermore, Autodock Vina requires a structure file format change for both ligand and receptor from PDB format which only contains atomic coordinates to PDBQT format which contains atomic coordinates, partial charges and atom types in one file. The docking was carried out using the Virtual Screening tool, PyRx, which integrates, Autodock tools, the software needed to assign the binding grid box, the torsion angles that determine degrees of freedom used in the flexible ligand docking, the structure file format change from PDB to PDBQT and the vina calculations, all bundled in one package. Once PyRx is launched, local or remote execution mode is selected, followed by selecting "start" option. Multiple ligand structure files are then loaded into the PyRx workspace using the "Add Ligand(s)" option, and the receptor structure file was also loaded using the "Add Macromolecule(s)" option. This is followed by selecting the "Forward" option. At this point the structure file format for both the ligands and the macromolecule are converted from the original PDB format to the PDBQT format and a grid box window appears allowing the assignment of the coordinates of the center and dimensions of the grid box that define the predicted binding site. This is followed by selecting the "Forward"

option which launched vina's calculations. Once completed, a table summarizing the binding affinities of the top 10 complexes of each ligand to the receptor is shown. All data was stored under the PyRx workspace the location of which can be assigned from preferences under the Edit pulldown menu.

Design the 3D structures by using AutoDock:

After docking the ligands to Macromolecule by using vina, we chose the top 10 structures which have the lowest free energy that analyze vina result. AutoDock Tools used to analyze the docking results. This was done by first choosing the Docking selection from the Docking pulldown menu. The docked structures were accessed using the Open option and by choosing Single Molecule from the Load Models window. Next, under analyze → molecules and select RRM_C that the structure of amino acids which docked. Click on interaction from display pulldown menu. A window to specify two group of nodes for interactions display will pop up. Choose the molecule list from the first molecule option that is the ligand and in the second molecule option, click on the receptor, which is RRM_C then OK. Therefore, three windows will open, one of them to set hydrogen bonds, the other two interactions display options and setback Ground Color. Finally, save the Image.

CHAPTER III

Results and Discussion:

Efficacy of Auto Dock in docking RNA molecules to RRM domain of hnRNP C protein:

The crystal structure of the RNA recognition motif (RRM) of Heterogeneous Nuclear Ribonucleoprotein (hnRNP C) bound to AUUUUUC. AUUUUUC, a sequence identified by SELEX to have the highest binding affinity to the RRM region of hnRNP C used to test the efficacy of AutoDock in predicting the binding specificity and to evaluate whether the RRM could distinguish between AUUUUUC in a primary structural conformation as opposed to the sequence being presented in a secondary structural context. This done by first removing AUUUUUC from the protein of the crystal structure. The structures formatted for AutoDock use and a Grid Box that loosely defined the general search space for the program to use as an initial step in the calculations. Once prepared, an AutoDock configuration file was generate which specified information such as the search algorithm to be used in the calculations and the energy minimization scheme to be conducted after the docking is complete. The top docked structure we obtained using AutoDock has a binding affinity, K_{eq} , of 3×10^7 (figure1). The conformation of the RNA molecule after docking (Panel B) remained fairly identical to that found in the crystal structure (Panel A), also the binding site of the docked RNA closely matches that found in the crystal structure. This can be seen by comparing the H-bonding interactions observed between the RNA molecule and the RRM amino acids in the co-crystal structure and the Docked RNA (Tables 1 and 2, respectively). We observed 9 hydrogen-bonds in both structures and that the same residues on the RNA and protein are involved in the formation of the hydrogen bonds.

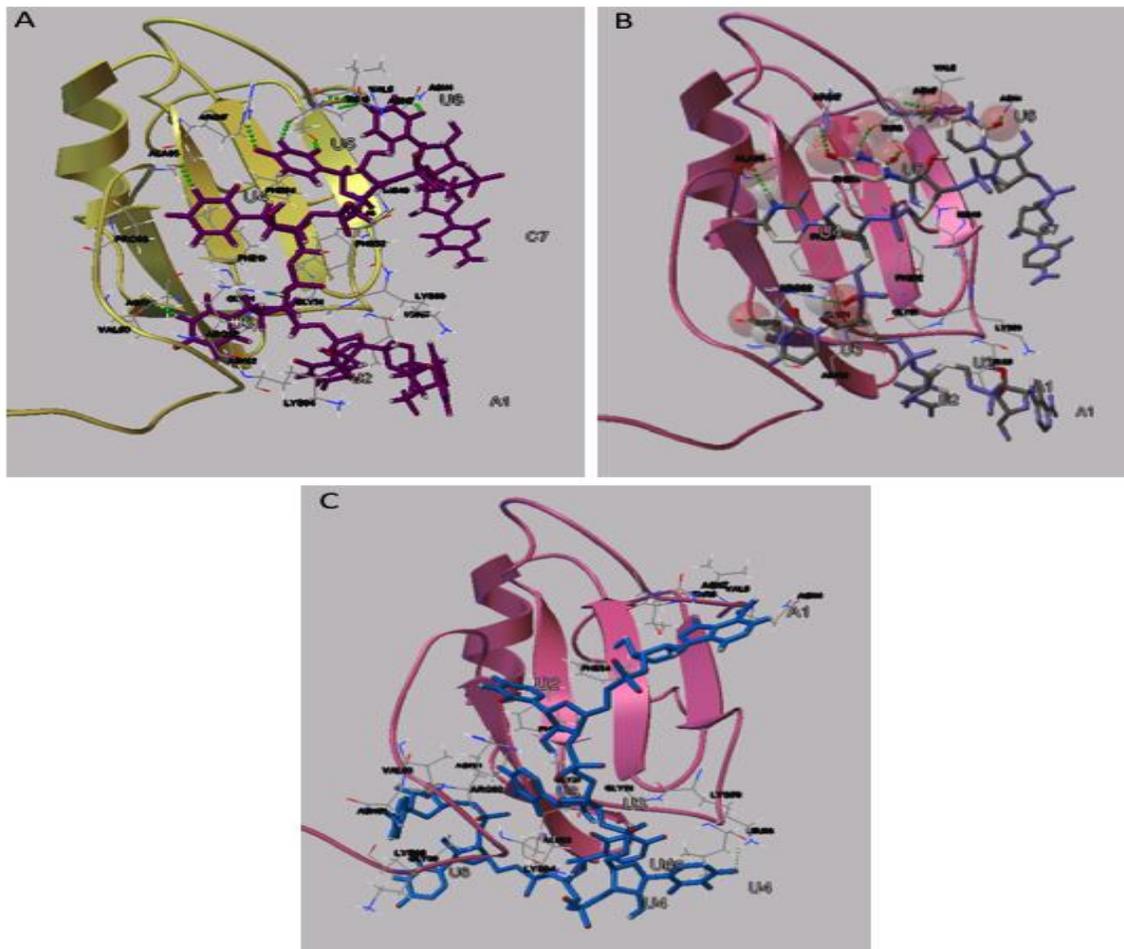


Figure1: The structures of RRM Bound to the SELEX Generated RNA Sequence, AUUUUUC. Panel A is the co-crystal structure of RRM with AUUUUUC, Panel B is the Docked structure of RRM with the linear form of AUUUUUC. Panel C is the Docked structure of RRM with the kinked AUUUUUC.

It is worth noting that the RNA molecule used to generate the docking results shown in Figure 1B is obtained as mentioned above by simply removing the RNA found in the co-crystal structure and was used as is in the docking procedure without any conformational modifications. To check whether doing so biased the results of the docking, we generated the AUUUUUC sequence using our RNA sequence generator, the sequence was then submitted to Rosetta Docking in order to generate the structure file (PDB) to be docked to the RRM. The resulting PDB structure of the RNA was surprising in that this RNA has a secondary structural element (a loop-like structure)

that resulted from a kink in the structure due to a π - π stacking between the U4 and U5 nucleotide rings (Figure 1C). We decided to still go ahead and dock the kinked RNA structure (K-RNA) to the RRM domain to see if the structural alteration changes the outcome of the docking procedure. In addition, as we suspected, the two docked structures were very different. Unlike the docked structure of the linear sequence, the RRM –K-RNA complex looked drastically different from the co-crystal structure. The hydrogen-bonding interactions in the latter were limited to two, instead of the nine observed in both the co-crystal structure and the docked RRM-linear RNA structure (Table3). Moreover, the binding affinity of the RRM-K-RNA was 10 fold lower than observed for the docked complex between the RRM and the linear AUUUUUC. The lowering in the binding affinity of the RRM-K-RNA complex as compared to the RRM-linear complex can be explained by noting that all nine hydrogen bonds between the RRM domain and the linear RNA involve nucleotides uracil 3-uracil 6. Those interactions were nullified in the kinked RNA the π - π stacking between the U4 and U5 nucleotide rings making most of the uracil inaccessible for hydrogen bonding to the RRM moiety.

Table 1: H-Bonding interactions between amino acids in the RRM domain and the nucleotides of the RNA sequence, AUUUUUC found in the co-crystal structure.

| RRM | Crystal Structure AUUUUUC | # H-bond |
|--------------|---------------------------|----------|
| GLY21, ARG92 | U3 | 2 |
| ASP81, ALA85 | U4 | 2 |
| TH6, ARG17 | U5 | 3 |
| ASN4, TH6 | U6 | 2 |

Table 2: H-Bonding interactions between amino acids in the RRM domain and the nucleotides of the RNA sequence, AUUUUUC found in the Docked structure.

| RRM | Docked Structure AUUUUUC | # H-bond |
|--------------|-----------------------------|----------|
| GLY21, ARG92 | U3 | 2 |
| ASP81, ALA85 | U4 | 2 |
| TH6, ARG17 | U5 | 3 |
| ASN4, TH6 | U6 | 2 |

Table 3: H-Bonding interactions between amino acids in the RRM domain and the nucleotides of the RNA sequence, AUUUUUC found in the co-crystal structure.

| RRM | Docked Structure K-RNA | # H-bond |
|-------|---------------------------|----------|
| VAL5 | A1 | 1 |
| LYS50 | U4 | 1 |

This observation addresses a heated debate among scientists that study RNA-protein interactions, which is whether the protein identifies the primary RNA sequence, or the secondary and tertiary structure of the RNA. The results observed here have a huge impact on the RNA-protein interaction debate in that in this case where we have two identical sequences, one is linear where the other possesses secondary structural elements and the protein favors the linear structure. Our results suggest that the protein identifies the structure of the RNA as well as its sequence. However, we should point out that previous experiments conducted with RRMs from other proteins have identified high affinity targets that do have secondary structure ^[23]. This will complicate our ability to use computerized selection in that it will necessitate the ability to generate libraries that contain randomized RNA molecules in both primary and secondary structural format. Currently, the scientists that are building these libraries using SELEX have focused on the hypothesis that only RNA secondary structures are relevant to protein binding. If our observation regarding structural preference being protein specific then we must have the

ability to generate randomized sequences that are both linear and structured. We believe that with the advances in computer calculation power and technologies and structural calculations protocols, this would be easier to accomplish than through experimental techniques such as SELEX.

To continue with the project, we generated well over 1000 of 7-nucleotides RNA structures. The reason we limited the size of the RNA molecules to 7-nucleotides is two-fold. Originally, it is because we wanted to see if we could pick out the AUUUUUC from our pool of RNA molecule to be the one with the highest affinity to the RRM domain of HnrNPC as determined by SELEX. Our second reason is because conducting molecular simulations on large RNA molecules would be daunting specially since our docking protocol allows the ligand (RNA molecule) flexibility around its many backbone torsion angles (47 for a 7 nucleotides sequence) while keeping a rigid protein structure. As an example, the calculations using a 7-nucleotides RNA took 30 minutes to complete on our fastest computer and that out docking results to only 100 RNA-RRM structures. Of the one hundred structures, we analyzed the structural conformation of the top 10 RNA molecules that displayed the highest binding affinity to the RRM domain. It is worth noting that in the pool of RNA molecules that we tested, none showed a binding affinity as high or higher than that obtained for the SELEX determined RNA aptamer, AUUUUUC (binding affinity of 3×10^7) (Table4). This observation gives further validity to our docking method. In comparison, the canonical RNA aptamer, AUUUUUC, using our docking study was determined to bind to HNRNPC with a binding constant of 3×10^7 , which is a 10-fold increase in binding affinity over RNA144, which has the highest binding affinity in our list of 10 aptamers.

Table 4: The RNA sequences that resulted in the top 10 RRM-binding affinities.

| Label of the RNA Molecule | RNA sequence | K_{eq} | ΔG values (kcal/mol) |
|---------------------------|--------------|-----------------|------------------------------|
| Rna9 | CCGAUCU | 2×10^6 | -8.5 |
| Rna16 | UACGUCC | 1×10^6 | -8.1 |
| Rna94 | UGUCGGU | 1×10^6 | -8.1 |
| Rna102 | UUUAAGU | 2×10^6 | -8.5 |
| Rna104 | GAUGCCC | 2×10^6 | -8.5 |
| Rna105 | GGGAGGU | 1×10^6 | -8.1 |
| Rna106 | UAAUAGC | 9×10^5 | -8.1 |
| Rna125 | GAAACAU | 3×10^6 | -8.8 |
| Rna137 | CAUUUCA | 3×10^6 | -8.8 |
| Rna144 | UAUUUUG | 3×10^6 | -8.8 |
| SELEX RNA | AUUUUUC | 3×10^7 | -10.1 |

The resulting ten RNA molecule structures each had their own docking structure and binding affinity (Table 2, Figures 2 - 11). These are shown to emphasize that none of the RNA molecules have the linear structure shown to be favored by the RRM domain of HnRNP C. We analyzed the docked structures of the top 10 RNA molecules for relevant hydrogen-bonding interactions between the RNA molecules and the RRM domain. Those interactions are summarized in tables 4-14. Also given in the tables are nucleotide-nucleotide interactions that gave rise to the observed secondary structures of the RNA molecules.

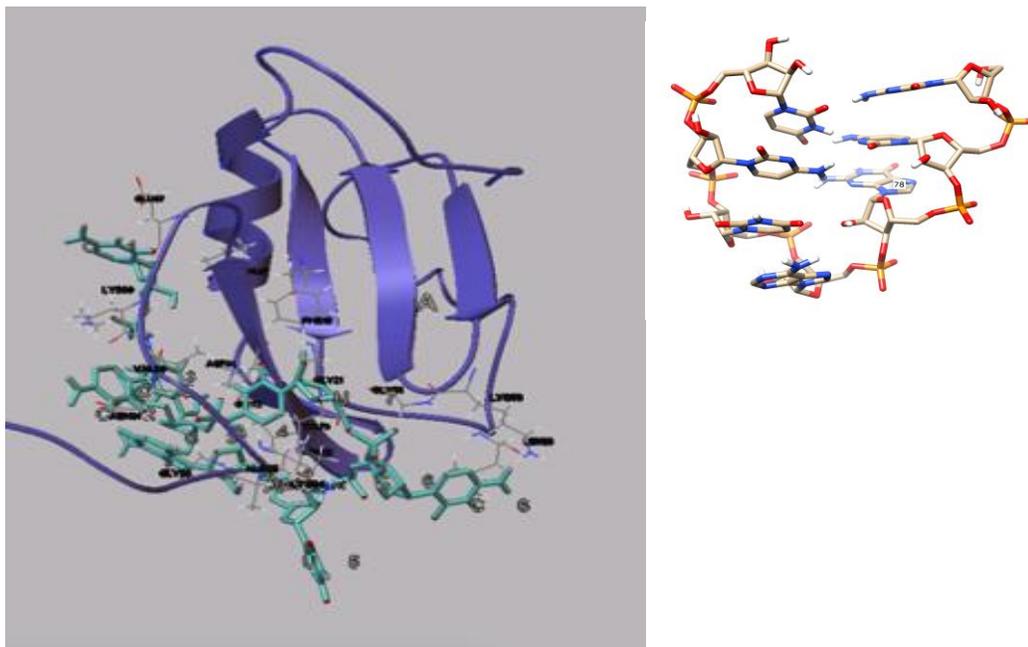


Figure 2: The Molecular Docking Structure of the RNA 9 sequence bound to RRM domain of HnRNP C (Panel A). RNA9 adopts a secondary structure in this docking protocol (Panel B).

Table 5: The Hydrogen Bonds observed between RNA 9 Nucleotides and the RRM Domain of HnRNPC. In addition, included are Nucleotide-Nucleotide Hydrogen Bonds observed as a result of the formation of Nucleic Acid Secondary Structure.

| Amino Acid | Nucleotide | Nucleotide |
|------------|------------|------------|
| LYS89 | C2 | - |
| VAL79 | U5 | - |
| ASN22 | U5 | - |
| ASN22 | U7 | - |
| ASP81 | U7 | - |
| ALA95 | A4 | - |
| - | G3 | C7 |
| - | U5 | U5 |
| - | C6 | U7 |
| - | U2 | U7 |
| - | U3 | C6 |
| - | U5 | U6 |
| - | C7 | G3 |

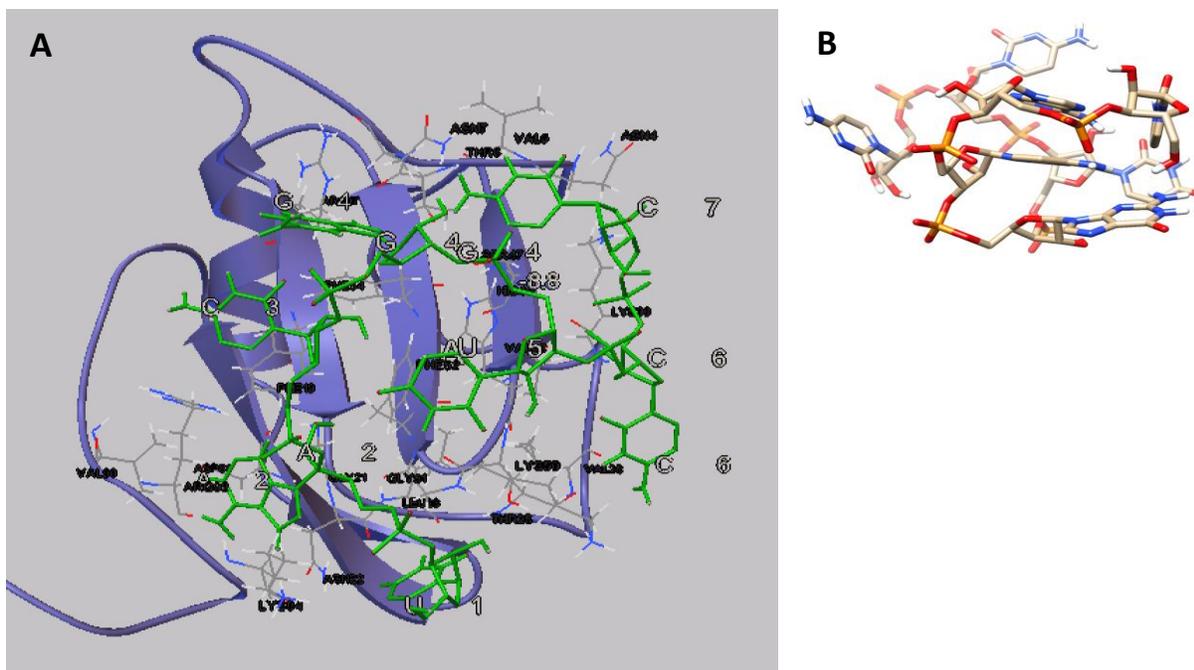


Figure 3: The Molecular Docking Structure of the RNA 16 sequence bound to RRM domain of HnRNP C (Panel A). Also shown is the secondary structure that RNA16 adopts in this docking protocol (Panel B).

Table 6: The Hydrogen Bonds observed between RNA 16 Nucleotides and the RRM Domain of HnRNP C. In addition, included are Nucleotide-Nucleotide Hydrogen Bonds observed as a result of the formation of Nucleic Acid Secondary Structure.

| Amino Acid | Nucleotide | Nucleotide |
|------------|------------|------------|
| LEU23 | U1 | - |
| ASP81 | A2 | - |
| ARG17 | G4 | - |
| LYS50 | U5 | - |
| THR25 | C6 | - |
| - | U5 | U5 |
| - | G4 | U2 |
| - | U2 | A2 |
| - | A1 | G4 |
| - | U3 | A2 |
| - | U5 | U6 |

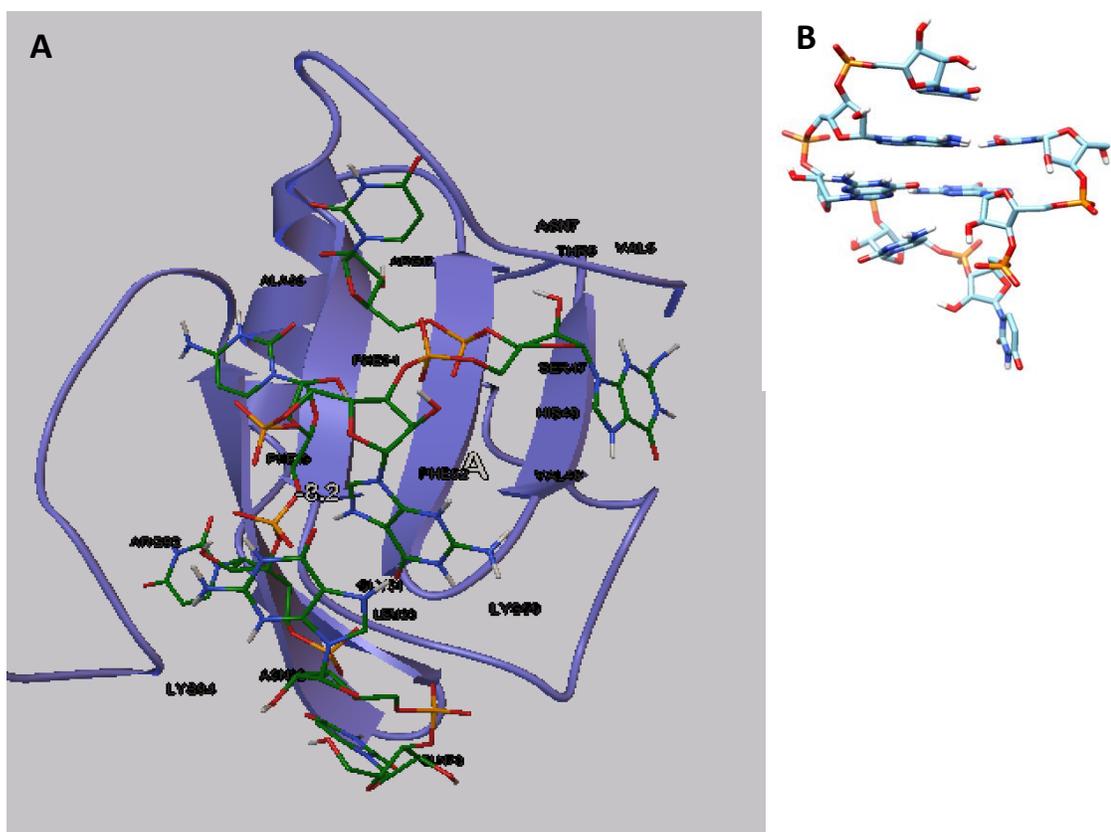


Figure 4: The Molecular Docking Structure of the RNA 94 sequence bound to RRM domain of HnRNP C (Panel A). Also shown is the secondary structure that RNA94 adopts in this docking protocol (Panel B).

Table 7: The Hydrogen Bonds observed between RNA 94 Nucleotides and the RRM Domain of HnRNPC. In addition, included are Nucleotide-Nucleotide Hydrogen Bonds observed as a result of the formation of Nucleic Acid Secondary Structure.

| Amino Acid | Nucleotide | Nucleotide |
|------------|------------|------------|
| LYS50 | G2 | - |
| ASP81 | U3 | - |
| ASN22 | U3 | - |
| ARG92 | C4 | - |
| - | G5 | C4 |
| - | U3 | C4 |
| - | U1 | U5 |
| - | U5 | U6 |
| - | U3 | U3 |

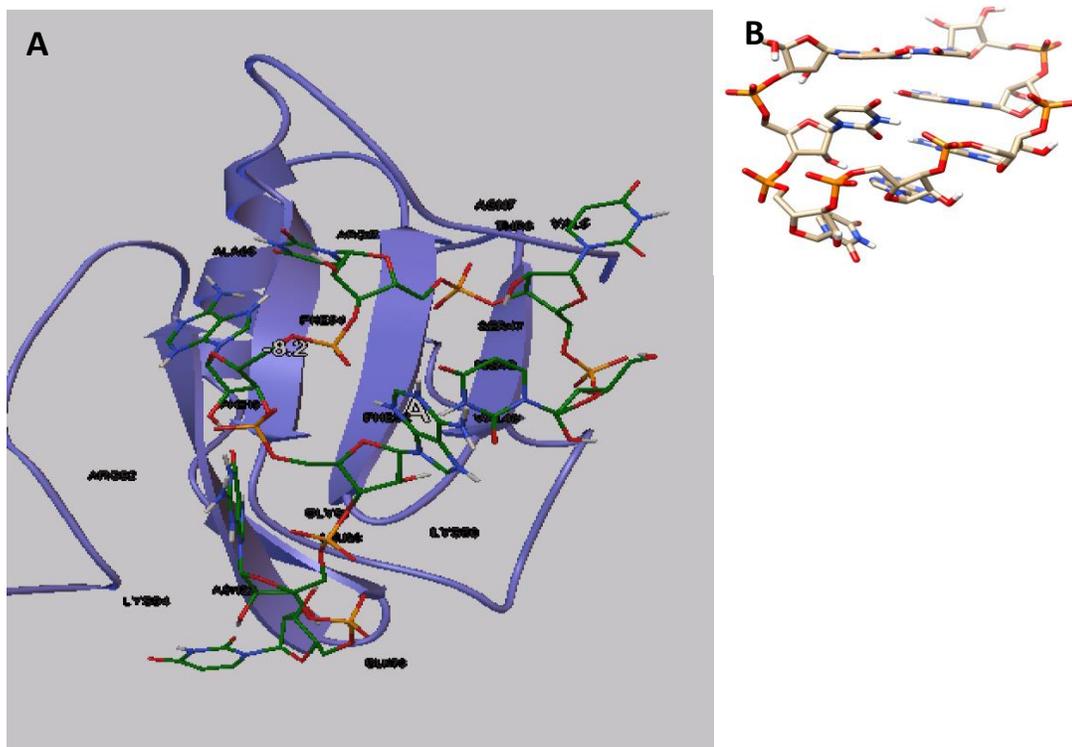


Figure 5: The Molecular Docking Structure of the RNA 102 sequence bound to RRM domain of HnRNP C (Panel A). Also shown is the secondary structure that RNA102 adopts in this docking protocol (Panel B).

Table 8: The Hydrogen Bonds observed between RNA 102 Nucleotides and the RRM Domain of HnRNPC. In addition, included are Nucleotide-Nucleotide Hydrogen Bonds observed as a result of the formation of Nucleic Acid Secondary Structure.

| Amino Acid | Nucleotide | Nucleotide |
|------------|------------|------------|
| VAL48 | U2 | - |
| ARG17 | U3 | - |
| ARG92 | A4 | - |
| ALA85 | A4 | - |
| LYS50 | A5 | - |
| LEU23 | U7 | - |
| ASN22 | U7 | - |
| - | G6 | A5 |
| - | U6 | A4 |
| - | U7 | U6 |
| - | U1 | U2 |
| - | A4 | U3 |
| - | U2 | G6 |
| - | A1 | U3 |
| - | U5 | U6 |

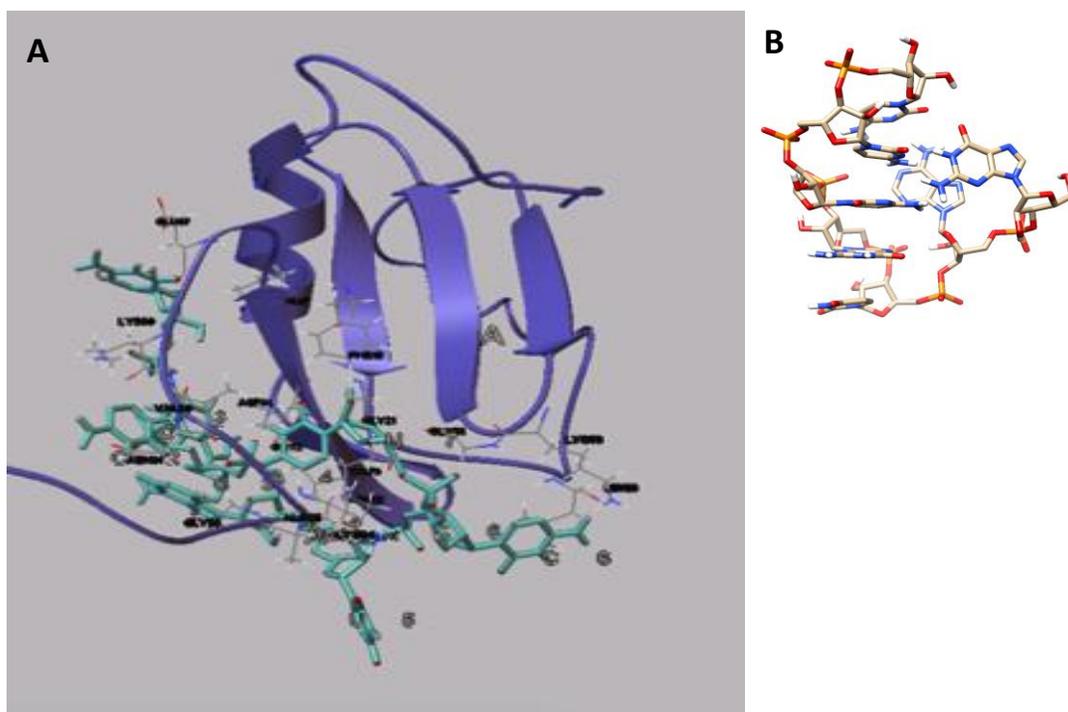


Figure 6: The Molecular Docking Structure of the RNA104 sequence bound to RRM domain of HnRNP C (Panel A). Also shown is the secondary structure that RNA104 adopts in this docking protocol (Panel B).

Table 9: The Hydrogen Bonds observed between RNA 104 Nucleotides and the RRM Domain of HnRNPC. In addition, included are Nucleotide-Nucleotide Hydrogen Bonds observed as a result of the formation of Nucleic Acid Secondary Structure.

| Amino Acid | Nucleotide | Nucleotide |
|------------|------------|------------|
| ASP81 | A2 | - |
| THR25 | C5 | - |
| ASN4 | C5 | - |
| ASN4 | C7 | - |
| - | G1 | U3 |
| - | C7 | A1 |
| - | G4 | G4 |
| - | G4 | U2 |
| - | U3 | G1 |
| - | U5 | U6 |
| - | G1 | U3 |

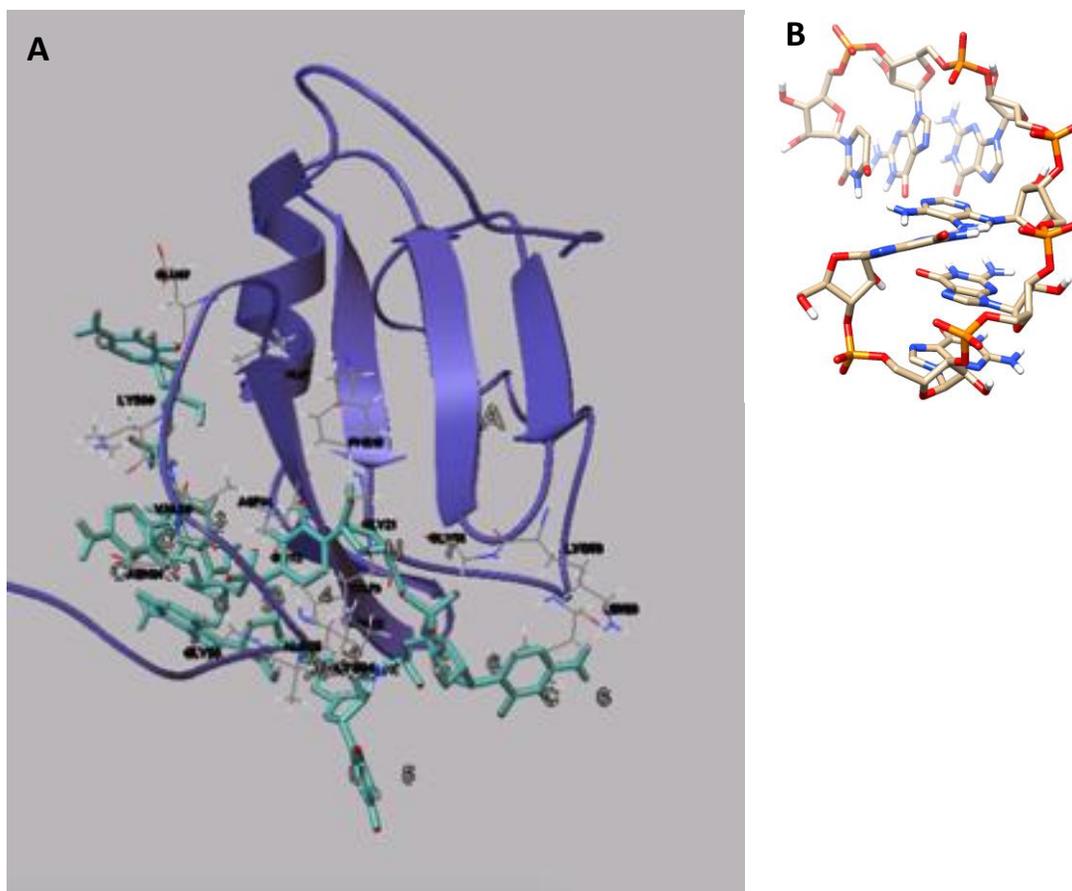


Figure 7: The Molecular Docking Structure of the RNA 105 sequence bound to RRM domain of HnRNP C (Panel A). Also shown is the secondary structure that RNA105 adopts in this docking protocol (Panel B).

Table10: The Hydrogen Bonds observed between RNA 105 Nucleotides and the RRM Domain of HnRNPC. In addition, included are Nucleotide-Nucleotide Hydrogen Bonds observed as a result of the formation of Nucleic Acid Secondary Structure.

| Amino Acid | Nucleotide | Nucleotide |
|------------|------------|------------|
| ASN22 | A4 | - |
| ASP71 | G5 | - |
| Serine100 | G6 | - |
| GLY96 | G6 | - |
| GLU87 | U7 | - |
| - | G2 | U4 |
| - | G2 | U5 |
| - | G5 | A4 |
| - | G6 | C7 |
| - | A4 | G3 |
| - | U3 | G2 |

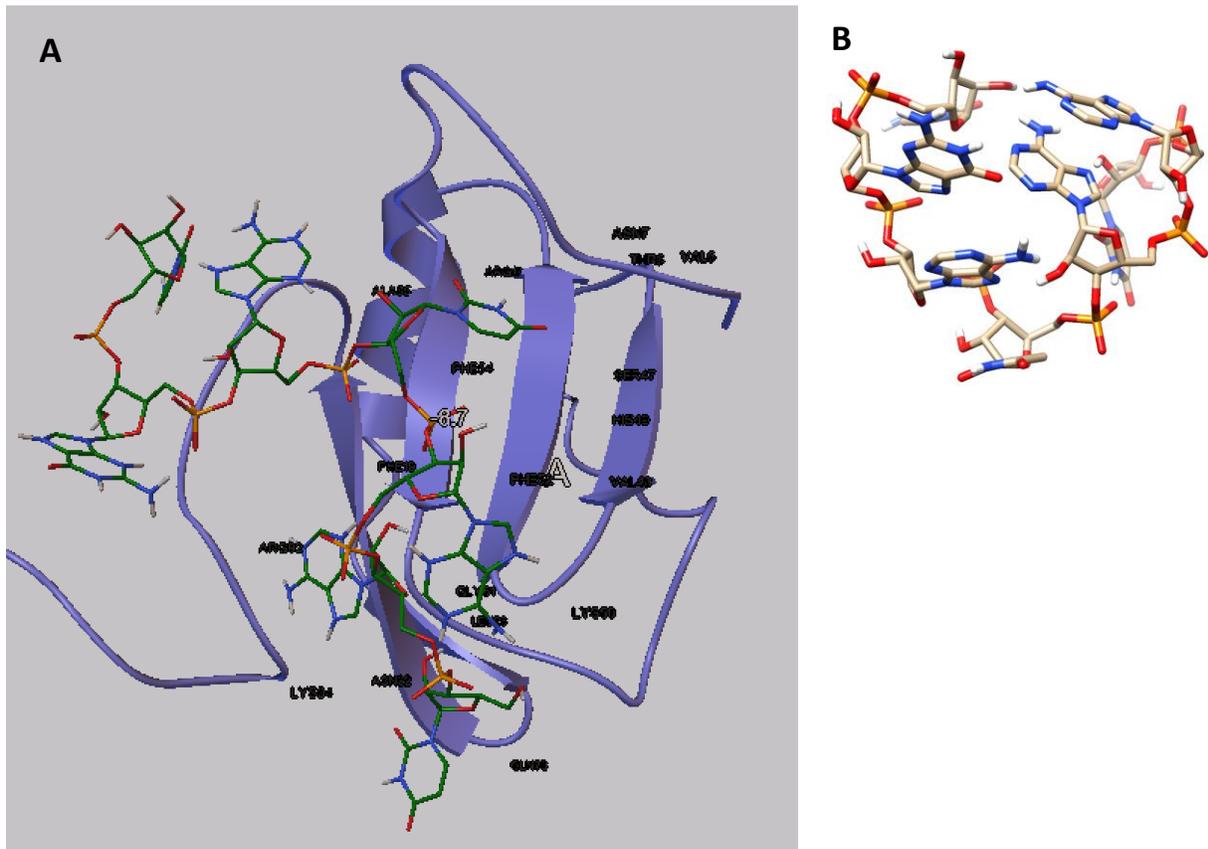


Figure 8: The Molecular Docking Structure of the RNA 106 sequence bound to RRM domain of HnRNP C (Panel A). Also shown is the secondary structure that RNA106 adopts in this docking protocol (Panel B).

Table 11: The Hydrogen Bonds observed between RNA 106 Nucleotides and the RRM Domain of HnRNPC. In addition, included are Nucleotide-Nucleotide Hydrogen Bonds observed as a result of the formation of Nucleic Acid Secondary Structure.

| Amino acid | Nucleotide | Nucleotide |
|------------|------------|------------|
| ASP81 | A2 | - |
| THR6 | U4 | - |
| ARG17 | U4 | - |
| ARG92 | A5 | - |
| ASN91 | G6 | - |
| - | A5 | C4 |
| - | A3 | C4 |
| - | U2 | A2 |
| - | U3 | U1 |
| - | U5 | U6 |

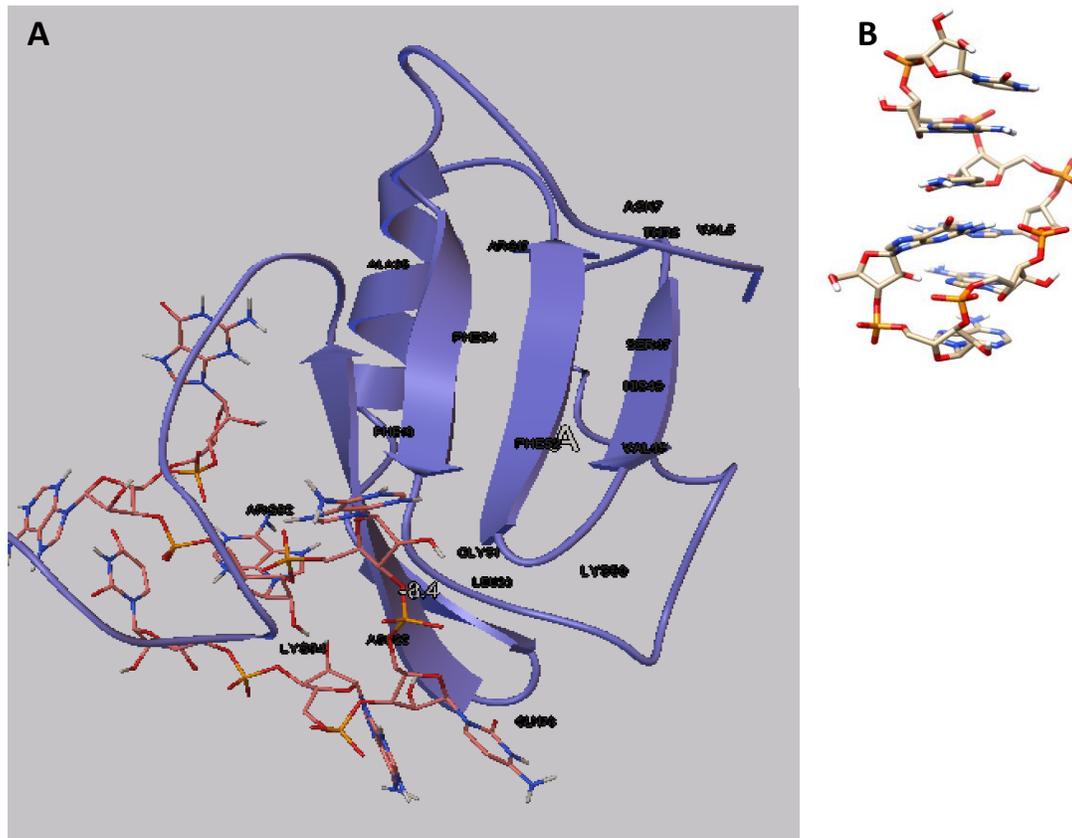


Figure 9: The Molecular Docking Structure of the RNA125 sequence bound to RRM domain of HnRNP C (Panel A). Also shown is the secondary structure that RNA125 adopts in this docking protocol (Panel B).

Table 12: The Hydrogen Bonds observed between RNA 125 Nucleotides and the RRM Domain of HnRNPC. In addition, included are Nucleotide-Nucleotide Hydrogen Bonds observed as a result of the formation of Nucleic Acid Secondary Structure.

| Amino acids of RRM domain | Nucleotide | Nucleotide |
|---------------------------|------------|------------|
| GLU87 | G1 | - |
| ASP71 | A3 | - |
| ASN22 | A6 | - |
| - | G1 | A2 |
| - | A3 | U6 |
| - | A6 | A3 |
| - | U3 | C5 |
| - | U5 | A6 |
| - | U6 | U7 |

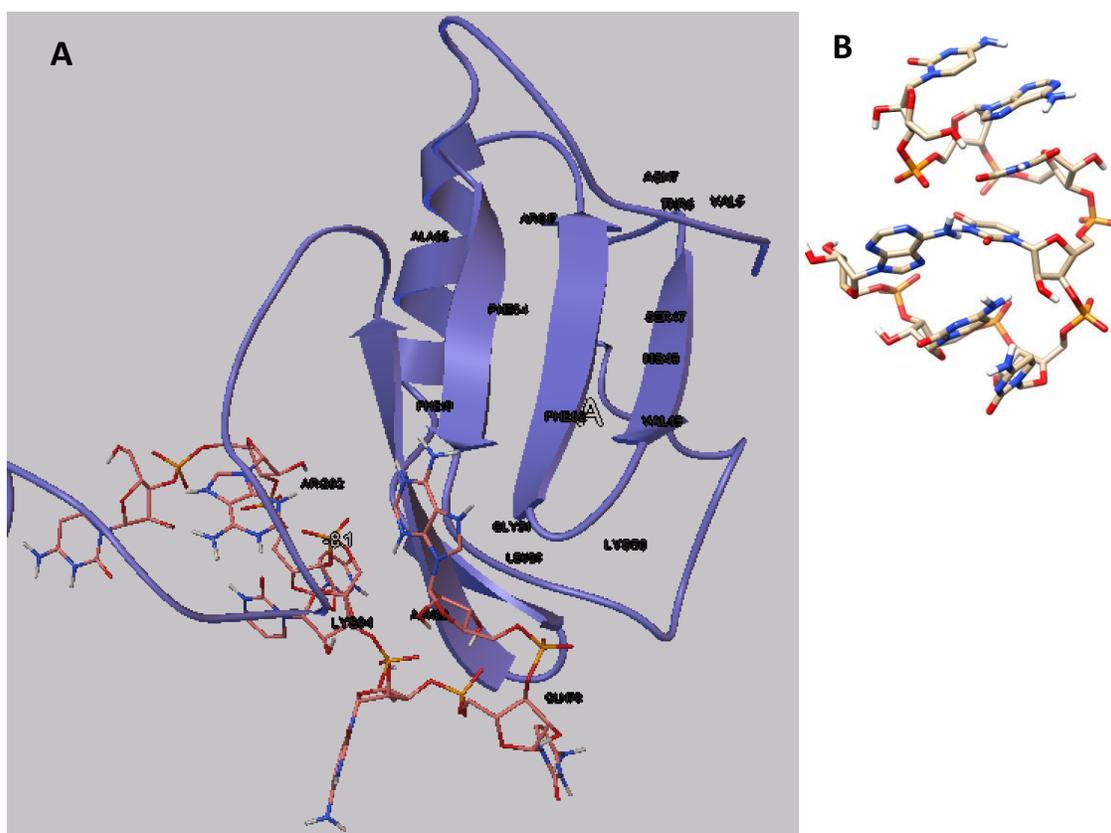


Figure 10: The Molecular Docking Structure of the RNA 137 sequence bound to RRM domain of HnRNP C (Panel A). Also shown is the secondary structure that RNA137 adopts in this docking protocol (Panel B).

Table 13: The Hydrogen Bonds observed between RNA 137 Nucleotides and the RRM Domain of HnRNPC. In addition, included are Nucleotide-Nucleotide Hydrogen Bonds observed as a result of the formation of Nucleic Acid Secondary Structure.

| Amino acids | Nucleotide | Nucleotide |
|-------------|------------|------------|
| ARG99 | C1 | - |
| GLY96 | A2 | - |
| ASP81 | A2 | - |
| VAL79 | U5 | - |
| - | A7 | U3 |
| - | A7 | U2 |
| - | U3 | A7 |
| - | U5 | U6 |
| - | C7 | U3 |

Upon careful examination of these figures and the associated tables, it observed that each RNA molecule has an extensive secondary structure. This explains why this group of RNA molecules, which were picked to have the highest affinity to the RRM motif, fail to match the affinity of the linear AUUUUUC structure. This supports our early conclusion that structural specificity of the RNA molecule is more important to the protein's recognition of the RNA molecule than the sequence itself.

In this project, we set out to develop an RNA aptamer library that can be used to find a RNA molecule that binds tightly to any biological target. Our original scheme was to generate a random pool of RNA molecules base on the assumption that when we conduct a virtual screening using a specific protein, the protein will then bind its preferred RNA sequence. However, in light of significant results this project yielded, our random pool of RNA aptamers must also contain randomized conformations as well.

Conclusion:

In this work we were able to devise an in silico method that enabled the development of an RNA aptamer library consisting of more than 5000 7-nucleotide RNA molecules of random sequences, using homology modeling and molecular simulations techniques we were able to generate coordinate files (PDB) for each RNA molecule in this pool. Our goal was to be able to use this pool as a database of ligands for the selection of a top binding RNA to any cellular target. We conducted a test screening using the RRM domain of HnRNP C as the cellular target. We used this particular protein because its SELEX RNA aptamer is known and we can use it as a check of the success of our in silico method. We were able to pick out the SELEX RNA sequence as the RNA sequence that has the highest affinity to RRM domain of HnRNP C. We also determined that for this sequence to bind tightly to the RRM domain, it much be presented to the protein as a linear sequence. When we subjected this RNA sequence to a molecular simulation program that used comparative modeling to determine the secondary structure of nucleic acids,

the resulting outcome shows an RNA with a secondary structure that involves interactions between its nucleotides. When this secondary structure was docked to the RRM domain, the resulting structure was drastically different from the co-crystal structure and the binding affinity of RNA molecule was 10-fold lower than that observed for the linear RNA. Based on these results, it was concluded that in order for us to be able to develop a library of RNA aptamers to be used as a database for the selection of sequences that binding specific targets. We not only have to have random sequence of RNA but our library pool must also consist of various structural forms of the RNA sequences. This task is involving but with the help of more computing power and already existing molecular dynamics software, we believe that an RNA library that contains randomized sequence and randomized structures is feasible.

REFERENCES:

- 1- Gold, L., Janjic, N., Jarvis, T., Schneider, D., Walker, J. J., Wilcox, S. K., & Zichi, D. (2012). Aptamers and the RNA world, past and present. *Cold Spring Harbor perspectives in biology*, 4(3), a003582.
- 2- Germer, K., Leonard, M., & Zhang, X. (2013). RNA aptamers and their therapeutic and diagnostic applications. *International journal of biochemistry and molecular biology*, 4(1), 27.
- 3- Tsao, S. M., Lai, J. C., Horng, H. E., Liu, T. C., & Hong, C. Y. (2017). Generation of aptamers from a primer-free randomized ssDNA library using magnetic-assisted rapid aptamer selection. *Scientific reports*, 7, 45478.
- 4- Brockstedt, U., Uzarowska, A., Montpetit, A., Pfau, W., & Labuda, D. (2004). In vitro evolution of RNA aptamers recognizing carcinogenic aromatic amines. *Biochemical and Biophysical Research Communications*, 313(4), 1004-1008.
- 5- Santulli-Marotto, S., Nair, S. K., Rusconi, C., Sullenger, B., & Gilboa, E. (2003). Multivalent RNA aptamers that inhibit CTLA-4 and enhance tumor immunity. *Cancer research*, 63(21), 7483-7489.
- 6- Khaled, A., Guo, S., Li, F., & Guo, P. (2005). Controllable self-assembly of nanoparticles for specific delivery of multiple therapeutic molecules to cancer cells using RNA nanotechnology. *Nano letters*, 5(9), 1797-1808.
- 7- Lee, J. H., Canny, M. D., De Erkenez, A., Krilleke, D., Ng, Y. S., Shima, D. T., ... & Jucker, F. (2005). A therapeutic aptamer inhibits angiogenesis by specifically targeting the heparin binding domain of VEGF165. *Proceedings of the National Academy of Sciences*, 102(52), 18902-18907.
- 8- Gupta, N., Mansoor, S., Sharma, A., Sapkal, A., Sheth, J., Falatoonzadeh, P., ... & Kenney, M. C. (2013). Diabetic retinopathy and VEGF. *The open ophthalmology journal*, 7, 4.

- 9- Catuogno, S., & Esposito, C. L. (2017). Aptamer cell-based selection: Overview and advances. *Biomedicines*, 5(3), 49.
- 10- Zhuo, Z., Yu, Y., Wang, M., Li, J., Zhang, Z., Liu, J., ... & Zhang, B. (2017). Recent advances in SELEX technology and aptamer applications in biomedicine. *International journal of molecular sciences*, 18(10), 2142.
- 11- Piva, F., & Principato, G. (2006). RANDNA: a random DNA sequence generator. *In silico biology*, 6(3), 253-258.
- 12- Lorenz, R., Bernhart, S. H., Zu Siederdissen, C. H., Tafer, H., Flamm, C., Stadler, P. F., & Hofacker, I. L. (2011). ViennaRNA Package 2.0. *Algorithms for Molecular Biology*, 6(1), 26.
- 13- Watkins, A. M.; Das, R. "An automated and customizable RNA fragment assembly protocol in Rosetta. *bioRxiv* 223305; Lyskov, S., Chou, F. C., Conchúir, S. Ó., Der, B. S., Drew, K., Kuroda, D., ... & Borgo, B. (2013). Serverification of molecular modeling applications: the Rosetta Online Server that Includes Everyone (ROSIE). *PloS one*, 8(5), e63906.
- 14- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry*, 25(13), 1605-1612.
- 15- Krieger, E., & Vriend, G. (2014). YASARA View—molecular graphics for all devices—from smartphones to workstations. *Bioinformatics*, 30(20), 2981-2982.
- 16- Trott, O., & Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry*, 31(2), 455-461.
- 17- Small-molecule library screening by docking with PyRx. *Dallakyan S, Olson AJ Methods Mol Biol*. 2015; 1263():243-50.
- 18- Schames, J. R., Henchman, R. H., Siegel, J. S., Sotriffer, C. A., Ni, H., & McCammon, J. A. (2004). Discovery of a novel binding trench in HIV integrase. *Journal of medicinal chemistry*, 47(8), 1879-1881.

- 19- Forli, S., Huey, R., Pique, M. E., Sanner, M. F., Goodsell, D. S., & Olson, A. J. (2016). Computational protein–ligand docking and virtual drug screening with the AutoDock suite. *Nature protocols*, *11*(5), 905.
- 20- Huey, R., & Morris, G. M. (2003). AutoDock tools. *La Jolla, CA, USA: The Scripps Research Institute*.
- 21- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., ... & Bourne, P. E. (2000). The protein data bank. *Nucleic acids research*, *28*(1), 235-242.
- 22- Chushak, Y., & Stone, M. O. (2009). In silico selection of RNA aptamers. *Nucleic acids research*, *37*(12), e87-e87.
- 23- Coburn, K., Melville, Z., Aligholizadeh, E., Roth, B. M., Varney, K. M., Carrier, F., ... & Weber, D. J. (2017). Crystal structure of the human heterogeneous ribonucleoprotein A18 RNA-recognition motif. *Acta Crystallographica Section F: Structural Biology Communications*, *73*(4), 209-214.