



DEMOCRITUS UNIVERSITY OF THRACE
SCHOOL OF HEALTH SCIENCES
DEPT. OF MOLECULAR BIOLOGY AND
GENETICS

Homology modeling of the complex between the Repressor of Primer protein and its cognate RNA kissing complex

Bachelor Thesis

Drosoula Nikolopoulou

Supervisor:

Dr. Nicholas M. Glykos

Structural and Computational Biology Lab

Alexandroupolis, Thrace, Greece

October 2022

Table of contents

Table of contents	1
Abstract	2
<i>Key words</i>	3
1. Introduction	4
1.1 <i>ROP - RNAI-RNAlI kissing loop complex in nature</i>	4
1.2 <i>ROP protein</i>	5
1.3 <i>RNA kissing complex</i>	7
1.4 <i>ROP - RNAI-RNAlI complex modeling</i>	9
2. Materials and Methods	11
2.1 <i>ROP protein models selection</i>	12
2.2 <i>Docking preparation</i>	17
2.3 <i>Docking principles</i>	19
2.4 <i>Docking Methodology</i>	21
2.4.1 <i>HDOCK program</i>	21
2.4.2 <i>Scoring methods</i>	26
2.4.3 <i>Additional Test Cases</i>	27
2.5 <i>Docking results evaluation</i>	32
3. Results	33
3.1 <i>Metrics</i>	38
3.1.1 <i>ΔiG P-value</i>	38
3.1.2 <i>Solvation free energy</i>	41
3.1.3 <i>Interface Area</i>	42
3.2 <i>Interface residues evaluation</i>	50
3.3 <i>Introducing docking parameters</i>	56
3.3.1 <i>Results</i>	56
4. Discussion	62
5. Bibliography	69

Abstract

The ROP protein is a dimeric protein with an α -helix bundle conformation that is found in bacteria. ROP protein is involved in the replication regulation mechanism of the colicinogenic factor EI) ColEI plasmid. The ROP protein restricts the copy numbers of the aforementioned plasmid by interacting with a RNA kissing loop structure created by the RNAI and RNAII molecules coded by the same plasmid [1, 2, 3]. The ROP protein by itself, has been studied very widely for its characteristic conformation and the structural, thermodynamic and functional changes that various mutations can provoke on the protein [21, 22, 23]. Although, there are numerous experimental data for the ROP protein and its mutations available in the literature and both the ROP protein and the RNA kissing loop 3D structures can be found in the Protein Data Bank, the complex 3D structure between the ROP protein and the RNAI-RNAII kissing loop is still to be determined. The literature only provides a small number of suggestions regarding the complex conformation [23, 29]. This thesis study, aims to model this molecular complex utilizing computational modeling tools. The modeling is achieved through the docking method and the evaluation of the complexes created is based on their comparison with the ROP protein mutants - RNAI-RNAII kissing loop complexes created with the same methods within this study. Finally, it proved that the complex models formed between the non-functional (lacking RNA binding affinity) ROP mutants and the RNA kissing loop produce low-quality docking decoys (models) compared to the models created from the wild type ROP protein binding upon the RNAI-RNAII kissing loop. This study lastly suggests one model complex structure that is more likely to resemble the naturally occurring ROP protein - RNA kissing loop complex although a few more steps are recommended for future docking studies of this complex, that can improve the reliability of the modeling results.

Key words

ROP (Repressor of Primer) protein

RNA kissing loop

Molecular Docking

Computational Structure Modeling

1. Introduction

1.1 ROP - RNAI-RNAII kissing loop complex in nature

ROP is a dimeric protein that plays a key participatory role in the underlying mechanism that regulates the copy number of the (colicinogenic factor EI) ColE1 plasmid. The plasmid is present in bacteria, with approximately 24 copies present in *E. coli*. ROP acts so as to inhibit plasmid replication by stabilizing the kissing complex of RNAI-RNAII molecules, increasing the affinity between them. The RNAI, RNAII and the ROP protein are all encoded by the ColE1 plasmid genes. The presence of the RNAII molecule is crucial for initiating plasmid replication; the RNAII molecule hybridizes with the plasmid's origin of replication, providing DNA polymerase with a starting point to kickstart the polymerization process. The RNAII molecule can act as the DNA polymerase starting point, however, only after it is cleaved by RNase H. The RNAI molecule can be hybridized with the RNAII molecule, driving the DNA-RNAII interaction to destabilization, by occupying the DNA replication starting point. This is the main mechanism governing ColE1 plasmid replication regulation [1, 2, 3].

The ROP protein does not play a direct role on plasmid replication regulation. Its main function is stabilizing the RNAI-RNAII interaction (the RNA kissing loop), thus inhibiting the plasmid replication initiation, since the DNA-RNAII interaction is now blocked. In concrete terms, the role of the ROP protein is to reduce the equilibrium dissociation constant of the initial RNA complex, rendering the possibility for DNA-RNAII association more unlikely. [3, 4, 5].

1.2 ROP protein

The ROP protein structure was uncovered all the way back in 1986 through the application of crystallography methods, while its soluble form was defined in 1990 using NMR [11, 12].

The protein is dimeric, with each monomer consisting of 63 amino acids that shape 2 α -helices folding together in a coiled-coil structure, due to the hydrophobic forces prevalent in the structure interior. A turn is formed close to the midpoint of the monomer sequence, allowing for the folding and subsequent formation of the coiled-coil. The two monomers are antiparallel and form a α -helix bundle, sitting at a height of approximately 45 Å [11, 12].

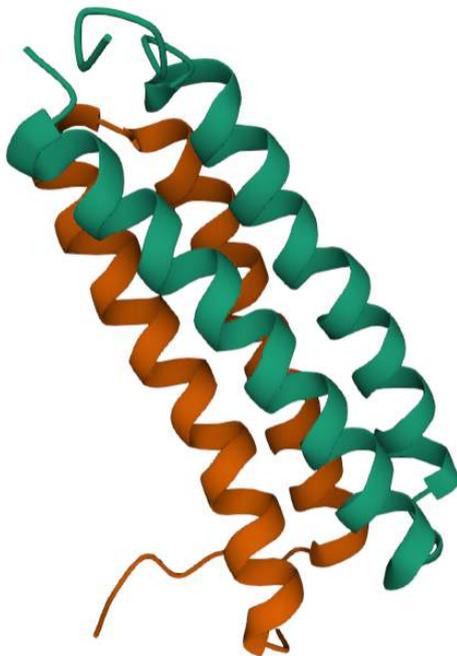


Figure 1| *PDB entry: 1RPR ROP protein*

While the exact workings of the interaction between the protein and the RNA complex are yet to be elucidated, some key components are now known, with examples being some (important for the protein-RNA complex interaction) ROP amino acids including the Lys-3, Asn-10, Gin-18, and Lys-25, as well as Phe-14 being one of the most important for the protein - RNA loop interaction. Notably, the way this interaction is governed appears to be dominated by the structure of the complex, instead of its sequence, since the protein has been found linking to other loop-loop complexes, such as the kissing complex present in HIV TAR [13,14, 15].

1.3 RNA kissing complex

The RNA loop-loop interactions are involved in various biological processes including the regulation of viral genetic material and plasmid DNA replication, the initiation of the RNA tertiary structure formation and others [6]. Both the RNAI and the RNAII genes are found on the upstream of the plasmid's origin of replication. The 5'-end coding region of the RNAII gene is antiparallel to the RNAI coding region. The length of the mature RNAII molecule is 555 nucleotides long and the RNAI length is 108 nucleotides.

The tertiary structure of the RNA molecules determines their ability to interact with other molecules (DNA, RNA, proteins etc.). The mature RNAI forms 3 stem loops ending up to a free 5'-end that lacks a specific structure. For the RNAII to be able to function as the DNA replication starting point for the ColE1 plasmid, its 3' -end is required to be enzymatically cut by RNase H. The RNAII molecule forms stem loops just like RNAI. For the kissing complex to be formed, the unhybridized nucleotides at the edge of the stem loops of the two RNA molecules interact. These stem loop regions bind each other in a reversible way, forming Watson and Crick bonds. The length of the binding region on each of the stem loops is 6-8 nucleotides for the RNAI-RNAII kissing complex formation [7]. This is the phase when the ROP protein can bind the RNAI-RNAII complex to stabilize their connection. The RNAI molecule is linearized on the next step of the RNAI-RNAII interaction, forming a full-length binding with the 5'-end of the RNAII molecule [8, 9, 10]. The process is depicted below:

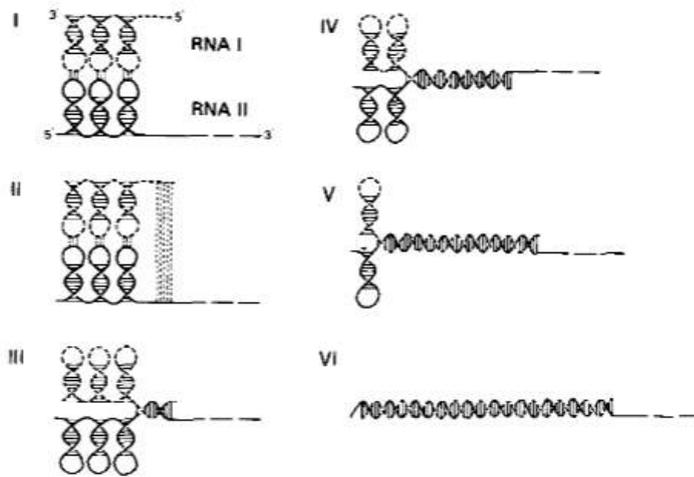


Figure 2| *RNAI - RNAII kissing interaction*
(Tomizawa, J. I. 1985)

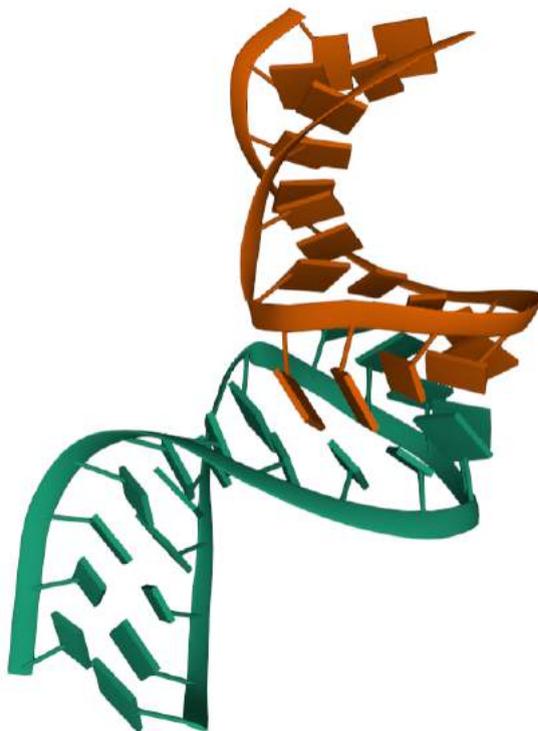


Figure 3| *PDB entry: 2BJ2*
The solution structure of the RNAI -RNAII loop-loop complex (Lee, A. J., & Crothers, D. M. 1998)

1.4 ROP - RNAI-RNAII complex modeling

Molecular docking is a modeling process that aims to predict the ways of interaction between two or more molecules, as is needed for the ROP and the kissing complex. Docking, specifically, is a method to generate the possible orientations the two molecules forming a complex can have in relation to each other, and then selecting the most preferred, according to the degree by which it corroborates the other experimental data regarding the complex present. Docking is computationally taxing when the linkage is formed between a protein and an RNA molecule, both because the latter is flexible (especially when in single-strand form) and because of the structural data in regards to RNA being severely lacking. Nonetheless, a handful of various software to apply docking methods in protein-RNA complexes have been developed, and are usually modified versions of the corresponding software used for protein docking [18]. The fundamental input needed for the docking algorithms is the structure files that can come with other parameters such as the interacting regions and the number of protein residues and ribonucleotides that form the contact regions [19, 20].

There are many attempts described in the related literature that are focused on the ROP protein - RNAI-RNAII kissing loop interaction and the complex formation study [23, 29, 39]. Although today the NMR structures of the ROP protein and the one of the RNA kissing loop are available at the PDB, the complex has not been crystallized [7, 9]. However, due to the high scientific interest that ROP protein used to attract, mostly for structural studies due to its characteristic α -helix bundle conformation, scientific data regarding its structural, physicochemistry features and others are available. Consequently, there are many homology structures available in the PDB for the ROP protein, making it easier to perform a comparison analysis of the ROP protein and its mutants binding upon the RNA kissing loop, and evaluate the docking results by referring to the literature data about the functionality of the ROP and its mutants on binding RNA. More specifically, after the complex modeling using the docking method, the docking models (decoys) produced, will be compared. The ROP mutants that lack their functionality are expected to create poor docking decoys (model) when combined with the RNAI-RNAII kissing loop, contrasted to the decoys resulting

from the wild type ROP protein, known for its binding upon the RNA kissing loop [16, 17, 21, 22, 23].

This study is mostly based on research findings regarding the ROP protein and not the RNAI-RNAII kissing loop since the bibliography related to the RNA kissing loop and in general, to RNA structures is still nascent [35].

2. Materials and Methods

In the following section the whole docking protocol is described step by step. The first step of the process is selecting the ROP structure file to be studied, as well as the mutants of the protein of considerable scientific interest in the context of homology modeling. The mutants can be used as evidence for the evaluation of the docking results, since the difference among the docking decoys originating from the wild type ROP or from the mutants of the protein should reflect the mutants' distinctiveness from the native ROP structure. For example, some of the mutants are known for their characteristic, complete loss of function, which is expected to affect the decoys' metrics when compared with the ones generated from the native structure [21, 22, 23]. After selecting the structures of interest, the protocol continues with the preparation of the receptors and ligands that will be used for the docking trials. The preparation step is of high importance, since it is necessary to achieve high quality docking decoys that do not contain any other molecules such as solvents and water that may interfere with the docking process [24, 25, 26]. Finally, the structures are ready to be introduced to the docking algorithm. The docking software produces a list of results including various docking decoys (possible docking solutions) [32, 33]. Most of the docking algorithms include an intrinsic scoring function that can be used to evaluate the docking decoys and categorize them according to their scoring values, selecting those with the higher values for further analysis [25]. The last step of the process is the evaluation of the decoys using other algorithms for interface structural analysis such as the PDBePISA [49]. However, it is important to further assess the docking results from the biological aspect, taking into account all the relevant literature available.

2.1 ROP protein models selection

At this point it is important to mention that, since the ROP protein structure was uncovered through the NMR method, the ROP protein structure file provided by the Protein Data Bank (PDB), with the (PDB ID: 1RPR) contains 10 structural models in total. However, in the Validation Document for the ROP protein structure, it is mentioned that the 8th model is the average model that resembles all the other 9 structure models most [7]. Thus, the model 8 of the 1RPR is the one that will be used for all analyses throughout this thesis for both ROP and ROP - (57-63 [residue]) tail protein models [27].

The ROP - (57-63 [residue]) tail, is a case that will be tested where the ROP protein lacks its last 7 residues from each of its monomers. The investigation of this case is required, in order to test the hypothesis that these peptide tails are preventing the docking process. The peptide tails that lack a specific secondary structure are flexible in nature, however the docking algorithm is not expected to predict such important conformation changes and thus the ROP protein tails were removed manually.

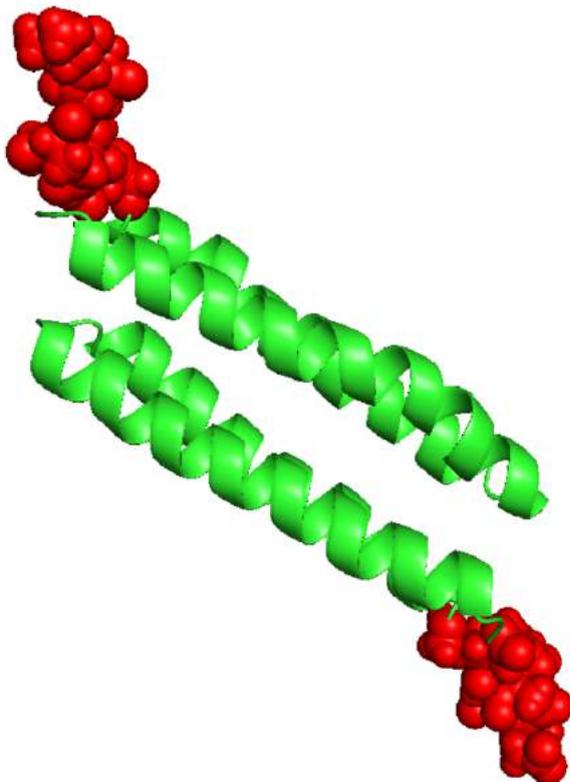


Figure 4| *ROP protein 3D structure with highlighted red 57-63 residue tails (image created by PYMOL [31]).*

The PDB ID of the protein and RNA structures will be used from now on when referring to the according structures to avoid any misinterpretation between the data presented in the study and the modified text files of PDB structures. Therefore, the term ROP protein and 1RPR as well as the 2BJ2 and the RNAI-RNAII kissing loop will be used alternatively throughout the present study.

To select the ROP mutants that will be used in the context of the homology modeling of the complex, the PROTEIN BLAST server was utilized. The input protein sequence was the one provided by the PDB FASTA file for the 1RPR NMR structure model of the ROP protein. The BLAST search for homologs was restricted among the homologs that exist as structure files in the Protein Data Bank, and the results are the following [27, 50]:

INPUT PROTEIN SEQUENCE:

>1RPR_1|Chains A, B|ROP|Escherichia coli (562)

MTKQEK TALN MARFIRSQTLTLLEKLNELDADEQADICESLHDHADELYRSCLARFG
DDGENL

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	Structure Of The ColE1 Rop Protein At 1.7 Angstroms Resolution [Escherichia coli]	Escherichia coli	129	129	100%	5e-41	100.00%	63	1ROP_A
<input checked="" type="checkbox"/>	ALANINE 31 PROLINE MUTANT OF ROP PROTEIN [Escherichia coli]	Escherichia coli	127	127	100%	2e-40	98.41%	63	1B8Q_A
<input checked="" type="checkbox"/>	Chain A, Regulatory protein rop [Escherichia coli]	Escherichia coli	126	126	98%	5e-40	100.00%	63	2LJK_A
<input checked="" type="checkbox"/>	Crystal Structure of the Rop protein mutant D30P/A31G at resolution 1.4 resolution [Escherichia coli]	Escherichia coli	125	125	100%	1e-39	96.83%	70	4DQ2_A
<input checked="" type="checkbox"/>	Chain A, Regulatory protein rop [Escherichia coli]	Escherichia coli	125	125	98%	1e-39	98.39%	63	2LJJ_A
<input checked="" type="checkbox"/>	Chain A, Regulatory protein rop [Escherichia coli]	Escherichia coli	125	125	98%	1e-39	98.39%	63	2LJH_A
<input checked="" type="checkbox"/>	Chain A, Regulatory protein rop [Escherichia coli]	Escherichia coli	123	123	98%	6e-39	98.39%	63	2LJI_A
<input checked="" type="checkbox"/>	Chain A, ROP [Escherichia coli]	Escherichia coli	122	122	96%	1e-38	98.36%	62	1GTO_A
<input checked="" type="checkbox"/>	Chain A, Regulatory protein rop [Escherichia coli]	Escherichia coli	122	122	98%	2e-38	96.77%	63	3K79_A
<input checked="" type="checkbox"/>	ATOMIC RESOLUTION (1.07 ANGSTROMS) STRUCTURE OF THE ROP MUTANT <2AA> [Escherichia coli]	Escherichia coli	121	121	100%	5e-38	95.38%	65	1NKD_A
<input checked="" type="checkbox"/>	Crystal structure of a five-residue deletion mutant of the Rop protein [Escherichia coli]	Escherichia coli	113	113	100%	5e-35	92.06%	58	1QXB_A
<input checked="" type="checkbox"/>	Chain A, Regulatory protein rop [Escherichia coli]	Escherichia coli	115	170	96%	6e-35	91.80%	120	1YQ7_A
<input checked="" type="checkbox"/>	Chain A, ROP ALA2ILE2-6 [Escherichia coli]	Escherichia coli	111	111	98%	4e-34	85.48%	63	1F4M_A
<input checked="" type="checkbox"/>	Chain A, Regulatory protein rop [Escherichia coli]	Escherichia coli	103	103	88%	7e-31	89.29%	65	7KAE_A

Figure 5| BLAST PROTEIN alignment for the ROP (1RPR) protein search in PDB [50]

Most of the BLAST run results are reliable, however only some of the proteins should be picked to continue with the docking protocol. These proteins should be of high scientific interest that in this case, is their inability to bind to the RNA kissing loop due to the mutations they carry:

1. **1F4N**: It is a ROP protein mutant carrying the Ala2Ile2-6 mutation. This mutation causes a conformational flip to the protein's core that leads to complete loss of the RNA binding ability [22].

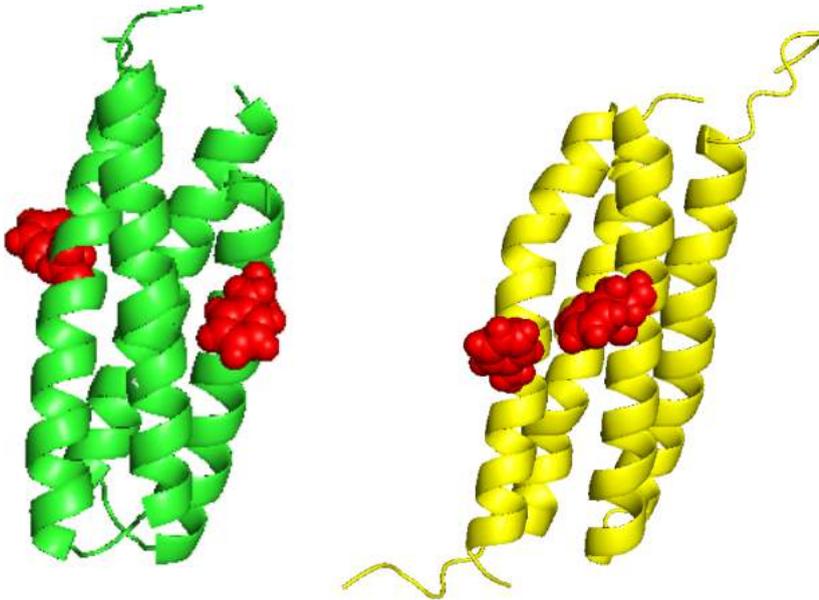


Figure 6| On the left the 1F4N ROP mutant is shown and wild type ROP protein is placed on the right. The Phe-14 residue is highlighted with red color in both structures (image created by PYMOL [22, 31]).

2. **1QX8 (RM6)**: It is a ROP mutant known as the loopless Rop. This mutant presents a dramatic change of its structural conformation after a deletion of the residues 30-34 and designed to restore the heptad periodicity at the turn region. This mutant completely lacks the RNA binding property [21].

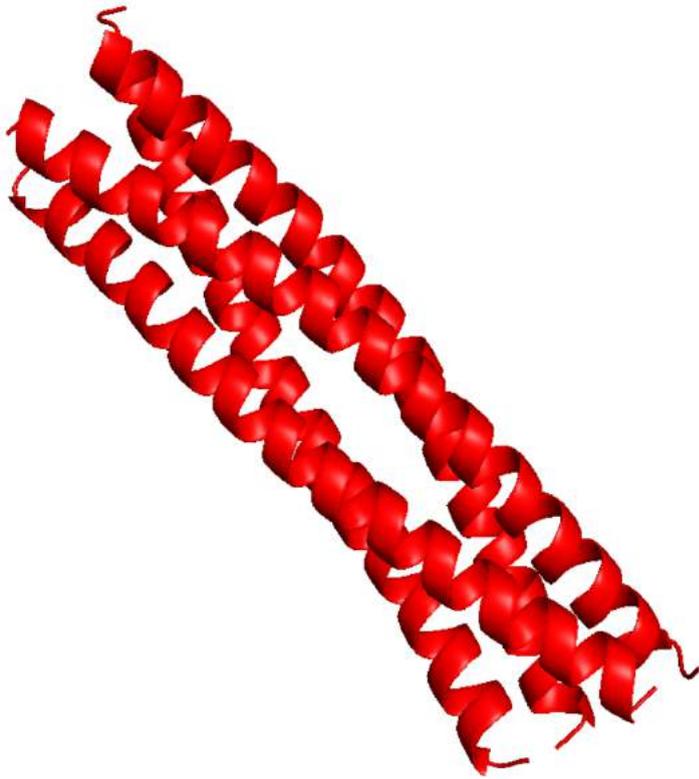


Figure 7| *The 1QX8 ROP mutant (image created by PYMOL [21, 31])*

3. **2IJH**: It is a ROP protein mutant that contains only one residue substitution in both of its monomers where the Phe-14 (phenylalanine) is changed to Trp (tryptophan). This mutant also lacks the binding affinity for the RNA kissing loop [23].

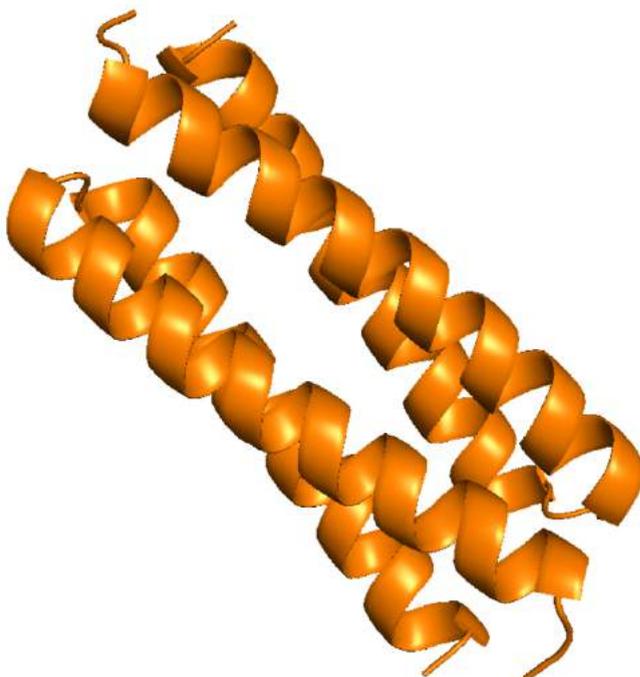


Figure 8| *The 2IJH ROP mutant (image created by PYMOL [23, 31]).*

An RNA kissing loop is used as a negative marker for the docking trials. This molecule is the well-characterized TAR-TAR kissing loop (PDB ID: 1KIS) that has already been studied for each binding upon the ROP protein [28, 29]. However, the naturally occurring protein-RNA complex is the one containing the ROP protein and the RNAI-RNAII kissing complex and thus the results coming from these molecular combinations are expected to be more reliable and consistent than those resulting from the docking between the ROP protein and the TAR-TAR kissing loop. That is also the case for the docking trials performed between the ROP mutants and the RNAI-RNAII kissing loop. Poor quality results are expected from the docking trials between the mutants and the RNA kissing loop in contrast with the docking decoys from the native ROP protein and the RNAI-RNAII kissing loop.



Figure 9| *PDB entry: 1KIS*
TAR-TAR KISSING HAIRPIN
COMPLEX (Chang, K. Y., &
Tinoco Jr, I. 1997)

2.2 Docking preparation

In order to prepare a structure properly before any docking simulations, a series of actions are required, including removing solvent molecules and other ligands, adding hydrogen molecules, replacing non-regular amino acids with their regular forms — such as changing methylselenyl-dUMP into typical uracil —, fulfilling incomplete side-chains or other parts of the structures if needed, etc [25, 30]. There are many standardized protocols for docking preparation available with the one provided from Chimera UCSF [53] being among the most widely used. However, it is worth mentioning that the NMR structures such as the ROP structure 1RPR and the ligand structure 2BJ2 (RNA kissing complex) as well as the TAR-TAR RNA complex (1KIS), do not contain additional ligands or solvents in their structure files. The aforementioned structure files do not lack hydrogen atoms, a problem that is mostly present in X-ray crystallography-determined structures [54] and, moreover, these two structure files are complete and thus both the polypeptide chains and the ribonucleic acid chains are continuous. Lastly, no rare amino acids are found in the ROP structure, meaning the docking preparation protocol does not need to be applied on these structure files.

On the other hand, the structure files for the ROP mutants used in this study should be edited using the Chimera docking protocol [53]. These structures are the following:

1. 1F4N
2. 1QX8
3. 2IJH

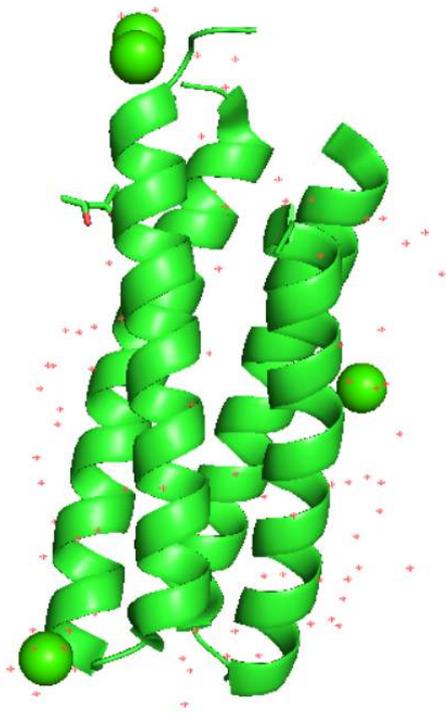


Figure 10| 1F4N ROP mutant **before** its modification through the Chimera Docking Preparation Protocol (the picture was obtained through the PYMOL program [22, 31])

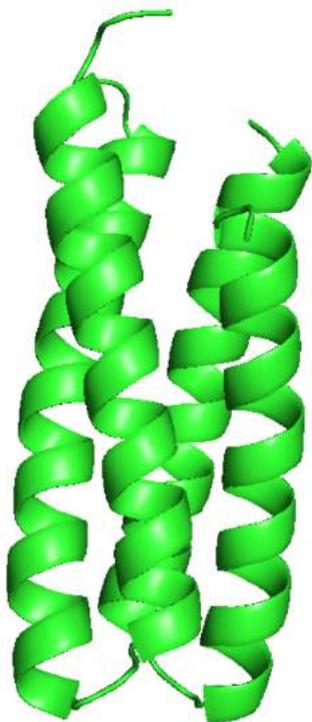


Figure 11| 1F4N ROP mutant **after** its modification through the Chimera Docking Preparation Protocol (the picture was obtained through the PYMOL program [22, 31, 53])
Not all the modifications caused by the Docking Preparation Protocol are obvious from this picture (only the water, and ligand molecular are shown to be missing).

2.3 Docking principles

Docking is one of the most useful methods used for molecular modeling with strong applications in the field of drug discovery and development. As it is already briefly mentioned in the first chapter, molecular docking is a procedure that generates all the possible orientations that a molecule can obtain over (an)other molecule(s). This method is commonly applied in the field of drug development when a computational approximation of the binding site is required before the research protocol moves to the experimental measurements. These computational methods can help the scientific community to save both time and money during the performance of large drug discovery protocols [32, 33, 55, 56]. Regarding the protein to RNA molecular docking, the methods can be divided into two distinct categories according to the relative literature: rigid body and flexible docking. The rigid body methods can generate a preliminary picture of the complex under investigation and are preferred when limited information is available for the structure to be defined. Additionally, the rigid body docking category is mostly utilized in cases where no major conformational changes are expected from the complex formation process. On the other hand, the flexible docking algorithms are constructed so as to conduct more detailed docking analysis and thus require more strict input parameters for the docking process [25, 56].

The most popular docking algorithms are provided below together with the indication of their docking type (rigid/flexible).

Name	Modified from Protein-Protein Docking Method	Docking Method (Rigid/Flexible)	Availability	
			Web Server	Standalone
3dRPC	×	Rigid	✓	✓
ClusPro	✓	Rigid	✓	×
FTDock	✓	Rigid	×	✓
GRAMM	✓	Rigid	✓	✓
Hex	✓	Rigid	✓	✓
ICM	✓	Rigid	×	✓
NPDock	×	Rigid	✓	×
PatchDock	✓	Rigid	✓	✓
PEPSI-DOCK	✓	Rigid	×	✓
pyDock	✓	Rigid	✓	✓
RosettaDock	✓	Rigid	✓	✓
ZDOCK	✓	Rigid	✓	✓
ATTRACT	✓	Flexible	✓	✓
HADDOCK	✓	Flexible	✓	✓
HDOCK	×	Flexible	✓	×
PIPER	✓	Flexible	×	✓
Prime	✓	Flexible	×	✓

Figure 12| Docking software for protein - RNA complexes modeling [25].

2.4 Docking Methodology

The first docking trials were performed using various docking algorithms from those mentioned in the table above, both in the online server or the standalone form. Some of the docking algorithms were more user friendly, providing plenty of relevant documentation on how to use them and were easy to use even for non-experienced users, others were more demanding, requiring various input parameters regarding the independent molecule to be introduced to the docking program and the expected results etc. Some of them require a tool package of considerably large size to be installed in order to use them and some others are slow, with one docking trial lasting for about one day with a typical personal computer setup. Some of them were created only for protein-protein docking and are now modified in order to be able to receive both proteins and DNA as input structure (e.g. pyDOCKDNA) [25, 32, 33, 34, 35]. After numerous docking attempts, the docking program selected for the ROP - RNAI-RNAII kissing loop complex studies is the HDOCK program, created by the Huang Lab [35].

2.4.1 HDOCK program

The docking algorithm that has been selected for the docking runs is the HDOCK program. HDOCK is based on a hybrid docking algorithm, working both with free docking and template based docking. The HDOCK program provides an intrinsic scoring method and it is also able to accept proteins and nucleic acids as input structure files. During the docking process, the algorithm places the receptor molecule in a fixed orientation and the ligand molecule performs rotation by an interval of 15° Euler angles in the rotational space, within a grid spacing of 1.2 Å implemented for molecular shape complementarity research. This procedure is executed with the Fast Fourier Transformation grid-based algorithm. The specialized, scoring function used from the HDOCK program for protein-nucleic acid decoys evaluation is the ITScore-PR function. ITScore-PR is characterized as a statistical mechanics-based and knowledge-based scoring method, achieving high success rates and allowing for RMSD measurements of the ligand molecule, a feature of high importance regarding the flexibility of the RNA complex used as the ligand for the present study [33, 35, 36, 37, 38].

The HDock server was selected for the docking studies due to the following advantages that it provides to the user:

1. Easy handling for inexperienced users
2. No prerequisites for binding restraints or pre-oriented structure file of the receptor and the ligand
3. Docking Calculations of high speed
4. Intrinsic scoring method
5. RMSD calculations for the receptor molecule/accounting for RNA's characteristic flexibility
6. Compatible output data/easy to introduce them for other calculations/evaluation

The picture below outlines the workflow of the HDock algorithm:

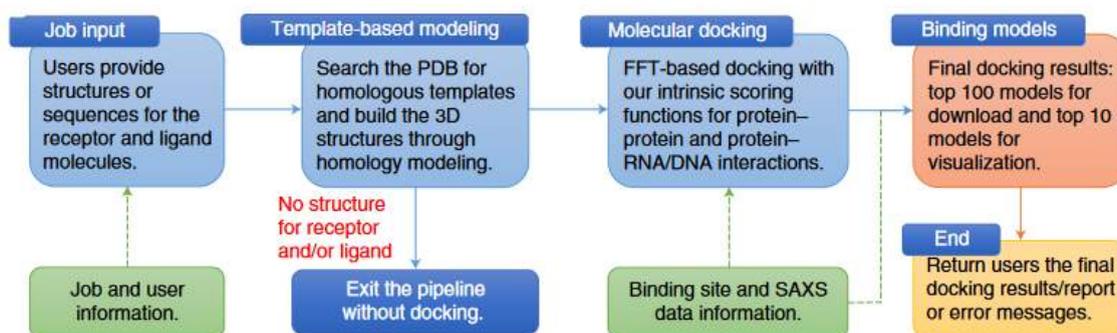


Figure 13| The workflow of the HDock server [35].



HDOCK SERVER

Protein-protein and protein-DNA/RNA docking based on a hybrid algorithm of template-based modeling and *ab initio* free docking.

[\[Huang Lab\]](#) [\[HDOCK\]](#) [\[Help\]](#) [\[Output example\]](#)

Input Receptor Molecule using ONE of the following four options: [\[help\]](#)

- Upload your **pdb** file in **PDB format**: No file selected. [\[example\]](#)
OR provide your **pdb** file in PDB ID:ChainID: (Example: [1CGI:E](#))
OR copy and paste your **sequence** below in **FASTA format** (Sample input: [1CGI:E](#), [1HCJ:A](#))

OR upload your **sequence** file in **FASTA format**: No file selected. [\[example\]](#)

Input Ligand Molecule using ONE of the following four options: [\[help\]](#)

- Upload your **pdb** file in **PDB format**: No file selected. [\[example\]](#)
OR provide your **pdb** file in PDB ID:ChainID: (Example: [1CGI:I](#))
OR copy and paste your **sequence** below in **FASTA format** (Sample input: [1CGI:I](#)) [\[help\]](#)

OR upload your **sequence** file in **FASTA format**: No file selected. [\[example\]](#)

Advanced Options (Optional):

- Template-free docking only [\[Explanation\]](#)
- Symmetric multimer docking: (e.g., **C2** or **C3** for Cyclic; **D2** or **D3** for Dihedral) [\[Note\]](#)
- SAXS experimental data file: No file selected. [\[help\]](#) [\[example\]](#)
- [Specify the residues of the binding site.](#)

Optional:

Enter your email:

Enter your jobname:

Figure 14| HDOCK home page [35].

After the docking preparation step, the molecules for docking were introduced to the HDock server. The docking attempts and the crossovers among the independent molecules are shown in the table below. All the results obtained were saved, however only the first 10, best rated docking decoys, will be used for further studies. All the output pdb formatted files contain information about the docking energy (mentioned as docking score) and the RMSD score of the ligand. In the case that there are decoys with the same RMSD values, the one with the lower docking energy is considered as the higher-confidence prediction achieved. The following picture is an example of the results page of the HDock web server [33, 35]:

Download Files

[Receptor PDB file](#) [Ligand PDB file](#)
[\[1\]](#) [\[2\]](#) [\[3\]](#) [\[4\]](#) [\[5\]](#) [\[6\]](#) [\[7\]](#) [\[8\]](#) [\[9\]](#) [\[10\]](#) [\[11\]](#) [\[12\]](#) [\[13\]](#) [\[14\]](#) [\[15\]](#) [\[16\]](#) [\[17\]](#) [\[18\]](#) [\[19\]](#) [\[20\]](#)
[Top 10 Predictions](#) [Top 100 Predictions](#) [All the results in a package](#)



Receptor style
Cartoon

Receptor color
 Pure
 Rainbow

MODEL No.

Ligand style
Cartoon

Action

Complex Template Information [\(Click to Hide\)](#)

Molecule	PDB ID	Chain ID	Align_length	Coverage	Seq_ID (%)
Receptor	1YO7	A	120	0.952	68.3

Note: The built model of "Model 0" based on the above PDB complex template has a confidence.

Summary of the Top 10 Models

Rank	1	2	3	4	5	6	7	8	9	10
Docking Score	-297.10	-284.13	-283.86	-280.32	-280.23	-278.05	-274.26	-267.98	-266.86	-266.20
Confidence Score	0.9499	0.9360	0.9357	0.9313	0.9312	0.9283	0.9231	0.9137	0.9119	0.9108
Ligand rmsd (Å)	67.35	72.77	84.58	70.40	71.97	71.80	73.25	84.31	72.66	64.10
Interface residues	model_1	model_2	model_3	model_4	model_5	model_6	model_7	model_8	model_9	model_10

Note: The models are ranked according to the docking scores. Please click [help](#) for the explanations of evaluation metrics

Quality of Docking Structures/Input Data [\(Click to Hide\)](#)

Receptor: The protein structure was submitted by users.

```

=====
Quality report of input model by ProQ v1.2
=====
LGscore:      2.218
MaxSub:       0.168
    
```

```

Different ranges of protein model quality by LGscore or MaxSub:
Connect      Good      Very good
1.5 < LGscore  3.0 ≤ LGscore < 5.0  5.0 ≤ LGscore
0.1 < MaxSub   0.5 ≤ MaxSub < 0.8   0.8 ≤ MaxSub
    
```

Ligand: The RNA structure was submitted by users.

NOTE: Quality checking is not conducted for this user-input RNA structure.

Figure 15| HDOCK docking results page [35].

Protein ID	RNA ID		Restrains
	2BJ2	1KIS	
1RPR	✓		-
1RPR		✓	-
1RPR - (57-63 tail)	✓		-
1RPR - (57-63 tail)	✓		ligand res: A[9]:U[12], U[10]:A[11], G[11]:C[10], G[12]:C[9]
1F4N	✓		-
1QX8	✓		-
2IJH	✓		-

Table 1| All the docking crossovers that will be used for docking purposes.

2.4.2 Scoring methods

Some of the docking programs/web-servers that are available, provide an intrinsic scoring algorithm - as it is already noticed regarding the HDOCK program. These algorithms are implementing a different kind of scoring method, based on various structure or bio-physical features to evaluate how close to the native protein-RNA complex are the decoys produced during the docking algorithm run. Among the most widely known scoring methods the DARS-RNP [41] and the ITScore-PR [37] are included with both of them regarded as knowledge based algorithms. The ITScore-PR, is the scoring algorithm that the HDOCK program utilizes for the docking decoys assessment. Nevertheless, the known RNA 3D structures are still limited and thus the scoring algorithms that aim specifically to the protein-RNA complexes' evaluation are prone to errors, especially when they are based on a knowledge based method [35]. Consequently, a second round of docking decoys evaluation is needed, to further filter the results and strengthen the reliability of the docking decoys that may include the docking solution. There are a variety of

independent scoring programs and servers that can be used to assess the results' validity by performing complex interface analyses. For the present study both independent evaluation algorithms as well as docking-intrinsic scoring methods will be used and the results will be compared when possible [42, 43, 49]

2.4.3 Additional Test Cases

In every case of the resulting docking decoys for wild type ROP - RNAI-RNAII kissing loop complex, the ROP protein was placed close to the inner side of the kissing loop as suggested from this publication see the pictures below [23]:

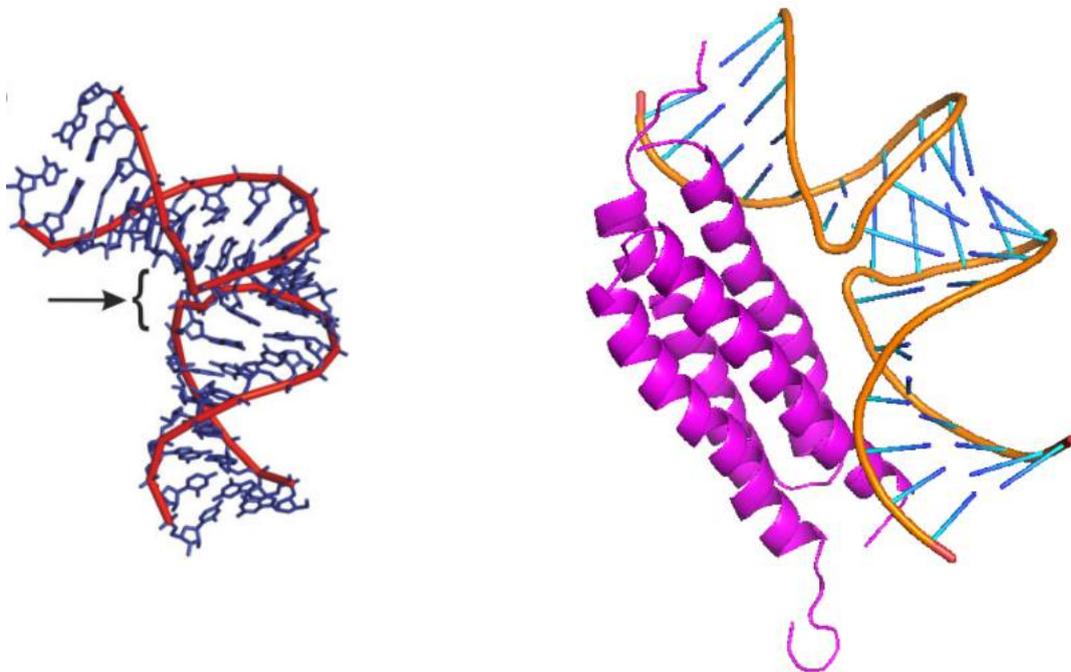


Figure 16| On the left the suggested location for ROP interaction upon the RNAI-RNAII kissing loop is depicted. On the right side the first rated docking decoy between 1RPR (NMR file of the ROP) and 2BJ2 (RNAI-RNAII kissing complex) is shown [23].

However, the literature is inconclusive regarding the relative orientation between the ROP protein and the RNA kissing complex, with the Comolli, L. R., Pelton, J. G., & Tinoco Jr, I. (1998) [29] study focused on the interaction between the

ROP protein and the Tar-Tar kissing complex claiming that the most possible relative orientation between the ROP protein and interaction RNA kissing loop being the one with the ROP placed on the outer side of the kissing loop as shown below [23, 29].

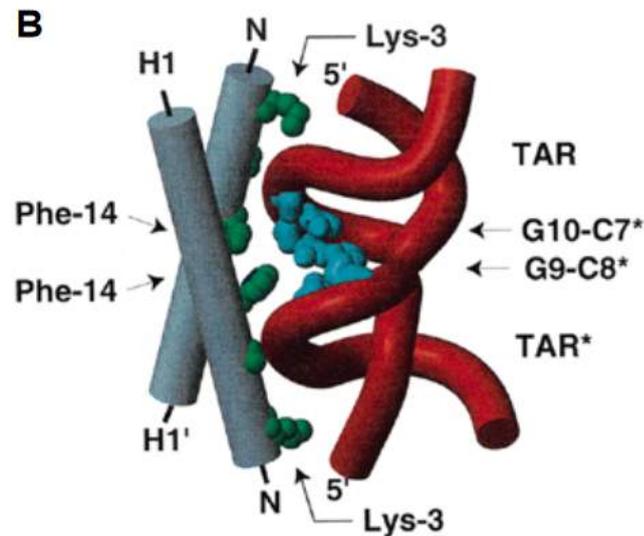


Figure 17| *The suggested complex of ROP protein and Tar-Tar kissing loop [29].*

Consequently, additional docking trials were required, aiming to test the hypothesis of the ROP protein being placed at the outer side of the RNA kissing loop as it is suggested from the publication above. For this purpose, the first step was to investigate the docking crossover between the ROP protein and the TAR-TAR RNA kissing loop *1KIS*, which was used for the studies of [29]. Even in this case, the ROP-TAR RNA complex took the same orientation as the ROP-RNAI - RNAII kissing complex. More details are discussed later in the present thesis study.

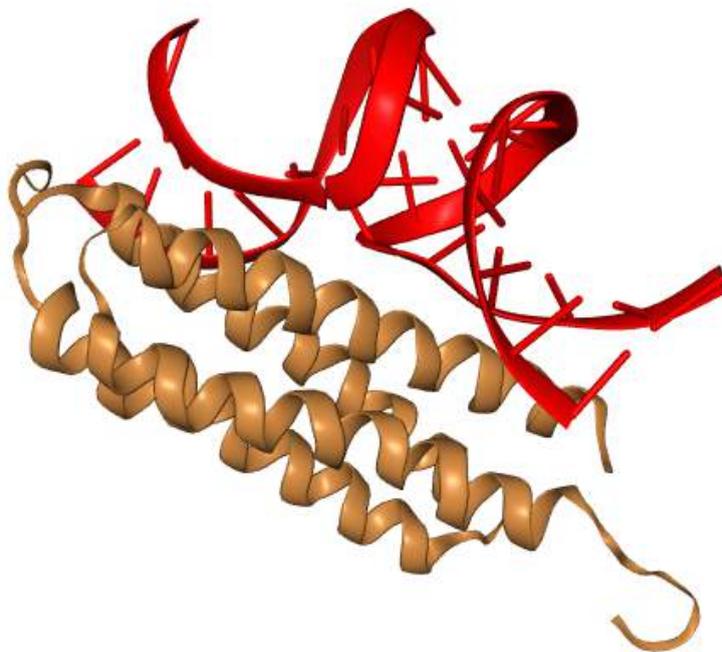


Figure 18| *ROP protein - TAR-TAR kissing loop complex decoy [35].*

In any case, the need for introducing docking restraints for the docking studies on the ROP protein and RNA kissing loop of interest complex is obvious - as the second step to study the hypothesis of the ROP protein being placed on the outer side of the RNA loop [29]. Thus, some more docking trials should be performed. However, there is a lack of relevant data regarding the exact number and identity of the RNA bases that are responsible for the interaction between the ROP protein and the RNAI-RNAII kissing complex. In this context, it is worth mentioning that the binding of ROP protein upon the RNA-kissing complexes is known as a structural-dependent event [39]. Using the MultiSETTER server [40] for multiple structure alignment, it gets obvious that there is an important structural similarity with an RMSD = 1.706 between the RNAI-RNAII kissing complex (PDB ID: 2BJ2) and the TAR-TAR kissing complex (PDB ID: 1KIS).

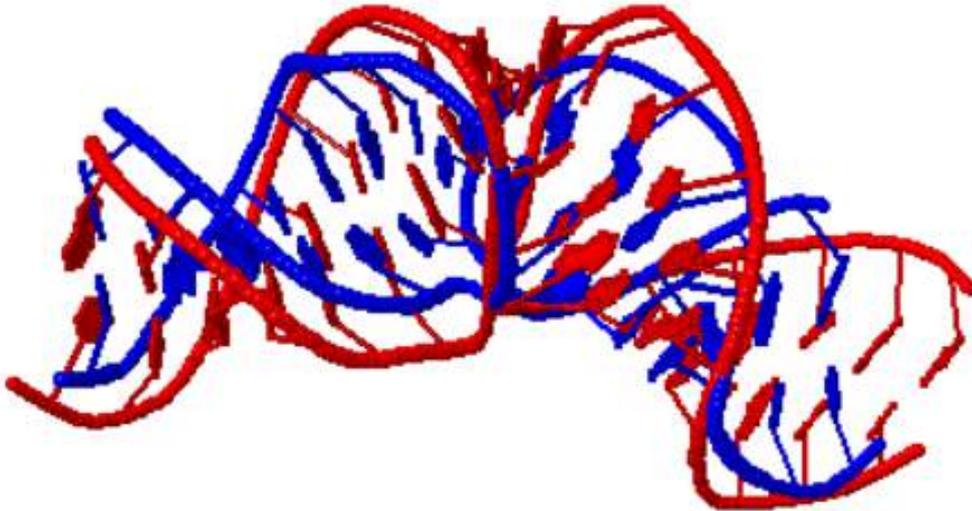


Figure 19| *Structure alignment between the RNAI-RNAII kissing complex (PDB ID: 2BJ2) in red color and the TAR-TAR kissing complex (PDB ID: 1KIS) in blue color [40].*

According to the publication suggesting that the ROP protein binds to the outer side of the RNA kissing loop, the main ribonucleotides that are involved in this interaction are the following: U(uracil):7 and G(guanine):8-10 [29]. From the alignment prepared from the SETTER server aiming to assess the structural similarity between the RNAI-RNA kissing complex and the TAR-TAR kissing complex, the region including the aforementioned RNA bases of the TAR-TAR complex, corresponds to the nucleotides of the RNAI-RNAII complex below:

I.A(adenine):9, I.U(uracil):10, I.G(guanine):11-12

Thus, one more round of docking trials between the ROP protein and the RNAI-RNAII kissing complex was performed, this time adding the restrains of the residues in contact with the protein:

Docking attempt case: I.A[9]:II.U[12], I.U[10]:II.A[11], I.G[11]:II.C[9], I.G[12]:II.C[10]

The nucleotide residue restraints can be easily introduced to the HDock server. Using the **Advanced Options** section, the restraints can be inserted for the receptor or the ligand molecule and the molecular distance parameters can be added as well. In this case, the restraints are mentioned above, concerning only the ligand molecule.

2.5 Docking results evaluation

After gathering a huge amount of docking decoys coming from various docking cases, the results can be further evaluated using tools for structural analysis of pre-calculated docking results. Some of the most used servers for relevant purposes are the PDBePISA [49] and the COCOMAPS [43] servers. These programs get the docking decoys as an input and prepare an interaction analysis of the surface of interest. However, not all of the programs allow the user to choose a specific interface surface of interest but they automatically recognize the molecular surfaces to analyze. That is exactly the problem of the PDBePISA, since the server considers each of the polypeptide or nucleic acid chains as independent molecule and thus it measures the interface characteristic between each of the chains of the input files, not allowing the user to select the surface that is shaped from molecules including more than one polypeptide or nucleic acid chain. To figure out this problem the .pdb files of the structures to be analyzed, were modified using a simple text editor, as each of the chains is continuous for the ROP protein and the RNA kissing loop. This means that in the input files of the ROP protein someone can notice that the protein does not include two separate polypeptide chains of 63 residues, but instead the modified protein now consists of one polypeptide chain of (63×2) 126 residues. The same is true for the RNA kissing complex as well. The RNA-RNA kissing loop structure files that were modified accordingly, do not contain two ribonucleic acid chains of 21 and 19 nucleotides, now the RNA kissing loop seems like a uniform RNA molecule composed of $(21+19)$ 40 nucleotides.

The evaluation results coming from the analysis carried out by COCOMAPS server and from the PDBePISA comparison are presented in the Results section of the present thesis study.

3. Results

The scoring results of the docking decoys analysis are mostly focused on the Accessible Solvent Area (ASA), Buried Area upon the molecular complex formation (BSA), Interface Area of the touching molecules and the ΔiG value which is actually the measurement of the solvation free energy gain resulting from the formation of the interface measured in kcal/M. This value is calculated as the difference in total solvation energies of isolated and interfacing structures. This means that the lower (negative values) the ΔiG value is the higher is the complex stability (however, bonds free energy contributions that are formed upon the complex formation are not included in this measurement). Moreover, complex evaluation can be conducted specifically for each of the interacting residues of the molecules that form the complex. The same values as these mentioned above can be measured for each of the interfacing residues. Both PDBePISA and COCOMAPS servers measure the BSA, ASA, and Interfacing Area for the molecular interfaces, and the PDBePISA additionally calculates the ΔiG value for each of the docking decoys [42, 43, 49].

Each docking trial using HDOCK, produces 100 complex models (decoys) that are categorized according to their scoring values. The top 10 scored decoys will be used for further analysis as it is suggested from the HDOCK server [33, 35]. The decoy models are sorted in a decreasing manner, meaning that the first docking model shows a higher docking score than the following one and thus, the possibility of a binding event between its independent docking structures is higher. However, the scoring value does not imply the binding affinity of each of the models. Another scoring value that labels each of the resulting docking decoys is the *Confidence Score*. This value is empirically defined by the HDOCK program, it is based on the Docking Score value and it is used as an indicator of the binding probability, resulting in complex formation. The Confidence Score is calculated according to the following equation:

$$[1] \quad \left[\textit{Confidence_score} = 1.0/[1.0+e^{0.02*(\textit{Docking_Score}+150)}] \right]$$

The HDOCK creators are referring to the empirical value of the Confidence Score of 0.7 and more for the complexes that are very likely to be formed. However, this threshold is empirical and further investigation is suggested from the HDOCK team [35, 63].

The figures below, show the Confidence Score for the first 10 best scored docking decoys, for each of the test cases that were used throughout the present study. The data labels present the Docking Score for each of the Decoy models:

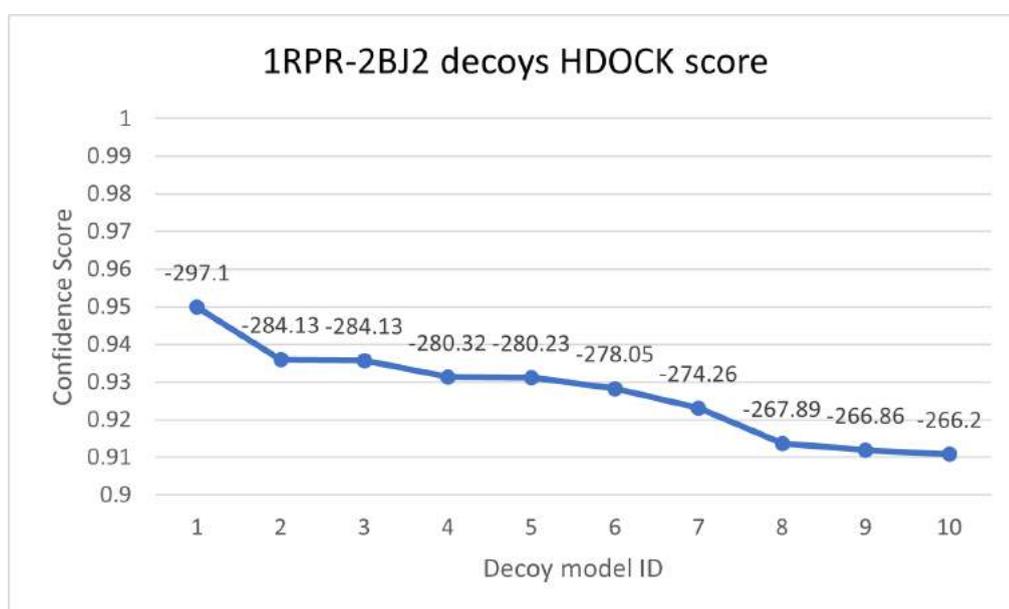


Figure 20| **HDOCK docking decoy confidence scoring** | [Crossover: ROP - RNAI-RNAII kissing loop] (the data labels are depicting the intrinsic HDOCK scoring values) [35].

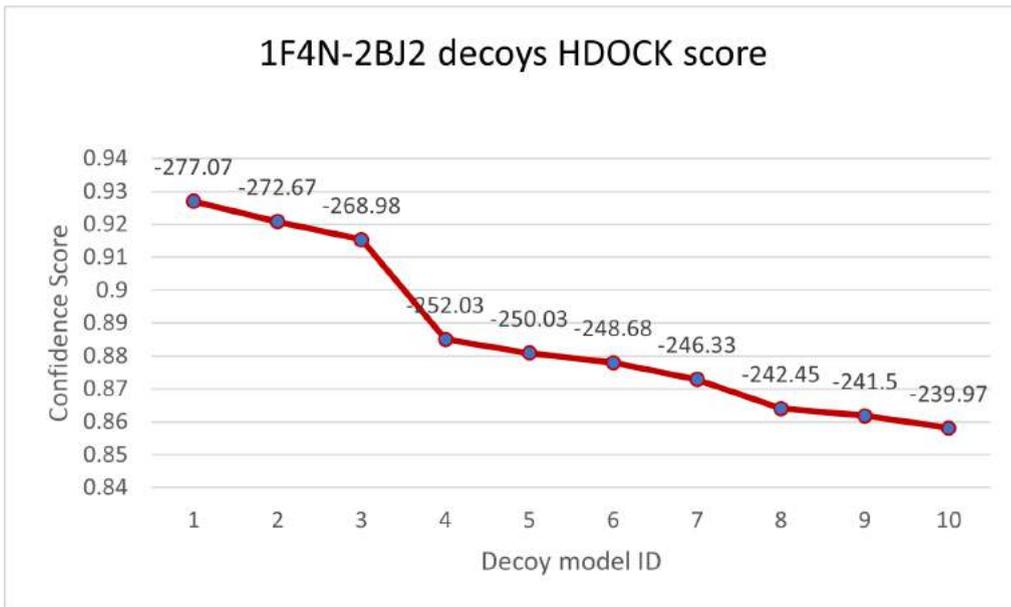


Figure 21| **HDOCK docking decoy confidence scoring** | [Crossover: 1F4N mutant - RNAI-RNAII kissing loop] (the data labels are depicting the intrinsic HDOCK scoring values) [35].

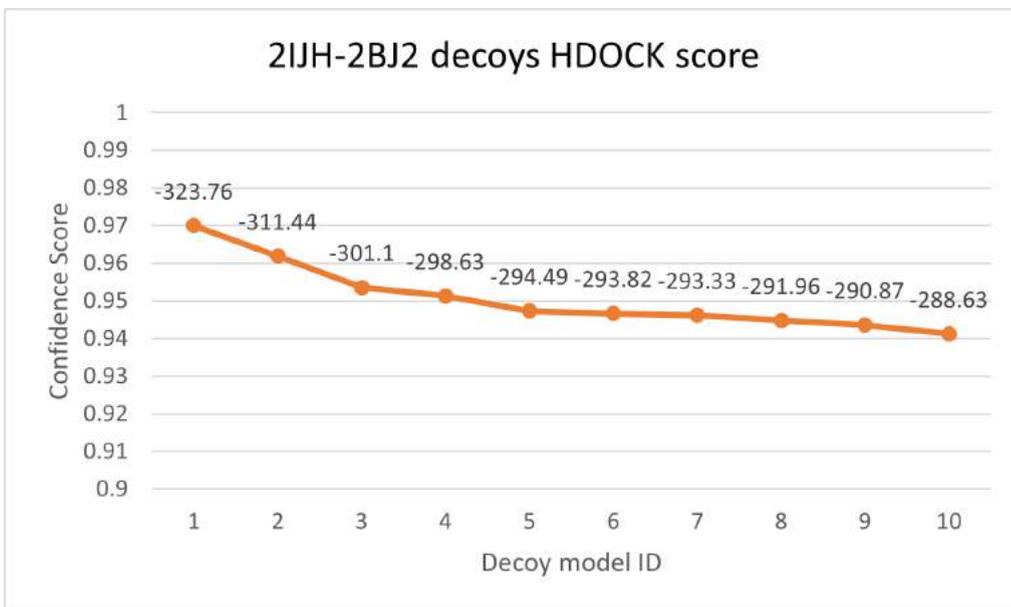


Figure 22| **HDOCK docking decoy confidence scoring** | [Crossover: 2IJH mutant - RNAI-RNAII kissing loop] (the data labels are depicting the intrinsic HDOCK scoring values) [35].

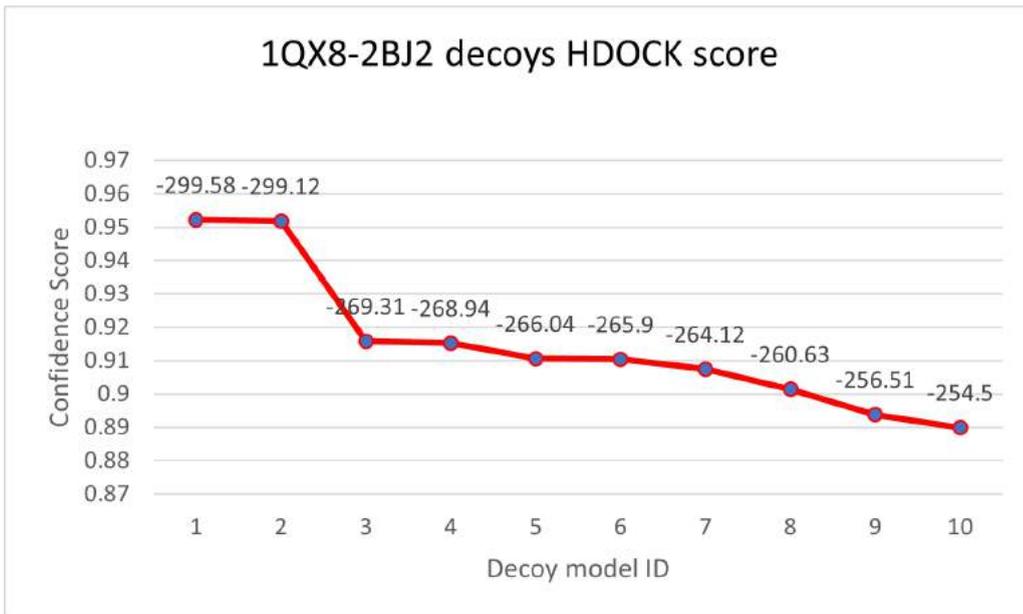


Figure 23| **HDOCK docking decoy confidence scoring** | [Crossover: 1QX8 mutant - RNAI-RNAII kissing loop] (the data labels are depicting the intrinsic HDOCK scoring values) [35].

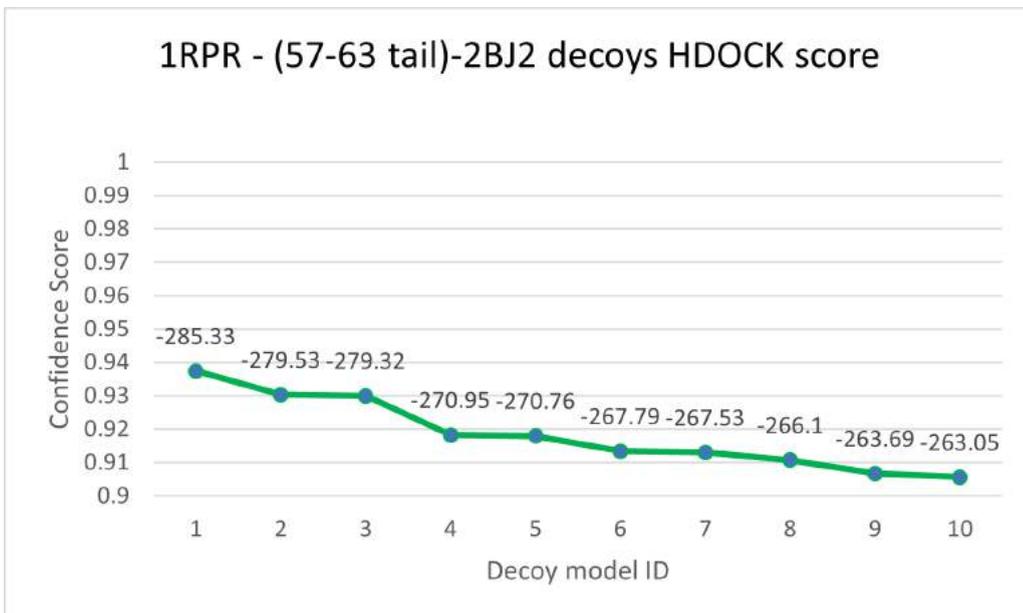


Figure 24| **HDOCK docking decoy confidence scoring** | [Crossover: ROP - (57-63 [residue]) tail - RNAI-RNAII kissing loop] (the data labels are depicting the intrinsic HDOCK scoring values) [35].

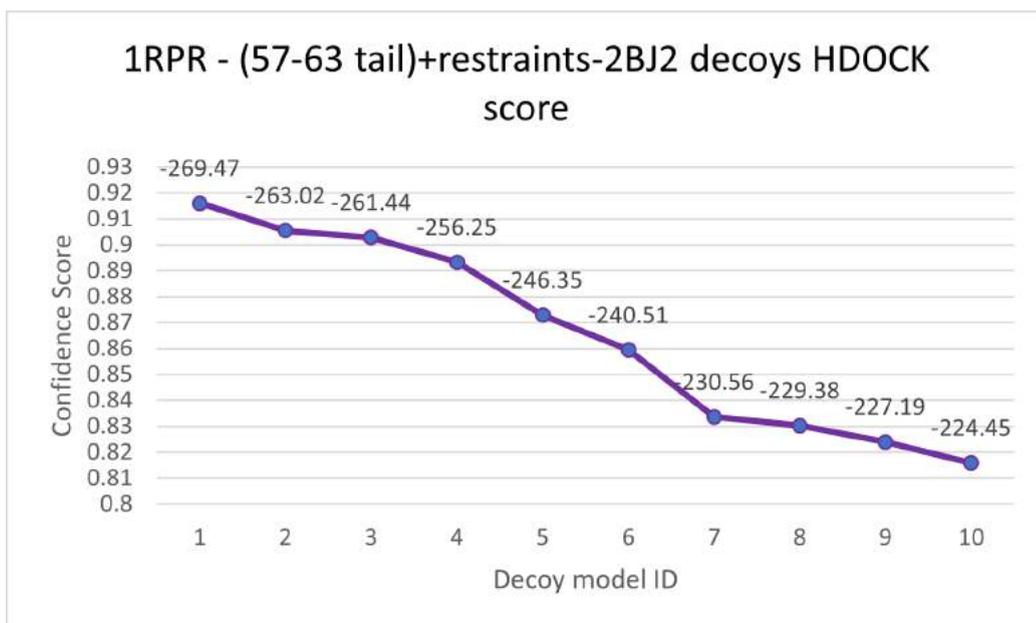


Figure 25| **HDOCK docking decoy confidence scoring** | [Crossover: ROP - (57-63 [residue]) tail (+restraints) - RNAI-RNAII kissing loop] (the data labels are depicting the intrinsic HDOCK scoring values) [35].

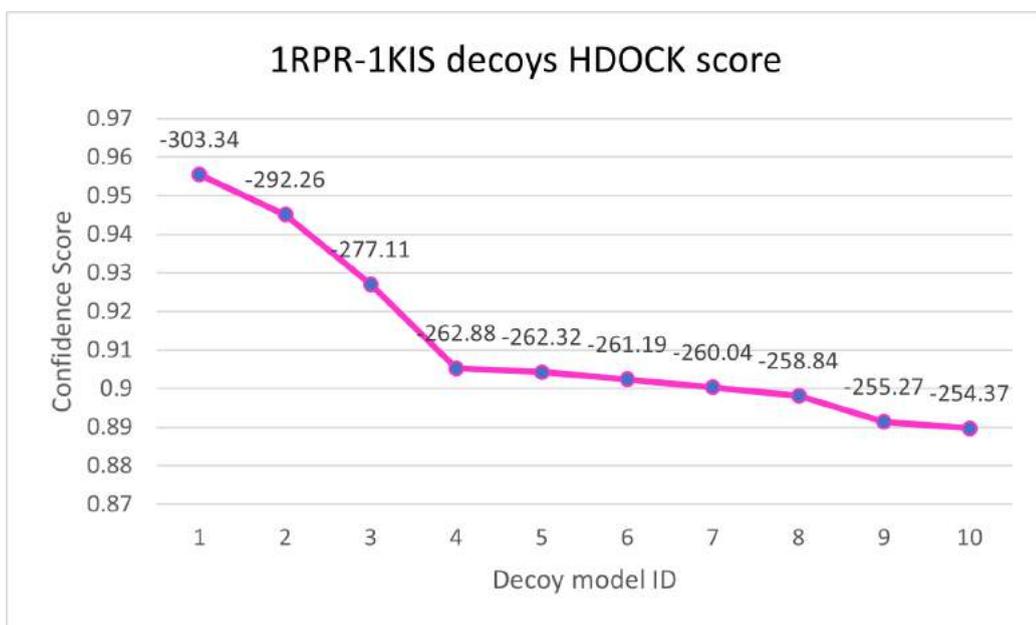


Figure 26| **HDOCK docking decoy confidence scoring** | [Crossover: ROP - TAR-TAR kissing loop] (the data labels are depicting the intrinsic HDOCK scoring values) [35].

3.1 Metrics

3.1.1 Δ^iG P-value

Δ^iG P-value is a metric for the decoy complex quality. This value provides an indication of the probability (P-value) that the Δ^iG calculated for the according docking decoy could not present a lower value in the case that the interface atoms are picked randomly from the molecular surface. It also works as a metric for complex specificity. As the P-value calculated is getting higher ($P \geq 0.5$), the probability of the complex formation being a result of an artifact is increasing. However, when the complex decoys present a Δ^iG P-value lower than 0.5 ($P < 0.5$), there is a strong indication of the structure's uniqueness and reliability. Thus, the Δ^iG P-value was used as a rough filter, to remove the low-quality decoys from the docking complex decoys pool of most likely for being the solution of the ROP-RNAI-RNA II kissing loop complex [42, 49].

In the graph shown below, all the docking decoys coming from the native ROP protein and the mutants docking on the RNAI - RNAII kissing complex are plotted. On the x axis the number of the model is indicated (starting from the model 1 to the model 10 according to the scoring of the docking protocol) and on the y axis the Δ^iG P-value for each of the decoys is shown. Δ^iG P-value is a metric that can only take values from 0 to 1.

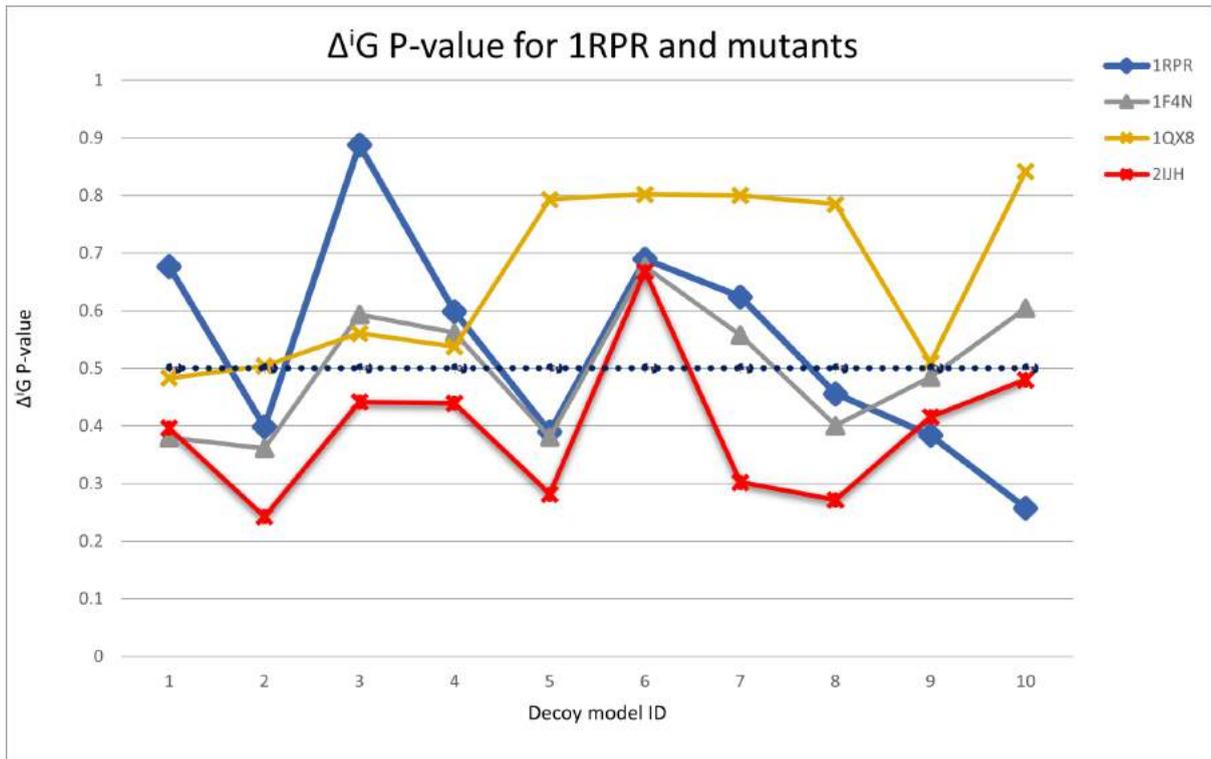


Figure 27| Δ^iG P-value for the first 10 decoys produced from the crossovers between the wild type ROP (1RPR) and its mutants with the RNAI-RNAII kissing loop.

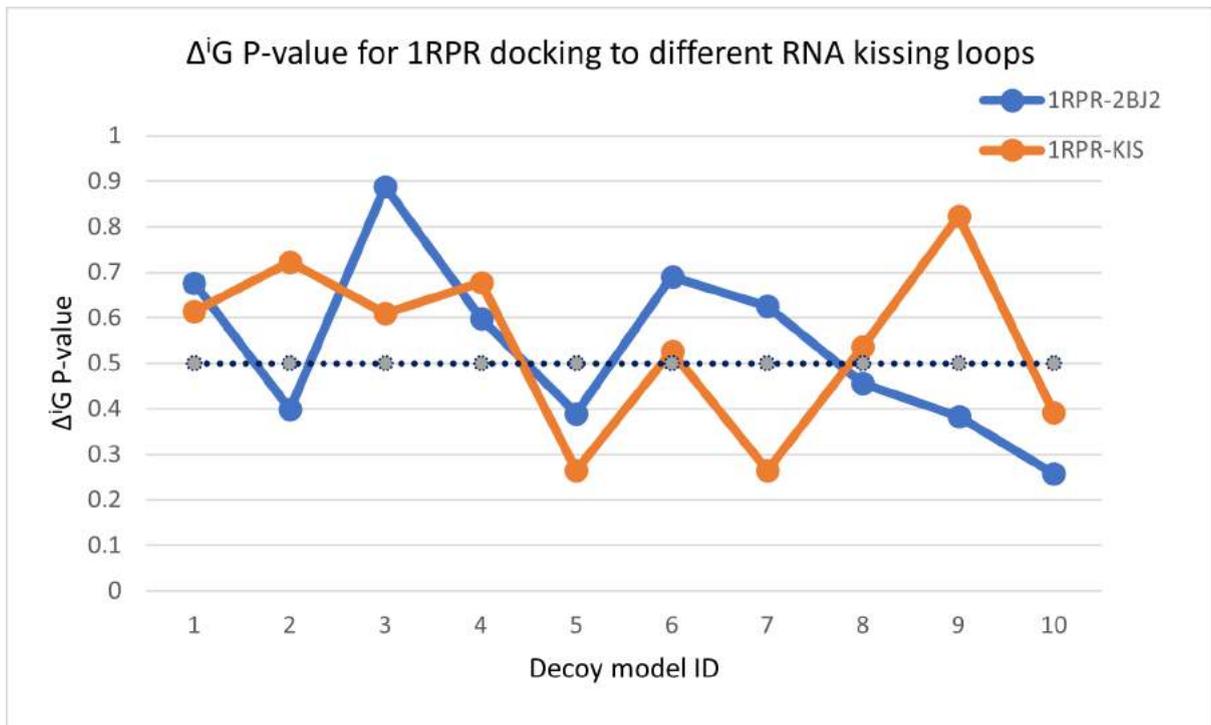


Figure 28| Δ^iG P-value for the first 10 decoys produced from the crossovers between the RNAI-RNAII kissing loop (2BJ2) and the TAR-TAR kissing loop (1KIS) with the wild type ROP (1RPR).

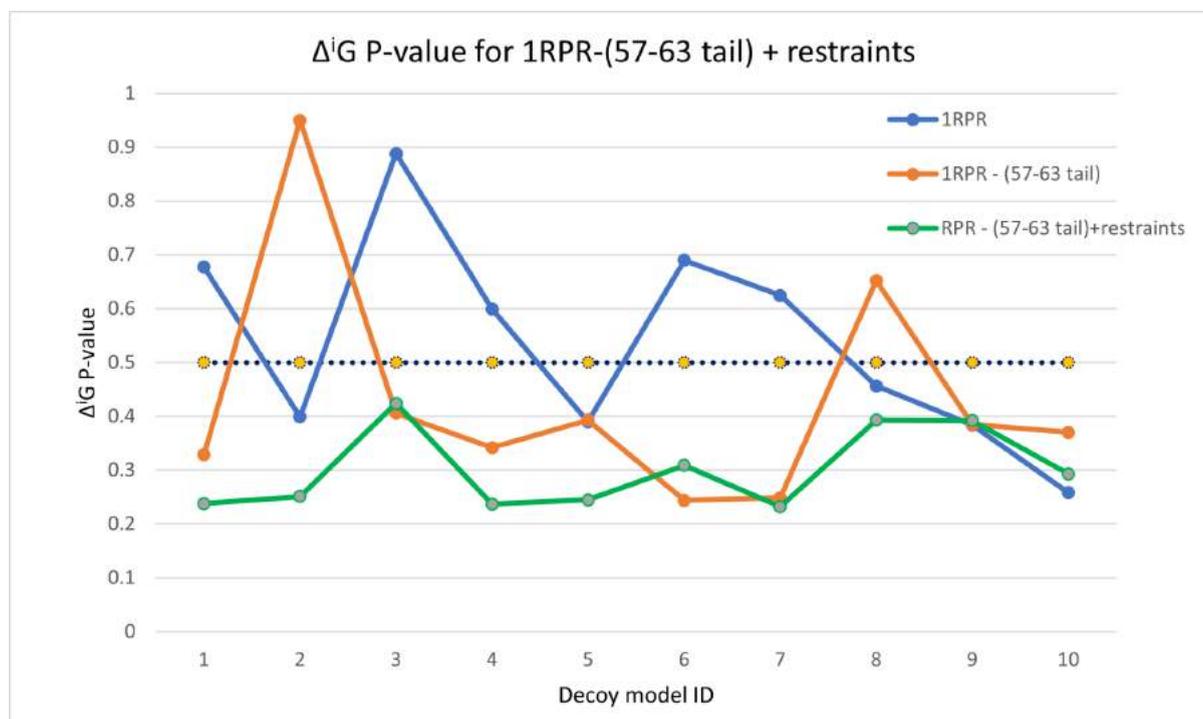


Figure 29| Δ^iG P-value for the first 10 decoys produced from the crossovers between the wild type ROP (1RPR) and the ROP - (57-63) tail with the RNAI-RNAII kissing loop (1 trial including additional residue parameters).

It is now obvious that some of the ROP mutants present a range of low quality decoys that can barely reach the Δ^iG P-value of 0.5. This result was expected for the mutants since, for example, the 1QX8 mutant of the ROP protein lacks the ROP's native activity and consequently the docking decoys' poor quality (due to inefficient protein-RNA binding) was expected [21, 23]. Moreover the docking results from the crossover between the ROP protein (1RPR) and the RNAI-RNAII kissing loop (2BJ2) present higher quality (5 models out of 10 have an accepted P-value) than those resulting from the crossover between the ROP protein and the TAR-TAR kissing loop (1KIS) (3 models out of 10 present accepted P-value). Finally from the

last P-value plot, it is shown that ROP structures missing the 57-63 tail gave higher quality docking structures than the unmodified ROP protein.

The fact that, the docking results coming from the docking attempts between the 1QX8 and the RNAI-RNAII kissing loop, as well as the decoys from the docking trial between the ROP protein and the TAR-TAR kissing loop are characterized from their low quality, is a good indicator of the evaluation program reliability. The 1QX8 ROP mutant is known for its complete loss of function and thus no binding upon the RNA kissing loop is expected. For the almost same reason, neither the ROP - TAR-TAR kissing loop is expected to present high quality docking decoys, since the nature-occurring ROP - RNA kissing loop complex is the ROP - RNAI-RNAII kissing loop for the ColEI plasmid replication regulation. Therefore, the docking models from the crossovers of the 1QX8 ROP mutant and those from the TAR-TAR kissing loop docking are suitable for negative controls of docking solution structure selection.

3.1.2 Solvation free energy

Solvation free energy gain is indicated by the symbol ΔiG and measured in kcal/M. This value is estimated from the solvation energy of the independent molecular components of the complex minus the solvation energy of the complex structure in constant temperature and density (or pressure). ΔiG depicts the energy that is needed to dissolve the molecular complex and consequently the more negative the ΔiG value is, the more stable the docking decoy is considered [44, 45]. The solvation free energy can therefore be implemented as the second evaluation metric for the docking decoys created. From the previous step of the decoys evaluation, many of the complexes were filtered out and thus the docking decoys now left for further assessment are the following colored in green:

ROP - 2BJ2	1RPR-(57-63 tail)- 2BJ2	1RPR-(57-63 tail) + restraints- 2BJ2	1F4N- 2BJ2	2IJH- 2BJ2
model 1	model 1	model 1	model 1	model 1
model 2	model 2	model 2	model 2	model 2
model 3	model 3	model 3	model 3	model 3
model 4	model 4	model 4	model 4	model 4
model 5	model 5	model 5	model 5	model 5
model 6	model 6	model 6	model 6	model 6
model 7	model 7	model 7	model 7	model 7
model 8	model 8	model 8	model 8	model 8
model 9	model 9	model 9	model 9	model 9
model 10	model 10	model 10	model 10	model 10

Table 2| *The docking decoys with acceptable P-values that can be used for further analysis.*

3.1.3 Interface Area

The interface area is an important metric for the evaluation of molecular complexes. Most of the time it is calculated in Å² and it is typically considered as the difference between the total accessible surface areas (ASA) of each of the interacting molecules minus the (ASA) of the interfacing structures, divided by two. Interface Area is also correlated with the Buried Surface Area created upon the complex formation. The value of the Buried Surface Area, indicates the part of solvent accessible area of the molecule or the residue that is buried after the complex formation and it is measured in percentages or in Å². There is an inconvenience in the relevant literature regarding the Buried Surface Area (BSA), however it is a widely used metric from molecular complexes' structural analysis. However, in both cases, a complex is considered as more stable when characterized

with high BSA and Interface Area values combined with a low negative value for the Solvation free energy (ΔiG) [46, 47, 48, 49].

The analysis results of the docking decoys prepared from the PDBePISA and the COCOMAPS server are almost similar. This fact is a good indicator of the docking quality although the docking decoys were modified before their introduction to the PDBePISA server as to continuous polypeptide and nucleic acid chains. This step was required since the PDBePISA does not provide the option of selecting the interface of interest and analyzing each polypeptide and nucleic acid chain separately, although a whole protein may be composed of several peptide and nucleic acid chains (polymere). From the chart graphs for the example cases shown below (one graph is referring to the wild type ROP interface area (\AA^2) with the 2BJ2 RNA kissing loop and the other one is based on the interface area values of the ROP lacking the 57-63 [residue] tail) docking with the RNAI-RNAII kissing loop) it is shown that especially for the second case, the according Interface Area values are very close.

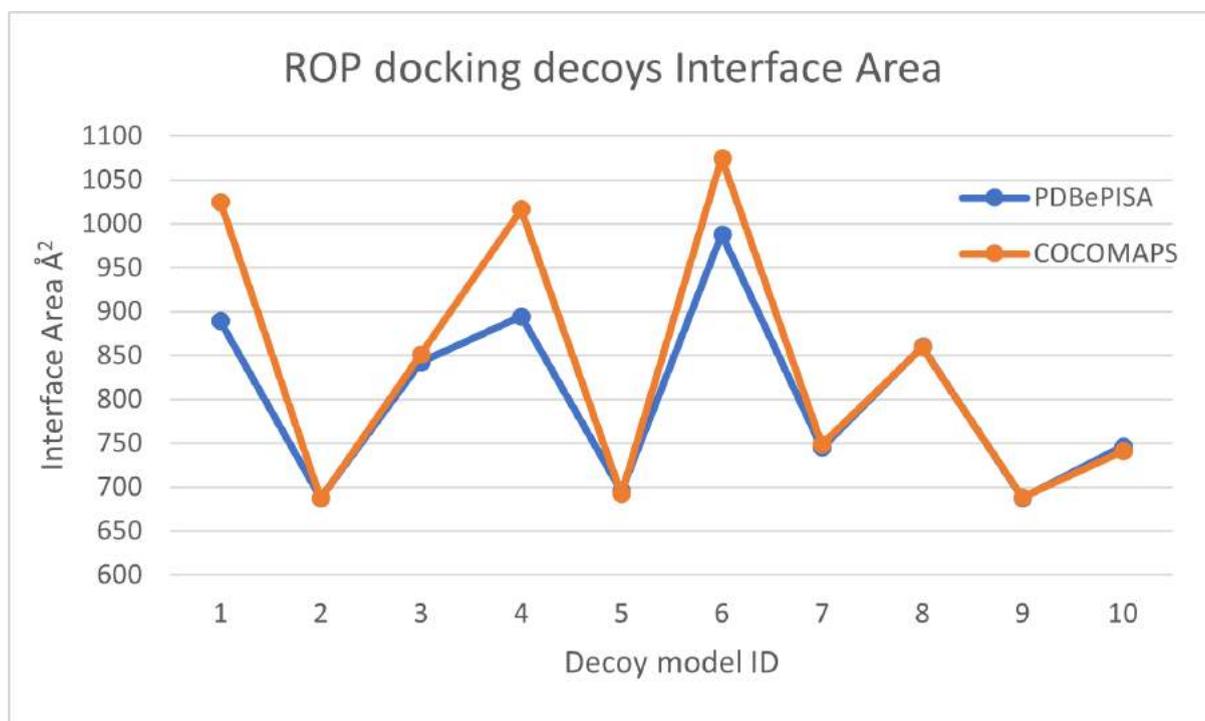


Figure 30| Comparison graph of the Interface Area calculated from the PDBePISA and the COCOMAPS server for the docking decoys of the ROP protein (1RPR) and the RNAI-RNAII kissing loop docking. It is evident that the results do not comply with each other.

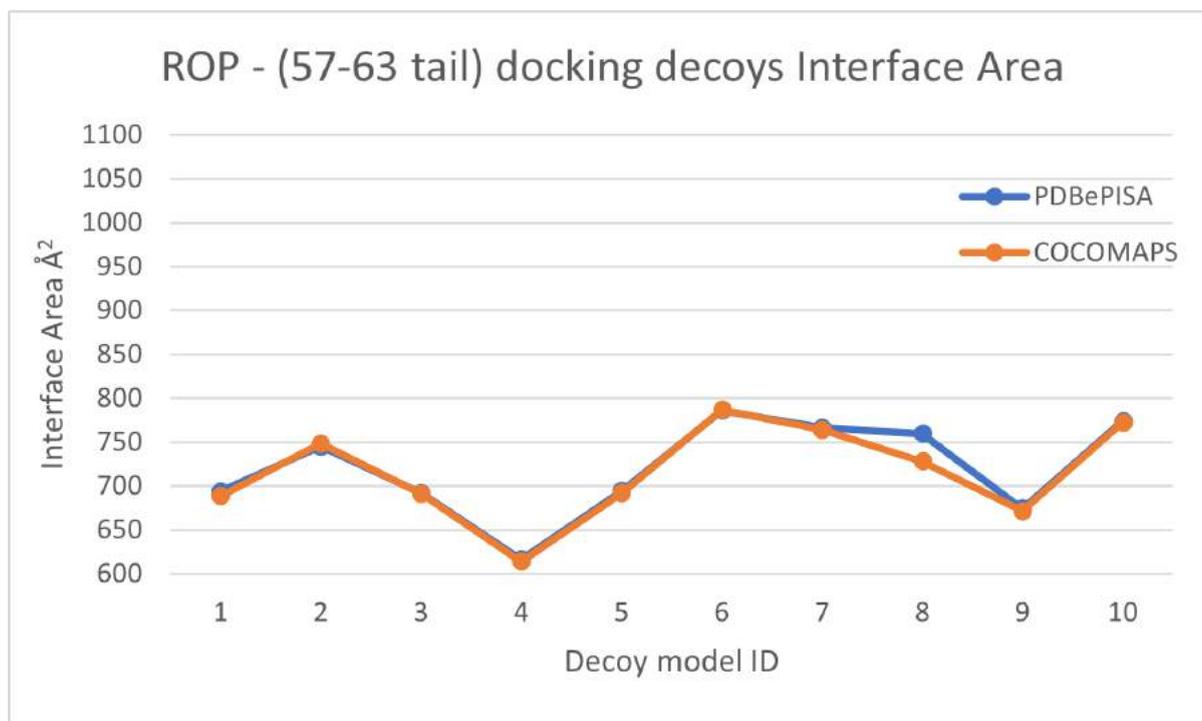


Figure 31| Comparison graph of the Interface Area calculated from the PDBePISA and the COCOMAPS server for the docking decoys of the ROP protein (1RPR) missing the 57-63 [residue] tail and the RNAI-RNAII kissing loop docking. The results prepared from the two different servers are consistent.

From the graphs shown above, it can be noticed that the Interface area values are more consistent when it comes to the ROP - (57-63 [residue]) tail docking decoys. This is probably a proof of the theory that the tails for the 57th till the 63rd residues of the ROP protein can disturb the docking results. However, this issue will be inspected later in the present thesis.

Moreover, in the case of the ROP interface area values calculated from both structural analysis servers, it is shown that the error bars for the values measured by the COCOMAPS server are wider. This fact in combination with the frequent outlier values calculated from the COCOMAPS server contributes to the choice of the PDBePISA server for the implementation of further analysis. Below, the graphs are used to highlight the error bars of the Interface Area values calculated from the COCOMAPS server.

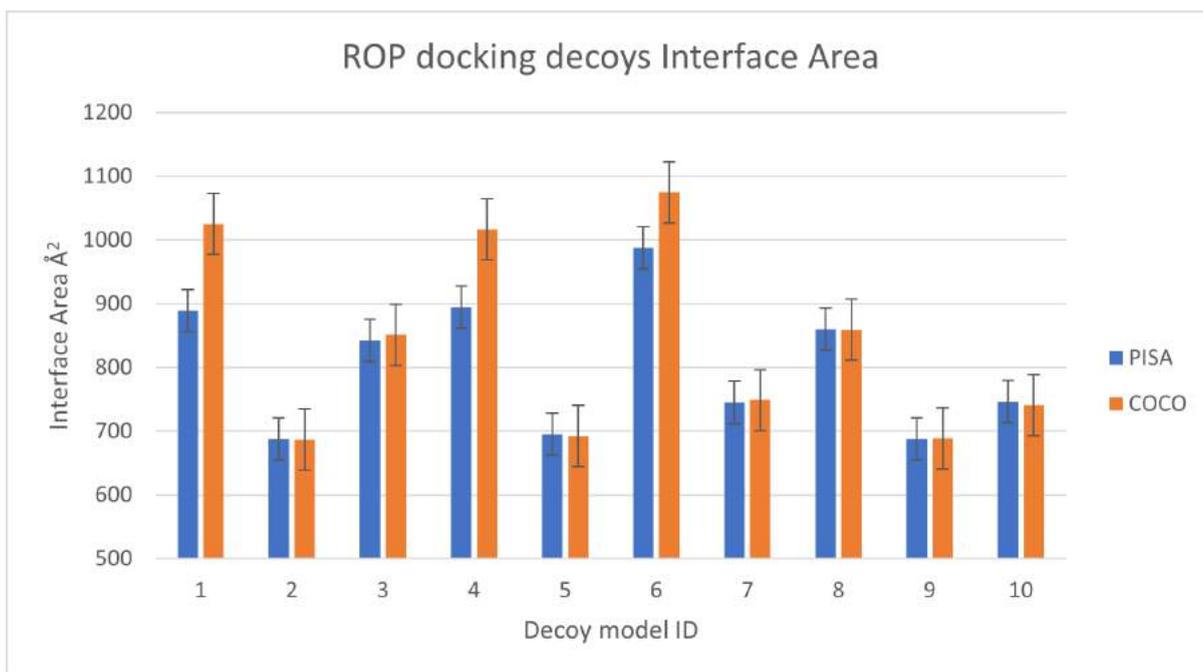


Figure 32| Comparison column graph (with error bars) of the Interface Area calculated from the PDBePISA and the COCOMAPS server for the docking decoys of the ROP protein (1RPR) and the RNAI-RNAII kissing loop docking.

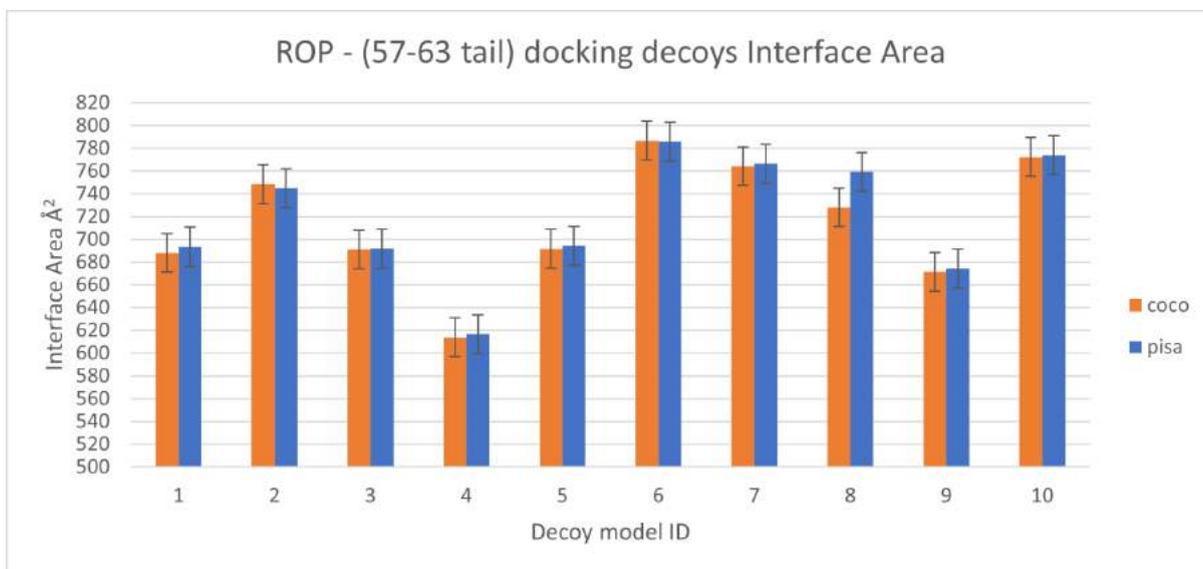


Figure 33| Comparison column graph (with error bars) of the Interface Area calculated from the PDBePISA and the COCOMAPS server for the docking decoys of the ROP protein (1RPR) missing the 57-63 [residue] tail and the RNAI-RNAII kissing loop docking.

Considering the aforementioned data it becomes evident that the ROP protein missing the 57-63 [residue] tail in crossover with the RNAI-RNAII kissing loop docking decoys as well as those generated from the crossover between the 2IJH ROP mutant and the 2BJ2, are those that present the most stable and consistent picture. At first, most of the top 10 docking decoys (8 out of 10 structures) and (9 out of 10) accordingly, have an accepted P-value (lower than 0.5) compared to the accepted decoys coming from the wild type ROP (5 out 10 structures are accepted) and the 1F4N mutant (5 out of 10 accepted structures). Furthermore, it was shown that the Interface Area measurements calculated from PDBePISA and from COCOMAPS servers come to agree with each other for the case of the 1RPR - (57-63 [residue]) tail with the RNAI-RNAII kissing loop.

Additionally, for the case of the ROP missing its tail, and the 2IJH ROP mutant, after filtering the structures that present high P-values, the data becomes even more consistent in contrast to the non-sense results of the 1F4N mutant and the wild type ROP protein dockings. As it is shown from the following diagrams, for the ROP - (57-63 [residue] tail) decoys, as the Interface Area score is elevating, the ΔG value becomes more negative, indicating the stability of the complexes when the interface area is extended. However, the same happens with the Interface Area and ΔG values for the 2IJH ROP mutant docking decoys which is not an expected result, since the binding ability of this protein upon the RNAI-RNAII kissing loop is absent. Thus, additional investigation is needed for this test case.

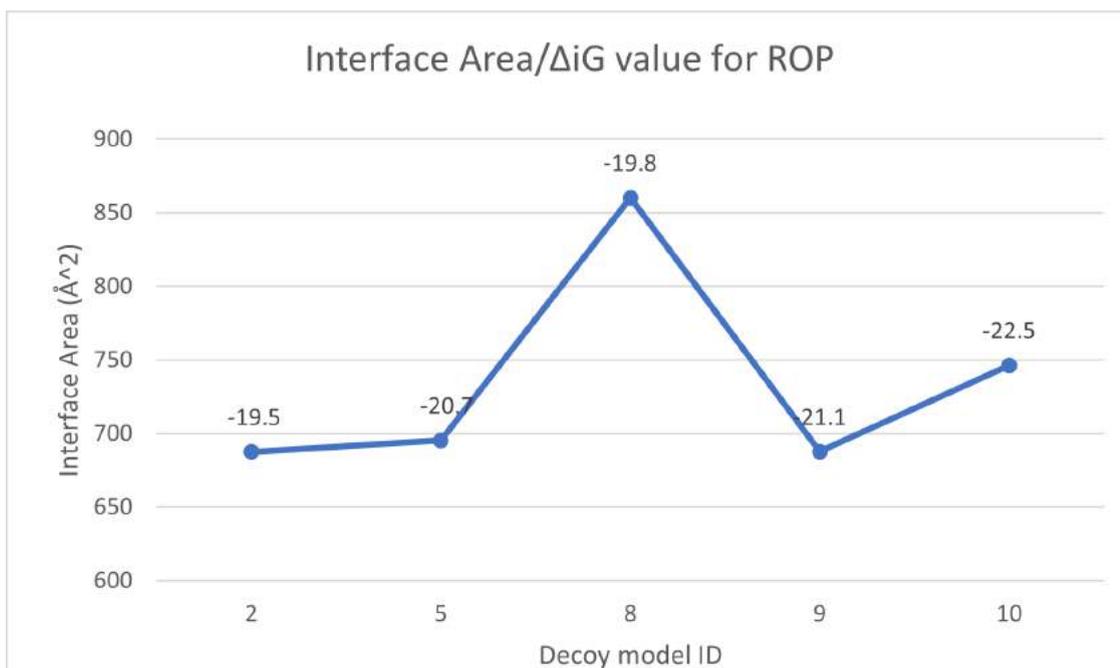


Figure 34| *Interface Area (Å²) to ΔiG (kcal/M) for the filtered decoys (P-value<0.5) of the docking between the ROP protein and the RNAI-RNAII kissing loop.*

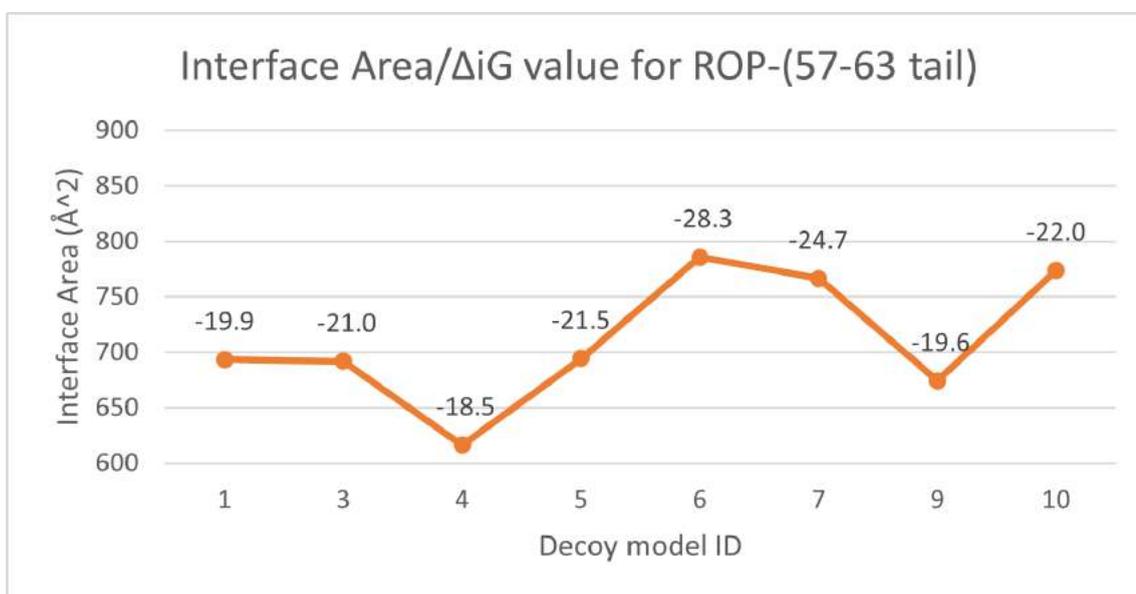


Figure 35| *Interface Area (Å²) to ΔiG (kcal/M) for the filtered decoys (P-value<0.5) of the docking between the ROP - (57-63) tail protein and the RNAI-RNAII kissing loop.*

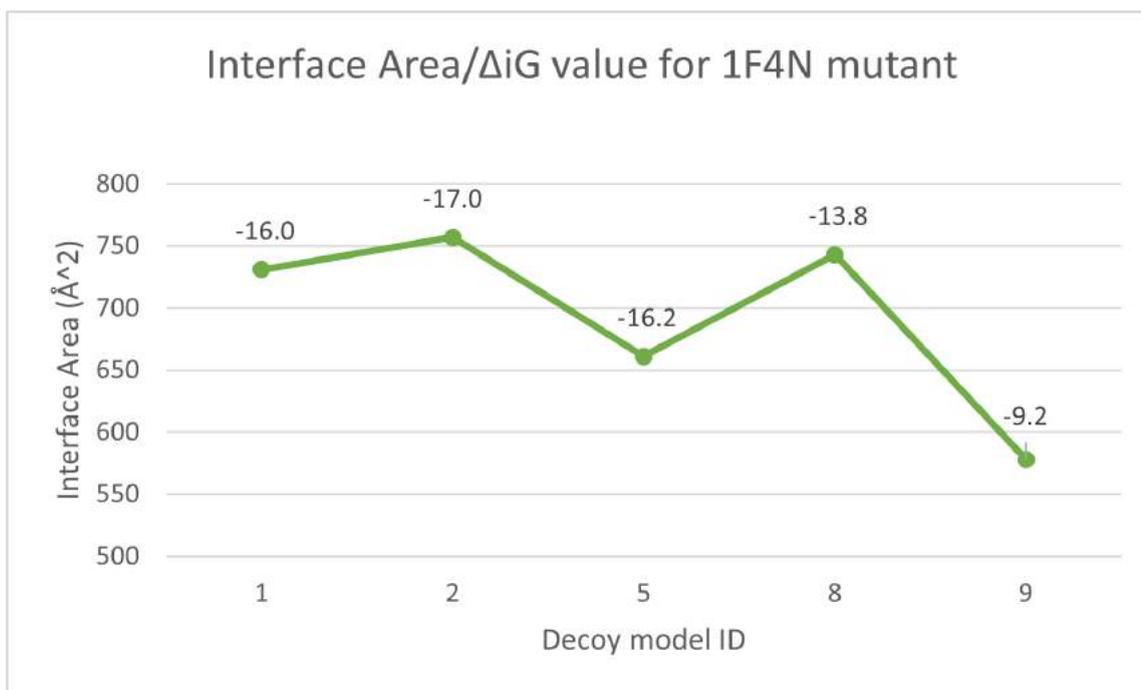


Figure 36| *Interface Area (Å²) to ΔiG (kcal/M) for the filtered decoys (P-value<0.5) of the docking with between the 1F4N ROP protein mutant and the RNAI-RNAII kissing loop.*

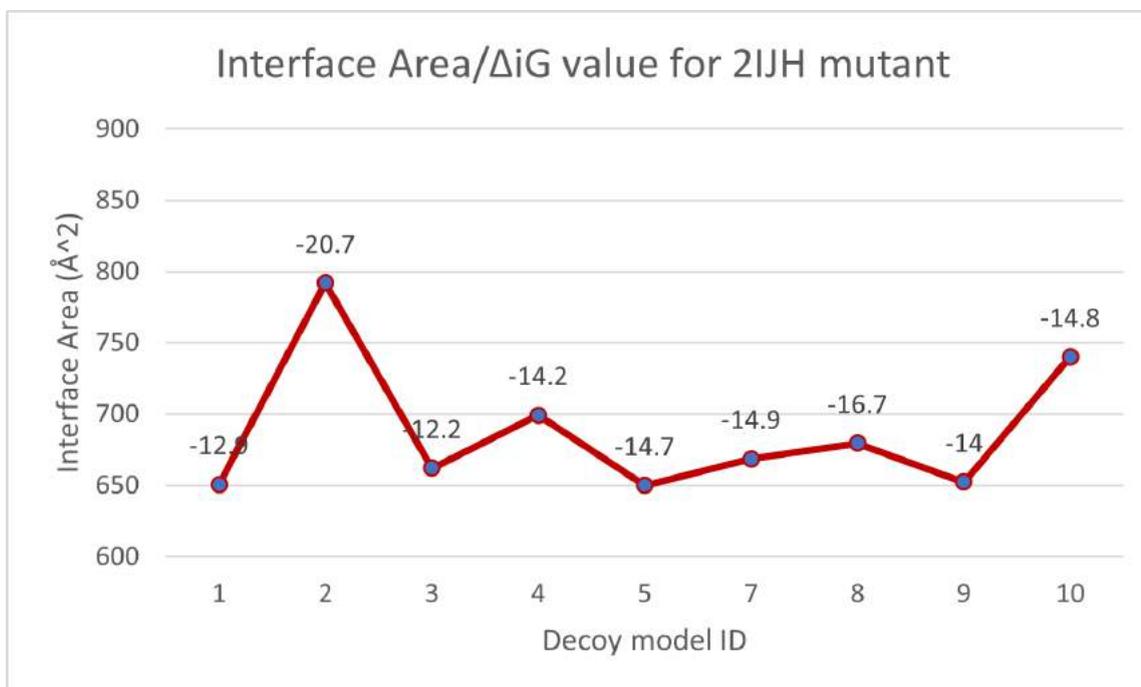


Figure 37| *Interface Area (Å²) to ΔiG (kcal/M) for the filtered decoys (P-value<0.5) of the docking with between the 2IJH ROP protein mutant and the RNAI-RNAII kissing loop.*

The above graphs can help decide on the docking decoys for further analysis, since they provide a straight indication of the docking results quality. The graphs that present a stable picture of the Interface Area and the ΔiG value of the produced docking models are; the one referring to the docking between the ROP - (57-63) tail protein and the RNAI-RNAII kissing loop and the last one for the analysis of the 2IJH mutant - 2BJ2 crossover and thus, these docking decoys are the most reliable. More specifically, the model 6 from the crossover case the ROP protein - (57-63) tail protein and the RNAI-RNAII kissing loop and the model 2 from the 2IJH mutant with the RNAI-RNAII kissing loop crossover are suggested for additional analysis.

3.2 Interface residues evaluation

All the metrics that have been already mentioned regarding the complex structure evaluation coming from the docking trials, can be applied to a residue scale. The Solvent Accessible Surface Area as well as the Buried Surface Area and the Interface Area can be calculated for each protein residue of nucleotide separately. Most of the programs for molecular interface structural analysis provide this function to the user. In this case, the PDBePISA output page contains a detailed map of the interactions in a residue/nucleotide-base. An example part of the map is shown in the picture below [49]:

Interfacing residues (not a contact table) Display level:

Inaccessible residues HSDC Residues making Hydrogen/Disulphide bond, Salt bridge or Covalent link
Solvent-accessible residues Interfacing residues

ASA Accessible Surface Area, Å² **BSA** Buried Surface Area, Å² **ΔG** Solvation energy effect, kcal/mol ||||| Buried area percentage, one bar per 10%

#	Structure 1	HSDC	ASA	BSA	ΔG	#	Structure 2	HSDC	ASA	BSA	ΔG
1	A:MET 1		201.66	0.00	0.00	1	C: G 1		265.82	0.00	0.00
2	A:THR 2		77.02	0.00	0.00	2	C: G 2		187.90	0.00	0.00
3	A:LYS 3		132.91	0.00	0.00	3	C: C 3		175.29	0.00	0.00
4	A:GLN 4		68.84	0.00	0.00	4	C: A 4		180.64	2.77	0.06
5	A:GLU 5		17.89	0.00	0.00	5	C: A 5	H	179.64	96.82	3.11
6	A:LYS 6		142.91	0.00	0.00	6	C: C 6		183.04	88.85	1.68
7	A:THR 7		67.36	34.95	0.55	7	C: G 7		140.08	45.51	2.17
8	A:ALA 8		5.70	0.00	0.00	8	C: G 8		119.79	39.45	1.13
9	A:LEU 9		9.38	0.00	0.00	9	C: A 9	H	162.81	82.91	2.86
10	A:ASN 10		87.07	52.62	-0.41	10	C: U 10	H	174.99	92.12	1.03
11	A:MET 11		26.56	25.05	0.39	11	C: G 11		192.93	8.84	0.37
12	A:ALA 12		0.12	0.00	0.00	12	C: G 12		186.46	0.00	0.00
13	A:ARG 13	H	92.02	32.02	0.23	13	C: U 13		164.38	0.00	0.00
14	A:PHE 14		100.04	99.42	1.59	14	C: U 14		175.15	0.00	0.00
15	A:ILE 15		0.17	0.00	0.00	15	C: C 15		198.79	0.00	0.00
16	A:ARG 16		52.59	0.00	0.00	16	C: G 16		181.87	0.00	0.00
17	A:SER 17		39.01	32.32	0.40	17	C: U 17		169.34	0.00	0.00
18	A:GLN 18		18.50	9.15	0.04	18	C: U 18		162.77	0.00	0.00
19	A:THR 19		0.49	0.00	0.00	19	C: G 19		179.56	0.00	0.00
20	A:LEU 20		94.90	23.28	0.37	20	C: C 20		170.16	0.00	0.00
21	A:THR 21	H	56.51	43.47	0.25	21	C: C 21		257.43	0.00	0.00
22	A:LEU 22		1.81	0.00	0.00	22	C: G 22		265.57	0.00	0.00
23	A:LEU 23		51.71	0.00	0.00	23	C: C 23		185.58	0.00	0.00
24	A:GLU 24		67.02	6.87	-0.11	24	C: A 24		188.30	0.00	0.00
25	A:LYS 25	H	55.00	42.82	0.49	25	C: C 25		174.35	0.00	0.00

Figure 38| *Interface Analysis Results page PDBePISA [49].*

The map presents the ASA, BSA, the hydrogen bond that is likely to be formed upon the complex creation, as well as the Δ_iG value for each of the residues of the ROP protein and the nucleotides of the RNA kissing loop. These tables can be used to pursue further information regarding the molecular interface formation. The

graphs shown below, depict the Buried Surface Area values for each of the protein residues and the nucleotides of the RNA kissing loop independently. In this case, the BSA value is referred to the solvent-accessible surface area of the corresponding residue that is now buried due to the molecular interface formation between the ROP -(57-63) tail protein and the RNA kissing complex measured in Å². The color code shown in the graphs is an indicator of the percentage of the total solvent-accessible surface area that is buried upon the complex formation. The columns colored red indicate the residues/nucleotides that lose more than the 50% of their solvent accessible area that is now buried from the complex formation [49].

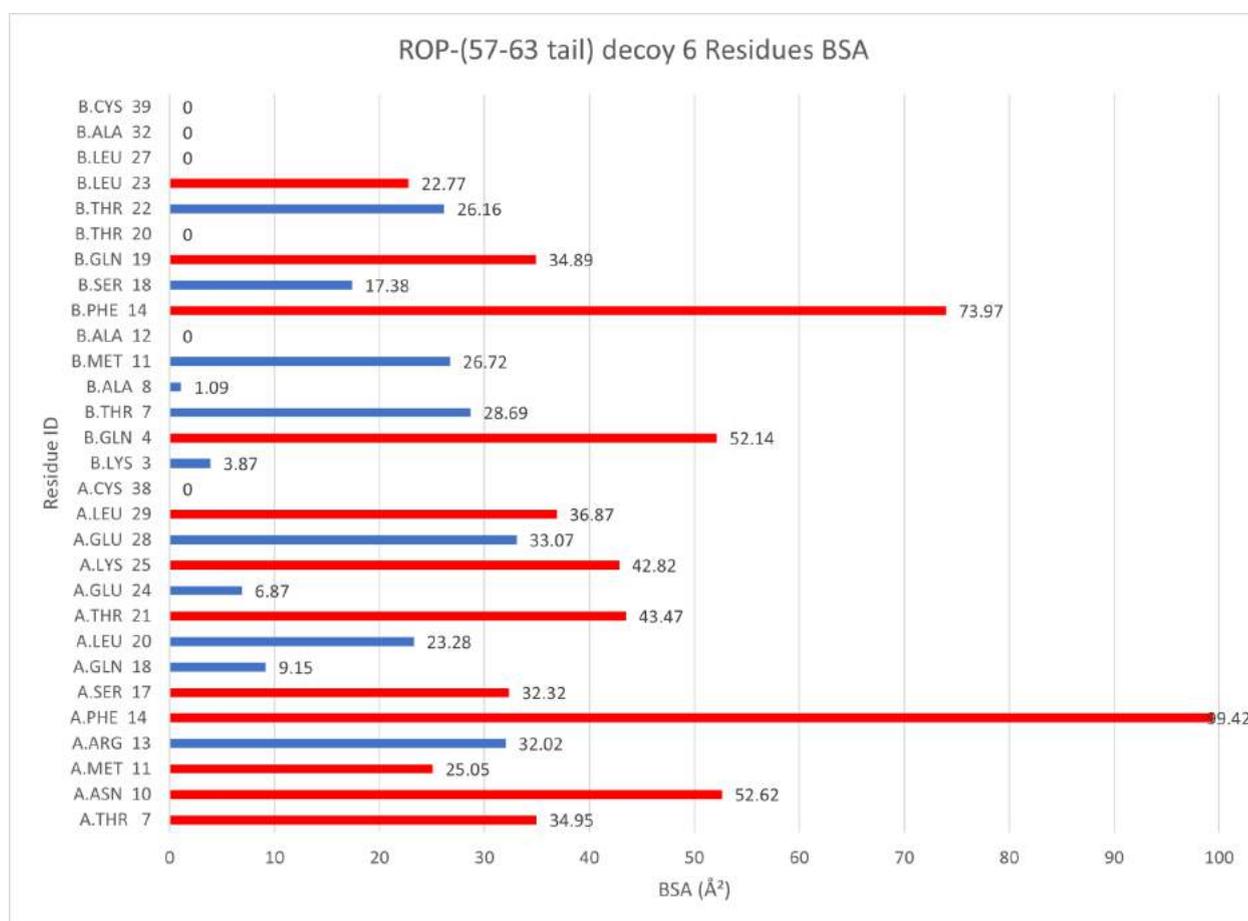


Figure 39| Buried Surface Area (Å²) table for each of the protein residues contributing to the protein - RNA kissing loop interface formation [Crossover: ROP - (57-63) tail - RNAI-RNAII kissing loop, Decoy: 6]

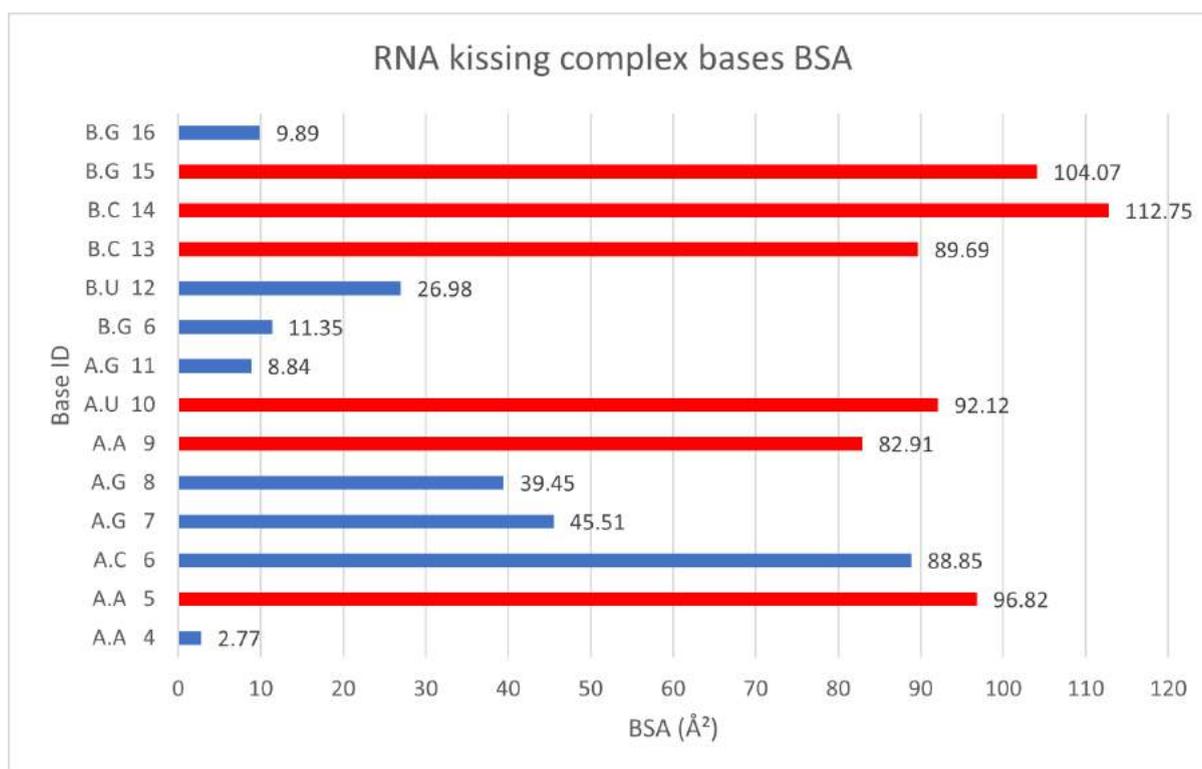


Figure 40| *Buried Surface Area (Å²) table for each of the ribonucleotide bases contributing to the protein - RNA kissing loop interface formation [Crossover: ROP - (57-63) tail - RNAI-RNAII kissing loop, Decoy: 6]*

Studying the interface characteristics in a residue-scale for the docking decoy 2 resulting from the crossover [2IJH - 2BJ2], it is getting clear that the decoy generated from this crossover are not reliable. More specifically, the protein residues that are shown to play a major role in the protein mutant - RNA kissing loop formation are the residues Trp-14 from both protein chains. From the literature, it is known that the ROP protein residue that is determinant for the protein binding upon the RNA kissing loop, is the Phe-14 residue from both protein polypeptide chains. This is exactly the residue that is replaced in the 2IJH mutant, leading to the total loss of protein ability to bind to the RNAI-RNAII kissing loop. As it is highlighted below (the A.Trp-14, B.Trp-14 bars are colored in magenta), the Trp-14 residues are claimed to have a principal role for the complex formation, indicating the inaccuracy of the docking results of this test case.

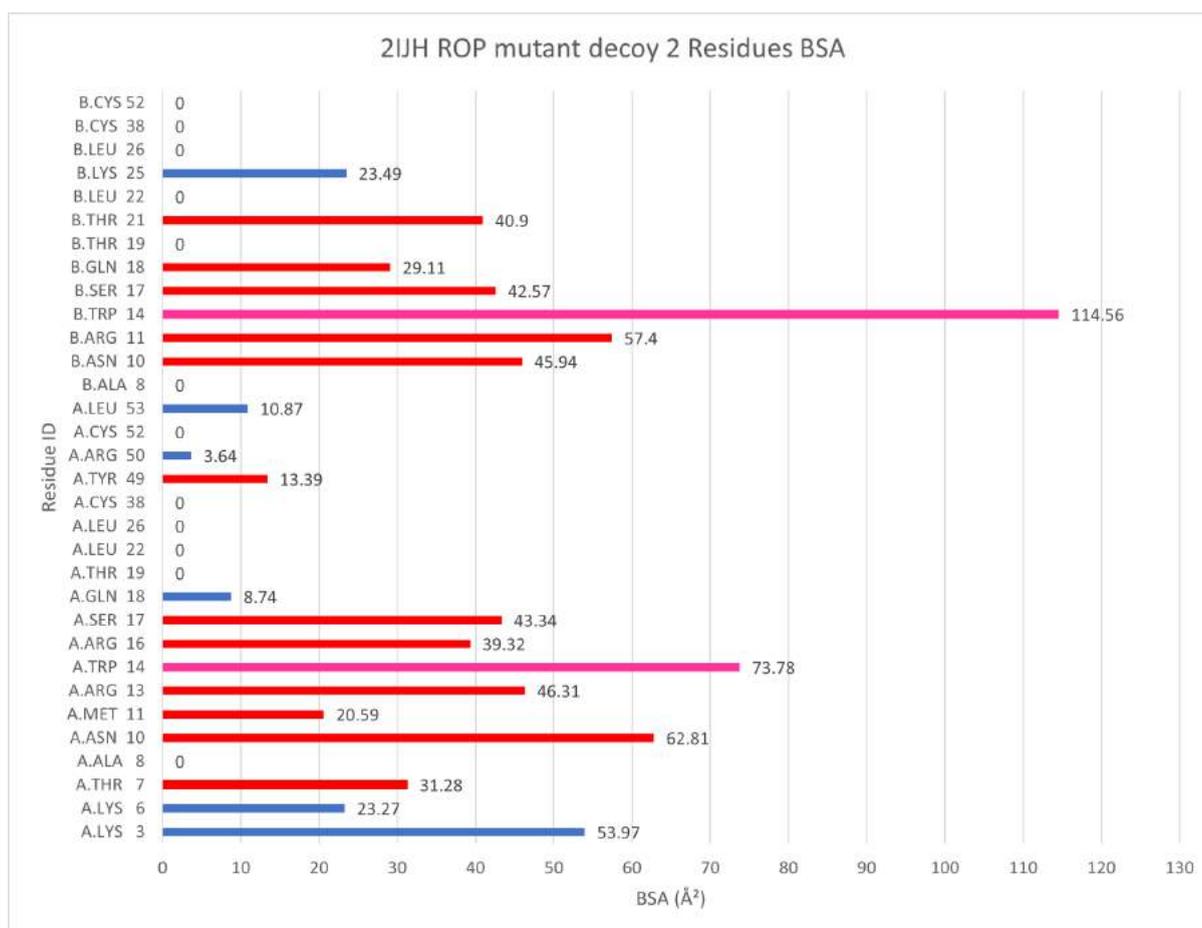


Figure 41| *Buried Surface Area (Å²) table for each of the protein residues contributing to the protein - RNA kissing loop interface formation [Crossover: 2IJH - RNAI-RNAII kissing loop, Decoy: 2]*

Taking into account all the data that have been analyzed till now, the docking decoy that is more possible to resemble the naturally occurring complex between the ROP protein and the RNAI-RNAII kissing loop in bacteria in the 6th model from the crossover: *ROP - (57-63) tail - RNAI-RNAII kissing loop*. The characteristics of this decoy structure are the following, as well as the 3D structure shown below, indicating the protein residues that more than the 50% of their solvent accessible area is buried after the complex formation, in red and the RNA kissing loop residues in cyan accordingly.

<i>Metrics</i>	<i>Decoy 6 [crossover: ROP - (57-63) tail - RNAI-RNAII kissing loop]</i>
<i>P-value</i>	<i>0.244</i>
<i>Delta<i>G</i> (kcal/M)</i>	<i>-28.3</i>
<i>Interface Area (Å²)</i>	<i>785.8</i>
<i>Hydrogen Bonds (N)</i>	<i>5</i>

Table 3| *Features of the Docking Decoy 6 resulting from the crossover of the ROP - (57-63) tail - RNAI-RNAII kissing loop.*

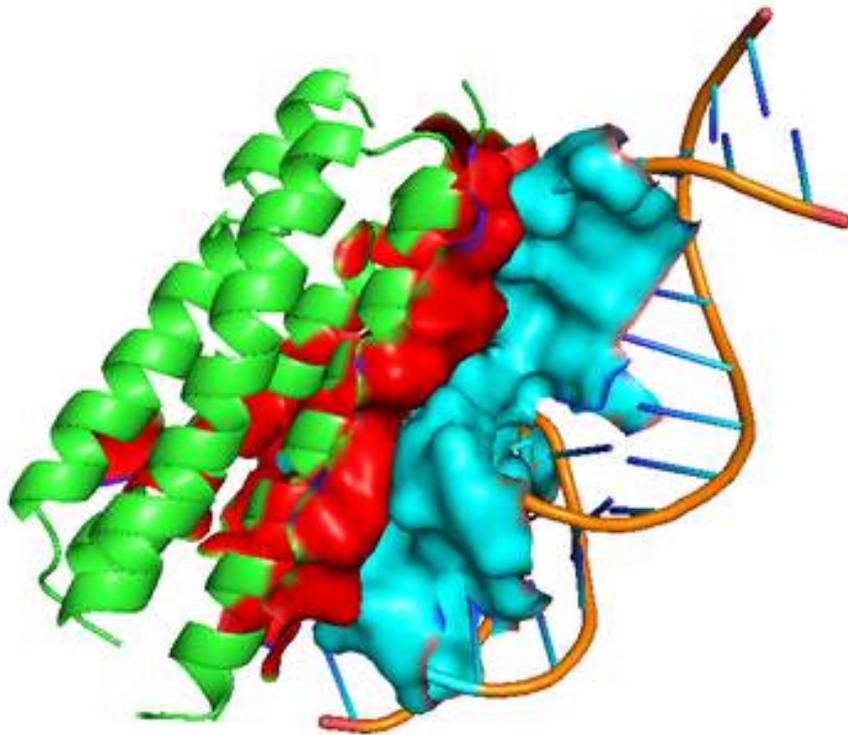


Figure 42| *The docking decoy 6 [Crossover: ROP - (57-63) tail - RNAI-RNAII kissing loop]*
(image created by PYMOL [31, 35])

3.3 Introducing docking parameters

As it has already been established, the HDOCK server allows for restraints introduction to the docking process [35]. These particular parameters introduced for the ligand part, were selected according to the publication for the ROP protein - TAR-TAR kissing loop complex structural analysis (Comolli, L. R., Pelton, J. G., & Tinoco Jr, I. 1998) and they are the following:

I.A[9]:II.U[12], I.U[10]:II.A[11], I.G[11]:II.C[9], I.G[12]:II.C[10]

3.3.1 Results

For this docking trial, the ROP - (57-63) tail protein was used, since it has already shown a consistent and stable picture regarding docking results. The following graph is depicting the the P-values for the Solvation free energy gain calculated for the docking decoys resulting from the crossover between the ROP - (57-63 [residue]) tail protein and the RNAI-RNAII kissing loop in combination with the additional parameter of the specific binding restrains mentioned above:

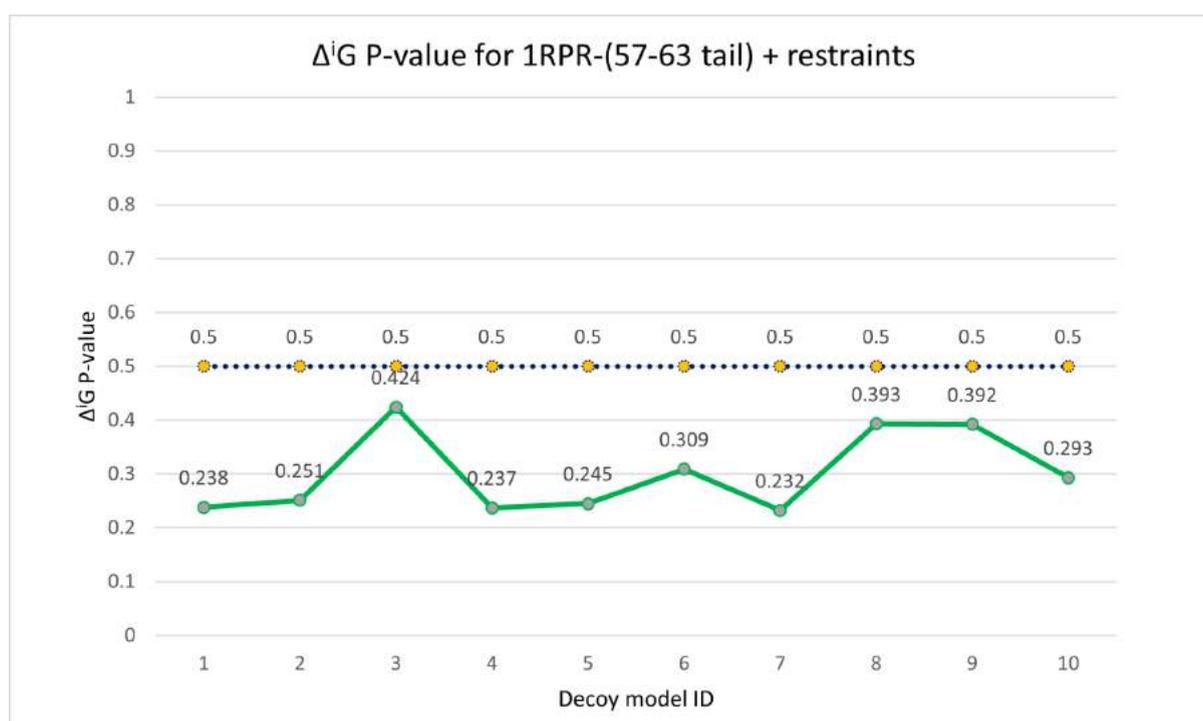


Figure 43| ΔiG P-values of docking decoys for the Crossover: ROP - (57-63) tail - RNAI-RNAII kissing loop with additional residue restraints.

It is obvious that all of the suggested models from the HDock server for further analysis present an accepted P-value for the solvation free energy for this study case. Thus, the interface area and the solvation free energy are now plotted as shown below:

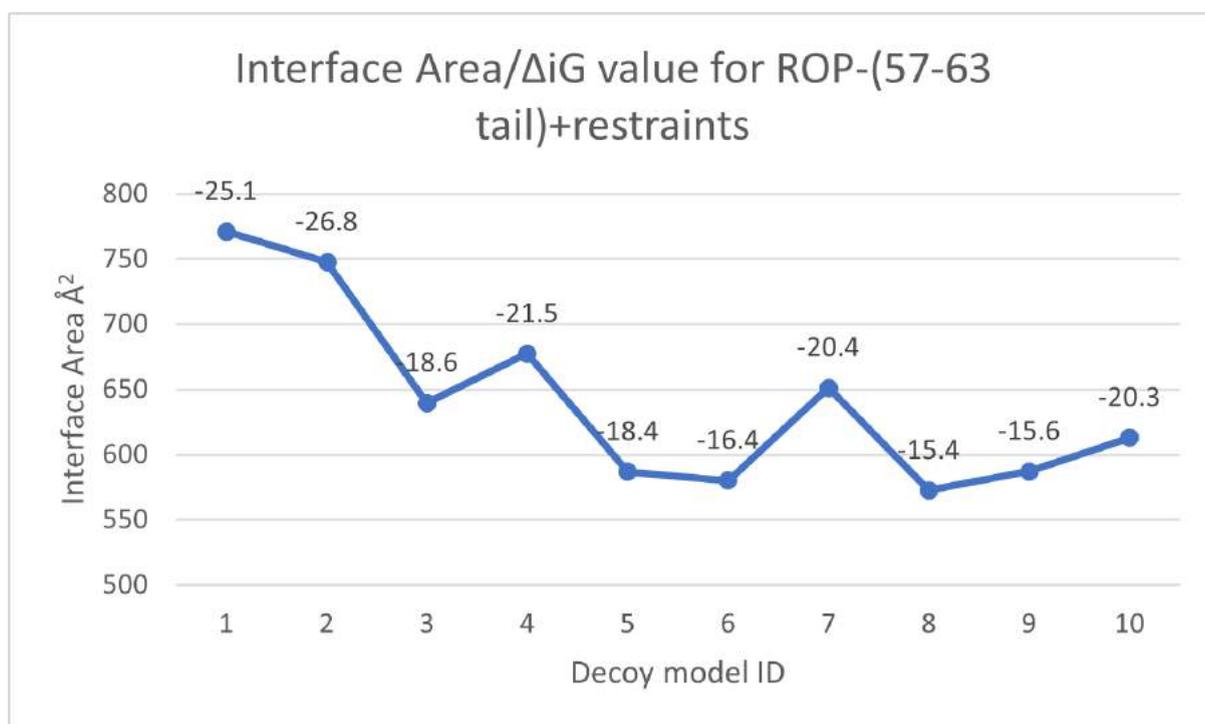


Figure 44| Interface Area (\AA^2) to ΔiG (kcal/M) for the filtered decoys (P -value <0.5) of the docking with additional restraints between the ROP - (57-63) tail protein and the RNAI-RNAII kissing loop.

<i>Docking Decoy 1</i>	
<i>P-value</i>	<i>0.238</i>
<i>Delta<i>G</i> (kcal/M)</i>	<i>-25.1</i>
<i>Interface Area (Å²)</i>	<i>771.1</i>
<i>Hydrogen Bonds (N)</i>	<i>4</i>

Table 4| *Features of the decoy 1 created from the docking crossover ROP - (57-63 [residue] tail - RNA kissing loop with the additional docking restraints for the ligand (RNA kissing loop).*

<i>Docking Decoy 2</i>	
<i>P-value</i>	<i>0.251</i>
<i>Delta<i>G</i> (kcal/M)</i>	<i>-26.8</i>
<i>Interface Area (Å²)</i>	<i>747.6</i>
<i>Hydrogen Bonds (N)</i>	<i>3</i>

Table 5| *Features of the decoy 2 created from the docking crossover ROP - (57-63 [residue] tail - RNA kissing loop with the additional docking restraints for the ligand (RNA kissing loop).*

In this case, the 1st model is the one with the best features regarding the molecular Interface Area score (higher value) and the solvation free energy gain upon complex formation (most negative value) in combination with the 4 hydrogen bonds formed upon the complex formation and thus it is the one selected for further analysis. The table below, shows the structural analysis in a residue-scale as it is already done with the ROP - (57-63 [residue] tail - RNA kissing loop complex without restraints. The color code shown in the graphs is an indicator of the percentage of the total solvent-accessible surface area that is buried upon the complex formation.

The columns colored red indicate the residues/nucleotides that more than the 50% of their solvent accessible area is buried after the complex formation.:

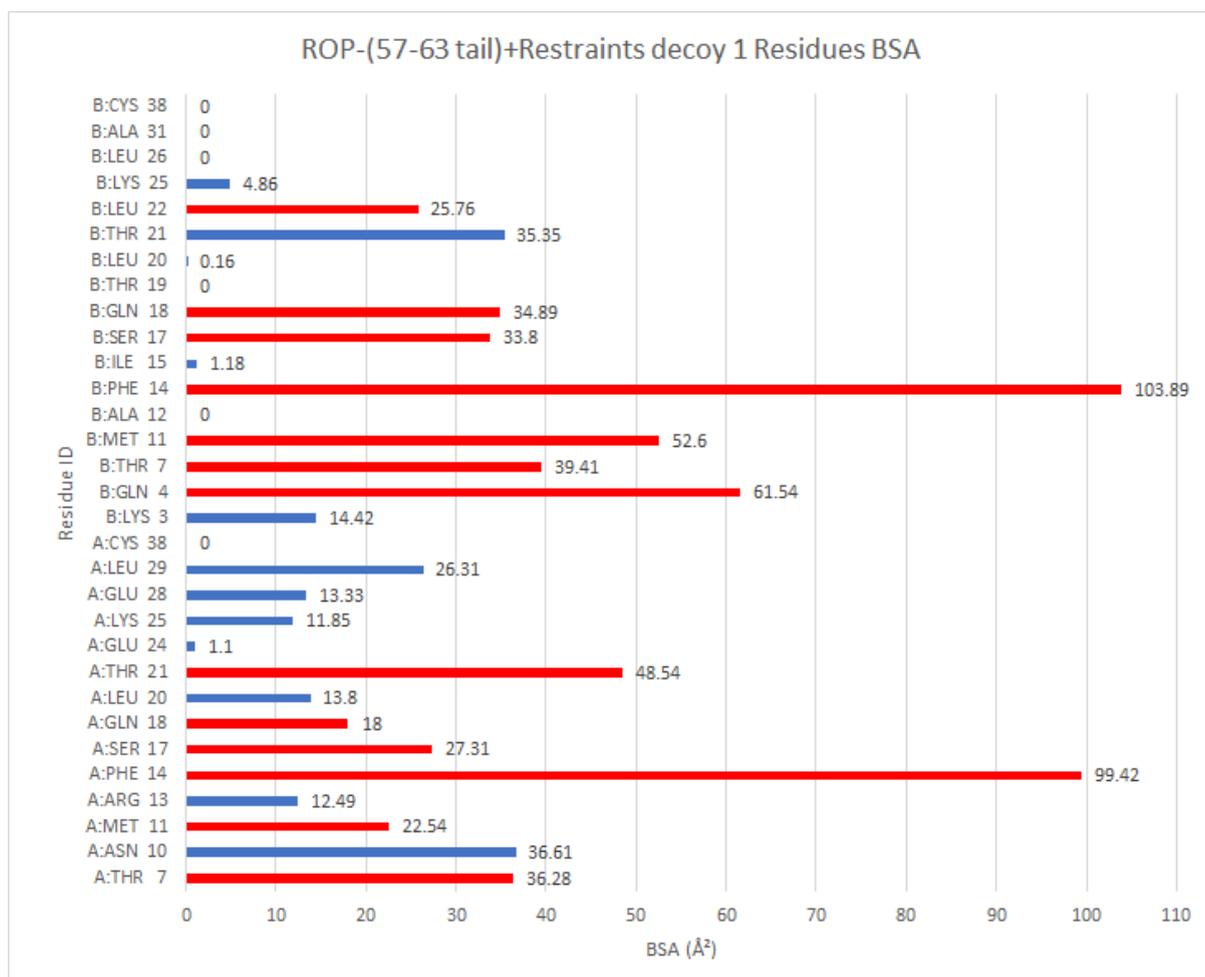


Figure 45| Buried Surface Area (Å²) table for each of the protein residues contributing to the protein - RNA kissing loop interface formation [Crossover: ROP - (57-63) tail - RNAI-RNAII kissing loop, +restraints, Decoy: 1]

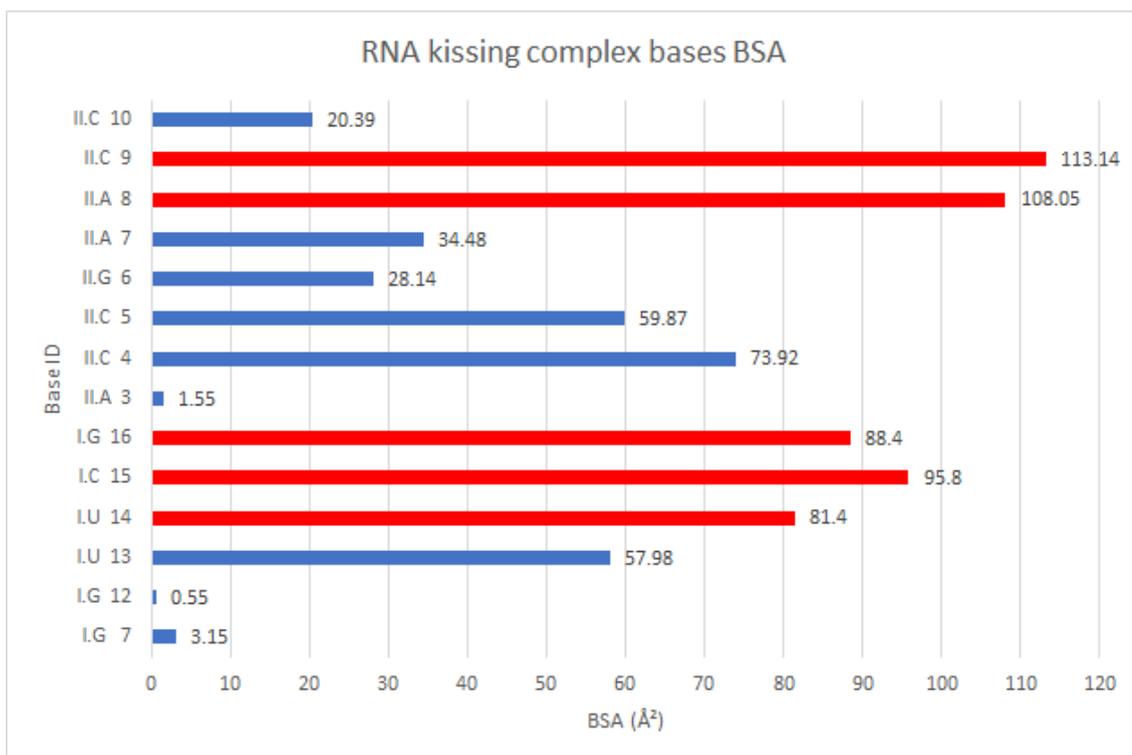


Figure 46| *Buried Surface Area (Å²) table for each of the ribonucleotide bases contributing to the protein - RNA kissing loop interface formation [Crossover: ROP - (57-63) tail - RNAI-RNAII kissing loop, +restraints, Decoy: 1]*

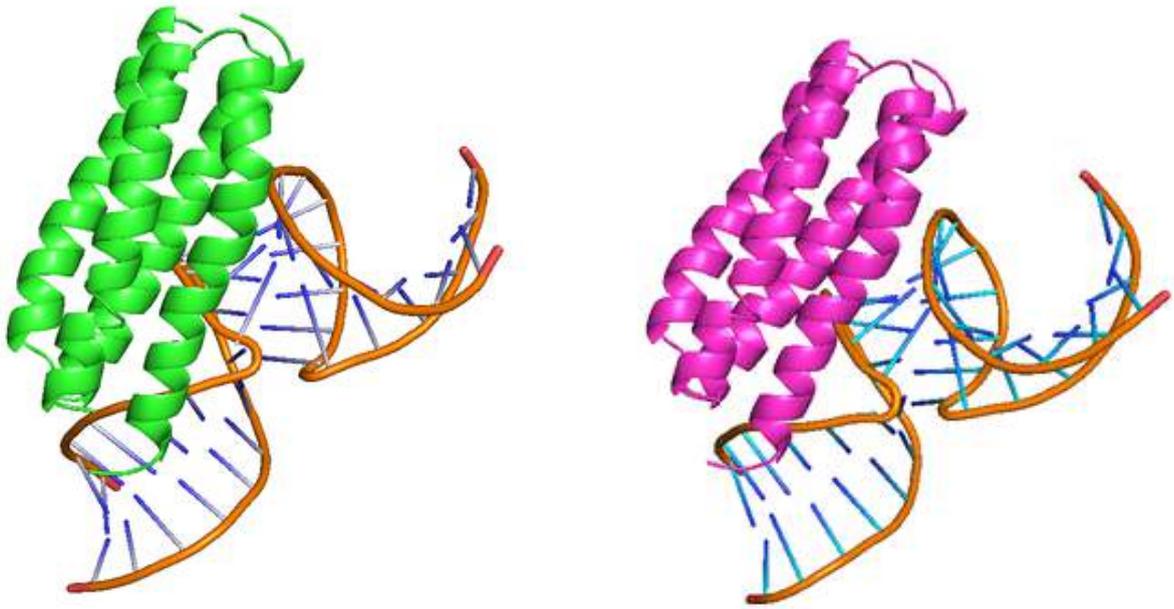


Figure 47| **Complex structures comparison** | The docking decoy 1 [Crossover: ROP - (57-63) tail - RNAI-RNAII kissing loop + restraints] on the right side and the docking decoy 6 [Crossover: ROP - (57-63) tail - RNAI-RNAII kissing loop] shown on the left side of the picture (image created by PYMOL [31, 35]).

4. Discussion

After creating, filtering and evaluating a huge number of docking decoys, there is one complex structure that is more likely to represent the ROP - RNAI-RNAII kissing complex found in bacteria. This model was created by the crossover between the ROP protein missing its 57-63 [residue] tail with the RNAI-RNAII kissing loop without introducing any additional docking parameters. The docking decoys resulting from these trials presented a stable picture with the 8 out of the 10 best scored docking models from the HDock program having an accepted P-value below the threshold of 0.5 [49]. Moreover, the interface analysis features for these docking models occur in a consistent fashion, where the solvation free energy was taking even more negative values while the complex Interface Area value was increasing - as it was expected for a pool of high quality decoy structures. One of the resulting models from this specific docking trial was selected for further investigation and it was the model 6 highlighted below.

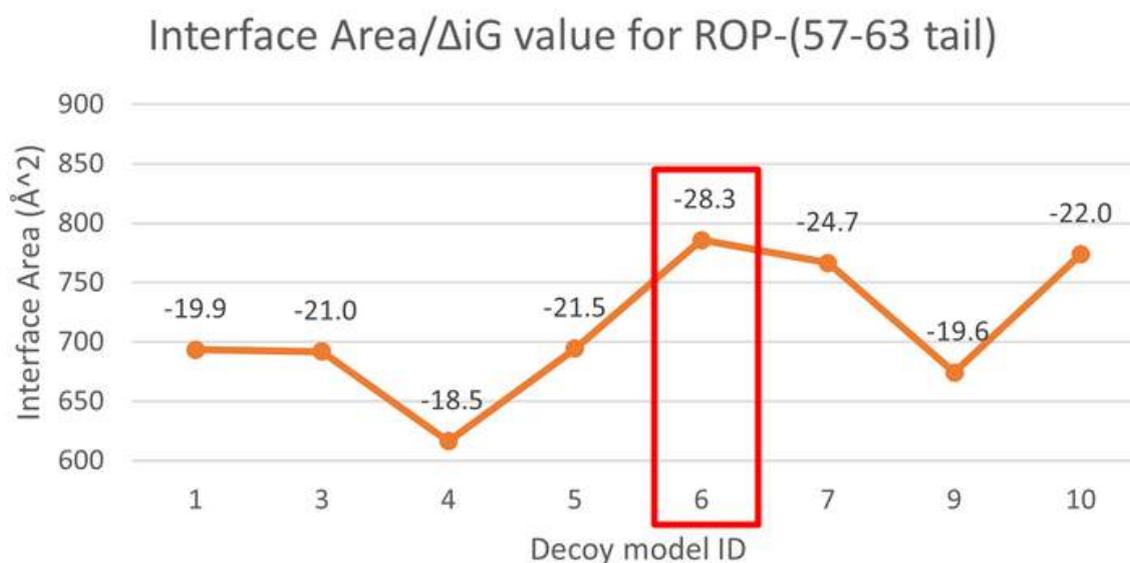


Figure 48| *Interface Area (\AA^2) to ΔiG (kcal/M) for the filtered decoys (P -value<0.5) of the docking between the ROP - (57-63) tail protein and the RNAI-RNAII kissing loop.*

The number of the ROP - (57-63 tail) protein contributing to the interface formation corresponds to the 20.5% of the total number of the protein residues and the nucleotides of the RNA kissing loop on the interface corresponds to the 35%. According to the relevant bibliography, the protein residues that are involved in the protein - RNA kissing loop interaction are the following: Lys-3, Asn-10, Gln-18, Phe-14 and Lys-25, with the Phe-14 being the most important residue for the RNA binding affinity [39]. From the graph below it is evident that the Phe-14 residues from both ROP monomers significantly contribute to the complex formation and the same is true for the Asn-10 and the Lys-25 of the first monomer. Hence, the interface analysis results of this decoy model between the ROP - (57-63 tail) protein and the RNAI-RNAII kissing loop come to agree with the literature, enhancing the reliability of the result.

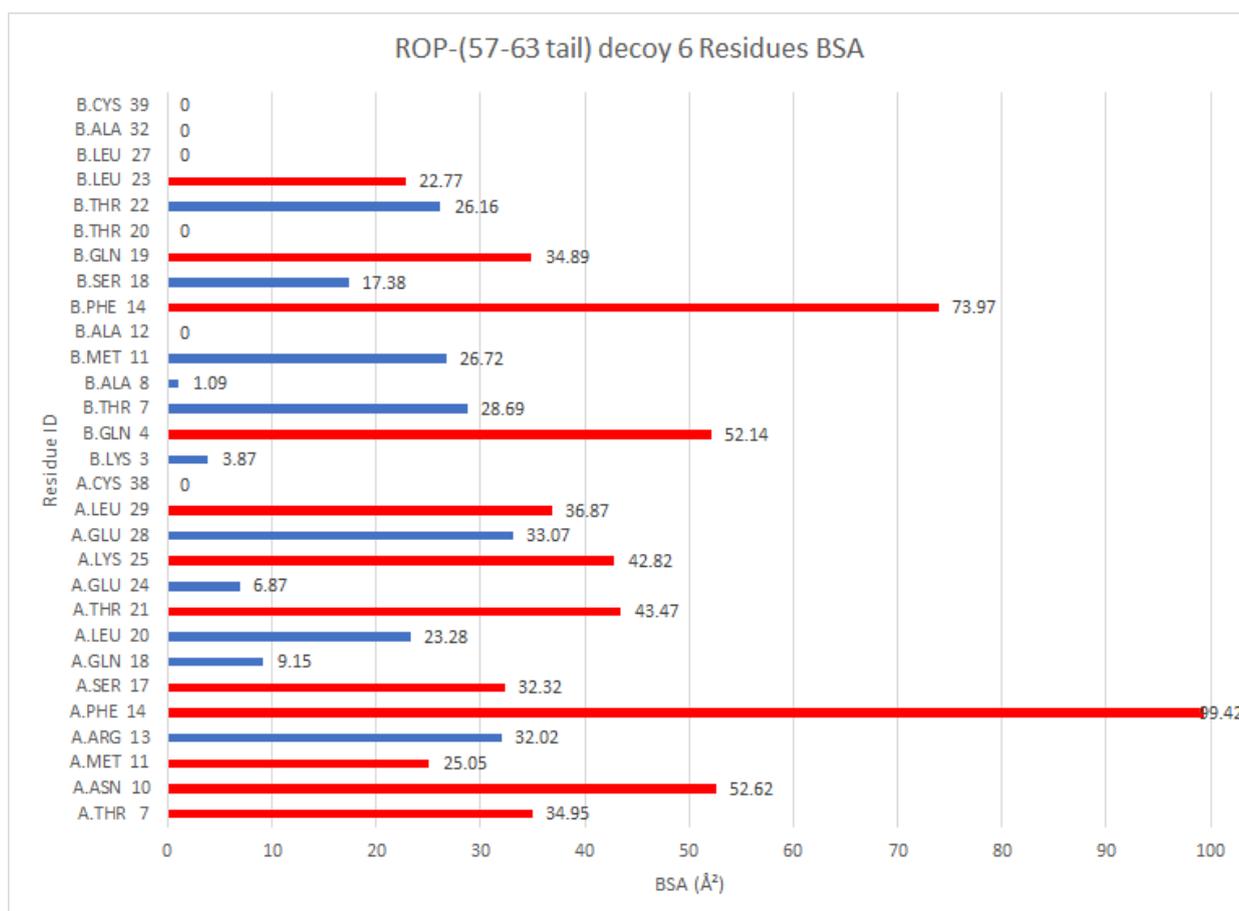


Figure 49| Buried Surface Area (Å²) table for each of the protein residues contributing to the protein - RNA kissing loop interface formation [Crossover: ROP - (57-63) tail - RNAI-RNAII kissing loop, Decoy: 6]

The best-scored decoy number 6, resulting from the molecular crossover: ROP - (57-63 tail) protein with the RNAI-RNAII kissing loop shows a lateral orientation between the ROP protein and the loop of the RNA kissing complex. This orientation leads to the formation of a structure that is an intermediate of the two orientations recommended from the literature (the one hypothesis is placing the ROP protein on the outer side of the RNA kissing loop [29] and the other one suggesting the the ROP protein is placed on the exact opposite side, the inner side of the RNA kissing loop [23]). Furthermore, the ROP - (57-63 tail) protein horizontal axis occupies a close-to-parallel position relative to the RNA kissing loop horizontal axis as shown below. This is also a feature that complies with the complex models suggested from the literature [23, 29].

The decoy number 1 resulting from the docking crossover: ROP - (57-63) tail - RNAI-RNAII kissing loop with the additional restraints I.A[9]:II.U[12], I.U[10]:II.A[11], I.G[11]:II.C[9], I.G[12]:II.C[10], was finally rejected, since the decoy 6 from the crossover ROP - (57-63) tail - RNAI-RNAII kissing loop described above, presents better features. The table below, displays the difference in the features of the two decoys:

<i>Metrics</i>	<i>Decoy 6 [crossover: ROP - (57-63) tail - RNAI-RNAII kissing loop]</i>	<i>Decoy 1 [crossover: ROP - (57-63) tail - RNAI-RNAII kissing loop + restraints]</i>
<i>P-value</i>	0.244	0.238
<i>Delta<i>G</i> (kcal/M)</i>	-28.3	-25.1
<i>Interface Area (Å²)</i>	785.8	771.1
<i>Hydrogen Bonds (N)</i>	5	4

Table 6| Comparison table for the features of the Docking Decoy 6 resulting from the crossover of the ROP - (57-63) tail - RNAI-RNAII kissing loop and the Decoy 1 resulting from the crossover of the ROP - (57-63) tail - RNAI-RNAII kissing loop with additional docking parameters.

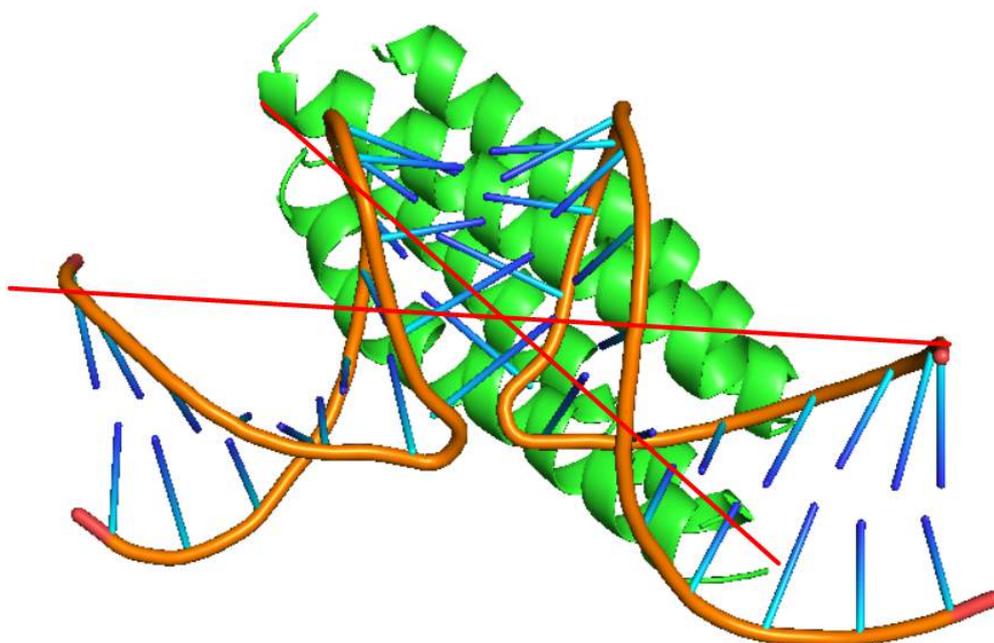


Figure 50| The decoy number 6 [crossover: ROP - (57-63 tail) protein - RNAI-RNAII kissing loop] with the horizontal axis of both ROP protein and RNA kissing loop highlighted in red color (image created by PYMOL [31, 35]).

Throughout the present study, it is finally getting clear that the docking model 6, resulting from the crossover between the ROP - (57-63 [residue]) tail and the RNAI-RNAII kissing loop is the one gathering the higher evaluation score - taking into account all the evaluation tools described so far. Nevertheless, further research is suggested upon the present study issue. Firstly, one important step that should be conducted in order to improve the results reliability for the complex evaluation, is the screening of all the decoy structures produced from the docking program as well as using more ROP protein mutants for the homology comparison. The first step can help the user to investigate if the scoring method of the docking program implemented, really complies with the results from the evaluation programs. Using only the first 10 best-scored model created by the docking algorithm, does not give a statistically significant comparison result. The second step can help to increase the volume of the comparison structures set, again scaling up the results' reliability. Moreover, another docking program could be used for further analysis of the complex. The HDock server calculations for the RMSD (Root Mean Square Deviation) value of the ligand molecule are performed by comparing the initial ligand structure position with the ligand conformation presented at the docking decoys. Thus the output RMSD can not be considered as an accurate metric for the docking decoy evaluation (HDock RMSD values > 10 Å). However, the fact that the RNA molecules are characterized as highly flexible structures, a docking or structure evaluation tool, able to estimate the RMSD value for structural changes of the RNA ligand due to its binding upon a receptor, is required [25, 35]. Lastly, utilizing docking programs that use template-based docking algorithms inclusively, such as the P3Dock server, could be considered, aiming to create docking decoys of higher quality [51].

Complex Template Information ([Click to Hide](#))

Molecule	PDB ID	Chain ID	Align_length	Coverage	Seq_ID (%)
Receptor	1Y07	A	120	0.952	68.3

Note: The built model of "Model 0" based on the above PDB complex template has a confidence.

Summary of the Top 10 Models

Rank	1	2	3	4	5	6	7	8	9	10
Docking Score	-297.10	-284.13	-283.86	-280.32	-280.23	-278.05	-274.26	-267.98	-266.86	-266.20
Confidence Score	0.9499	0.9360	0.9357	0.9313	0.9312	0.9283	0.9231	0.9137	0.9119	0.9108
Ligand rmsd (Å)	67.35	72.77	84.58	70.40	71.97	71.80	73.25	84.31	72.66	64.10
Interface residues	model 1	model 2	model 3	model 4	model 5	model 6	model 7	model 8	model 9	model 10

Note: The models are ranked according to the docking scores. Please click [help](#) for the explanations of evaluation metrics

Quality of Docking Structures/Input Data ([Click to Hide](#))

Receptor: The protein structure was submitted by users.

```

=====
Quality report of input model by ProQ v1.2
=====
LGscore:          2.210
MaxSub:           0.160

Different ranges of protein model quality by LGscore or MaxSub:
Correct          Good          Very good
1.5 < LGscore   3.0 ≤ LGscore < 5.0     5.0 ≤ LGscore
0.1 < MaxSub    0.5 ≤ MaxSub < 0.8     0.8 ≤ MaxSub

```

Ligand: The RNA structure was submitted by users.

NOTE: Quality checking is not conducted for this user-input RNA structure.

Figure 51| *HDOCK* results page from the docking trial between the ROP protein and the RNAI-RNAII kissing loop including the Ligand rmsd (Å) [35].

Another problem faced during the docking structures evaluation is the zero value of the CSS. The CSS value represents the Complexation Significance Score, which indicates the significance of the interface resulting from the complex formation [49]. The molecules contributing to the complex formation are indeed very loosely bound to each other, leading to the zero value measurement of the CSS metric. This is a problem that could probably be mitigated if the user provides the acceptable area for the docking, creating boundaries such as the grid options that can be introduced through the AutoDock VINA docking software [52].

Lastly, it is important to notice that building a protocol for protein - RNA docking is really challenging, since this field is still in its infancy, especially when the RNA structure consists of two distinct nucleic acid chains that present an unusual conformation such as the RNA kissing loops [25]. Most of the docking programs for protein - RNA docking are still under development and the documentation is limited, making it even more difficult for inexperienced users to utilize them [42]. At the same time, many tools for molecular docking, structural and interface evaluation and molecular handling in general, are not updated or they are even not available

anymore, adding one more obstacle to the docking and evaluation protocol construction.

5. Bibliography

- [1] Polisky, Barry. "ColE1 replication control circuitry: sense from antisense." *Cell* 55.6 (1988): 929-932.
- [2] Hershfield, Vickers, et al. "Plasmid ColE1 as a molecular vehicle for cloning and amplification of DNA." *Proceedings of the National Academy of Sciences* 71.9 (1974): 3455-3459.
- [3] Tomizawa, Jun-ichi, and Tapan Som. "Control of cole 1 plasmid replication: enhancement of binding of RNA I to the primer transcript by the rom protein." *Cell* 38.3 (1984): 871-878.
- [4] Eguchi, Yutaka, and Jun-ichi Tomizawa. "Complex formed by complementary RNA stem-loops and its stabilization by a protein: function of ColE1 Rom protein." *Cell* 60.2 (1990): 199-209.
- [5] Lacatena, R. M., and G. Cesareni. "Base pairing of RNA I with its complementary sequence in the primer precursor inhibits ColE1 replication." *Nature* 294.5842 (1981): 623-626.
- [6] Singh, Abhishek, Latsavongsakda Sethaphong, and Yaroslava G. Yingling. "Interactions of cations with RNA loop-loop complexes." *Biophysical journal* 101.3 (2011): 727-735.
- [7] Lee, Anna J., and Donald M. Crothers. "The solution structure of an RNA loop-loop complex: the ColE1 inverted loop sequence." *Structure* 6.8 (1998): 993-1007.
- [8] Tomizawa, Jun-ichi. "Control of cole 1 plasmid replication: the process of binding of RNA I to the primer transcript." *Cell* 38.3 (1984): 861-870.
- [9] Tomizawa, Jun-ichi. "Control of ColE1 plasmid replication: initial interaction of RNA I and the primer transcript is reversible." *Cell* 40.3 (1985): 527-535.
- [10] Hjalt, Tord, E. Gerhart, and H. Wagner. "The effect of loop size in antisense and target RNAs on the efficiency of antisense RNA control." *Nucleic acids research* 20.24 (1992): 6723-6732.
- [11] Banner, David W., Michael Kokkinidis, and Demetrius Tsernoglou. "Structure of the ColE1 Rop protein at 1.7 Å resolution." *Journal of molecular biology* 196.3 (1987): 657-675.
- [12] Eberle, Wolfgang, et al. "The structure of ColE1 rop in solution." *Journal of biomolecular NMR* 1.1 (1991):71-82.

- [13] Chang, Kung-Yao, and Ignacio Tinoco Jr. "The structure of an RNA "kissing" hairpin complex of the HIV TAR hairpin loop and its complement." *Journal of molecular biology* 269.1 (1997): 52-66.
- [14] Eguchi, Yutaka, and Jun-ichi Tomizawa. "Complexes formed by complementary RNA stem-loops: Their formations, structures and interaction with ColE1 Rom protein." *Journal of molecular biology* 220.4 (1991):831-842.
- [15] Castagnoli, L., et al. "Genetic and structural analysis of the ColE1 Rop (Rom) protein." *The EMBO journal* 8.2 (1989): 621-629
- [16] Hari, S. B., Byeon, C., Lavinder, J. J., & Magliery, T. J. (2010). Cysteine-free rop: A four-helix bundle core mutant has wild-type stability and structure but dramatically different unfolding kinetics. *Protein science*, 19(4), 670-679.
- [17] Glykos, N. M., Cesareni, G., & Kokkinidis, M. (1999). Protein plasticity to the extreme: changing the topology of a 4- α -helical bundle with a single amino acid substitution. *Structure*, 7(6), 597-603.
- [18] Dominguez, Cyril, Rolf Boelens, and Alexandre MJJ Bonvin. "HADDOCK: a protein- protein docking approach based on biochemical or biophysical information." *Journal of the American Chemical Society* 125.7 (2003): 1731-1737.
- [19] Chauvot de Beauchene, Isaure, Sjoerd J. de Vries, and Martin Zacharias. "Binding site identification and flexible docking of single stranded RNA to proteins using a fragment-based approach." *PLoS computational biology* 12.1 (2016): e1004697.
- [20] Huang, Yangyu, et al. "A novel protocol for three-dimensional structure prediction of RNA-protein complexes." *Scientific reports* 3.1 (2013): 1-7.
- [21] Glykos, N. M., Papanikolaou, Y., Vlassi, M., Kotsifaki, D., Cesareni, G., & Kokkinidis, M. (2006). Loopless Rop: Structure and dynamics of an engineered homotetrameric variant of the repressor of primer protein. *Biochemistry*, 45(36), 10905-10919.
- [22] Willis, M. A., Bishop, B., Regan, L., & Brunger, A. T. (2000). Dramatic structural and thermodynamic consequences of repacking a protein's hydrophobic core. *Structure*, 8(12), 1319-1328.
- [23] Struble, E. B., Ladner, J. E., Brabazon, D. M., & Marino, J. P. (2008). New crystal structures of ColE1 Rom and variants resulting from mutation of a surface exposed residue: Implications for RNA-recognition. *Proteins: Structure, Function, and Bioinformatics*, 72(2), 761-768.

- [24] Bender, B. J., Gahbauer, S., Lutten, A., Lyu, J., Webb, C. M., Stein, R. M., ... & Shoichet, B. K. (2021). A practical guide to large-scale docking. *Nature protocols*, 16(10), 4799-4832.
- [25] Nithin, C., Ghosh, P., & Bujnicki, J. M. (2018). Bioinformatics tools and benchmarks for computational docking and 3D structure prediction of RNA-protein complexes. *Genes*, 9(9), 432.
- [26] Madan, B., Kasprzak, J. M., Tuszynska, I., Magnus, M., Szczepaniak, K., Dawson, W. K., & Bujnicki, J. M. (2016). Modeling of protein–RNA complex structures using computational docking methods. In *Computational Design of Ligand Binding Proteins* (pp. 353-372). Humana Press, New York, NY.
- [27] Eberle, W., Pastore, A., Sander, C., & Rösch, P. (1991). The structure of ColE1 rop in solution. *Journal of biomolecular NMR*, 1(1), 71-82.
- [28] Chang, K. Y., & Tinoco Jr, I. (1997). The structure of an RNA “kissing” hairpin complex of the HIV TAR hairpin loop and its complement. *Journal of molecular biology*, 269(1), 52-66.
- [29] Comolli, L. R., Pelton, J. G., & Tinoco Jr, I. (1998). Mapping of a protein-RNA kissing hairpin interface: Rom and Tar-Tar. *Nucleic acids research*, 26(20), 4688-4695.
- [30] Madhavi Sastry, G., et al. "Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments." *Journal of computer-aided molecular design* 27.3 (2013): 221-234.
- [31] The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.
- [32] Tuszynska, I., Magnus, M., Jonak, K., Dawson, W., & Bujnicki, J. M. (2015). NPdock: a web server for protein–nucleic acid docking. *Nucleic acids research*, 43(W1), W425-W430.
- [33] Yan, Y., Zhang, D., Zhou, P., Li, B., & Huang, S. Y. (2017). HDock: a web server for protein–protein and protein–DNA/RNA docking based on a hybrid strategy. *Nucleic acids research*, 45(W1), W365-W373.
- [34] Cheng, T. M. K., Blundell, T. L., & Fernandez-Recio, J. (2007). pyDock: electrostatics and desolvation for effective scoring of rigid-body protein–protein docking. *Proteins: Structure, Function, and Bioinformatics*, 68(2), 503-515.
- [35] Yan Y, Tao H, He J, Huang S-Y.* The HDock server for integrated protein-protein docking. *Nature Protocols*, 2020; doi: <https://doi.org/10.1038/s41596-020-0312-x>.

- [36] Yan Y, Wen Z, Wang X, Huang S-Y. Addressing recent docking challenges: A hybrid strategy to integrate template-based and free protein-protein docking. *Proteins* 2017;85:497-512.
- [37] Huang S-Y, Zou X. A knowledge-based scoring function for protein-RNA interactions derived from a statistical mechanics-based iterative method. *Nucleic Acids Res.* 2014;42:e55.
- [38] Huang S-Y, Zou X. An iterative knowledge-based scoring function for protein-protein recognition. *Proteins* 2008;72:557-579.
- [39] Predki, P. F., Nayak, L. M., Gottlieb, M. B., & Regan, L. (1995). Dissecting RNA-protein interactions: RNA-RNA recognition by Rop. *Cell*, 80(1), 41-50.
- [40] Čech, P., Svozil, D., & Hoksza, D. (2012). SETTER: web server for RNA structure comparison. *Nucleic acids research*, 40(W1), W42-W48.
- [41] Tuszynska, I., & Bujnicki, J. M. (2011). DARS-RNP and QUASI-RNP: new statistical potentials for protein-RNA docking. *BMC bioinformatics*, 12(1), 1-16.
- [42] Paxman, J. J., & Heras, B. (2017). Bioinformatics tools and resources for analyzing protein structures. In *Proteome Bioinformatics* (pp. 209-220). Humana Press, New York, NY.
- [43] Vangone, A., Spinelli, R., Scarano, V., Cavallo, L., & Oliva, R. (2011). COCOMAPS: a web application to analyze and visualize contacts at the interface of biomolecular complexes. *Bioinformatics*, 27(20), 2915-2916.
- [44] Choi, H., Kang, H., & Park, H. (2013). New solvation free energy function comprising intermolecular solvation and intramolecular self-solvation terms. *Journal of cheminformatics*, 5(1), 1-13.
- [45] Gurung, A. B., Bhattacharjee, A., & Ali, M. A. (2016). Exploring the physicochemical profile and the binding patterns of selected novel anticancer Himalayan plant derived active compounds with macromolecular targets. *Informatics in Medicine Unlocked*, 5, 1-14.
- [46] Chakravarty, D., Guharoy, M., Robert, C. H., Chakrabarti, P., & Janin, J. (2013). Reassessing buried surface areas in protein-protein complexes. *Protein Science*, 22(10), 1453-1457.
- [47] Connolly, M. L. (1983). Solvent-accessible surfaces of proteins and nucleic acids. *Science*, 221(4612), 709-713.
- [48] Richmond, T. J. (1984). Solvent accessible surface area and excluded volume in proteins: Analytical equations for overlapping spheres and implications for the

- hydrophobic effect. *Journal of molecular biology*, 178(1), 63-89.
- [49] Krissinel, E., & Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *Journal of molecular biology*, 372(3), 774-797.
- [50] Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) "Basic local alignment search tool." *J. Mol. Biol.* 215:403-410.
- [51] Zheng, J., Hong, X., Xie, J., Tong, X., & Liu, S. (2020). P3DOCK: a protein–RNA docking webserver based on template-based and template-free docking. *Bioinformatics*, 36(1), 96-103.
- [52] Huey, R., Morris, G. M., & Forli, S. (2012). Using AutoDock 4 and AutoDock vina with AutoDockTools: a tutorial. *The Scripps Research Institute Molecular Graphics Laboratory*, 10550, 92037.
- [53] [UCSF Chimera--a visualization system for exploratory research and analysis.](#) Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. *J Comput Chem.* 2004 Oct;25(13):1605-12.
- [54] Educational portal of PDB. Guide to understanding pdb data/missing coordinates and biological assemblies. (09. 2022) <https://pdb101.rcsb.org/learn/guide-to-understanding-pdb-data/missing-coordinates-and-biological-assemblies>
- [55] Puton, T., Kozłowski, L., Tuszynska, I., Rother, K., & Bujnicki, J. M. (2012). Computational methods for prediction of protein–RNA interactions. *Journal of structural biology*, 179(3), 261-268.
- [56] Tuszynska, I., Matelska, D., Magnus, M., Chojnowski, G., Kasprzak, J. M., Kozłowski, L. P., ... & Bujnicki, J. M. (2014). Computational modeling of protein–RNA complex structures. *Methods*, 65(3), 310-319.
- [57] Remmert M, Biegert A, Hauser A, Soing J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat Methods.* 2011;9:173-175.
- [58] W. R. Pearson and D. J. Lipman. Improved tools for biological sequence comparison. *Proc Natl Acad Sci USA* 1988;85:2444-2448.
- [59] Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, Thompson JD, Higgins DG. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol.* 2011;7:539.
- [60] Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG.

Clustal W and Clustal X version 2.0. *Bioinformatics*. 2007;23:2947-8.

[61] Marti-Renom MA, Stuart A, Fiser A, Sanchez R, Melo F, Sali A. Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct* 2000;29:291-325.

[62] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. *Nucleic Acids Res* 2000;28:235-242.

[63] HDOCK SERVER. Help for using HDOCK server (09/2022)
<http://hdock.phys.hust.edu.cn/help.php#metric>