Predicting and stabilising the 3D structure of aptamers using computational methods

by Kristy Montalbo

A thesis submitted in partial fulfilment for the requirements for the degree of MSc in Cancer Biology and Therapy at the University of Central Lancashire

Supervisor: Dr. Lisa Shaw
Co-supervisor: Dr. David Michael Griffiths

August 2020

Abstract

The purpose of this dissertation was to predict and stabilise the three-dimensional (3D) structure of aptamers. The focus was in the area of bioinformatics. A guideline was provided on how to obtain 3D images of aptamers, predict the binding site and stabilise the aptamer's structures. Such a study is of importance to facilitate in vivo experiments by using computational methods to determine if stabilising agents and shortened aptamer sequences interfere with the predicted binding site. Research methods included a literature search and creating a protocol on which software and webserver to use, combined with the collection and analysis of figures. The 2D (two dimensional) structures were obtained through Mfold alongside RNAComposer to obtain the PDB file containing the predicted 3D structure. The aptamer's 3D structures were visualised with YASARA, PyMOL and Chimera. The findings provided evidence that shortening the aptamer's sequence stabilised the structures. The predicted binding sites show that the ligand binds in the stem-loop region. The main conclusions drawn from this project were that computational methods can be used to analyse and gain insight into 3D aptamer structures. Some common challenges and technological difficulties, such as modifying an aptamer, were discussed. This project recommended Mfold to be used to obtain the secondary structure of aptamers, whereas for the tertiary structure PyMOL and YASARA were the easiest to navigate in providing a clear and detailed structure.

Contents

1. Introduction

Aptamers were first introduced in the 1990s and have been used in various fields of science (Sefah et al., 2010; Liu et al., 2014). Aptamers are artificially made DNA (ssDNA) or RNA (ssDNA) that consist of different numbers of bases (Liu et al., 2014). DNA is more stable in vitro compared to RNA and facilitates the aptamer selection procedure (Yan et al., 2019). An aptamers molecular weight is approximately 10 to 30 kDA hence, it is smaller than an antibody (Gomes de Castro et al., 2017). Aptamers have been used and selected against a wide range of targets such as viruses (Song et al., 2012), whole cells (Opazo et al., 2015) and small organic molecules (Ellington et al., 1990). Additionally, they can fold into numerous secondary and tertiary structures (Chandola et al., 2016). Due to the aptamer's 3D structure it binds to certain targets (Caroli et al., 2016) with high affinity, specificity and low immunogenicity (Berg et al., 2016). Aptamers interacts with their targets by recognizing intermolecular interactions e.g. van der Waals interactions and hydrogen bonds (Rahimizadeh et al., 2017, Ruiz Ciancio et al., 2018). SELEX (Systemic Evolution for Ligands by Exponential Enrichment) is the technology used to select aptamers (Tuerk, Gold, Ellington and Szostak, 1990). At current, aptamers have multiple purposes in biomedical settings e.g. targeted drug delivery (Ma et al., 2015). The development of aptamers is primarily restricted to laboratory settings. They are used in cancer diagnostic due to their high potential for cancer detection (Wu et al., 2015). Aptamers increase treatment efficacy and disease management by targeting overexpressed proteins in tumours (Chiang et al., 2013).

Due to aptamers' characteristics they are suitable candidates for computational methods (Emami et al., 2020). The advantages of utilizing bioinformatical methods for aptamers is the use of software and webservers without requiring sophisticated devices (Ahirwar et al., 2016). Looking at the 3D structure of aptamers is fairly new. However, various computational experiments with aptamers were performed, such as the design of efficient aptamers (Hamada, 2018) and predicting interactions (Li et al., 2014). The detailed prediction of the 2D structure is important to use as an input for 3D modelling (Popenda et al., 2012).

The simple structures of aptamers facilitate modifying it for various applications (Wang et al., 2016). Moreover, by altering its chemical and physical properties the aptamers' characteristics can be improved to influence its stability (Röthlisberger and Hollenstein; 2018; Zhou and Rossi, 2017; Kim et al., 2017) by modifying e.g. backbone and sugar changes in the blood due to nuclease degradation (Flamme et al., 2019). Researchers increased aptamers' half-life in vivo by conjugating with different values of PEG (Polyethylene glycol) (Floege et al, 1999; Healy et al., 2004). Aptamers smaller than 30 kDA are excreted more easily via the kidneys (Gupta et al., 2017). A study proofed that using 20 kDa PEG for aptamer modification lead to decreased rates of renal excretion and increased half-life in vivo (Shaw et al., 1991; Jäschke et al., 1994). Modified aptamers can be seen in Fig.1 such as F, OCH_3, Ch_2OH or SH being substituted for 2'OH (H). Thiol (S) or borane (BH3) groups were applied to an α phosphate to reinforce oligonucleotide backbones (Horhota et al., 2005; Tsai et al., 2007; Kuwahara et al., 2008; Kasahara et al., 2013). Additionally, modifications are seen in 2'-F-pyrimidine, 5-chloro-2'-deoxyuridine (5ClU), 2'-O-methoxyethyl (2'-MOE), 2'-(deoxy-)ribose sugar, 2'-O-methyl (2'-OMe), 2'-fluorocytidines, 2'-aminouridines, 2'-aminonucleotides, , 2'-phosphate groups and different nucleic acid designs (Pieken et al.,1991; Kuwahara et al., 2013; Eremeeva and Herdewijin, 2019).

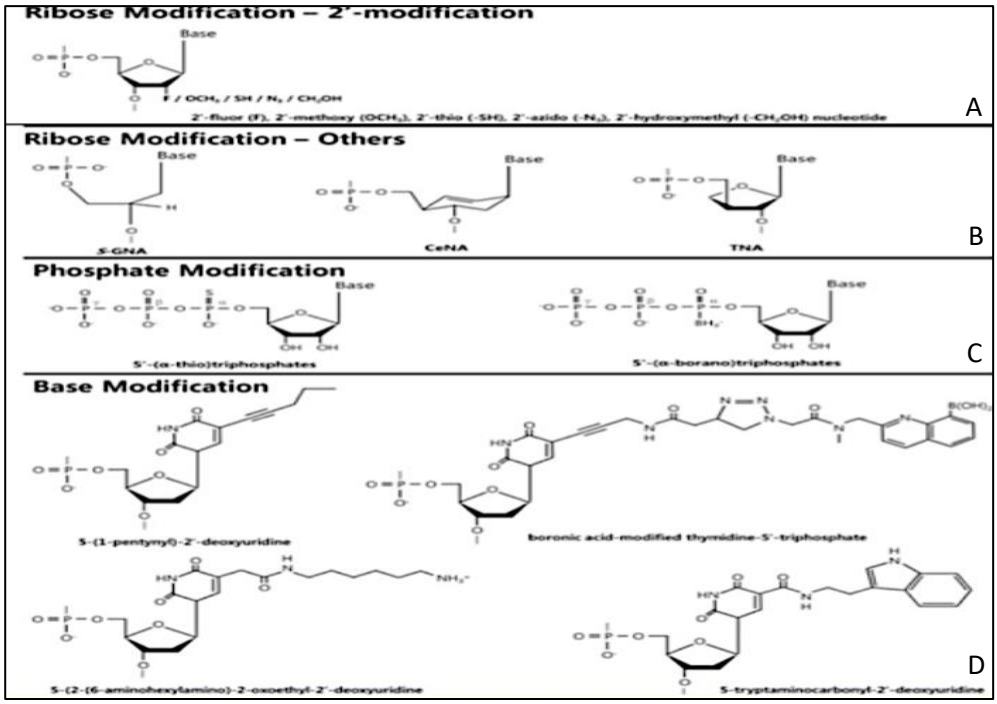


Figure 1. A-D show chemically modified nucleotide structures (Kuwahara et al., 2008; Kasahara et al., 2013, Kong and Byun, 2013). Ribose 2' modification (A), Phosphate (C) and base (D) modification are used in vivo to increase aptamer stability. A shows the often used substitution of F, OCH_3, SH or CH_2OH for 2'-OH (H). Additionally, to strengthen oligonucleotide backbone thiol (S) or borane (BH_3) group is introduced to α phosphate. Furthermore, functional groups can also be introduced into the base.

1.1 Aim and Objectives

The main research focus is to visualise the 3D structure and provide a guideline on how to use freely available webservers and software to obtain the 2D and 3D structures of the given aptamers (SA43, S1p, EGFR8, XApt004, TDM1 and hTMA10). The research area is bioinformatics and hence computational methods are used.

Firstly, the 2D structure was obtained by using the webservers Mfold and Vienna RNAfold (Appendix A-B). The Mfold web server application predicts the secondary structure folding of single-stranded nucleotides (Zuker, 2003). Vienna RNAfold webserver predicts and compares secondary RNA structures (Lorenz et al., 2011). These webservers were compared to validate which is more reliable. Then, the 2D structure and its dot-bracket notation were verified and obtained. The dot-bracket notation represents secondary structures by representing base pairs as brackets () and unpaired bases are symbolized as dots 3 (Ramlan and Zauner, 2008). The 3D structure was visualised using the software Yet Another Artificial Reality Application (YASARA), PyMOL and Chimera (Appendix C).

Further objectives of this project were to stabilise the aptamers e.g. by shortening the sequence by removing the unpaired bases at the 5'-3' ends (Appendix D). Aptamers are shortened to increase penetration (Ferreira and Missailidis, 2007; Lakhin et al., 2013) and to observe a change in structure. The stem-loop region was predicted to be the target's main binding site (Moody and Bevilaqua, 2003; Debiais et al., 2020). Stem-loop structures consist of a base paired stem structure and a loop sequence with unpaired or non-Watson-Crick-paired nucleotides (Hirao et al., 1994). The sequences of aptamers can be long hence less stable and more expensive impeding preparation. Therefore, it is

recommended to shorten aptamers at their 'floppy' ends, which consist of unpaired bases (Gopinath et al., 2017). Knowing the 3D structure enables understanding functions and interactions of molecules (Laing and Schlick, 2010). Additionally, it was analysed if stabilising agents interfere with the predicted binding sites. Chimera was used to add stabilising agents by uploading the PDB file of a chosen structure e.g. thiol (S) as this strengthens the oligonucleotide backbone (Appendix J) (Kuwahara et al., 2008; Kasahara et al., 2013). PyMOL provided the editing option to add chemical compounds and proteins to the aptamer structure (Appendix H). AutoDockTools was used to prepare the aptamer, ligand and the grid for docking. Vina and Terminal were used for docking (Appendix F). Additionally, a SA43-KU Heterodimer (Ku 70 and Ku 80) docking was attempted (Appendix G).

This research project is of value as it establishes a guideline on how to obtain and visualise the 2D and 3D structure of aptamers. This enables including bioinformatical information prior to a wet lab project to predict the success of an experiment and save time and material costs. This project establishes an insight into the computational predictive structural approaches of six aptamer sequences (Table 1). The predictions of the structures were carried out in three main categories: 2D and 3D structure-based prediction, aptamer stabilization and aptamer-ligand docking.

2. Materials and Methods

The identification of a gap in existing research in the bioinformatical analyses of aptamers suggested the need of evidence and information on how to obtain visualised aptamer structures especially that of the tertiary. The research methods attempted to provide an insight into bioinformatical approaches to aptamers by implementing a guideline and collecting figures of the given aptamers (Table 1) from various software and webservers. The step by step guidelines for each method can be found in Appendix A-J.

Table 1. Shows the original aptamer sequences with their dot-bracket notation obtained by a former MSc student at UCLan as provided by Dr. Lisa Shaw.

Aptamer	**Sequence**
SA43	5'-ACGTTACTCTTGCAACACAAACTTTAATAGCCTCTTATAGTTC-3' ..(((.....)))...((((.(((.....))).)))).
S1p	5'- ATCCTAATACGACTCACTATGCCGCCGGAGACCAAAGAGCAACACGAGGGAAC GCAAGAGGGGGGATCGACAGAATTCTGCGGTAAACTCGAGG-3' .....(((.....)))..........((((....((((.....))))).....))))...
EGFR8	3'CCCTCGAGTTTACCGCAGAATTCTTGATTTTCTGTACCTCGGTCGGACCATAC CTGCCCCGTTTCGTTGCGTCGAATTCTCTCCCTATAGTGAGTCGTATTAGGAT-5' .....(((((((((.....)))))).....))))(((..((....))..))).((((......))))..(((((.....)))))..
XApt004	5'- CCCTCGAGTTTACCGCAGAATTCCACTCCTTCCCTCCCCGTCTTCACTTCCTCATC CGTACGACACTGCCGTGGAATTCTCTCCCTATAGTGAGTCGTATTAGGAT-3' .....((((((((((.....(((.....))).....)))))))))...((((.....))))..
TDM1	5'- ATCCTAATACGACTCACTATGTAGGGACTTAGCTCTCACCGTATGGCATGCGTA GTACTGAGAGTGTGGGAATTCTGCGGTAAACTCGAGG-3' ..((((.....))))..(((((((..(((.....)))))))))).....((.....))...
hTMA10	5'- ATCCTAATACGACTCACTATGCGGACACGGGGCAACGGCGAGAGACGGGATG GGCGGCAACACGGGAATTCTGCGGTAAACTCGAGG-3' .....((.(((.....((....))..((....))..))).)).....((.....)).....((.....))...

A Literature Search was performed on how to obtain 2D and 3D structures establishing which software and webservers to use. The chosen criteria for the computational methods were; easily accessible, obtains results quickly, free to download and lastly,

immediate available license. The required data was obtained by taking screenshots of the structures for each aptamer in various webservers and software to compare which software had the best and most accurate visualisation and to analyse changes in the aptamer structure.

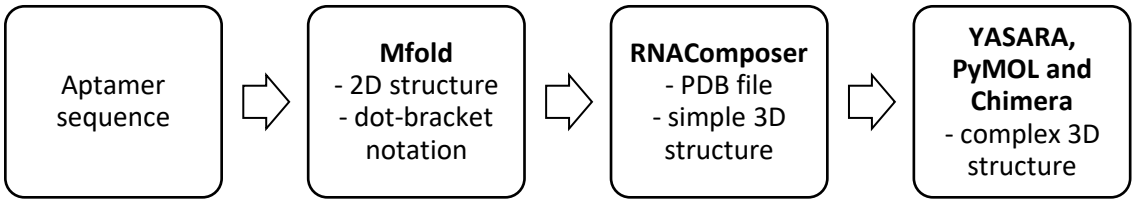


Figure 2. Step by Step guide on how to obtain the 3D structure.

2.1 2D Structure

The secondary structure was obtained via webserver Mfold by inserting each given aptamer sequence from Table 1 individually into the DNA Folding Form Application using the default setting (Appendix A). The crawling method deleted bases from the 5' and 3'ends of an aptamer. Stem-loop regions were checked to see if they were present in the obtained secondary structures (Gopinath et al., 2017). The next step was to obtain the dot-bracket notations (Appendix B). The dot-bracket notations were created by analysing the obtained 2D structure. Another webserver called Vienna RNAfold was used to compare the dot-bracket notations from Mfold (Appendix C). It was chosen as it provided results quickly. The dot-bracket notation that matched the one from Table 1 were used for further steps ensuring validity. The dot-bracket notations from Mfold that did not match were manually created by analysing the obtained 2D structure with the most hairpin loop and fewest bulge, external and interior loops in Mfold (Moody and Bevilaqua, 2003; Debiais et al., 2020). The aptamer sequence and dot-bracket notation were added to RNAComposer to get a PDB file incorporating the 3D structure.

2.2 3D Structure

The webserver RNAComposer was used to obtain a PDB file by pasting the aptamer sequence and its dot-bracket into the server. When using the RNAComposer, DNA is treated as RNA, hence thymine (T) is replaced by uracil (U). 3D structure elements suitable to these fragmentations are searched in the RNA FRABASE database, which consists of a 3D structural element with structural properties (Popenda et al., 2010). RNAComposer provided a simple 3D structure and the PDB file. The software YASARA, PyMOL and Chimera were downloaded to upload the PDB file (Appendix C). The 3D structures were obtained and still figures were taken for each aptamer.

2.3 Aptamer Stabilisation

Each six aptamers were stabilised by performing a sequence shortening (Appendix D) or by adding stabilising agents to its structure (Appendix E) to analyse aptamer modifications.

2.3.1 *Shortened Aptamers*

The aptamer structures from Mfold were shortened to obtain a new sequence with no to little floppy ends. The number of removed nucleotides at the 5'-3' end of each six aptamers varied, hence new shortened sequences for each aptamer were designed. The matching dot-bracket notation for each aptamer was also shortened by removing the same number of dots replicating an unmatched nucleotide at the 5'-3'-end of the sequence. This shortening method was chosen as it was convenient and reliable when comparing the original structures as the same settings were applied. As many structures are provided on Mfold, the structure which resembled the original one the most was chosen. For each shortened aptamer sequence, a matching dot-bracket notation was created. The shortened aptamer sequence and the matching dot-bracket notation were pasted to RNAComposer to obtain the PDB file and a simple visualised 3D structure preview. To obtain the final 3D structure the PDB file of an aptamer was added to

YASARA, PyMOL and Chimera. Screenshots of each aptamer in all the used computational programs were obtained.

2.3.2 *Addition of Stabilising Agents*

The aim is to add functional groups to the floppy ends, which are the unpaired nucleotides at the start and end of the sequences, to modify the aptamer (Appendix E) and see if they interfered with the stem-loop region, which are the predicted binding site (Debiais et al., 2020). Chimera was chosen as it gave the option to edit the structures. Hence, a functional group e.g. thiol (S) was added. The structure is generated using the SMILES translator. PyMOL was used to edit the 3D structure by using the provided 'Builder' option. Stabilising agents, such as available chemical compounds or proteins were added to see changes in the structure (Appendix H).

2.5 Aptamer and Ligand Docking

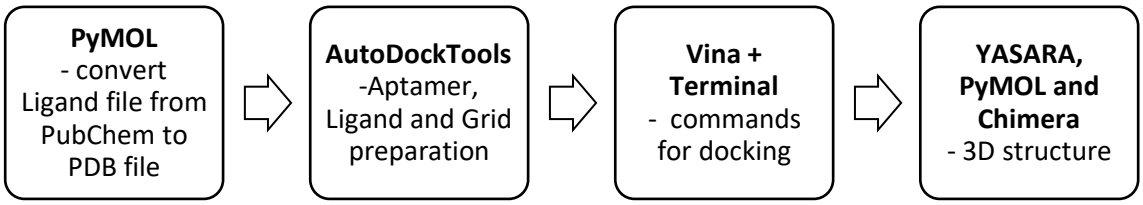


Figure 3. Step by Step guide on how to perform a docking.

The binding site was predicted by performing an aptamer-ligand docking (Appendix F). PubChem provided various chemical structures which were easily and freely downloadable, hence PEG was used as ligand as it had a simple structure. PyMOL was used to convert files into PDB format. Each aptamer and the ligand were prepared using AutoDockTools and PyMOL. The Grid showed where the ligand was most likely to bind. For the final docking step Vina and Terminal were used. Still pictures were obtained for each aptamer.

Additionally, to check if the ligand would interfere with the aptamers binding site, a docking was performed for SA43 (Appendix J).

3. Results

3.1 2D and 3D Structure

The figures obtained through the webservers Mfold, RNAComposer and software YASARA, PyMOL and Chimera were shown as a collection of figures for each six aptamers. The original aptamer structure can be compared to the modified versions such as the shortened structure or with the added stabilising agent. The binding site is also provided to see if the binding took place in the stem-loop region.

3.1.1 *SA43*

A1 in Fig.4 showed the 2D structure of SA43 with two hairpin loops between base-pairs G_3-C_{16} as well as A_{20}-T_{42} (Appendix J). Therefore, A_1, C_2 and C_{43} were cut off for the shortening step (Appendix D). 3C-3E (Fig.4) shows that the stabilising agent was located automatically on the first nucleotide (adenine) of SA43 by the software Chimera. Editing its position in the aptamer was not provided (Appendix I). 4C-4E (Fig.4) shows the predicted binding site was located in the stem-loop region between T_{25}-A_{37} (Appendix J).

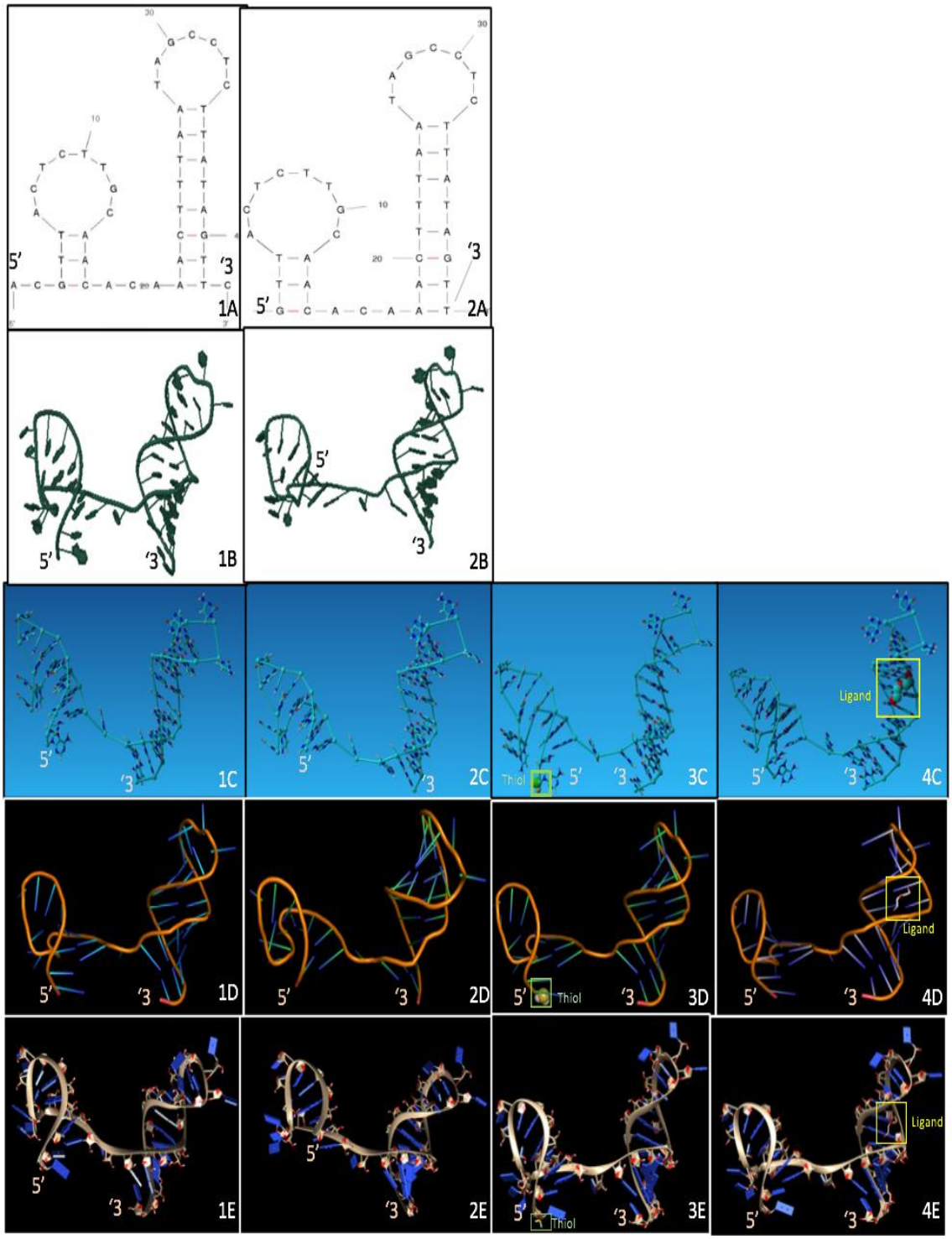


Figure 4. The 2D and 3D structures of SA43 were summarised. 1A-2A are the 2D structures from Mfold. 1B and 2B are the 3D structures using Vienna RNAfold. C1-C4 are the 3D structures using YASARA. D1-D4 are the 3D structures obtained through PyMOL and the 3D structures in 1E-4E are from Chimera. 1A-1E are the structures from the original SA43. 2A-2E is the shortened version. 3C-3E is the aptamer with the stabilising agent (green). 4C-4E shows the predicted binding site (yellow).

3.1.2 *S1p*

The 2D structure of S1p showed two hairpin loops (A1 in Fig.5). One between C_{13}-G_{30} and C_{53}-G_{82} (Appendix J). A_1-A_{12} and A_{92}-G_{94} were cut off from the original sequence (Appendix D). 2B in Fig.5 shows the obtained structure after removing unpaired bases at the 5'-3' ends of the sequence. In 3C-3E (Fig.5) the stabilising agent was located on the first nucleotide (adenine) of S1p, but its position could not be changed (Appendix I). In 4C-4E (Fig.5) the predicted binding site was located in the stem-loop region between C_{45}-G_{91} (Appendix J).

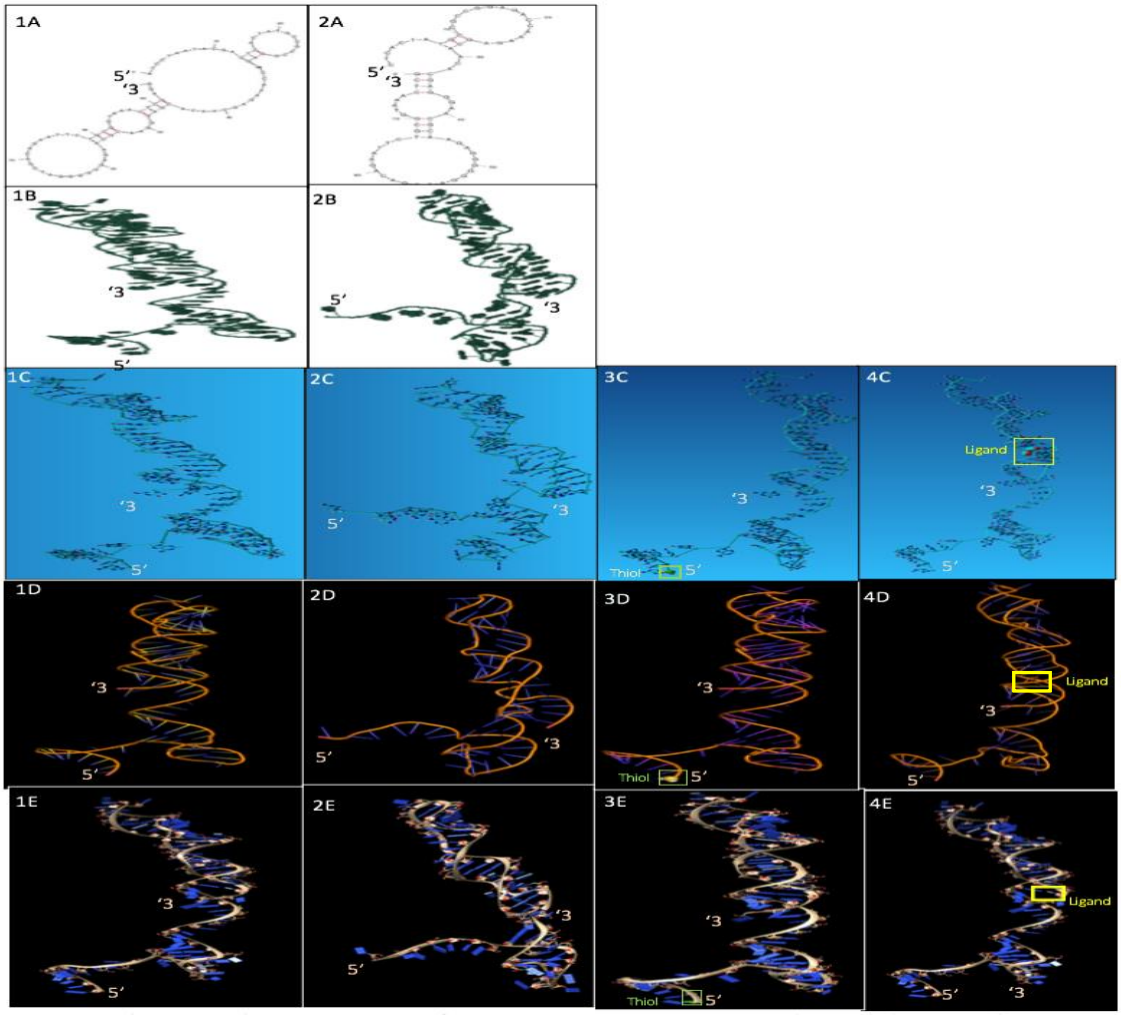


Figure 5. The 2D and 3D structures of aptamer S1p were collected. A1 and A2 are the 2D structures from Mfold. 1B and 2B are the 3D structures using Vienna RNAfold. 1C-4C are the 3D structures using YASARA. 1D-4D are the 3D structures obtained through PyMOL and the 3D structures in 1E-4E are from Chimera. 1A-1E are the structures from the original S1p. 2A-E is the shortened version. 3C-3E is the aptamer with the stabilising agent (green). 4C-4E shows the predicted binding site (yellow).

3.1.3 *EGFR8*

The 2D structure of EGFR8 had four hairpin structures between C_{16}-G_{34}, C_{44}-G_{61}, T_{62}-A_{76} and C_{84}-G_{104} (Appendix J). Hence, C_1-T_{11} and A_{106} and T_{107} are cut off for the shortening method (Appendix D). In 3C-3E (Fig.6) the stabilising agent was shown on the first nucleotide (cytosine) of EGFR8, but it was not possible to change its location (Appendix I). 4C-4E (Fig.6) shows the predicted binding site was located in stem-loop region between A_{12}-T_{43} (Appendix J).

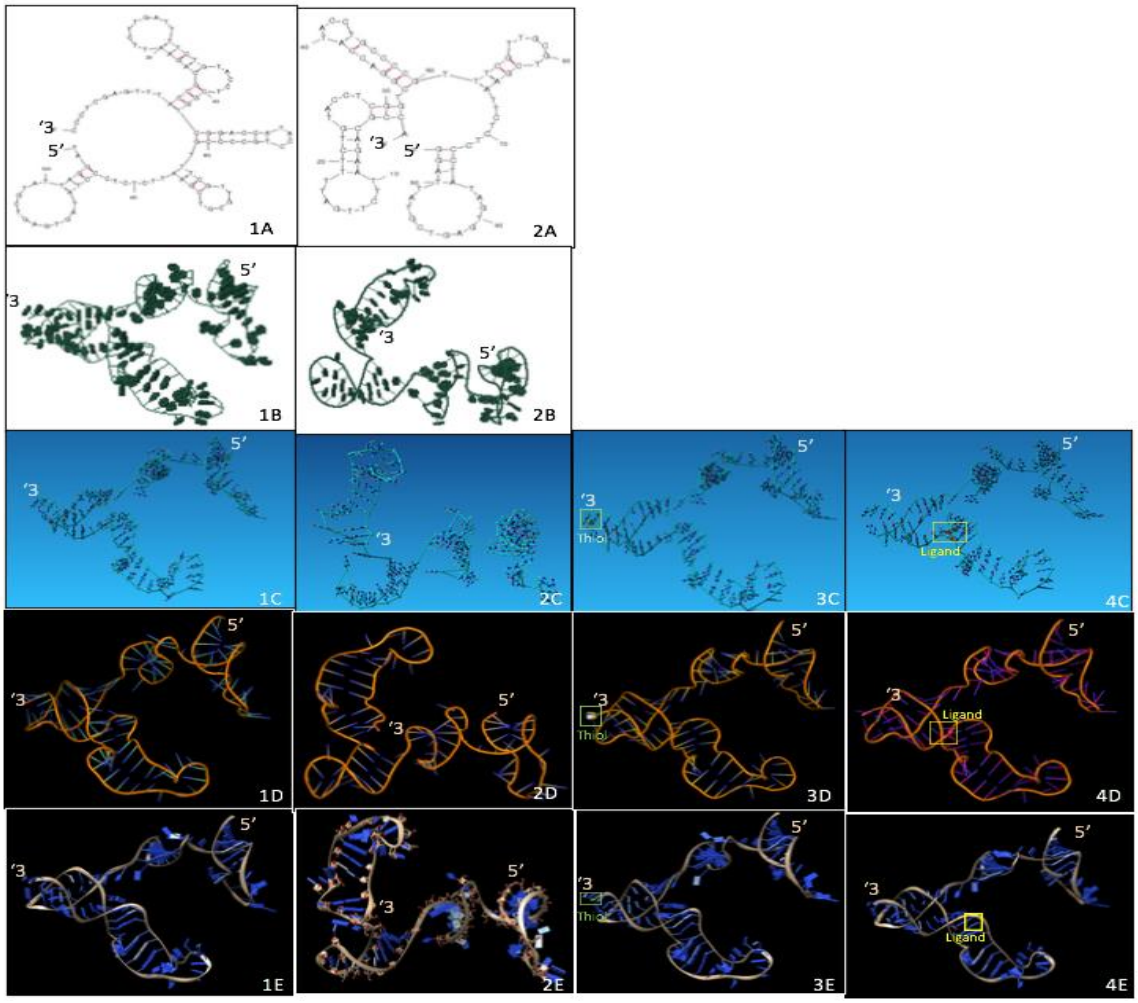


Figure 6. The 2D and 3D structures of aptamer EGFR8 are summarised. 1A and 2A are the 2D structures from Mfold. 1B and 2B are the 3D structures using Vienna RNAfold. 1C-4C are the 3D structures using YASARA. 1D-4D are the 3D structures obtained through PyMOL and the 3D structures in 1E-4E are from Chimera. 1A-1E are the structures from the original EGFR8. 2C-2E is the shortened version. 3C-3E is the aptamer with the stabilising agent (green). 4C-4E shows the predicted binding site of the ligand (yellow).

3.1.4 XApt004

The 2D structure of XApt004 had two hairpin loops between C_{39}-G_{62} and C_{84}-G_{104}. Hence, C_1-G_{16} and A_{105} and T_{106} were deleted to create a shortened sequence (Appendix D). 3C-3E (Fig.7) the stabilising agent was found in the first nucleotide (cytosine) of the sequence. Adjusting its location was not possible (Appendix I). 4C-4E (Fig. 7) shows the predicted binding site was located in the stem-loop region between C_{40}-C_{61} (Appendix J).

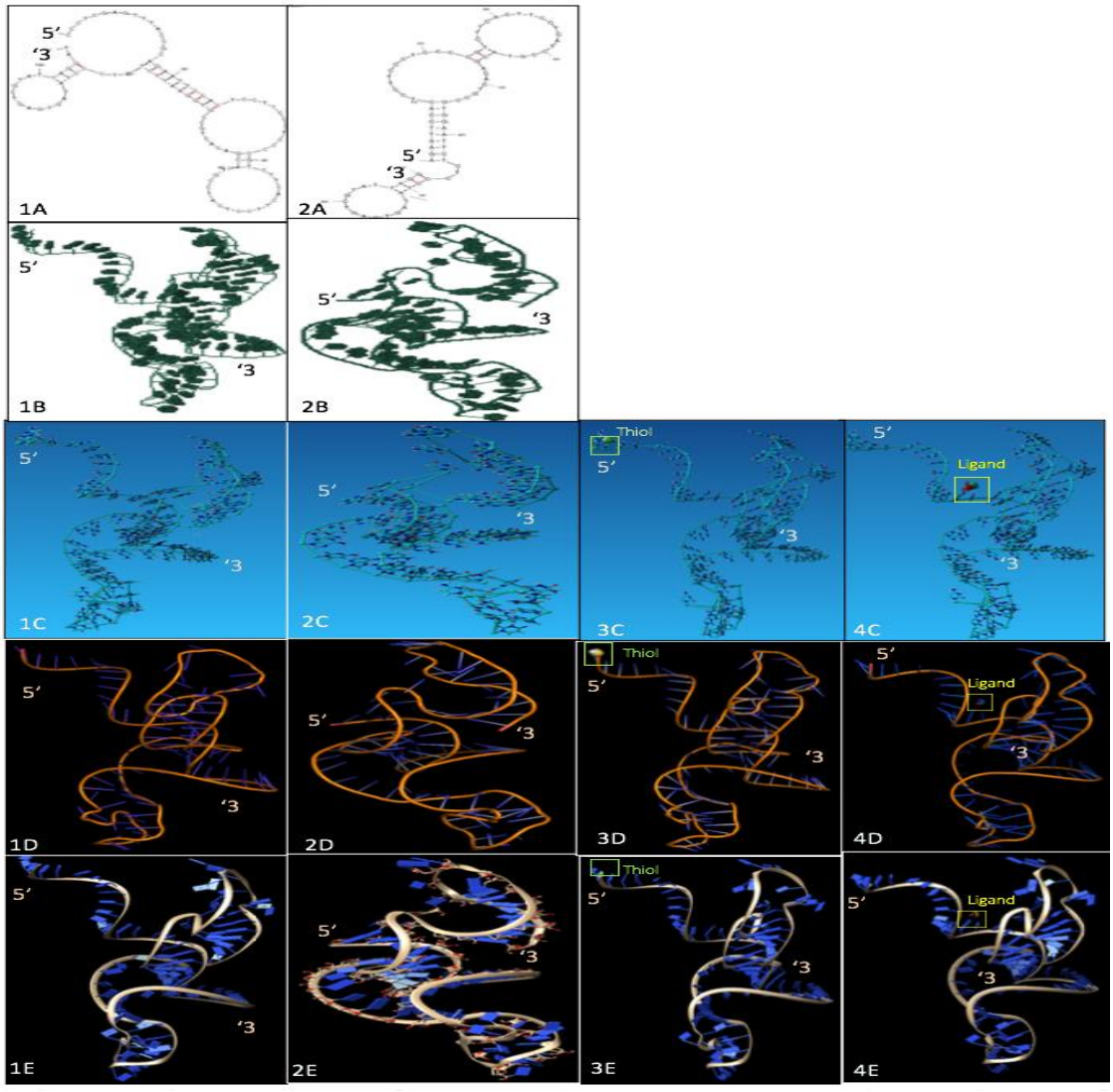


Figure 7. The 2D and 3D structures of aptamer XApt004 are summarised. 1A and 2A are the 2D structures from Mfold. 1B and 2B are the 3D structures using Vienna RNAfold. 1C-4C are the 3D structures using YASARA. D1-D4 are the 3D structures obtained through PyMOL and the 3D structures in 1E-4E are from Chimera. 1C-1E are the structures from the original XApt004. 2C-2E is the shortened version. 3C-3E is the aptamer with the stabilising agent (green). 4C-4E shows the predicted binding site (yellow).

3.1.5 *TDM1*

The 2D structure of TDMA1 (A1 in Fig.8) had three hairpin loops between C_3-G_{25}, G_{41}-C_{58} and C_{82}-G_{88}. A_1, T_2 and A_{89}-G_{91} were removed to shorten the TDM1 sequence (Appendix D). In 3C-3E (Fig.8) the stabilising agent was placed on the first nucleotide (adenine) of TDM1. It was not possible to adjust the location of the built structure (thiol) (Appendix I). 4C-4E (Fig.8) shows the predicted binding site was located in the stem-loop between T_{36}-A_{61} (Appendix J).

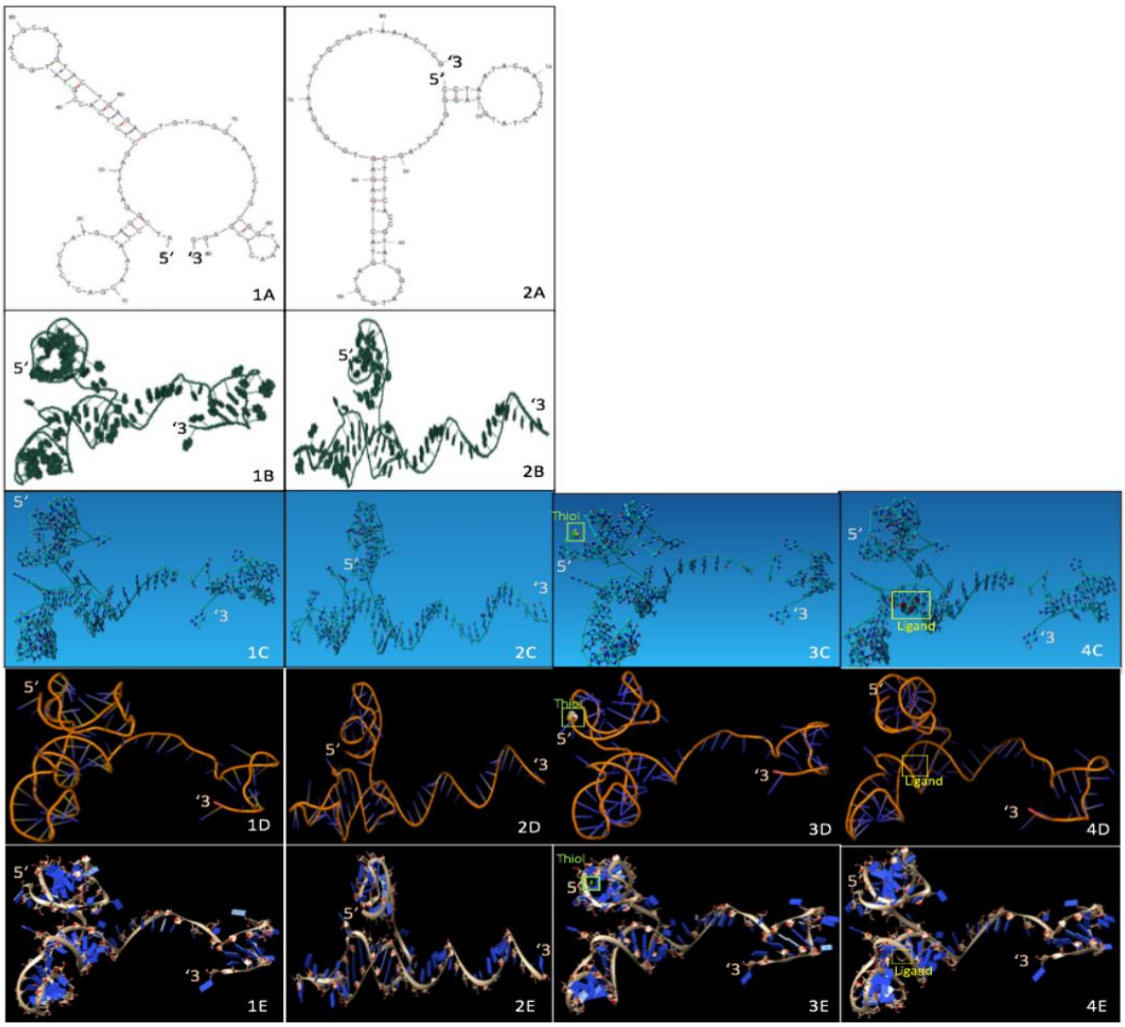


Figure 8. The 2D and 3D structures of aptamer TDM1 are summarised. 1A and 2A are the 2D structures from Mfold. 1B and 2B are the 3D structures using Vienna RNAfold. C1-C4 are the 3D structure using YASARA. 1D-4D are the 3D structures obtained through PyMOL and the 3D structures in E1-E4 are from Chimera. 1A-E are the structures from the original TDM1. 2C-2E is the shortened version. 3C-3E is the aptamer with the stabilising agent (green). 4C-4E shows the predicted binding site (yellow).

3.1.6 *hTMA10*

The 2D structure of hTMA10 (A1 in Fig.9) The 2D sequence of hTMA10 had three hairpin loops between T_{20}-A_{34}, C_{55}-G_{64} and C_{74}-G_{84}. A_1-A_9 and A_{85}-G_{87} were removed for the shortening process (Appendix D). 3C-3E (Fig.9) shows that stabilising agent was automatically located on the first nucleotide (adenine) (Appendix I). In 4C-4E (Fig.9) the binding site was located in the stem-loop region between C_{13}-G_{44} (Appendix J).

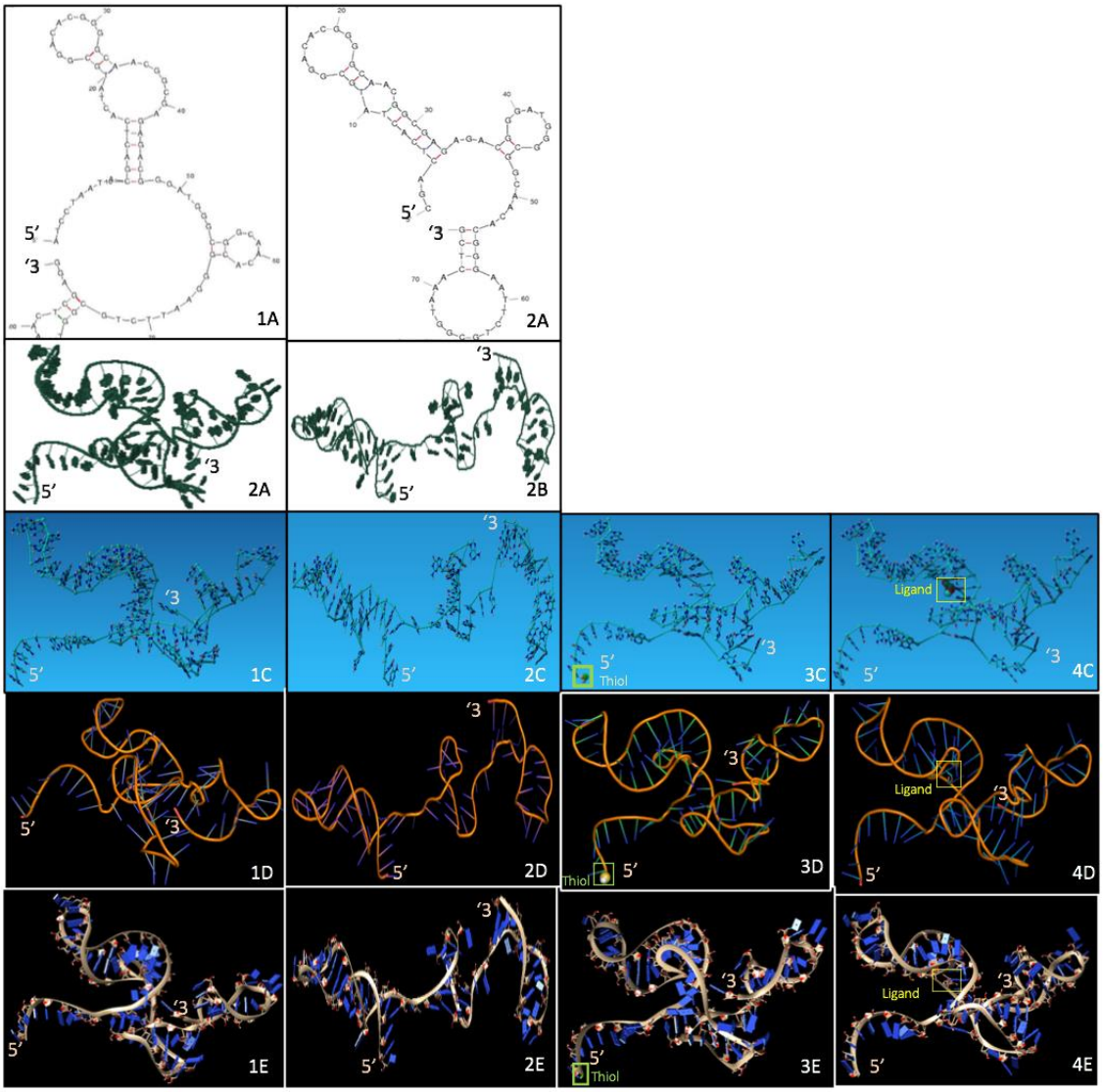


Figure 9. The 2D and 3D structures of aptamer hTMA10 are summarised. 1A and 2A are the 2D structures from Mfold. 1B and 2B are the 3D structures using Vienna RNAfold. 1C-4C are the 3D structure using YASARA. 1D-4D are the 3D structures obtained through PyMOL and the 3D structures in 1E-4E are from Chimera. 1A-1E are the structures from the original hTMA10. 2A-2E is the shortened version. 3C-3E is the aptamer with the stabilising agent (green). 4C-4E shows the predicted binding site of the ligand (yellow).

4. Discussion

Though there are various, insightful and free modelling programs on the internet, many of these do not run efficiently on some computers and have limited options to analyse aptamers (Bruno, 2017). Considering that most aptamers have been widely used in vivo, they are not well studied in the area of bioinformatic. In the last three decades, many computational methods have been developed to investigate aptamers and substantial achievements were obtained including the design of efficient aptamers via SELEX (Tuerk and Gold, 1990; Ellington and Szostak, 1990), simulating aptamer selection and interaction and structure prediction (Emami et al., 2020). To comprehend the possible implications of aptamers for future applications, a comprehensive characterization of theses single-stranded oligonucleotides is needed (Albada et al., 2015). This is important to obtain reliable 3D aptamer models (Zhong et al., 2020). By comparing theory with bioinformatical practice, the research will gain a fuller insight into aptamers, facilitate in vivo experiments by lowering costs with shortened aptamer sequences (Emami et al., 2020),

A study presented a method to perform an inexpensive 3D DNA molecular modelling using computational applications, which had been mainly used for protein–protein or protein–ligand interactions (Bruno, 2017). A challenge was obtaining PDB files of chosen structures. In this project, the most convenient way to obtain PDB files were by using PubChem to download structures and convert them in PyMOL or use RNAComposer. A study recommended web-based file translator sites e.g. National Cancer Institutes (NCI's) and SMILES Translator (Bruno, 2017). A limitation with the currently free and easily accessible software is that further modifications require more advanced computers and software that require bigger budgets. Additionally, software for chemical sketching e.g. ChemSketch, did not provide a PDB file download (Hunter, 1997; Wan et al., 2004).

The 2D structure was obtained by comparing Vienna RNAfold and Mfold. Mfold provided multiple structures whereas Vienna RNAfold provided one. These two webservers gave

the same structure results for SA43, but not the other aptamers. Vienna RNAfold did not require manually typed dot-bracket notations. Hence, Mfold was more reliable than Vienna RNAfold, as it was specially made for DNA folding. The dot-bracket notations of structures from Mfold had to be analysed and typed carefully to ensure correctness. Nonetheless, four out six obtained dot-bracket notations from this project agreed with the provided ones. Mfold (Zuker, 2003) was used to analyze low-throughput sequence data, but it is difficult to perform structure prediction on the HTS (high-through output sequencing) level (Kinghorn et al., 2017).

Another limitation was that DNA aptamers were modelled as RNA in RNAComposer. This issue was also encountered in a study (Bruno, 2017). It was suggested changing uridine back to thymine using Discovery Studio Visualizer software. For this project this was not necessary as the aptamer structures were not affected by the change of thymine to uracil. The 2D and 3D structure of all six given aptamer sequences were obtained as well as visualising a shortened version of each six aptamers.

It was not established how stabilizing agents would affect the aptamer's structure. Neither PyMOL nor Chimera were able to edit the chemical compounds of the aptamer as well as see changes when adding a new structure e.g. thiol. The aptamer domain commonly needs to form a unique ligand-binding pocket for specifically binding its ligand e.g. stabilising agent (Heiat, 2016). This locks the conformation of the aptamer domain. In this project Chimera provided the insight that the stabilising agent thiol (S) was automatically located at the first nucleotide of the sequence. In order to explain the nature of aptamer-ligand interactions, a broad characterization of the complexes is required (Albada et al., 2015). The results were obtained in a few seconds for the mentioned software.

For the 3D visualisation YASARA, PyMOL and Chimera were used. YASARA gave a good overview of all the components of a chemical structure, with the option to add multiple PDB files allowed viewing and hiding to keep a clean working space. Unfortunately, YASARA only allowed to view and 'hide' parts of the aptamers without a change in structure. YASARA gave no option to bind the aptamer and the newly added structure

to create a new structure. PyMOL was the easiest to use as files can be easily dragged into the editing box. PyMOL was able to label the structure facilitating the finding of specific nucleotides. The ends of the sequences were also coloured red making them easy to locate. Chimera can be used if a third 3D visualisation is needed. It had options to find and select specific parts of the structure e.g. backbones. Without the specific selection, the structures seem cluttered and the loops were hard to differentiate.

This project predicted the binding site by using AutoDockTools in combination with Vina and Terminal. The structures were visualised with YASARA, PyMOL and Chimera. Albada et al. (2015) showed that using AutoDock in the YASARA Structure software package was useful in performing docking simulations of ligand-aptamer models. The binding site for each aptamer was predicted. This validated that the loop region is the binding site of targets in aptamers as studies suggested (Kaur and Yung, 2012; Debiais et al., 2020). More studies have to be performed to establish and further validate computational aptamer optimization methods (Nonaka et al., 2012). Albada et al. (2015). validated the use of the YASARA Structure software package with respect to the docking of ligands to DNA aptamers by studying the DNA-aptamers for argininamide. Furthermore, Bruno (2017) provided a theoretical shape complementarity-based docking model using PatchDock and YASARA. These were examined for their agreement with the empirical ELASA colorimetric binding assay data.

The aptamer-ligand prediction of SA43 and KU Heterodimer was not possible due to the limitation of given errors by AutoDockTools. A study showed that the KU Heterodimer (Ku 70 and Ku 80) was identified as the most abundant target in the SA43 samples analysed by mass spectrometry, followed by subsequent confirmation by Western blot using commercially available antibodies (Aptekar et al., 2015).

Aptamers are suited for bioinformatic techniques and the developed guideline and methods from this project can benefit aptamer scientists. Future in vivo experiments can make use of computational methods to quickly simulate and freely obtain results prior to starting an experiment. As discussed above, some limitations of these computational methods exist. A constant development in computational methods and

sophisticated technology are expected to overcome these limitations in the future (Gong et al., 2017).

5. Conclusion

It is critical to not only perform research of aptamers in vivo but also develop computational methods for visualizing and modifying aptamers. It is apparent that the full potential of aptamer technology has not been yet established. Addressing the challenges presented in this project are vital for improving the 3D predictions of aptamers. Therefore, the webservers and software used in this project provided free and time saving methods to predict and give insight into aptamer structures with the potential to be used prior for future in vivo experiments to lower material cost and save time.

For the 2D prediction, Mfold is recommended as it gives a detailed description table of the location and the number of hairpins, which was important when deciding which Mfold structure to use. The aptamer can be stabilized by shortening and removing the unpaired bases at the start and end of the sequence. By docking the aptamers to a ligand, the binding site was predicted to be in the loop region. Concluding that the aptamers can be shortened for stabilisation without interfering with the binding site. The interference of stabilising agents, such as available chemical compounds in PyMOL, were visualised but did not determine a change of structure. Chimera is not able to provide evidence that a stabilising agent changes the structure. YASARA and PyMOL provide the clearest and detailed 3D aptamer structure.

6. Acknowledgments

My appreciation goes to my supervisors Dr. Lisa Shaw and Dr. David Michael Griffiths for their guidance. Thank you to my family especially my parents Maggie and Manolo Montalbo for their unconditional love and support. I am grateful to my boyfriend

Cameron Collison, his family and Celine-Justine Matibag for being my support systems and making the UK a home away from home.

References

Ahirwar, R., Nahar, S., Aggarwal, S., Ramachandran, S., Maiti, S. and Nahar, P., 2016. In silico selection of an aptamer to estrogen receptor alpha using computational docking employing estrogen response elements as aptamer-alike molecules. *Scientific reports*, *6*, p.21285.

Albada, H.B., Golub, E. and Willner, I., 2015. Computational docking simulations of a DNA-aptamer for argininamide and related ligands. *Journal of computer-aided molecular design*, *29*(7), pp.643-654.

Aptekar S, Arora M, Lawrence CL, Lea RW, Ashton K, Dawson T, et al. (2015) Selective Targeting to Glioma with Nucleic Acid Aptamers. PLoS ONE 10(8): e0134957.

Berg, K., Lange, T., Mittelberger, F., Schumacher, U. and Hahn, U., 2016. Selection and Characterization of an α6β4 Integrin blocking DNA Aptamer. *Molecular Therapy-Nucleic Acids*, *5*, p.e294.

Bruno, J.G., 2017. Do it yourself 3-dimensional aptamer-ligand molecular modeling. *Journal of Bionanoscience*, *11*(3), pp.183-188.

Caroli, J., Taccioli, C., De La Fuente, A., Serafini, P. and Bicciato, S., 2016. APTANI: a computational tool to select aptamers through sequence-structure motif analysis of HT-SELEX data. *Bioinformatics*, *32*(2), pp.161-164.

Cerchia, L., Esposito, C.L., Jacobs, A.H., Tavitian, B. and De Franciscis, V., 2009. Differential SELEX in human glioma cell lines. *PloS one*, *4*(11), p.e7971.

Chiang, S.C., Han, C.L., Yu, K.H., Chen, Y.J. and Wu, K.P., 2013. Prioritization of cancer marker candidates based on the immunohistochemistry staining images deposited in the human protein atlas. *PloS one*, *8*(11), p.e81079.

Chandola, C., Kalme, S., Casteleijn, M.G., Urtti, A. and Neerathilingam, M., 2016. Application of aptamers in diagnostics, drug-delivery and imaging. *Journal of biosciences*, *41*(3), pp.535-561.

Debiais, M., Lelievre, A., Smietana, M. and Müller, S., 2020. Splitting aptamers and nucleic acid enzymes for the development of advanced biosensors. *Nucleic Acids Research*, *48*(7), pp.3400-3422.

Ellington, A.D. and Szostak, J.W., 1990. In vitro selection of RNA molecules that bind specific ligands. *nature*, *346*(6287), pp.818-822.

Emami, N., Pakchin, P.S. and Ferdousi, R., 2020. Computational predictive approaches for interaction and structure of aptamers. *Journal of Theoretical Biology*, p.110268.

Eremeeva, E. and Herdewijn, P., 2019. Reprint of: Non Canonical Genetic Material. *Current opinion in biotechnology*.

Feigon, J., Dieckmann, T. and Smith, F.W., 1996. Aptamer structures from A to ζ. *Chemistry & biology*, *3*(8), pp.611-617.

Ferreira, C.S. and Missailidis, S., 2007. Aptamer-based therapeutics and their potential in radiopharmaceutical design. *Brazilian archives of biology and technology*, *50*(SPE), pp.63-76.

Flamme, M., McKenzie, L.K., Sarac, I. and Hollenstein, M., 2019. Chemical methods for the modification of RNA. *Methods*.

Floege, J., Ostendorf, T., Janssen, U., Burg, M., Radeke, H.H., Vargeese, C., Gill, S.C., Green, L.S. and Janjic, N., 1999. Novel approach to specific growth factor inhibition in vivo: antagonism of platelet-derived growth factor in glomerulonephritis by aptamers. *The American journal of pathology*, *154*(1), pp.169-179.

Gong, S., Wang, Y., Wang, Z. and Zhang, W., 2017. Computational methods for modeling aptamers and designing riboswitches. *International journal of molecular sciences*, *18*(11), p.2442.

Gopinath, S.C., Lakshmipriya, T., Md Arshad, M.K., Voon, C.H., Adam, T., Hashim, U., Singh, H. and Chinni, S.V., 2017. Shortening full-length aptamer by crawling base deletion–Assisted by Mfold web server application. *Journal of the Association of Arab Universities for Basic and Applied Sciences*, *23*(1), pp.37-42.

Gupta, S., Drolet, D.W., Wolk, S.K., Waugh, S.M., Rohloff, J.C., Carter, J.D., Mayfield, W.S., Otis, M.R., Fowler, C.R., Suzuki, T. and Hirota, M., 2017. Pharmacokinetic properties of DNA aptamers with base modifications. *nucleic acid therapeutics*, *27*(6), pp.345-353.

Hamada, M., 2018. In silico approaches to RNA aptamer design. *Biochimie*, *145*, pp.8-14.

Hanwell, M.D., Curtis, D.E., Lonie, D.C., Vandermeersch, T., Zurek, E. and Hutchison, G.R., 2012. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *Journal of cheminformatics*, *4*(1), p.17.

Healy, J.M., Lewis, S.D., Kurz, M., Boomer, R.M., Thompson, K.M., Wilson, C. and McCauley, T.G., 2004. Pharmacokinetics and biodistribution of novel aptamer compositions. *Pharmaceutical research*, *21*(12), pp.2234-2246.

Heiat, M., Najafi, A., Ranjbar, R., Latifi, A.M. and Rasaee, M.J., 2016. Computational approach to analyze isolated ssDNA aptamers against angiotensin II. *Journal of biotechnology*, *230*, pp.34-39.

Horhota, A., Zou, K., Ichida, J.K., Yu, B., McLaughlin, L.W., Szostak, J.W. and Chaput, J.C., 2005. Kinetic analysis of an efficient DNA-dependent TNA polymerase. *Journal of the American Chemical Society*, *127*(20), pp.7427-7434.

Hunter, A.D., 1997. ACD/ChemSketch 1.0 (freeware); ACD/ChemSketch 2.0 and its tautomers, dictionary, and 3D plug-ins; ACD/HNMR 2.0; ACD/CNMR 2.0.

Jäschke, A., Fürste, J.P., Nordhoff, E., Hillenkamp, F., Cech, D. and Erdmann, V.A., 1994. Synthesis and properties of oligodeoxyribonucleotide—polyethylene glycol conjugates. *Nucleic acids research*, *22*(22), pp.4810-4817.

Kasahara, Y., Irisawa, Y., Ozaki, H., Obika, S. and Kuwahara, M., 2013. 2′, 4′-BNA/LNA aptamers: CE-SELEX using a DNA-based library of full-length 2′-O, 4′-C-methylene-bridged/linked bicyclic ribonucleotides. *Bioorganic & medicinal chemistry letters*, *23*(5), pp.1288-1292.

Kaur, H. and Yung, L.Y.L., 2012. Probing high affinity sequences of DNA aptamer against VEGF 165. *PloS one*, *7*(2), p.e31196.

Kim, M.W., Jeong, H.Y., Kang, S.J., Choi, M.J., You, Y.M., Im, C.S., Lee, T.S., Song, I.H., Lee, C.G., Rhee, K.J. and Lee, Y.K., 2017. Cancer-targeted nucleic acid delivery and quantum dot imaging using EGF receptor aptamer-conjugated lipid nanoparticles. *Scientific reports*, *7*(1), pp.1-11.

Kinghorn, A.B., Fraser, L.A., Liang, S., Shiu, S.C.C. and Tanner, J.A., 2017. Aptamer bioinformatics. *International journal of molecular sciences*, *18*(12), p.2516.

Kong, H.Y. and Byun, J., 2013. Nucleic Acid aptamers: new methods for selection, stabilization, and application in biomedical science. *Biomolecules & therapeutics*, *21*(6), p.423.

Krieger, E., Vriend, G. and Spronk, C., 2013. YASARA–Yet Another Scientific Artificial Reality Application. *YASARA. org*, *993*.

Kuwahara, M., Obika, S., Nagashima, J.I., Ohta, Y., Suto, Y., Ozaki, H., Sawai, H. and Imanishi, T., 2008. Systematic analysis of enzymatic DNA polymerization using oligo-DNA templates and triphosphate analogs involving 2', 4'-bridged nucleosides. *Nucleic acids research*, *36*(13), pp.4257-4265.

Kuwahara, M. and Obika, S., 2013. In vitro selection of BNA (LNA) aptamers. *Artificial DNA: PNA & XNA*, *4*(2), pp.39-48.

Laing, C. and Schlick, T., 2010. Computational approaches to 3D modeling of RNA. *Journal of Physics: Condensed Matter*, *22*(28), p.283101.

Lakhin, A.V., Tarantul, V.Z. and Gening, L., 2013. Aptamers: problems, solutions and prospects. *Acta Naturae (англоязычная версия)*, *5*(4 (19)).

Li, B.Q., Zhang, Y.C., Huang, G.H., Cui, W.R., Zhang, N. and Cai, Y.D., 2014. Prediction of aptamer-target interacting pairs with pseudo-amino acid composition. *PLoS One*, *9*(1), p.e86729.

Li, Z., Wan, H., Shi, Y. and Ouyang, P., 2004. Personal experience with four kinds of chemical structure drawing software: review on ChemDraw, ChemWindow, ISIS/Draw, and ChemSketch. *Journal of Chemical Information and Computer Sciences*, *44*(5), pp.1886-1890.

Lorenz, A.R., Bernhart, S.H., Neubock, R. and Hofacker, I.L., 2008. The vienna rna websuite. *Nucleic Acids Res*, *36*, pp. W70-4.

Liu, B., Zhang, J., Liao, J., Liu, J., Chen, K., Tong, G., Yuan, P., Liu, Z., Pu, Y. and Liu, H., 2014. Aptamer-functionalized nanoparticles for drug delivery. *Journal of biomedical nanotechnology*, *10*(11), pp.3189-3203.

Lorenz, R., Bernhart, S.H., Zu Siederdissen, C.H., Tafer, H., Flamm, C., Stadler, P.F. and Hofacker, I.L., 2011. ViennaRNA Package 2.0. *Algorithms for molecular biology*, *6*(1), p.26.

Ma, H., Liu, J., Ali, M.M., Mahmood, M.A.I., Labanieh, L., Lu, M., Iqbal, S.M., Zhang, Q., Zhao, W. and Wan, Y., 2015. Nucleic acid aptamers in cancer research, diagnosis and therapy. *Chemical Society Reviews*, *44*(5), pp.1240-1256.

Madden, T.L., Tatusov, R.L. and Zhang, J., 1996. [9] Applications of network BLAST server. In *Methods in enzymology* (Vol. 266, pp. 131-141). Academic Press.

Mathews, D.H., Disney, M.D., Childs, J.L., Schroeder, S.J., Zuker, M. and Turner, D.H., 2004. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proceedings of the National Academy of Sciences*, *101*(19), pp.7287-7292.

Mei, H., Bing, T., Yang, X., Qi, C., Chang, T., Liu, X., Cao, Z. and Shangguan, D., 2012. Functional-group specific aptamers indirectly recognizing compounds with alkyl amino group. *Analytical chemistry*, *84*(17), pp.7323-7329.

Moody, E.M. and Bevilacqua, P.C., 2003. Thermodynamic coupling of the loop and stem in unusually stable DNA hairpins closed by CG base pairs. *Journal of the American Chemical Society*, *125*(8), pp.2032-2033.

Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K. and Olson, A.J., 1998. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of computational chemistry*, *19*(14), pp.1639-1662.

Nonaka, Y., Yoshida, W., Abe, K., Ferri, S., Schulze, H., Bachmann, T.T. and Ikebukuro, K., 2013. Affinity improvement of a VEGF aptamer by in silico maturation for a sensitive VEGF-detection system. *Analytical chemistry*, *85*(2), pp.1132-1137.

Opazo, F., Eiden, L., Hansen, L., Rohrbach, F., Wengel, J., Kjems, J. and Mayer, G., 2015. Modular assembly of cell-targeting devices based on an uncommon G-quadruplex aptamer. *Molecular Therapy-Nucleic Acids*, *4*, p.e251.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. and Ferrin, T.E., 2004. UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry*, *25*(13), pp.1605-1612.

Pieken, W.A., Olsen, D.B., Benseler, F., Aurup, H. and Eckstein, F., 1991. Kinetic characterization of ribonuclease-resistant 2'-modified hammerhead ribozymes. *Science*, *253*(5017), pp.314-317.

Popenda, M., Szachniuk, M., Blazewicz, M., Wasik, S., Burke, E.K., Blazewicz, J. and Adamiak, R.W., 2010. RNA FRABASE 2.0: an advanced web-accessible database with the capacity to search the three-dimensional fragments within RNA structures. *Bmc Bioinformatics*, *11*(1), p.231.

Popenda, M., Szachniuk, M., Antczak, M., Purzycka, K.J., Lukasiak, P., Bartol, N., Blazewicz, J. and Adamiak, R.W., 2012. Automated 3D structure composition for large RNAs. *Nucleic acids research*, *40*(14), pp.e112-e112.

Purzycka, K.J., Popenda, M., Szachniuk, M., Antczak, M., Lukasiak, P., Blazewicz, J. and Adamiak, R.W., 2015. Automated 3D RNA Structure Prediction Using the RNAComposer Method for Riboswitches1. In *Methods in enzymology* (Vol. 553, pp. 3-34). Academic Press.

Quirós, M., Gražulis, S., Girdzijauskaitė, S., Merkys, A. and Vaitkus, A., 2018. Using SMILES strings for the description of chemical connectivity in the Crystallography Open Database. *Journal of cheminformatics*, *10*(1), pp.1-17.

Rahimizadeh, K., AlShamaileh, H., Fratini, M., Chakravarthy, M., Stephen, M., Shigdar, S. and Veedu, R.N., 2017. Development of cell-specific aptamers: recent advances and insight into the selection procedures. *Molecules*, *22*(12), p.2070.

Ramlan, E.I. and Zauner, K.P., 2008. An extended dot-bracket-notation for functional nucleic acids.

Rouillard, J.M., Zuker, M. and Gulari, E., 2003. OligoArray 2.0: design of oligonucleotide probes for DNA microarrays using a thermodynamic approach. *Nucleic acids research*, *31*(12), pp.3057-3062.

Röthlisberger, P. and Hollenstein, M., 2018. Aptamer chemistry. *Advanced drug delivery reviews*, *134*, pp.3-21.

Ruiz Ciancio, D., Vargas, M.R., Thiel, W.H., Bruno, M.A., Giangrande, P.H. and Mestre, M.B., 2018. Aptamers as diagnostic tools in cancer. *Pharmaceuticals*, *11*(3), p.86.

Sefah, K., Shangguan, D., Xiong, X., O'donoghue, M.B. and Tan, W., 2010. Development of DNA aptamers using Cell-SELEX. *Nature protocols*, *5*(6), p.1169.

Shaw, J.P., Kent, K., Bird, J., Fishback, J. and Froehler, B., 1991. Modified deoxyoligonucleotides stable to exonuclease degradation in serum. *Nucleic acids research*, *19*(4), pp.747-750.

Song, K.M., Lee, S. and Ban, C., 2012. Aptamers and their biological applications. *Sensors*, *12*(1), pp.612-631.

The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC

Trott, O. and Olson, A.J., 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry*, *31*(2), pp.455-461.

Tsai, C.H., Chen, J. and Szostak, J.W., 2007. Enzymatic synthesis of DNA on glycerol nucleic acid templates without stable duplex formation between product and template. *Proceedings of the National Academy of Sciences*, *104*(37), pp.14598-14603.

Tuerk, C. and Gold, L., 1990. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *science*, *249*(4968), pp.505-510.

Wang, S., Zhang, Y.H., Lu, J., Cui, W., Hu, J. and Cai, Y.D., 2016. Analysis and identification of aptamer-compound interactions with a maximum relevance minimum redundancy and nearest neighbor algorithm. *BioMed research international*, *2016*.

Yan, J., Xiong, H., Cai, S., Wen, N., He, Q., Liu, Y., Peng, D. and Liu, Z., 2019. Advances in aptamer screening technologies. *Talanta*, *200*, pp.124-144.

Wu, X., Chen, J., Wu, M. and Zhao, J.X., 2015. Aptamers: active targeting ligands for cancer diagnosis and therapy. *Theranostics*, *5*(4), p.322.

Zhong, Y., Zhao, J. and Chen, F., 2020. Advances of aptamers screened by Cell-SELEX in selection procedure, cancer diagnostics and therapeutics. *Analytical Biochemistry*, p.113620.

Zhou, G., Latchoumanin, O., Bagdesar, M., Hebbard, L., Duan, W., Liddle, C., George, J. and Qiao, L., 2017. Aptamer-based therapeutic approaches to target cancer stem cells. *Theranostics*, *7*(16), p.3948.

Zhou, J. and Rossi, J., 2017. Aptamers as targeted therapeutics: current potential and challenges. *Nature reviews Drug discovery*, *16*(3), pp.181-202.

Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic acids research*, *31*(13), pp.3406-3415.

Zuker, M., Mathews, D.H. and Turner, D.H., 1999. Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide. In *RNA biochemistry and biotechnology* (pp. 11-43). Springer, Dordrecht.

Appendix

Appendix A: How to obtain the 2D structure of an aptamer

Mfold
Link: http://unafold.rna.albany.edu/?q=mfold

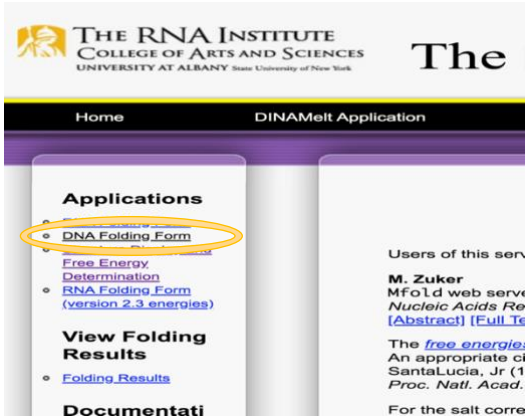


Figure 1. Shows the available applications. Click DNA Folding Form circled in orange.

Steps to follow:

- Insert sequence name e.g. SA43

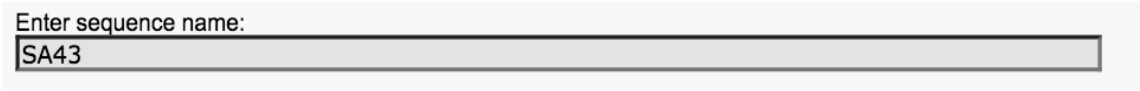


Figure 2. Shows the aptamer name in the box.

- Paste aptamer sequence

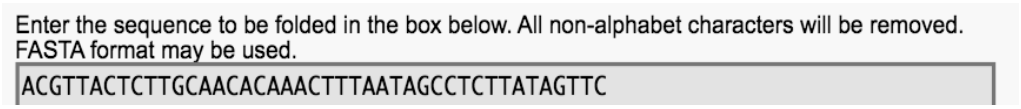


Figure 3. Shows the sequence added to the box.

- Click Format Sequence

Figure 3. Shows what to click to format the aptamer sequence.

- Obtain Sequence Detail

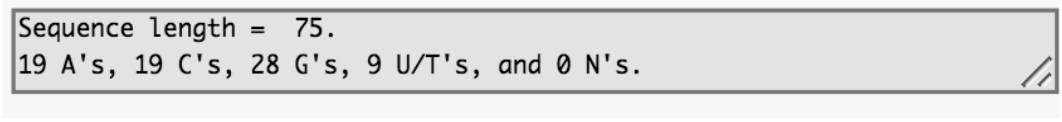


Figure 4. Shows the result when formatting the sequence.

- No constraints were chosen

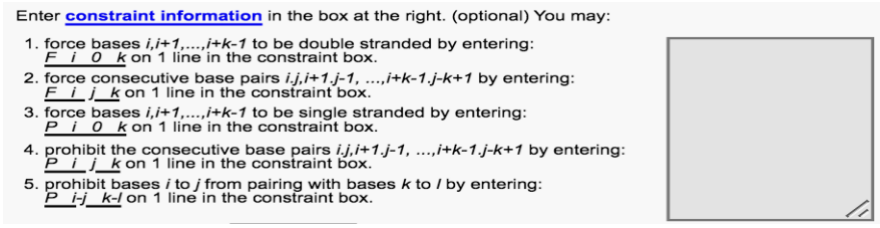


Figure 5. Shows that no constraints were added.

- Default settings (no need to change settings)

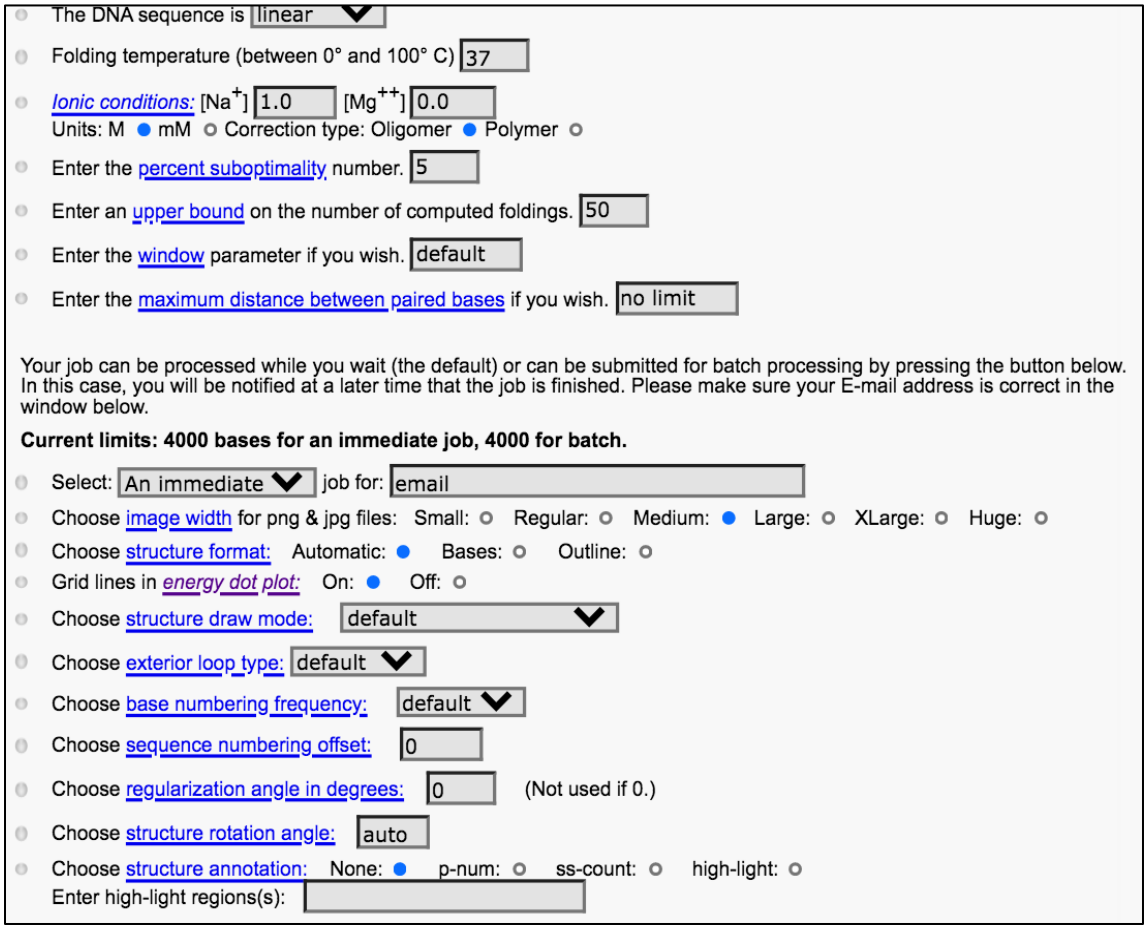


Figure 6. Shows the default setting in Mfold, when folding a DNA.

- Click Fold DNA

Figure 7. Shows what to click to obtain result.

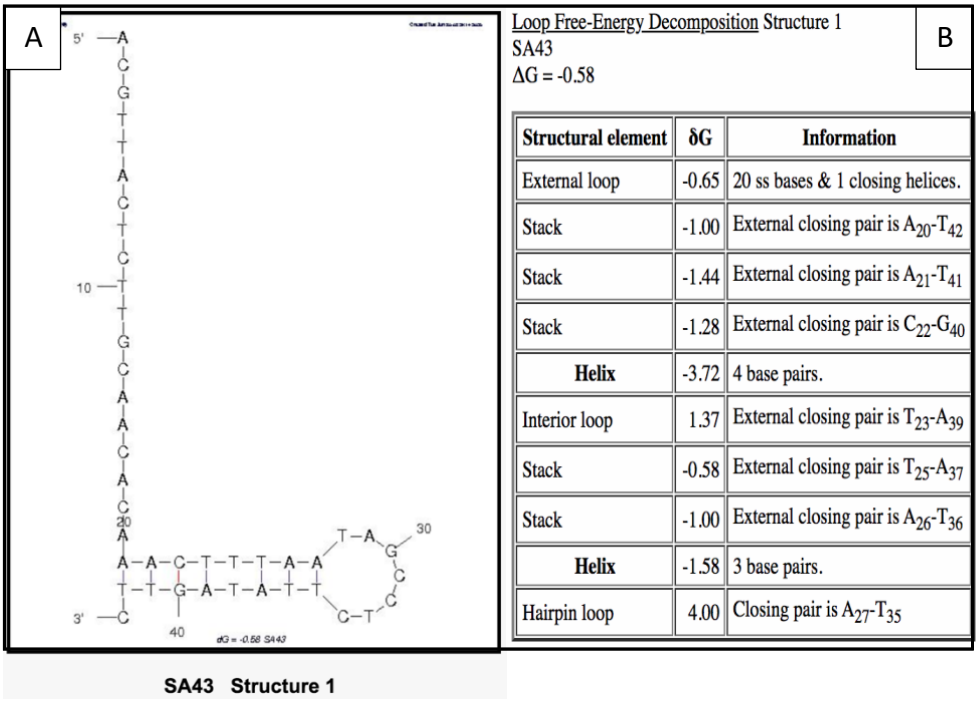
- Obtain predicted structures

Figure 8. Shows the available and predicted structures for aptamer SA43.

- Select a preferred file format to view the structure e.g. jpg

Figure 9. Shows what to click to obtain the 2D structure of the aptamer.

- Obtain 2D Structure and its detail



Structural element	δG	Information
External loop	-0.65	20 ss bases & 1 closing helices.
Stack	-1.00	External closing pair is A_{20}-T_{42}
Stack	-1.44	External closing pair is A_{21}-T_{41}
Stack	-1.28	External closing pair is C_{22}-G_{40}
Helix	-3.72	4 base pairs.
Interior loop	1.37	External closing pair is T_{23}-A_{39}
Stack	-0.58	External closing pair is T_{25}-A_{37}
Stack	-1.00	External closing pair is A_{26}-T_{36}
Helix	-1.58	3 base pairs.
Hairpin loop	4.00	Closing pair is A_{27}-T_{35}

Figure 10. A shows Structure 1 of SA43 obtained from Mfold. B shows the structural element of SA43.

- The dot-bracket notation has to be manually written by looking at the 2D structure. An unpaired nucleotide is written as a dot whereas a pair is a bracket (. If a paired base belongs to another stem-loop further bracket styles { are

used. When a base pair starts the bracket used is this ***(*** when the stem-loop ends the brackets need to be close ***)***.

As an example, the dot-bracket notation for SA43 (structure 1) would be as followed:

.....(((.(((.....))).))).

Vienna RNAfold

Link: http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi

The Vienna RNAfold Package predicts and compares secondary RNA structures.

- Paste the aptamer sequence of e.g. SA43

Figure 11. Shows the aptamer sequence pasted into the box.

- No constraints were chosen

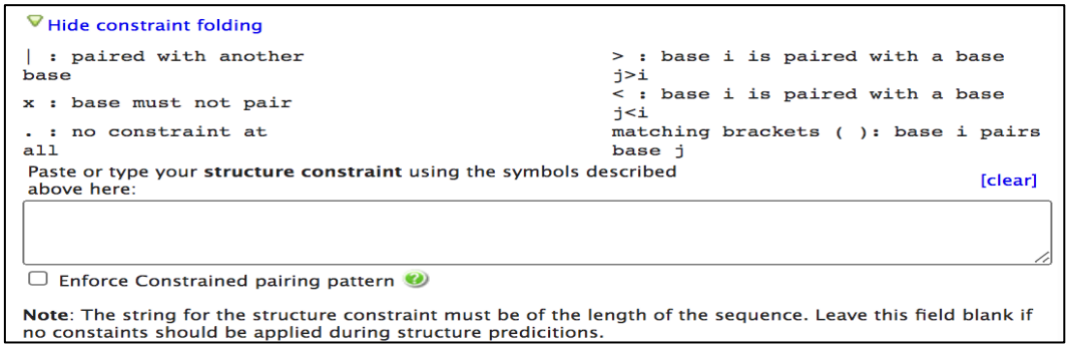


Figure 12. Shows that no constraints were added.

- Choose following settings

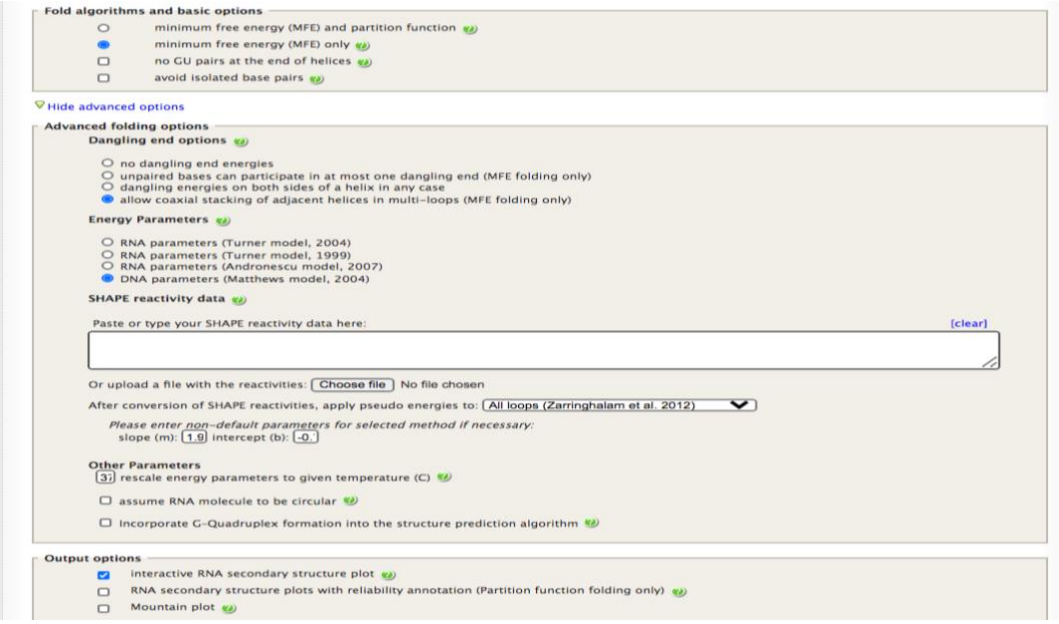


Figure 13. Shows the chosen setting for Vienna RNAfold.

- Click Proceed

Figure 14. Shows the button to click to obtain results.

- Obtain the 2D structure and the dot-bracket notation

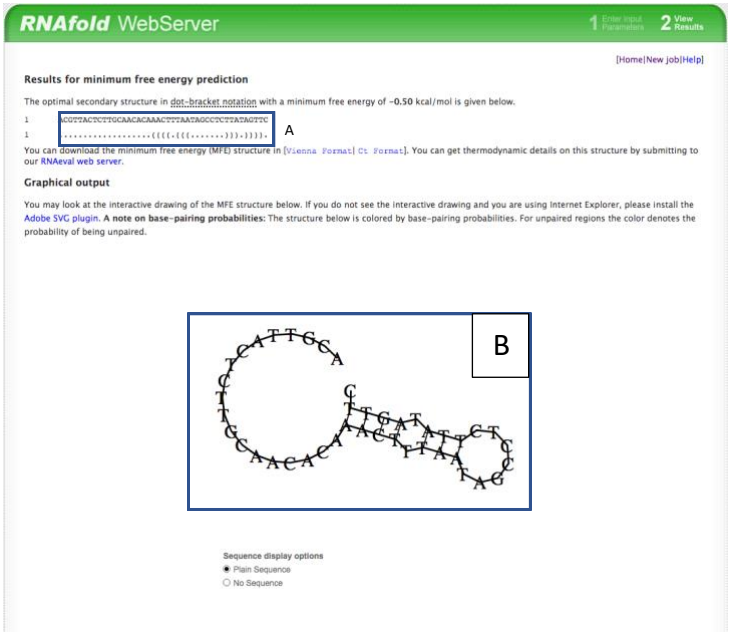


Figure 15. Shows the dot-bracket notation from Vienna RNAfold and the aptamers 2D structure.

Appendix B: Dot-bracket Mfold vs. Vienna RNAfold

Mfold predicts possible structures that are used to obtain the 3D structure.

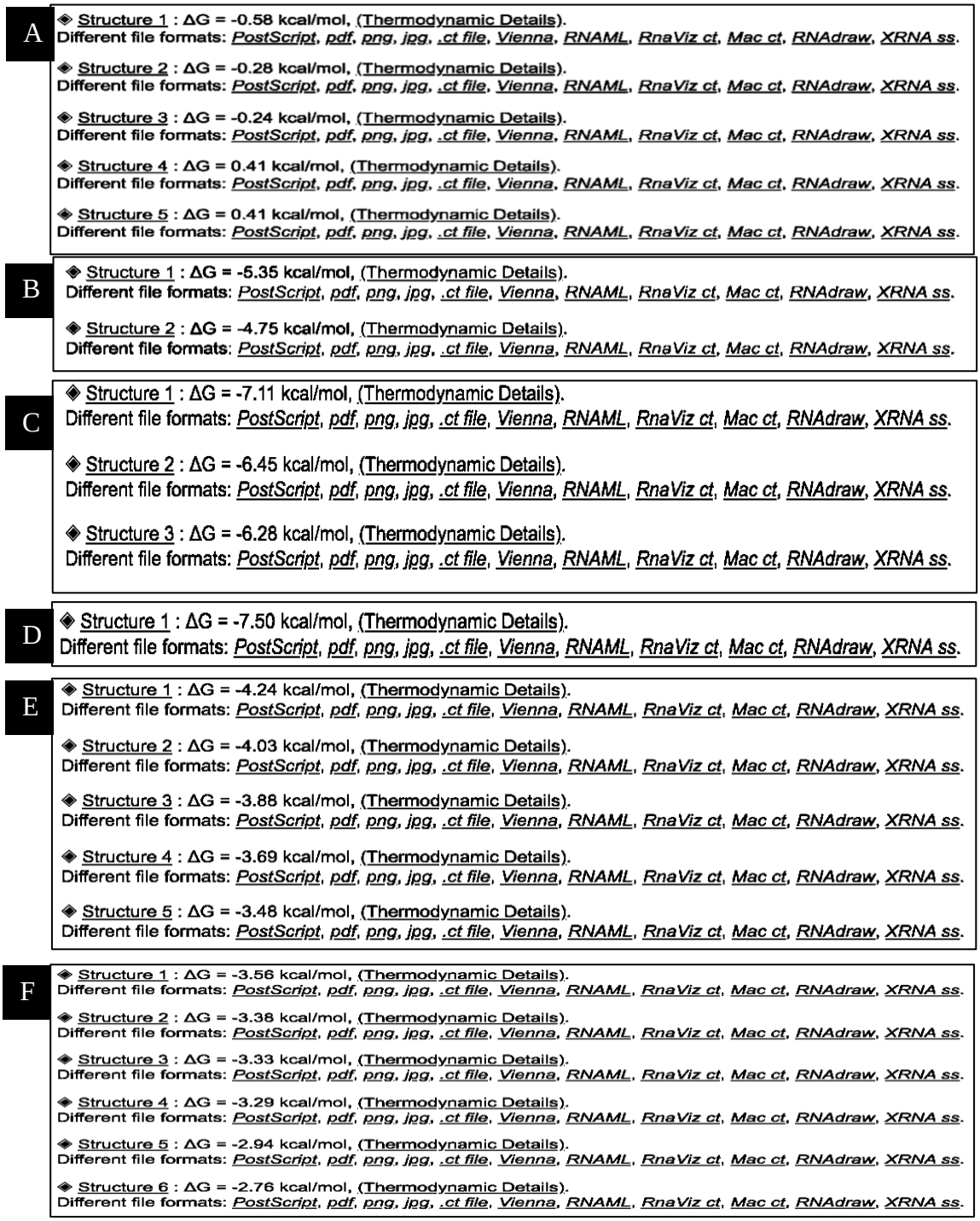


Figure 1. Shows the obtained structures for each aptamer (A-F) from the webserver Mfold. One or more structures for aptamers SA43 (A), S1p (B), EGFR8 (C), XApt004 (D) TDM1 (E) and hTMA10 (F) were chosen for the 3D visualisation.

The structures that matched the provided dot-bracket notations were chosen. For SA43 Structure 2 (Fig. 1A) was chosen to be visualised (Fig. 2A). S1p (Fig. 1B) (Fig. 2D) For EGFR8 Structure 1 (Fig. 1B) can be seen in Fig. 2F. The only predicted structure for XApt004 (Fig. 1D) can be seen in Fig. 2H. Structure 2 (Fig. 1E) for TDM1 was chosen to be visualised in Fig. 2J and Structure (Fig. 1F) for hTMA10 can be seen in Fig. 2L.

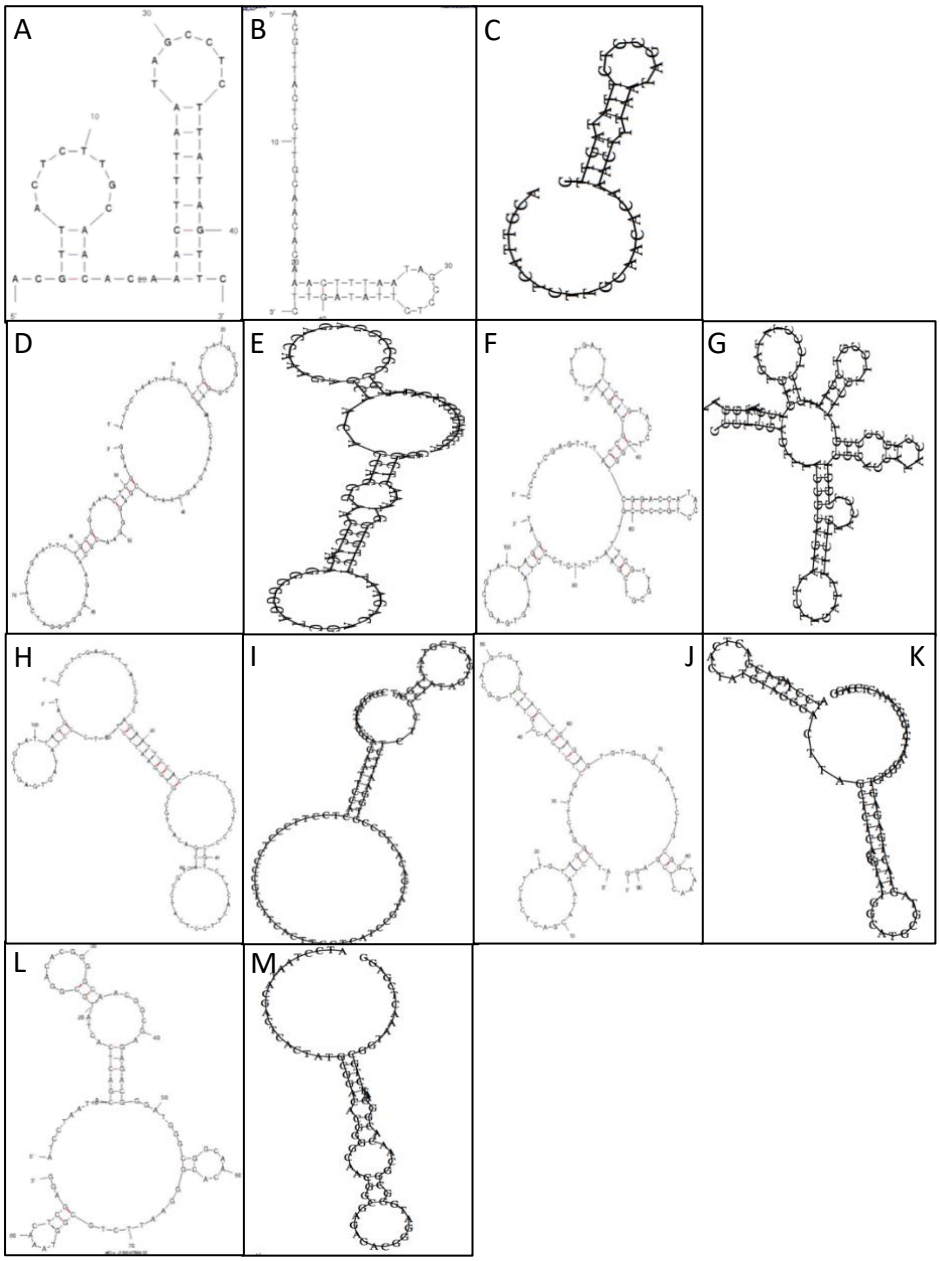


Figure 2. Shows the comparison between the chosen secondary structures of all the aptamers using Mfold (A, B, D, F, H, J and L) and Vienna RNAfold (C, E, G, I, K and M). A-C are the obtained 2D structures for aptamer SA43. A and B were obtained from Mfold whereas C was from Vienna RNAfold. Structure A gives the exact dot-bracket notation that was provided. B and C have the same dot-bracket notation. D and E are the 2D structures for aptamer S1p. D is from Mfold

whereas E is from Vienna RNAfold. F and H are the 2D structures for aptamer EGFR8. The structure F and G are described as 5'-3'- sequences as the given aptamer sequence is an anti-sense sequence 5' and 3' in each structure has to be switched. H and I are the 2D structures for aptamer XApt004. J and K are the 2D sequences for TMD1 L and M are the 2D sequences for hTMA10.

Structures A, D, F, were and their matching dot-bracket notations were used for predicting the 3D structure. The obtained SA43, S1p, XApt004 and TDM10, dot-bracket notation matched the given dot-bracket notation. EGFR8 and hTMA10 had to be newly written (Table 1.) due to the Structures F-H from Mfold not matching the given dot-bracket notation. B and C were the only structures that matched using both Mfold and Vienna RNAfold. Green base-pair binding between T and G was excluded and symbolized with a dot (Tabel 1).

Table 1. Shows the obtained Mfold and Vienna RNA dot-bracket notations for each aptamer.

Aptamer	Mfold	Vienna RNAfold
SA43	5'- ACGTTACTCTTGCAACACAAACTTTAATAG CCTCTTATAGTTC-3' ..(((........)))...((((.(((.......))).)))).	5'- ACGTTACTCTTGCAACACAAACTTTAATAGC CTCTTATAGTTC-3' .....((((.(((.......))).)))).
S1p	5'- ATCCTAATACGACTCACTATGCCGCCGGA GACCAAAGAGCAACACGAGGGAACGCAA GAGGGGGGATCGACAGAATTCTGCGGTA AACTCGAGG-3' .....(((.....))).....((((....((((..... .....))))).....))))...	5'- ATCCTAATACGACTCACTATGCCGCCGGAGA CCAAAGAGCAACACGAGGGAACGCAAGAG GGGGGATCGACAGAATTCTGCGGTAAACTC GAGG-3' .....(((.....))).....((((..(.((((.((..... .....))))))).)...))))....
EGFR8	3'- CCCTCGAGTTTACCGCAGAATTCTTGATTT TCTGTACCTCGGTCGGACCATACCTGCCC CGTTTCGTTGCGTCGAATTCTCTCCCTATA GTGAGTCGTATTAGGAT-5' .....((((((.....}}}}}.....))))(((..((....))..))).((((.....)))).....((((.....))))..	3'- CCCTCGAGTTTACCGCAGAATTCTTGATTTTC TGTACCTCGGTCGGACCATACCTGCCCCGTT TCGTTGCGTCGAATTCTCTCCCTATAGTGAGT CGTATTAGGAT-5' .((((((....(((((((((.....)))))).....))))(((..((....))..))). ((((.....)))).((((.....)))))).....)))..
XApt004	5'- CCCTCGAGTTTACCGCAGAATTCCACTCCT TCCCTCCCCGTCTTCACTTCCTCATCCGTAC	5'- CCCTCGAGTTTACCGCAGAATTCCACTCCTTC CCTCCCCGTCTTCACTTCCTCATCCGTACGAC

	GACACTGCCGTGGAATTCTCTCCCTATAG TGAGTCGTATTAGGAT-3’ .....((((((((((.....(((.....)))).....))))))))))...((((.....))))..	ACTGCCGTGGAATTCTCTCCCTATAGTGAGT CGTATTAGGAT-3’ .....((((((((((..... ...))))))))))...((((.....))))..
TDM1	5’- ATCCTAATACGACTCACTATGTAGGGACT TAGCTCTCACCGTATGGCATGCGTAGTAC TGAGAGTGTGGGAATTCTGCGGTAAACTC GAGG -3’ ..((((.....)))).....((((((..(((.....))))))))).....((.....))..	5’- ATCCTAATACGACTCACTATGTAGGGACTTA GCTCTCACCGTATGGCATGCGTAGTACTGAG AGTGTGGGAATTCTGCGGTAAACTCGAGG - 3’ .((((..((((.....)))))))....(((((((..((((.....))))))))))....
hTMA10	5’- ATCCTAATACGACTCACTATGCGGACACG GGGCAACGGCGAGAGACGGGATGGGCG GCAACACGGGAATTCTGCGGTAAACTCG AGG-3’ .....((.(((....(((.....)))).....))).))....((..... .))).....((.....))..	5’- ATCCTAATACGACTCACTATGCGGACACGGG GCAACGGCGAGAGACGGGATGGGCGGCAA CACGGGAATTCTGCGGTAAACTCGAGG-3’ .....((((((.((.(....((.(.....).))....).)).)...)))))....

The obtained dot-bracket from Mfold (Table 1.) were used to obtain the 3D structure.

Appendix C: How to predict the 3D structure of an aptamer

After obtaining the dot-bracket notation the 3D structure can be created using:

RNA Composer

Link: http://rnacomposer.cs.put.poznan.pl/

- Go to a web browser and use the webserver RNAComposer
- Delete the example in the 'interactive mode box'.
- Write > (name of your aptamer) e.g. >SA43 and paste the aptamer sequence and its dot-bracket notation.

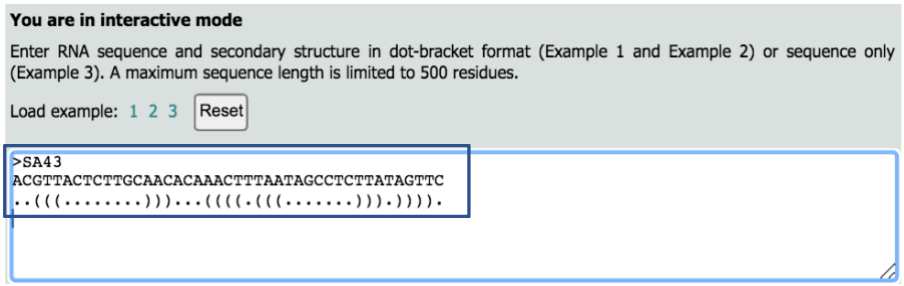


Figure 1. Shows the SA43 sequence and its dot-bracket notation pasted into the box.

- Click Compose

  Note: A message will appear that Characters T, t (Thymine) are replaced by U, u (Uracil) due to the webserver being used for RNA, but this will not affect the structure.

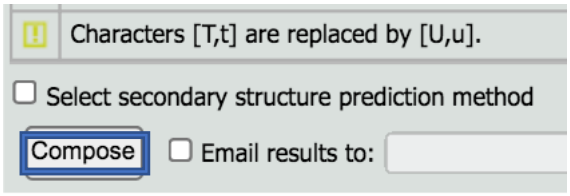


Figure 2. Shows the message about nucleotide T being replaced by U.

- Obtain the PDB file with the 3D structure by clicking SA43.pdb. The file will be automatically downloaded. A simple moveable preview of how the 3D structure of the aptamer can be viewed (right).

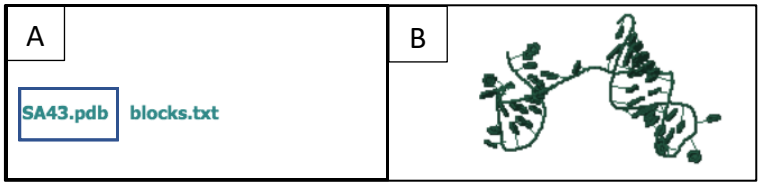


Figure 3. A shows what to click to download the aptamer SA43's PDB file. B shows a preview of the 3D structure of the aptamer.

YASARA (Yet Another Scientific Artificial Reality Application)
Link: http://www.yasara.org/downloads.htm

Figure 4. Shows the symbol for the software YASARA.

YASARA View was downloaded by registering (type in name and e-mail address).

Instruction on how to download YASARA was sent via email with a license to use it.

- Open YASARA

- Upload the PDB file of the chosen structure -> Click File -> Load -> PDB file -> Select the needed file.

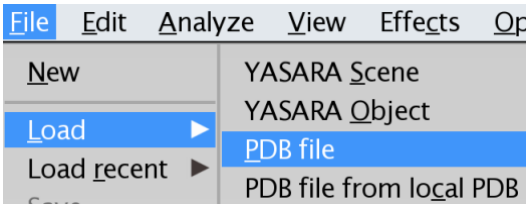


Figure 5. Shows what to click to upload the aptamer's file.

- The 3D structure is available to be viewed.

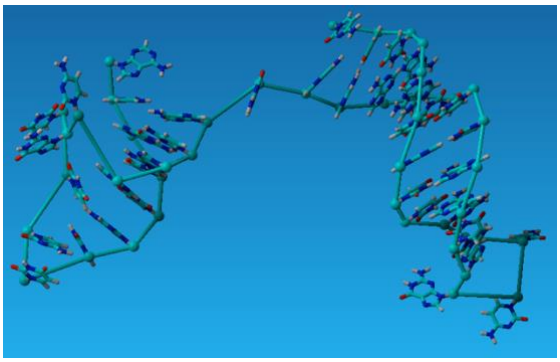


Figure 6. Shows the 3D structure of SA43.

- The nucleotide sequence can be viewed by hoovering the pointer of the mouse at the bottom of the screen.

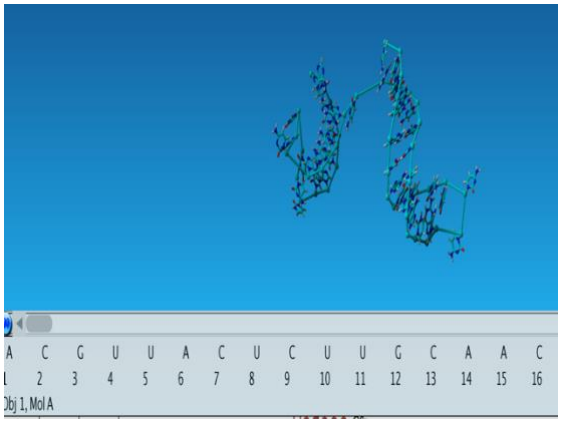


Figure 7. Shows the aptamer's nucleotides sequence.

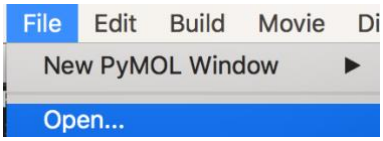
- Multiple PDB files can be added at the same time and visualised or made invisible by using the SCENE CONTENT and use the column Vis -> Click Yes and the structure will 'disappear -> Click No -> it appears again.

Figure 8. Shows how to view or hide the structures.

PyMOL

Download Link: https://pymol.org/2/

- Open PyMOL
- Drag file into black editing box or by clicking File -> Open -> Select the chosen



PDB file.

Figure 9. Shows how to upload the PDB file.

- Obtain the 3D structure.
- View nucleotide sequence by clicking -> L (Label) -> residues -> Click L and clear to remove again.

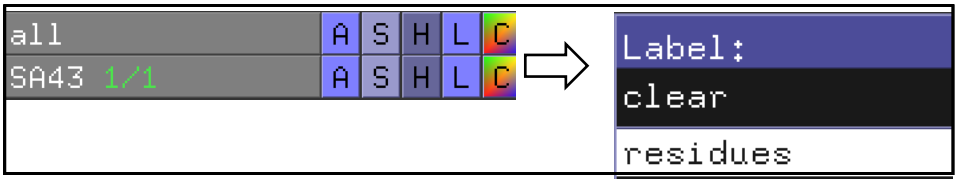


Figure 10. Shows how to obtain the residues to label each nucleotide in the structure.

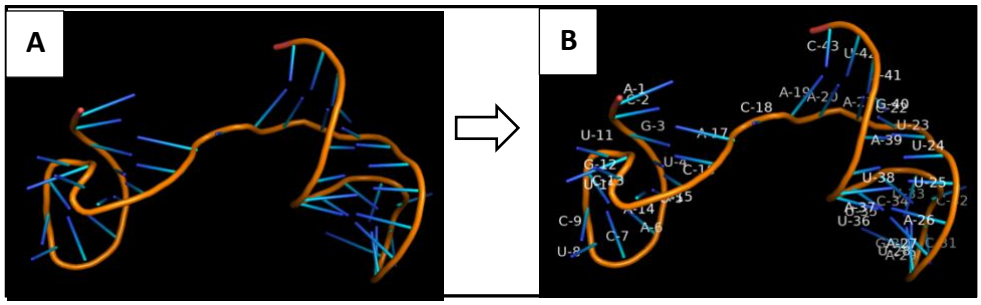


Figure 11. A shows the 3D structure of SA43. B shows the labelled aptamer with the nucleotide and the sequence number.

- Add Surface -> Click S (Shows) -> Surface.

Figure. Shows where to obtain the surface option.

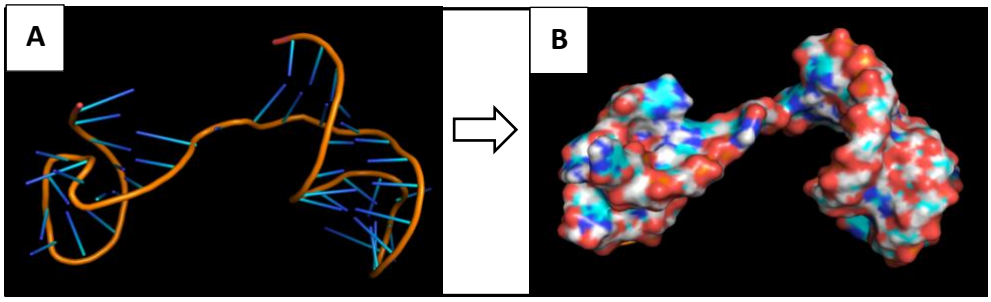


Figure. A shows the simple 3D SA43 structure, whereas B shows the aptamer with the surface visible.

Chimera

Link: https://www.cgl.ucsf.edu/chimera/download.html

- Open Chimera
- Click Browse in the bottom right corner to upload a file.

- Add nucleotide description -> Click Actions -> Label -> residue -> name or 1 - letter code.

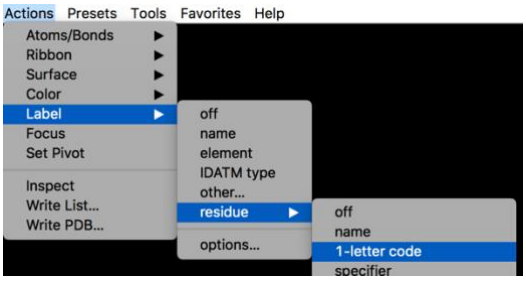


Figure. Shows what to click to add the labelling on the 3D structure of an aptamer.

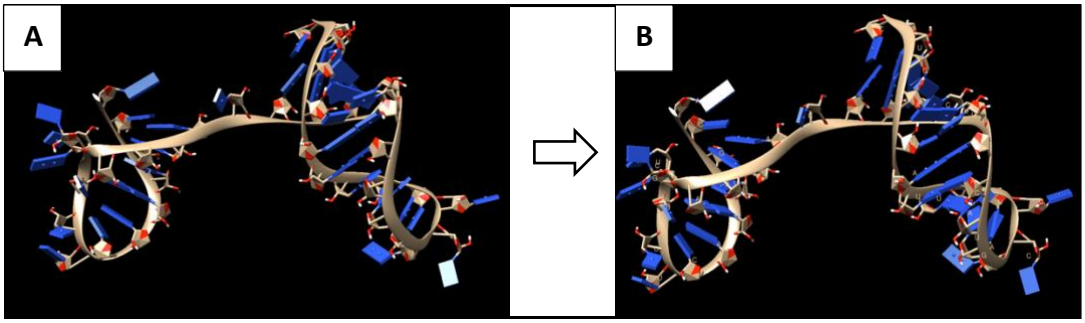


Figure. A shows the 3D structure without the nucleotide being labelled. B is the aptamer with the added nucleotide labels.

Appendix D: How to shorten an aptamer for stabilisation

- Obtain 2D structure by using the webserver Mfold
- Choose the structure with the best and most hairpin stem-loops (few to no bulge loops)
- Write down dot-bracket notation of the chosen structure
- Look at the structure and remove nucleotides from the 5' and 3' end that are unpaired from the original aptamer. These are considered to be the 'floppy' ends.
- Insert the newly written and shortened aptamer into Mfold and obtain the new 2D structure.
- Write down the new shortened dot-bracket notation.
- Insert the shortened aptamer sequence and its dot-bracket notation into the webserver RNAComposer for the 3D structure.
- Obtain the PDB file, which can be opened in various software (Appendix C).

Table 1. Shows the shortened aptamer sequences and their matching dot-bracket notation obtained through Mfold.

Aptamer Name	**Shortened Sequence + dot-bracket notation**
SA43	GTTACTCTTGCAACACAAACTTTAATAGCCTCTTATAGTT (((........)))...((((.(((.......))).))))
S1p	ACCGCAGAATTCTTGATTTTCTGTACCTCGGTCGGACCATACCTGCCCCGTTT CGTTGCGTCGAATTCTCTCCCTATAGTGAGTCGTATTAGG .......(((.................)))...((((....((((.....................))))).....))))
EGFR8	ACCGCAGAATTCTTGATTTTCTGTACCTCGGTCGGACCATACCTGCCCCGTTT CGTTGCGTCGAATTCTCTCCCTATAGTGAGTCGTATTAGG .(((({{{{{.........}}}}}.....))).(((..((....))..))).((((......)))).......((((.............))))
XApt004	AGAATTCCACTCCTTCCCTCCCCGTCTTCACTTCCTCATCCGTACGACACTGCC GTGGAATTCTCTCCCTATAGTGAGTCGTATTAGG ((((((((((.............(((...................)))........))))))))))...((((.............))))
TDM1	CCTAATACGACTCACTATGTAGGGACTTAGCTCTCACCGTATGGCATGCGTA GTACTGAGAGTGTGGGAATTCTGCGGTAAACTCG ((((..............))))..........((((((..((((..........)))))))))).....
hTMA10	CGACTCACTATGCGGACACGGGGCAACGGCGAGAGACGGGATGGGCGGCA ACACGGGAATTCTGCGGTAAACTCG ...(((.(..(((.........)))...).)))...((.......))......((.(..............).))

Appendix E: How to build and add a structure (stabilising agent) to an aptamer

SMILES (Simplified Molecular Input Line Entry Specification)
Link: https://search.molinstincts.com/search/searchTextList.ce

Chimera
Link: https://www.cgl.ucsf.edu/chimera/download.html

Figure 1. Shows the symbol for the Chimera application.

- Open Chimera.
- Upload the aptamer PDB file.
- Click Tools -> Structure Editing -> Build Structure

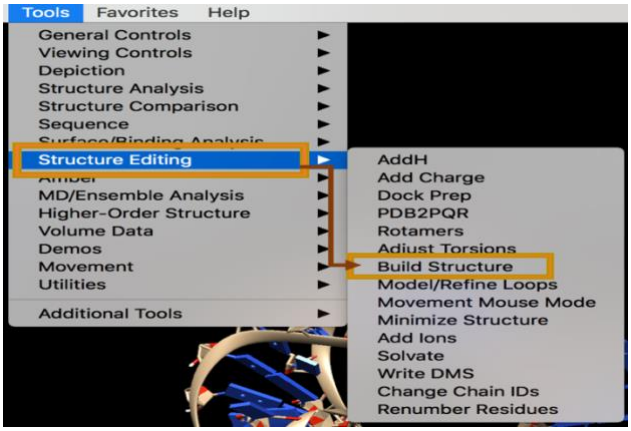


Figure 2. Shows what to click to start building a structure.

- Create functional group e.g. thiol (S) to be added as a stabilising agent using the SMILES string parameter -> Click Apply.

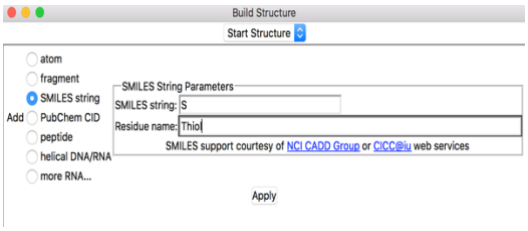
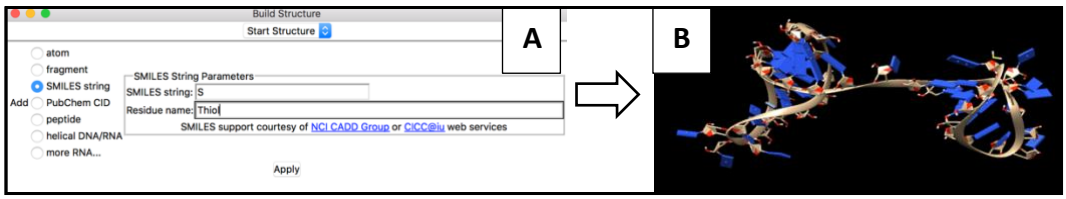


Figure 3. Shows the window for the structure building.

- The created structure e.g. thiol is randomly located in the structure.



- Figure 4. A shows what to click to start building a structure. B shows the aptamer with thiol being added to the structure.

- No further editing such as binding it to a specific location in the aptamer structure is available.

- Click Favourites -> Model Panel -> Click copy/combine ->

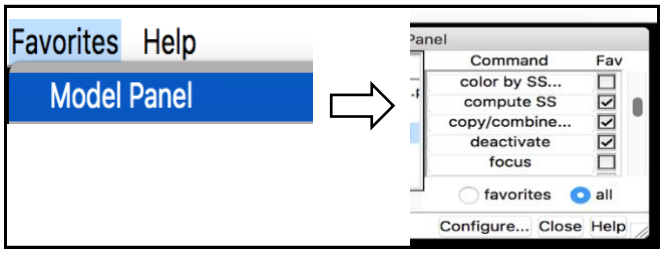


Figure 5. Shows what to click to get the copy/combine option.

- A window will pop open -> name the new model e.g. SA43Thiol to configure the thiol and SA43 to obtain the aptamer with the built structure.

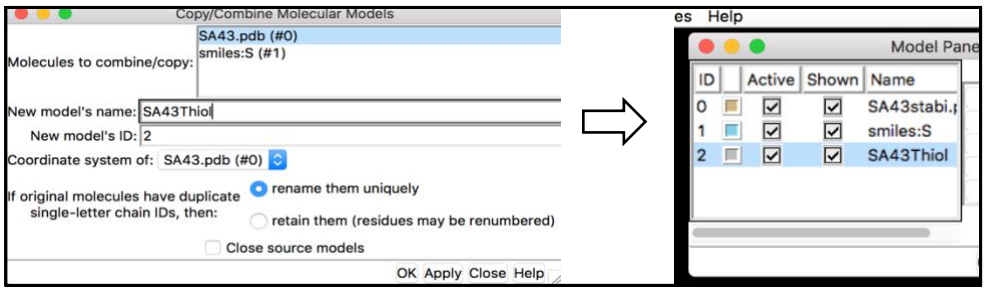


Figure 6. Shows the window for the copy/combine option.

- To save the new Click File -> Save PDB -> Click OK

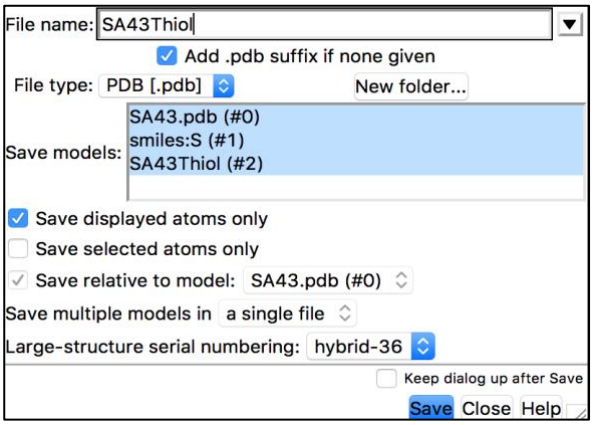


Figure 7. Shows the window for obtaining the PDB file.

- Choose a folder for the new structures and name -> Click OK

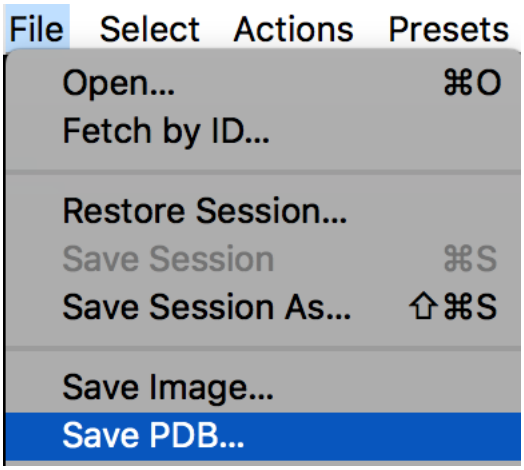


Figure 8. Shows what to click to save the structure as a PDB file

- Obtain PDB file containing the structure and aptamer.

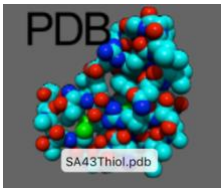


Figure 9. Shows how the symbol for a PDB file looks like.

- Visualise the obtained structure using YASARA.

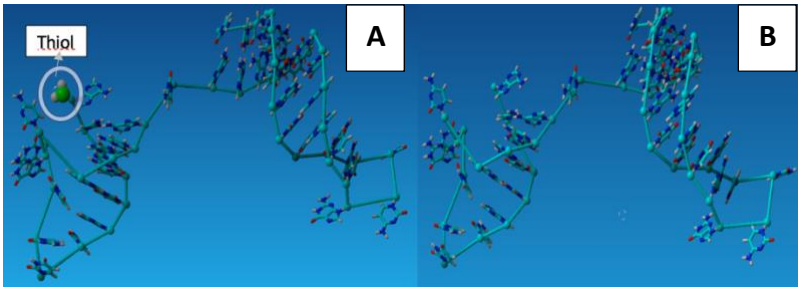


Figure 10. The structure from Chimera in which thiol and SA43 was combined (left) has not changed in YASARA. The 'original' SA43 can be seen on the right. The thiol was automatically positioned at the first nucleotide (A) of the sequence.

Appendix F: How to dock an aptamer with a ligand

Software and Webservers

AutoDockTools

Link: http://autodock.scripps.edu/resources/adt

This program is needed in combination with PyMOL and YASARA if a ligand is bound to the aptamer.

Figure 1. Shows the software symbol for AutoDockTools.

Command Prompt

- The command application that is available on a Microsoft computer.

MGLTools

Link: http://mgltools.scripps.edu/

This is the software needed to obtain AutoDockTools. The aim is to prepare the receptor and ligand for docking as well as obtaining the PDBQT file.

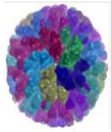


Figure 2. Shows the software symbol for MGLTools.

PyMOL

Download Link: https://pymol.org/2/

Can be used to convert it into a PDB file and visualise 3D structures.

Figure 3. Shows the software symbol for PyMOL.

Terminal

- The command application that is available on a Mac.

Figure 4. Shows the software symbol for Terminal.

XQuartz

Link: https://www.macosforge.org

XQuartz was downloaded as my Mac had no X11 application, opening

AutoDockTools would have not worked.

Figure 5. Shows the software symbol for AutoDockTools.

VINA

Link: http://vina.scripps.edu/download.html

VINA is not a software. It is used via Terminal (Mac) or Command Prompt (Windows) by typing in commands.

IMPORTANT:

'Bin' is the pre-made folder provided after downloading VINA. It only contains two files. To facilitate the command in Terminal, move the two into a folder with your aptamer and ligand e.g. folder named docking. To ensure that the docking will work all the PDBQT files of the prepared aptamer and ligand, txt file with the x y z coordinates and the downloaded vina and vina_split have to be put into a folder.

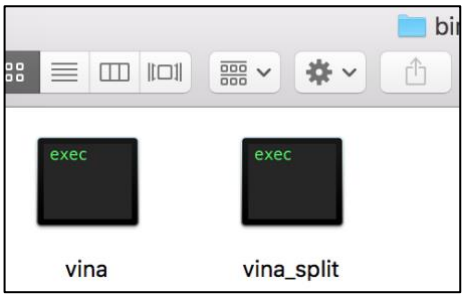


Figure 6. Shows the needed files that have to be in the same folder as all the PDBQT files for docking.

1. <u>Preparation of Aptamers</u>

- Upload the aptamer PDB file.

Figure 7. Shows the symbol to upload files.

- Click Edit.

- Delete Water molecules, they are in the pocket region and can interfere with the docking of the ligand, ligand won't sit comfortably.

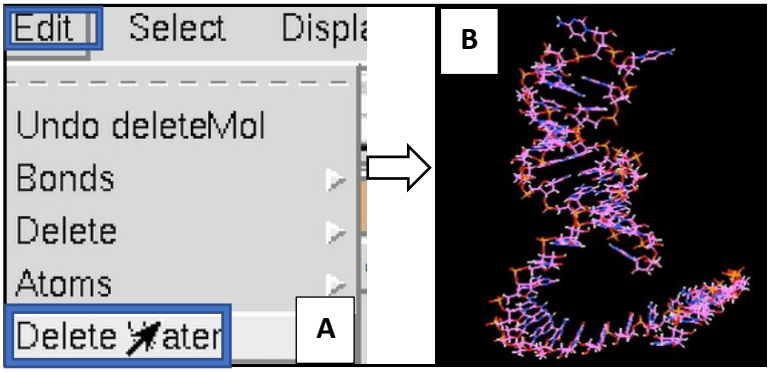


Figure 8. A shows what to click to prepare the aptamer for docking (A). B shows how the structure looks after deleting the water molecules.

- Add charges: Kollmann Charge is the most frequently used setting.

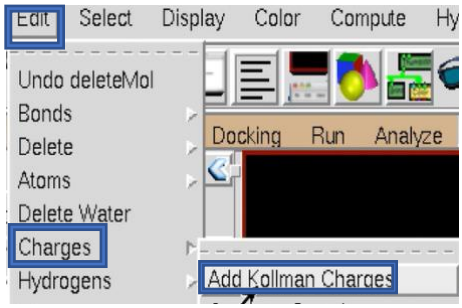


Figure 9. Shows the steps on how to add Kollman Charges. These charges are default values for each amino acid derived from the corresponding electrostatic potential using quantum mechanics

- A window will pop up -> Press OK.

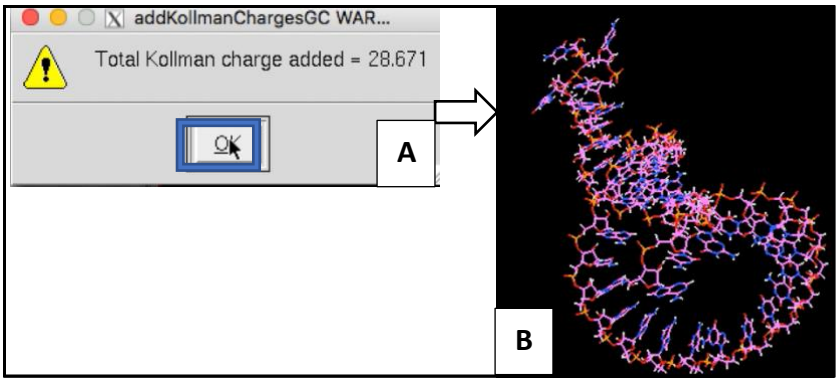


Figure 10. A shows the pop-up window when adding Kollman Charges. B shows how the structure looks after adding the Kollman Charges.

- Click Hydrogens -> Add -> A window will pop up -> Click only polar -> Press OK.

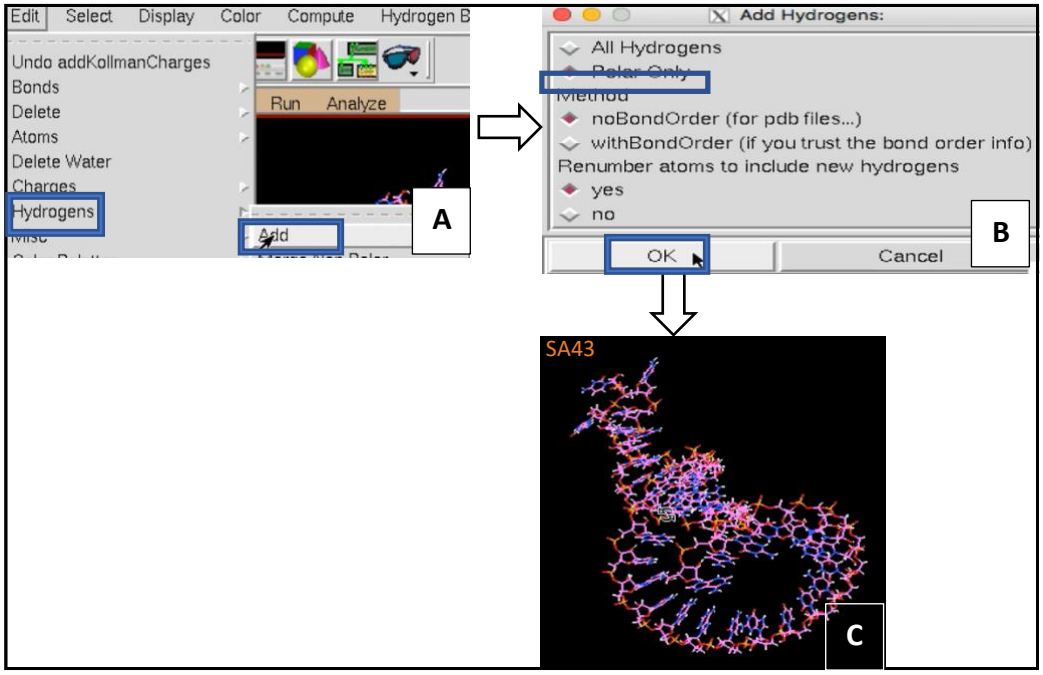


Figure 11. A shows the pop-up window when adding Kollman Charges (A and B). C shows how the structure looks after adding the Kollman Charges.

The prepared aptamer has to be saved.

- Click Grid -> Select Macromolecules -> Select aptamer -> Select molecules -> save file It is important to convert the file into a PDBQT file.

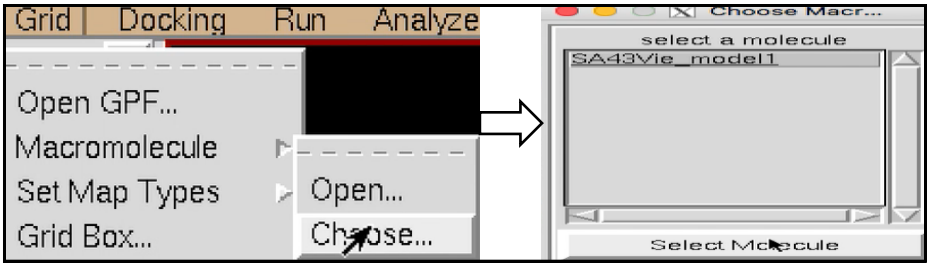


Figure 12. Shows the step on how to save the structure as a PDBQT file.

2. <u>Preparation of Ligand</u>

- Choose a e.g. Ligand: PEG DI(HYDROXYETHYL)ETHER from PubChem.

  Link: https://pubchem.ncbi.nlm.nih.gov/substance/7889788

- Downloaded the chosen ligand's SDF file, if a PDB file is available download this. A SDF file will not be recognised by AutoDockTools.

- Use PyMOL to convert SDF file into a PDB file by dragging the ligand file into the software window.

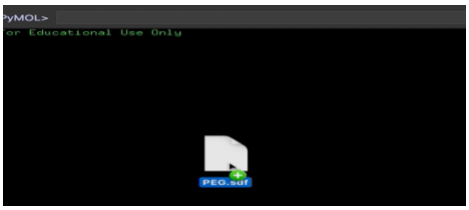


Figure 13. Shows how the SDF file is dragged into PyMOL.

- PyMOL is able to visualise the 3D structure of the ligand PEG using a SDF file

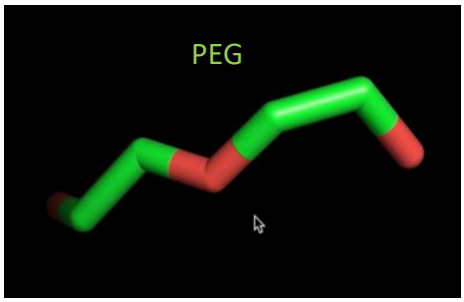


Figure 14. Shows the 3D structure of the chosen ligand PEG.

- Click File in PyMOL and select Export Molecule.

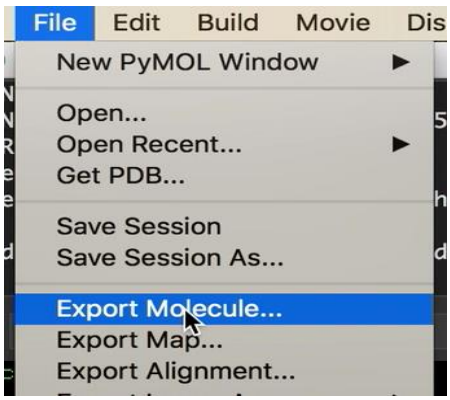


Figure 15. Shows how to save the ligand as a PDB file.

- Select the Molecule to be saved.

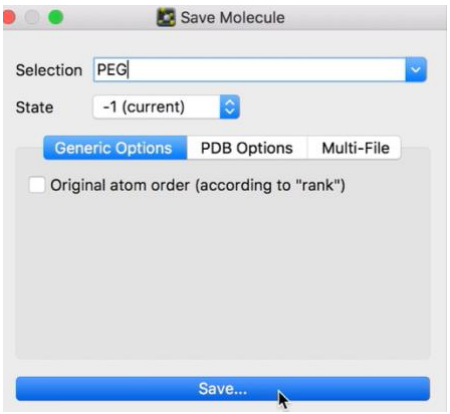


Figure 16. Shows how to save the ligand as a PDB file.

- Type in the ligand's name and choose PDB as a file type.

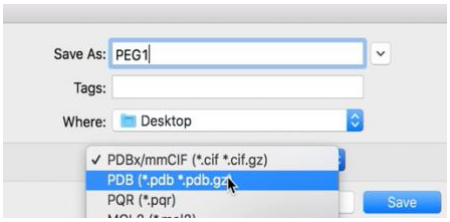


Figure 17. Shows how to save the ligand as a PDB file.

Once the ligand has been converted into a PDB file it can be opened in AutoDockTools to be prepared for docking.

- Open AutoDockTools
- Upload the Ligand in the AutoDockTools software.

Figure 18. Shows the symbol to upload files.

- Click Ligand -> Input -> Choose -> A new window will appear -> Select the Ligand for AutoDock 4.

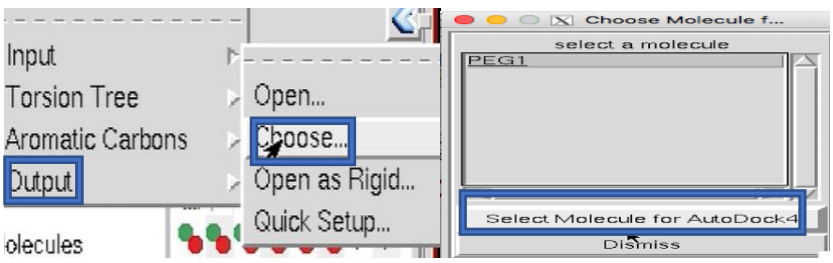


Figure 19. Shows how to select the ligand e.g. PEG for docking.

- A new window will pop up -> Click OK.

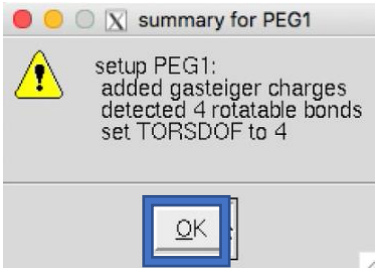


Figure 20. Shows the pop-up window with the summary of the ligand.

- The ligand will be ready for docking.

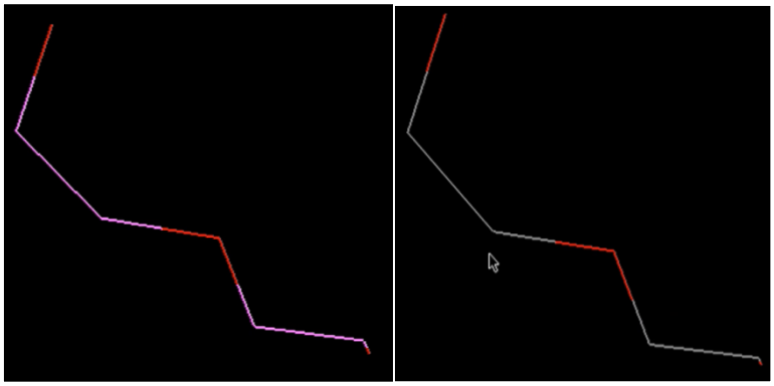


Figure 21. Shows the unprepared ligand PEG (left) and the prepared PEG (right).

To save the prepared ligand:

- Click Ligand -> Output -> Save as PDBQT

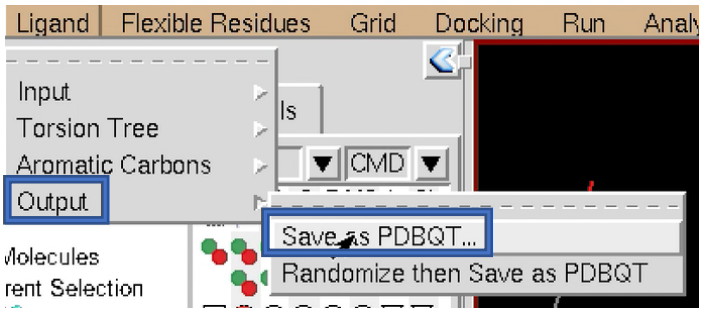


Figure 22. Shows how to save a file as a PDBQT.

- Do not close AutoDockTools and do not delete the prepared ligand.
-

3. GRID Preparation

The Grid shows where the ligand is most likely to bind by showing a square box. Continue with the already opened AutoDockTools and the prepared ligand (PEG) from Step 2.

- Upload the aptamer PDB file obtained through RNAComposer (Appendix c).

Figure 23. Shows the symbol to access the files to be uploaded.

- Click Grid-> Macromolecules-> Choose -> A window will pop up -> Select molecule (the uploaded aptamer) for the docking.

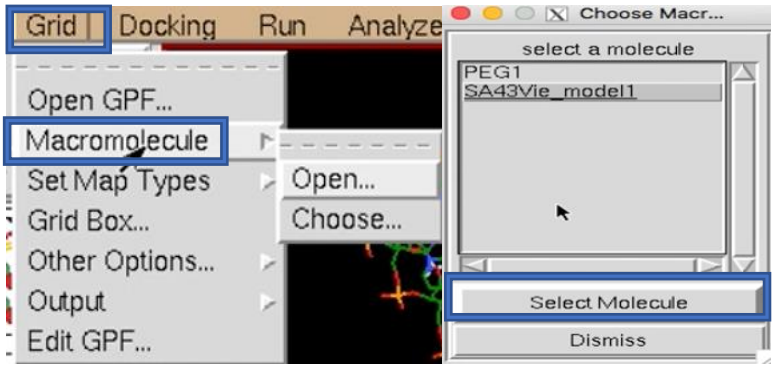


Figure 24. Shows how to select the molecule for docking.

- A new window will pop up -> Press Yes.

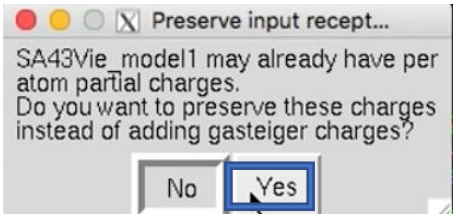


Figure 25. Shows the pop-up window asking if the charge of the structure should be preserved.

- Another window will pop up -> Press OK.

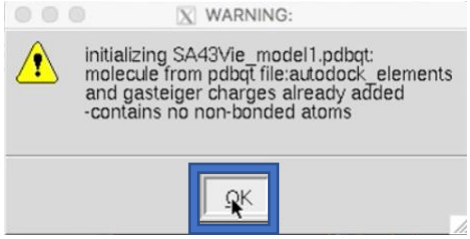


Figure 26. Shows the warning when selecting a molecule for docking.

- This is an example of how the structures will look like.

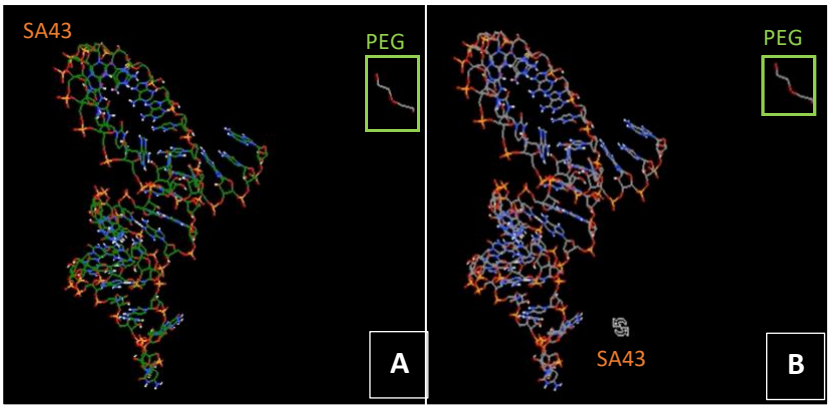


Figure 27. Shows SA43 (A) before being selected for docking. SA43 and PEG are ready for docking and for the grid box step in B.

- Click Grid -> Click Grid Box.

Figure 28. Shows how to obtain the Grid Box.

- A window will pop up with the generated Grid Box details.

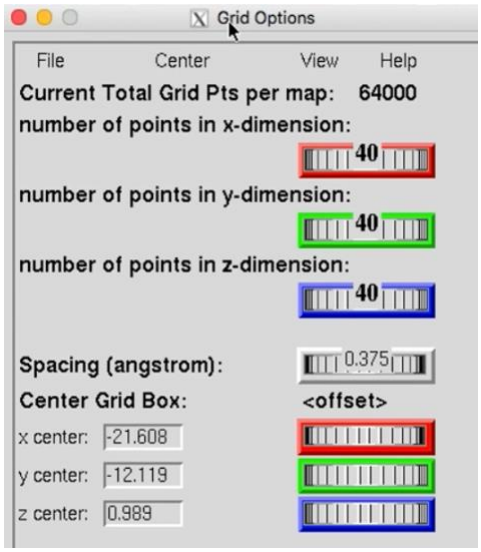


Figure 29. Shows the Grid Options Box with the grid details.

- The Grid Box appears in the aptamer structure.

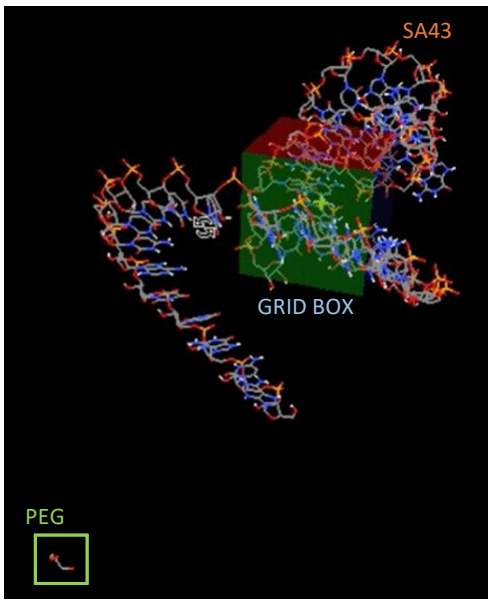


Figure 30. Shows the predicted Grid Box in aptamer SA43 and the ligand PEG. PEG is not yet docked to SA43.

- Click File -> Output grid dimensions file to obtain the x, y and z centre in a txt file.

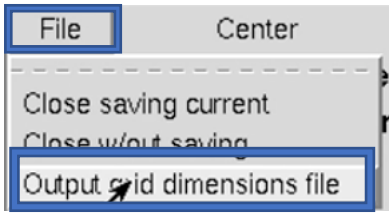


Figure 31. Shows how to obtain the Grid txt file.

- The file will contain the needed details needed for the docking. It gives you a summary of the Grid box dimension. The npts value represent the size of the square for x,y and z. The center represents x=-21.608, y=-12.119 and z=0.989.

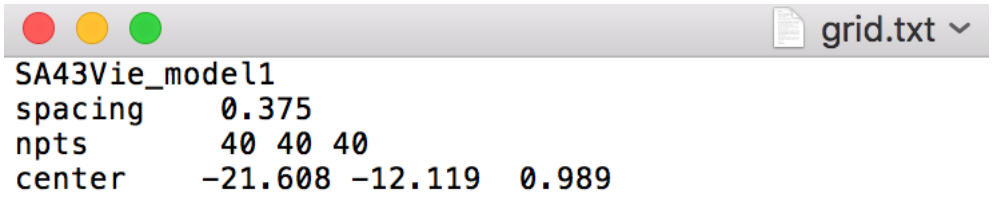


Figure 32. Shows the obtained Grid Box information saved in a txt file.

- Create a txt file named config.txt to facilitate the docking process and shorten the amount of commands that have to be typed. The energy range shows the

value of highest and lowest predicted docking score. By, default it is considered to be 4.

The exhaustiveness lets you decide how many times the docking is going to take place. By default, 8 times is set until shows final pose.

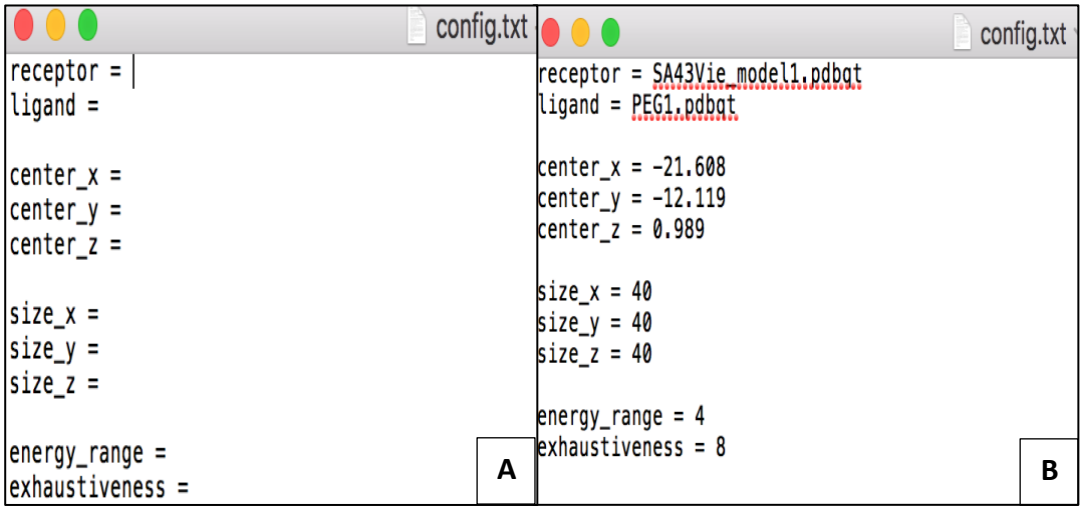


Figure 33. A shows the needed description for the configuration txt file. B shows added grid txt details.

4. Ligand and Aptamer Docking

Vina for docking is not a software. It runs on commands.

- Open the available and already installed Command application on the used computer. For Mac it is Terminal.

- To check if Vina is ready for docking type these commands. Press Enter after each command. The command 'cd' decided which folder you are accessing. As can be seen a summary will be given if downloaded correctly.

```
pc-195-214:~ kristy$ cd Desktop
pc-195-214:Desktop kristy$ cd VINA
pc-195-214:VINA kristy$ cd bin
pc-195-214:bin kristy$ ./VINA
```

Figure 34. Shows the command typed into Vina.

- This information will appear when Vina was properly installed.

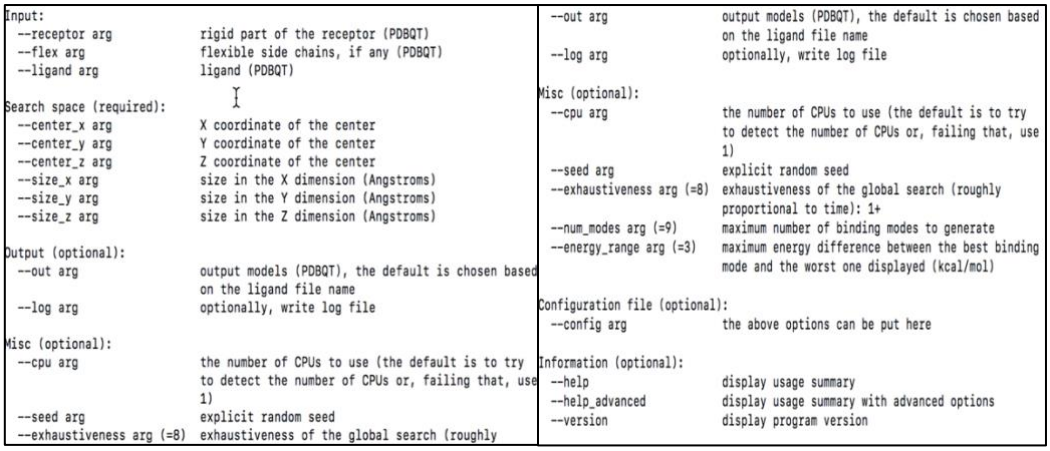
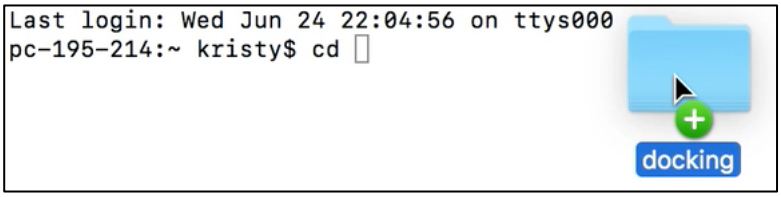


Figure 35. Shows the information that appear when vina was installed properly.

- Type cd and drag the folder with the obtained files into the Terminal



application.

Figure. Shows how the folder is dragged into Terminal.

- Type command ./vina --config config.txt --out output.pdbqt. The output PDBQT file contains the ligand with the x,y and z coordinates obtained through the grid box preparation in Step 3.

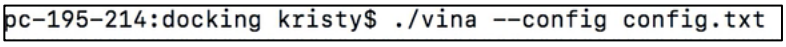


Figure 36. Shows the command typed into Vina.

- Obtain the docking poses and the PDBQT file, which contains the ligand with the coordinates of the Grid box.

```
WARNING: The search space volume > 27000 Angstrom^3 (See FAQ)
Output will be PEG1_out.pdbqt
Detected 4 CPUs
Reading input ... done.
Setting up the scoring function ... done.
Analyzing the binding site ... done.
Using random seed: -1825260560
Performing search ...
0%    10    20    30    40    50    60    70    80    90    100%
|----|----|----|----|----|----|----|----|----|----|
**************************************************
done.
Refining results ... done.

mode |    affinity | dist from best mode
     | (kcal/mol) | rmsd l.b.| rmsd u.b.
-----+------------+----------+----------
   1          -3.8      0.000      0.000
   2          -3.7      0.300      4.383
   3          -3.6      1.943      2.494
   4          -3.6      2.488      5.257
   5          -3.5     10.944     12.270
   6          -3.4     11.042     11.899
   7          -3.4      2.258      4.783
   8          -3.4      2.964      3.646
   9          -3.3      3.475      4.857
Writing output ... done.
```

Figure 37. Shows the result of the output command. The obtained poses are saved as PDBQT file to put into PyMOL as ligand poses.

5. Docking visualisation

- Open PyMOL and drag the aptamer PDBQT file and the output PDBQT file.

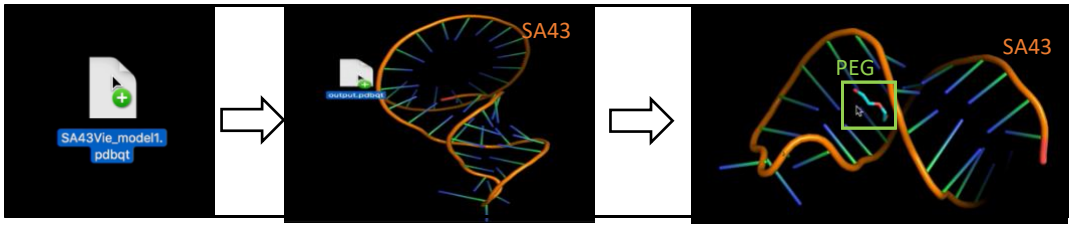
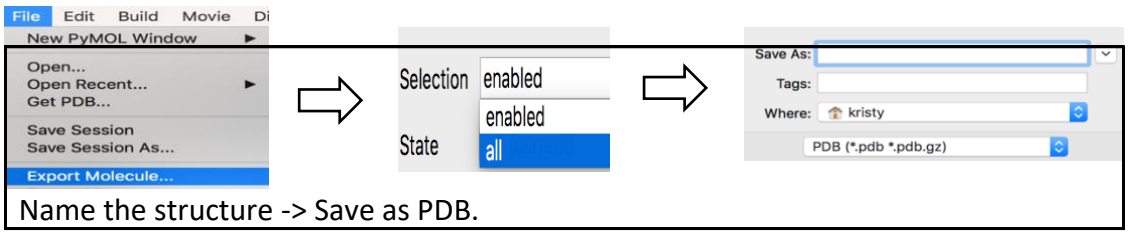


Figure 38. Shows the steps on how to combine the aptamer with the docked ligand.

- Save the docked structure as a PDB file. Click Export Molecule -> Selection (all) ->



- Name the structure -> Save as PDB.

Figure 39. Shows the steps to needed to save the docked structure as a PDB file.

Appendix G: Where does KU Heterodimer bind in an aptamer SA43?

- Download the KU Heterodimer PDB file from the RCSB PDB website
  Link: https://www.rcsb.org/structure/1JEQ

- Use YASARA to see the 3D structure of the KU Heterodimer.

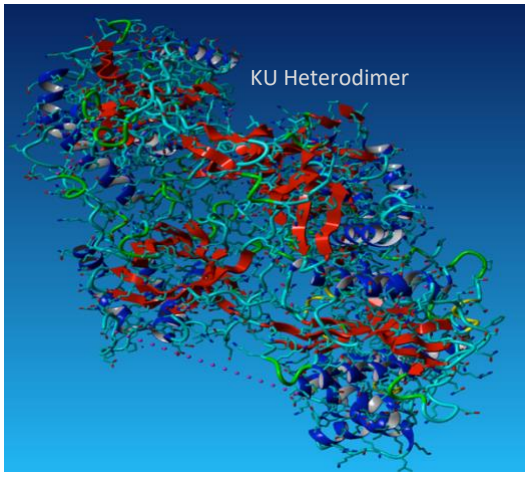


Figure 1. Shows the 3D structure of the KU heterodimer including its molecule KU70 and KU80 visualised using YASARA.

- Follow Docking Steps for the Aptamer, Ligand and Grid preparation from Appendix F.
- Obtain the prepared aptamer and ligand with the predicted Grid box.

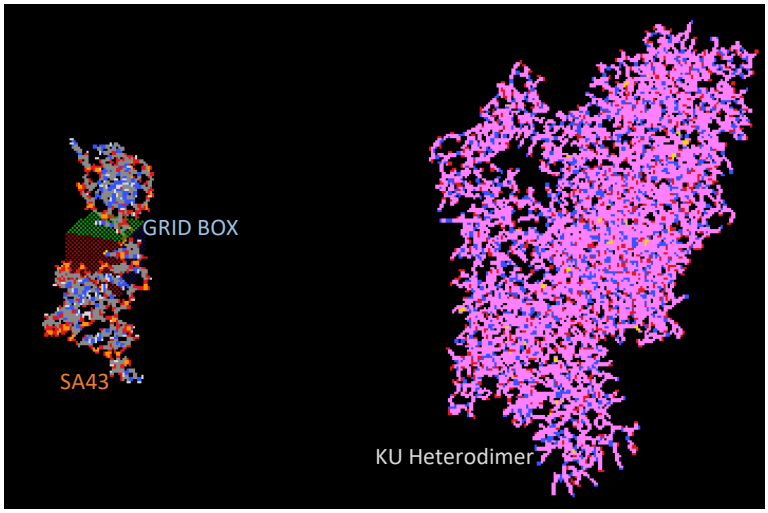


Figure 2. Shows the prepared Aptamer (left) and the KU Heterodimer (right). The Grid box was added shows the binding occurs.

- A Problem that occurred is that the prepared KU Heterodimer cannot be saved as a PDBQT, which is needed for the docking step using Vina and the Command application. The KU Heterodimer PDB. file cannot be converted into a PDBQT file. The given Error code is that no root can be found (Fig). By selecting Torsion Tree is says that no root less than 1 chain is allowed.

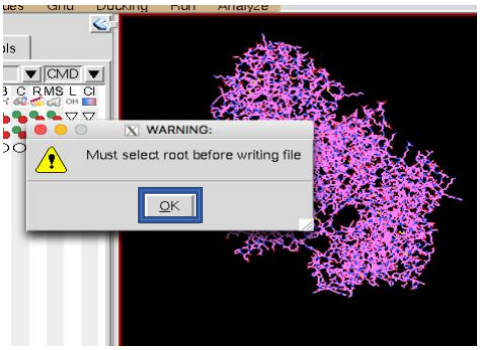


Figure 3. Shows that the PDB file cannot be converted into a PDBQT file.

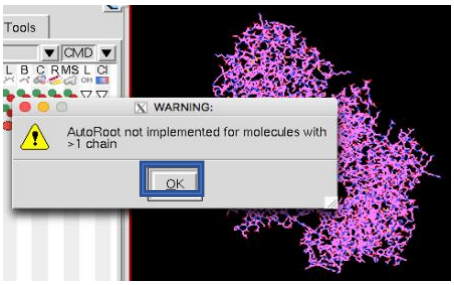


Figure 4. Shows the error/warning that no root was found.

The docking of the KU Heterodimer and the SA43 was not possible due to the file not being able to be converted into a PDBQT file. But looking at Figure 2 the Grid Box is approximately located where the PEG would bind to it.

Appendix H: How to rebuild an aptamer using PyMOL

The software PyMOL has the option to rebuild a structure. The limitation is that the overall structure such as the stem-loop region itself, do not change.

- Open PyMOL
- Upload your aptamer (PDB file)
- Click Builder

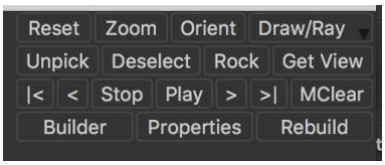


Figure 1. Shows the ‘Builder’ option.

- A window will pop up -> Press Yes.

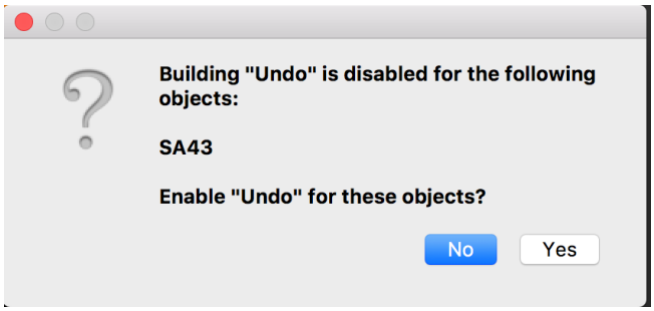


Figure 2. Shows the pop-up window option.

- A window will pop-up, which shows the options how the structure can be edited.

  Each Chemical, Protein and Nucleoic Acid can be used to replace a location.

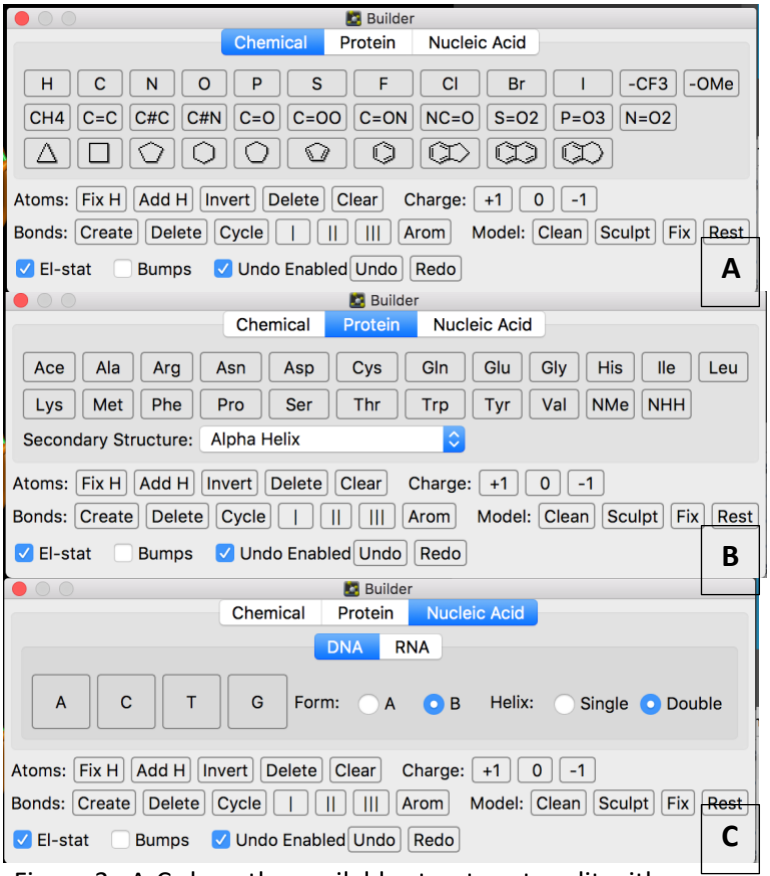


Figure 3. A-C show the available structure to edit with.

- The structure can be viewed with different labelling.

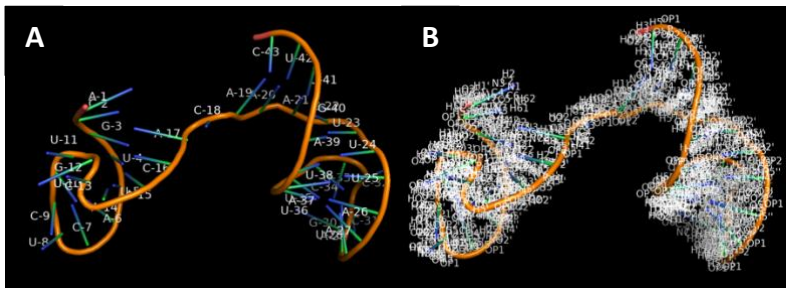


Figure 4. A shows the structure with the nucleotide labelling. B shows the structure with the atom names.

- To see if the structure changes when adding a chemical compound.

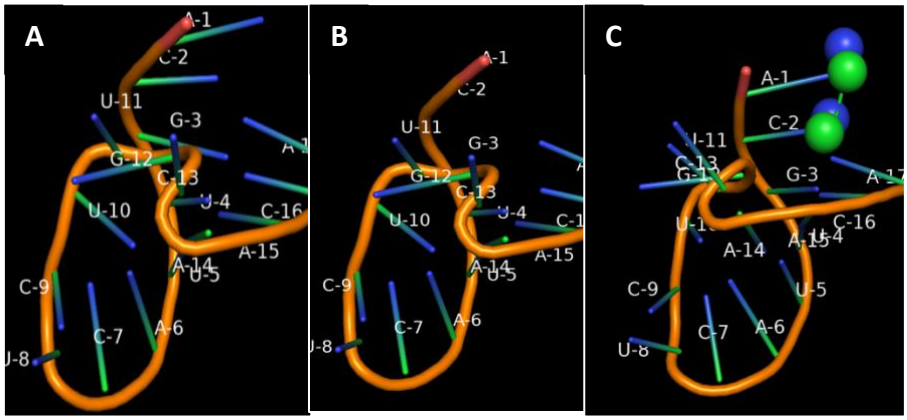


Figure 4. A is the chosen 'floppy end' at the 5'-sequence of SA43. Shows the structure when a chemical (B), Protein (C) is added to the floppy ends which are the unpaired bases.

B in Fig 4. Various chemical compounds were used to replace C3 in the bases A1, C2 and G3, but the only change in the structure was that the green blue 'sticks' disappeared. In C Tyrosine was used to stabiles the structure.

Appendix I: Where does the built structure bind to in the aptamer?

The results for each aptamer with a stabilising agent (thiol) was provided. Thiol was used as a stabilising agent as it reinforces oligonucleotides. Appendix E shows the guideline on how to add a stabilising agent to the 3D structure of the aptamer. It was automatically added to the first nucleotide in the sequence, without the option to change its location.

Figure 1. The summary of the location of the added structure thiol in aptamer SA43 (A), S1p (B), EGFR8 (C), XApt004 (D), TDM1 (E) and hTMA10 (F) in PyMOL. Each thiol is indicated with a green circle together with the nucleotide A1 or C1 to which it is bound to.

Appendix J: Where is the predicted binding site in the aptamer?

After the docking steps in Appendix F it can be established where the ligand's binding site Is located (Fig.1). The final 3D aptamer-ligand docking structure was visualised in PyMOL to add nucleotide labels. The ligand was located and analysed which stem-loop region it bound to. Appendix B was used to count the nucleotide sequence number and to determine if the ligand was in the stem-loop region. Mfold provided details of the aptamer's structural element and information on their exact location (Fig.2).

- Open PyMOL.
- Upload the PDB file containing the docked structure of the aptamer and ligand.
- Label the structure by adding the residues = nucleotide (Click L -> residues)
- Analyse structure and determine where the ligand is located.

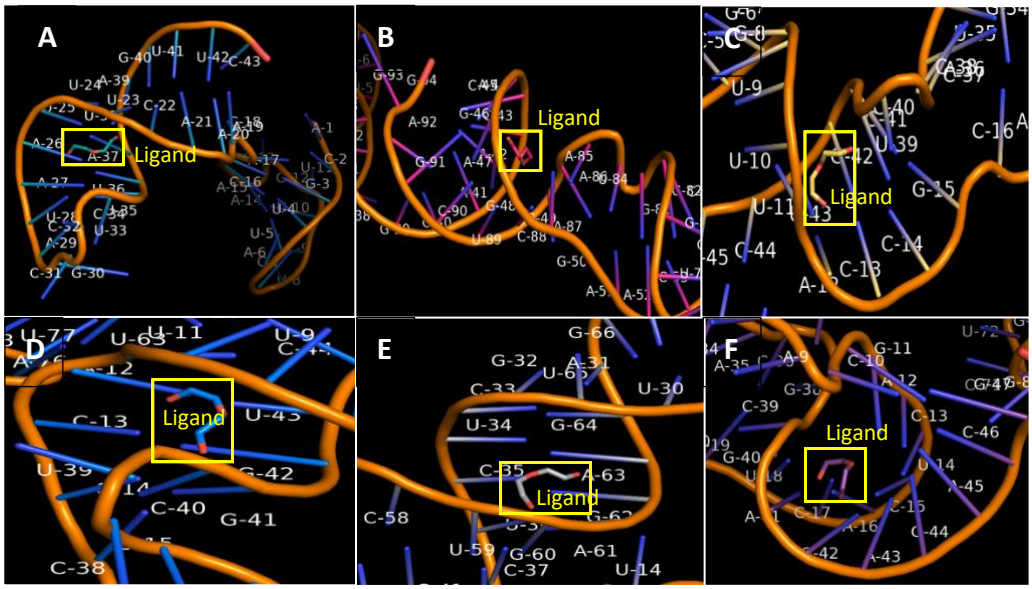


Figure 1. The summary of the predicted binding sites of aptamer SA43 (A), S1p (B), EGFR8 (C), XApt004 (D), TDM1 (E) and hTMA10 (F). in PyMOL. Each ligand is indicated with a yellow square for each aptamer (A-F).

Structural element	δG	Information
External loop	-1.81	6 ss bases & 2 closing helices.
Stack	-1.00	External closing pair is A_{20}-T_{42}
Stack	-1.44	External closing pair is A_{21}-T_{41}
Stack	-1.28	External closing pair is C_{22}-G_{40}
Helix	-3.72	4 base pairs.
Interior loop	1.37	External closing pair is T_{23}-A_{39}
Stack	-0.58	External closing pair is T_{25}-A_{37}
Stack	-1.00	External closing pair is A_{26}-T_{36}
Helix	-1.58	3 base pairs.
Hairpin loop	4.00	Closing pair is A_{27}-T_{35}
Stack	-1.44	External closing pair is G_3-C_{16}
Stack	-1.00	External closing pair is T_4-A_{15}
Helix	-2.44	3 base pairs.
Hairpin loop	3.90	Closing pair is T_5-A_{14}

A

Structural element	δG	Information
External loop	-3.76	29 ss bases & 2 closing helices.
Stack	-2.17	External closing pair is C_{45}-G_{91}
Stack	-1.30	External closing pair is G_{46}-C_{90}
Stack	-1.28	External closing pair is A_{47}-T_{89}
Helix	-4.75	4 base pairs.
Interior loop	3.00	External closing pair is G_{48}-C_{88}
Stack	-2.17	External closing pair is C_{53}-G_{82}
Stack	-2.24	External closing pair is G_{54}-C_{81}
Stack	-1.45	External closing pair is C_{55}-G_{80}
Helix	-5.86	4 base pairs.
Hairpin loop	5.00	Closing pair is A_{56}-T_{79}
Stack	-1.28	External closing pair is C_{13}-G_{30}
Stack	-1.30	External closing pair is T_{14}-A_{29}
Helix	-2.58	3 base pairs.
Hairpin loop	3.60	Closing pair is C_{15}-G_{28}

B

Structural element	δG	Information
External loop	-2.40	21 ss bases & 4 closing helices.
Stack	-1.84	External closing pair is C_{84}-G_{104}
Stack	-1.28	External closing pair is C_{85}-G_{103}
Stack	-0.58	External closing pair is T_{86}-A_{102}
Helix	-3.70	4 base pairs.
Hairpin loop	4.20	Closing pair is A_{87}-T_{101}
Stack	-1.00	External closing pair is T_{63}-A_{76}
Stack	-1.30	External closing pair is T_{64}-A_{75}
Stack	-2.17	External closing pair is C_{65}-G_{74}
Helix	-4.47	4 base pairs.
Hairpin loop	3.10	Closing pair is G_{66}-C_{73}
Stack	-2.17	External closing pair is C_{44}-G_{61}
Stack	-1.84	External closing pair is G_{45}-C_{60}
Helix	-4.01	3 base pairs.
Interior loop	1.70	External closing pair is G_{46}-C_{59}
Stack	-1.45	External closing pair is C_{49}-G_{56}
Helix	-1.45	2 base pairs.
Hairpin loop	3.20	Closing pair is A_{50}-T_{55}
Stack	-1.44	External closing pair is A_{12}-T_{43}
Stack	-1.84	External closing pair is C_{13}-G_{42}
Stack	-2.17	External closing pair is C_{14}-G_{41}
Helix	-5.45	4 base pairs.
Bulge loop	3.30	External closing pair is G_{15}-C_{40}
Stack	-1.45	External closing pair is C_{16}-G_{34}
Stack	-1.28	External closing pair is A_{17}-T_{33}
Stack	-1.30	External closing pair is G_{18}-C_{32}
Stack	-1.00	External closing pair is A_{19}-T_{31}
Helix	-5.03	5 base pairs.
Hairpin loop	3.90	Closing pair is A_{20}-T_{30}

C

Structural element	δG	Information
External loop	-2.00	21 ss bases & 2 closing helices.
Stack	-1.84	External closing pair is C_{84}-G_{104}
Stack	-1.28	External closing pair is C_{85}-G_{103}
Stack	-0.58	External closing pair is T_{86}-A_{102}
Helix	-3.70	4 base pairs.
Hairpin loop	4.20	Closing pair is A_{87}-T_{101}
Stack	-1.28	External closing pair is A_{17}-T_{80}
Stack	-1.30	External closing pair is G_{18}-C_{79}
Stack	-1.00	External closing pair is A_{19}-T_{78}
Stack	-0.88	External closing pair is A_{20}-T_{77}
Stack	-1.00	External closing pair is T_{21}-A_{76}
Stack	-1.30	External closing pair is T_{22}-A_{75}
Stack	-1.84	External closing pair is C_{23}-G_{74}
Stack	-1.45	External closing pair is C_{24}-G_{73}
Stack	-1.44	External closing pair is A_{25}-T_{72}
Helix	-11.49	10 base pairs.
Interior loop	4.60	External closing pair is C_{26}-G_{71}
Stack	-2.17	External closing pair is C_{39}-G_{62}
Stack	-1.44	External closing pair is G_{40}-C_{61}
Helix	-3.61	3 base pairs.
Hairpin loop	4.50	Closing pair is T_{41}-A_{60}

D

Structural element	δG	Information
External loop	-3.73	25 ss bases & 3 closing helices.
Stack	-2.17	External closing pair is C_{78}-G_{88}
Stack	0.08	External closing pair is G_{79}-C_{87}
Helix	-2.09	3 base pairs.
Hairpin loop	3.10	Closing pair is G_{80}-T_{86}
Stack	-1.28	External closing pair is C_{33}-G_{64}
Stack	-1.30	External closing pair is T_{34}-A_{63}
Stack	-1.28	External closing pair is C_{35}-G_{62}
Stack	-1.30	External closing pair is T_{36}-A_{61}
Stack	-1.45	External closing pair is C_{37}-G_{60}
Helix	-6.61	6 base pairs.
Bulge loop	2.95	External closing pair is A_{38}-T_{59}
Stack	-1.44	External closing pair is G_{41}-C_{58}
Stack	-0.58	External closing pair is T_{42}-A_{57}
Stack	0.07	External closing pair is A_{43}-T_{56}
Helix	-1.95	4 base pairs.
Hairpin loop	3.80	Closing pair is T_{44}-G_{55}
Stack	-1.84	External closing pair is C_3-G_{25}
Stack	-1.28	External closing pair is C_4-G_{24}
Stack	-0.58	External closing pair is T_5-A_{23}
Helix	-3.70	4 base pairs.
Hairpin loop	4.20	Closing pair is A_6-T_{22}

E

Structural element	δG	Information
External loop	-4.20	28 ss bases & 3 closing helices.
Stack	-2.17	External closing pair is C_{74}-G_{84}
Stack	0.08	External closing pair is G_{75}-C_{83}
Helix	-2.09	3 base pairs.
Hairpin loop	3.10	Closing pair is G_{76}-T_{82}
Stack	-2.17	External closing pair is C_{55}-G_{64}
Helix	-2.17	2 base pairs.
Hairpin loop	3.00	Closing pair is G_{56}-C_{63}
Stack	-2.17	External closing pair is C_{10}-G_{47}
Helix	-2.17	2 base pairs.
Interior loop	0.34	External closing pair is G_{11}-C_{46}
Stack	-1.28	External closing pair is C_{13}-G_{44}
Stack	-1.30	External closing pair is T_{14}-A_{43}
Helix	-2.58	3 base pairs.
Interior loop	3.50	External closing pair is C_{15}-G_{42}
Stack	-1.45	External closing pair is T_{20}-A_{34}
Stack	-2.24	External closing pair is G_{21}-C_{33}
Helix	-3.69	3 base pairs.
Hairpin loop	3.40	Closing pair is C_{22}-G_{32}

F

Figure 2. Shows the structural element for aptamer SA44 (A), S1p (B), EGFR (C), XApt004 (D), TDM1 (E) and hTMA10 (F) obtained through Mfold.

The findings show that the binding site is in the stem-loop region for each aptamer, when comparing Fig.1 and Fig.2. The binding site was located in one of the 'stack' for each aptamer. These 'stack' are the regions that form a stable duplex (Helix) (Aalberts et al., 2002). In SA43 it was in the stack of an interior loop with its external closing area of T_{25}-A_{37}. For S1p it was the stack of an external with its external closing pair between C_{45}-G_{91}. For EGFR8 the binding site was in the stack of a hairpin loop with its external closing pair A_{12}-T_{43}. In XApt004 the binding site is located in an interior loop at its second stack with the external closing pair G_{40}-C_{61}. In TDM1 the binding site was in one of the stacks of a hairpin loop with the external closing pair T_{36}-A_{61}. The binding site of hTMA10 was in an interior loop in one of its stack with the external closing pair C_{13}-G_{44}.

Name of Student: Kristy Montalbo

REPORT A: Student Self-Assessment Report

Provide comments below. Sections can be expanded as required.

Comment on progress made over this assessment period. Key points to comment on include:

- ✓ **Summarise the work you have undertaken during the project, including your objectives, achievements and plans (recommended word limit 500 words).**

A literature Research was conducted to establish that a focus on using bioinformatical methods for aptamers is needed especially in the 3D (three dimensional) structure visualization of aptamers. Then a protocol was created on which software and webserver to use combined with the collection and analysis of figures containing the original aptamer structure and its modified versions. The aim of this Research Project is to obtain the 3D structure of aptamers and stabiles its structure by shortening as well as by adding stabilizing agents at the 5'-3'end of the sequence. Furthermore, the binding site of each aptamer was determined to see if stabilizing agents will interfere with the stem-loop region, which are predicted to be the main binding site of aptamers. Bioinformatical methods such as using webservers and software were used to obtain the 3D structure of six given aptamers. Firstly, two webservers Mfold and Vienna RNAfold were compared to predict the 2D (two dimensional) structure. The dot-bracket notation is an easy way to represent the secondary structure by using dots for unpaired bases and matching pairs with parenthesis (). It was established that Mfold is more reliable as it focuses on DNA folding and the obtained 2D structures matched the ones given by the supervisor Dr. Lisa Shaw provided by a former Master student. After obtaining the 2D structure with its corresponding dot-bracket the notation the 3D structure was obtained by using the webserver RNAComposer. RNAComposer provided the PDB file of the structure as well as a simple 3D preview of the aptamer. The software YASARA, PyMOL and Chimera were used to upload the PDB (Program Database) file containing the 3D structure for visualization. To obtain a shortened version of each aptamer the 2D structure obtained through Mfold were analysed. The unpaired bases at the 5'-3'-end of the sequence were removed, and a new and shortened aptamer sequences was designed. For each new aptamer sequence a new dot-bracket notation was created and pasted into the webserver RNAComposer together with its matching sequence. The PDB file containing the new shortened aptamer was then uploaded onto the software YASARA, PyMOL and Chimera to obtain a detailed 3D structure. To see if bioinformatical methods can be used to stabilize the original aptamer structure, Chimera and PyMOL were used to add and build the structure of a stabilizing agents such as chemical compounds e.g. thiol. The obtained structure did not change the original structure. Therefore, establishing that there was a limitation in free and easily available software when editing an aptamer sequence. To predict the binding site, the software AutoDockTools was used for the aptamer, ligand and grid preparation for the docking procedure. Vina and the command application Terminal were used to perform the final docking by using commands. PyMOL was used to convert the ligand file from a SDF file into a PDB as well as to visualise the final structure with the docked ligand. The binding was predicted for each aptamer and it was established that the stem-loop region is the main target site.

- ✓ **Have you met the research objectives agreed with your supervisory team the start of your studies?**

Yes, the research objectives were met. The main research objective was to provide a step-by step guide on how to visualize the 3D structures of the six given aptamers (SA43, S1p, EGFR8, XApt004, TDM1 and hTMA10), which was done. Additionally, the dot-bracket notation 2D structures were also analysed and obtained by using computational methods. Furthermore, the obtained original structures were modified by either shortening or adding stabilizing agents to see if there is a change in structure. The shortened 2D and 3D structures were obtained from various webservers and software and visualized in a collection of figures for the six aptamers. The addition of stabilizing agents was performed, but due to the lack of editing options in the software YASARA, PyMOL and Chimera no change in structure was analysed. Additionally, the binding site of each six aptamers were predicted, proving that the main binding site in an aptamer was in the stem-loop region. Additionally, it was suggested to analyse if SA43 and the KU Heterodimer can be docked, due to the limitation given by the used server this was not established.

✓ **Identify your strengths, weaknesses, opportunities, and problems identified during the year**

Prior to the pandemic my strengths were being dedicated to the course by attending all of the scheduled lectures. Another strength of mine is being able to work under difficult and stressful circumstances and being resilient such as working on my assessments, undertaking the final exam and conducting my Research Project during the pandemic. I can adapt to situations very well and work independently such as doing a dry lab in the area of bioinformatic and teaching myself how to use each webserver and software. I was confident and enthusiastic on finding the easiest to use and free to download software as the project time limit was 10 weeks. I am well organised and have a good time management when doing a project. Overall, I am a diligent student who is committed to achieving a career in Cancer Research. I am aware that this career is competitive and has its challenges. However, pursuing a careershows that I am willing to take the necessary steps to achieve this goal.

A weakness of mine is to ask for help as I always want to be able to find solutions to a problem on my own, this was overcome by having supportive and helpful supervisors during my MSc project. Furthermore, I sometimes have the inclination of not being confident in my ability, which was eliminated by getting good progress report feedbacks during the MSc project.

Specialising in the area of Cancer Biology and Therapy, my master's degree has given me great foundational knowledge in Oncology. Furthermore, studying in England gave me the opportunity to study the course I am most passionate about. Several aspects of my postgraduate education have enabled me to gain further knowledge on cancerous diseases, their treatment and research skills, their treatment and research skills. For example, this course required a Research Proposal submission in which I demonstrated my intellectual purpose, originality and awareness of research contexts. Subsequently, this developed my ability to communicate scientific ideas clearly and critically appraise scientific papers and data. My proficiency in technical writing and sourcing were also skills I was able to develop and apply to diverse laboratory techniques and research. Another beneficial aspect of my postgraduate degree was the requirement of writing case studies in which I had to differentiate the cellular and molecular events associated with the development and spread of specific cancer examples. These tasks additionally involved evaluating cancer treatment modalities with an emphasis on novel therapeutics such as targeted therapies and their supporting evidence. Most significantly within these case studies, I showed my ability to critically evaluate the concepts of patient centred therapy.

I was offered the opportunity on creating a protocol for predicting and visualising the 3D structure of aptamers, which will be of use for future papers by my supervisor Dr. Lisa Shaw. I had the opportunity to expand my knowledge in the area of bioinformatics which is a valuable skill during my MSc project.

The limitation with the currently free and easily accessible software was that further modification on the 3D structure required more advanced computers. The problem was that these require a bigger budget and sophisticated technology, which I did not have access to. I wanted to gain more 'hands on' experience during the course. Although I had experienced laboratory work during my MSc it was limited to five lab course days, which did not enable me to gain enough practical lab skills.