

# Development and validation of anti-human Alpha synuclein DNA aptamer using computer modelling techniques—an *in silico* study

Marcus Rock<sup>1</sup>, Georgios D. Zouganelis<sup>1,\*</sup>, Ana Flavia Belchior de Andrade<sup>1</sup>, Samantha J. Drake<sup>1</sup>, Athanasios Alexiou<sup>2,3</sup>, Ashraf Albrakati<sup>4</sup>, Gaber El-Saber Batiha<sup>5</sup>, Thomas Arron Illingworth<sup>1</sup>

<sup>1</sup>Human Sciences Research Centre, University of Derby, DE22 1GB Derby, UK

<sup>2</sup>Department of Science & Engineering, Novel Global Community Educational Foundation, 2770 Hebersham, Australia

<sup>3</sup>AFNP Med Austria, 1010 Wien, Austria

<sup>4</sup>Department of Human Anatomy, College of Medicine, Taif University, P.O. BOX 11099, 21944 Taif, Saudi Arabia

<sup>5</sup>Department of Pharmacology & Therapeutics, Faculty of Veterinary Medicine, Damanshour University, 22511 Damanshour, AlBeheira, Egypt

\*Correspondence: [g.zouganelis@derby.ac.uk](mailto:g.zouganelis@derby.ac.uk) (Georgios D. Zouganelis)

DOI: [10.31083/j.jin2101005](https://doi.org/10.31083/j.jin2101005)

This is an open access article under the CC BY 4.0 license (<https://creativecommons.org/licenses/by/4.0/>).

Submitted: 1 August 2021 Revised: 26 August 2021 Accepted: 18 September 2021 Published: 21 January 2022

**Background:** Biomarker detection strategies have, in recent years, been moving towards nucleic acid-based detection systems in the form of aptamers, short oligonucleotide sequences which have shown promise in pre-clinical and research settings. One such aptamer is M5-15, a DNA aptamer raised against human alpha synuclein ( $\alpha$ -syn) the causative agent in Lewy body and Parkinson's disease (PD) associated dementia. While this aptamer has shown promise, *in silico* methodologies have demonstrated a capacity to produce aptamers that have higher affinities for their targets than in vitro generated sequences. **Methods:** A Python script random generated library of DNA sequences were screened based on their thermodynamic stability with the use of DINAMelt server-QuickFold web server. The selected sequences were examined with MFold in order to generate secondary structure data that were used to produce 3D data with the use of RNA composer software. Further on, the structure was corrected and RNA was replaced with DNA and the virtual screening for  $\alpha$ -syn aptamer took place with a series of molecular docking experiments with the use of CSD-Discovery-GOLD software. **Results:** Herein we propose an alternative *in silico* generated aptamer we call TMG-79 which demonstrates greater affinity for the target compared to M5-15 (M5-15 =  $-15.9$  kcal/mol, TMG-79 =  $-17.77$  kcal/mol) as well as better ChemPLP fitness scoring between the top poses (M5-15 = 32.33, TMG-79 = 53.32). Structural analysis suggests that while there are similarities, the greater potential flexibility of TMG-79 could be promoting greater affinity for the  $\alpha$ -syn compared to M5-15. **Conclusions:** *In silico* methods of aptamer generation has the potential to revolutionise the field of aptamer design. We feel that further development of TMG-79 and validation in vitro will make it a viable candidate for diagnostic and research use in the future.

## Keywords

DNA aptamers; *In silico* design; Computer aided design; Alpha synuclein; Biomarker detection

## 1. Introduction

Biomarker detection has become a critical component in the diagnosis of disease and has emerged as a critical factor in deciding courses of treatment of patients. Over 30 years ago, the first monoclonal antibody was developed for biomarker detection with the first licensed products being deployed in 1986 [1]. Global market value for monoclonal antibodies surpassed \$95 billion in 2017, with an expectation that it will reach \$131 billion by 2023 [2]. High specificity and utility of monoclonal antibodies in both research and clinical settings fuels the demand for these detection strategies [3]. While monoclonal antibodies are still considered the gold standard in both clinical and research settings, being involved in a number of key molecular biology techniques like western blotting and immunofluorescence microscopy [4, 5], there are a number of drawbacks in relation to the technology. Monoclonal antibodies are exceptionally expensive to produce and purchase depending on the target protein, ranging between €53/g and €84/g depending on the production method chosen [6]. Reliance on animal models to produce these antibodies, which has been reduced to an extent due to the widespread adoption of hybridoma production strategies, also introduces problems. Species specific glycosylation of the antibodies can lead to batch-to-batch variations in antibody affinity [7] and the variable fragment crystallisable (Fc) domain derived from the host species can cause immunogenicity problems in clinical settings [8]. Despite these issues, monoclonal antibodies have seen widespread adoption, but in recent years industry and individual researchers have been looking at potential alternative strategies for biomarker detection. Aptamers have emerged as a potential candidate to replace monoclonal antibodies.

### 1.1 Aptamers

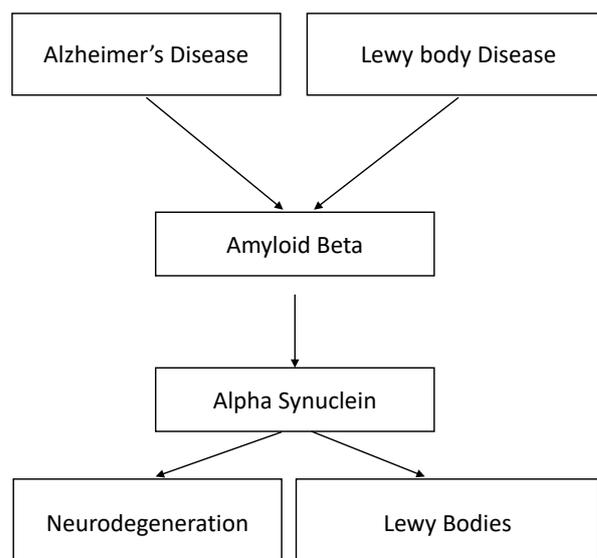
Aptamers are defined as synthetic ligands composed of single stranded RNA or DNA which have unique structures that can be designed to target specific biomolecules with high affinity [9]. The production of aptamers is achieved through a process of systematic evolution of ligands by exponential enrichment (SELEX) and was discovered by three groups independently at the beginning of the 1990s, beginning with RNA aptamers then expanding into DNA and peptide-based aptamers (referred to as affimers) developed today [10–12]. Nucleic acid-based aptamers are single stranded nucleic acid structures which can form secondary structures by folding in on themselves into thermodynamically stable shapes [13]. These structures consist of double helices, and hair-pin structures of varying sizes governed by the number of mis-match pairing of bases which produce internal loops and three-way helical regions [14]. While SELEX has been used previously for the development of aptamers, there have been movements towards the use of *in silico* computer-based modelling to streamline the production and design of aptamers [15]. Techniques like M.A.W.S (making aptamers without SELEX) are the most notable recent examples of *in silico* aptamer development and has been seen as a gold standard for the development of aptamers *in silico* [16]. The focus of aptamer generation *in silico* looks to remove some of the initial high costs associated with aptamer selection, as well as looking to produce high affinity aptamers prior to the purchase of reagents needed for *in vitro* validation of aptamer binding [17].

One of the main advantages of aptamers over their monoclonal antibody counterparts is tissue penetration, as well as a capability to be raised against targets that do not elicit an immune response, such as charged ions, intracellular proteins and vitamins [18]. The use of aptamers has emerged as a potential detection strategy for a wide number of biomarkers, with particular interest in the detection of markers for neurological disease [19]. In particular, alpha synuclein ( $\alpha$ -syn) has emerged as a candidate in the diagnosis of neurological diseases such as Parkinson's (PD) and Alzheimer's (AD).

### 1.2 Alpha synuclein ( $\alpha$ -syn)

$\alpha$ -syn is a 140 amino acid protein encoded by the *SNCA* gene located on chromosome 4 and has been implicated in the progression of Lewy body, Alzheimer's and Parkinson's disease associated dementia [20]. Under physiological conditions,  $\alpha$ -syn is maintained in an unfolded state but in disease states it is capable of forming toxic oligomers and aggregates with local proteins and has been linked to toxicity via mitochondrial damage, inflammation and axonal transport dysfunction (Fig. 1, Ref. [21]) [22].

The link between AD and  $\alpha$ -syn has been established, however the link with PD has not been fully confirmed but has the potential to be used to establish differences in diagnosis with lower cerebrospinal fluid (CSF) levels being detected in PD compared to AD [23]. Efforts to prevent the aggregation of  $\alpha$ -syn, and by extension alleviate symptoms of AD,



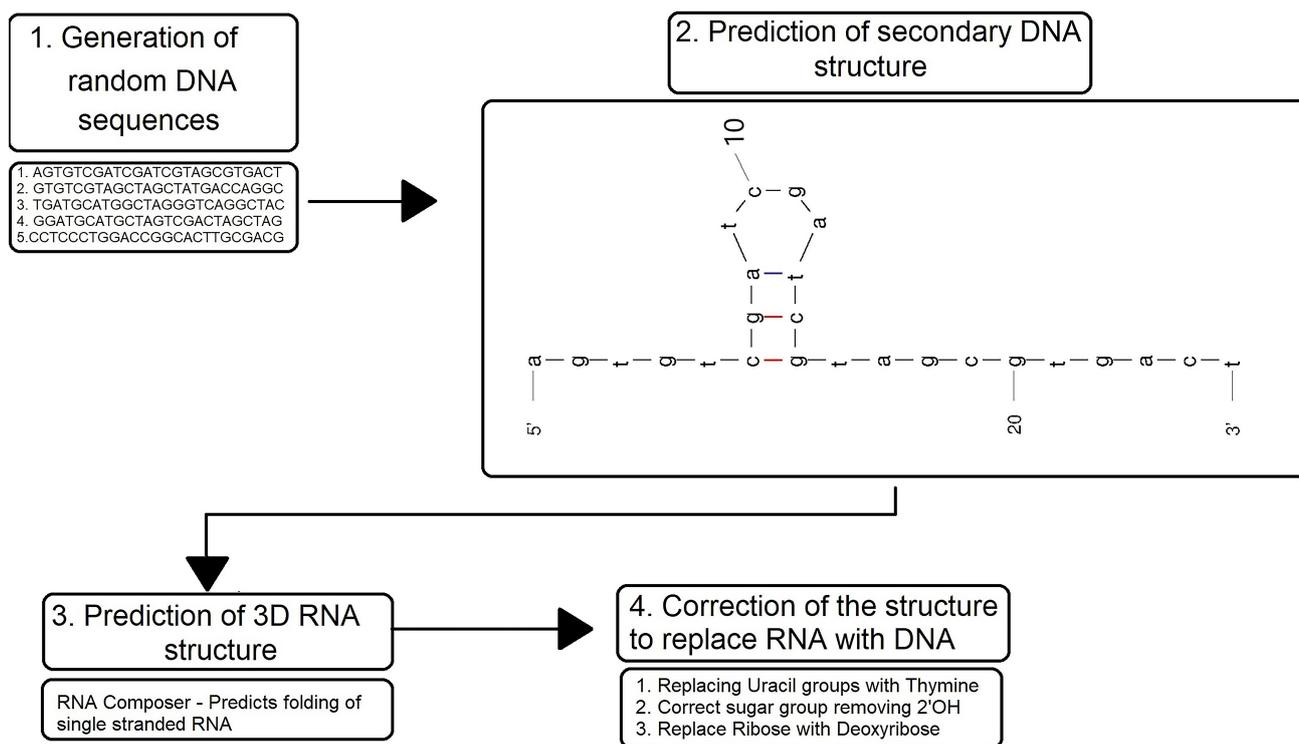
**Fig. 1. Summary of the interactions in AD.** Notably  $\alpha$ -syn is associated directly with synapse loss and neurodegeneration in AD, however this is separate from the formation of amyloid plaques which are typically associated with the condition as described in [21].

have focused on the use of aptamers due to their inherent ability to penetrate the blood brain barrier. Several aptamers have been identified which have successfully targeted  $\alpha$ -syn and have demonstrated mitochondrial protective effects as well promoting  $\alpha$ -syn degradation *in vitro* [22, 24, 25]. The approaches taken when producing anti  $\alpha$ -syn aptamers has been the traditional SELEX process, with minor modifications that have been incorporated over time to improve the methodology. With the advances in computer simulations and molecular docking experimentation, interest in improving the experimentally derived aptamers has increased exponentially.

### 1.3 In silico design

The concept of computer aided drug design, and by extension *in silico* aptamer design, has been gaining traction in recent years, specifically due to its ability to streamline the design of innovative treatment and detection strategies for a wide variety of applications. Advances in computer algorithms capable of predicting the folding characteristics of single stranded DNA, as well as increasing availability of these software, have been critical to the development of *in silico* aptamer methods [26]. The standardised approach to *in silico* aptamer design is summarised in Fig. 2 (Ref. [27]).

Sequence generation can be upscaled to be comparable with *in vitro* SELEX, allowing for the screening of an exceptionally large number of sequences. The difficulties come when establishing selection criteria for the computer-generated sequences, and which ones to proceed with in a docking experiment as the large number of sequences make it impractical to dock every aptamer produced. A common approach is to look at the free energy potential of the aptamers



**Fig. 2. Summary of steps within *in silico* aptamer design.** RNA folding algorithms have proven to be effective at predicting DNA aptamer structure despite minor chemical differences. It is common practice to correct these structures in post. Image based on information contained within [27].

secondary structure [26] and secondary structural motifs [28] to screen potential aptamer candidates. The second option is particularly useful when looking at aptamer optimization studies, as the original sequence can be mutated at specific points to see what impact this will have on overall binding.

Following sequence generation, aptamers are then docked to the target molecule in question. There are a number of programmes which are available for use in this field, some which are freely available and some requiring a subscription [28]. Irrespective of the programme used, the objective is to create a pose (two molecules which have formed a complex) with the lowest Gibbs free energy ( $\Delta G$ ) binding energy [27]. Experiments themselves can take two main approaches, fixed or flexible docking which can be further sub-divided into 5 categories. The approaches are summarised in Table 1 [29].

*In silico* docking is not without its limitations however, despite the advances in computer modelling and DNA/RNA folding algorithms. Issues with computer processing load [27], particularly with more flexible approaches to docking, as well as the difficulties in accurately mimicking the electrochemical environment these molecules would experience *in vivo* [30] do mean that at this stage, predictions need to be validated *in vitro* before definitive conclusions can be made. However, this does not detract from the usefulness of *in silico* methods for they have the potential to both streamline and improve the aptamer generation process.

#### 1.4 Aims

In this study, an experimentally produced aptamer was docked alongside an *in silico* developed aptamer against human  $\alpha$ -syn. The intention of this study was to demonstrate a novel method by which aptamers can be developed *in silico* and validated against previously observed interactions, with the hope of refining that aptamer to improve the binding efficiency to its target. To our knowledge, this is the first investigation looking into the refinement of an anti-human  $\alpha$ -syn DNA aptamer to produce a structure with a predicted higher binding affinity to the experimentally derived sequence.

## 2. Methods

### 2.1 Generation and selection of sequences and structures

Aptamer sequences were generated using Python script to create a library of 1000 sequences with an N30 wobble region flanked by constant primer binding regions (ATAGTCCCATCATTCATT-N30-AGATATTAGCAAGTGTCA). Sequences were retained following prediction of thermodynamic stability using the DINAMelt server-QuickFold web server [31]. Retained sequences were required to have negative  $\Delta G$  values, using the parameters listed in Table 2.

Secondary structural information for 50 structures, selected based on their  $\Delta G$ , was obtained using MFold [32]. Structure prediction in MFold used the same parameters as were used within the DINAMelt server-QuickFold.

**Table 1. Summary of docking common docking approaches *in silico* design. Information in the table is based on [29].**

Docking approach	Key characteristic
Rigid docking - fixed receptor and ligand	Both receptor and ligand bond angles and atomic positioning are fixed. This mimics the traditional “lock and key” theory for ligand/receptor interactions. Is the least resource intensive out of the approaches, but does not take into consideration structure flexibility. It also limits understanding of the solvents impact on docking
“Soft docking” - rigid receptor, semi-flexible ligand	Small degree of flexibility within a small area for the ligand, the receptor remains fixed
“Side chain flexibility” - rigid receptor, side chain flexibility only	Majority of the structures (both ligand and receptor) are fixed in position, with only the amino acid side chains able to flex
Composite receptor modelling - flexible receptor	Combined structural information breaking up larger receptors into smaller subunits. These subunits are then contained within a single docking “box”. Allows for degree of flexibility in the receptor while minimising computer load
Ensemble docking - Flexible receptor	Generation of a composite receptor structure from multiple components, with each sub-group capable of rotating freely through its structure. Ligand can be either fixed or flexible in these scenarios. Critically in this instance they are not confined to a single “box”, but are made up of multiple components that work independently. This is a processor intensive position and requires time and resources to run effectively

**Table 2. DINAMelt server-QuickFold parameters for 2D sequence generation.**

Folding temperature	25 °C
Ionic conditions	[Na <sup>+</sup> ] = 150 mM [Mg <sup>++</sup> ] = 0 mM
Distance between paired bases	No limit on distance
Sequence type	Linear
Maximum possible foldings	50

## 2.2 Creation of 3D structures using RNAcomposer

Vienna sequence outputs from MFold were used to assist in the production of 3 dimensional models using RNAcomposer [33]. To account for this being a programme which folds RNA, all thymine residues within the sequences were replaced with uracil groups for the purpose of folding. Vienna sequences were input along with the base sequence data, with the top 50 sequences being run in batch format within RNAcomposer. Structure prediction used default parameters for the purposes of folding.

Protein data bank (PDB) output files of the predicted sequences were converted into mol2 format using CSD-Discovery-Mercury [34]. Structures were then modified to reintroduce the deoxyribose groups, replacing the ribose sugar groups within their structures using PyMol [35] via command line editing within the software (Schrödinger, New York, USA). Commands were set to remove O2' and HO2' atoms from all residues and to add a hydrogen to the C2' atom of each residue. Uracil groups were replaced with thymine groups using the mutagenesis wizard also contained within the PyMol software. Files were then exported back into CSD-Discovery-Mercury to add any hydrogens to the structure. In addition, bond angles and types were corrected as some had been incorrectly modelled during the file conversion process. Of the 50 structures that were originally generated, the top 10 were selected to dock with the target molecule. Primary structural similarity to M5-15 was used

as the primary criteria when selecting the aptamer, followed by thermodynamic stability. This is due in part to the nature and composition of free hydrogen linkage points within the short structure determining how the single stranded nucleotide will fold in on itself [36].

## 2.3 Virtual screening of anti-human $\alpha$ -syn aptamer using CSD-Discovery-GOLD

Docking experiments were analysed determining the predicted interactions between the components of the final pose. Bond distance, nature of interactions and type of interaction were all documented and analysed. Overall fitness of poses was determined using ChemPLP fitness scoring within CSD-Discovery. Structures were virtually screened in blocks of 10 using LYS 96, O as a central atom and a radius of 81 Å, encompassing the target protein completely. All rotatable bonds were fixed to create a rigid dock and search efficiency was set to low/virtual screening. Fitness scoring function was set to ChemScore with no re-scoring function selected.

Binding energies were determined using follow up confirmation docking experiments using AutoDock Vina. Docking parameters between the programmes were kept constant. The docking positions generated by CSD-Discovery were used to inform the choice of grid location for docking in AutoDock Vina. The top 3 aptamers selected from CSD-Discovery were docked against  $\alpha$ -syn using the following settings: grid size: 110, 110, 110; coordinates: center\_x = 23.673, center\_y = 56.090, center\_z = -15.985; energy range = 4; exhaustiveness = 8; ligand torsions set to zero during setup. A summary of data files used within the work can be found in Table 3 [37, 38]. A spinach protein aptamer, 4ts2, was chosen as a control as it has no specificity to the target protein. Comparisons with a sequence which binds to a protein which is in an unrelated organism gives an indication of the baseline interaction between the target and a nucleotide sequence.

**Table 3. Record of all PDB files used in study and any modifications made.**

File name	Modifications to file	Reference
1xq8 (Human $\alpha$ -syn)	No modifications made	[37]
4ts2 (Spinach RNA aptamer in complex with DFHBI ((5Z)-5-(3,5-difluoro-4-hydroxybenzylidene)-2,3-dimethyl-3,5-dihydro-4H-imidazol-4-one), magnesium ions)	Metal ions (Magnesium and Potassium) removed Ligands (beta-D-fructofuranose-(2-1)-alpha-D-glucopyranose and Cytidine-5'-Phosphate-2',3'-cyclic phosphate, (5Z)-5-(3,5-difluoro-4-hydroxybenzylidene)-2,3-dimethyl-3,5-dihydro-4H-imidazol-4-one) removed Hydrogens added	[38]

**Table 4. Data displaying the Gibbs free energy of the reference aptamer and the top 10 aptamer sequences based on  $\Delta G$  in ascending order.**

Aptamer	Sequence	$\Delta G$ kJ/mol
M5-15	Forward -GTATGGTACGGCGCGGTGGCGGGTGCCTGG- Reverse	-6.39
120	Forward -CGGGTTAGGTGGAAGATCCCACCTATTCCG- Reverse	-8.52
123	Forward -GTGGGGTCCAGGCCTAGTACTGCTAAGGC- Reverse	-7.70
36	Forward -CATAAGCCCATGCTAGACTTGCAGAAAGGGT- Reverse	-7.16
TMG-79	Forward -ACGTTGTTGCATGCCGAAAGCCGGCGAATGC- Reverse	-6.96
60	Forward -GGGGATACTAACCCGTCCCCGGCGAGCGG- Reverse	-6.27
54	Forward -ATAAGGGCCCGTGCCGTCTACGGTTCCTCG- Reverse	-6.22
195	Forward -CCTTTCGGGGGCGCCCCCTTCCGGTAT- Reverse	-5.96
116	Forward -TAACTTTTAAAATCTGTAGAGCAGGTGCGA- Reverse	-5.79
125	Forward -ACGTTTGGGAAATAGTATCCCCCAACAT- Reverse	-5.28
30	Forward -CTGAAAGGGGCTGTGCCGCCGTAATAATA- Reverse	-4.83

**Table 5. Breakdown of GOLD docking data including H-bond contribution, number of H-bonds and the average bond length of these polar interactions.**

Aptamer	PLP. Chemscore. Hbond (3.dp)	Number of H bonds	Average bond length (3.dp)	Top pose binding energy (kcal/mol)	Mean Binding energy (kcal/mol) (2.dp)	Mean dist from rmsd l.b.
Reference	0.008	2	2.948	-16.8	-15.90	27.312
4ts2	0.000	0	-	-17.6	-16.22	21.583
79	2.000	2	2.884	-18.7	-17.77	13.801
125	1.089	3	2.944	-17.0	-16.40	16.830
123	4.976	5	2.734	-18.4	-16.48	16.314

rmsd, root mean squared distance.

#### 2.4 Data analysis/structure analysis

Top poses from CSD GOLD were used to determine the best aptamers to move forward to docking in Autodock Vina based on a combination of Fitness score, number of hydrogens and average bond length of interactions. Analysis of variance was conducted on the data produced by Autodock Vina to determine any significant changes in binding energies in comparison to M5-15.

### 3. Results

#### 3.1 *In silico* 2D aptamer structure prediction

The *in silico* selection of ssDNA aptamers against  $\alpha$ -syn using simulated SELEX produced numerous potential candidates which were shown to be comparable to the reference aptamer made by [37] within the limitations of the model. Virtual screening of the top 50 aptamers based on  $\Delta G$  produced a pool of 10 aptamers which are shown to bind to  $\alpha$ -syn (see Table 4). The pool of candidates displays a large variation in  $\Delta G$  due to differences in GC content and the presences of

Watson-Crick base pairing.

The known sequence which binds to  $\alpha$ -syn was shown to form three hairpin/stem loop structures when folded at 25 °C under the ionic conditions [37]. Many of the *in silico* sequences selected based on  $\Delta G$  have shown the formation of similar binding motifs which may be effective at targeting  $\alpha$ -syn due to structural homology (see Fig. 2). Lower  $\Delta G$  values were also shown to be heavily correlated to the formation of Watson-Crick base pairing as opposed to higher  $\Delta G$  values which are the result of linear regions within secondary structures.

Pairwise alignment (Fig. 3) was conducted to determine the sequence homology between each candidate and M5-15. When aligned, TMG-79 was shown to have the largest percentage similarity (51.4%), suggesting a degree of similarity between the structural motifs of TMG-79 and the M5-15 (see Figs. 2,4). Aptamers 123 and 125 both show lower similarities than TMG-79 which would explain structural differences within aptamer 123 and 125 such as the appearance of struc-

reference	1	--GTATGGT--ACGGCG-CGGTGGCGGGTGCCTGG	30
		. .   . .     .	
TMG-79	1	ACGT-TGGTGCATGCCGAAGCCGGCGAATGC----	30
reference	1	-----GTA-TGGTACGGCGCGGTGGCGGGTGCCTGG	30
		. .	
Aptamer_123	1	GTGGGGTCCAGGCCTAGTACTGCTAAGGC-----	30
reference	1	GTATGGTACGGCGCGGTGGCGGGTGC--TGG-----	30
		.	
Aptamer_125	1	-----ACGTTTGGGGAAATAGTATCCCCCA	26
reference	31	----	30
Aptamer_125	27	ACAT	30

**Fig. 3. Pairwise alignment of N30 wobble region for top three selected aptamers based on PLP Fitness score against experimentally created aptamer.** Similarity: TMG-79 = 51.4%; Aptamer 123 = 20.8%; Aptamer 125 = 9.3%.

tural motifs not present in M5-15 and the change in relative position of stem loops.

### 3.2 Docking data

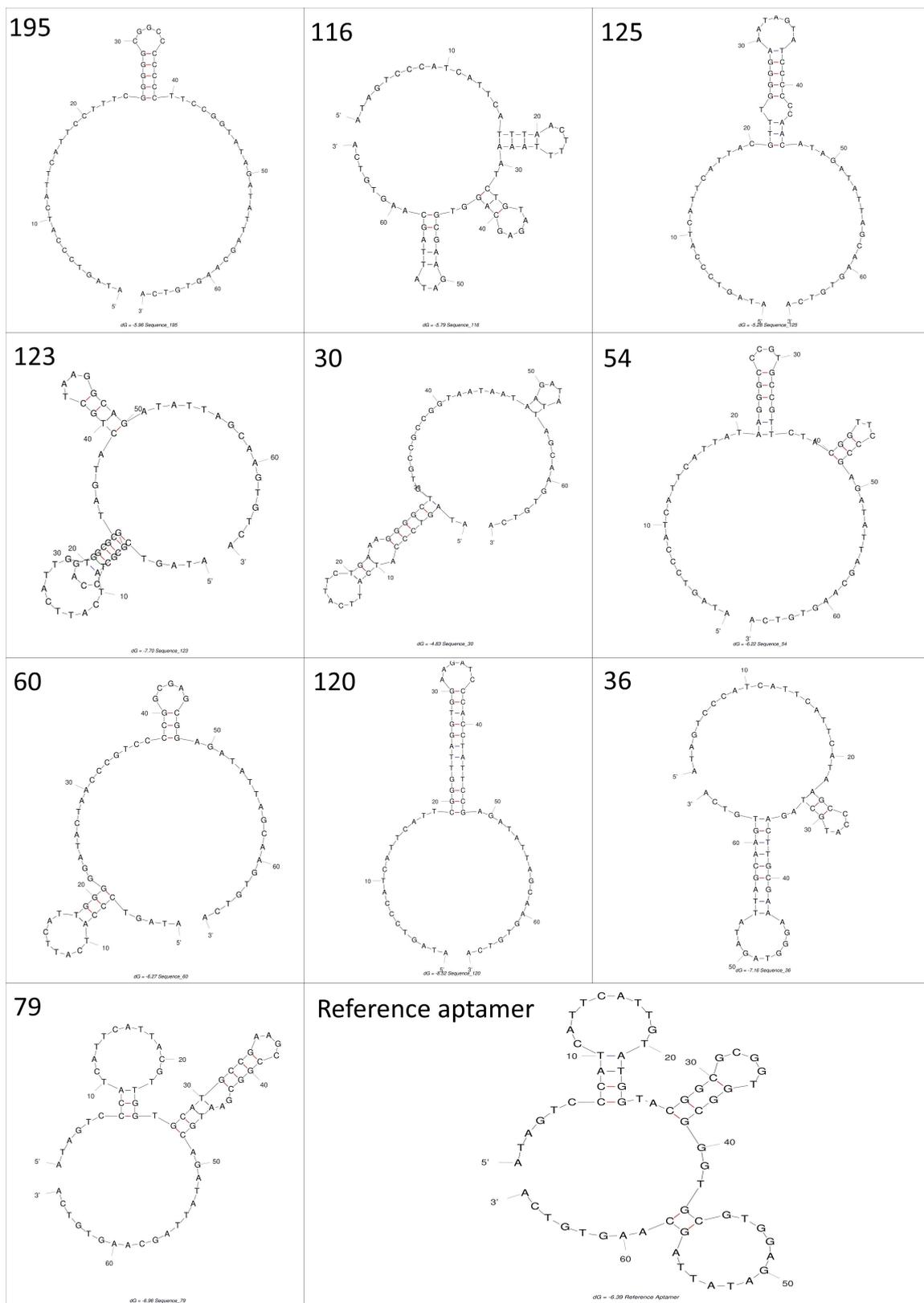
GOLD docking results show a large improvement in fitness scores when comparing positive and negative controls to *in silico* aptamers. TMG-79 PLP fitness (Fig. 5) indicates it could be a candidate for greater binding potential. Further analysis of docking data shows that aptamer complexes for candidates 123 and 125 have a larger number of predicted H bonds with a comparable average bond length (Table 5).

Despite having a lower fitness score than that of TMG-79, aptamer 123 was predicted to form a larger number of H-bonds making it a promising candidate. This aptamer also appears to show some minor structural similarities to the reference aptamer, particularly when examining its hairpin loop structures. Large variations in ChemPLP score were observed across the poses for each aptamer due to the low search efficiency used for virtual screening.

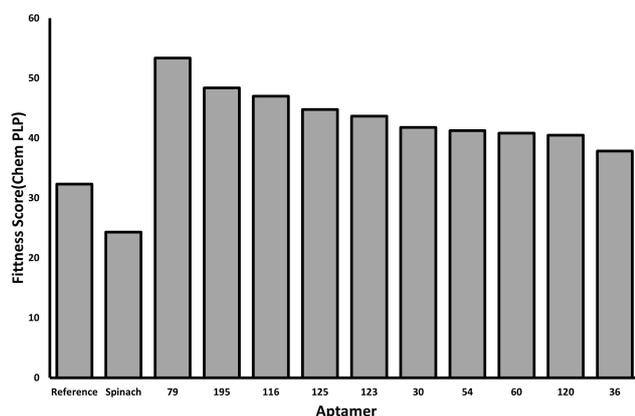
In addition, follow up experiments using AutoDock Vina demonstrating top pose and average binding affinity for the reference, spinach and the top 3 *in silico* generated aptamer sequences. Binding free energy was lower for all generated sequences compared to the M5-15, these changes were not statistically significant (TMG-79,  $p = 0.18$ ; aptamer 123,  $p = 0.17$ ; aptamer 125,  $p = 0.42$ ; Spinach,  $p = 0.18$ ).

Further analysis of the reference aptamer and TMG-79 showed that interactions occurred mainly in or around the stem loop structural motifs of each aptamer. The reference aptamer was predicted to form H-bonds at residues

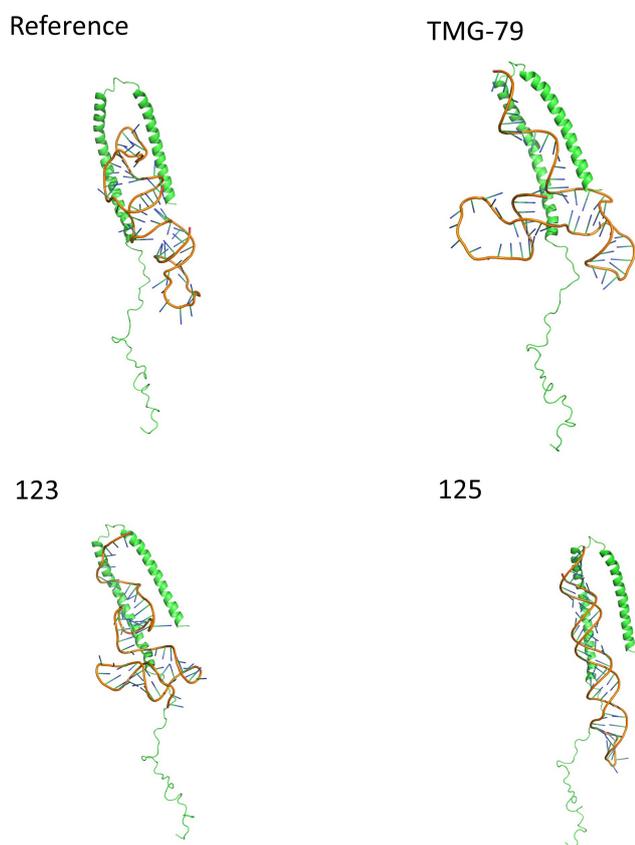
LYS97 and LYS21. In contrast, TMG-79 was predicted to form H-bonds at protein residues LYS80 and GLN99 suggesting that it may interact in a similar region to the reference aptamer. Differences in the overall structure of TMG-79 compared to the reference aptamer support the observation that the interactions are similar, but at different points within the protein. Aptamer 123, whilst having a lower overall ChemPLP fitness score, had a greater number of interaction points in these simulations, importantly with polar side chains of amino acids such as THR81, suggesting stability of these bonds within the complex. However, given that ChemPLP score is a summary of critical factors, including binding energy, this suggests that while there are greater interaction points, these interactions are less likely to form spontaneously and be stable. Bond distances observed from the reference aptamer are also greater than in 2 of the 3 aptamers generated, and mean binding energy predicted greater affinity across all aptamers screened. This indicates a tighter binding potential with the generated sequences compared to M5-15. Structural comparisons of the poses suggests that all aptamers bind to a similar region within  $\alpha$ -syn, primarily to residues contained within the 2 alpha helix motif containing LYS21 LYS43 LYS45 THR59 LYS80 THR81 SER87 LYS96 LYS97 GLN99 ASN103. These residues appear to be critical to the stabilisation of aptamer binding within the protein; the poses containing the *in silico* aptamers (Fig. 6) share remarkable similarity to the reference aptamer. Prediction of the final location of the aptamer within the central cleft within the target molecule also suggests that this may be a preferential location for a DNA molecule within  $\alpha$ -syn.



**Fig. 4. Predicted secondary structures for top 10 aptamers based on  $\Delta G$  predicted by mfold for sequences in digitally created library. M5-15, referred to here as the reference aptamer [37] raised against  $\alpha$ -syn, is included as a control within the process. (Energy rules: DNA, temp = 24 °C, 125 mM [Mg], 0 mM [Na]).**



**Fig. 5. Data showing the ChemPLP fitness score for the top poses of the previously selected aptamer sequences.** The reference aptamer and spinach aptamer serve as positive and negative controls, respectively. Each selected aptamer shows a higher docking score for their top pose when compared to the reference aptamer.



**Fig. 6. Showing the binding position and orientation of each aptamer in their top pose when docked against alpha synuclein.** Each aptamer is predicted to bind within the alpha helices of  $\alpha$ -syn similarly to the reference aptamer.

## 4. Discussion

The purpose of this investigation was to determine if it was possible, based on a template sequence derived experimentally, to improve upon the binding affinity of an ap-

tamer to a given target. While there has been significant advancement in aptamer generation in recent years, any improvements that can be made will naturally be of benefit in the future. As part of this work, the team has managed to generate several potential aptamer candidates which demonstrated higher affinity for the target molecule than the experimentally derived aptamer within our model. Data presented here suggests that it is possible to produce an aptamer which has the potential to be comparable or better than the experimentally derived sequence. While the data presented here does not suggest a significant difference in predicted binding affinity, it does demonstrate that there is a possibility with a greater data set and use of the aptamer in real world settings that the aptamer could either perform as well or outperform the reference aptamer. Analysis of the predicted interactions could shed light on the reasons for the observed increases in affinity, despite this not being statistically significant in this case.

### *Nature of the observed interactions*

When comparing the reference aptamer interaction with the generated sequences, there are similarities between the way they interact with  $\alpha$ -syn, specifically around the preference to bind within the alpha helix flanking structure within the protein. However, there are differences in the predicted binding between the reference and generated aptamers. Molecular docking relies on predicting the presence of intermolecular forces between two molecules, including van der Waals forces, hydrogen bonds, electrostatic interactions, side chain interactions, hydrophobic interactions and interaction of the solvent on docking (usually water) [39]. Scoring of a docking pose is a combination of all these factors, based on the assumptions and conditions that were imposed during the docking experiment [40]. The likelihood of a docking pose being true can be argued based on known interactions between amino acid residues, the involvement of any side chains and how they interact with DNA molecules. Within the docking poses proposed here, there are a number of interaction points that are observed particularly within amine groups and aromatic side chains between the DNA and the amino acids. These observations have been documented previously. Within all of the structures proposed, there were 5 amino acid residues which were consistently involved in some capacity. These amino acids were, threonine (Thr), lysine (Lys), glutamine (Gln), asparagine (Asn) and serine (Ser).

Thr and Lys interactions with DNA are well documented, with Thr interacting predominantly with phosphate backbone groups [41], and Lys predominantly interacting with guanine groups [42]. Specifically, when looking at Lys, it has been reported to interact with oxidized guanine residues (8-Oxo-G). Binding induces conformational change in the 8-oxoguanine DNA glycosylase (Ogg) required for the removal of 8-Oxo-G, an important DNA repair mechanism [42]. While there are no 8-Oxo-G residues present within the structure of the aptamer proposed here, it appears that there is a natural interaction between Lys and guanine. Thr

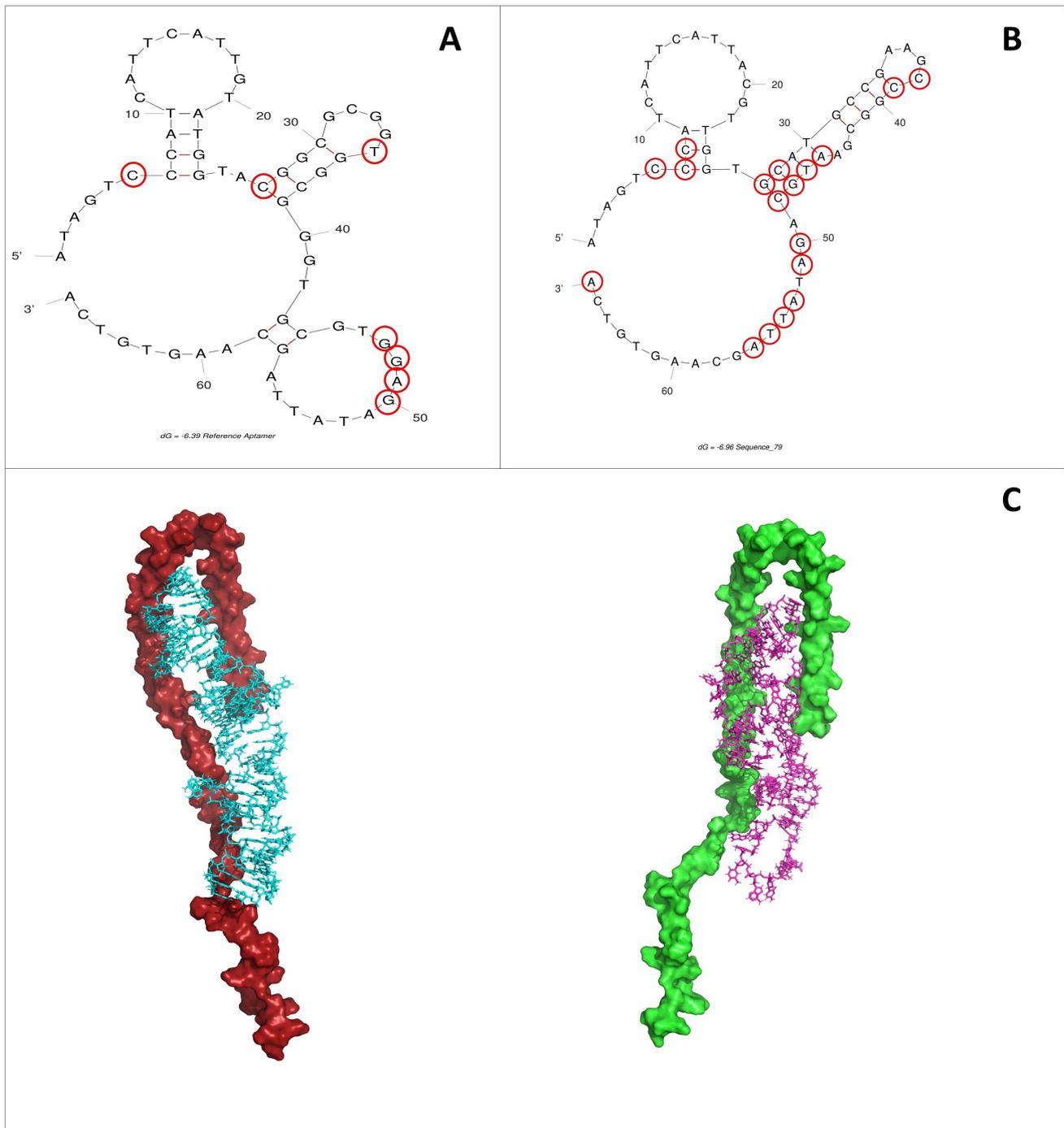
based hydrogen bonding has the highest propensity in highly hydrophilic environments [43], and since the target protein is water-soluble and primarily found in neuronal tissue, there is a suggestion that this interaction may occur *in vivo*. Following on from this, we should consider the possibility of the predicted Gln, Asn and Ser interactions between the generated aptamer and the target. Gln has been reported previously to have strong association with polar side chains and phosphates contained within the backbone, as well as interacting with the nitrogenous base component of adenosine [41]. Asn is generally associated with base recognition binding, predominantly to adenosine, and has been demonstrated in synthetic zinc finger constructs like 1meyC [44]. These interactions have also been observed in B-DNA binding of transcription factor-like elements [45]. Ser rich proteins are essential pre mRNA splicing factors. They regulate constitutive splicing and influence alternative splicing processes through an extensive and controlled phosphorylation of the Ser residues [46]. Ser residues are highly conserved within the DNA binding domains of bZip transcription factors of *Arabidopsis thaliana*, and were reported to be in physical contact with DNA [47] specifically providing a binding point through the hydroxymethyl side chain, something that was observed here in our model.

Interestingly, the non-species aptamer 4ts2 demonstrated a greater binding affinity for the target M5-15 which was an unexpected outcome of the work. This may be in part due to the very nature of  $\alpha$ -syn and its ability to associate with other “sticky” proteins. The unfolded nature of  $\alpha$ -syn allows it to function as a molecular chaperone assisting in the transport of vesicles and facilitating SNARE formation [48]. This function could extend to the adhesion of nucleotides to  $\alpha$ -syn, and could also suggest that RNA has a naturally greater affinity for  $\alpha$ -syn and might be a viable alternative to the DNA aptamers raised here and experimentally. This has been observed experimentally, with  $\alpha$ -syn being capable of binding its own mRNA to prevent translation of further protein [49]. Using catRAPID, an algorithm used to predict protein/RNA interactions, demonstrated  $\alpha$ -syn easily binds to its own mRNA without the need for an intrinsic riboswitch sequence within the mRNA [50].

When considering the structures of the aptamers, it is important to consider the presence and size of key structural motifs within the aptamer. Loops, both external and internal, tend to be much more flexible in nature when compared to standard Watson and Crick paired sections or other loops such as stems and bulges [51]. This inherent flexibility, combined with the relatively small size of aptamers in three-dimensional space, allows for the aptamer to integrate with regions of proteins that larger more complicated structures, like antibodies, are simply incapable of achieving [14]. This has been demonstrated previously by Autiero *et al.* [52] when looking into RNA aptamer binding to S8 ribosomal proteins. Here the ability of the aptamer to adjust its structure to bind to the target not only improves the binding efficiency of the

aptamer but is also a critical property of aptamers that warrants further investigation.

When looking at the reference aptamer and the top aptamer generated as part of this study, displayed in Fig. 7, the most notable differences between them are the extension of the stem loop between C7 and G25 and the incorporation of a greater number of Watson and Crick pairing between G27 and C48 in TMG-79. This increase in bulge size, and the incorporation of unpaired G44 and A45, gives greater flexibility to this extended region in comparison to the reference aptamer's similar structure between C26 and G39. An additional consideration is the increase in the number of potential binding points which have been predicted within TMG-79. This may be due to the increased flexibility of the aptamer compared to the reference aptamer which, while not being a rigid structure, has a greater number of structural motifs limiting its flexibility [52]. It is also important to remember that, while the increase in flexibility is important, the structural motifs themselves provide unique binding properties which allow for aptamer/protein interactions. However, there is evidence to suggest that both stable and unstable secondary motifs are important in aptamer binding. Adachi *et al.* [53] demonstrated that their RNA aptamer's inherent instability actually promoted tighter binding with higher specificity compared to aptamers which had more stable structures. This conflicts with Troisi *et al.* [54], where the structural motifs interact in a specific fashion to promote greater affinity for thrombin of NU172. The difference between these studies and their approach to the development of the aptamer may be down to either differences in complexity of the proteins or their specific locations. In both examples presented, they are looking at relatively complex proteins, either thrombin [54] or interleukin 17 receptor A/F (IL17A/F) inhibition [53]. Within the second example, the hope is to inhibit formation of IL17A/F, requiring only to be able to bind to one component to competitively inhibit this function. In contrast, Troisi *et al.* [54] identified sequence specific binding to the thrombin protein which could be significantly reduced based on single amino acid mutations. The conclusion that can be drawn from this is that location, structure and the complexity of the proteins involved dictates the requirement for either rigidity or flexibility within the aptamer's structure which influences its ability to bind to its target. Considering the inherent properties of  $\alpha$ -syn being a mainly unstructured protein with the exception of the two alpha helix motifs, it is interesting to see that both the reference aptamer and TMG-79 both bind within this alpha helix “sandwich”. Incorporation into this region and the increased affinity for the target observed with TMG-79 suggests that, while the protein may be mainly in an unstructured state, this is a critical recognition element for biomarker detection. While this data is promising, the requirement to get a solved crystal structure for the interaction is necessary to confirm the interactions predicted as part of this work.



**Fig. 7. Binding interaction points highlighted in red.** (A) M5-15. (B) TMG-79. (C) While there are similarities between the two aptamer structures (51.4% sequence homology), the notable difference is the lack of stem loop between G43 and C58 within the reference aptamer. This change results in a larger bulge region in TMG-79, allowing for increased flexibility of the overall aptamer structure. Top pose predictions of 4ts2 (left) and M5-15 (right) against  $\alpha$ -syn are shown.

## 5. Conclusions

While these models are based on artificial, computer-generated structures it does present an interesting case for moving towards a hybrid *in silico/in vitro* development strategy for aptamer selection. Within this work we have demonstrated using *in silico* methodologies that you can refine the structures of experimentally derived aptamers to provide a

structure which has the potential to have greater affinity for the target. Predicted interactions, as well as calculated free energy, suggest that TMG-79 has the potential to be a more effective means of detecting  $\alpha$ -syn. Moving forward, this work opens up opportunities to develop aptamers exclusively *in silico* with all the potential time and cost saving benefits this could bring.

## Author contributions

All authors contributed to the study conception and design. Model design, python scripting for library generation, docking simulations and data collection was performed by MR, TAI. MR and GDZ assisted in data interpretation and writing of the 1st draft of the manuscript. AF-BDA, SD, AAlbrakati, GESB and AAlexiou assisted with final manuscript writing. All authors approve the final version of this manuscript.

## Ethics approval and consent to participate

Not applicable.

## Acknowledgment

The authors would like to thank the University of Derby for supporting the project. We would like to thank the Taif University Researchers Supporting Program (project number: TURSP -2020/151), Taif University, Saudi Arabia.

## Funding

This work was supported by Taif University Researchers Supporting Program (project number: TURSP -2020/151), Taif University, Saudi Arabia.

## Conflict of interest

The authors declare no conflict of interest.

## References

- [1] Liu JKH. The history of monoclonal antibody development - Progress, remaining challenges and future innovations. *Annals of Medicine and Surgery*. 2015; 3: 113–116.
- [2] Shoaie M, Khorashadizadeh M, Derakhshani A, Safarnejad MR, Safarpour H. An overview of the current status of engineered therapeutic monoclonal antibodies. *International Pharmacy Acta*. 2019; 2: e9.
- [3] Little M, Kipriyanov SM, Le Gall F, Moldenhauer G. Of mice and men: hybridoma and recombinant antibodies. *Immunology Today*. 2000; 21: 364–370.
- [4] Zhou H, Pisitkun T, Aponte A, Yuen PST, Hoffert JD, Yasuda H, *et al*. Exosomal Fetuin-a identified by proteomics: a novel urinary biomarker for detecting acute kidney injury. *Kidney International*. 2007; 70: 1847–1857.
- [5] Desplat-Jégo S, Johanan C, Escande A, Goetz J, Fabien N, Olsson N, *et al*. Update on Anti-Saccharomyces cerevisiae antibodies, anti-nuclear associated anti-neutrophil antibodies and antibodies to exocrine pancreas detected by indirect immunofluorescence as biomarkers in chronic inflammatory bowel diseases: results of a multicenter study. *World Journal of Gastroenterology*. 2007; 13: 2312–2318.
- [6] Klutz S, Holtmann L, Lobedann M, Schembecker G. Cost evaluation of antibody production processes in different operation modes. *Chemical Engineering Science*. 2016; 141: 63–74.
- [7] Zhang P, Woen S, Wang T, Liau B, Zhao S, Chen C, *et al*. Challenges of glycosylation analysis and control: an integrated approach to producing optimal and consistent therapeutic drugs. *Drug Discovery Today*. 2018; 21: 740–765.
- [8] Mayer G. The chemical biology of aptamers. *Angewandte Chemie*. 2009; 48: 2672–2689.
- [9] Blind M, Blank M. Aptamer Selection Technology and Recent Advances. *Molecular Therapy-Nucleic Acids*. 2019; 4: e223.
- [10] Ellington AD, Szostak JW. In vitro selection of RNA molecules that bind specific ligands. *Nature*. 1990; 346: 818–822.
- [11] Tuerk C, Gold L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*. 1990; 249: 505–510.
- [12] Robertson DL, Joyce GF. Selection in vitro of an RNA enzyme that specifically cleaves single-stranded DNA. *Nature*. 1990; 344: 467–468.
- [13] Acquah C, Danquah MK, Yon JLS, Sidhu A, Ongkudon CM. A review on immobilised aptamers for high throughput biomolecular detection and screening. *Analytica Chimica Acta*. 2015; 888: 10–18.
- [14] Gelinat AD, Davies DR, Janjic N. Embracing proteins: structural themes in aptamer-protein complexes. *Current Opinion in Structural Biology*. 2016; 36: 122–132.
- [15] Krüger A, Zimbres FM, Kronenberger T, Wrenger C. Molecular Modeling Applied to Nucleic Acid-Based Molecule Development. *Biomolecules*. 2018; 8: 83.
- [16] Team HeidelbergTeam. 2021. Available at: <http://2015.igem.org/Team:Heidelberg/software/maws> (Accessed: 29 June 2021).
- [17] Sabri MZ, Abdul Hamid AA, Sayed Hitam SM, Abdul Rahim MZ. In Silico Screening of Aptamers Configuration against Hepatitis B Surface Antigen. *Advances in Bioinformatics*. 2020; 2019: 6912914.
- [18] Tombelli S, Minunni M, Mascini M. Analytical applications of aptamers. *Biosensors & Bioelectronics*. 2005; 20: 2424–2434.
- [19] Lecocq S, Spinella K, Dubois B, Lista S, Hampel H, Penner G. Aptamers as biomarkers for neurological disorders. Proof of concept in transgenic mice. *PLoS ONE*. 2018; 13: e0190212.
- [20] Stefanis L.  $\alpha$ -Synuclein in Parkinson's disease. *Cold Spring Harbor Perspectives in Medicine*. 2013; 2: a009399.
- [21] Crews L, Tsigelny I, Hashimoto M, Masliah E. Role of Synucleins in Alzheimer's Disease. *Neurotoxicity Research*. 2009; 16: 306–317.
- [22] Zheng Y, Qu J, Xue F, Zheng Y, Yang B, Chang Y, *et al*. Novel DNA Aptamers for Parkinson's Disease Treatment Inhibit  $\alpha$ -Synuclein Aggregation and Facilitate its Degradation. *Molecular Therapy-Nucleic Acids*. 2018; 11: 228–242.
- [23] Twohig D, Nielsen HM.  $\alpha$ -synuclein in the pathophysiology of Alzheimer's disease. *Molecular Neurodegeneration*. 2019; 14: 1–19.
- [24] Tsukakoshi K, Harada R, Sode K, Ikebukuro K. Screening of DNA aptamer which binds to  $\alpha$ -synuclein. *Biotechnology Letters*. 2010; 32: 643–648.
- [25] Ren X, Zhao Y, Xue F, Zheng Y, Huang H, Wang W, *et al*. Exosomal DNA Aptamer Targeting  $\alpha$ -Synuclein Aggregates Reduced Neuropathological Deficits in a Mouse Parkinson's Disease Model. *Molecular Therapy. Nucleic Acids*. 2020; 17: 726–740.
- [26] Zulkeflee Sabri M, Azzar Abdul Hamid A, Mariam Sayed Hitam S, Zulkhairi Abdul Rahim M. In-Silico Selection of Aptamer: a Review on the Revolutionary Approach to Understand the Aptamer Design and Interaction through Computational Chemistry. *Materials Today: Proceedings*. 2019; 19: 1572–1581.
- [27] Navien TN, Thevendran R, Hamdani HY, Tang T, Citartan M. In silico molecular docking in DNA aptamer development. *Biochimie*. 2021; 180: 54–67.
- [28] Hamada M. In silico approaches to RNA aptamer design. *Biochimie*. 2018; 145: 8–14.
- [29] Nabuurs SB, Wagener M, de Vlieg J. A flexible approach to induced fit docking. *Journal of Medicinal Chemistry*. 2008; 50: 6507–6518.
- [30] Wang T, Chen C, Larcher LM, Barrero RA, Veedu RN. Three decades of nucleic acid aptamer technologies: Lessons learned, progress and opportunities on aptamer development. *Biotechnology Advances*. 2019; 37: 28–50.
- [31] Markham NR, Zuker M. DINAMelt web server for nucleic acid melting prediction. *Nucleic Acids Research*. 2005; 33: 577–581.
- [32] Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*. 2003; 31: 3406–3415.
- [33] Popena M, Szachniuk M, Antczak M, Purzycka KJ, Lukasiak P,

- Bartol N, *et al.* Automated 3D structure composition for large RNAs. *Nucleic Acids Research*. 2012; 40: e112.
- [34] Macrae CF, Bruno IJ, Chisholm JA, Edgington PR, McCabe P, Pidcock E, *et al.* Mercury CSD 2.0– new features for the visualization and investigation of crystal structures. *Journal of Applied Crystallography*. 2008; 41: 466–470.
- [35] Schiffrin B, Radford SE, Brockwell DJ, Calabrese AN. PyXlinkViewer: A flexible tool for visualization of protein chemical crosslinking data within the PyMOL molecular graphics system. *Protein Science*. 2020; 29:1851–1857.
- [36] Liang X, Kuhn H, Frank-Kamenetskii MD. Monitoring single-stranded DNA secondary structure formation by determining the topological state of DNA catenanes. *Biophysical Journal*. 2006; 90: 2877–2889.
- [37] Ulmer TS, Bax A, Cole NB, Nussbaum RL. Structure and Dynamics of Micelle-bound Human  $\alpha$ -Synuclein. *Journal of Biological Chemistry*. 2005; 280: 9595–9603.
- [38] Warner KD, Chen MC, Song W, Strack RL, Thorn A, Jaffrey SR, *et al.* Structural basis for activity of highly efficient RNA mimics of green fluorescent protein. *Nature Structural & Molecular Biology*. 2014; 21: 658–663.
- [39] Meng X, Zhang H, Mezei M, Cui M. Molecular docking: a powerful approach for structure-based drug discovery. *Current Computer-Aided Drug Design*. 2011; 7: 146–157.
- [40] Nguyen ED, Norn C, Frimurer TM, Meiler J. Assessment and challenges of ligand docking into comparative models of G-protein coupled receptors. *PLoS ONE*. 2014; 8: e67302.
- [41] Luscombe NM, Laskowski RA, Thornton JM. Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Research*. 2001; 29: 2860–2874.
- [42] Faucher F, Wallace SS, Doublé S. The C-terminal lysine of Ogg2 DNA glycosylases is a major molecular determinant for guanine/8-oxoguanine distinction. *Journal of Molecular Biology*. 2010; 397: 46–56.
- [43] Jeong E, Kim H, Lee S, Han K. Discovering the interaction propensities of amino acids and nucleotides from protein-RNA complexes. *Molecules and Cells*. 2004; 16: 161–167.
- [44] Sathyapriya R, Vishveshwara S. Interaction of DNA with clusters of amino acids in proteins. *Nucleic Acids Research*. 2004; 32: 4109–4118.
- [45] Rooman M, Liévin J, Buisine E, Wintjens R. Cation- $\pi$ /H-bond Stair Motifs at Protein-DNA Interfaces. *Journal of Molecular Biology*. 2002; 319: 67–76.
- [46] Sanford JR, Ellis J, Cáceres JF. Multiple roles of arginine/serine-rich splicing factors in RNA processing. *Biochemical Society Transactions*. 2005; 33: 443–446.
- [47] Kirchler T, Briesemeister S, Singer M, Schütze K, Keinath M, Kohlbacher O, *et al.* The role of phosphorylatable serine residues in the DNA-binding domain of Arabidopsis bZIP transcription factors. *European Journal of Cell Biology*. 2010; 89: 175–183.
- [48] Spinelli KJ, Taylor JK, Osterberg VR, Churchill MJ, Pollock E, Moore C, *et al.* Presynaptic alpha-synuclein aggregation in a mouse model of Parkinson's disease. *Journal of Neuroscience*. 2014; 34: 2037–2050.
- [49] Surguchev AA, Surguchov A. Synucleins and Gene Expression: Ramblers in a Crowd or Cops Regulating Traffic? *Frontiers in Molecular Neuroscience*. 2019; 10: 224.
- [50] Zanzoni A, Marchese D, Agostini F, Bolognesi B, Cirillo D, Botta-Orfila M, *et al.* Principles of self-organization in biological pathways: a hypothesis on the autogenous association of alpha-synuclein. *Nucleic Acids Research*. 2013; 41: 9987–9998.
- [51] Musafia B, Oren-Banaroya R, Noiman S. Designing anti-influenza aptamers: novel quantitative structure activity relationship approach gives insights into aptamer-virus interaction. *PLoS ONE*. 2015; 9: e97696.
- [52] Autiero I, Ruvo M, Improta R, Vitagliano L. The intrinsic flexibility of the aptamer targeting the ribosomal protein S8 is a key factor for the molecular recognition. *Biochimica Et Biophysica Acta. General Subjects*. 2018; 1862: 1006–1016.
- [53] Adachi H, Ishiguro A, Hamada M, Sakota E, Asai K, Nakamura Y. Antagonistic RNA aptamer specific to a heterodimeric form of human interleukin-17a/F. *Biochimie*. 2011; 93: 1081–1088.
- [54] Troisi R, Napolitano V, Spiridonova V, Russo Krauss I, Sica F. Several structural motifs cooperate in determining the highly effective anti-thrombin activity of NU172 aptamer. *Nucleic Acids Research*. 2018; 46: 12177–12185.