Journal of
Medicinal Chemistry

Article

An investigation of the factors that dictate the preferred orientation of lexitropsins in the minor groove of DNA

Hasan Y. Alniss, Ini-Isabée Witzel, Mohamad H. Semreen, Pritam Panda, Yogendra Kumar Mishra, Rajeev Ahuja, and John A. Parkinson

J. Med. Chem., **Just Accepted Manuscript** • DOI: 10.1021/acs.jmedchem.9b01534 • Publication Date (Web): 28 Oct 2019

Just Accepted

is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

An investigation of the factors that dictate the preferred orientation of lexitropsins in the minor groove of DNA

Hasan Y. Alniss[a,b*], Ini-Isabée Witzel[c], Mohammad H. Semreen [a,b], Pritam K. Panda[d], Yogendra Kumar Mishra[e], Rajeev Ahuja[d,f], and John A. Parkinson[g]

[a]College of Pharmacy, University of Sharjah, P.O. Box 27272, Sharjah, United Arab Emirates.
[b]Sharjah Institute for Medical Research, University of Sharjah, P.O. Box 27272, Sharjah, United Arab Emirates.
[c]Core Technology Platform, New York University of Abu Dhabi, P.O. Box 129188, Saadiyat Island, Abu Dhabi, United Arab Emirates
[d]Department of Physics and Astronomy, Uppsala University, Box 516, SE-75120, Uppsala Sweden
[e]Functional Nanomaterials, Institute for Materials Science, Kiel University, Kaiserstr. 2, D-24143 Kiel, Germany
[f]Department of Materials and Engineering, Royal Institute of Technology (KTH), SE-10044, Stockholm Sweden
[g]WestCHEM, Department of Pure and Applied Chemistry, University of Strathclyde, 295 Cathedral Street, Glasgow G1 1XL, United Kingdom

Keywords: minor groove binders (MGB), lexitropsins, Androgen Response Element (ARE), Isothermal Titration Calorimetry (ITC), Molecular modeling

*Corresponding author: Phone: +(971)-6-5057427. Fax: +(971)-6-5585812. E-mail address halniss@sharjah.ac.ae

Abstract

Lexitropsins are small molecules that bind to the minor groove of DNA as antiparallel dimers in a specific orientation. These molecules have shown therapeutic potential in the treatment of several diseases; however, the development of these molecules to target particular genes requires revealing the factors that dictate their preferred orientation in the minor grooves, which to date have not been investigated. In this study, a distinct structure (thzC) was carefully designed as an analog of a well-characterized lexitropsin (thzA) to reveal the factors that dictate the preferred binding orientation. Comparative evaluations of the biophysical and molecular modeling results of both compounds showed that the position of the dimethylaminopropyl group and the orientation of the amide links of the ligand with respects to the 5'-3'-ends; dictate the preferred orientation of lexitropsins in the minor grooves. These findings could be useful in the design of novel lexitropsins to selectively target specific genes.

1-Introduction

The cationic lexitropsins, which bind reversibly to the minor groove of DNA, have been shown to have therapeutic efficacy in the treatment of cancer, bacterial, viral, and fungal diseases.[1] The biological significance of these compounds arises from their ability to bind reversibly in the minor groove of DNA and to control gene expression without inducing permeant DNA damage, which is commonly encountered with the use of currently available toxic chemotherapy. The significance of minor groove binders (MGBs) in anti-infective and anti-cancer therapy is currently being investigated as a promising new class of antibiotics for the treatment of *Clostridium difficile* infections, [2-4] and as novel transcription factor antagonists for the treatment of prostate cancer.[5-7]

A remarkable advancement in the MGB research has come from the discovery that distamycin binds to the minor groove as an antiparallel side-by-side dimer, in a head to tail fashion.[8-10] This opened new perspectives for developing novel MGBs with base-pair selectivity.[11-14] Thiazotropsin A (thzA), which is a synthetic analog of the natural polyamide distamycin A (Figure 1), can bind selectively to the 5'-WCWWGW-3' sequence (where W is A or T).[15-17] Its structure has a crescent shape that matches the curvature of the DNA minor groove. This geometrical fitness allows the ligand to bind with the minor groove of duplex DNA as an antiparallel dimer via hydrogen bonding with DNA bases (Figure 2B). This binding causes alterations in the DNA structure by widening the minor grooves and bending the DNA helix toward the major groove leading to a compression of the major groove. The ring slippage of the side-by-side minor groove binders (such as thzA, Figure 2A) has enabled these ligands to span six base pair DNA sequences, contrary to the expectation of four base pairs with respect to the ligand dimensions.[12, 18, 19] The importance of this in terms of gene targeting is that such ligands with small molecular weight and enhanced lipophilicity can be used to disrupt the binding of transcription factors to the response element of the target gene, which is composed of six base pair sequences. A response element is a short DNA sequence within the promoter region of a gene, which can bind to a specific receptor complex (transcription factors) and therefore regulate transcription of genes. There are many examples of transcription factors that recognize hormone response elements composed of a six base pair DNA sequence, such as the androgen receptor (AR), progesterone receptor (PR), glucocorticoid receptors (GR), and mineralocorticoid receptor (MR).[20] The androgen receptor recognizes the androgen response element (ARE), which is organized as repeats of the consensus DNA sequence 5′-WGWWCW-3' (W is A or T), with a three-nucleotide spacer.[20, 21] Malfunction of these transcription factors can lead to dysregulated gene expression, which is observed in many human cancers such as prostate cancer.[22-24]

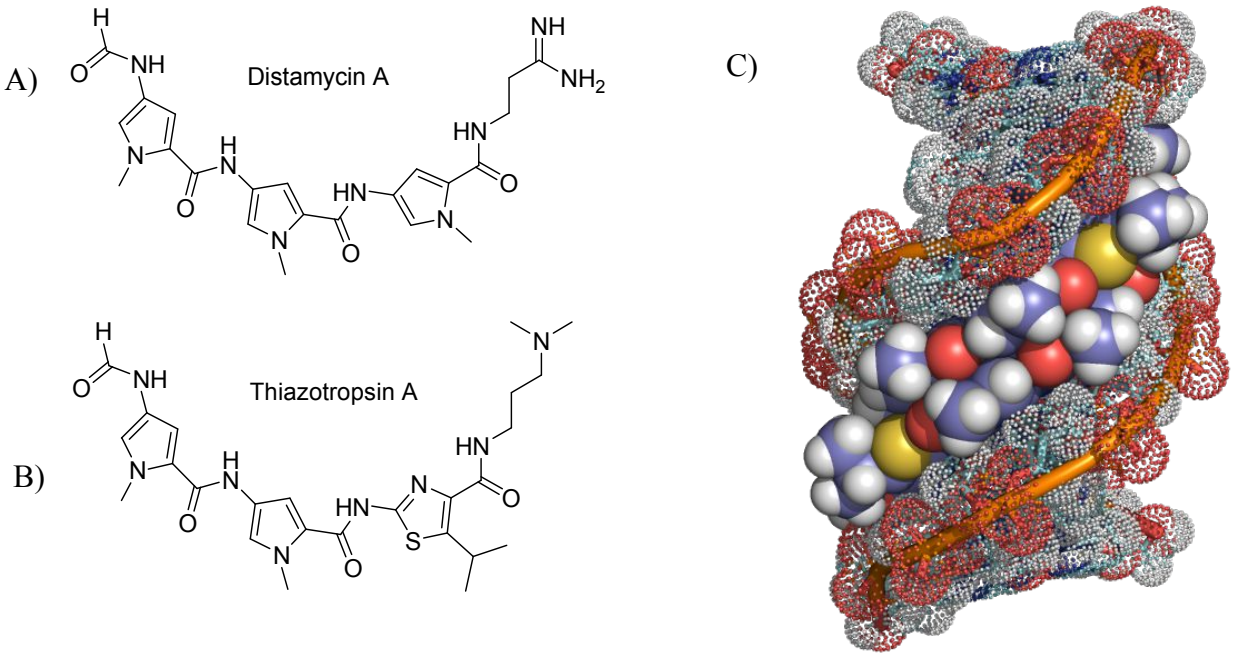


Figure 1. Structures of (**A**) distamycin, (**B**) thiazotropsin A and (**C**) NMR refined solution structure of a 2:1 complex of thiazotropsin A: DNA (PDB 1RMX).[19] The figure was developed using PyMOL.

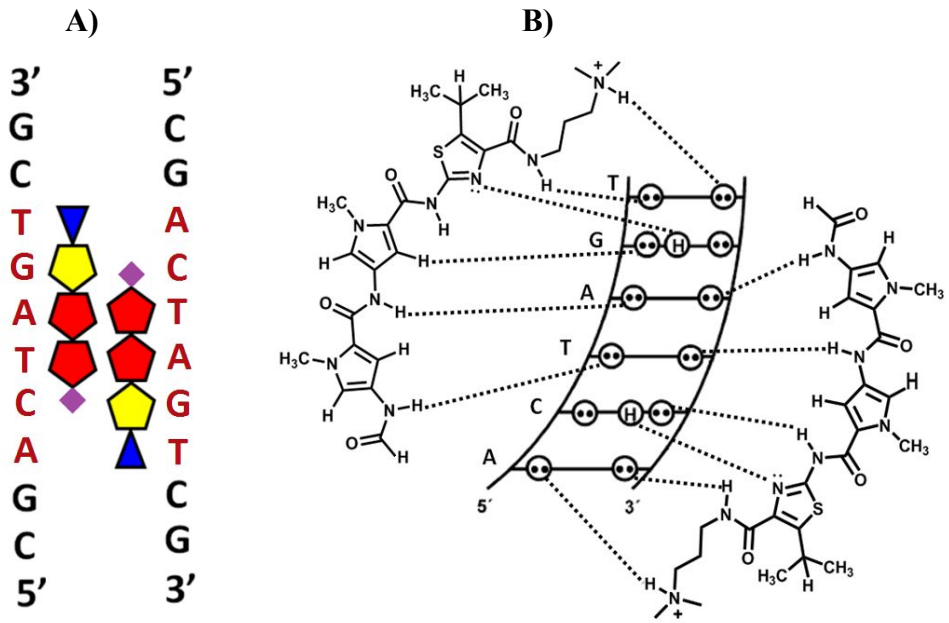


Figure 2. A) Schematic showing the alignment of thiazotropsin A dimer relative to the DNA sequence. **B)** The arrangement of hydrogen bonding between thiazotropsin A and the DNA sequenced(CG**ACTAGT**CG)$_2$.[19] Color coding: magenta diamond, formyl head; yellow pentagon, isopropyl thiazole; red pentagon, N-methylpyrrole; blue triangle, dimethylaminopropyl group.

The ability of thzA and its analogs to recognize a DNA sequence very similar to that of the ARE (for instance, thzA can recognize the 5′-WCWWGW-3′ sequence W is A or T) is encouraging', and the development of these ligands is therefore of value in order to target the ARE sequence and other promoter sequences. The small molecular weight of the lexitropsins, which are composed of only three aromatic rings, and their enhanced lipophilicity via the isopropyl thiazole system, could offer them additional beneficial physical properties in terms of absorption and oral bioavailability compared with the large molecular weight cyclic polyamide ligands or hairpin structures, which were developed by the Dervan group.[5, 18, 25] Moreover, although one might expect that the larger polyamide ligands (composed of 8 heteroaromatic rings) would have much higher binding affinity as they could form larger number hydrogen bonds with DNA, however, the fact that lexitropsins are able recognize the minor grooves as dimers enable them to form comparable number of hydrogen bonds with same binding site (5'-WGWWCW-3'). Tremendous research efforts have focused on revealing the factors that drive thiazotropsins-DNA associations, particularly the ligand's structural features that dictate DNA sequence selectivity, and significant progress was achieved in exploring structure-activity relationships (SARs) of these compounds, which is vital for the rational design and development of new MGBs. For instance, the binding properties of thzA[16, 17] (Figure 1) and its analogs, P22, P36, thiazotrobcin B and AIK 18-51 (Figure 3) are well characterized. [15-17, 26, 27] NMR, circular dichroism (CD) and isothermal titration calorimetry (ITC) studies revealed that thzA and its analogs recognize the minor groove of DNA as a slipped dimer where two of the three aromatic rings from each ligand stack on top of each other, allowing the ligand to read six base pairs (Figure 2 & 4). The ligand binds specifically to the central four base pairs (CWWG, W is A or T) while the other two bases at the edge of the binding site are covered by the dimethylaminopropyl (DAP) tail, which cannot differentiate between the four DNA bases.

The pairs of Thz/Pyr or Thz/Fmyl recognize G·C; Pyr/Thz or Fmyl/Thz recognize C·G but not A·T or T·A; the Pyr/Pyr pair does not distinguish

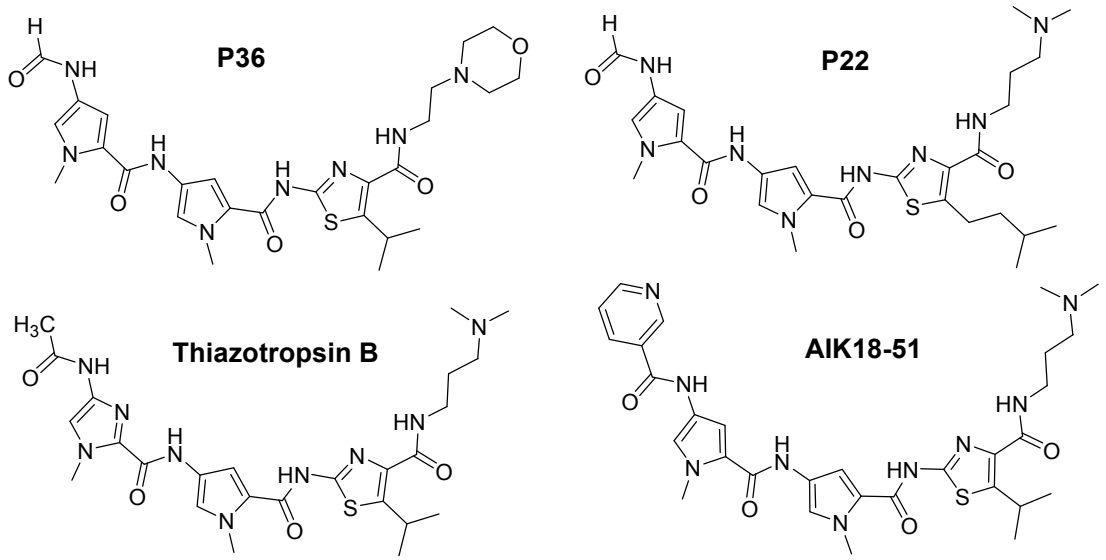


Figure 3. Structures of the thiazotropsin A analogs.

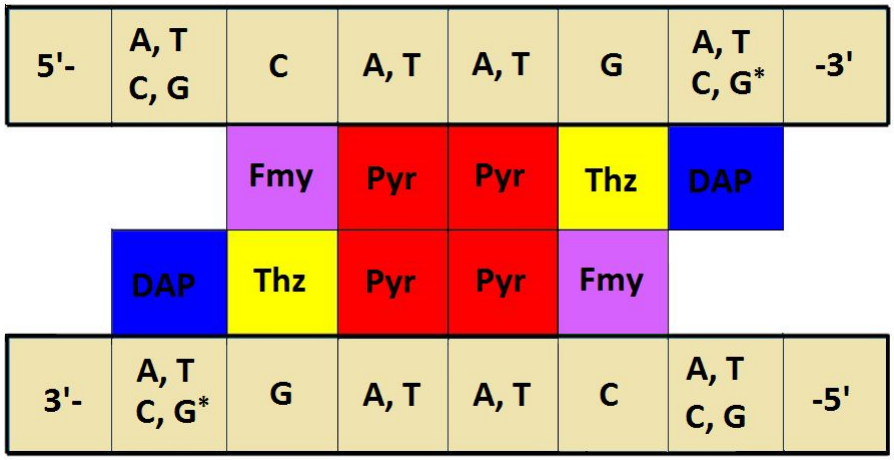


Figure 4. A schematic shows all the possible DNA recognitions by the slipped dimer of thiazotropsin A. * A steric clash between the NH_2 of guanine and the DAP of ligand reduces the affinity. Fmy, formyl "head"; Pyr, N-methylpyrrole; Thz, thiazole; DAP, dimethylaminopropyl "tail."

T·A and A·T but will not bind C·G or G·C; and the DAP tail covers the flanking bases, A, T, C, or G. A reduced affinity was observed when the DAP tail lies adjacent to G's because of a steric clash with the exocyclic NH_2 group.[17] These findings are consistent with Dervan's pairing rules for minor groove recognition.[28]

The development of these small molecules to selectively target specific DNA sequences requires an understanding of the factors that dictate the preferred ligand orientation in the minor groove concerning the 3' and 5' ends, which have not been thoroughly investigated to date. In all the studies that have been conducted, thzA and its analogs (P22, P36, and AIK 18-51) bind to 5′-WCWWGW-3′ in the minor groove as dimers in an anti-parallel side-by-side fashion and only a single orientation of the ligand dimer has been observed: the positively charged DAP groups are orientated towards the 3'-ends of the DNA strands (Figure 4). The role of isopropyl thiazole in guanine recognition is clearly understood; however, the question that remained unanswered is the role of the DAP group in dictating the preferred dimer orientation toward the 3'-ends, and whether reversing the position of the DAP group (from N- to C-terminus) in the ligand structure will change the preferred dimer orientation toward the 5'-ends or not?

In the current work, a new compound, thiazotropsin C (thzC), Figure 5, was carefully designed and synthesized to investigate the role of the DAP tail in dictating the preferred orientation of the lexitropsins anti-parallel dimer in the minor groove of DNA with respect to the 5'- and 3'-ends. The proposed structure, thzC, was developed from the structure of thiazotropsin A. In the new structure, the positions of both the DAP tail and the isopropyl thiazole in thzA were changed: the DAP group is attached to the C-terminus, and the isopropyl thiazole replaces the pyrrole ring located towards the C-terminus. Studying the DNA binding behavior of this compound compared

to that of thzA helps in both revealing the factors that dictate the preferred orientation of lexitropsins anti-parallel dimers in the minor groove of DNA, and the development of these small molecules to target specific genes such as the androgen response (ARE).

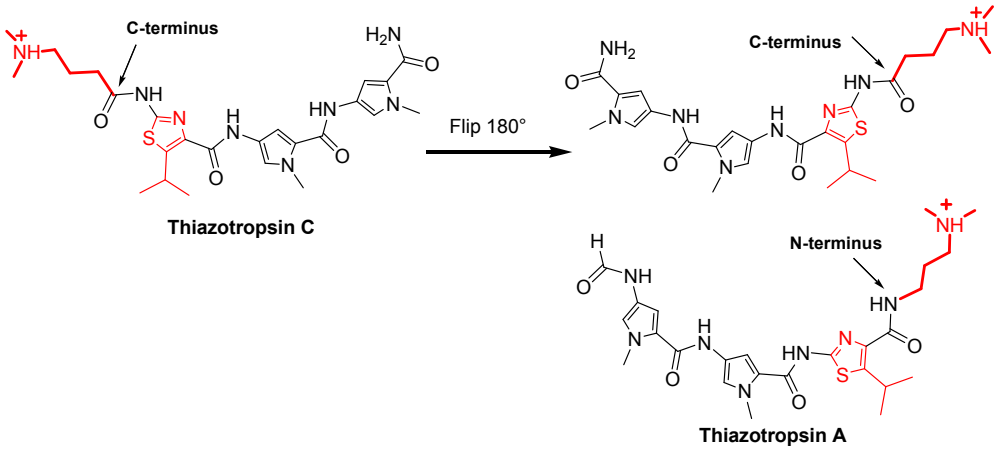


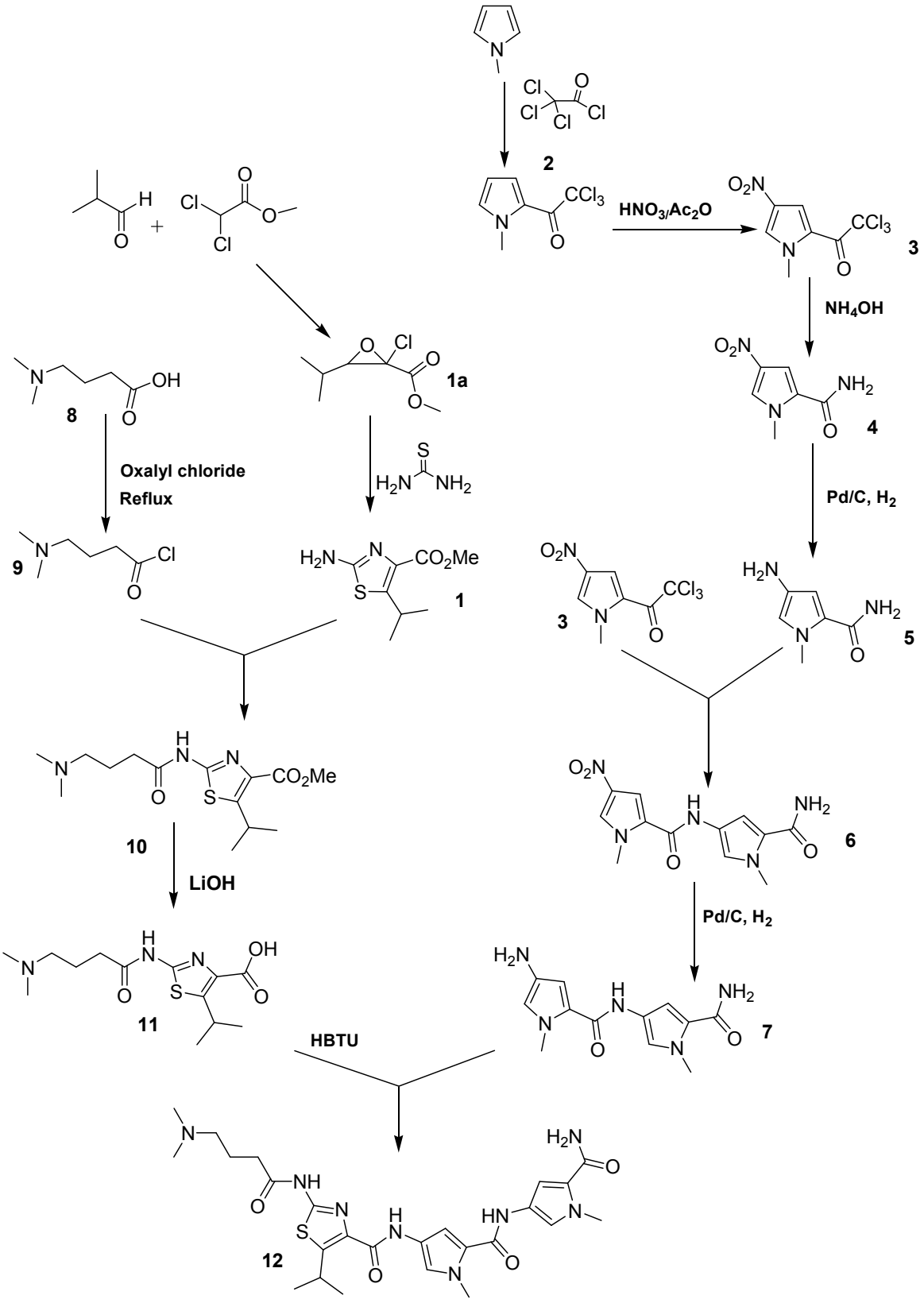
Figure 5. Structure of thiazotropsin A (DAP at the N-terminus) and its proposed analog, thiazotropsin C (DAP at the C-terminus).

3. Results and discussion

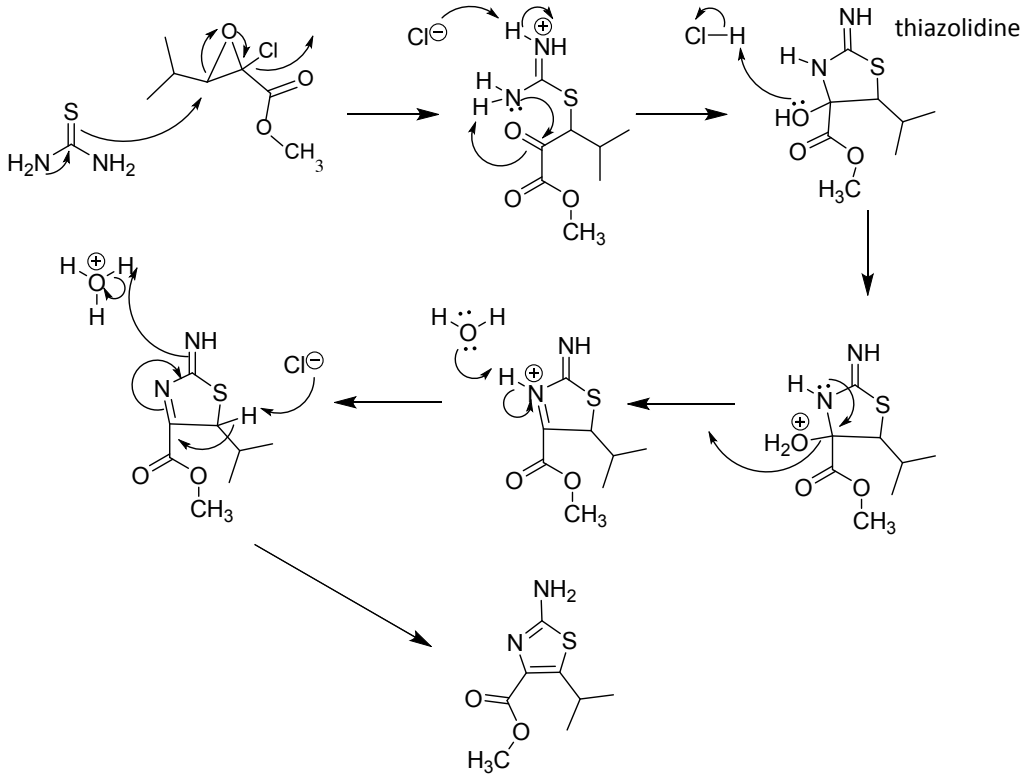
3.1 Chemical Synthesis

The synthesis of the target compound, thiazotropsin C, was achieved by using convergent synthesis. The detailed synthetic PDB pathway is summarized in Scheme 1. To synthesize thiazotropsin C, several MGB monomers were prepared (**1, 2, 3, 4, and 5, Scheme 1**). Methyl-2-amino-5-isopropyl-1,3-thiazole-4-carboxylate[29] **1** was prepared in good yield (70 %) by reacting methyl dichloroacetate with isobutyl aldehyde at 0 ºC in diethyl ether in the presence of sodium methoxide as a strong base to give the α-chloro glycidic ester intermediate **1a (**Darzens reaction, Supporting Information S2**).**[30] Since the α-chloro glycidic ester is highly reactive, it was directly used for the second step without further purification. The second step involves the coupling between the α-chloro glycidic intermediate **1a** and thiourea in dry methanol to give the methyl 2-

amino-5-isopropyl-1,3-thiazole-4-carboxylate **1.** The reaction proceeds via nucleophilic attack of the thiourea sulfur atom on the epoxide carbon which leads to epoxide ring-opening and the formation of new α-carbonyl moiety. The formation of the isopropylthiazole ring was achieved by subsequent nucleophilic attack of the thiourea amino group on the α-carbonyl carbon in a common Schiff's base reaction[31] to form the thiazolidine ring (**Scheme 2**). Two dehydration steps of the thiazolidine ring then restore the aromaticity of the thiazole ring. 2-Trichloroacetyl-1-methylpyrrole (**2)** was prepared in good yield (77%) by the reaction of trichloroacetylchloride and 1-methylpyrrole. **2-Trichloroacetyl-1-methyl-4-nitropyrrole (**Compound **3)** was prepared by the nitration of 2-trichloroacetyl-1-methylpyrrole (**2)** with nitric acid in acetic anhydride, in good yield (70%). The presence of the deactivating group, trichloroacetyl, attached to the pyrrole ring plays an important role in improving the yield of nitro product. Nitration of the carboxylic acid derivative of **3** produced much lower yield (25 %)**.** These results suggest that the trichloroacetyl group is more efficient than the carboxylic acid in deactivating the electron-rich pyrrole nucleus, thus reducing the poly-nitration of N-methylpyrrole ring at position C3 and C5. The amide product **4** was obtained after preparing **3**, which was reacted with ammonium hydroxide to give **4** in a very good yield (81%).



Scheme 1: Synthetic pathway for thiazotropsin C.



Scheme 2: Mechanism of thiazole ring formation by the coupling between α-chloro glycidic ester and thiourea.

Compound **5** was prepared by reducing the nitro compound **4** by catalytic hydrogenation. This compound was directly used for the next step of synthesis due to the lack of stability of the amino pyrrole product. The amide-linked pyrrole dimer **6**, N-(5-carbonyl-1-methyl-1H-pyrrol-3-yl)-1-methyl-4-nitro-1H-pyrrole-2-carboxamide, was prepared by the coupling between **3** and **5**. The reaction mixture was heated to 70 ºC overnight, and that was enough to form the amide linkage directly by releasing the trichloromethyl group (it leaves as chloroform) from the carbonyl group. This route was efficient, quick and less laborious. It saved the time required for the synthesis of

the carboxylic acid derivative of **3** and the use of coupling agents to activate the carboxylic acid to form the amide link. Compound **7** was prepared by reducing the nitro group of **6** via catalytic hydrogenation using Pd-C/H_2. The amino compound **7** was used directly for the next step of synthesis due to the lack of stability of the amino group attached to the pyrrole ring.

4-(Dimethylamino) butanoyl chloride **9** (Scheme 1) was prepared by refluxing **8** in oxalyl chloride to produce the acid chloride, which was directly coupled with the amine compound **1** to obtain **10** in good yield (70 %). Compound **11** was obtained by the basic hydrolysis of the methyl ester in compound **10** to corresponding carboxylic acid in good yield (80%). Mild basic hydrolysis conditions (lithium hydroxide at room temperature) were used for the hydrolysis of the methyl ester in order to avoid the hydrolysis of amide group in the compound.

Thiazotropsin C**, 12,** N-(5-(5-carbonyl-1-methyl-1H-pyrrol-3-ylcarbomyl)-1-methyl-1H-pyrrol-3-yl)-2-(4-(dimethylamino)butanamid)-5-isopropylthiazole-4-carboxamide, was prepared by coupling of the pyrrole dimer amine **7** with the carboxylic acid **11** using HBTU as a coupling agent. Structures of all intermediates and final products were characterized by using ^{1}H and ^{13}C NMR spectroscopy and mass spectrometry.

3.2 The rationale of ligand design and interpretation of possible binding orientation

Thorough structural and biophysical studies have been previously conducted for the characterization of thzA-DNA associations. These studies revealed that thiazotropsin A binds to the 5′-WCWWGW-3' sequence (W is A or T) in the minor groove of DNA as a dimer in an anti-parallel side-by-side fashion, and most interestingly, only one orientation of the ligand dimer was observed: the positively charged DAP groups are orientated towards the 3’-ends of the DNA strands (Figure 2A). However, the ligand's structural features that dictate the preferred dimer

orientation with respect to the 5'-3'-ends, especially the role of the DAP group, are still not understood.

The proposed ligand structure, thiazotropsin C, was therefore specifically designed to investigate the factors that dictate the preferred orientation of lexitropsin anti-parallel dimers in the minor groove of DNA by changing the positions of both the DAP tail (dimethylaminopropyl) and the isopropyl thiazole in thzA (Figure 5). In theory, there are four possibilities for the binding of thiazotropsin C to the minor groove if it is compared with the binding properties of the lead compound thzA with 5′-WCWWGW-3′. Each possibility would suggest certain ligand structural features responsible for DNA minor groove recognition, which might reveal the factors that dictate the preferred orientation of lexitropsins anti-parallel dimers with respect to 5′-3′ ends.

1- ThzC recognizes only the 5′-WCWWGW-3′ sequence as a dimer, with its positively charged DAP group lying towards the 3'-end, and the carbonyl group of each amide (colored blue) linking the aromatic rings is closer to the 3'-end than the partnering NH (Figure 6A). If the biophysical experiments confirm the binding of thzC to the 5′-WCWWGW-3' sequence, such results would suggest that the orientation of lexitropsins in the minor groove is dictated by the positively charged DAP group of the ligand. In this case, the DAP group always prefers to point towards the 3′-end regardless of its position in the ligand structure at the N- or C- terminus, and regardless of the two possible orientations of the amide links (blue/red colored amides) with respect to the 5'-3'-ends. These conclusions are based on the fact that the DAP group in thzA, which is linked to the N-terminus, always points toward the 3'-end (Figure 7A). Consequently, if thzC, which has its DAP group linked to the C-terminus, also prefers to point toward the 3'-end, such

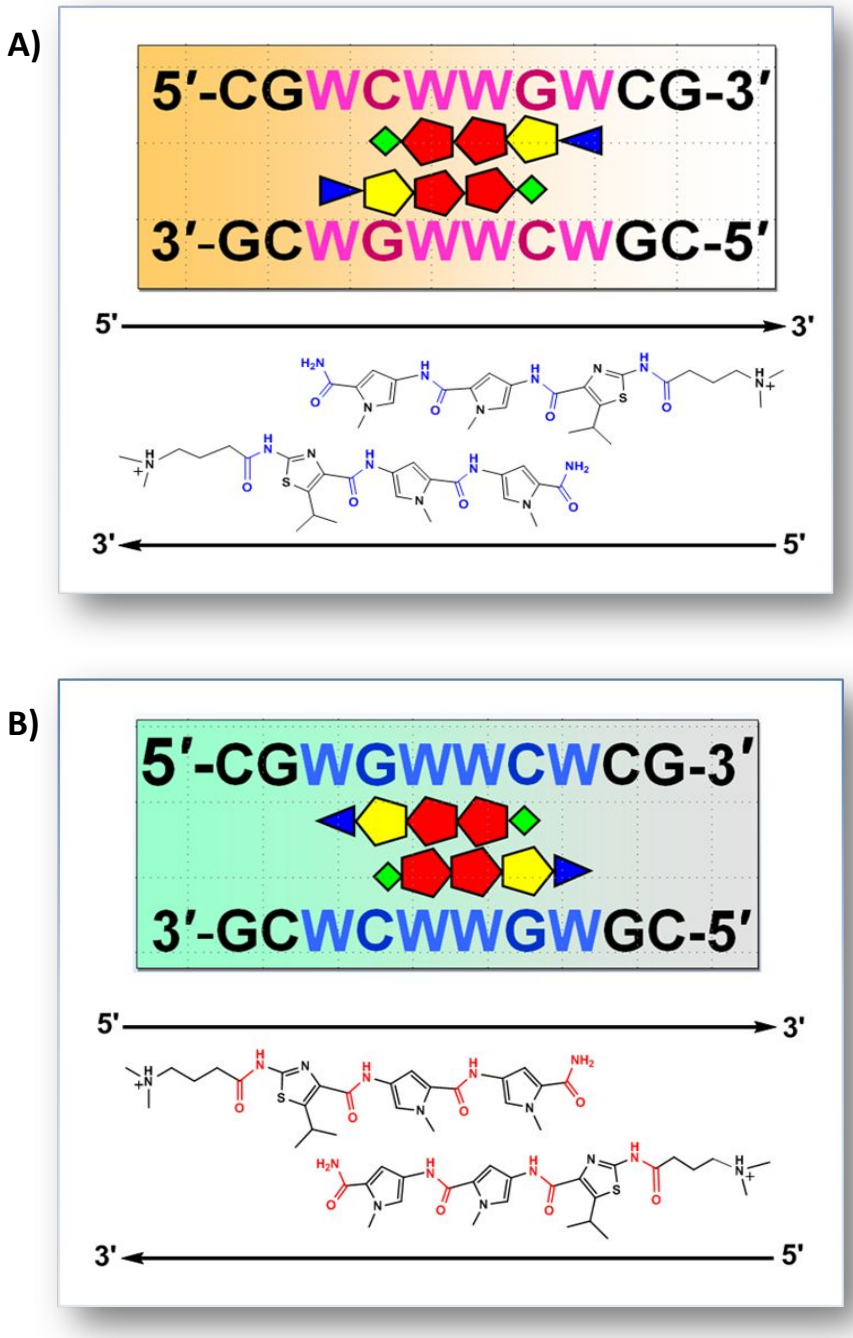


Figure 6 Representation of the possible binding orientations that may be adopted by thiazotropsin C in the minor groove of DNA. **A)** The ligand is aligned in a 5'-3' direction and recognizes the 5'-WCWWGW-3' sequence. **B)** The ligand is aligned in a 3'-5' direction and recognizes the 5'-WGWWCW-3' sequence.

results will indicate that the DAP group is responsible for dictating the preferred orientation of lexitropsin dimers in the minor groove of DNA.

2- ThzC recognizes only the 5′-WGWWCW-3′ sequence as a dimer, with its positively charged DAP group lying toward the 5'-end and the carbonyl group of each amide (colored red) linking the aromatic rings is closer to the 5'-end than its partnering NH (Figure 6B). If the biophysical experiments confirm the binding of thiazotropsin C to the 5′-WGWWCW-3′ sequence, such results would suggest that the factor that dictates the orientation of lexitropsin ligands in the minor groove is the orientation of the amide links with respects to the 5'-3'-ends. In this case, the amide link always prefers an orientation where its carbonyl group is closer to the 5'-end than the partnering NH regardless of the DAP group position in the ligand structure, whether it is attached to the *N*- or *C*- terminus. This conclusion is supported by the fact that thzA recognizes 5′-WCWWGW-3′, where its carbonyl group is closer to the 5'-end than the partnering NH (Figure 7A).

3- ThzC exhibits both orientations. If the biophysical experiments confirm the binding of thzC to both the 5′-WGWWCW-3′ and 5′-WCWWGW-3′, such results would indicate that the DAP tail and the orientation of the amide links with respects to the 5'-3'-ends play no role in dictating the orientation of lexitropsin dimers in the minor groove of DNA. This lack of sequence selectivity could happen if the ligand dimer is highly flexible and can adopt multiple conformations to match the helical structure of the DNA minor grooves at different binding sites. However, based on the flexibility of thzC, which is similar to that of thzA, this scenario is not expected, because the biophysical studies on thzA have shown that lexitropsins prefer to adopt a single orientation (Figure 7).

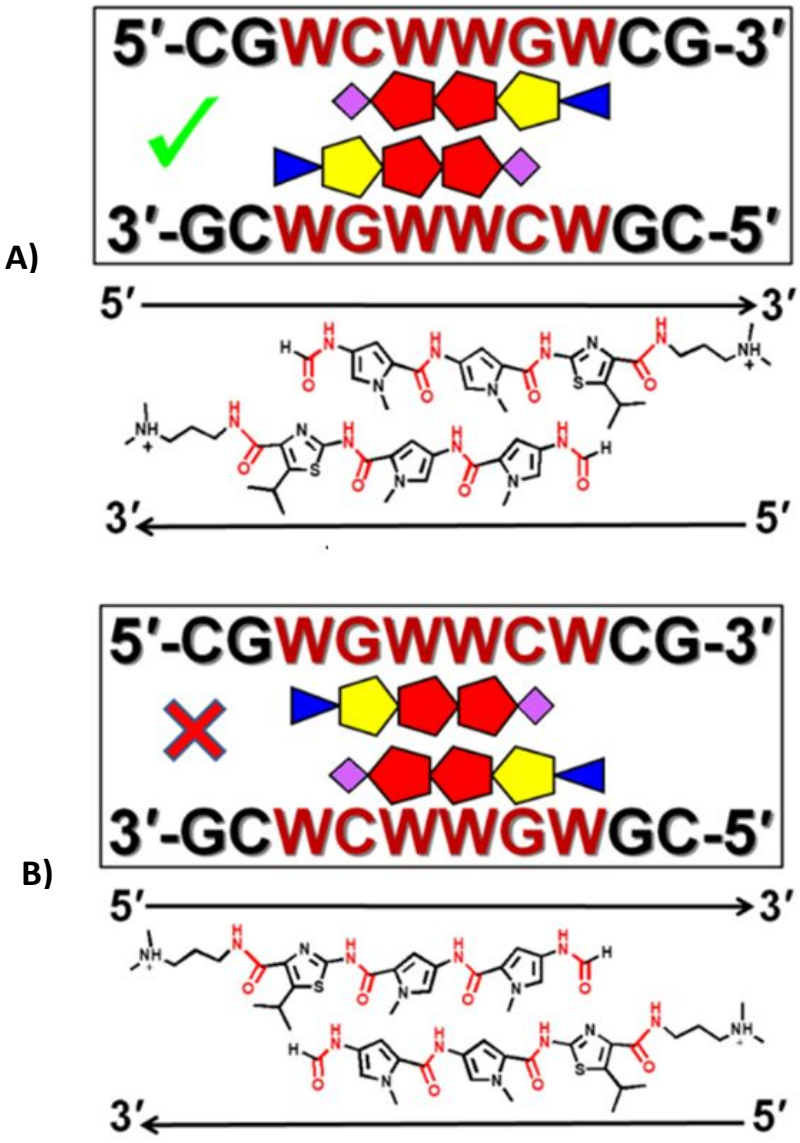


Figure 7 Representation of the confirmed single binding orientations adopted by thiazotropsin A in the minor groove of DNA **A)** The ligand aligned in a 5'-3' direction, recognizing the 5'-WCWWGW-3' sequence **B)** Biophysical studies showed that thzA does not recognize the 5'-WGWWCW-3' sequence, thus ligand alignment in a 3'-5' direction is not possible.

4- In the fourth scenario, the ligand does not recognize both orientations. If the biophysical experiments show that thzC neither binds to 5′-WGWWCW-3′ nor 5′-WCWWGW-3′ or very weak binding is observed, such results would suggest that both the position of the DAP group and the orientation of the amide links with respects to the 5'-3'-ends dictate the preferred orientation of lexitropsin ligands in the minor groove of DNA. The inability of thzC to recognize the 5'-WGWWCW-3′ presumably indicates a significant steric clash between the DAP group and the backbone of DNA when the DAP, which is attached to the C-terminus, is lying toward the 5'-end (Figure 6B). These results would suggest that the orientation of lexitropsins in the minor groove is dictated by the position of the positively charged DAP of the ligand. In this case, the DAP group should be attached to the N-terminus and always prefers to point towards the 3′-end. The inability of thzC to recognize the 5'-WCWWGW-3′ sequence would indicate that the orientation of the amide links with respects to the 5'-3'-ends also dictates the preferred orientation of lexitropsin dimers in the minor groove. In this case, the amide link always prefers an orientation where its carbonyl group is closer to the 5'-end than the partnering NH, which implies that this orientation is crucial for the formation of hydrogen bonds with DNA bases.

Recent data from Dervan's group have shown that the hairpin and cyclic polyamide minor groove binders, comprised of 8 imidazole/pyrrole rings which bind to the ARE (5′-WGWWCW-3′), can regulate the expression of androgen receptor target genes in cell culture studies.[5] Although it was concluded from these studies that the relatively large and polar polyamide structures are cell-permeable, absorption studies carried out on cyclic polyamide compound showed low Caco-2

permeability, suggesting that it might not be orally bioavailable.[25] The ability of lexitropsins to recognize a DNA sequence very similar to that of the ARE (for instance, thzA[17] can recognize the 5′-WCWWGW-3′ sequence, where W is A or T) is encouraging and understanding the factors that dictate the preferred dimer orientation of lexitropsin in the minor groove is crucial for the development of these ligands in order to target the ARE sequence. The small molecular weight of the lexitropsins, which are composed of only three aromatic rings and their enhanced lipophilicity via the isopropyl thiazole system, could offer them additional beneficial physical properties compared with the large molecular weight polyamide ligands. Furthermore, additional lipophilic functional groups could be introduced without sacrificing binding affinity as these substituents point out from the minor grooves.

3.3 Biophysical testing and evaluation of thiazotropsin C-DNA association

The interaction of thiazotropsin C with ten dodecamer ODNs (5'-GCG**AGTACT**CGC-3')d, (5'-GCG**AGATCT**CGC-3')d, (5'-GCG**TGATCA**CGC-3')d, (5'-GCG**TGTACA**CGC-3')d, (5'-GCG**ACTAGT**CGC-3')d, (5'-GCG**ACATGT**CGC-3')d, (5'-GCG**TCATGA**CGC-3') d, (5'-GCG**TCTAGA**CGC-3')d, (5'-GCG**AGAACA**CGC-3')d, (5'-GCG**TGTTCT**CGC-3')d were evaluated using Isothermal Titration Calorimetry (ITC). ITC is used to directly determine the thermodynamic profile of intermolecular binding at constant temperature and that includes the determination of enthalpy (ΔH), entropy (ΔS), binding affinity (K), stoichiometry (N), and free energy of binding (ΔG).[32-35] ITC experiments were carried out at low salt concentration in order to avoid the competition between the ions of salt and thzC molecules, which would be more pronounced at higher salt concentration. The ODNs were selected based on the 5'-WCWWGW-3' sequence, the binding sequence of thzA, and its complementary sequence, 5'-WGWWCW-3',

which might be recognized by thzC if its orientation is reversed in the minor grooves. The ITC results showed that thzC does not recognize both the 5'-WGWWCW-3' and 5'-WGWWCW-3' frame (Figure 8). A control experiment was first performed by titrating the ligand into the buffer solution, in which endothermic peaks appeared as a result of ligand dilution in the buffered solution. ThzC was then titrated to solutions containing ten different DNA dodecamers (supporting information S3); however, no change in the heat was observed; only endothermic peaks appeared which were identical to those of ligand dilution in the buffer. The inability of thzC to recognize both 5'-WGWWCW-3' and 5'-WGWWCW-3' suggests that the factors that dictate the orientation of lexitropsin dimers in the minor grooves are both the position of the DAP group in the structure of the ligand and the orientation of the amide links with respect to the 5'-3'-ends. This implies that the DAP group must be attached to the N-terminus and point toward the 3'-ends to avoid a steric clash with the backbone of DNA, and the carbonyl of amides must be closer to 5'-ends than its partnering NH to effectively form hydrogen bonds with the DNA bases.

For the 5'-WGWWCW-3' sequence, given that the heterocyclic units of thzC dimer are compatible with the central GWWC sequence, the inability of thzC to recognize this sequence indicates a significant steric clash between the DAP group and the backbone of DNA or DNA bases. To evaluate this conclusion, the active conformation of the DAP group in the resolved NMR structure of the complex between $d(CGACTAGTCG)_2$ and thzA was inspected.[19] The NMR structure of thzA-DNA complex shows that DAP is folded in a way to enable the protonated tertiary amine to form an intra hydrogen bond with the adjacent carbonyl oxygen (Figure 9). This folded conformation makes the DAP group of each ligand in the dimer lie in the middle of the minor grooves, deviated toward the opposite DNA strand. The folded conformation of the DAP group may explain why there is one preferred orientation for the thzA dimer to recognize the minor

groove with respect to the 5-' and 3'-end. Reversing the position of ligand 1 and 2 of the thzA dimer (flipped thiazotropsin A dimer-FthzA), as shown in Figure 10B is expected to allow FthzA to read the 5'-WGWWCW-3' sequence, however biophysical studies showed that this recognition is not possible[17, 19] and that this is presumably due to a significant steric clash between the DAP group and the walls of the minor groove when the DAP group points toward the 5'-end. Moreover, in the FthzA dimer, the orientations of the ligand amide links with respect to the 5' and 3' ends

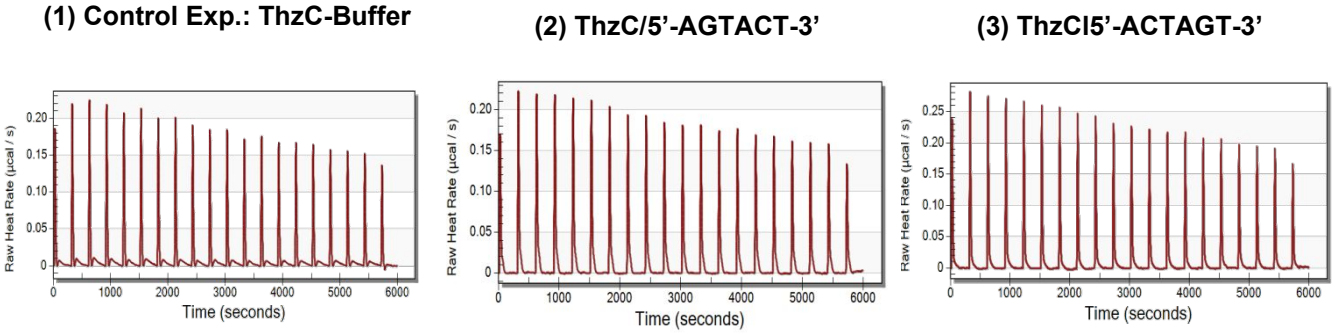


Figure 8. ITC titrations of thzC to (1) PIPES Buffer (control exp.) (2) d(5'-GCG**AGTACT**CGC-3') (3) d(5'-GCG**ACTAGT**CGC-3'). No change in the heat was observed upon ligand titration to the DNA dodecamers suggesting the absence of interaction.

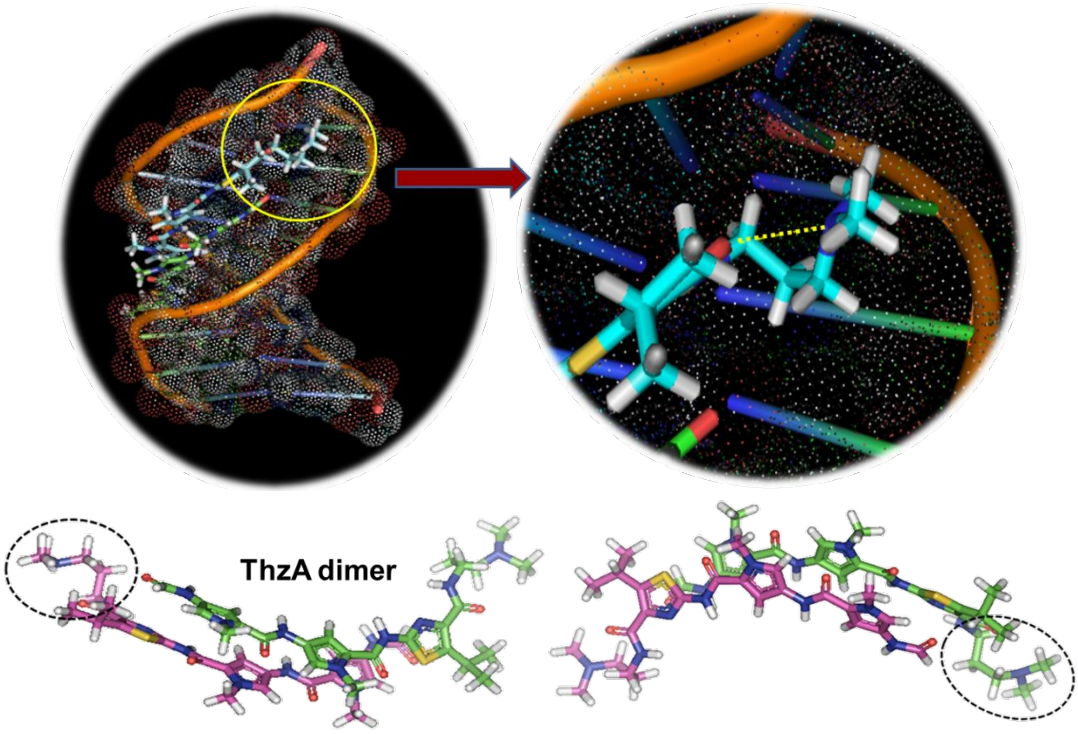


Figure 9. Representation of the NMR refined solution structure of thzA complex with 5′-ACTAGT-3′ (1rmx) showing the confirmation of the DAP group. The protonated nitrogen of the DAP group is attracted to the carbonyl oxygen to form a strong intra hydrogen bond.

are reversed, and this might disrupt the hydrogen bonds formed between the ligand amide groups and DNA bases. Hydrogen bonds are directional, and their strength is also profoundly affected by the distance between hydrogen bond acceptor and donor. These assumptions were supported by a molecular modeling study, as shown in the next section.

The inspection of the NMR structure of thiazotropsin B-DNA complex also shows that the DAP group of thzB adopts the folded conformation.[15] If we assume that the DAP group of thzC, which is attached to the C-terminus, adopts the folded confirmation, this can explain why thzC does not recognize both 5'-WGWWCW-3' and 5'-WCWWGW-3'. The alignment of the thzC dimer in the minor groove of 5′ -WGWWCW-3′ is shown in Figure 10C. Here the heterocyclic units of ligand dimer are compatible with the central GWWC. The carbonyl of the amide groups is closer

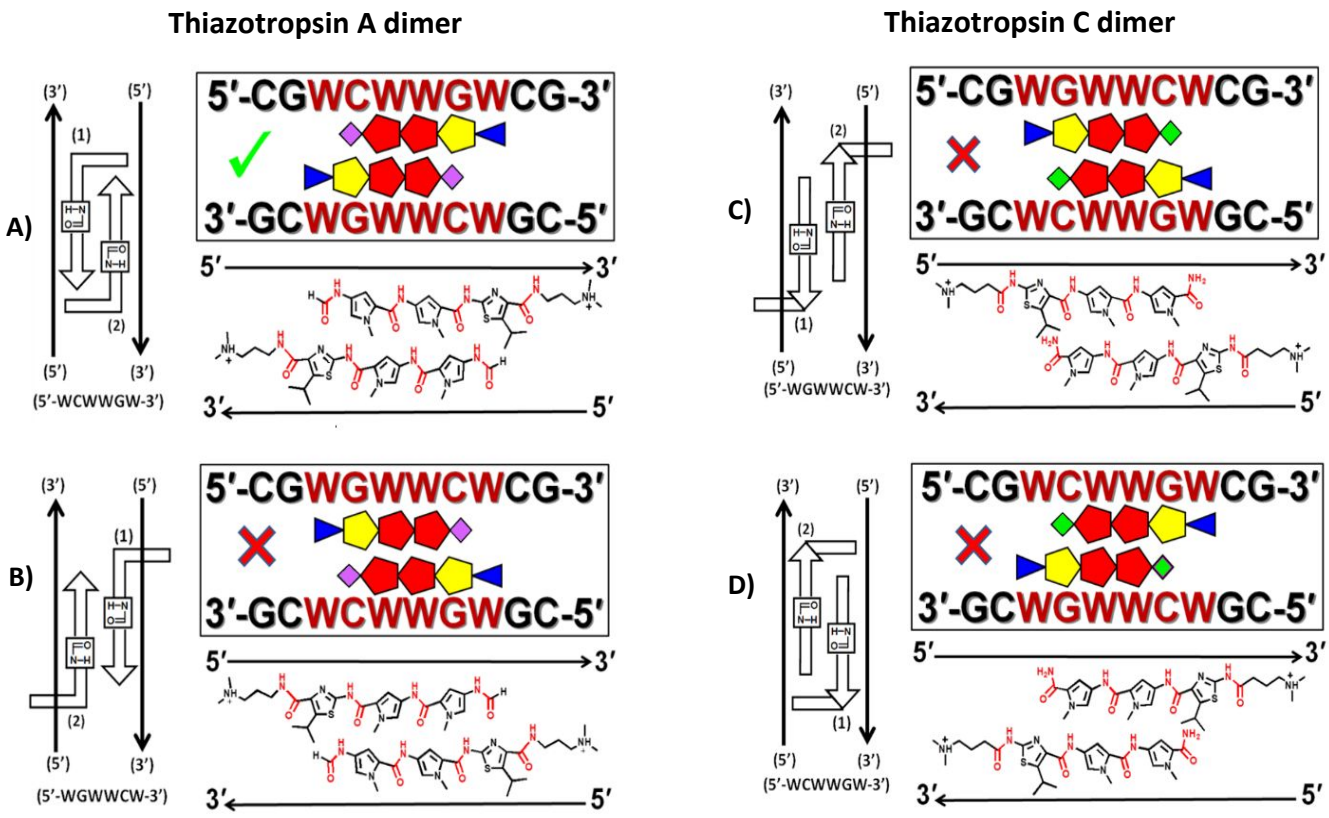


Figure 10 Illustrations of the feasible and unfeasible dimer orientations of thzA and thzC in the minor groove of DNA. **A)** thzA-WCWWGW (confirmed recognition): the DAP group is attached to the N-terminus and points toward the 3'-ends, and the carbonyl of amides is closer to 5'-ends than its partnering NH. **B)** Flipped thiazotropsin A (FthzA)- WGWWCW (no recognition): the DAP group points toward the 5'-ends (steric clash), and the carbonyl of amides is closer to 3'-ends than its partnering NH (disruption of hydrogen bonds). **C)** thzC-WGWWCW (no recognition): the DAP group points toward the 5'-ends (steric clash). **D)** thzC-WCWWGW (no recognition): the carbonyl of amides is closer to 3'-ends than its partnering NH (disruption of hydrogen bonds).

to the 5'-end than its partnering NH, which is similar to the known and confirmed orientation of amide groups in the thzA complex with 5′-WGWWCW-3′ (Figure 10A). However, although this dimer configuration is supposed to enable the amide groups of thzC to form hydrogen bonds with DNA bases, the folded conformation of the DAP group, which is lying toward the 5'-end,

causes significant steric clash with the backbone of DNA thus preventing the recognition of the 5'-WGWWCW-3' sequence (Figure 10 C).

Reversing the position of ligand 1 and 2 of thzC dimer (flipped thiazotropsin C dimer- FthzC), as shown in Figure 10D, is expected to make it compatible with the 5'-WCWWGW-3' sequence. In this dimer configuration, the DAP group is lying toward the 3' end in the middle of minor grooves with no steric clash with the backbone of DNA. However, the inverted configuration of FthzC dimer changes the orientation of amide groups with respect to the 5' and 3' ends, which is expected to disrupt the hydrogen bonds responsible for the ligand -DNA interactions, thus preventing the recognition of 5'-WCWWGW-3' sequence.

3.3 Molecular modeling

3.3.1 Static structural analysis

An insight into the intrinsic atomic structure of a docked complex to depict and analyze its stability can reveal inter- and intramolecular bonding, which contributes to its binding affinity and stability. Hydrogen bonds play a key role in the intermolecular interactions, and in predicting the shapes and properties of biomolecules. UCSF Chimera[36] program was used for the analysis of hydrogen bonds and steric clashes associated with thzA-DNA and FthzA-DNA complexes (Supporting Information, S4) respectively. The formation of different types of intermolecular hydrogen bonds between H-bond donor (D) and acceptor (A) (D...A and D–H...A) are well established. The average distances of hydrogen bond in nucleic acid structures are close to 3.0 Å according to the Watson and Crick pairing model. The strong hydrogen bonds with D...A distance of 2.2–2.5 Å are mostly covalent, 2.5–3.2 Å are moderate and mostly electrostatic, and 3.2–4.0 Å are weak and mainly electrostatic.[37, 38] The availability of a refined NMR structure for thzA association with 5'-ACTAGT-3' (1rmx)[19] facilitated the assessment of potential steric clashes and the formation of

hydrogen bonds in the complexes under investigation. Analysis of hydrogen bonds for thzA and FthzA complexes with DNA were depicted in Table 1 and Figure 11. Ten hydrogen bonds were formed in the thzA-DNA complex, whereas, only two weak hydrogen bonds were observed in the FthzA-DNA complex indicating the disruption of hydrogen bonds in this complex when the carbonyl of amide links of the ligand is closer to the 3'-ends of DNA than its partnering NH (Figure 10 B).

In ligand-bound complexes, the unnatural overlap of any two non-bonding atoms or moieties in a complex structure often leads to steric clashes that may affect the stability of the macromolecule. Such a steric clash was observed in case of FthzA-WGWWTW complex (Table 2), whereas in 1rmx (thzA-ACTAGT complex), no steric clashes were noticed. Both the formyl head and DAP tail of thzA encountered steric clashes with C7-G14 and G4-C17 base pairs regions respectively (Figure 12, Table2) , and this may explain the inability of the flipped thzA dimer to recognize the 5'-WGWWCW-3' sequence when the DAP tail is lying toward the 3'-ends (Figure 10 B).

Table 1: Hydrogen bond analysis using Chimera

H-bonds	Donor	Acceptor	Hydrogen	D..A dist (Å)	D-H..A dist (Å)
1rmx	DG 7.A N2	AIK 11.A N21	DG 7.A H22	2.89	2.084
	AIK 11.A N16	DG 7.A N3	AIK 11.A H16	2.864	1.854
	AIK 11.A N2	DT 5.A O2	AIK 11.A H2	3.437	2.601
	AIK 11.A N26	DT 8.A O2	AIK 11.A H26	2.703	1.912
	AIK 11.A N9	DA 6.A N3	AIK 11.A H9	3.464	2.492
	AIK 2.B N16	DG 17.B N3	AIK 2.B H16	2.97	1.954
	AIK 2.B N2	DT 15.B O2	AIK 2.B H2	2.713	1.701
	AIK 2.B N26	DT 18.B O2	AIK 2.B H26	2.646	1.814
	AIK 2.B N9	DA 16.B N3	AIK 2.B H9	2.888	1.884
	DG 17.B N2	AIK 2.B N21	DG 17.B H22	2.876	1.922
FtzhA	AIK 11.A N2	DT 18.B O2	AIK 11.A H2	3.368	2.596
	AIK 2.B N9	DA 6.A N3	AIK 2.B H9	3.242	2.388

Table 2: Steric clashes in FthzA-DNA complex

Atom1 (ligand)	Atom2 (DNA)	Overlap (Å)	Distance (Å)
AIK 2.B H2	DG 14.B N2	2.106	0.534
AIK 2.B N2	DG 14.B N2	2.012	1.253
AIK 2.B C7	DC 7.A O2	1.675	1.505
AIK 11.A H7	DC 17.B O2	1.289	1.191
AIK 11.A 2H32	DG 14.B C4'	1.203	1.677
AIK 11.A C7	DC 17.B O2	1.181	1.999
AIK 2.B C3	DG 14.B N2	1.16	2.18
AIK 2.B C1	DG 14.B N2	1.159	2.181
AIK 2.B H2	DG 14.B C2	1.136	1.564
AIK 2.B C3	DC 7.A O2	1.075	2.105
AIK 11.A C29	DT 15.B H5'	1.069	1.631
AIK 2.B H7	DC 7.A O2	1.02	1.46
AIK 11.A H2	DG 4.A N2	0.996	1.644
AIK 11.A N2	DG 4.A N2	0.907	2.358
AIK 11.A C32	DG 14.B C4'	0.868	2.712
AIK 2.B N2	DG 14.B C2	0.834	2.491
AIK 11.A 1H28	DG 14.B C1'	0.824	2.056
AIK 11.A 2H32	DG 14.B C5'	0.813	2.067
AIK 11.A C32	DG 14.B C5'	0.796	2.784
AIK 2.B C7	DC 7.A C2	0.792	2.608
AIK 2.B C6	DC 7.A O2	0.759	2.421
AIK 11.A C29	DT 15.B C5'	0.736	2.664
AIK 2.B C28	DT 5.A H5'	0.73	1.97
AIK 2.B 1H28	DT 5.A H5'	0.693	1.307
AIK 11.A C28	DT 15.B H5'	0.652	2.048
AIK 11.A C28	DG 14.B C1'	0.603	2.977
AIK 2.B C7	DG 14.B N2	0.601	2.739

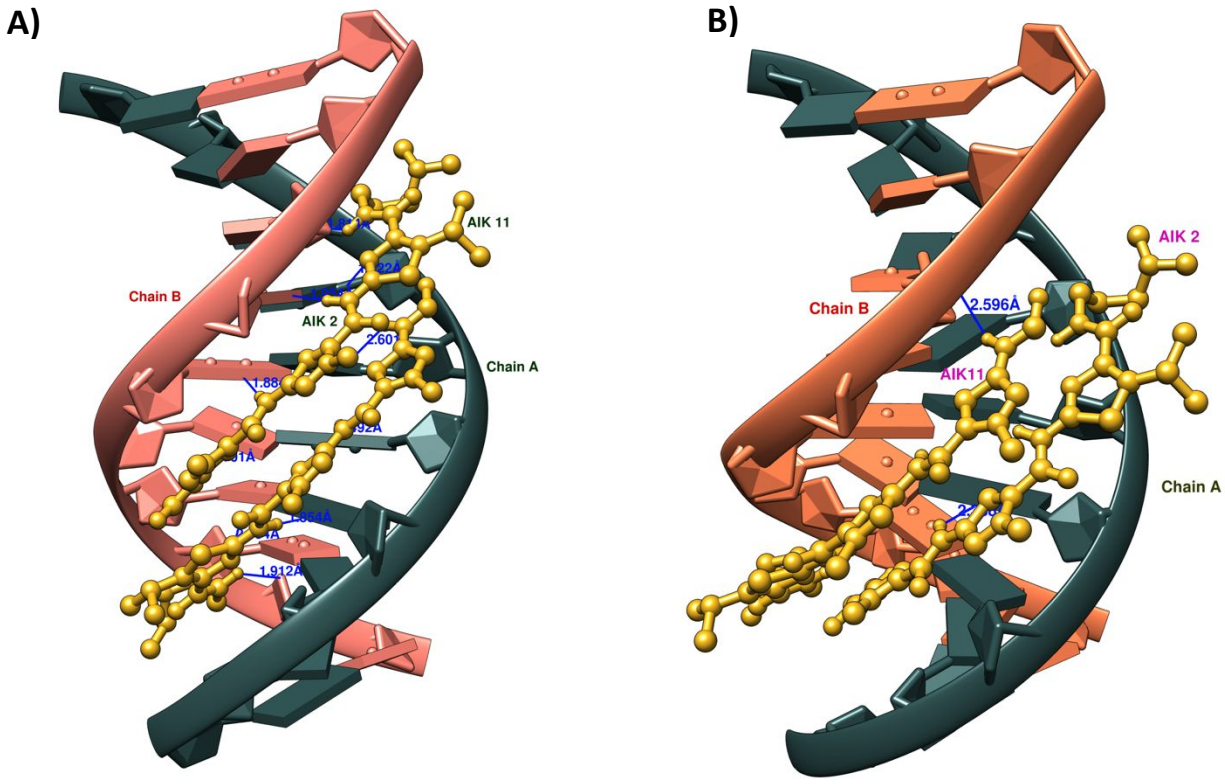


Figure 11: Intermolecular hydrogen bonding analysis. **A)** H bond analysis in thzA-DNA complex (PDB 1RMX).[19]. **B)** H bonds in FthzA-DNA complex (PDB FthzA, Supporting Information).

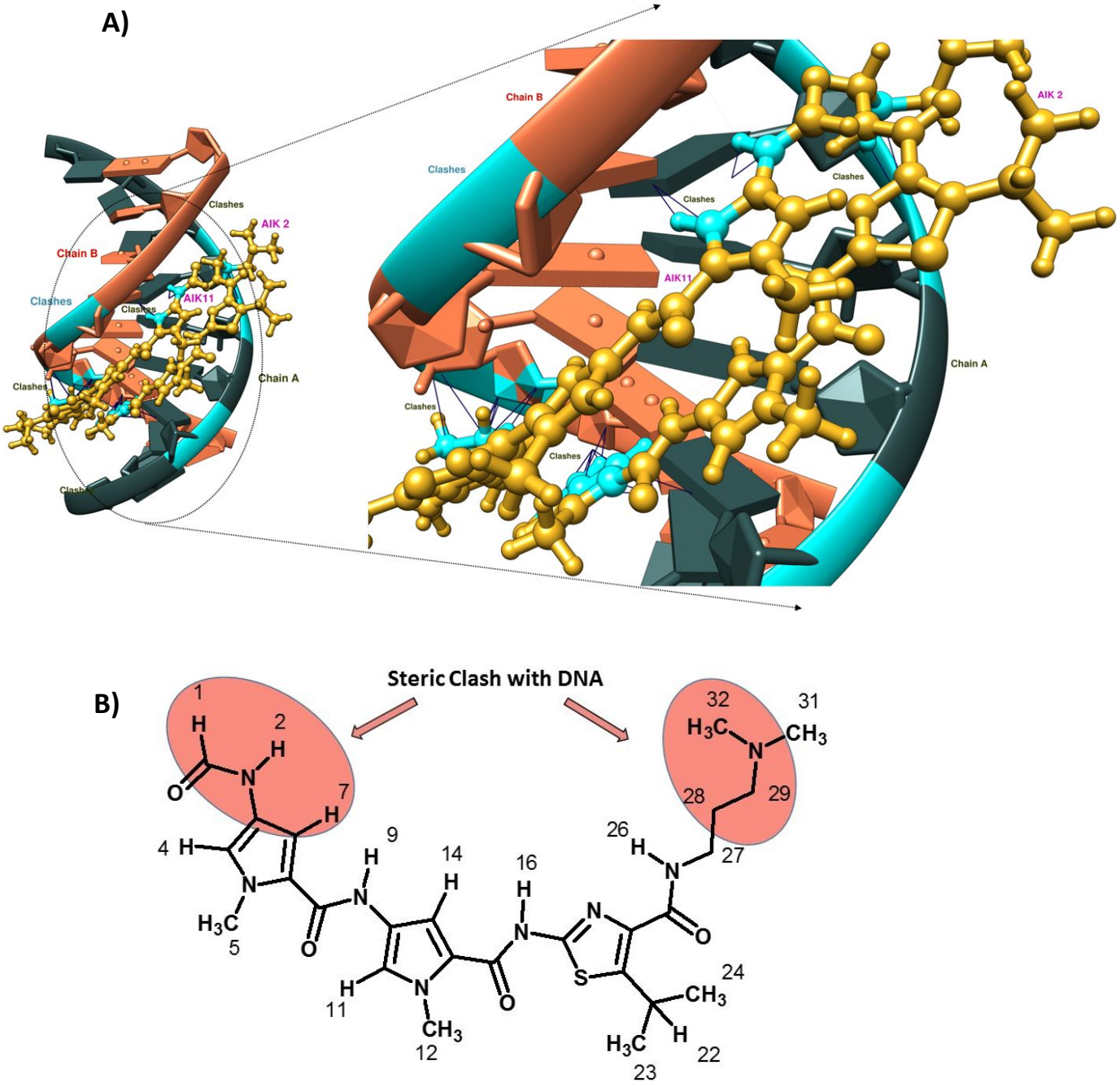


Figure 12: A) Steric clashes in FthzA-DNA complex (PDB 1RMX)[19] **B)** A drawing of thzA structure showing the moieties that encountered a steric clash with DNA and the numbering scheme used for the ^{1}H-atoms of thzA in the PDB file of FthzA-DNA complex (PDB FthzA, Supporting Information).

If we compare the DNA binding results of thzC with thzA, and supposing that the DAP group of thzC adopts a similar conformation to that of thzA, similar explanations can be reached: the inability of thzC to recognize the 5'-WGWWCW-3′ is due to a steric clash with the DAP group when it is attached to the C-terminus and point toward the 5'-ends (Figure 10 C), and the inability of thzC to recognize the 5'-WCWWGW-3′ is due to the disruption of hydrogen bonds with DNA when the carbonyl of the amide links of thzC is closer to the 3'-end than the partnering NH (Figure 10 D).

3.3.2 Stability analysis

To assess the stability of the structures (DNA-ligand), Root Mean Square Deviation (RMSD's) of both complexes were analyzed for the total equilibrated system containing ions and solvent and individual trajectories without solvent and ions. From the stability point of view, the whole system containing solvent, ions, DNA & ligand were observed to be stable during the course of 20 ns simulation and the average RMSDs were obtained for both thzA-DNA and FthzA-DNA complexes. The average RMSDs for the whole system were 3.58 ± 0.1 nm and 3.59 ± 0.1 nm for thzA-DNA & FthzA-DNA respectively (Figure 13A). Variation in RMSDs of individual DNA decamer duplex trajectories, ligand dimers and complex systems (DNA-ligand) were depicted in (Figure 13 B-D). It was evident that when thzA and FthzA dimers were bound to DNA, a significant difference in RMSDs was observed, thus, from the viewpoint of stability, the preferred binding orientation of lexitropsins in the minor grooves of DNA can be established, and that is consistent with experimental results. Based on the MD results, consistent stability was achieved for the preferred orientation of lexitropsin dimers (thzA) in the minor groove of the DNA whereas for the flipped orientation (FthzA), the stability seems to be diverged from the original orientation.

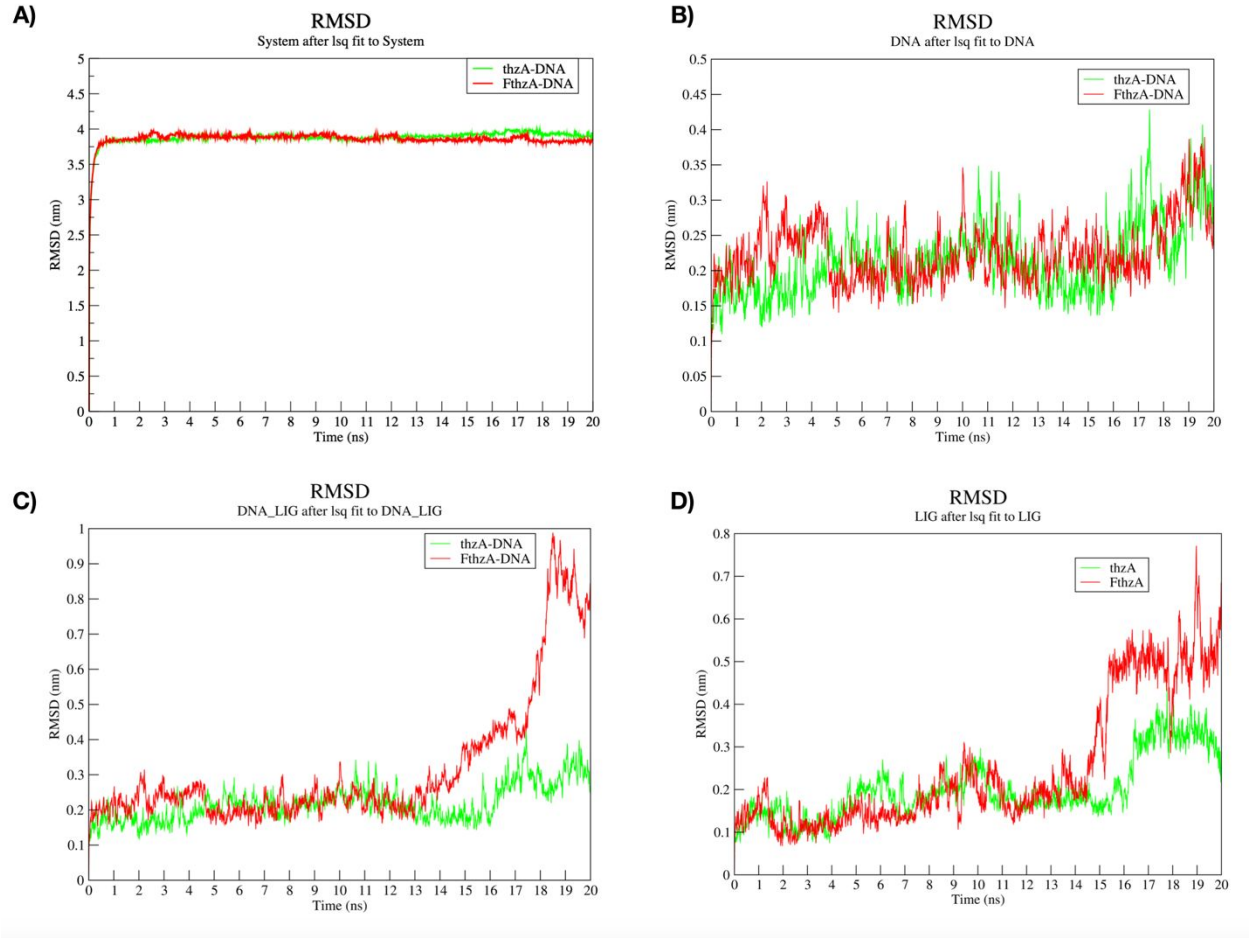


Figure 13: **A)** RMSD for the whole system containing all the trajectories of solvent, ions, DNA and ligand dimers. **B)** the individual trajectories for the DNA decamer duplex. **C)** the complex DNA-ligand dimers. **D)** the individual ligand dimers.

In order to establish the significant conformational rearrangements of thzA and FthzA bound to the DNA duplex, the energy profiles have been obtained during 20 ns simulation. The total energy of the complex systems obtained were averaged around -2.04e+05 ± 0.01 kJ/mol for both the systems, however, a little deviation has been observed in case of FthzA (Figure 14A). Similarly, the potential energy profile of both systems were observed to be -2.46e+05 ± 0.1 kJ/mol (Figure 14B). Root mean square fluctuations (RMSF) were analyzed for thzA-DNA and FthzA-DNA complex along with individual ligand dimers. The greater oscillations were observed for the

complex system FthzA-DNA in comparison to thzA-DNA (Figure 14 C-D), which were noted to be around 1.9 to 2.6 nm and 1.6 to 3.1 nm respectively. When evaluating the individual ligand dimers, larger fluctuations have been observed in case of FthzA ligand dimer. The lower fluctuations, which were seen in the bound complex trajectories, imitate the known stiffening effect of MGB binding to nucleic acids, whilst the larger fluctuations for the free DNA replicate the inherent flexibility, which was observed in the experimental results.[16] The energies that were obtained during the course of 20ns simulations have been tabulated in Table S30 (Supporting information S5). Moreover, ligand-DNA interaction energy calculations have been performed for both systems and the energies obtained were listed in Table S31 (Supporting information S6).

3.3.3 Structural analysis based on molecular dynamics

Hydrogen bond formation between the duplex DNA and both ligand dimers were determined. It was observed that during the course of the simulation, thzA formed consistent H-bonds until 15 ns but reduced thereafter. Similar observations were elucidated in Figure 15A, where FthzA bounded to DNA duplex through H-bonds, but with 5 H-bonds less than thzA, which maintained around 10 H-bonds. The minimum distance between the ligand dimers and the DNA seems to be consistent until 15 ns but after 15 ns, it starts deviating due to the dynamics of DNA duplex, for instance, the groove (minor and major) geometries are highly affected by movement of DNA bases, which may lead to groove compression or expansion via, for instance, a rolling action of the adjacent base pairs.[16] These findings are also clearly apparent when comparing the step parameters of the free DNA duplex. We have also analyzed the solvent accessible surface area and radius of gyration which were depicted in (Figure 15 C-D)

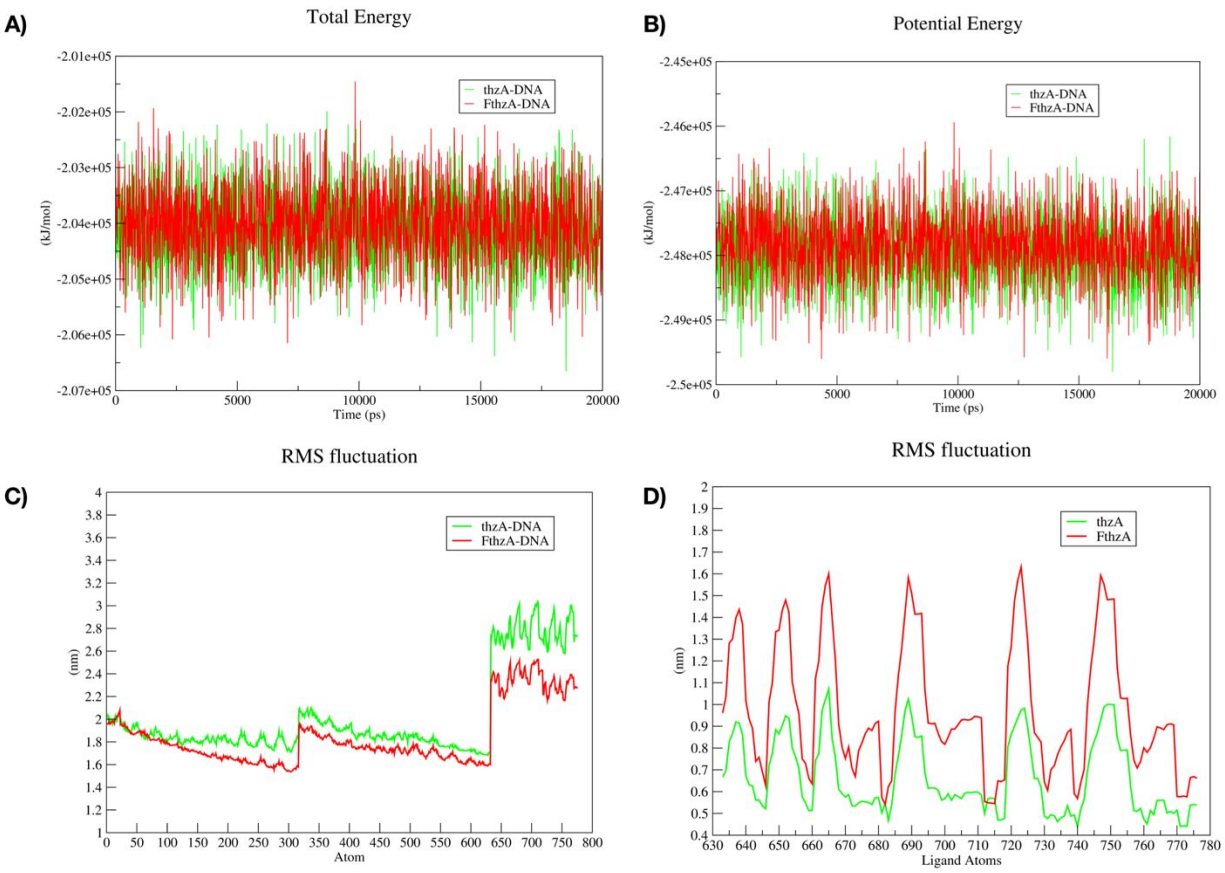


Figure 14: A) Total energy of the system simulated for 20 ns **B)** Potential energy of the system **C)** RMSF for the conjugated DNA-ligand dimer **D)** RMSF for the individual trajectories of the ligand dimers.

The intrinsic molecular levels of the studied macromolecular complexes have been validated using molecular dynamics (MD) simulation, which plays a pivotal role in deciphering the structural properties. The analysis revealed some of the intriguing facts which complemented our experimental results. The experimental analysis revealed that the preferred orientation of lexitropsin dimers in the minor groove of DNA is dictated by the position of the DAP group, which should be attached to the N-terminus of the ligand and always prefers to point towards the 3′-end to avoid steric clashes with the backbone of DNA. According to the molecular dynamics of the ligand bound complex, the clashes or the contacts have been appeared only in FthzA-DNA

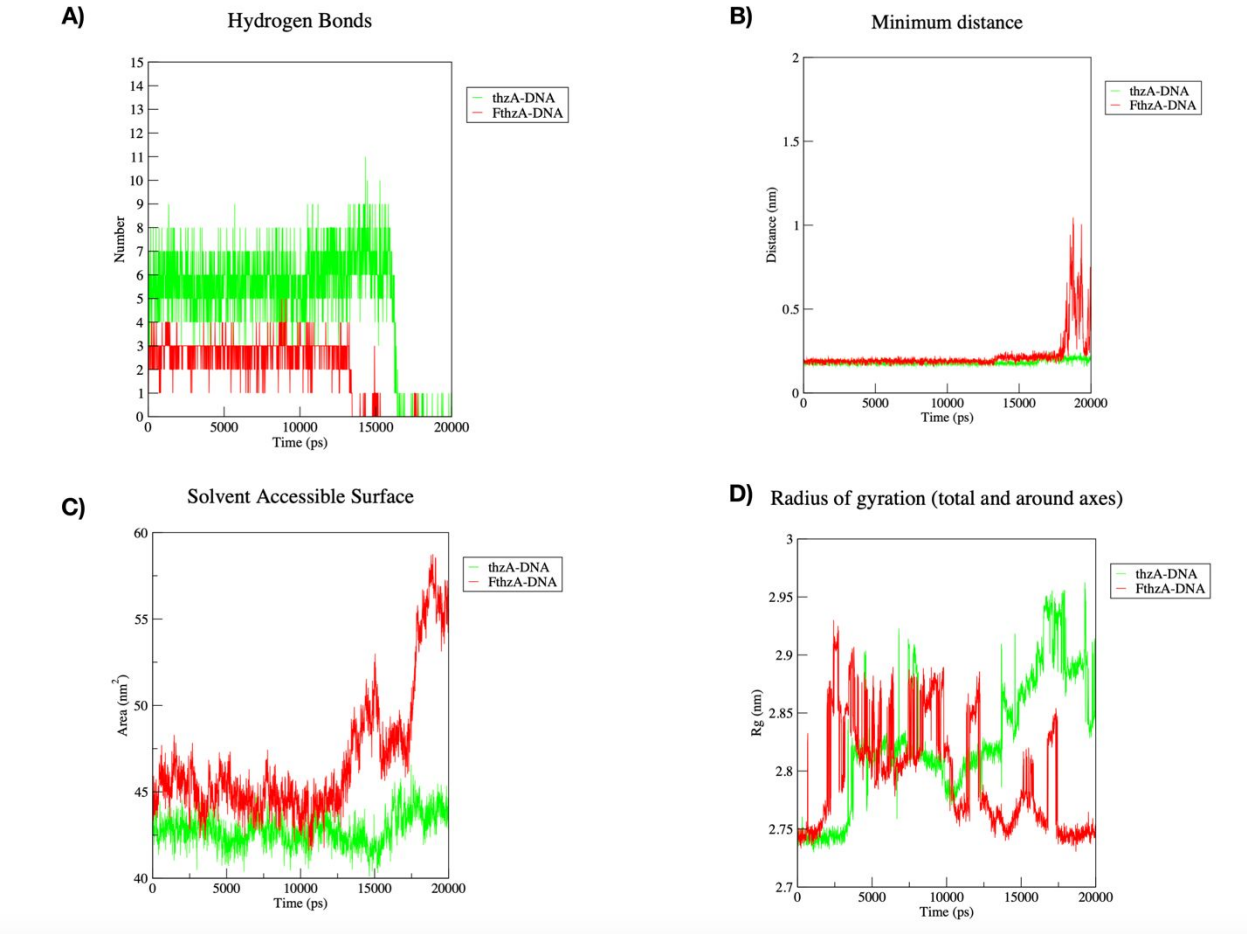


Figure 15: **A)** Hydrogen bonds between the DNA and the ligand dimer during 20ns simulation. **B)** Minimum distance observed between the DNA and the ligand during 20ns simulation **C)** Solvent Accessible Surface Area of the complex systems **D)** Radius of gyration of both complex systems.

complex (Figure 16) but not in thzA complex. Moreover, thzA remains bounded to DNA during 20 ns (Figure 17) whereas FthzA seems to be separated from its respective position as observed in Figure 15A and 15D. The clashes /contacts that were observed in FthzA-DNA complex have been listed in Table S7 (Supporting Information S7), whereas for thzA-DNA complex no clashes were observed. Thus, it can be concluded that FthzA dimer (DAP group attached to C-terminus of the ligand and lying toward the 5'-end) is a non-preferable orientation in the minor groove of DNA and that is mainly due to steric clashes with the backbone of DNA.

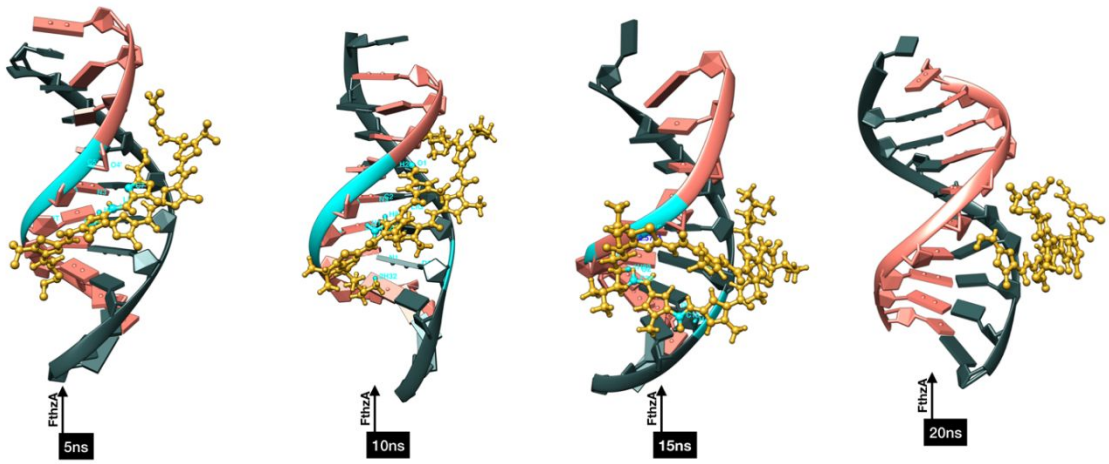


Figure 16: Simulated structures of the complex FthzA-DNA showing the clashes represented in Cyan color. The 20th ns structure seems to be distorted as the FthzA dimer separated from the DNA duplex structure (PDB FthzA, Supporting Information).

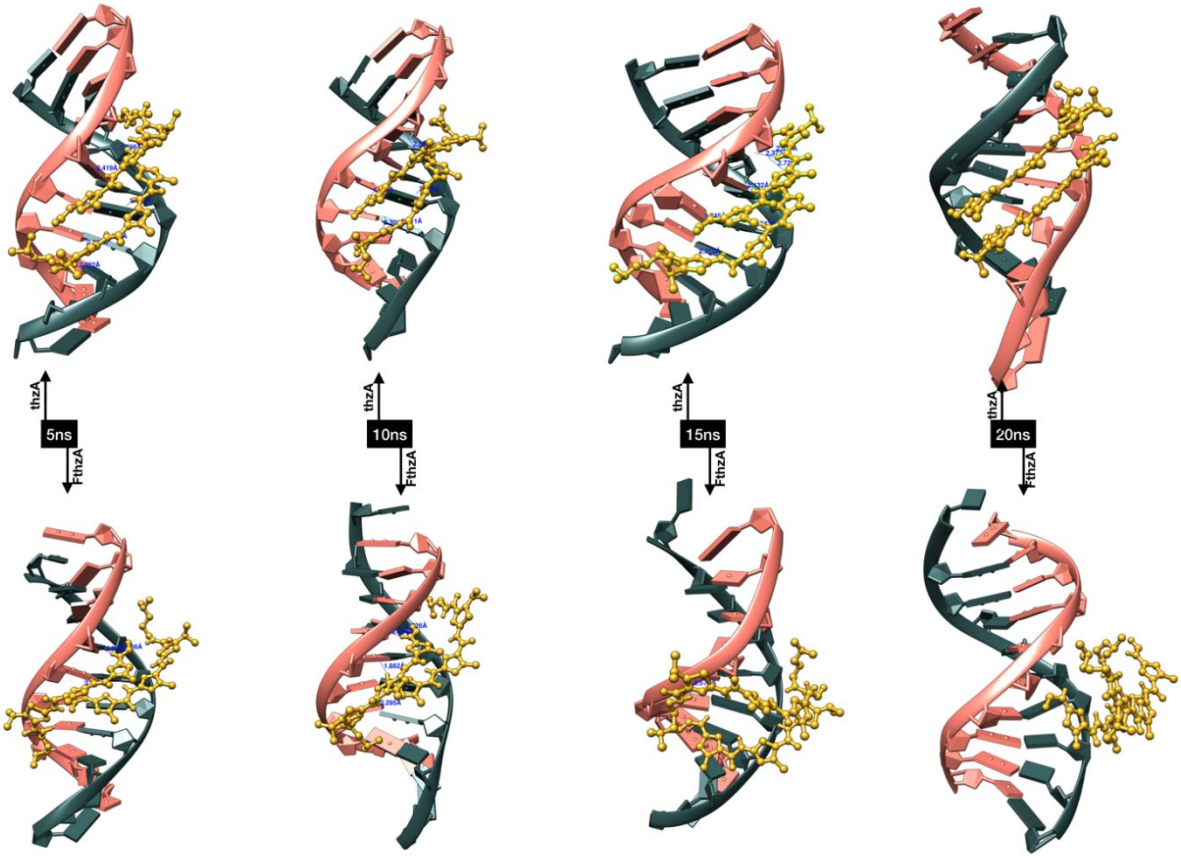


Figure 17: Simulated structures of the complex thzA-DNA (PDB 1rmx) and FthzA-DNA (PDB FthzA, Supporting Information) during the course of 20ns. The upper row depicts the Hydrogen bonds in thzA-DNA retained at intervals e.g. 5ns, 10ns, 15ns, 20ns. The lower row represents the H-bonds retained in FthzA-DNA complex. The structural representations were illustrated using Chimera. Chain A of DNA has been represented in Dark slate gray, Chain B in Salmon and the ligand dimer in Golden color. H bonds have been depicted in blue color.

4. Conclusions

The studies described here provide clear evidence that the preferred orientation of lexitropsin dimers in the minor groove of DNA is dictated by both the position of the DAP group and the orientation of the amide links of the ligand with respects to the 5'-3'-ends. To avoid steric clashes with DNA, the DAP group should be attached to the N-terminus of the ligand and always prefers to point towards the 3′-end. To avoid the disruption of hydrogen bonding between the amide links of the ligand and DNA bases, the ligand should adopt an orientation that keeps the carbonyl group of amid links closer to the 5'-end than the partnering NH.

This study clearly shows the ability of lexitropsin dimers to recognize the DNA minor grooves exclusively in a single orientation with respect to the 5'-3'-ends, and in a sequence-specific manner, which is considered an advantage for improving the DNA binding selectivity. However, reversing the orientation of lexitropsin binding in order to recognize different palindromic DNA sequences might require modifying the ligand structure to improve its flexibility. One such modification could be the incorporation of a β-alanine unit in the place of one of the heterocyclic units (*N*-methylpyrrole) to increase the ligand flexibility, which may help in resetting the curvature of lexitropsin's structure to better match the helical shape of DNA minor grooves and thus avoiding steric clashes with DNA and disruptions of intermolecular hydrogen bonds that drive minor groove recognition. However, although the flexibility of the ligand might allow it to form better interactions with the DNA, it may also lead to multiple inactive ligand conformations, which might decrease sequence specificity and DNA binding affinity, however that is probably dependent on the position of β-alanine substitution and the target DNA sequence.[4] Further investigations in this context are therefore being carried out within our laboratory, and the findings of these investigations will be reported in due course.

5. Experimental section

5.1 General

Solvents and chemical reagents were purchased from Sigma-Aldrich (Steinheim, Germany) and were used without further purification. Both 1H and 13C NMR experiments were carried out on a Bruker 500 MHz instrument. Chemical shifts were measured in parts per million (ppm) relative to the internal tetramethylsilane standard. NMR data were processed by using ACD NMR processor (academic edition). NMR data are presented as follows: s (singlet), bs (broad singlet), d (doublet), dd (double doublet), t (triplet), q (quartet), or m (multiplet). Coupling constants, J, are reported in hertz (Hz). The structures of intermediate compounds and the final structure were confirmed by mass spectroscopy NMR analyses (Supporting Information S8-S31). TLC (Thin-layer chromatography) experiments were carried out using silica gel plates (0.25 mm, E. Merck, 60 F254) and a UV lamp for visualization. Electrospray ionization (ESI) mass spectrometry was performed on a triple-quadrupole tandem mass spectrometer (MicromassW Quattro microTM Waters Corp., Milford, MA, USA) coupled with an electrospray ionization (ESI) interface operated in positive ionization mode for mass spectrometric analysis. HPLC purification was performed on a Shimadzu LC-2040C 3D system (Duisburg, Germany) coupled with a photodiode array detector (PDA), or Agilent 1260 infinity liquid chromatograph (Agilent Technologies, Palo Alto, CA, USA) using Zorbax-SB-C18 column (5 µm; 25 cm × 9.4 mm i.d). The purity of the final tested compound (thzA) was confirmed by HPLC (purity: 99%), Supporting Information S32.

5.2 ITC experiments

For ITC measurements, the oligodeoxynucleotide (ODNs) sequences were purchased from Alpha DNA Ltd. (Montreal, Canada) as HPLC-purified salt-free ODNs, which were used without further purification. Distilled and deionized Millipore-filtered water was used in the preparation of all buffer solutions. PIPES, ACES, EDTA, and NaCl, were purchased from Sigma-Aldrich (Steinheim, Germany) and were used to prepare buffers. DNA solutions were prepared by annealing the ODNs dodecamers at 95 °C for 10 min, then gradually cooling them to room temperature. The ligand and the DNA samples were dissolved in the same batch of the degassed buffer (0.01m PIPES, 0.02m NaCl, 0.001m EDTA, pH 7). The concentrations of the DNA solutions were determined spectroscopically at 260 nm using the OD values supplied by the supplier. For DNA solutions, aliquots of the stock solution were taken and diluted to achieve the concentration required for the ITC experiments (10 mM). The ligand solution was prepared at 0.25 mM. ITC measurements were performed at 25 °C by using a Nano ITC instrument (TA Instruments, New Castle, Delaware, USA). Ligand mixing with DNA was performed by stirring the sample cell at 250 rpm. The ITC unit was interfaced to a PC equipped with the ITCRun and NanoAnalyze software packages for instrument control and data manipulation, respectively. A 50 μL rotating ITC syringe was used to perform 20 injections of the ligand solution (2.5 μL) with a 300-sec delay between injections. Control experiments were carried out under similar conditions using buffer only in the sample cell in order to subtract the heat of ligand dilution from the subsequent heat obtained from ligand-DNA titrations.

5.3 Molecular Modeling

5.3.1 Static structural analysis

Hydrogen bond and steric clashes were analyzed through the UCSF Chimera program for thzA-DNA and FthzA-DNA complexes respectively. The PDB file of FthzA-DNA complex was generated from NMR refined solution structure of thzA-DNA complex (1rmx) by exchanging the position of the ligands (AIK11 & AIK2) and replacing the following DNA bases : C4, G7, C14, and G17 (in the 1rmx) with G4, C7, G14, and C17 respectively to generate the FthzA-DNA complex structure. This was done by using the function, Swapna, "mutates" nucleic acid residues in Chimera.

The relaxed H-bond constrains of 0.4 Angstroms, and 20^{o} were employed that indicated tolerances to Relax constraints that should be applied to meet the precise geometric criteria. Steric clashes were based on VDW radii that can be defined by:

$$\boldsymbol{overlap}_{ij} = \boldsymbol{r}_{VDWi} + \boldsymbol{r}_{VDWj} - \boldsymbol{d}_{ij} - \boldsymbol{allowance}_{ij} .$$

Reasonable cutoff values of 0.4-1.0 Å and allowance values of 0.2-0.6 Å were employed.

5.3.2 Methodology of molecular dynamics simulation

The NMR solution structure of the DNA decamer duplex d(CGACTAGTCG)$_2$ bound to thiazotropsin A dimer (thzA) and Flipped thiazotropsin A dimer (FthzA) were subjected for molecular dynamics simulation using Gromacs v2019.2.[39] The DNA and ligand atoms (thzA & FthzA) were parameterized by CHARMM36[40-42] additive forcefields. CHARMM forcefield for ligand atoms were obtained from Swiss PARAM[43, 44] that provides topology and parameters for small organic molecules compatible with the CHARMM all atoms force field, for use with CHARMM and GROMACS. The complex systems were placed in a periodic cubic box solvated with TIP3P solvent molecules. Periodic boundary conditions with a 15 Å cut-off for non-bonded

interactions were applied, with the particle mesh Ewald (PME)[45-47] method applied to account for the long- range electrostatic interactions. The system was neutralized with 18 Na+ counterions to attain equilibration. Energy minimization and equilibration were carried out in three steps e.g. i) The whole system containing ions, solvent, DNA and ligand were minimized using Verlet cut-off scheme for up to 50,000 steps using steepest-descent algorithm and PME for treatment of long-range electrostatic interactions. ii) Constraints were added to DNA and the ligand dimer for 100ps during heating using NVT ensemble with leap-frog integrator and LINCS holonomic constrains. Modified Berendsen thermostat has been used with Verlet cutoff scheme and Particle Mesh Ewald for long range electrostatics (PME) for equilibration phase 1. iii) NPT ensemble was used at a constant pressure (1 bar) and temperature (300K) for 100ps using a timestep of 2fs for equilibration phase 2 and SHAKE algorithm was used to constraint hydrogen to heavy atom bonds. The MD production phase for both the systems have been simulated for 20 ns each with time step of 2fs. Furthermore, after 20 ns simulation, the DNA-ligand interaction energy was evaluated to compute the non-bonded interaction energy and short-range non-bonded energies, which were quantitatively reproduced with energy profiles generated by Gromacs tools.

5.4 Chemical Synthesis

5.4.1 Preparation of methyl 2-amino-5-isopropyl-1, 3-thiazole-4-carboxylate (1).

Sodium (3.21 g, 139.8 mmol) was dissolved in dry methanol (60 ml) at 0 ºC and was allowed to stir for a further 20 min. The solution was then added dropwise to a solution of isobutyraldehyde (15.0 ml, 165 mmol), and methyl dichloroacetate (22 ml, 212.3 mmol) in ether (60 ml) over 40 min at 0 ºC, and the solution allowed to stir for 20 min. Diethyl ether (60 ml) and brine (60 ml) were then added. The ether layer was collected and dried over anhydrous magnesium sulfate, filtered and the solvent removed under reduced pressure to yield pale-yellow oil. The oil was then

dissolved in a solution of thiourea (9.5 g, 124.8 mmol) in methanol (60 ml), and the solution refluxed for 5 hours. The solvent was then removed under reduced pressure and the crude product taken up in distilled water (50ml), neutralized with ammonium hydroxide solution, and extracted with dichloromethane (2 x 60 ml) to yield the crude product, which was dissolved in 30 mL of methanol and charcoal was added; the mixture was then hot-filtrated to remove charcoal. Excess methanol was removed under reduced pressure, and the product was recrystallized from methanol /water (1:3) to give the desired product as shiny white crystals. Yield (17.5 g, 70%). ^{1}H NMR (methanol-d4) δ 1.15 (6H, d, CH3), 3.72 (3H, s, OCH3), 4.06 (1H, m); ^{13}C NMR δ 23.7, 27.4, 50.6, 133.9, 147.5, 162.6, 165.4; LC-MS (ESI): m/z calcd for $C_8H_{12}N_2O_2S$, 200.0619, found 201.0935 $[M+H]^+$.

5.4.2 Preparation of 2-trichloroacetyl-1-methylpyrrole (2). Trichloroacetylchloride (27.1g, 150.2 mmol) in DCM (120 ml) was placed in a round-bottomed flask at room temperature under nitrogen. A solution of 1-methylpyrrole (12.15 g, 150.3 mmol) in DCM (50 ml) was then added dropwise over two hours. The solution was allowed to stir overnight, and the solvent was then removed under reduced pressure to yield the crude product, which was filtered through a silica column to yield the product as white-gray crystals. Yield (26.1 g, 77 %); ^{1}H NMR ($CDCl_3$) δ 4.00 (3H, s, CH_3), 6.25 (1H, dd, Ar-H), 6.99 (1H, t, Ar-H), 7.53 (1H, dd, Ar-H); ^{13}C NMR (CDCl3): δ 38.56 (CH_3), 96.34 (CCl_3), 108.91 (C), 121.86 (C-H), 124.04 (C-H), 133.64(C-H), 172.92(C=O); LC-MS (ESI): m/z calcd for $C_7H_6Cl_3NO$, 224.9515, found 225.6725 $[M+H]^+$.

5.4.3 Preparation of 2-trichloroacetyl-1-methyl-4-nitropyrrole (3). Nitric acid (70%, v/v, 7.0 ml) was added dropwise to acetic anhydride (50 mL) at -25 ºC and allowed to stir for a further 30 minutes. The solution was then added dropwise to a solution of 2-trichloroacetyl-1-methylpyrrole (**2**) (9.00 g, 40.18 mmol), in acetic anhydride (30 ml) at -25 ºC and allowed to return gradually to

0 ˚C. The solution was cooled to -50 ˚C and water added slowly at which point the product precipitated as an off-white solid. The solid was collected and washed with hexane, before being dried under reduced pressure. Yield (7.6 g, 70%); ^{1}H NMR ($CDCl_3$) δ 4.08 (3H, s, CH_3), 7.78 (1H, d, Ar-H), 7.97.56 (1H, s, Ar-H); ^{13}C NMR (CDCl3): δ 39.78 (CH_3), 94.85 (CCl_3), 117.55 (C), 121.43 (C-H), 130.26 (C-H), 135.35 (C-H), 173.71(C=O); LC-MS (ESI): m/z calcd for $C_7H_5Cl_3N_2O_3$, 269.9366, found 270.9842 $[M+H]^+$.

5.4.4 Preparation of 1-methyl-4-nitro-1H-pyrrole-2-carboxamide (4). 2-trichloroacetyl-1-methyl-4-nitropyrrole (3) (1.0 g, 3.7 mmol) was dissolved in THF (10 mL). Ammonium hydroxide (1M, 10 mL) was then added to this solution and allowed to stir for two hours. During this time, the product precipitated and was filtered using a Buchner funnel to obtain a crude product as an off-white solid. 20 mL of distilled water was added to the crude product, which was extracted with ethyl acetate twice (2x20 mL). The organic solvent was separated and dried over anhydrous magnesium sulfate; ethyl acetate was then removed under reduced pressure, and the product was obtained as a solid white powder. Yield (0.51g, 81.5 %);^{1}H NMR (DMSO-d_6) δ 3.91 (3H, s, CH_3), 7.32 (1H, d, Ar-H), 7.47 (1H, d, Ar-H), 7.80 (1H, bs, N-H), 8.1 (1H, bs, N-H); ^{13}C NMR (DMSO-d_6): δ 37.95, 108.46, 126.62, 128.54, 134.17, 162.11; LC-MS (ESI): m/z calcd for $C_6H_7N_3O_3$, 169.1381, found 170.0726 $[M+H]^+$.

5.4.5 Preparation of N-(5-carbonyl-1-methyl-1H-pyrrol-3-yl)-1-methyl-4-nitro-1H-pyrrole-2-carboxamide (6)

1-Methyl-4-nitro-1H-pyrrole-2-carboxamide (0.51 g, 3.01mmol) was dissolved in methanol (30 mL), and THF (10 mL) and the solution was placed in a 100 mL round-bottomed flask. 0.3 g of Pd/C was then added slowly to this solution at 0 °C. The suspension was then placed under

hydrogen and allowed to stir for 4 hours (until the disappearance of the reactant's TLC spot). The suspension was then filtered through celite (6 g), and the solvent was removed under reduced pressure to yield the product as pale-yellow oil, which was used directly in the next step of synthesis due to the lack of stability for this aminopyrrole product (**5**). The aminopyrrole product (**5**) was then dissolved in THF (20 mL), and triethylamine (1 mL), followed by drop-wise addition of 2-trichloroacetyl-1-methyl-4-nitropyrrole **(3)** (0.81 g, 3.01 mmol) dissolved in DCM (10 mL) and the solution was allowed to stir overnight at 70 ºC. The solvent was then removed under reduced pressure. Water (15 ml) was added, and the crude product was extracted twice with ethyl acetate (2x20 ml). The ethyl acetate layer was dried over magnesium sulfate and then removed under reduced pressure to obtain the product as a solid yellow powder. Yield (0.49 g, 56 %); ^{1}H NMR (DMSO-d6) δ 3.82 (3H, s, CH_3), 3.96 (3H, s, CH_3), 6.84 (1H, d, Ar-H), 7.24 (1H, d, Ar-H), 7.59 (1H, d, Ar-H), 10.24 (1H, s, N-H); ^{13}C NMR (DMSO-d_6): δ 36.63, 37.95, 105.28, 108.04, 118.97, 121.69, 123.29, 126.77, 128.67, 134025, 157.36, 163.42; ; LC-MS (ESI): m/z calcd for $C_{12}H_{13}N_5O_4$, 291.0968., found 292.1059 $[M+H]^+$.

5.4.6 Preparation of methyl 2-(4-(dimethylamino) butanamido)-5-isopropylthiazole-4-carboxylate (10). 4-(Dimethylamino)butanoic acid (1.00 g, 7.62mmol) was placed in a round-bottomed flask (100 ml), and a solution of oxalyl chloride (2 ml, 4 mmol), DMF (0.6 ml) in DCM (20 ml) was added to the flask. The reaction mixture was then refluxed for 4 hours. DCM solvent was then removed under vacuum. The reaction flask was placed in an ice bath, and a solution of triethylamine TEA (1.6 ml, 11.0 mmol), DCM (20 ml) and isopropylthiazole (1.2 g, 5.99 mmol) were added and left stirring overnight. DCM solvent was removed under vacuum, and the reaction mixture was then made alkaline by adding a solution of $NaHCO_3$, then extracted with ethyl acetate (2x40 ml). The solution was dried over anhydrous magnesium sulfate; ethyl acetate

was then removed under reduced pressure, and the crude product was obtained as a pale-yellow brownish oil. The crude product was purified on a silica column using ethyl acetate as eluting solvent. Yield (1.3 g, 70%); ^{1}H NMR ($CDCl_3$) δ 1.34 (6H, d, CH_3), 2.3 (2H, m, CH_2), 2.89 (2H t, CH_2), 2.93 (6H, d, N-CH_3), 3.28 (2H t, CH_2), 3.88 (3H, s, OCH_3), 4.03 (1H, m, CH), 11.95 (1H, bs, N-H); ^{13}C NMR ($CDCl_3$) δ 21.05, 24.95, 27.86, 34.84, 40.72, 51.83, 60.39, 135.37, 153.72, 155.39, 163.15, 170.35; LC-MS (ESI): m/z calcd for $C_{14}H_{23}N_3O_3S$, 313.1460, found 314.911 [M +H]$^{+}$.

5.4.7 Preparation of 2-(4-(dimethylamino) butanamido)-5- isopropylthiazole-4-carboxyli acid (11). A solution of water (20 ml), lithium hydroxide (0.0378 g, 1.58 mmol) and tetrahydrofurane (5 ml) were added to the methyl 2-(4(dimethylamino) butanamido)-5-isopropylthiazole-4-carboxylate (0.5g) prepared previously; the reaction mixture was then stirred for 3 h at room temperature. The organic solvent THF was removed under reduced pressure using rotary evaporator. The aqueous reaction mixture was acidified by adding a dilute solution of HCl to reach pH 3-4.5. Part of the product was precipitated and collected by filtration as brown solid. The rest of the product remained soluble in the aqueous solvent, which was removed under vacuum, and the crude product was purified using reverse-phase chromatography and methanol/water (30:70 %) as eluting solvent. Yield(0.4g, 80%); ^{1}H NMR (DMSO-d) δ 1.34 (6H, d, CH_3), 2.30 (2H, m, CH_2), 2.87 (2H t, CH_2), 2.93 (6H, s, N-CH_3), 3.28 (2H, t, CH_2), 4.03 (1H, m, CH), 11.95 (1H, bs, N-H); ^{13}C NMR (DMSO-d) δ 21.05, 24.95, 27.86, 34.84, 40.72, 51.83, 60.39, 135.37, 153.37, 153.72, 155.72, 163.15, 170.35; LC-MS (ESI): m/z calcd for $C_{13}H_{21}N_3O_3S$, 299.1304, found 300.2766 [M +H]$^{+}$.

5.4.8 Preparation of N-(5-(5-carbonyl-1-methyl-1H-pyrrol-3-ylcarbomyl)-1-methyl-1H-pyrrol-3-yl)-2-(4-(dimethylamino)butanamid)-5-isopropylthiazole-4-carboxamide (12)

N-(5-carbonyl-1-methyl-1H-pyrrol-3-yl)-1-methyl-4-nitro-1H-pyrrole-2-carboxamide (**6**) (0.35g, 1.2 mmoles) was dissolved in methanol (15 ml) and THF (5 mL) and the solution was placed in a 100 mL round-bottomed flask. 0.2 g of palladium (Pd/C) was then added gradually to the solution at 0 C°. The suspension was placed under hydrogen and allowed to stir for 4 hours. The suspension was then filtered through celite (6 g), and the solvent removed under reduced pressure to yield the amino product (**7**) as a pale-yellow oil that was used directly in the next step of synthesis due to the lack of stability of this compound. Compound **7** was then dissolved in DMF (2 ml) and added slowly to a solution containing compound **11** (0.359g, 1.2 mmoles), diisopropylethylamine (3.6 mmoles, 0.465 g) and HBTU (1.5 mmol, 0.569 g) dissolved in 4 ml DMF. This solution was allowed to stir overnight under nitrogen at room temperature. The reaction was then quenched by addition of 20 mL sodium bicarbonate (1 M), and the product was extracted with ethyl acetate (2x20 ml). The crude final product was then purified by semi-preparative HPLC. The purity of the final product was confirmed by HPLC (purity: 99%), Supporting Information S32. Yield (0.1 g, 15.3 %); ^{1}H NMR (DMSO-d_6) δ 1.28 (6H, d), 1.94 (2H, m), 2.55 (2H, t), 2.79 (6H, d), 3.09 (2H, m), 3.80 (3H, s), 3.85 (3H, s), 4.19 (1H, m), 6.84 (1H, d, J = 1.8 Hz), 7.08 (1H, d, J = 1.8 Hz), 7.20 (1H, d, J = 1.8 Hz), 7.27 (1H, d, J = 1.8 Hz), 9.64 (1H, s), 9.86 (1H, s), 12.20 (1H, s); ^{13}C NMR (DMSO-d_6): δ 19.44, 24.81, 26.48, 31.59, 36.17, 36.28, 42.38, 56.11, 104.66, 105.19, 118.52, 121.34, 121.94, 122.65, 123.19, 136.18, 145.76, 153.35, 158.33, 158.49, 159.26, 163.21, 170.55; LC-MS (ESI): m/z calcd for $C_{25}H_{34}N_8O_4S$, 542.2424, found 543.2804.

ASSOCIATED CONTENT

Supporting information

The Supporting Information is available free of charge on the ACS Publications website at DOI: Mechanism of Darzens reaction; ITC measurements; structures of thzA-DNA and FthzA-DNA complexes; steric clashes and the energies obtained during the 20 ns simulation; NMR spectra; mass spectra; HPLC purity of thiazotropsin C (PDF).

Molecular formula strings (CSV).

FthzA (PDB).

AUTHOR INFORMATION

Corresponding author

*** E-mail: halniss@sharjah.ac.ae, Phone: +(971)-6-5057427.**

ORCID

Hasan Alniss: 0000-0001-8639-9531

Ini-Isabée Witzel: 0000-0001-9675-1877

Mohammad Semreen: 0000-0002-0169-7538

Pritam K. Panda: 0000-0003-4879-2302

Yogendra Kumar Mishra: 0000-0002-8786-9379

Rajeev Ahuja: 0000-0003-1231-9994

John Parkinson: 0000-0003-4270-6135

Notes

The authors declare no competing financial interest. All authors have given approval to the final version of the manuscript.

ACKNOWLEDGMENTS

This work was supported by a grant from the Research Funding Department, University of Sharjah-UAE (grant number: 1601110113-P). Swedish Research Council (VR) is acknowledged for financial support. SNIC and HPC2N are acknowledged for providing the computing facilities.

ABBREVIATIONS USED

MGBs, minor groove binders; ODNs, oligodeoxynucleotides; ITC, isothermal titration calorimetry; thzA, thiazotropsin A; thzC, thiazotropsin C; ΔG, Gibbs free energy; TΔS, entropy changes; ΔH, enthalpy changes; *K*, binding constant; Thz, isopropylthiazole; Pyr, N-methylpyrrole; DAP, dimethylaminopropyl; ACES, 2-(carbamoylmethylamino)ethanesulfonic acid; EDTA, ethylenediaminetetraacetic acid; PIPES, 1,4-piperazinediethane-sulfonic acid; A, adenine; T, thymine; G, guanine; C, cytocine; W, adenine or thymine; ARE, androgen response element; Py, N-methylpyrrole; AR, androgen receptor; PR, progesterone receptor; GR, glucocorticoid receptor; MR, mineralocorticoid receptor; DCM, dichloromethane; DMAD, DMF, dimethylformamide; SAR, structure-activity relationship; TEA, triethylamine; THF, tetrahydrofuran.

REFERENCES

1. Chaires, J. B.; Leng, F.; Przewloka, T.; Fokt, I.; Ling, Y. H.; Perez-Soler, R.; Priebe, W. Structure-Based Design of a New Bisintercalating Anthracycline Antibiotic. *J Med Chem* **1997,** 40, 261-266.

2. Scott, F. J.; Puig-Sellart, M.; Khalaf, A. I.; Henderson, C. J.; Westrop, G.; Watson, D. G.; Carter, K.; Grant, M. H.; Suckling, C. J. An Evaluation of Minor Groove Binders as Anti-Lung Cancer Therapeutics. *Bioorg Med Chem Lett* **2016,** 26, 3478-3486.

3. Barrett, M. P.; Gemmell, C. G.; Suckling, C. J. Minor Groove Binders as Anti-Infective Agents. *Pharmacology & therapeutics* **2013,** 139, 12-23.

4. Alniss, H. Y. Thermodynamics of DNA Minor Groove Binders. *J Med Chem* **2019,** 62, 385-402.

5. Kurmis, A. A.; Yang, F.; Welch, T. R.; Nickols, N. G.; Dervan, P. B. A Pyrrole-Imidazole Polyamide Is Active against Enzalutamide-Resistant Prostate Cancer. *Cancer Res* **2017,** 77, 2207-2212.

6. Ridge, S. M.; Bhattacharyya, D.; Dervan, E.; Naicker, S. D.; Burke, A. J.; Murphy, J. M.; O'Leary, K.; Greene, J.; Ryan, A. E.; Sullivan, F. J.; Glynn, S. A. Secreted Factors from Metastatic Prostate Cancer Cells Stimulate Mesenchymal Stem Cell Transition to a Pro-Tumourigenic 'Activated' State That Enhances Prostate Cancer Cell Migration. *Int J Cancer* **2018,** 142, 2056-2067.

7. Raskatov, J. A.; Szablowski, J. O.; Dervan, P. B. Tumor Xenograft Uptake of a Pyrrole-Imidazole (Py-Im) Polyamide Varies as a Function of Cell Line Grafted. *J Med Chem* **2014,** 57, 8471-8476.

8. Pelton, J. G.; Wemmer, D. E. Structural Characterization of a 2:1 Distamycin A.D(Cgcaaattggc) Complex by Two-Dimensional Nmr. *Proc Natl Acad Sci U S A* **1989,** 86, 5723-5727.

9. Geierstanger, B. H.; Jacobsen, J. P.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. Structural and Dynamic Characterization of the Heterodimeric and Homodimeric Complexes of Distamycin and 1-Methylimidazole-2-Carboxamide-Netropsin Bound to the Minor Groove of DNA. *Biochemistry* **1994,** 33, 3055-3062.

10. Youngquist, R. S.; Dervan, P. B. Sequence-Specific Recognition of B-DNA by Oligo(N-Methylpyrrolecarboxamide)S. *Proc Natl Acad Sci U S A* **1985,** 82, 2565-2569.

11. Mrksich, M.; Wade, W. S.; Dwyer, T. J.; Geierstanger, B. H.; Wemmer, D. E.; Dervan, P. B. Antiparallel Side-by-Side Dimeric Motif for Sequence-Specific Recognition in the Minor Groove of DNA by the Designed Peptide 1-Methylimidazole-2-Carboxamide Netropsin. *Proc Natl Acad Sci U S A* **1992,** 89, 7586-7590.

12. White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B. Recognition of the Four Watson-Crick Base Pairs in the DNA Minor Groove by Synthetic Ligands. *Nature* **1998,** 391, 468-471.

13. Kostjukov, V. V.; Khomytova, N. M.; Davies, D. B.; Evstigneev, M. P. Electrostatic Contribution to the Energy of Binding of Aromatic Ligands with DNA. *Biopolymers* **2008,** 89, 680-690.

14. Dervan, P. B.; Burli, R. W. Sequence-Specific DNA Recognition by Polyamides. *Current opinion in chemical biology* **1999,** 3, 688-693.

15. Alniss, H. Y.; Salvia, M. V.; Sadikov, M.; Golovchenko, I.; Anthony, N. G.; Khalaf, A. I.; MacKay, S. P.; Suckling, C. J.; Parkinson, J. A. Recognition of the DNA Minor Groove by Thiazotropsin Analogues. *Chembiochem : a European journal of chemical biology* **2014,** 15, 1978-1990.

16. Treesuwan, W.; Wittayanarakul, K.; Anthony, N. G.; Huchet, G.; Alniss, H.; Hannongbua, S.; Khalaf, A. I.; Suckling, C. J.; Parkinson, J. A.; Mackay, S. P. A Detailed Binding Free Energy Study of 2:1 Ligand-DNA Complex Formation by Experiment and Simulation. *Physical chemistry chemical physics : PCCP* **2009,** 11, 10682-10693.

17. Alniss, H. Y.; Anthony, N. G.; Khalaf, A. I.; MacKay, S. P.; Suckling, C. J.; Waigh, R. D.; Wheate, N. J.; Parkinson, J. A. Rationalizing Sequence Selection by Ligand Assemblies in the DNA Minor Groove: The Case for Thiazotropsin A. *Chemical Science* **2012,** 3, 711-722.

18. Dervan, P. B. Design of Sequence-Specific DNA-Binding Molecules. *Science* **1986,** 232, 464-471.

19. Anthony, N. G.; Johnston, B. F.; Khalaf, A. I.; MacKay, S. P.; Parkinson, J. A.; Suckling, C. J.; Waigh, R. D. Short Lexitropsin That Recognizes the DNA Minor Groove at 5'-Actagt-3':

Understanding the Role of Isopropyl-Thiazole. *Journal of the American Chemical Society* **2004,** 126, 11338-11349.

20. Culig, Z.; Klocker, H.; Bartsch, G.; Hobisch, A. Androgen Receptors in Prostate Cancer. *Endocr Relat Cancer* **2002,** 9, 155-170.

21. Shaffer, P. L.; Jivan, A.; Dollins, D. E.; Claessens, F.; Gewirth, D. T. Structural Basis of Androgen Receptor Binding to Selective Androgen Response Elements. *Proc Natl Acad Sci U S A* **2004,** 101, 4758-4763.

22. Kurmis, A. A.; Dervan, P. B. Sequence Specific Suppression of Androgen Receptor-DNA Binding in Vivo by a Py-Im Polyamide. *Nucleic Acids Res* **2019,** 47, 3828-3835.

23. Hwang, D. J.; He, Y.; Ponnusamy, S.; Mohler, M. L.; Thiyagarajan, T.; McEwan, I. J.; Narayanan, R.; Miller, D. D. New Generation of Selective Androgen Receptor Degraders: Our Initial Design, Synthesis, and Biological Evaluation of New Compounds with Enzalutamide-Resistant Prostate Cancer Activity. *J Med Chem* **2019,** 62, 491-511.

24. Chenoweth, D. M.; Dervan, P. B. Structural Basis for Cyclic Py-Im Polyamide Allosteric Inhibition of Nuclear Receptor Binding. *J Am Chem Soc* **2010,** 132, 14521-14529.

25. Chenoweth, D. M.; Harki, D. A.; Phillips, J. W.; Dose, C.; Dervan, P. B. Cyclic Pyrrole-Imidazole Polyamides Targeted to the Androgen Response Element. *J Am Chem Soc* **2009,** 131, 7182-7188.

26. Wittayanarakul, K.; Anthony, N. G.; Treesuwan, W.; Hannongbua, S.; Alniss, H.; Khalaf, A. I.; Suckling, C. J.; Parkinson, J. A.; Mackay, S. P. Ranking Ligand Affinity for the DNA Minor Groove by Experiment and Simulation. *ACS medicinal chemistry letters* **2010,** 1, 376-380.

27. Salvia, M. V.; Addison, F.; Alniss, H. Y.; Buurma, N. J.; Khalaf, A. I.; Mackay, S. P.; Anthony, N. G.; Suckling, C. J.; Evstigneev, M. P.; Santiago, A. H.; Waigh, R. D.; Parkinson, J. A. Thiazotropsin Aggregation and Its Relationship to Molecular Recognition in the DNA Minor Groove. *Biophysical chemistry* **2013,** 179, 1-11.

28. Dervan, P. B. Molecular Recognition of DNA by Small Molecules. *Bioorg Med Chem* **2001,** 9, 2215-2235.

29. Khalaf, A. I.; Waigh, R. D.; Drummond, A. J.; Pringle, B.; McGroarty, I.; Skellern, G. G.; Suckling, C. J. Distamycin Analogues with Enhanced Lipophilicity: Synthesis and Antimicrobial Activity. *J Med Chem* **2004,** 47, 2133-2156.

30. Liu, Y.; Provencher, B. A.; Bartleson, K. J.; Deng, L. Highly Enantioselective Asymmetric Darzens Reactions with a Phase Transfer Catalyst. *Chem Sci* **2011,** 2, 1301-1304.

31. Grazi, E.; Rowley, P. T.; Cheng, T.; Tchola, O.; Horecker, B. L. The Mechanism of Action of Aldolases. Iii. Schiff Base Formation with Lysine. *Biochem Biophys Res Commun* **1962,** 9, 38-43.

32. Ababou, A.; Ladbury, J. E. Survey of the Year 2004: Literature on Applications of Isothermal Titration Calorimetry. *J Mol Recognit* **2006,** 19, 79-89.

33. Cliff, M. J.; Gutierrez, A.; Ladbury, J. E. A Survey of the Year 2003 Literature on Applications of Isothermal Titration Calorimetry. *J Mol Recognit* **2004,** 17, 513-523.

34. Cliff, M. J.; Ladbury, J. E. A Survey of the Year 2002 Literature on Applications of Isothermal Titration Calorimetry. *J Mol Recognit* **2003,** 16, 383-391.

35. Alniss, H.; Zamiri, B.; Khalaj, M.; Pearson, C. E.; Macgregor, R. B., Jr. Thermodynamic and Spectroscopic Investigations of Tmpyp4 Association with Guanine- and Cytosine-Rich DNA and Rna Repeats of C9orf72. *Biochemical and biophysical research communications* **2018,** 495, 2410-2417.

36. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. Ucsf Chimera--a Visualization System for Exploratory Research and Analysis. *J Comput Chem* **2004,** 25, 1605-1612.

37. Jeffrey, G. A. *An Introduction to Hydrogen Bonding*. Oxford University Press: 1997.

38. Soni, N.; Swain, S. K.; Kant, R.; Singh, A.; Ravichandran, R.; Verma, S. K.; Panda, P. K.; Suar, M. Landscape of Rod9 Island: Functional Annotations and Biological Network of Hypothetical Proteins in Salmonella Enterica. *Comput Biol Chem* **2019,** 83, 107110.

39. Abraham, M.; Van Der Spoel, D.; Lindahl, E.; Hess, B.; Spoel, D. v. d.; Lindahl, E. *Gromacs User Manual Version 5.0.4.* ; 2014.

40. Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmuller, H.; MacKerell, A. D., Jr. Charmm36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nature methods* **2017,** 14, 71-73.

41. Lee, S.; Tran, A.; Allsopp, M.; Lim, J. B.; Henin, J.; Klauda, J. B. Charmm36 United Atom Chain Model for Lipids and Surfactants. *The journal of physical chemistry. B* **2014,** 118, 547-556.

42. Huang, J.; MacKerell, A. D., Jr. Charmm36 All-Atom Additive Protein Force Field: Validation Based on Comparison to Nmr Data. *Journal of computational chemistry* **2013,** 34, 2135-2145.

43. Zoete, V.; Cuendet, M. A.; Grosdidier, A.; Michielin, O. Swissparam: A Fast Force Field Generation Tool for Small Organic Molecules. *Journal of computational chemistry* **2011,** 32, 2359-2368.

44. Knight, J. L.; Yesselman, J. D.; Brooks, C. L., 3rd. Assessing the Quality of Absolute Hydration Free Energies among Charmm-Compatible Ligand Parameterization Schemes. *Journal of computational chemistry* **2013,** 34, 893-903.

45. Cerutti, D. S.; Duke, R. E.; Darden, T. A.; Lybrand, T. P. Staggered Mesh Ewald: An Extension of the Smooth Particle-Mesh Ewald Method Adding Great Versatility. *Journal of chemical theory and computation* **2009,** 5, 2322.

46. Darden, T.; Perera, L.; Li, L.; Pedersen, L. New Tricks for Modelers from the Crystallography Toolkit: The Particle Mesh Ewald Algorithm and Its Use in Nucleic Acid Simulations. *Structure* **1999,** 7, R55-60.

47. Perera, L.; Li, L.; Darden, T.; Monroe, D. M.; Pedersen, L. G. Prediction of Solution Structures of the Ca2+-Bound Gamma-Carboxyglutamic Acid Domains of Protein S and Homolog Growth Arrest Specific Protein 6: Use of the Particle Mesh Ewald Method. *Biophysical journal* **1997,** 73, 1847-1856.

Table of Contents graphic: