

1 **Spectroscopic, biochemical and computational studies of bioactive DNA**
2 **minor groove binders targeting 5'-WGWWCW-3' motif**

3 Hasan Y. Alniss*^{1,3}, Bryony M. Kemp², Elizabeth Holmes², Joanna Hoffmann², Rafal M.
4 Ploch², Wafaa S. Ramadan³, Yousef A. Msallam^{1,3}, Hadeel M. Al-Jubeh³, Moustafa M.
5 Madkour^{1,3}, Bekir C. Celikkaya², Fraser J. Scott², Raafat El-Awady^{1,3} and John A.
6 Parkinson*²

7
8
9 ¹College of Pharmacy, Department of Medicinal Chemistry, University of Sharjah, Sharjah,
10 27272, UAE; ²Department of Pure and Applied Chemistry, University of Strathclyde, 295
11 Cathedral Street, Glasgow, G1 1XL, Scotland, UK; ³Research Institute for Medical and
12 Health Sciences, University of Sharjah, Sharjah, 27272, UAE.

13
14
15
16
17
18
19 *Corresponding Addresses:

20 -Hasan Y. Alniss, College of Pharmacy, University of Sharjah, Sharjah, 27272, United Arab
21 Emirates Arab Emirate. Tel: +(971)-6-5057427. E-mail: halniss@sharjah.ac.ae.

22
23 -John A. Parkinson Department of Pure and Applied Chemistry, University of Strathclyde,
24 295 Cathedral Street, Glasgow, G1 1XL, Scotland, UK john.parkinson@strath.ac.uk

31 **Abstract**

32 Spectroscopic, biochemical, and computational modelling studies have been used to assess the
33 binding capability of a set of minor groove binding (MGB) ligands against the self-
34 complementary DNA sequences 5'-d(CGCACTAGTGCG)-3' and 5'-
35 d(CGCAGTACTGCG)-3'. The ligands were carefully designed to target the DNA response
36 element, 5'-WGWWCW-3', the binding site for several nuclear receptors. Basic 1D ¹H NMR
37 spectra of the DNA samples prepared with three MGB ligands show subtle variations
38 suggestive of how each ligand associates with the double helical structure of both DNA
39 sequences. The variations among the investigated ligands were reflected in the line shape and
40 intensity of 1D ¹H and ³¹P-¹H NMR spectra. Rapid visual inspection of these 1D NMR
41 spectra proves to be beneficial in providing valuable insights on MGB binding molecules. The
42 NMR results were consistent with the findings from both UV DNA denaturation and molecular
43 modelling studies. Both the NMR spectroscopic and computational analyses indicate that the
44 investigated ligands bind to the minor grooves as antiparallel side-by-side dimers in a head-to-
45 tail fashion. Moreover, comparisons with results from biochemical studies offered valuable
46 insights into the mechanism of action, and antitumor activity of MGBs in relation to their
47 structures, essential pre-requisites for future optimization of MGBs as therapeutic agents.

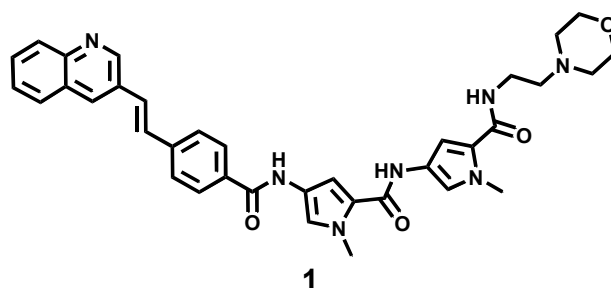
48

49 **Key Words**

50 NMR Spectroscopy, Nucleic Acids, Molecular Modelling, Minor Groove Binders (MGBs),
51 DNA Denaturation, Topoisomerase I.

52 1. Introduction

53 Sequence-specific molecular recognition of the minor groove in double-helical DNA by low
54 molecular weight, aromatic peptides continues as a buoyant field of investigative research.¹⁻¹¹ The
55 broad class of ligands known as lexitropsins has been widely modified chemically and variously
56 investigated structurally as probes for DNA sequence recognition and particularly as potential
57 therapeutic agents for a variety of diseases¹²⁻¹⁵. Some of these compounds are now showing promise
58 as future antibiotics capable of overcoming the pervasive and very threatening advance of antibiotic
59 resistance¹⁶. Among these is the first in class compound known as MGB-BP-3, **1**, belonging to the
60 Strathclyde Minor Groove Binder family (S-MGBs), a substance that has successfully completed
61 Phase IIa clinical trials as a treatment against *Clostridioides difficile* infections and that could
62 outperform the industry standard vancomycin¹⁷.



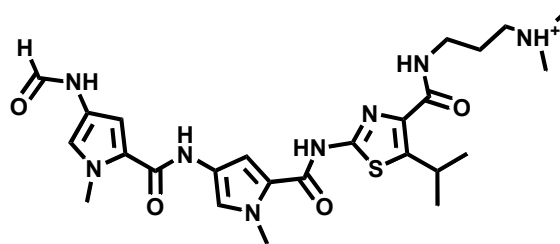
67 Molecules including **1** emerge from a sustained program of empirical investigation into the design,
68 synthesis and biological testing of a range of molecules,^{16, 18} which stem from a class of compounds
69 for which the name thiazotropins was originally coined and which included the first in class
70 molecule Thiazotropin A, **2**.^{3, 19, 20} Since DNA is considered to be the fundamental recognition
71 target of these compounds, the details of DNA structural and sequence recognition subtleties within
72 the overall development profile of these small molecule ligands has played a critical role in our
73 understanding of how these ligands behave with respect to DNA sequence recognition.

74

75

76

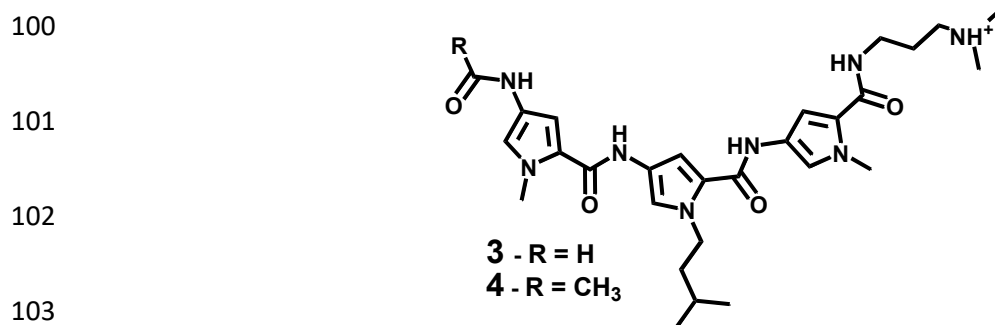
77

**2 (Thiazotropsin A)**

78 Among these investigations, solution-phase NMR spectroscopy studies continue to play a key role
79 in the biophysical exploration of the structure, dynamics and molecular recognition profiles of the
80 biomolecular complexes formed. When combined with results arising from related biophysical
81 studies, computational modelling and biochemical assays, NMR spectroscopy proves to be a well-
82 placed method that reports subtle structural variations which occur in molecular recognition
83 processes between ligands and nucleic acids. This becomes clearer through detailed evaluation of
84 three-dimensional solution-phase structures of these complexes.^{3, 21} Recognition relies on relative
85 complementarity between hydrogen bond partners, hydrophobic features, electrostatics and related
86 energetic considerations.^{3, 4} The hand-in-glove induced steric fit between ligand and biomolecule
87 target is key and detailed NMR studies, whilst being the gold standard for acquiring intricate
88 insights into the subtleties within such three-dimensional solution structures, are nevertheless costly
89 in terms of both the magnet time required to generate the data as well as the labour costs required
90 to interpret the data and reduce it to a form that is suitable for solution-phase three-dimensional
91 structure calculations.

92 Initial assessment of the state of interaction between ligand and DNA can be had by comparing one-
93 dimensional NMR data for a series of ligand analogues and their potential nucleic acid targets.
94 When taken in the context of studies across related ligand analogues and their DNA recognition
95 sequences, the significance of the findings takes on greater weight. The method lends itself to
96 screening newly synthesized ligands for their ability to target selected DNA sequences.

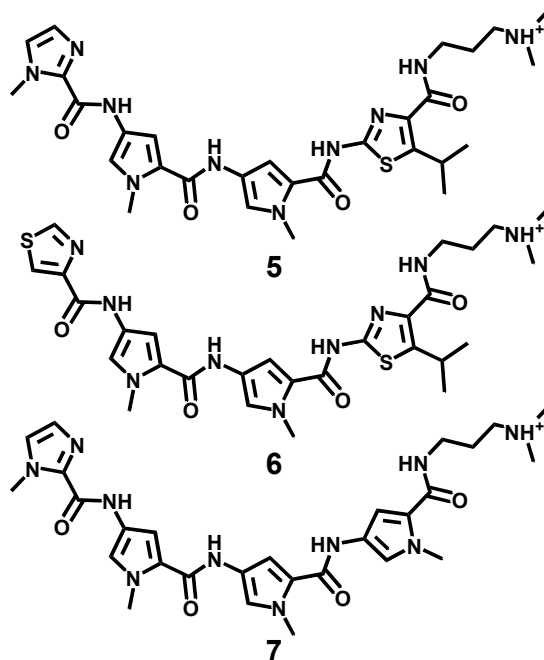
97 Illustration of this principle, reported for **3** and **4** with respect to DNA duplex d(CGATATATCG)₂,
98 allowed conclusions to be drawn concerning the relative fit between ligands and proposed DNA
99 binding site in each case.²²



104

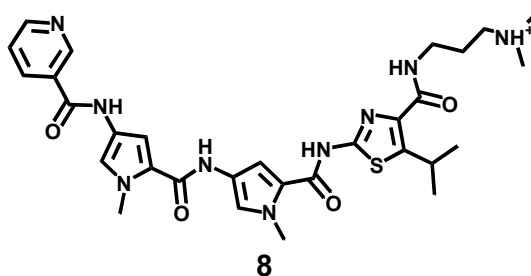
105 Findings from allied studies⁵ guided the design of new analogues (**5-7**) to target the DNA response
106 element sequence, 5'-WGWWCW-3' (W is A or T) in the current study, which serves as the binding
107 promoter site for the transcription factors of several nuclear receptors, including the androgen
108 response element (ARE) sequence. Extensive development work has taken place in designing and
109 testing hairpin and cyclic polyamide ligands in this regard.^{12, 23-26} including for use during *in vitro*
110 cell culture studies. These investigations reveal the ability of such molecules to regulate expression
111 of the androgen receptor, a factor associated with prostate cancer. High polarity connected with
112 these large molecules and their low cell permeability renders them orally inactive.²⁶ If lexitropsin
113 analogues **5-7** bind with 5'-WGWWCW-3', this would offer them better physical and biological
114 properties due to their enhanced lipophilicity (*via* thiazole rings) and lower molecular weights.

115
116
117
118
119
120
121



122 Here we report assessment of lexitropsin analogues **5-7** binding to DNA recognition sites in a
123 sequence-specific manner using a combination of biophysical, theoretical and biochemical
124 approaches. It is an example of the use of straightforward tools for assessing binding capability with
125 the added bonus that some of the information reports at the atomic scale. Variations in ligands
126 include replacement of isopropyl thiazole in **5** and **6** by *N*-methyl pyrrole in **7** and replacement of
127 headgroup *N*-methyl imidazole in **5** and **7** with thiazole in **6**. The selected compounds have been
128 screened against the DNA sequences 5'-ACTAGT-3' or 5'-AGTACT-3' within the context of 12-
129 mer, self-complementary duplexes and supported by preliminary findings for non-self-
130 complementary 15-mer, 18-mer and 20-mer DNA duplexes containing the same recognition sites
131 and based on former studies showing how thiazotropsin analogue **8** is bound securely to 5'-
132 ACTAGT-3'.^{4, 25}

133
134
135



136 In the case of **8**, the calculated ligand-DNA complex revealed no particular evidence for hydrogen
137 bond interaction between nitrogen in the pyridyl headgroup and the DNA recognition site.⁴ It was
138 speculated that repositioning nitrogen in the headgroup along with reorientation of selected DNA
139 base-pairs could subtly alter the recognition profile of these molecules and create opportunities for
140 hydrogen bonding that would secure the ligand to create energetically favourable complexes with
141 enhanced K_d .

142 Screening for the tell-tale signs of slow chemical exchange on the NMR chemical shift timescale
143 between ligand and DNA and comparison with data arising from DNA denaturation and molecular
144 modelling studies confirm the structural features and forces that influence stabilization of these
145 ligand-DNA complexes.

146 NCI (National Cancer Institute, USA) screening results for the investigated MGBs showed a potent
147 antiproliferative activity for the ligand with thiazole head group (ligand **6**). The study therefore
148 discusses the potential mechanisms of action and structure-activity relationships for antitumor
149 activity, which could be helpful for future optimisation and development of MGBs as potential
150 anticancer agents.

151

152 **2. Results and Discussion**

153 **2.1 Chemical synthesis**

154 Ligands were synthesised according to published procedures^{2, 5, 15} being reproduced for
155 convenience to the reader in the Supplementary Information at **C1** (description) and by summary
156 in **Scheme C1** (see Supporting Information).

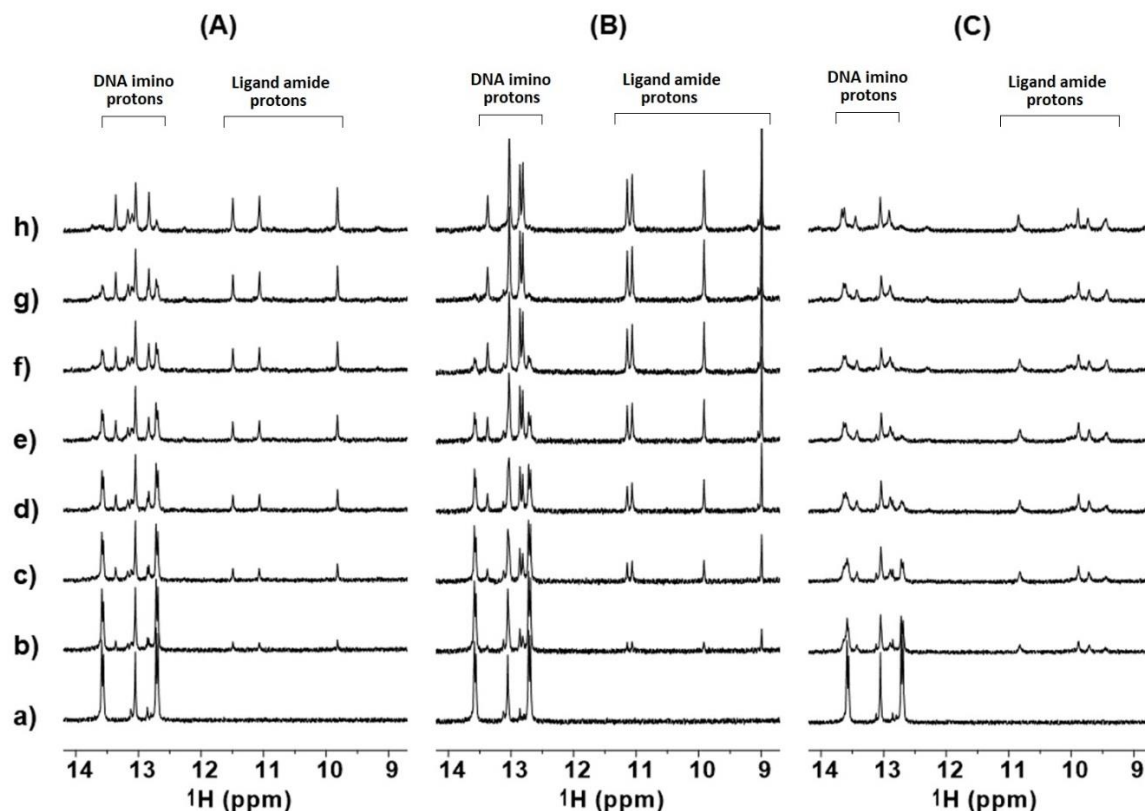
157

158

159 2.2 One-dimensional NMR data

160 Stock solutions of DNA samples, prepared alongside concentrated stock solutions of **5-7**, were
161 divided into aliquots. Each DNA sample was titrated using the concentrated stock ligand solutions.
162 Six samples were created using self-complementary 12-mer oligonucleotides. Further samples were
163 prepared using 15-mer, 18-mer and 20-mer oligonucleotides as detailed in the Supplementary
164 Information. Titrating DNA samples with ligand rather than attempting to load the relevant number
165 of equivalents of ligand into each sample allows nuances within the data to be observed as titrations
166 proceed. Acquisition of 1D ^1H and $^{31}\text{P}\{-^1\text{H}\}$ NMR data enable changes to be observed at the
167 molecular level for both ligand and DNA (*via* ^1H NMR spectra) and DNA alone (*via* $^{31}\text{P}\{-^1\text{H}\}$ NMR
168 spectra). Focus centres on the $\delta^1\text{H} = 8.0 - 14.0$ ppm region of the ^1H NMR spectrum for samples
169 solubilized in 90% H_2O / 10% D_2O for observation of DNA imino and peptide NH NMR resonances
170 arising from ligands containing peptide bonds.

171 As shown for the titration of **5-7** with $\text{d}(\text{CGCACTAGTGCG})_2$ (Figure 1) three quite similar ligands
172 result in somewhat different 1D ^1H NMR spectral properties: the introduction of **5** to
173 $\text{d}(\text{CGCACTAGTGCG})_2$ yields a clear set of peptide NH signals arising from the ligand by the end
174 point of the titration (Figure 1 (A), h). A mixture of sharp higher intensity signals among a set of
175 low intensity, broad resonances appear in the DNA imino proton resonance region. Contrastingly,
176 addition of **6** to $\text{d}(\text{CGCACTAGTGCG})_2$ results in a definitive change in the spectrum, showing as
177 clean, sharp responses with no evidence of low level, broadened NMR signals (Figure 1 (B), h).
178 Further marked differences show with **7** compared with the results for **5** and **6** where signal intensity
179 is low and definition for the DNA imino and ligand peptide NH resonances is poor and ill defined
180 (Figure 1(C), h). These differences are noteworthy given the very subtle structural variations across
181 the three ligand molecules studied.



182

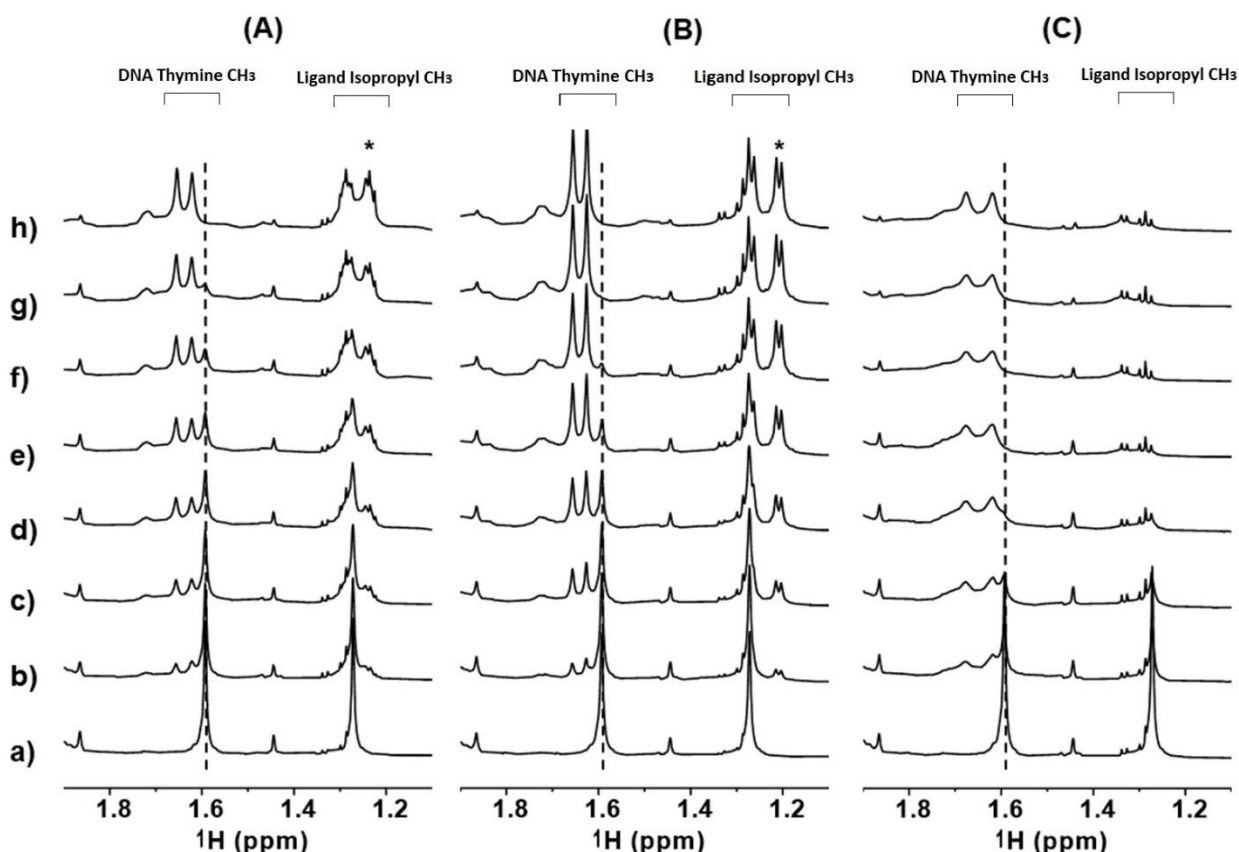
183 **Figure 1.** Titration of DNA duplex $d(CGCACTAGTGCG)_2$ with ligands **5-7**: Overview of
 184 equivalent regions of 600 MHz 1D 1H NMR spectra recorded for each titration point showing DNA
 185 duplex imino NH and ligand peptide NH resonances associated with each system. **(A)** Titration with
 186 **5**; **(B)** Titration with **6**; **(C)** Titration with **7**. Titration points are defined in number of molar
 187 equivalents of ligand with respect to DNA duplex: **a**) 0:1; **b**) 0.3:1; **c**) 0.7:1; **d**) 1:1; **e**) 1.3:1; **f**) 1.7:1;
 188 **g**) 2:1; **h**) 2.1:1. Resonance assignments: 1) Free DNA: 13.59 ppm - T^9H3 ; 13.56 ppm - T^6H3 ; 13.05
 189 ppm - G^2H ; 12.72 ppm - $G^{10}H1$; 12.69 ppm - G^8H . 2) Complex with **5**: 11.49 ppm - $5H16$; 11.02
 190 ppm - $5H24$; 9.62 ppm - $5H8$. 3) Complex with **6**: 13.38 ppm - T^9H3 ; 13.03 ppm - G^2H1 ; 13.02
 191 ppm - G^8H1 ; 12.86 ppm - T^6H3 ; 12.81 ppm - $G^{10}H1$; 11.14 ppm - $6H16$; 11.06 ppm - $6H24$; 9.91
 192 ppm - $6H8$; 8.99 ppm - $6H2$.

193

194 When ligands were added to the solution of $d(CGCACTAGTGCG)_2$, several new resonances
 195 emerged in the imino proton resonance region (Figure 1). These resonances steadily intensified
 196 throughout the titration until reaching a 2:1 ratio of ligand to DNA, at which point the resonances
 197 of the free DNA disappeared entirely (most notably with ligand 6). In the presence of free DNA,
 198 the 1H NMR spectrum showed well-defined imino proton resonances from both free and ligand-
 199 bound DNA concurrently, indicating a tight binding that induces a slow exchange between the free
 200 and ligand-bound DNA.²⁷⁻²⁹ However, with ligand 7, the NMR spectrum exhibits broadened peaks,

201 possibly due to intermediate exchange³⁰⁻³² between the free and bound states in case of weak
202 binding ligands. When all free DNA in solution is bound to ligand, the ¹H NMR spectrum for the
203 complex becomes simplified, indicating that two MGB molecules oriented themselves in opposite
204 directions within the minor groove of the palindromic sequence d(CGCACGCGAGCG)₂. This
205 mode of binding preserves the symmetry of the complex and limits variation in the chemical shifts
206 of both the DNA and ligand protons, thereby simplifying the NMR spectrum. The complexity of
207 the NMR spectrum alone can therefore serve as an indicator of the stoichiometry of interaction for
208 tight binding ligands. Having a ¹H NMR spectrum with only one peak corresponding to each ligand
209 proton and the appearance of only 3 and 4 imino proton resonances for the free and ligand bound
210 DNA, respectively (e.g., ligand 6 Figure 1B) indicated the formation of a 2:1 (ligand: DNA)
211 symmetric complex with no evidence for the presence of other complexes. These findings are
212 consistent with previous 1D and 2D NMR studies of analogous MGBs that showed the ability of
213 these compounds to bind to the minor groove of DNA in a 2:1 ratio as antiparallel side-by-side
214 dimer in a head-to-tail fashion.^{4, 27, 33} Additionally, computational studies (sections 2.6 & 2.7)
215 confirmed the 2:1 binding stoichiometry of the investigated MGBs.

216 This theme is extended into the aliphatic methyl proton resonance regions of the same NMR spectra
217 (Figure 2). Loss of signals arising from free DNA is clearly evident (dashed line position in each
218 case at the proton chemical shift of the T⁶ methyl group in self-complementary 12-mer). Loss of
219 free-DNA is clear in each case, but replacement by signals from ligand-bound DNA is variable from
220 sample to sample including loss of *J*-coupling definition for the isopropyl methyl proton NMR
221 resonances in the titration of **5** with d(CGCACTAGTGCG)₂ (Figure 2 (A), h, at *) compared with
222 that of **6**, where good definition of the retained *J*-coupling exists (Figure 2 (B), h, at *).



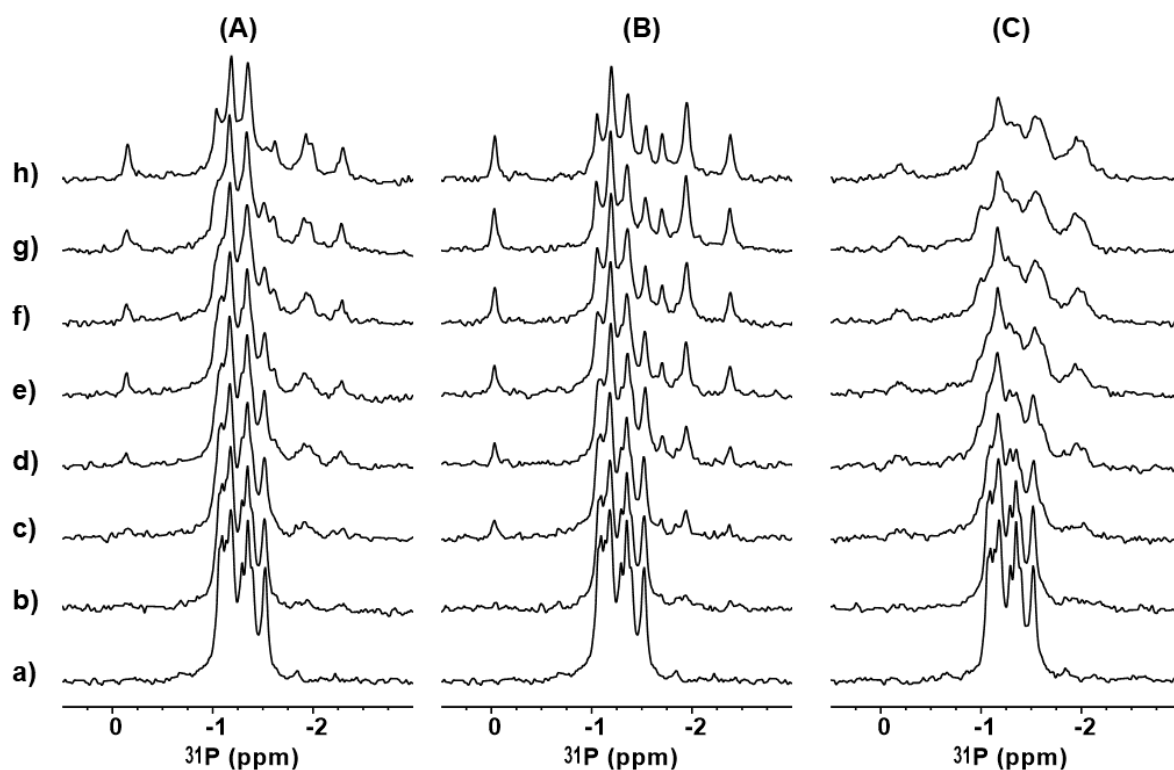
223

224 **Figure 2.** Titration of DNA duplex $d(CGCACT^6AGTGCG)_2$ with ligands **5-7**: Methyl resonance
 225 regions of 600 MHz 1D 1H NMR spectra recorded for each titration point. **(A)** Titration with **5**; **(B)**
 226 Titration with **6**; **(C)** Titration with **7**. Titration points are defined in number of molar equivalents
 227 of ligand with respect to DNA duplex: **a)** 0:1; **b)** 0.3:1; **c)** 0.7:1; **d)** 1:1; **e)** 1.3:1; **f)** 1.7:1; **g)** 2:1; **h)**
 228 2.1:1. Dashed line: Signal assigned to T^6CH_3 in free DNA at $\delta^{1H} = 1.592$ ppm. * Methyl 1H NMR
 229 resonance from one of two isopropyl methyl groups of ^{iPr}Th in **5** and **6**. Resonance assignments: 1)
 230 Free DNA: 1.59 ppm – T^6CH_3 ; 1.27 ppm – T^9CH_3 . 2) Complex with **5**: 1.66 ppm – T^6CH_3 ; 1.63
 231 ppm – T^9CH_3 ; 1.27 ppm, 1.21 ppm – $5^{31/32}CH_3$. 3) Complex with **6**: 1.66 ppm – T^6CH_3 ; 1.625 ppm
 232 – T^9CH_3 ; 1.27 ppm, 1.21 ppm – $6^{31/32}CH_3$. 4) Complex with **7**: 1.65 ppm – T^6CH_3 ; 1.61 ppm –
 233 T^9CH_3 .

234

235 The pattern is consistent for related one-dimensional $^{31}P\{-^1H\}$ NMR data. For the same systems
 236 (Figure 3) a resonance appears at higher chemical shift close to $\delta^{31}P = 0$ ppm compared with all of
 237 the other $^{31}P\{-^1H\}$ NMR signals, a feature known to be consistent with a specific phosphodiester
 238 group in the DNA backbone, which transitions from predominantly B_I backbone conformation in
 239 free DNA to predominantly B_{II} backbone conformation in the ligand bound form. Whilst this signal
 240 is clear and present in the complex with **5** (Figure 3 (A), h) and clearer still in the case of **6** (Figure

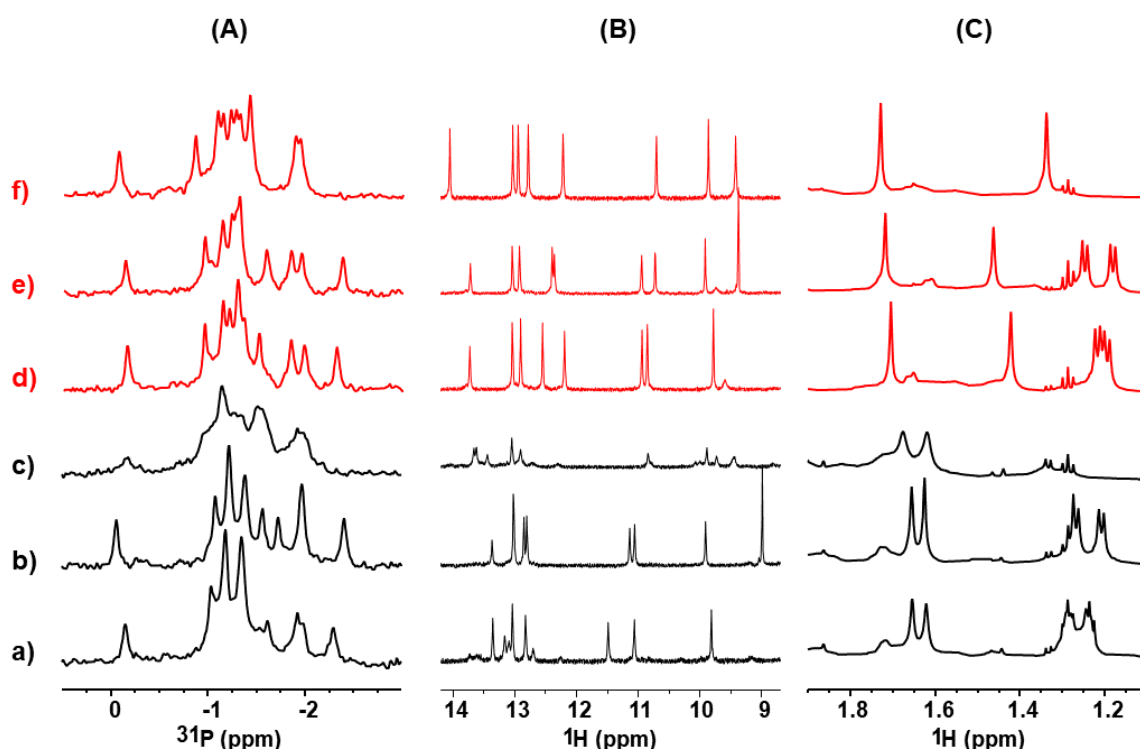
241 3 (B), h), a similar resonance present in the data for the DNA complex with **7** (Figure 3 (C), h) is
 242 much less obvious, appearing as a broadened response.



243
 244 **Figure 3.** Titration of DNA duplex d(CGCACTAGTGCG)₂ with ligands **5-7**: ³¹P-¹H} NMR
 245 spectra recorded for each titration point. **(A)** Titration with **5**; **(B)** Titration with **6**; **(C)** Titration
 246 with **7**. Titration points are defined in number of molar equivalents of ligand with respect to DNA
 247 duplex: **a)** 0:1; **b)** 0.3:1; **c)** 0.7:1; **d)** 1:1; **e)** 1.3:1; **f)** 1.7:1; **g)** 2:1; **h)** 2.1:1.

248
 249 Marked contrast can be drawn between the sets of data presented for the titration of
 250 d(CGCACTAGTGCG)₂ and **5-7** with those associated with the titration of
 251 d(CGCACTAGTGCG)₂ by the same ligands (Figure 4 and Supporting Information Figures S1-S3
 252 for contrasting data sets to those shown in Figures 1-3). The difference in appearance of the data
 253 for the two DNA sequences is notable, especially for **7**. Substantially broadened NMR spectra are
 254 apparent with d(CGCACTAGTGCG)₂ in the presence of **7** (Figure 4c, A, B, and C) in contrast to
 255 the sharp, well defined NMR spectra apparent with d(CGCACTAGTGCG)₂ in the same ligand
 256 context (Figure 4f, A, B, and C). The variations between complexes of **5** with the two DNA

257 sequences is less obvious on first inspection but is apparent when compared closely with the
 258 complexes of **6** in the two DNA contexts.



259
 260 **Figure 4.** Contrasting behaviours of DNA complexes of ligands **5-7** through observation of 1D ^1H
 261 and ^{31}P - $\{^1\text{H}\}$ NMR spectra: (A) 1D ^{31}P - $\{^1\text{H}\}$ NMR data; (B) 1D ^1H NMR data in the DNA imino
 262 proton and peptide NH resonance region; (C) 1D ^1H NMR data in the aliphatic methyl resonance
 263 region. a)/d): Complexes of **5**; b)/e): Complexes of **6**; c)/f): Complexes of **7**. Black spectra:
 264 complexes with d(CGCACTAGTGCG) $_2$; Red spectra complexes with d(CGCAGTACTGCG) $_2$.
 265 ^1H resonance assignments for complexes with d(CGCAGTACTGCG) $_2$: d) Complex with **5**: 13.75
 266 ppm - T 9 H3; 13.01 ppm - G 2 H1; 12.85 ppm - G 8 H1; 12.51 ppm - T 6 H3; 12.20 ppm - G 10 H1; 10.95
 267 ppm - 5H16; 10.80 ppm - 5H24; 9.61 ppm - 5H8; 1.71 ppm - T 6 CH $_3$; 1.45 ppm - T 9 CH $_3$. e)
 268 Complex with **6**: 13.75 ppm - T 9 H3; 13.01 ppm - G 2 H1; 12.86 ppm - G 8 H1; 12.45 ppm - T 6 H3;
 269 12.40 ppm - G 10 H1; 10.95 ppm - 6H16; 10.70 ppm - 6H24; 9.95 ppm - 6H8; 9.40 ppm - 6H2; 1.72
 270 ppm - T 6 CH $_3$; 1.46 ppm - T 9 CH $_3$. f) Complex with **7**: 14.05 ppm - T 9 H3; 13.01 ppm - G 2 H1; 12.86
 271 ppm - G 8 H1; 12.75 ppm - T 6 H3; 12.20 ppm - G 10 H1; 10.75 ppm - 7H16; 9.80 ppm - 7H24; 9.45
 272 ppm - 7H8; 1.73 ppm - T 6 CH $_3$; 1.34 ppm - T 9 CH $_3$.

273
 274 Through visual inspection of the NMR line shape, it is evident that ligands **5-7** bind tightly with 5'-
 275 AGTACT-3' site, as indicated by the sharpness of the NMR peaks (Figure 4, red spectra).
 276 Conversely, the significant peak broadening observed in the association of ligand **7** with 5'-
 277 ACTAGT-3' site (Figure 4 c)) suggests a weak binding interaction. These findings are consistent

278 with previous studies that have reported ^1H NMR data for both strong and weak binding MGBs.^{31,}
279 ³⁴

280 **2.3 Use of 1D NMR as a qualitative assessment tool for MGB binding**

281 When ligand binding to DNA occurs, changes in chemical shifts and the sharpness of NMR peaks
282 occur due to both the close contact between the ligand and DNA and the ligand-induced
283 conformational changes of the DNA helix, with binding influenced by both the number and strength
284 of the non-covalent interactions formed between the ligand and DNA.

285 MGBs with dimensions similar to distamycin typically bind to 5-6 base pairs. The observed
286 alterations in chemical shifts therefore extend across multiple base pairs within a binding site.^{31, 34}
287 NMR peak shape reflects the behaviour of a complex at the atomic scale. This can guide subsequent
288 exploration of dynamics and stability using complementary analyses. Sharp NMR peaks may be
289 associated with changes in molecular dynamics, chemical exchange rates and molecule-on-
290 molecule association characteristics including aggregation. It may report on increased rigidity^{35, 36}
291 arising from changes caused by formation of non-covalent interactions within a complex. It can also
292 be used, without recourse to extensive and time-consuming data interpretation, as a first measure in
293 providing atomic level information on molecular recognition following careful study designs as
294 illustrated for **6** and **7** when screened against 15-mer, 18-mer and 20-mer DNA duplexes containing
295 both of the DNA recognition sequences screened in this study (see Supplementary Information).
296 The ligand induced-bending of the DNA helix toward the major groove results in major groove
297 compression (see Figure 14). Consequently, the chemical shift of the thymine methyl group can be
298 influenced by minor groove binders, although the methyl group is located in the major groove. This
299 is attributed to structural changes in the DNA helix, leading to alterations in the chemical
300 environment of certain nucleotide protons, including the thymine methyl group within the major
301 groove.

302 Distinct resonances for free and bound DNA states in slow chemical exchange on the NMR
303 chemical shift timescale^{35, 37} provide a guide when assessing ligand binding. Numerous studies in
304 the literature report the utility of 1D NMR to study DNA minor groove binders and to determine
305 the exact binding site in the DNA sequence from the changes in chemical shifts.^{4, 38-42} Among the
306 investigated MGBs are the conjugated biaryl minor groove binders (e.g., Hoechst 33258, DB226,
307 DB224, DB60, and DB75)⁴³⁻⁴⁵ distamycin analogues (e.g., thiazotropsin A, thiazotropsin B, AIK18-
308 51)⁴, hairpin and cyclic polyamide³⁴ ligands. Despite the structural diversity of these MGBs, they
309 share the common characteristic of a curved shape that fits the helical structure of DNA. This snug
310 and close association with the minor grooves comes with disruption of the groove dimensions upon
311 complexation.⁴⁶⁻⁴⁸ Therefore, the described NMR method may be widely applied to screen for minor
312 groove recognition by leveraging chemical shifts, line shape and signal resolution, regardless of the
313 specific structures of the ligands and the DNA sequences involved in complexation. The unique
314 binding mode of MGBs, along with structural alterations in the DNA helix, result in distinct changes
315 in ¹H and ³¹P-¹H NMR spectra. This approach proves beneficial for more complex cases when
316 larger DNA fragments, with mixtures of target sites, are screened for minor groove binding by a
317 given class of ligand, as illustrated in the supplement to this study with longer DNA fragments. Our
318 findings show suitability for distinguishing binding site specificity through straightforward
319 comparison of data arising from such model systems. If the ligand-DNA titration experiment is
320 well-designed and executed, the likelihood of artifacts other than binding causing changes in line
321 shape and chemical shifts is negligible. NMR spectroscopy is a well-established, robust, and reliable
322 technique, offering superior reproducibility of results compared to other analytical methods. Visual
323 inspection of the 1D NMR spectra by the analyst is helpful in confirming binding and allows
324 selection of those systems which the researcher can confidently expect to yield high-resolution,
325 solution-phase NMR structures. Such findings do not require validation by other analytical

326 techniques, allowing the analyst to proceed directly to lengthier 2D/3D NMR data collection for
327 full signal assignment.

328 It should be emphasized that the conclusions related to the NMR peak shape are valid for interaction
329 of ligands with short, well-defined ODN duplexes. Long DNA molecules, such as those found in
330 salmon sperm and calf thymus DNA, may contain multiple binding sites, or some overlapping sites,
331 that severely complicates any analyses.^{2, 11, 49} This complexity arises from competing effects and
332 uncertainties regarding the stoichiometry of ligand interaction with long DNA molecules.

333 In the context of ligand binding to short ODNs, the sharpness of peaks in an NMR spectrum reflects
334 the rate at which molecules transition between free and bound states. Higher binding affinity often
335 leads to slower exchange rates between free and bound states,^{28,29} where the equilibrium is shifted
336 toward the bound state. As the ligand is titrated, new signals corresponding to the complexed state
337 gradually emerge, while the signals for free DNA gradually diminish.⁵⁰ This results in sharp NMR
338 peaks as the NMR instrument can distinguish between the separate resonances of the different states
339 due to the slow transition that generates a stable and dominant bound state. As molecules transition
340 between states slowly, the NMR signals from each state are observed separately, resulting in well-
341 defined NMR peaks.³ It is important to note that following the binding interaction, the resulting
342 complexes retain the ability to move freely within the solution. Consequently, the nuclei encounter
343 a relatively consistent and homogeneous magnetic environment, leading to clearly defined
344 resonance frequencies and narrower peaks.

345 Peak broadening in NMR spectra can be directly related to the dynamics of binding interactions
346 between molecules. Specifically, the rate of exchange between the free and bound states of
347 molecules involved in binding interactions influences the extent of peak broadening. If the binding
348 interaction is moderate or weak, where there is continuous equilibrium between the free and bound
349 states (intermediate exchange), the peaks in the NMR spectrum become broadened.^{30, 32, 51-53} This

350 broadening occurs because the NMR signals from the free and bound states are not distinguishable
351 and hence averaged over time, resulting in reduced resolution and broader peaks in the spectrum.

352 Peak broadening in NMR spectra can also be attributed to aggregation phenomena within the
353 sample. When molecules aggregate or form larger structures, such as dimers, oligomers, or
354 aggregates, the resulting NMR signals can become broadened.⁵⁴ This broadening occurs because
355 the aggregated species generally undergo slower molecular tumbling rates compared to individual
356 molecules. As a result, the NMR signals from these aggregated species encounter heterogeneous
357 magnetic environments, resulting in broader resonance peaks in the spectrum.

358 Ultimately, in NMR spectroscopy, the sharpness of peaks can provide insights into the binding of
359 small molecules with short ODNs: Sharp peaks with no changes in the chemical shifts indicate no
360 binding. Changes in the chemical shifts, which are associated with the emergence of sharp peaks
361 may indicate both higher binding affinity and a good chance to obtain a high-resolution NOESY
362 NMR structure of the ligand-ODN complex. Peak broadening may indicate weak binding affinity
363 and imply challenge in obtaining clear NOE contacts between the ligand and ODN, as supported by
364 numerous previous studies.^{3, 4, 31}. However, it is essential to note that other factors, such as the
365 temperature, sample conditions, concentration and the nature of interacting molecules, can also
366 influence peak broadening. Hence, while the sharpness of NMR peaks can offer valuable qualitative
367 information about binding affinity, quantitative analysis often requires utilizing additional
368 experimental techniques

369 **2.4 DNA denaturation studies (T_m)**

370 By studying the DNA melting profiles, the effects of ligand-DNA complexation on the stability of
371 short DNA sequences can be assessed.⁵⁵ Since the stability of the DNA double helix is increased
372 by ligand binding, elevating the DNA melting temperature (T_m) significantly, ΔT_m values can be
373 used as an indicator for the strength of binding interaction.^{11, 56}

374 We determined the UV melting characteristics of the two self-complementary 12-mer DNA
 375 sequences in the absence and presence of 2 equimolar quantities of each ligand (**5**, **6** and **7**) in PIPES
 376 buffer. T_m values were determined from melting curves at the half height when half of the DNA is
 377 denatured. The absorbance was determined in the temperature range 30 – 100 °C. The production
 378 of air bubbles in solution at high temperature (>90 °C) caused an increase in the absorbance
 379 measurements and for this reason these readings were ignored in the final analysis.

380 In the absence of any added ligand, the melting temperature (T_m) of the two DNA sequences was
 381 found to be 48 ± 0.5 °C. The melting temperature of the ACTAGT sequence increased by 15 and
 382 23 °C after the addition of ligands **5** and **6**, respectively, while a slight increase in T_m was observed
 383 (3 °C) upon the addition of **7** (Table 1, Supplementary Information Figure S5) indicating weak or
 384 negligible interaction. A large increase was observed in the DNA melting temperature of the
 385 AGTACT sequence by 24, 23, and 18.5 °C, for complexes with **5**, **6**, and **7**, respectively. These
 386 findings clearly indicate that the studied ligands can stabilize the DNA helix *via* formation of strong
 387 non-covalent interactions and the strength of these interactions, as indicated from the measured ΔT_m
 388 values, are consistent with the NMR findings.

389 **Table 1.** Changes in the DNA melting temperature of the studied ODNs, d(5'-GCGACTAGTCGC-
 390 3')₂ and d(5'-GCGAGTACTCGC-3')₂, after the addition of two equivalents of **5**, **6** and **7**.

391	Binding ligand	5	6	7
	ΔT_m (°C) of the ODNs [†]			
392	d(5'-CGCACTAGTGCG-3') ₂ [‡]	15	23	3
	d(5'-CGCAGTACTGCG-3') ₂	24	23	18.5

393 [†]Absorbance was measured in the temperature range 30-100 °C.

394 [‡]The solutions contained 10 mM PIPES buffer, 20 mM NaCl, and 1 mM EDTA at
 395 pH 7.00. ([DNA] = 10 μM).

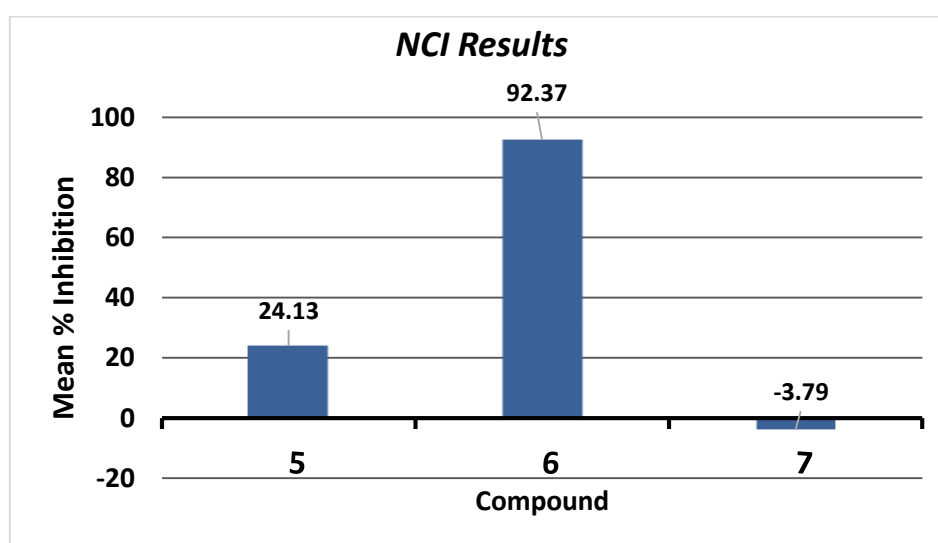
396

397

398

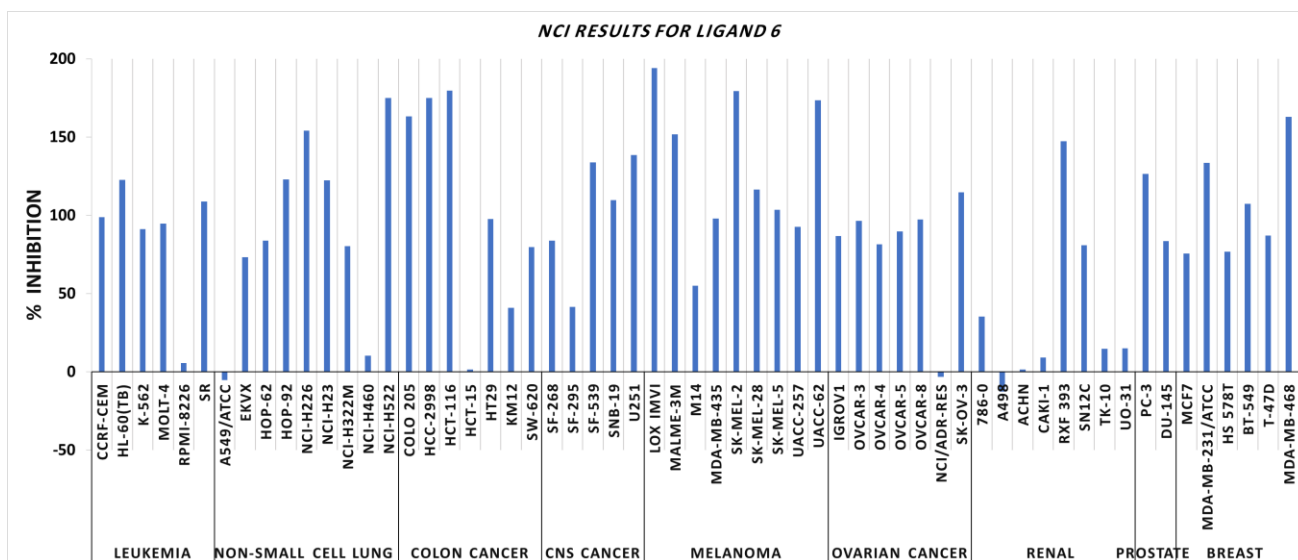
399 2.5 Antitumor activity

400 In line with the protocol of the Drug Evaluation Branch, National Cancer Institute (Bethesda, USA),
401 60 human tumor cell lines were studied *in vitro* against the target MGBs, **5**, **6** and **7** using nine
402 different tissues (Supplementary Information, Figure S6). A single-dose (10 μ M) was used for
403 estimating the percentage of growth inhibition of the NCI-60 panel cell lines. The NCI results
404 showed that ligand **6** induced the highest level of growth inhibition against the NCI-60 panel (92.73
405 %), whereas compound **5** and **7** showed weak and negligible inhibition respectively (Figure 5).



406
407 **Figure 5.** Mean % inhibition of the tested compounds (**5-7**) over NCI-60 cancer cell line panel. The
408 percentage of growth inhibition was calculated by subtracting mean growth percentage values obtained from
409 the NCI from 100.

410 Ligand **6** showed more than 75 % inhibition on 45 cell lines including melanoma, colon, non-small
411 cell lung, breast, renal, CNS, leukemia, ovarian, and prostate cancer cell lines (Figure 6). It showed
412 the highest activity against melanoma, colon, breast, and non-small cell lung cancer subpanels. Due
413 to its broad-spectrum antiproliferative activity, the NCI selected ligand **6** for five-dose testing to
414 determine its GI₅₀ and total growth inhibition (TGI) values (provided in the Supplementary
415 Information, Figure S7).



416

417 **Figure 6.** Percentage of growth inhibition induced by ligand **6** against all tested cancer cell-
 418 lines.

419

420 Although our biophysical studies confirmed the direct binding of ligands **5-7** with different extents
 421 to short DNA sequences, the huge variation in their antitumor activity can be attributed to another
 422 crucial factor — the lipophilicity of the ligand. The degree of lipophilicity impacts the ligand's
 423 efficiency in penetrating the cell membrane and reaching its molecular target within the nucleus.
 424 This conclusion is consistent with the calculated lipophilicity (LogP) for **5**, **6**, and **7**, which were
 425 found to be 0.55, 1.11, and -1.34 respectively at pH 7.4 (determined by Marvin,
 426 <https://www.chemaxon.com>). The computational permeability tool, PerMM, ⁵⁷ indicated that **6**
 427 exhibits a 26-fold higher cell permeability than **7** in PAMPA model and a 2.5-fold greater
 428 permeability in Caco-2 model (**S1**, **S13**). Additionally, **6** exhibits a 5.4-fold higher cell permeability
 429 than **5** in PAMPA model and a 1.6-fold greater permeability in the Caco-2 model. These results
 430 align with the logP values of MGB compounds and highlight the crucial influence of lipophilicity
 431 on the biological activity of MGBs. For instance, the relatively polar ligands **5** and **7** can physically
 432 bind to DNA. However, their low cell permeability renders them inactive. A similar observation on
 433 the effect of lipophilicity on lung tumour potency has previously been reported for another structural

434 class of MGBs.¹³ Such results indicate that the lipophilicity plays a vital role in the anticancer
435 activity of the investigated MGBs and shall be considered in future optimization and development
436 of novel DNA minor groove binders.

437 **2.6 Molecular Docking**

438 Qualitative ¹H NMR data showed the binding of ligands **5-7** to different extents to the two DNA
439 sequences d(CGCACTAGTGCG)₂ and d(CGCAGTACTGCG)₂. To perform reliable molecular
440 modeling studies for these complexes, the starting 3D structures of the two DNA sequences and the
441 investigated ligands were derived from an experimental NMR refined structure for the complex
442 between d(CGACTAGTCG)₂ and the related thiazotropsin analogue **8**, which was obtained from
443 the protein data bank (PDB code: 2MNE).⁴ HADDOCK 2.4 web server was used to determine the
444 best binding pose of the ligands **5-7** with the two DNA sequences and the generated structures were
445 then used as inputs for the molecular dynamic simulation studies to assess the stability of the studied
446 complexes. The HADDOCK docking results (Figure 7) of ligands **5-7** with the DNA duplex
447 d(CGACTAGTCG)₂ showed that these MGBs bind to 5'-ACTAGT-3' in the minor groove of DNA
448 as dimers in an antiparallel side-by-side fashion, except the complex with ligand **7**. This showed a
449 distorted dimer that did not fit deeply and snugly within the minor groove (Figures 7C). The docking
450 results of ligands **5-7** with d(CGACTAGTCG)₂ are listed in Table 4. A higher negative
451 HADDOCK score indicates a more stable complex. According to the HADDOCK results, ligands
452 **5** and **6** showed higher stabilities than **7** with the ACTAGT binding site (Table 2), which is
453 consistent with NMR results.

454

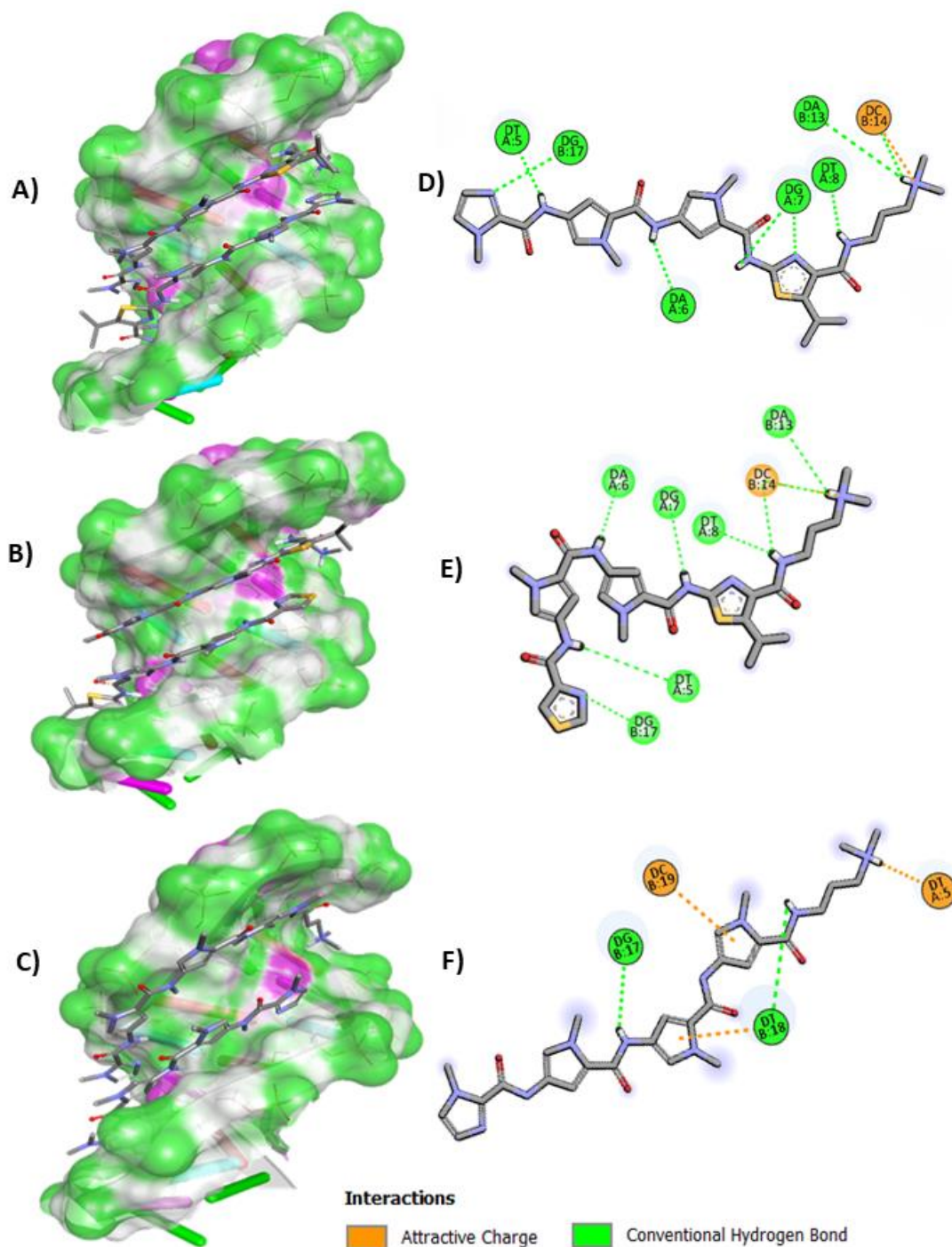
455

456

457 **Table 2.** Docking results and water refinement of ligands **5-7** with d(CGACTAGTCG)₂, obtained
 458 from HADDOCK docking server.

460	<i>Ligand</i>	5	6	7
461	<i>HADDOCK score</i>	-202.4 ± 3.4	-203.2 ± 1.2	-181.4 ± 3.6
462	<i>RMSD (Å)</i>	3.9 ± 0.1	0.6 ± 0.4	0.6 ± 0.4
463	<i>van der Waals energy (kcal/mol)</i>	-145.4 ± 2.7	-147.3 ± 3.3	-128.6 ± 3.6
464	<i>Electrostatic energy (kcal/mol)</i>	-374.1 ± 11.7	-392.5 ± 13.2	-379.5 ± 10.2
465	<i>Desolvation energy</i>	17.8 ± 0.5	22.6 ± 0.5	23.0 ± 1.2
466	<i>Buried surface area (Å²)</i>	2875.7 ± 38.6	2925.1 ± 18.8	2804.0 ± 47.3

467 According to the HADDOCK results, electrostatic energy is the major driving force for
 468 complexation. The stability of these complexes is mainly dictated by a series of hydrogen bonds
 469 formed between the NHs of the amide links, the nitrogen of imidazole and thiazole rings in the
 470 MGB structures and the DNA bases in the minor groove. In **5** and **6**, six hydrogen bonds were
 471 formed with the DNA bases, in addition to the electrostatic interaction between the DMAP tail of
 472 the ligand and negatively charged backbone of DNA, whereas **7** exhibited only two weak hydrogen
 473 bonds *via* the NH's of the amide links and that might explain the lack of stability of this complex.
 474 The interaction of ligands **5-7** with d-(5'-CGAGTACTCG-3')₂ DNA duplex was also studied using
 475 HADDOCK 2.4 (see Supplementary Information Figures **S8** and **S9** for the correlated data sets to
 476 those shown in Figure 7 and Table 2). The HADDOCK docking results showed that the three MGB
 477 ligands bind to 5'-AGTACT-3' in the minor groove of DNA as dimers in an antiparallel side-by-
 478 side fashion, which is similar to the binding mode observed with the 5'-ACTAGT-3' binding site.
 479 However, ligand **7** formed a more stable complex with 5'-AGTACT-3'. This can be related to the



480

481 **Figure 7.** 3D diagram of the best docked structure for (A) 5, (B) 6, and (C) 7, with
 482 d(CGACTAGTCG)₂. 2D drawing demonstrating the non-covalent interactions for the complexation
 483 of (D) 5, (E) 6, and (F) 7 with d(CGACTAGTCG)₂. The binding interactions shown in D, E and F
 484 were generated automatically by BIOVIA Discovery Studio Visualizer (2021).⁵⁸ (Note: the
 485 conformations of the ligands shown on the right-hand side of the figure [D, E and F] do not represent
 486 the correct curved conformation of MGB ligands that matches the curvature of the minor grooves
 487 of DNA, which are correctly defined in ligand-DNA complexes shown on the left-hand side of the
 488 figure [A, B and C]).

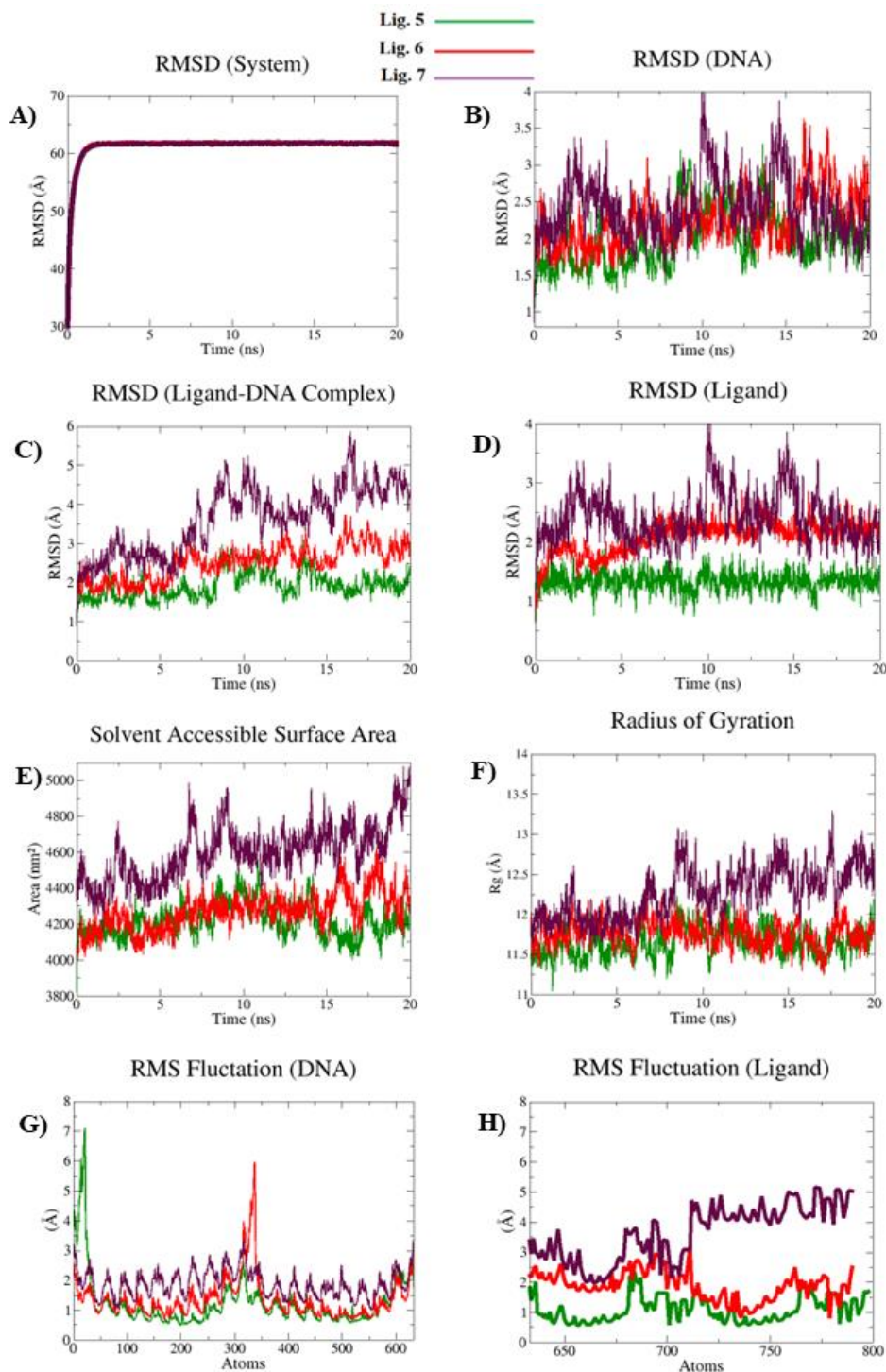
489 formation of five hydrogen bonds *via* the amide NH's and the nitrogen of imidazole in **7**
490 (Supplementary Figure S8, F), which were able to form five hydrogen bonds with the DNA bases
491 compared to only two hydrogen bonds with the 5'-ACTAGT-3' binding site.

492 Remarkably, the nitrogen in head group imidazole and thiazole rings in **5** and **6** respectively, which
493 sit close to cytosine base number 4 (Base Pair C₄-G₁₇) in the first strand of the DNA duplex
494 d(CGAC⁴TAGTCG)₂ was capable of forming a hydrogen bond with the guanine base number 17
495 on the parallel strand, C₄-G₁₇ (Figure 7, D & E). The imidazole and thiazole nitrogen also formed
496 a direct hydrogen bond with the guanine base number 4 in the first strand (G₄-C₁₇) of the DNA
497 duplex d(CGAG⁴TACTCG)₂ (see Supplementary Figure S8, D & E). The same trend was noticed
498 for the isopropylthiazole ring, which formed selective hydrogen bonds with the guanine bases
499 located at position 7 (G₇-C₁₄) and position 14 (C₇-G₁₄) in the DNA duplexes
500 d(CGACTAG⁷TCG)₂ and d(CGAGTAC⁷TCG)₂ respectively.

501 These findings indicate that the imidazole and thiazole rings in the ligand dimer can recognize both
502 GC and CG base pairs and explain the ability of ligands **5** and **6** to bind with both the ACTAGT
503 and AGTACT sites.

504 **2.7 Molecular Dynamic Simulations**

505 To determine the stability of the examined ligand-DNA complexes in aqueous solution, molecular
506 dynamic (MD) simulations were performed. Root Mean Square Deviation (RMSD) was used to
507 identify the structural stability of **5**, **6** and **7** complexes with both DNA duplexes,
508 d(CGACTAGTCG)₂ and d(CGAGTACTCG)₂. For the d(CGACTAGTCG)₂ duplex, the whole
509 system was found to be stable during the 20 ns simulation (Figure 8A). Furthermore, it was clearly
510 noted that complexes **5** and **6** exhibited more stable average RMSDs of 0.19 nm and 0.25 nm
511 respectively, and no significant changes in RMSD occurred during the whole run (Figure 8C). By
512 contrast, complex **7** showed an average RMSD of 0.36 nm with a substantial change from 0.2 nm
513 to almost 0.5 nm after 8 ns of simulation until the end of the run (Figure 8C). This indicates the lack



514

515 **Figure 8:** **A)** RMSD for the whole system including DNA (ACTAGT sequence), ligand, ions and
 516 solvent molecules. **B)** RMSD of the DNA duplex only. **C)** RMSD of the ligand-DNA complex. **D)**
 517 RMSD of the ligand only. **E)** Solvent accessible surface area for the ligand-DNA complex. **F)**
 518 Radius of gyration for the ligand-DNA complex. **G)** RMSF for complexed DNA. **H)** RMSF for the
 519 complexed ligand.

520

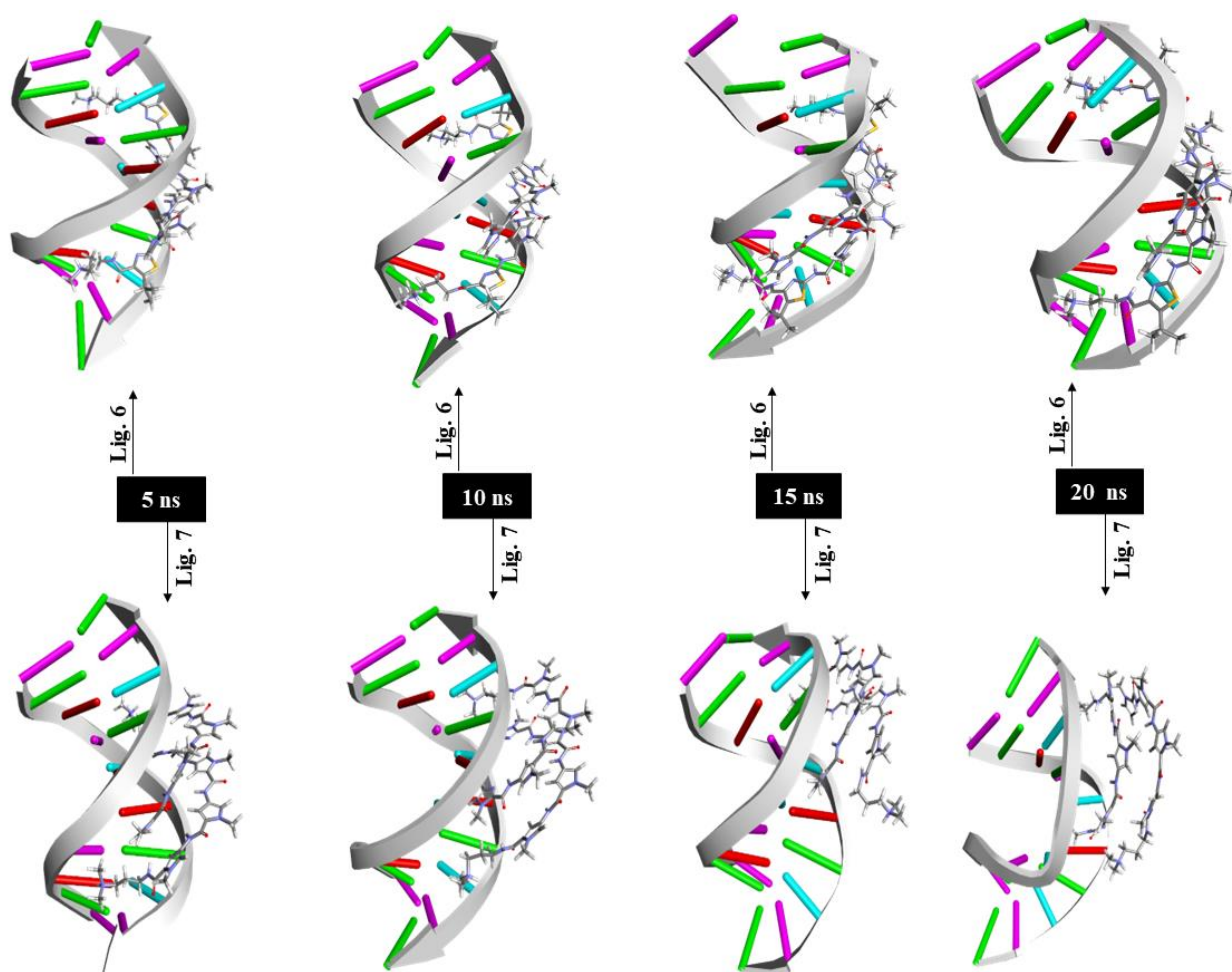
521 of stability of **7** complexed with d(CGACTAGTCG)₂. Root Mean Square Fluctuations (RMSFs),
522 which indicate the mobility of the DNA or ligand atoms during the MD simulation were calculated
523 for the three complexes. It was apparent that the majority of the DNA atoms in the central region of
524 the DNA sequence (the binding region) for the three complexes exhibited RMSF below 0.2 nm with
525 an average of 0.12 nm, 0.13 nm and 0.19 nm for **5**, **6** and **7** respectively (Figure 8 G), while the
526 atoms of the terminal DNA bases showed the highest fluctuation values as these atoms are remote
527 from the central binding site. The fluctuation of the ligands' atoms was found in the ranges of 0.05-
528 0.2 nm, 0.1-0.3 nm and 0.2-0.5 nm for ligands **5**, **6**, and **7** respectively (Figure 8 H). The majority
529 of the ligands' atoms for **5** and **6** showed fluctuations less than 0.2 nm with an average of 0.1 nm
530 and 0.18 nm respectively. However, **7** had higher fluctuations with an average of 0.4 nm, which
531 indicate a weaker and unstable binding with DNA as the binding is expected to rigidify the complex
532 and decrease the flexibility of the bound ligand.

533 Radius of gyration (Rg), which evaluates the compactness of ligand-DNA association, was
534 calculated for the three complexes. Ligands **5** and **6** showed steady Rg scores with an average of
535 11.65 Å and 11.76 Å respectively. However, ligand **7** exhibited more fluctuations starting from 11.8
536 Å reaching to 13.25 Å (Figure 8 D), which also indicates a lower stability for the association of **7**
537 with 5'-ACTAGT-3' compared to **5** and **6**. Furthermore, Solvent Accessible Surface Area (SASA),
538 which measures the surface area of the ligand-DNA complex that is accessible to the solvent
539 molecules, was determined. Low SASA values indicate a more compact and stable complex. Ligand
540 **5** and **6** have lower SASA values, averaged at 4,216 nm² and 4,271 nm² respectively, than **7**, which
541 has an average of 4,599 nm² (Figure 8E). Eventually, **5** and **6** complexes showed more stable
542 RMSDs, RMSFs, SASAs and Rgs scores during 20 ns MD simulation compared to complex **7**,
543 which showed less stability.

544 A similar approach was followed to study the stability of **5**, **6** and **7** complexations with d-(5'-
545 CGAGTACTCG-3')₂ using Molecular dynamic (MD) simulation *via* the determination of RMSDs,

546 RMSFs, SASAs and Rgs scores (see Supporting Information S10 for the correlated data sets to
547 those shown in Figure 9). The results of all these four descriptors indicated the stability of the
548 investigated MGB complexes including ligand 7 with d-(5'-CGAGTACTCG-3')₂ in contrast to the
549 complex between ligand 7 and d-(5'-CGACTAGTCG-3')₂, which showed a lack of stability during
550 the course of 20 ns MD simulations (Figure 9). These findings are consistent with the experimental
551 NMR results.

552



553

554 **Figure 9.** Simulated structures of the complex between ligand 7 and d(5'-CGACTAGTCG-3')₂ (the
555 lower row) and the complex of ligand 6 with d-(5'-CGACTAGTCG-3')₂ (the upper row) during the
556 course of 20 ns MD simulation (i.e., at 5 ns, 10 ns, 15 ns, and 20 ns). After 10 ns, the structure of
557 ligand 7 seems to be distorted and the ligand dimer is separated from the DNA duplex structure,
558 contrary to ligand 6, which remained bound to the minor groove of DNA. The structural
559 representations were generated using BIOVIA Discovery Studio Visualizer (2021).

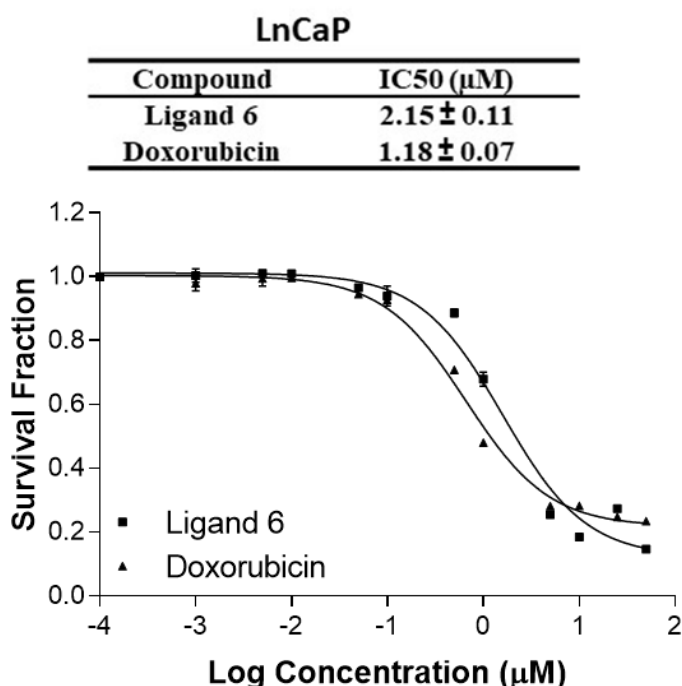
560

561

562 **2.8 Rationale of the ligands' design and their mechanism of action**

563 The MGBs formulated in this study are analogs of distamycin and exhibit three notable features.
564 Firstly, their structures are curved, enabling them to match the curvature of the minor groove.
565 Secondly, they form specific hydrogen bonds through the amide NH groups and the N of
566 thiazole/imidazole, with the DNA bases in the floor of the minor groove (Figure 7). Lastly, they are
567 monocationic, facilitating the binding of two molecules in an antiparallel orientation to prevent
568 electrostatic repulsions. These MGB compounds were designed with the purpose of targeting the
569 DNA response element, 5'-WGWWCW-3' (where W represents either A or T), for several nuclear
570 receptors to disrupt or inhibit the interaction of transcription factors with DNA. There are several
571 examples of transcription factors that bind with a DNA response element composed of six base
572 pairs, such as progesterone receptor (PR), mineralocorticoid receptor (MR), glucocorticoid
573 receptors (GR), estrogen receptor (ER), androgen receptor (AR), constitutive androstane receptor
574 (CAR), and thyroid hormone receptor (TR).⁵⁹⁻⁶³ While the DNA sequence 5'-WGWWCW-3' is
575 known in literature as the DNA response element for the androgen receptor (ARE), it's noteworthy
576 that other nuclear receptors, including glucocorticoid receptors (GR), progesterone receptor (PR),
577 and mineralocorticoid receptor (MR), also interact with the same DNA response element.^{59,64} This
578 DNA sequence is organized as repeats of the consensus DNA sequence 5'-WGWWCW-3', with a
579 three-nucleotide spacer. Malfunction of steroid receptors and their transcription factors may drive
580 cancer cell growth, proliferation, and metastasis, which is observed in many human tumours such
581 as breast and prostate cancers.⁵⁹ Remarkably, ligand **6** showed potent inhibition of breast cancer
582 cell lines that are sensitive (e.g., MCF7, T47D) and non-sensitive (e.g., MDAMB-231, MDAMB-
583 468), to the oestrogen receptor (Figure 7). Similarly, ligand **6** inhibited the androgen non-sensitive
584 prostate cancer cell lines, PC-3 and DU-145 according to the NCI screening results (Figure 6).
585 However, since there were no examples of androgen sensitive prostate cancer cells in the NCI-60
586 panel cell lines, we used the SRB cell viability assay to determine the IC₅₀ of ligand **6** against the

587 androgen-sensitive prostatic cancer cells, LNCaP. The SRB assay results showed that Ligand 6
 588 effectively suppressed LNCaP cell proliferation (IC_{50} : 2.15 μ M) (Figure 10).



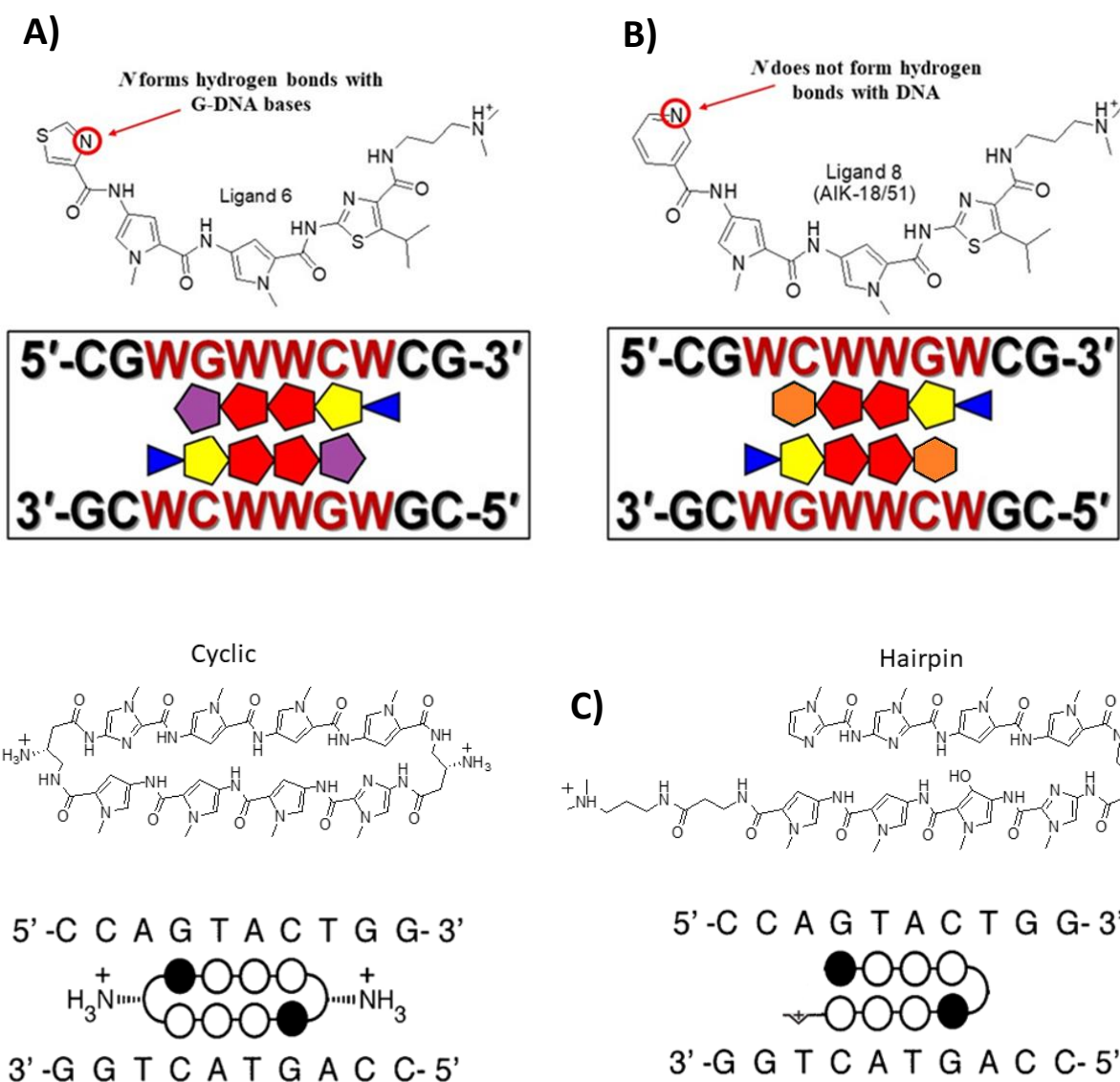
589

590 **Figure 10.** IC_{50} of ligand 6 and doxorubicin against LNCaP cells. The cell viability was determined
 591 after 48 hours using the SRB assay. Each value represents the mean \pm SEM of three independent
 592 experiments.

593

594 The design and development of novel MGBs to target specific genes requires revealing the factors
 595 that dictate preferred orientation of MGBs in the minor grooves. Ligand 6 was designed to target
 596 the 5'-WGWWCW-3' sequence and that was facilitated by the findings of our previous study, which
 597 indicated that the position of the dimethylaminopropyl group and the orientation of the amide links
 598 of the ligand with respect to the 5'-3'-ends dictate the preferred orientation of MGBs in the minor
 599 groove of DNA.⁵ Furthermore, the 3D NMR structure (PDB: 2MNE) for the complex between
 600 "AIK-18/51", 8, and $d(5'-CGACTAGTCG-3')_2$ was used to guide the design of ligand 5, 6, and 7.
 601 We speculated that replacing the pyridine head group in AIK-18/51 (ligand 8) with an imidazole or
 602 thiazole ring in the designed structures would allow the ligand to recognize a guanine base at
 603 position 2 in the binding site, 5'-W**G**WWCW-3' (Figure 11 A & B). Remarkably, the experimental

604 and molecular modelling simulations showed the ability of imidazole/thiazole nitrogen to
605 selectively recognize both G.C and C.G base pairs by forming a hydrogen bond with the exocyclic
606 amino group of guanine whether it is located in the first or second DNA strand. This study reports
607 the first MGBs with small molecular weight that can bind to the 5'-WGWWCW-3' sequence without
608 utilizing harpin or cyclic polyamide structures, which were developed by the Dervan group to target
609 the ARE sequence.^{12, 65, 66} According to the reported data from Dervan's lab, a cyclic polyamide
610 minor groove binder with 8 imidazole/pyrrole rings (Figure 11 C) that binds to the ARE (5'-
611 WGWWCW-3') can control the expression of AR target genes in studies conducted in cell culture.
612 The relatively large cyclic polyamide was found to be cell permeable in this study. However,
613 absorption tests on the substance revealed low Caco-2 permeability, indicating that it may not be
614 bioavailable orally.²⁶ According to the computational permeability tool PerMM, ligand **6** exhibits
615 55 billion times higher cell permeability than the cyclic polyamide in PAMPA model, and 933 times
616 higher permeability in Caco-2 model (Supplementary Information, **S13**). The high polarity and large
617 molecular weight of the cyclic and hairpin molecules (logP -12.03) adversely affected their cell
618 permeability.⁶⁵ The MGB compounds presented in this study, which contain only four aromatic
619 rings, possess extra advantageous physical properties due to their reduced molecular weight and
620 enhanced lipophilicity afforded by thiazole rings as in ligand **6**. Furthermore, the main advantage
621 of the described MGBs over the currently available chemotherapy lies in their ability to bind
622 reversibly in the minor groove *via* non-covalent bonds. This binding mechanism avoids the
623 induction of permanent DNA damage, a common issue with alkylating agents that create covalent
624 bonds with DNA, resulting in mutations and irreversible DNA damage. Additionally, MGBs exhibit
625 greater selectivity in their binding compared to other DNA-binding agents like intercalators and
626 alkylating agents, which are characterized by a lack of selectivity and increased toxicity. The
627 cytotoxicity of ligands 5-7 was evaluated against normal human dermal fibroblast (HDF) cells at a



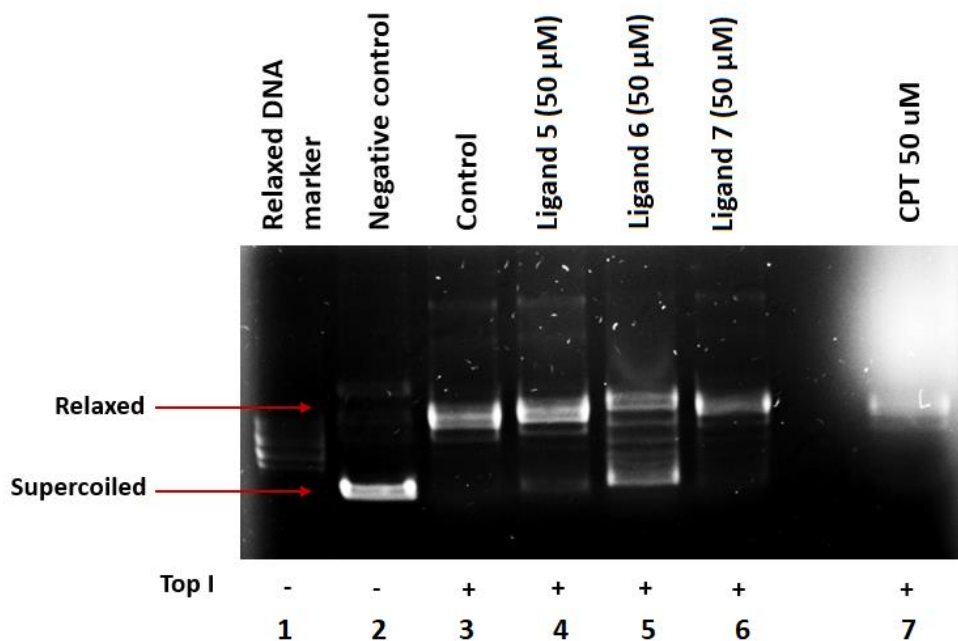
629

630 **Figure 11.** Chemical structures and schematic representation of the binding orientation adopted by
 631 **A)** ligand **6** dimer in the minor groove **A)** AIK-18/51, **8**, dimer in the minor groove ⁴. **C)** Cyclic and
 632 hairpin polyamide alignment in the minor groove relative to the DNA sequence.^{65, 66} Colour code:
 633 orange hexagon=pyridine; red pentagon=*N*-methylpyrrole; yellow pentagon=isopropyl thiazole;
 634 magenta pentagon=thiazole; blue triangle=dimethylaminopropyl “tail”.
 635

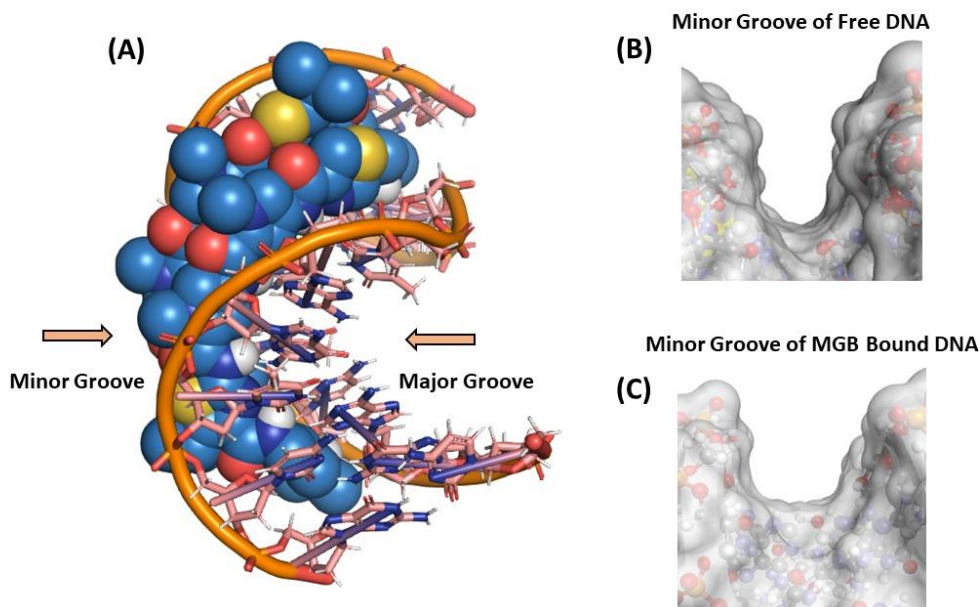
636 concentration of 50 μ M (Supplementary Information, **S14**). Ligand **7** demonstrated negligible
 637 toxicity (P-values >0.01), while ligands **5** and **6** exhibited limited toxicity, resulting in a 20-30%
 638 reduction in cell viability at the specified concentration. In accordance with the International
 639 Organization for Standardization (ISO) 10993-5 guidelines, cytotoxicity ranging from 20% to 40%
 640 is categorized as negligible (ISO, 2009). Consequently, these compounds can be deemed safe.

641 Ligand **6**, which was designed to target the 5'-WGWWCW-3' sequence, has shown potent
642 anticancer activity against androgen sensitive and non-sensitive prostate cancer cell lines.
643 Additionally, it exhibited potent activity against numerous other cancer cell lines, as illustrated in
644 Figure 6. These findings indicate that the ligand **6** might inhibit other oncogenes, transcription
645 factors or DNA binding proteins. To investigate whether our MGBs interfere with the activity of
646 DNA binding proteins such as the topoisomerase enzymes, the *in-vitro* DNA relaxation assay was
647 utilized to assess the activity of topoisomerase I. As shown in Figure 12, the supercoiled DNA was
648 completely relaxed in the presence of topoisomerase I (Lane 3) as indicated by the appearance of
649 relaxed DNA fragments. Camptothecin, a known topoisomerase I poison, inhibited topoisomerase
650 I activity *via* the formation of a nicked open circular DNA product as a result of stabilizing the
651 topoisomerase I-DNA cleavage complexes. The open circular DNA product appeared above the
652 relaxed products (Lane 7), but it is not clearly shown in the image due to Camptothecin's
653 fluorescence, which covered the area above the relaxed band. Ligand **6** showed an inhibitory effect
654 on topoisomerase I activity as indicated by the appearance of a supercoiled DNA band (Lane 5). A
655 minor inhibition effect was observed with ligand **5** (Lane 4), while no effect was detected for ligand
656 **7** (Lane 6). These findings are consistent with the results of growth inhibition of the NCI-60 panel
657 cell lines, which showed weak, potent, and negligible inhibition for ligand **5**, **6**, and **7**, respectively
658 (Figure 6). The topoisomerase I assay results are remarkably consistent with several studies in
659 literature. These studies have indicated that thiazole-containing minor groove binders (e.g.,
660 compounds similar to ligand **6**) can effectively inhibit topoisomerase I enzyme,⁶⁷⁻⁶⁹ which plays a
661 pivotal role in relaxing DNA supercoils during DNA replication. This inhibition has the potential
662 to disrupt both replication and transcription processes leading to various substantial biological
663 effects including cell death. The results of the topoisomerase I inhibition assay are consistent with
664 the DNA binding characteristics of minor groove binders.

665



666
 667 **Figure 12.** Supercoiled DNA relaxation by topoisomerase I in the presence of 50 μM ligands **5-7**
 668 or camptothecin. The reaction products were separated by agarose gel electrophoresis. The relaxed
 669 DNA (Top I activity) was indicated upward of the supercoiled DNA. Lanes 1 and 2: relaxed and
 670 supercoiled control DNA, respectively. Lane 3: supercoiled DNA with topoisomerase I and the
 671 vehicle. Lanes 4, 5, 6 and 7: supercoiled DNA with topoisomerase I in the presence of ligands **5, 6,**
 672 **7** and camptothecin, respectively.



683 **Figure 13.** **A)** 3D structure of the complex between ligand 6 (CPK) and $d(\text{CGAGTACTCG})_2$ (sticks
 684 and tubes) showing MGB-induced widening of the minor groove and bending of the DNA helix
 685 generating a compressed major groove. **B)** The width of the minor groove (11.7 \AA) for free DNA
 686 (narrow and deep). **C)** The width of the minor groove (15 \AA) for the complex between ligand **6** and
 687 $d(\text{CGAGTACTCG})_2$, (wide and shallow).

688 Another potential mechanism for DNA minor groove binders reported in the literature is the
689 allosteric modulation of transcription factors binding site⁶⁵ through the MGB induced widening of
690 the minor groove and compression of the major groove, which blocks the transcription factor from
691 recognizing their target in the major groove of DNA. To assess this hypothesis, the minor groove
692 width for both the free and complexed DNA structure with ligand **6** was calculated using the 3D
693 structures produced from the molecular docking studies (Figure 13). The minor groove width for
694 the free and complexed DNA with ligand **6** were found to be 11.70 Å and 15.00 Å respectively
695 (Figure 13 B & C). This widening of the minor groove twists the DNA structure towards the major
696 groove generating a compressed major groove (Figure 13 A). These structural alterations in the
697 DNA helix explain the molecular basis for the allosteric perturbations of transcription factor-DNA
698 interfaces by the MGB compounds. The findings related to ligand **6** indicate that the broad-
699 spectrum antiproliferative activities of this molecule might originate from its ability to target
700 multiple nucleic acid sequences and promoter sites of different oncogenes and hence might inhibit
701 various DNA binding proteins and transcription factors. Small molecules that bind to DNA can
702 generate significant biological effects and induce profound inhibition of different proteins as a result
703 of altering or disrupting the DNA structure or function.^{11, 70-72} Our study clearly indicates that the
704 observed multitarget activity of MGBs arises from their intervention at nucleic acids level. This
705 multi-target activity could be beneficial for the treatment of multifactorial diseases such as cancer.
706

707 **3. Conclusions**

708 In this study, we utilized a variety of techniques including spectroscopic (NMR spectroscopy and
709 DNA UV melting), computational (molecular docking and MD simulations) and biochemical
710 methods (NCI screening, SRB and topoisomerases I assays) to characterize the interaction of novel
711 MGBs with DNA and to gain insight on their mechanism of action. NMR spectroscopy was
712 effectively used to confirm the binding of the investigated MGBs with short well-defined DNA

713 sequences before proceeding to advanced NMR experiments, which is beneficial in screening
714 studies. The NMR results were consistent with the DNA UV melting and the molecular modeling
715 results, which confirmed the binding with different extents to short DNA sequences. The modeling
716 results showed the ability of the investigated MGBs to fit snugly in the minor groove as antiparallel
717 dimers *via* hydrogen bonding with the DNA bases and electrostatic interaction with backbone of
718 DNA. The NMR and modeling data indicated that the thiazole and imidazole rings can recognize
719 both G.C and C.G base-pairs, while the amide links of the ligand can bind to AT or TA bases pairs.
720 The ability of MGBs to target multiple DNA sequences including the promoter site (5'-
721 WGWWCW-3') of several nuclear receptors, may explain the observed broad spectrum anticancer
722 activity of ligand **6**. Disruption of DNA structure and function *via* the allosteric inhibition of
723 transcription factors and other DNA binding proteins (e.g., topoisomerase enzymes) can block the
724 replication and transcription processes leading to significant biological effects such as apoptosis,
725 necrosis, cell death and inhibition of protein synthesis^{73, 74}. Targeting the promoter sites of multiple
726 oncogenes, transcription factors or DNA binding proteins represents the potential mechanisms of
727 action of the MGB compounds at the molecular level. This multi-target activity could be
728 advantageous in the treatment of multifactorial diseases such as cancer.⁷⁵⁻⁷⁷ Although modern
729 approaches to drug discovery focus on the design of selective drugs to avoid side effects, treatment
730 of complex diseases requires the design of drugs that can interact with numerous biological targets.
731 This study presents the first small-molecular-weight MGBs that bind to the 5'-WGWWCW-3'
732 sequence without utilizing the high molecular weight, harpin or cyclic polyamide scaffolds. The
733 molecular weight of the MGBs introduced in the current study is approximately half that of the
734 hairpin and cyclic polyamide molecules developed by Dervan group to target the 5'-WGWWCW-
735 3' sequence. Whereas these compounds consist of 8 heterocyclic rings, our MGBs are comprised of
736 4 heterocyclic rings. The reported MGBs have additional beneficial physical properties due to their
737 low molecular weight, improved lipophilicity, and enhanced DNA binding selectivity toward G.C

738 base pairs provided by the thiazole rings. These features could potentially contribute to their
739 anticancer activity. Moreover, MGBs exhibit greater selectivity compared to other DNA-binding
740 agents such as intercalators and alkylating agents, which are characterized by a lack of selectivity
741 and increased toxicity. Additionally, the key advantage of MGBs over existing chemotherapy lies
742 in their ability to bind reversibly in the DNA minor groove *via* non-covalent bonds without inducing
743 permanent DNA damage, which is commonly associated with conventional chemotherapeutic
744 drugs.

745 This study indicates that the structural optimization of MGBs should prioritize enhancing the
746 ligand's lipophilicity while preserving the essential binding groups in the MGB scaffold. This
747 improvement could be achieved, for example, by introducing fluorine atoms to the aromatic rings
748 or incorporating more lipophilic tail groups like piperidine propylamine and morpholine
749 propylamine. Several cycles of structural optimization may be necessary to enhance the efficacy
750 and potency of MGBs. Ultimately, this study revealed the structural features that drive lexitropsin-
751 DNA interactions, and that could be useful to guide future design and development of novel MGBs
752 as potential therapeutic agents.

753

754 **4. Experimental Section**

755 **4.1 General**

756 The solvents and reagents used in this study were purchased from Sigma-Aldrich (Steinheim,
757 Germany). The self-complementary ODNs d(CGCACTAGTGCG)₂ and d(CGCAGTACTGCG)₂
758 were purchased from Alpha DNA Ltd. (Montreal, Canada) as desalted, cartridge-purified,
759 lyophilized powders, which were used without further purification. The chemical structures of MGB
760 intermediates and final products were confirmed by NMR spectroscopy and mass spectrometry
761 (Supplementary Information, S11).

762 The NMR Chemical shifts were measured using a Bruker 500 MHz instrument for both ^1H and ^{13}C
763 NMR experiments, and they were expressed as parts per million (ppm) compared to the internal
764 tetramethylsilane standard. The ACD NMR program (academic version)⁷⁸ was used to process
765 NMR data. The NMR data were displayed as follows: dd (double doublet), q (quartet), m
766 (multiplet), s (singlet), or bs (broad singlet), J, the coupling constant, is given in hertz (Hz).
767 Electrospray ionization-mass spectrometry (ESI-MS) experiments were performed for the
768 intermediate and final products using a triple-quadrupole tandem mass spectrometer (MicromassW
769 Quattro microTM Waters Corp., Milford, MA, United States) with an electrospray ionization
770 interface. The reactions were monitored using thin-layer chromatography (TLC) silica gel plates
771 (0.25 mm, E. Merck, 60 F254) and visualized under UV lamp. The melting points were measured
772 using a STUART SMP50 melting point apparatus. HRMS spectra of the final products were
773 acquired utilizing ESI timsTOF Pro instrument (Bruker Daltonics).

774 The final MGB products, **5-7** were prepared as TFA salts and purified by reverse-phase HPLC using
775 Agilent 1260 infinity liquid chromatography (Agilent Technologies, Palo Alto, CA, U.S.A.). All
776 compounds tested in this study are > 95% pure by HPLC analysis (see Supplementary Information,
777 S12). The C18 column used in HPLC (Zorbax-SB-C18) has a length of 25 cm, an internal diameter
778 of 9.4 mm and particle size of 5 μm . The HPLC gradient elution method used for separation and
779 purification is described below.

Time (mins)	Flow rate (ml/min)	% Water (0.1% TFA)	% Acetonitrile (0.1% TFA)
0	4	90	10
25	4	50	50
26	4	0	100
30	4	0	100
31	4	90	10
35	4	90	10

780

781

782

783 4.2 NMR Spectroscopy

784 Samples of ligands **5-7** were separately prepared as DMSO-*d*₆ solutions at concentrations of 10 mM
785 in a final volume of 550 μ L. Full NMR signal assignment of ¹H, ¹³C and ¹⁵N resonances for each
786 compound was carried out using classical NMR procedures for small molecule structure elucidation
787 (including acquisition and interpretation of 2D [¹H, ¹³C] and [¹H, ¹⁵N] HSQC and HMBC alongside
788 2D [¹H, ¹H] COSY, TOCSY and NOESY NMR data acquired at 600 MHz) to confirm the structures
789 and integrity of each compound (See Supporting Information Table S4 for details). DNA samples
790 were prepared as 1 mM stock solutions in 50 mM phosphate buffer at pH = 7.4 made with 90% H₂O
791 and 10% D₂O to provide a deuterium frequency lock. Ligands were prepared as concentrated
792 solutions in volumes of 50 μ L H₂O such that upon addition to the DNA NMR sample (550 μ L
793 volume admitted to a 5 mm diameter high-precision NMR tube) their final molar ratios would match
794 those of the DNA for ligand:DNA single strand ratios of 1:1 (e.g. 2:1 ligand:duplex ratio), thereby
795 taking into account dilution effects.

796 NMR spectra were acquired on a Bruker 600 MHz AVANCE II⁺ NMR spectrometer equipped with
797 a TBI [¹H, ¹³C, X]-z probe-head fitted with an actively shielded gradient coil. 1D ¹H NMR spectra
798 were acquired using excitation sculpting for efficient solvent signal suppression using Bruker pulse
799 program zgpg30. Spectra were centred at $\delta^1\text{H} = 4.694$ ppm (offset = 2816.74 Hz) and set for a
800 frequency width of 20.557 ppm (12335.53 Hz). Data were acquired with 64 transients and 4 dummy
801 transients using a data acquisition time of 1.328 s and a recycle time of 2.0 s in 32K data points.
802 Excitation sculpting was achieved using a sinc shaped pulse of duration 2 ms at a power of 2.49
803 mW and sets of paired z-pulsed field gradients with values of 31% and 11% of the maximum 50
804 G/cm value applied during periods of 1 ms each. 1D ³¹P-¹H NMR spectra were acquired using a
805 power-gated pulse sequence, Bruker pulse program zgpg30. Spectra were centred at $\delta^{31}\text{P} = 0$ ppm
806 with ¹H decoupler offset set to $\delta^1\text{H} = 4$ ppm. Data were acquired with 256 transients into 2K data
807 points over a frequency width of 10 ppm (2429.54 Hz) for an acquisition time of 421 ms using a

808 recycle delay of 0.5 s. Aliquots of ligand solutions (5 μ L) were added stepwise to DNA samples
809 (550 μ L starting volume). Following each addition, an initial white precipitate gave way to a clear
810 solution indicating take-up of the ligand by the DNA in the form of a complex. Both 1D ^1H and 1D
811 $^{31}\text{P}\{-^1\text{H}\}$ NMR spectra were acquired automatically using a sample changer running under
812 IconNMR prior to the first ligand aliquot addition and then subsequently after each 5 μ L ligand
813 aliquot addition. The imino-proton ^1H NMR resonance region ($\delta^1\text{H} = 12\text{-}14$ ppm) was monitored
814 for changes in the DNA after each addition along with the aromatic peptide NH resonance region
815 ($\delta^1\text{H} = 8\text{-}12$ ppm), which was used to assess the presence of ligand. Ligand aliquot additions were
816 stopped when the NMR data showed evidence of a titration end-point. Following sample
817 preparations, this whole process can be completed by a non-expert NMR user in a typical period of
818 2-3 hours.

819 **4.3 DNA UV melting studies**

820 The absorbance profiles of the free and complexed ODNs [$d(5'\text{-CGCACTAGTGCG-3}')_2$ and $d(5'\text{-}$
821 $\text{CGCAGTACTGCG-3}')_2$] at 260 nm versus temperature were measured in the range 30 – 100 $^\circ\text{C}$
822 at a heating rate of 1 $^\circ\text{C}/\text{min}$. The ligands and ODNs were dissolved in 10 mM PIPES buffer (pH
823 7.0), 20 mM NaCl, and 1 mM EDTA. Two equimolar quantities of the ligands were added to the
824 DNA solution ($[\text{DNA}] = 10$ μM). The measurements were carried out using a Shimadzu UV-2600
825 spectrophotometer equipped with S-1700 temperature control unit.

826

827 **4.4 Cell Culture and Cell Safety Assay**

828 Normal human dermal fibroblast cells (HDF, 106-05A, Sigma, EU) were cultured in complete
829 Dulbecco's Modified Eagle's Medium (DMEM, Sigma-Aldrich, USA). The cell line was obtained
830 from the European Collection of Cell Cultures (ECACC, UK). Cell line incubations were kept at
831 37 $^\circ\text{C}$ in a humidified incubator with 5% CO_2 . Cytotoxic effects of the ligands were assessed

832 through the 3-(4,5-dimethylthiazol-2)-2,5-diphenyltetrazolium bromide (MTT) assay. In summary,
833 100 μ L of the cell line was seeded at a density of 1×10^4 cells/well in a 96-well plate and incubated
834 at 37 °C with 5% CO₂ for 24 hours. Following cell adherence, the media were removed, and 100
835 μ L of the tested compound at a concentration of 50 μ M in fresh media was incubated for another
836 24 h. Subsequently, each well received 20 μ l of sterile filtered MTT reagent in PBS (5mg/mL), and
837 the mixture was incubated for 4 hours. After aspirating the aliquots, DMSO (100 μ L/well) was
838 added to dissolve the resulting formazan. DMSO (0.1%) was employed as a negative control. The
839 generated purple colour was quantified using the Multiskan Go machine (spectrophotometer) at
840 570nm. Each experiment was replicated three times, and the cell viability percentage was
841 determined using the following formula.

842 Cell viability (%)

$$843 = \frac{[(\text{OD test} - \text{OD blank}) \div (\text{OD negative control} - \text{OD blank})] \times 100$$

844 **4.5 Passive membrane permeability**

845 The PerMM web server (<https://permm.phar.umich.edu/>) was used for the evaluation of the passive
846 permeability of the investigated MGBs across a variety of membrane systems, including PAMPA,
847 BBB, and Caco-2 systems. The 3D structures of the tested MGBs 5-7 were generated using the PDB
848 structure 2mne,⁴ while the 3D structure of the cyclic polyamide ligand was obtained from the PDB
849 structure 3OMJ.⁶⁵ These 3D structures served as input files and were submitted for analysis by the
850 PerMM server to determine the passive membrane permeability, utilizing the default settings (T =
851 298 K, pH 7.4).⁵⁷

852 **4.6 Molecular Docking and Structure Refinement**

853 The starting 3D structures of the complexes between the studied MGBs (5-7) and the DNA duplexes
854 d(CGACTAGTCG)₂ & d(CGAGTACTCG)₂ were built from an NMR solution structure obtained

855 from the protein data bank (PDB code: 2MNE) for the interaction of an analogous compound
856 (AIK18/51, **8**) with d(CGACTAGTCG)₂. Discovery Studio Visualizer (2021)⁷⁹ was used to modify
857 the AIK18/51 structure to obtain the structures for ligands **5-7**. UCSF Chimera software was used
858 to generate the 3D structure of d(CGAGTACTCG)₂ from 2MNE by replacing the following DNA
859 bases: C4, G7, C14, and G17 (in the 2MNE) with G4, C7, G14, and C17 respectively. This was
860 performed using the function Swapna, which “mutates” nucleic acid residues in Chimera. The
861 generated pdb files for the studied complexes were then refined using the water refinement protocol
862 in HADDOCK webservice (version 2.4, <https://wenmr.science.uu.nl/haddock2.4/>)^{80, 81}. The
863 Haddock protocol of energy minimization for the initial structures was employed. Site-specific
864 docking was then performed to determine the best docked conformation for the three ligands with
865 the binding sites 5'-ACTAGT-3' and 5'-AGTACT-3'. The HADDOCK generated pdb structures
866 were then subjected to NAMD molecular dynamic simulations.

867

868 **4.7 Molecular dynamic simulations**

869 The interactions of ligands **5-7** with the DNA duplexes d(CGACTAGTCG)₂ &
870 d(CGAGTACTCG)₂ were studied by molecular dynamic simulations using NAMD v2.15
871 (<https://www.ks.uiuc.edu/Development/Download/download.cgi?PackageName=NAMD>).⁸²
872 Ligand atoms were parametrized by CHARMM General Force Field (CGenFF),⁸³ which performs
873 atom typing and assignment of parameters and charges by analogy in a fully automated fashion.
874 The DNA atoms were parametrized using CHARMM36 additive force fields. The molecular
875 dynamic simulation input files, including the PDB files, force field parameters, the solvated system
876 and the periodic boundary conditions, were generated using the CHARMM-GUI webservice
877 (<https://www.charmm-gui.org/>).⁸⁴ Each complex system was solvated with TIP3P solvent
878 molecules in a periodic cubic box with an average of (90 Å × 90 Å × 90 Å) and neutralized with 18
879 Na⁺ counter-ions. The long-range electrostatic interactions were calculated using the particle mesh

880 Ewald (PME)⁸⁵ method, and periodic boundary conditions with a 15 Å cutoff for non-bonded
881 interactions were applied. Lennard-Jones (LJ)⁸⁶ calculations were used to calculate the van der
882 Waals interactions. Three steps of energy minimization and equilibration were conducted: (i) Using
883 the conjugate gradient method, the potential energy of the entire system, which included DNA,
884 ligand, solvent, and ions, was minimized for 50,000 steps (100 ps). (ii) The system was heated using
885 the NVT ensemble and Langevin dynamics were used after adding constraints to the DNA and the
886 ligand. (iii) The SHAKE⁸⁷ algorithm and the NPT ensemble were employed at a constant pressure
887 of 1 bar and temperature of 300 K for 100 ps to restrict hydrogen to heavy atom bonds. NPT
888 ensemble was used to simulate the MD production runs for 20 ns for each complex at 300 K
889 temperature, 1 bar pressure, and a time step of 2 fs. Graphs were produced using the Qtgrace
890 program, and MD trajectories were displayed using the Visual Molecular Dynamics software
891 (VMD, version 1.9.4).⁸⁸

892

893 **4.8 Anticancer screening against the NCI-60 cancer cell lines**

894 Testing against the cancer cell lines was performed at the National Cancer Institute (Bethesda, MD,
895 USA) using a standard protocol.⁸⁹

896 **4.9 Cytotoxicity assay**

897 The cytotoxicity of **6** in LnCaP cells was measured by Sulforhodamine B (SRB) assay.⁹⁰ Both ligand
898 **6** and Doxorubicin powders were dissolved in sterile water. The serial dilutions to obtain the final
899 concentrations were prepared in culture medium. The cells were seeded in 96-well plates at a density
900 of 8,000 cells per well. The next day, the cells were treated with various concentrations of **6** (0.0001-
901 50µM). Doxorubicin was used as a reference drug. After 48 hours, the cells were fixed with 50%
902 trichloroacetic acid (Sigma Aldrich, Germany), washed and stained with 0.4% SRB (Sigma Aldrich,
903 Germany). The SRB dye was solubilized with 10 mM Tris base and colour intensity was measured
904 at 492 nm using a microplate reader Varioskan™ Flash (Thermo Fisher Scientific, USA). GraphPad

905 Prism 8 software (GraphPad Software, USA) was used to calculate the IC₅₀ value by sigmoidal
906 curve fitting models.

907 **4.10 Topoisomerase I relaxation assay**

908 The *in vitro* effect of ligands **5-7** on Top I activity was measured using the Human Topoisomerase
909 I Assay Kit (TopoGEN, USA) according to manufacturer's protocol.¹¹ In brief, the reaction
910 mixtures (20 µL) contained nuclease-free water, Top I reaction buffer (10 × TGS) (10 mM Tris-
911 HCl pH 7.9, 1 mM EDTA, 0.15 M NaCl, 0.1% BSA, 0.1 mM Spermidine, 5% glycerol), 0.25 µg of
912 the supercoiled plasmid DNA and 1-unit human Top I. The relaxation reaction was incubated at 37
913 °C for 30 minutes in the absence or in the presence of 50 µM of the tested ligand or Camptothecin
914 (positive control). After 30 minutes, the reaction was terminated by adding 4 µL of 5 × gel loading
915 buffer containing the loading dye (0.125% bromophenol blue, 25% glycerol, 5% Sarkosyl). The
916 DNA products were run in 1% agarose gels at 2 V/cm for 2 hours in 1X TAE buffer. After
917 electrophoresis, the gels were stained with RedSafe™ (iNtRON Biotechnology, Korea) for 30
918 minutes, de-stained with Millipore water and photographed by ChemiDoc™ imaging system
919 (Biorad, USA).

920 **4.11 Synthesis**

921 **4.11.1 Synthesis of the intermediate structures**

922 The synthesis of the intermediate structures (1a, 1b, 1e, 1d, 1f, 1g, 2a, 2d, 3a, 3b, 3c, 3d – see
923 Supplementary Information) was reported in reference^{#5, 15}.

924 **4.11.2 Synthesis of N-[3-(dimethylamino)propyl]-2- {1-methyl-4-[1-methyl-4-(1-methyl-1H- 925 imidazole-2-amido)-1H-pyrrole-2-amido]-1H-pyrrole-2-amido}-5-(propan-2-yl)-1,3-thiazole- 926 4-carboxamide (Ligand 5).**

927 Intermediate **3d** (0.136g, 0.3228 mmol) was dissolved in THF (5 ml) and cooled to 0 °C under ice
928 bath. 10%-Pd/C (0.05 g) was added slowly in small portions while the solution is stirring. The round

929 bottomed flask was emptied and refilled with hydrogen gas several times and the reaction mixture
930 was left stirring under hydrogen for 3 hours. Pd/C was then filtered and the organic solvent was
931 removed under vacuum to produce the amine product intermediate, which was dissolved in 1.5 ml
932 DMF and added dropwise to a solution of **1f** (0.077 mg, 0.3104 mmol), HBTU (0.1403 g, 0.3699
933 mmol), and triethylamine (0.185 g, 1.8271 mmol) in DMF (4 ml) and the reaction was left stirring
934 overnight at room temperature under nitrogen gas. The organic solvent was then removed under
935 vacuum and saturated sodium bicarbonate solution was added (12 ml) to the reaction mixture. The
936 crude product was extracted twice (2x20 ml) with ethylacetate and the organic layer was collected,
937 dried (over magnesium sulphate) and then evaporated under reduced pressure to obtain the crude
938 product, which was purified by HPLC (reverse-phase) using gradient elution as described in **section**
939 **4.1**. The product was then freeze-dried to obtain the final product as white solid. Yield (51%); MP:
940 202-204 °C; ¹H NMR (ACETIC ACID-d₄) δ 1.32 (6H, d, CH₃), 2.12 (2H, m, CH₂), 3.02 (6H, s,
941 NCH₃), 3.31 (2H, t, NCH₂), 3.49 (2H, q, CONH-CH₂), 3.97 (3H, s, NCH₃), 4.00 (3H, s, NCH₃),
942 4.11 (3H, s, NCH₃), 4.33 (1H, m, Ar-CH), 7.06 (1H, d, Ar-H), 7.08 (1H, d, Ar-H), 7.32 (1H, d, Ar-
943 H), 7.36 (1H, d, Ar-H), 7.38 (1H, d, Ar-H), 7.48 (1H, d, Ar-H); ¹³C NMR (ACETIC ACID-d₄): δ
944 15.31, 17.70, 26.29, 26.97, 27.33, 33.34, 46.08, 95.07, 97.48, 97.52, 109.75, 112.42, 114.42, 117.21,
945 126.71, 129.48, 137.78, 146.06, 149.82. HRMS (ESI): m/z calcd for C₂₉H₃₈N₁₀O₄S, 623.2876,
946 found 623.2880 [M + H]⁺.

947 **4.11.3 Synthesis of N-[3-(dimethylamino)propyl]-2-{1-methyl-4-[1-methyl-4-(1,3-thiazole-4-**
948 **amido)-1H-pyrrole-2-amido]-1H-pyrrole-2-amido}-5-(propan-2-yl)-1,3-thiazole-4-**
949 **carboxamide (Ligand 6)**

950 The synthetic procedure for preparing ligand 5 was applied in the synthesis of ligand 6. In brief, the
951 nitro intermediate 3d was reduced by catalytic hydrogenation then coupled directly with the acid 1g
952 using HBTU and triethylamine. The final product was purified by reverse phase HPLC, freeze dried,
953 and obtained as white solid. Yield (55 %); MP: 195-197 °C; ¹H NMR (ACETIC ACID-d₄) δ 1.31

954 (6H, d, CH₃), 2.12 (2H, m, CH₂), 3.03 (6H, s, NCH₃), 3.31 (2H, t, NCH₂), 3.48 (2H, q, CONH--
955 CH₂), 3.96 (3H, s, NCH₃), 3.98 (3H, s, NCH₃), 4.35 (1H, m, Ar-CH), 7.14 (1H, d, Ar-H), 7.42 (1H,
956 d, Ar-H), 7.42 (1H, d, Ar-H), 7.47 (1H, d, Ar-H), 8.33 (1H, d, Ar-H), 9.11(1H, d, Ar-H), 9.51 (1H,
957 s, CONH), 9.69 (1H, s, CONH). ¹³C NMR (ACETIC ACID-d₄): δ 15.15, 17.57, 26.03, 26.78, 27.14,
958 36.88, 45.94, 95.14, 97.43, 106.05, 108.36, 111.36, 112.64, 114.34, 126.48, 137.71, 142.23, 144.96,
959 148.56, 149.77, 150.31, 154.37. HRMS (ESI): m/z calcd for C₂₈H₃₅N₉O₄S₂, 626.2332, found
960 626.2342 [M + H]⁺.

961 **4.11.4 Synthesis of N-[5-({5-[(5-{[3-(dimethylamino)propyl] carbamoyl}-1-methyl-1H-pyrrol-**
962 **3-yl) carba moyl]-1-methyl-1H-pyrrol-3-yl}carbamoyl)-1-methyl-1H-pyrrol-3-yl]-1-methyl-**
963 **1H-imidazole-2-carboxamide (Ligand 7)**

964 The synthetic procedure for preparing ligand 5 above was applied in the synthesis of ligand 7. In
965 brief, the nitro intermediate 2d was reduced by catalytic hydrogenation then coupled directly with
966 the acid 1f using HBTU and triethylamine. The product was purified by reverse phase HPLC, freeze
967 dried, and the final product was obtained as yellow solid. Yield (39 %); MP: 210-212 °C; ¹H NMR
968 (ACETIC ACID-d₄) δ 2.12 (2H, m, CH₂), 3.04 (6H, s, NCH₃), 3.33 (2H, t, NCH₂), 3.36 (6H, s,
969 NCH₃), 3.49 (2H, q, CONH--CH₂), 3.90 (3H, s, NCH₃), 3.93 (3H, s, NCH₃), 3.96 (3H, s, NCH₃),
970 4.12 (1H, s, NH) 6.94 (1H, d, Ar-H), 6.98 (1H, d, Ar-H), 7.05 (1H, d, Ar-H), 7.22 (1H, d, Ar-H),
971 7.27 (1H, d, Ar-H), 7.33 (1H, d, Ar-H), 7.84 (1H, s, CONH), 9.27 (1H, s, CONH), 9.33 (1H, s,
972 CONH). ¹³C NMR (ACETIC ACID-d₄): δ 16.22, 20.82, 20.91, 26.31, 26.53, 26.93, 27.02, 27.11,
973 33.41, 46.03, 95.01, 95.39, 109.78, 112.64, 113.77, 113.96, 114.33, 114.78, 116.54, 117.43, 145.88,
974 149.79, 149.93, 154.02. HRMS (ESI): m/z calcd for C₂₇H₃₈N₁₀O₄, 576.2999, found 577.3000 [M +
975 H]⁺.

976 **ASSOCIATED CONTENT**

977 Supplementary Information is available online.

978 **AUTHOR INFORMATION**

979 **Corresponding Author**

980 *Phone: +(971)-65057427. Fax: +(971)-6-5585812. E-mail: halniss@sharjah.ac.ae

981 *Phone: +(44)-1415482820. E-mail: john.parkinson@strath.ac.uk

982 **ORCID:**

983 Hasan Y. Alniss:0000-0001-8639-953

984 Wafaa S. Ramadan: 0000-0002-7334-3093

985 Yousef A. Msallam: 0000-0002-8760-1508

986 Bekir C. Celikkaya: 0000-0002-8677-9101

987 Fraser J. Scott: 0000-0003-0229-3698

988 Raafat El-Awady: 0000-0002-7046-8587

989 John A. Parkinson:0000-0003-4270-6135

990 **Conflict of interest**

991 Hasan Alniss has a patent pending for compound **6**. The remaining authors have no conflicts of
992 interest to declare.

993 **ACKNOWLEDGEMENTS**

994 We would like to thank the National Cancer Institute (NCI), Bethesda, Maryland, USA, for
995 conducting the *in vitro* anticancer screening using a panel of 60 cancer cell lines.

996

997

998 **FUNDING**

999 This work was supported by a grant from the Research Funding Department, University of Sharjah-
1000 UAE (grant number: 2101110149 & 2401110195) and a doctoral scholarship grant by the Republic
1001 of Turkey Ministry of National Education. NMR studies were supported through UKRI EPSRC
1002 funding for the Scottish High Field NMR Facility under Grant EP/R030065/1.

1003 **Abbreviations used**

1004 A, Adenine; ARE, Androgen Response Element; C, Cytocine; DCM, dichloromethane; DMSO,
1005 dimethylsulfoxide; DMF, dimethylformamide; G, Guanine; HBTU, 2-(1H-benzotriazol-1-yl)-
1006 1,1,3,3-tetramethyluronium hexafluorophosphate, Hexafluorophosphate Benzotriazole Tetramethyl
1007 Uronium; NCI, National cancer Institute; Im, Imidazole; MeOH, methanol; MGB, minor groove
1008 binder; NMR, Nuclear Magnetic Resonance; LNCaP, epithelial cell line derived from a human
1009 prostate carcinoma; T, Thymine; TEA, trimethylamine; THF, tetrahydrofuran; Thz, thiazole; Pd/C,
1010 palladium on carbon; LogP, partition coefficient; LogD, partition coefficient at specific pH.

1011

1012 **References:**

- 1013 1. Liu, B.; Bashkin, J. K.; Poon, G. M. K.; Wang, S.; Wang, S.; Wilson, W. D., Modulating
1014 DNA by polyamides to regulate transcription factor PU.1-DNA binding interactions. *Biochimie*
1015 **2019**, *167*, 1-11.
- 1016 2. Alniss, H. Y., Thermodynamics of DNA Minor Groove Binders. *Journal of medicinal*
1017 *chemistry* **2019**, *62* (2), 385-402.
- 1018 3. Alniss, H. Y.; Anthony, N. G.; Khalaf, A. I.; MacKay, S. P.; Suckling, C. J.; Waigh, R.
1019 D.; Wheate, N. J.; Parkinson, J. A., Rationalising sequence selection by ligand assemblies in the
1020 DNA minor groove: the case for thiazotropsin A. *Chemical Science* **2012**, *3*, 711-722.
- 1021 4. Alniss, H. Y.; Salvia, M. V.; Sadikov, M.; Golovchenko, I.; Anthony, N. G.; Khalaf, A.
1022 I.; MacKay, S. P.; Suckling, C. J.; Parkinson, J. A., Recognition of the DNA minor groove by
1023 thiazotropsin analogues. *Chembiochem* **2014**, *15* (13), 1978-90.

- 1024 5. Alniss, H. Y.; Witzel, II; Semreen, M. H.; Panda, P. K.; Mishra, Y. K.; Ahuja, R.;
1025 Parkinson, J. A., Investigation of the Factors That Dictate the Preferred Orientation of Lexitropsins
1026 in the Minor Groove of DNA. *Journal of medicinal chemistry* **2019**, *62* (22), 10423-10440.
- 1027 6. Wittayanarakul, K.; Anthony, N. G.; Treesuwan, W.; Hannongbua, S.; Alniss, H.; Khalaf,
1028 A. I.; Suckling, C. J.; Parkinson, J. A.; Mackay, S. P., Ranking ligand affinity for the DNA minor
1029 groove by experiment and simulation. *ACS medicinal chemistry letters* **2010**, *1* (8), 376-80.
- 1030 7. Padroni, G.; Withers, J. M.; Taladriz-Sender, A.; Reichenbach, L. F.; Parkinson, J. A.;
1031 Burley, G. A., Sequence-Selective Minor Groove Recognition of a DNA Duplex Containing
1032 Synthetic Genetic Components. *J Am Chem Soc* **2019**, *141* (24), 9555-9563.
- 1033 8. Aman, K.; Padroni, G.; Parkinson, J. A.; Welte, T.; Burley, G. A., Structural and Kinetic
1034 Profiling of Allosteric Modulation of Duplex DNA Induced by DNA-Binding Polyamide
1035 Analogues. *Chemistry* **2019**, *25* (11), 2757-2763.
- 1036 9. He, G.; Vasilieva, E.; Harris, G. D., Jr.; Koeller, K. J.; Bashkin, J. K.; Dupureur, C. M.,
1037 Binding studies of a large antiviral polyamide to a natural HPV sequence. *Biochimie* **2014**, *102*, 83-
1038 91.
- 1039 10. Finn, P. B.; Bhimsaria, D.; Ali, A.; Eguchi, A.; Ansari, A. Z.; Dervan, P. B., Single
1040 position substitution of hairpin pyrrole-imidazole polyamides imparts distinct DNA-binding
1041 profiles across the human genome. *PLoS One* **2020**, *15* (12), e0243905.
- 1042 11. Alniss, H. Y.; Chu, C.; Ramadan, W. S.; Msallam, Y. A.; Srinivasulu, V.; El-Awady, R.;
1043 Macgregor, R. B., Jr.; Al-Tel, T. H., Interaction of an anticancer benzopyrane derivative with DNA:
1044 Biophysical, biochemical, and molecular modeling studies. *Biochim Biophys Acta Gen Subj* **2023**,
1045 *1867* (6), 130347.
- 1046 12. Kurmis, A. A.; Yang, F.; Welch, T. R.; Nickols, N. G.; Dervan, P. B., A Pyrrole-Imidazole
1047 Polyamide Is Active against Enzalutamide-Resistant Prostate Cancer. *Cancer Res* **2017**, *77* (9),
1048 2207-2212.
- 1049 13. Scott, F. J.; Puig-Sellart, M.; Khalaf, A. I.; Henderson, C. J.; Westrop, G.; Watson, D.
1050 G.; Carter, K.; Grant, M. H.; Suckling, C. J., An evaluation of Minor Groove Binders as anti-lung
1051 cancer therapeutics. *Bioorg Med Chem Lett* **2016**, *26* (15), 3478-86.
- 1052 14. Barrett, M. P.; Gemmell, C. G.; Suckling, C. J., Minor groove binders as anti-infective
1053 agents. *Pharmacology & therapeutics* **2013**, *139* (1), 12-23.
- 1054 15. Alniss, H. Y.; Khan, N. A.; Boghossian, A.; Akbar, N.; Al-Jubeh, H. M.; Msallam, Y.
1055 A.; Saeed, B. Q.; Siddiqui, R., Synthesis and Evaluation of Novel DNA Minor Groove Binders as
1056 Antiamoebic Agents. *Antibiotics* **2022**, *11* (7).

- 1057 16. Suckling, C. J.; Hunter, I. S.; Scott, F. J., Multitargeted anti-infective drugs: resilience to
1058 resistance in the antimicrobial resistance era. *Future drug discovery* **2022**, *4* (1), FDD73.
- 1059 17. MGB Biopharma. MGB Biopharma Announces Successful End-of-Phase II meeting with
1060 FDA for MGB-BP-3. . <https://bit.ly/3b6JdHe>.
- 1061 18. Suckling, C., From multiply active natural product to candidate drug? Antibacterial (and
1062 other) minor groove binders for DNA. *Future medicinal chemistry* **2012**, *4* (8), 971-89.
- 1063 19. Anthony, N. G.; Fox, K. R.; Johnston, B. F.; Khalaf, A. I.; Mackay, S. P.; McGroarty, I.
1064 S.; Parkinson, J. A.; Skellern, G. G.; Suckling, C. J.; Waigh, R. D., DNA binding of a short
1065 lexitropsin. *Bioorg Med Chem Lett* **2004**, *14* (5), 1353-6.
- 1066 20. Anthony, N. G.; Johnston, B. F.; Khalaf, A. I.; MacKay, S. P.; Parkinson, J. A.; Suckling,
1067 C. J.; Waigh, R. D., Short lexitropsin that recognizes the DNA minor groove at 5'-ACTAGT-3':
1068 understanding the role of isopropyl-thiazole. *Journal of the American Chemical Society* **2004**, *126*
1069 (36), 11338-49.
- 1070 21. Padroni, G.; Parkinson, J. A.; Fox, K. R.; Burley, G. A., Structural basis of DNA duplex
1071 distortion induced by thiazole-containing hairpin polyamides. *Nucleic Acids Res* **2018**, *46* (1), 42-
1072 53.
- 1073 22. Parkinson, J.; Khalaf, A.; Anthony, N.; MacKay, S.; Suckling, C.; Waigh, R., Comparison
1074 of DNA complex formation behaviour for two closely related lexitropsin analogues. *Helv. Chim.*
1075 *Acta* **2009**, (92), 795-822.
- 1076 23. Yang, F.; Nickols, N. G.; Li, B. C.; Szablowski, J. O.; Hamilton, S. R.; Meier, J. L.;
1077 Wang, C. M.; Dervan, P. B., Animal toxicity of hairpin pyrrole-imidazole polyamides varies with
1078 the turn unit. *J Med Chem* **2013**, *56* (18), 7449-57.
- 1079 24. Yang, F.; Nickols, N. G.; Li, B. C.; Marinov, G. K.; Said, J. W.; Dervan, P. B., Antitumor
1080 activity of a pyrrole-imidazole polyamide. *Proc Natl Acad Sci U S A* **2013**, *110* (5), 1863-8.
- 1081 25. Salvia, M. V.; Addison, F.; Alniss, H. Y.; Buurma, N. J.; Khalaf, A. I.; Mackay, S. P.;
1082 Anthony, N. G.; Suckling, C. J.; Evstigneev, M. P.; Santiago, A. H.; Waigh, R. D.; Parkinson, J.
1083 A., Thiazotropsin aggregation and its relationship to molecular recognition in the DNA minor
1084 groove. *Biophys Chem* **2013**, *179*, 1-11.
- 1085 26. Chenoweth, D. M.; Harki, D. A.; Phillips, J. W.; Dose, C.; Dervan, P. B., Cyclic pyrrole-
1086 imidazole polyamides targeted to the androgen response element. *J Am Chem Soc* **2009**, *131* (20),
1087 7182-8.
- 1088 27. Alniss, H. Y.; Anthony, N.; Khalaf, A. I.; Mackay, S. P.; Suckling, C. J.; Waigh, R. D.;
1089 Wheatea, N. J.; Parkinson, J. A., Rationalising sequence selection by ligand assemblies in the DNA
1090 minor groove: the case for thiazotropsin A. *Chem Sci* **2012**, *3* (3), 711-722.

- 1091 28. Lepre, C. A.; Moore, J. M.; Peng, J. W., Theory and applications of NMR-based screening
1092 in pharmaceutical research. *Chem Rev* **2004**, *104* (8), 3641-76.
- 1093 29. Tapley, T. L.; Cupp-Vickery, J. R.; Vickery, L. E., Structural determinants of HscA peptide-
1094 binding specificity. *Biochemistry* **2006**, *45* (26), 8058-66.
- 1095 30. Lou, Y. C.; Wei, S. Y.; Rajasekaran, M.; Chou, C. C.; Hsu, H. M.; Tai, J. H.; Chen, C.,
1096 NMR structural analysis of DNA recognition by a novel Myb1 DNA-binding domain in the
1097 protozoan parasite *Trichomonas vaginalis*. *Nucleic Acids Res* **2009**, *37* (7), 2381-94.
- 1098 31. Hannah, K. C.; Gil, R. R.; Armitage, B. A., ¹H NMR and optical spectroscopic investigation
1099 of the sequence-dependent dimerization of a symmetrical cyanine dye in the DNA minor groove.
1100 *Biochemistry* **2005**, *44* (48), 15924-9.
- 1101 32. Reibarkh, M.; Malia, T. J.; Wagner, G., NMR distinction of single- and multiple-mode
1102 binding of small-molecule protein ligands. *J Am Chem Soc* **2006**, *128* (7), 2160-1.
- 1103 33. Pelton, J. G.; Wemmer, D. E., Structural characterization of a 2:1 distamycin
1104 A.d(CGCAAATTGGC) complex by two-dimensional NMR. *Proc Natl Acad Sci U S A* **1989**, *86*
1105 (15), 5723-7.
- 1106 34. Zhang, Q.; Dwyer, T. J.; Tsui, V.; Case, D. A.; Cho, J.; Dervan, P. B.; Wemmer, D. E.,
1107 NMR structure of a cyclic polyamide-DNA complex. *J Am Chem Soc* **2004**, *126* (25), 7958-66.
- 1108 35. Pons, M.; Millet, O., Dynamic NMR studies of supramolecular complexes. *Progress in*
1109 *Nuclear Magnetic Resonance Spectroscopy* **2001**, *38* (4), 267-324.
- 1110 36. Perham, R. N.; Duckworth, H. W.; Roberts, G. C., Mobility of polypeptide chain in the
1111 pyruvate dehydrogenase complex revealed by proton NMR. *Nature* **1981**, *292* (5822), 474-7.
- 1112 37. Shusterman-Krush, R.; Grimm, L.; Avram, L.; Biedermann, F.; Bar-Shir, A., Elucidating
1113 dissociation activation energies in host-guest assemblies featuring fast exchange dynamics. *Chem*
1114 *Sci* **2020**, *12* (3), 865-871.
- 1115 38. Yang, X. L.; Kaenzig, C.; Lee, M.; Wang, A. H., Binding of AR-1-144, a tri-imidazole
1116 DNA minor groove binder, to CCGG sequence analyzed by NMR spectroscopy. *Eur J Biochem*
1117 **1999**, *263* (3), 646-55.
- 1118 39. Harika, N. K.; Paul, A.; Stroeve, E.; Chai, Y.; Boykin, D. W.; Germann, M. W.; Wilson,
1119 W. D., Imino proton NMR guides the reprogramming of A*T specific minor groove binders for
1120 mixed base pair recognition. *Nucleic Acids Res* **2016**, *44* (10), 4519-27.
- 1121 40. Krafcikova, M.; Dzatko, S.; Caron, C.; Granzhan, A.; Fiala, R.; Loja, T.; Teulade-Fichou,
1122 M. P.; Fessl, T.; Hansel-Hertsch, R.; Mergny, J. L.; Foldynova-Trantirkova, S.; Trantirek, L.,
1123 Monitoring DNA-Ligand Interactions in Living Human Cells Using NMR Spectroscopy. *J Am*
1124 *Chem Soc* **2019**, *141* (34), 13281-13285.

- 1125 41. Wang, A. H.; Cottens, S.; Dervan, P. B.; Yesinowski, J. P.; van der Marel, G. A.; van
1126 Boom, J. H., Interactions between a symmetrical minor groove binding compound and DNA
1127 oligonucleotides: ^1H and ^{19}F NMR studies. *J Biomol Struct Dyn* **1989**, *7* (1), 101-17.
- 1128 42. Haynes, D. H.; Kowalsky, A.; Pressman, B. C., Application of nuclear magnetic resonance
1129 to the conformational changes in valinomycin during complexation. *J Biol Chem* **1969**, *244* (2),
1130 502-5.
- 1131 43. Bunkenborg, J.; Behrens, C.; Jacobsen, J. P., NMR characterization of the DNA binding
1132 properties of a novel Hoechst 33258 analogue peptide building block. *Bioconjug Chem* **2002**, *13*
1133 (5), 927-36.
- 1134 44. Fede, A.; Labhardt, A.; Bannwarth, W.; Leupin, W., Dynamics and binding mode of
1135 Hoechst 33258 to d(GTGGGAATTCCAC)₂ in the 1:1 solution complex as determined by two-
1136 dimensional ^1H NMR. *Biochemistry* **1991**, *30* (48), 11377-88.
- 1137 45. W. David Wilson; Fariat A. Tanious; Daoyuan Ding; Arvind Kumar; David W. Boykin;
1138 Pierre Colson; Claude Houssier; Bailly, C., Nucleic Acid Interactions of Unfused Aromatic
1139 Cations: Evaluation of Proposed Minor-Groove, Major-Groove, and Intercalation Binding Modes.
1140 *J. Am. Chem. Soc.* **1998**, *120* (40), 10310-10321.
- 1141 46. Geierstanger, B. H.; Wemmer, D. E., Complexes of the minor groove of DNA. *Annu Rev*
1142 *Biophys Biomol Struct* **1995**, *24*, 463-93.
- 1143 47. Chenoweth, D. M.; Dervan, P. B., Allosteric modulation of DNA by small molecules. *Proc*
1144 *Natl Acad Sci U S A* **2009**, *106* (32), 13175-9.
- 1145 48. Kwok, Y.; Zhang, W.; Schroth, G. P.; Liang, C. H.; Alexi, N.; Bruice, T. W., Allosteric
1146 interaction of minor groove binding ligands with UL9-DNA complexes. *Biochemistry* **2001**, *40*
1147 (42), 12628-38.
- 1148 49. Bourassa, P.; Tajmir-Riahi, H. A., Folic acid binds DNA and RNA at different locations.
1149 *International journal of biological macromolecules* **2015**, *74*, 337-42.
- 1150 50. Zhang, X.; Sanger, A.; Hemmig, R.; Jahnke, W., Ranking of high-affinity ligands by NMR
1151 spectroscopy. *Angew Chem Int Ed Engl* **2009**, *48* (36), 6691-4.
- 1152 51. Huang, X.; de Vera, I. M.; Veloro, A. M.; Blackburn, M. E.; Kear, J. L.; Carter, J. D.;
1153 Rocca, J. R.; Simmerling, C.; Dunn, B. M.; Fanucci, G. E., Inhibitor-induced conformational shifts
1154 and ligand-exchange dynamics for HIV-1 protease measured by pulsed EPR and NMR
1155 spectroscopy. *J Phys Chem B* **2012**, *116* (49), 14235-44.
- 1156 52. Williamson, M. P., Using chemical shift perturbation to characterise ligand binding. *Prog*
1157 *Nucl Magn Reson Spectrosc* **2013**, *73*, 1-16.

- 1158 53. Giri, M.; Maulik, A.; Singh, M., Signatures of Specific DNA Binding by the AT-Rich
1159 Interaction Domain of BAF250a. *Biochemistry* **2020**, *59* (1), 100-113.
- 1160 54. Hjalte, J.; Hossain, S.; Hugerth, A.; Sjogren, H.; Wahlgren, M.; Larsson, P.; Lundberg,
1161 D., Aggregation Behavior of Structurally Similar Therapeutic Peptides Investigated by (1)H NMR
1162 and All-Atom Molecular Dynamics Simulations. *Mol Pharm* **2022**, *19* (3), 904-917.
- 1163 55. Shi, X.; Chaires, J. B., Sequence- and structural-selective nucleic acid binding revealed by
1164 the melting of mixtures. *Nucleic acids research* **2006**, *34* (2), e14.
- 1165 56. Paul, A.; Farahat, A. A.; Boykin, D. W.; Wilson, W. D., Thermodynamic Factors That
1166 Drive Sequence-Specific DNA Binding of Designed, Synthetic Minor Groove Binding Agents. *Life*
1167 **2022**, *12* (5).
- 1168 57. Lomize, A. L.; Hage, J. M.; Schnitzer, K.; Golobokov, K.; LaFaive, M. B.; Forsyth, A.
1169 C.; Pogozheva, I. D., PerMM: A Web Tool and Database for Analysis of Passive Membrane
1170 Permeability and Translocation Pathways of Bioactive Molecules. *J Chem Inf Model* **2019**, *59* (7),
1171 3094-3099.
- 1172 58. BIOVIA Dassault Systèmes, *Discovery Studio Visualizer 2021*, Dassault Systèmes: San
1173 Diego, 2021.
- 1174 59. Culig, Z.; Klocker, H.; Bartsch, G.; Hobisch, A., Androgen receptors in prostate cancer.
1175 *Endocr Relat Cancer* **2002**, *9* (3), 155-70.
- 1176 60. Khorasanizadeh, S.; Rastinejad, F., Nuclear-receptor interactions on DNA-response
1177 elements. *Trends in biochemical sciences* **2001**, *26* (6), 384-90.
- 1178 61. Claessens, F.; Gewirth, D. T., DNA recognition by nuclear receptors. *Essays in biochemistry*
1179 **2004**, *40*, 59-72.
- 1180 62. Weikum, E. R.; Liu, X.; Ortlund, E. A., The nuclear receptor superfamily: A structural
1181 perspective. *Protein Sci* **2018**, *27* (11), 1876-1892.
- 1182 63. Yang, F. W.; Li, Y. X.; Ren, F. Z.; Luo, J.; Pang, G. F., Assessment of the endocrine-
1183 disrupting effects of organophosphorus pesticide triazophos and its metabolites on endocrine
1184 hormones biosynthesis, transport and receptor binding in silico. *Food Chem Toxicol* **2019**, *133*,
1185 110759.
- 1186 64. Meinel, S.; Ruhs, S.; Schumann, K.; Stratz, N.; Trenkmann, K.; Schreier, B.; Grosse, I.;
1187 Keilwagen, J.; Gekle, M.; Grossmann, C., Mineralocorticoid receptor interaction with SP1
1188 generates a new response element for pathophysiologically relevant gene expression. *Nucleic Acids*
1189 *Res* **2013**, *41* (17), 8045-60.

- 1190 65. Chenoweth, D. M.; Dervan, P. B., Structural basis for cyclic Py-Im polyamide allosteric
1191 inhibition of nuclear receptor binding. *Journal of the American Chemical Society* **2010**, *132* (41),
1192 14521-9.
- 1193 66. Nickols, N. G.; Dervan, P. B., Suppression of androgen receptor-mediated gene expression
1194 by a sequence-specific DNA-binding polyamide. *Proceedings of the National Academy of Sciences*
1195 *of the United States of America* **2007**, *104* (25), 10418-23.
- 1196 67. Vasiutina, E. L.; Bugreev, D. V.; Riabinin, V. A.; Siniakov, A. N.; Buneva, V. N.;
1197 Nevinskii, G. A., Interaction of the human topoisomerase I-DNA complex with oligo-1,3-
1198 thiazolecarboxamides and their oligonucleotide conjugates. *Mol Biol (Mosk)* **2000**, *34* (3), 356-362.
- 1199 68. Bugreev, D. V.; Vasyutina, E. L.; Ryabinin, V. A.; Sinyakov, A. N.; Buneva, V. N.;
1200 Nevinsky, G. A., Inhibition of human DNA topoisomerase I by new DNA minor groove ligands:
1201 derivatives of oligo-1,3-thiazolecarboxamides. *Antisense Nucleic Acid Drug Dev* **2001**, *11* (3), 137-
1202 47.
- 1203 69. Ismail, M. A.; Tratat, C.; Haroun, M. G., Molecular modelling design, synthesis and
1204 cytotoxic evaluation of certain substituted 2-(3,4,5-triacetoxybenzoylamino)benzo[d]thiazole and
1205 2-(galloylamino)benzo[d]thiazole derivatives having potential topoisomerase-I inhibitory activity.
1206 *J Enzyme Inhib Med Chem* **2013**, *28* (6), 1331-45.
- 1207 70. Momparler, R. L.; Karon, M.; Siegel, S. E.; Avila, F., Effect of adriamycin on DNA, RNA,
1208 and protein synthesis in cell-free systems and intact cells. *Cancer Res* **1976**, *36* (8), 2891-5.
- 1209 71. van der Zanden, S. Y.; Qiao, X.; Neefjes, J., New insights into the activities and toxicities
1210 of the old anticancer drug doxorubicin. *FEBS J* **2021**, *288* (21), 6095-6111.
- 1211 72. Shi, J. H.; Chen, J.; Wang, J.; Zhu, Y. Y., Binding interaction between sorafenib and calf
1212 thymus DNA: spectroscopic methodology, viscosity measurement and molecular docking.
1213 *Spectrochim Acta A Mol Biomol Spectrosc* **2015**, *136 Pt B*, 443-50.
- 1214 73. Portugal, J.; Bataller, M.; Mansilla, S., Cell death pathways in response to antitumor
1215 therapy. *Tumori* **2009**, *95* (4), 409-21.
- 1216 74. Ricci, M. S.; Zong, W. X., Chemotherapeutic approaches for targeting cell death pathways.
1217 *The oncologist* **2006**, *11* (4), 342-57.
- 1218 75. Hilgeroth, A.; Tell, V.; Kramer, S.; Totzke, F.; Schachtele, C., Approaches to a
1219 multitargeting drug development: first profiled 3-ethoxycarbonyl-1-aza-9-oxafluorenes
1220 representing a perspective compound class targeting Alzheimer disease relevant kinases CDK1,
1221 CDK5 and GSK-3beta. *Medicinal chemistry* **2014**, *10* (1), 90-7.
- 1222 76. Proschak, E.; Stark, H.; Merk, D., Polypharmacology by Design: A Medicinal Chemist's
1223 Perspective on Multitargeting Compounds. *J Med Chem* **2019**, *62* (2), 420-444.

- 1224 77. Peters, J. U., Polypharmacology - foe or friend? *J Med Chem* **2013**, *56* (22), 8955-71.
- 1225 78. Kwan, E. E., ACD/Spectrus Processor review. *J Chem Inf Model* **2012**, *52* (7), 1898-900.
- 1226 79. BIOVIA, D. S., Discovery Studio Visualizer 2021, in, Dassault Systèmes, San Diego, 2021.
- 1227 80. Honorato, R. V.; Koukos, P. I.; Jimenez-Garcia, B.; Tsaregorodtsev, A.; Verlato, M.;
1228 Giachetti, A.; Rosato, A.; Bonvin, A., Structural Biology in the Clouds: The WeNMR-EOSC
1229 Ecosystem. *Front Mol Biosci* **2021**, *8*, 729513.
- 1230 81. van Zundert, G. C. P.; Rodrigues, J.; Trellet, M.; Schmitz, C.; Kastritis, P. L.; Karaca,
1231 E.; Melquiond, A. S. J.; van Dijk, M.; de Vries, S. J.; Bonvin, A., The HADDOCK2.2 Web Server:
1232 User-Friendly Integrative Modeling of Biomolecular Complexes. *J Mol Biol* **2016**, *428* (4), 720-
1233 725.
- 1234 82. Savelyev, A.; MacKerell, A. D., Jr., Competition among Li(+), Na(+), K(+), and Rb(+)
1235 monovalent ions for DNA in molecular dynamics simulations using the additive CHARMM36 and
1236 Drude polarizable force fields. *J Phys Chem B* **2015**, *119* (12), 4428-40.
- 1237 83. Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian,
1238 E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D., Jr., CHARMM general force field: A
1239 force field for drug-like molecules compatible with the CHARMM all-atom additive biological
1240 force fields. *J Comput Chem* **2010**, *31* (4), 671-90.
- 1241 84. Jo, S.; Kim, T.; Iyer, V. G.; Im, W., CHARMM-GUI: a web-based graphical user interface
1242 for CHARMM. *J Comput Chem* **2008**, *29* (11), 1859-65.
- 1243 85. Wang, H.; Dommert, F.; Holm, C., Optimizing working parameters of the smooth particle
1244 mesh Ewald algorithm in terms of accuracy and efficiency. *The Journal of chemical physics* **2010**,
1245 *133* (3), 034117.
- 1246 86. Sponer, J.; Kypr, J., Characterization of the base stacking interactions in DNA by means of
1247 Lennard-Jones empirical potentials. *General physiology and biophysics* **1989**, *8* (3), 257-72.
- 1248 87. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C., Numerical integration of the cartesian
1249 equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of*
1250 *Computational Physics* **1977**, *23* (3), 327-341.
- 1251 88. Humphrey, W.; Dalke, A.; Schulten, K., VMD: visual molecular dynamics. *Journal of*
1252 *molecular graphics* **1996**, *14* (1), 33-8, 27-8.
- 1253 89. https://dtp.cancer.gov/discovery_development/nci-60/methodology.htm (accessed 18 May
1254 2023).
- 1255 90. S Ramadan, W.; Saleh, E. M.; Menon, V.; Vazhappilly, C. G.; Abdu-Allah, H. H. M.; El-
1256 Shorbagi, A.-N. A.; Mansour, W.; El-Awady, R., Induction of DNA damage, apoptosis and cell

1257 cycle perturbation mediate cytotoxic activity of new 5-aminosalicylate–4-thiazolinone hybrid
1258 derivatives. *Biomedicine & Pharmacotherapy* **2020**, *131*, 110571.

1259

1260