Bioorg. Med. Chem. Lett. 73 (2022) 128886

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry Letters

journal homepage: www.elsevier.com/locate/bmcl

Identification of sugar-containing natural products that interact with i-motif DNA

Rupesh V. Chikhale, Dilek Guneri, Robert Yuan, Christopher J. Morris, Zoë A.E. Waller *

UCL School of Pharmacy, 29-39 Brunswick Square, London WC1N 1AX, UK

ARTICLE INFO

Keywords:
i-Motif
G-quadruplex
DNA
Sugars
Carbohydrates
Natural products
High throughput virtual screening
Molecular dynamics
MM-GBSA
Fluorescence intercalator displacement assay
Forster resonance energy transfer melting

ABSTRACT

There are thousands of compounds shown to interact with G-quadruplex DNA, yet very few which target i-motif (iM) DNA. Previous work showed that tobramycin can interact with iM- DNA, indicating the potential for sugar-molecules to target these structures. Computational approaches indicated that the sugar-containing natural products baicalin and geniposidic acid had potential to target iM-DNA. We assessed the DNA interacting properties of these compounds using FRET-based DNA melting and a fluorescence-based displacement assay using iM-DNA structures from the human telomere and the insulin linked polymorphic region (ILPR), as well as complementary G-quadruplex and double stranded DNA. Both baicalin and geniposidic acid show promise as iM-interacting compounds with potential for use in experiments into the structure and function of i-motif forming DNA sequences and present starting points for further synthetic development of these as probes for iM-DNA.

The stability of DNA varies depending on the sequence, which in turn affects the potential intermolecular interactions, base stacking and hydrogen bonding.[1–2] In particular GC-rich sequences have a higher thermal stability compared to AT-rich sequences and genome wide GC content varies between organisms due to many reasons: selection, life-history traits, genome size, mutational bias and the biased recombination-associated DNA repair.[3–5] In addition to B-form double stranded (ds) DNA, GC-rich sequences can form various alternative structures like Z-DNA, hairpins, G-quadruplex (G4) and i-motif (iM) DNA structures.[6–9] The G-quadruplex is a G-rich quadruple helical structure that has been shown to form in both the telomeric and gene promoter regions of the genome.[9] In GC-rich genomic DNA, opposite where there are G-rich regions the complementary C-rich regions can form another type of quadruplex DNA, known as i-motif.[10–12] G-quadruplex and i-motif DNA structures have been identified to form *in vivo* and are potential targets for development of novel drug molecules towards cancer and other genetic diseases[13–15] as well as targets for infectious diseases.[16–18].

X-ray crystallography and NMR spectroscopy has provided detailed information on the structures of double stranded DNA, G4-DNA, and iM-DNA (Fig. 1).[19–20] iMs and G4s are found not only in the telomeric region but also in promoter region of oncogenes including *c-MYC*, *BCL2*, *cKIT*, *HIF-1A* and other regulatory regions such as the insulin linked polymorphic region (ILPR)[21], which suggest possible roles in gene expression.[6]

Stabilisation of G4s and iMs in the human telomere can inhibit the telomerase activity leading to telomeric uncapping, DNA damage response and apoptosis.[22–23] The ILPR is known to affect the human insulin protein production.[24] The ILPR 5′-(ACAGGGGTGTGGGG)$_2$TGT sequence can fold into a G4 at physiological pH and iM at slightly acidic pH, but these two species have been shown to be mutually exclusive using laser tweezer experiments.[25–26].

The interaction between small molecule ligands and the alternative DNA structures is not completely understood, especially for iM-DNA. There are thousands of ligands known to interact with G4s but comparatively few to target iMs. Various techniques have been applied to identify ligands interacting with G4s and iMs, such as Fluorescence Resonance Energy Transfer (FRET) melting,[27] Fluorescent Intercalator Displacement (FID),[28] Electrospray Ionization Mass Spectrometry (ESI-MS),[29] Small-Molecule Microarrays (SMMs)[30] surface plasmon resonance (SPR),[28] and molecular modelling studies.[31–33] In our pursuit of iM-interacting compounds, we have previously investigated and reported that several compounds, including mitoxantrone, tobramycin, tilorone, harmalol and quinalizarin, interact with the telomeric and c-

* Corresponding author at: Department of Pharmaceutical and Biological Chemistry, UCL School of Pharmacy, 29-39 Brunswick Square, London WC1N 1AX, UK.
E-mail address: z.waller@ucl.ac.uk (Z.A.E. Waller).

https://doi.org/10.1016/j.bmcl.2022.128886
Received 31 May 2022; Received in revised form 5 July 2022; Accepted 7 July 2022
Available online 11 July 2022

MYC iMs.[28] Tobramycin[28] is an aminoglycoside antibiotic derived from *S. tenebrarius,* used in the treatment of Gram-negative bacterial infections. In our screen it was found to interact with the iM-forming sequences from the human telomere (hTeloC, SPR $K_d = 17 \pm 2.0$ μM), and the promoter region of c-MYC (SPR $K_d = 13 \pm 1.8$ μM) as well as ds-DNA (SPR $K_d = 18 \pm 1.1$ μM). Although tobramycin was observed to be a generic DNA binding compound, this work motivated us to investigate the further potential the naturally-occurring saccharides as possible iM-interacting compounds. Recent improvements in the molecular docking algorithms and forcefields for molecular dynamics (MD) simulations can help in screening large datasets of ligand libraries. Here we employed a well-established workflow of high throughput virtual screening[34] combined with molecular dynamics simulation and binding energy calculation[35] towards identification of iM-DNA binding ligands from a set of natural products containing sugars and related metabolites.

To begin looking at a large range of natural products containing sugar moieties, we devised an *in silico* workflow (Fig. 2) which involved steps starting from selection of databases and target DNA structures to Molecular Mechanics-Generalized Born Solvent Accessibility MM-GBSA calculations. Detailed methodological information is provided in the supplementary information (SI). A set of natural products containing sugars comprising 255 molecules were curated from the Selleckchem database and extracted from the database in.sdf format. This set comprises of glycosides, saponins, saccharides, flavones, pyranosides and various other conjugated sugar moieties (SI).[36] These ligand molecules were converted from 2D to 3D in the ligprep module of the Maestro-Glide and all possible conformations were generated to yield 2062 conformation in total. Our aim was to identify iM-DNA interacting compounds, but we wanted to have an indication of their potential interactions with other types of DNA structure. We therefore prepared models of B-form ds-DNA (1BNA)[37] as well as the G4 (1KF1)[13] and iM (1ELN)[12] structures from the human telomere. These were obtained from the protein databank and prepared in the Glide software using the protein preparation wizard. The grids for all the target structures were generated based on the Qsite tool.[38] Molecular docking of ligands was performed in the standard precision (SP) and extra precision (XP) modes. Based on the consensus from the SP and XP modes and ligand-DNA interactions and RMSD, a cut-off of −5.0 dock score and −30.0 kcal/mol of Glide energy score was considered for filtering and selecting the ligands. Accordingly, 25, 26 and 21 molecules with RMSD < 2.0 Å were selected for iM, G4 and ds-DNA respectively for further studies.

There were several high dock score molecules found common in all the three subsets obtained from the molecular docking studies, these were later included in the MDS. The molecular docking experiments provided with the Dock Score and Glide energy scores (Table 1). As we were interested in potential compounds that could be used as probes for iM-DNA, the docking results filtered out common sugars present in the human body. Several foodstuff-derived chemicals showed high dock scores towards the iM structure: baicalin was found to dock in the minor groove with several interactions between the surrounding bases, with a dock score of −6.50 and glide energy of −51.43 kcal/mol. The DNA-ligand interaction forms various contacts including hydrogen bonding and electrostatic interactions, the O8 and O9 in baicalin from the trihydroxyoxane ring forms a hydrogen bond interaction with the iM-Cytosine10-OP1 (OP1–O8 bond length = 2.70 Å and OP1–O9 bond length = 2.86 Å), and iM-Cytosine11-OP1 (OP2–O7 bond length = 2.77 Å) (Fig. 3 and SI16). The highest dock score molecules were mostly from the ds-DNA subset with daidzin as the highest scorer (Dock score = -8.74, Glide Energy = -68.14 kcal/mol) (Fig. SI3), it docks in the minor groove of the DNA, the O8 on glucopyranosyl ring forms a hydrogen bond with the N2 of DNA-Guanine17 (N–O bond length = 2.76 Å). The compound aesculin ranked highest from the G4 subset (Dock score = -6.26, Glide Energy = -40.82 kcal/mol) (Fig. SI8), it docked over the top of the loop forming hydrogen bond between G4-Adenine5 N3 and O6 of the glucopyranosyl ring (N–O bond length = 2.68 Å). After careful examination of the docked ligands from the SP and XP docking experiments (Fig. SI1-21) across the different DNA structures we identified baicalin and genoposidic acid as potential hits that were found to target iM-DNA that did not come up in the top-ranking compounds for G4 or ds-DNA (SI Table 5). Given their similarity in structure, we decided to also assess the analogues geniposide, daidzin, aesculin and genistin for comparison (Fig. 4). We also examined lactulose as it is a commonly-used drug and was found to score highly against iM and G4-DNA but not ds-DNA.

To understand the binding of these molecules on each of the DNA structures and understand the stability over a period of time we performed MD simulation on these DNA-ligand complexes for 100 ns in an explicit solvent model, and then performed MM-GBSA calculations on all the frames from the trajectories to calculate the binding energies (full details in the SI). The molecular docking studies of selected complexes showed good interactions between the ligand molecules and the iM. The aesculin-iM complex showed the formation of various hydrogen bond interactions. Aesculin binds in the major groove, the coumarin ketone oxygen (O3) forms interaction with the iM-Thymine N3 (O–N, bond length = 2.93 Å), the 7-OH forms hydrogen bond with iM-Cytosine 5-N4 (O–H–N, bond length = 2.84 Å) (Fig. SI 15). The glucopyranosyloxy ring formed three hydrogen bonds with the iM DNA; iM-Cytosine 6-N4 with aesculin O7 (O–H–N, bond length = 2.80 Å), iM-Cytosine 11-OP2 with aesculin O6 (O–H–O, bond length = 2.75 Å), and iM-Cytosine 11-OP2 with aesculin O9 (O–H–O, bond length = 2.64 Å). These intermolecular interactions, dock score and Glide energy score of −5.20 and −42.98 kcal/mol indicates formation of stable complex between

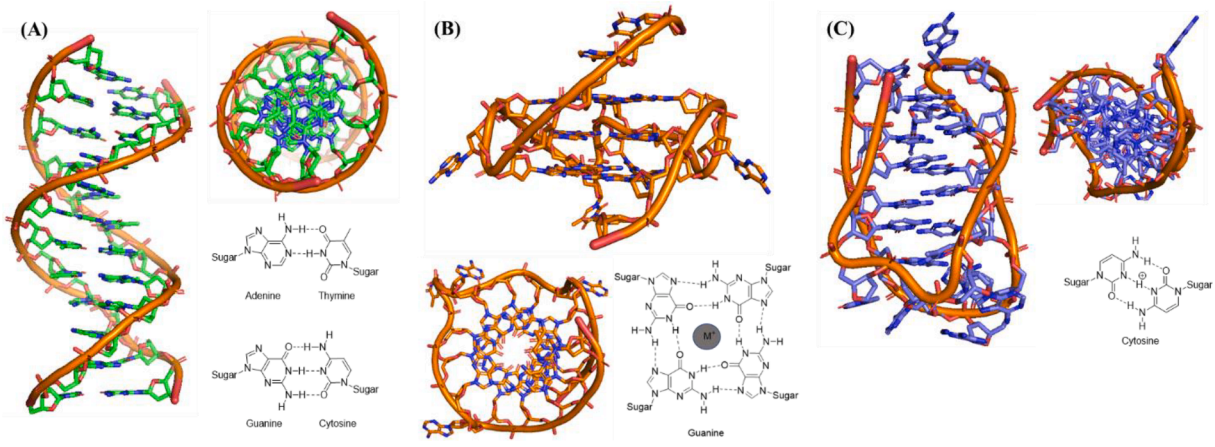

Fig. 1. Structures used in the *in silico* study (A) ds-DNA (PDB: 1BNA), structure of B-form DNA with Watson and Crick base pairing, (B) G-quadruplex DNA (PDB: 2BLY) structure and Hoogsteen base pairing in a G4 with metal ion (M^+), (C) i-Motif DNA (PDB: 1EL2) structure and hemi-protonated cytosine-cytosine base pairing.

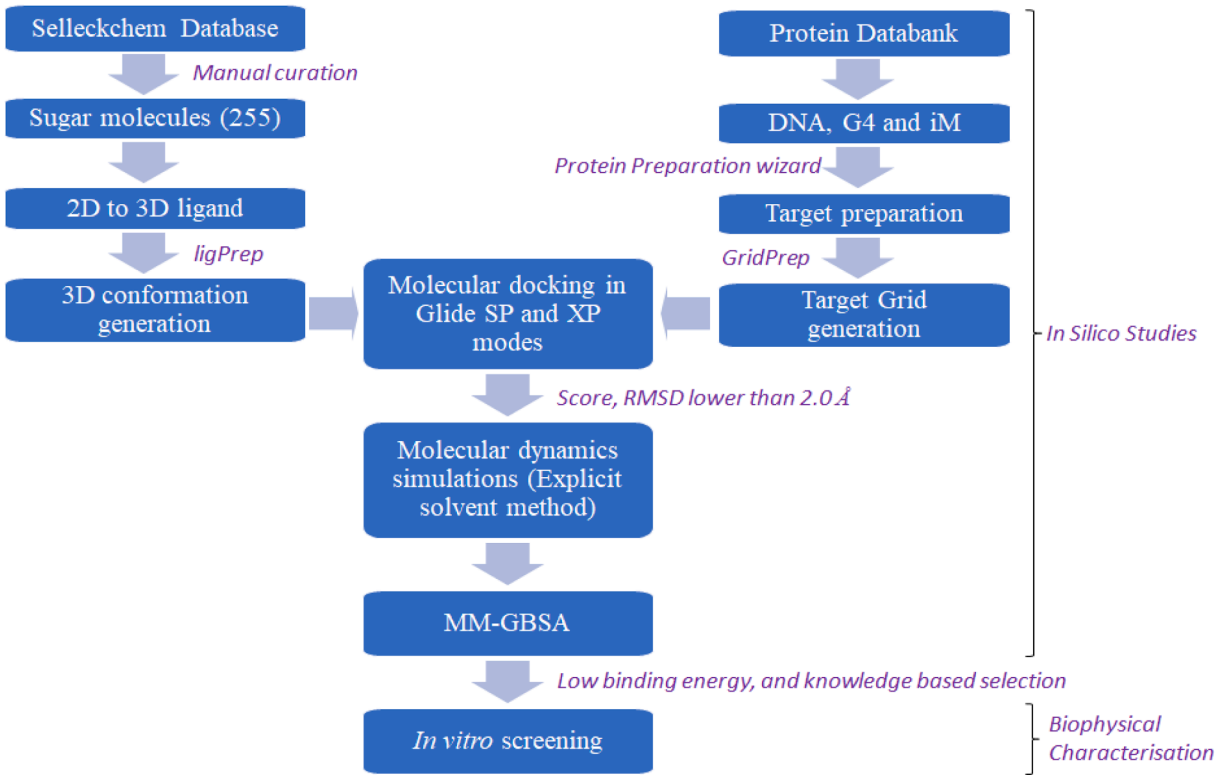

Fig. 2. Experimental workflow and steps involved in the screening of ligand molecules during *in silico* and *in vitro* experiments.

Table 1
Results from the molecular docking studies, molecular dynamics simulations and MM-GBSA binding energy calculations.

Compounds (CAS)	B-DNA (ds)					hTeloG (G4)					hTeloC (iM)				
	Mol. Docking*		MM-GBSA#	MD-RMSD (Avg)$		Mol. Docking		MM-GBSA	MD-RMSD (Avg)		Mol. Docking		MM-GBSA	MD-RMSD (Avg)	
	Dock Score	Glide Energy	ΔG_{bind} (SD)	DNA	Ligand	Dock Score	Glide Energy	ΔG_{bind} (SD)	DNA	Ligand	Dock Score	Glide Energy	ΔG_{bind} (SD)	DNA	Ligand
Aesculin (531–75-9)	−6.07	−41.05	−22.17 (3.87)	3.37	4.15	−6.26	−40.82	−27.77 (3.84)	2.48	3.18	−5.20	−42.98	−8.80 (3.24)	6.07	21.35
Baicalin (21967–41-9)	−6.9	−45.34	−15.12 (4.41)	2.81	5.59	−5.21	−43.22	−16.91 (3.21)	2.67	2.93	−6.50	−51.42	−7.78 (4.81)	3.19	2.84
Daidzin (552–66-9)	−8.74	−68.14	−25.78 (12.77)	3.78	5.92	−5.17	−40.96	−26.71 (4.93)	2.44	4.80	−5.67	−44.81	−26.50 (3.76)	1.75	3.79
Geniposide (24512–63-8)	−5.74	−41.64	−10.33 (8.10)	3.43	3.78	−5.58	−47.16	−14.54 (2.34)	3.52	9.67	−5.20	−38.41	−6.80 (1.60)	7.80	26.27
Geniposidic Acid (27741–01-1)	−7.71	−61.38	−23.03 (0.13)	3.18	3.83	−5.76	−46.71	−22.43 (3.53)	2.97	8.00	−5.03	−34.92	−19.70 (5.08)	4.83	17.07
Genistin (529–59-9)	−7.01	−49.74	−22.31 (5.11)	3.13	3.04	−5.87	−45.57	−27.75 (3.46)	3.27	10.05	−5.27	−45.70	−18.38 (2.61)	4.93	3.59
Lactulose (4618–18-2)	−5.32	−40.56	−12.65 (6.16)	6.60	33.28	−5.12	−46.04	−10.44 (4.46)	2.41	1.61	−5.82	−56.28	−8.01 (2.98)	6.22	25.59

*The docking results are presented as dock score and the Glide energy in kcal/Mol, #MM-GBSA binding energy is presented in ΔG_{bind} in kcal/mol with standard deviation in parentheses, $molecular dynamics root mean square deviation (RMSD) averages for the DNA and bound ligand is presented in Å.

aesculin and iM DNA. The dock score and the Glide energy score for the iM-diadzin complex were −5.67 and −44.81 kcal/mol respectively. The ligand DNA interaction was observed between the glucopyranosyloxy ring and the bases but the interaction with the flavone ring was not observed (Fig. SI 17). The iM-Cytosine 12-OP2 forms hydrogen bond with the daidzin O7 (O–H–O, bond length = 2.63 Å), iM-Cytosine 11-OP2 forms hydrogen bond with the daidzin O8 (O–H–O, bond length = 2.62 Å), and iM-Cytosine 4-O2 formed a hydrogen bond with the daidzin O9 (O–H–O, bond length = 2.62 Å). In this case the hydrogen bonds were short and thus considered strong interactions, but there is no interaction with the flavone ring which makes this complex weaker compared to other iM-ligand interactions.

The iM-geniposide complex showed dock score and Glide energy scores of −5.20 and −38.41 kcal/mol respectively with the ligand binding linearly forming various interactions (Fig. SI 18). The glycosidic hydroxy (O10) forms a hydrogen bond with the iM Cytosine 5-N4 (O–H–

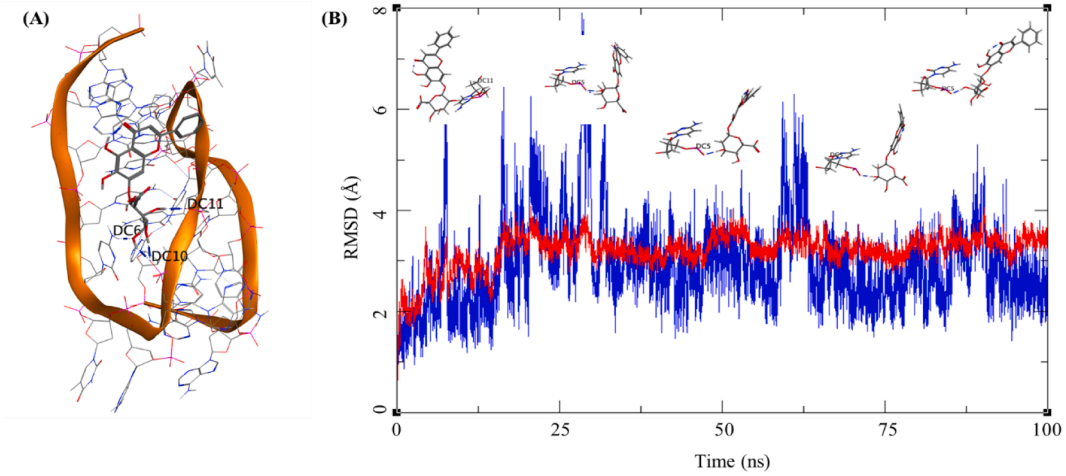

Fig. 3. (A) iM-Baicalin complex from the molecular docking experiment, (B) Ligand (baicalin) RMSD (Blue) and iM RMSD (Red) over 100 ns of molecular dynamics simulation. The ligand bound to the iM is presented with different conformations adopted during the MDS while retaining its hydrogen bond interaction with the iM.

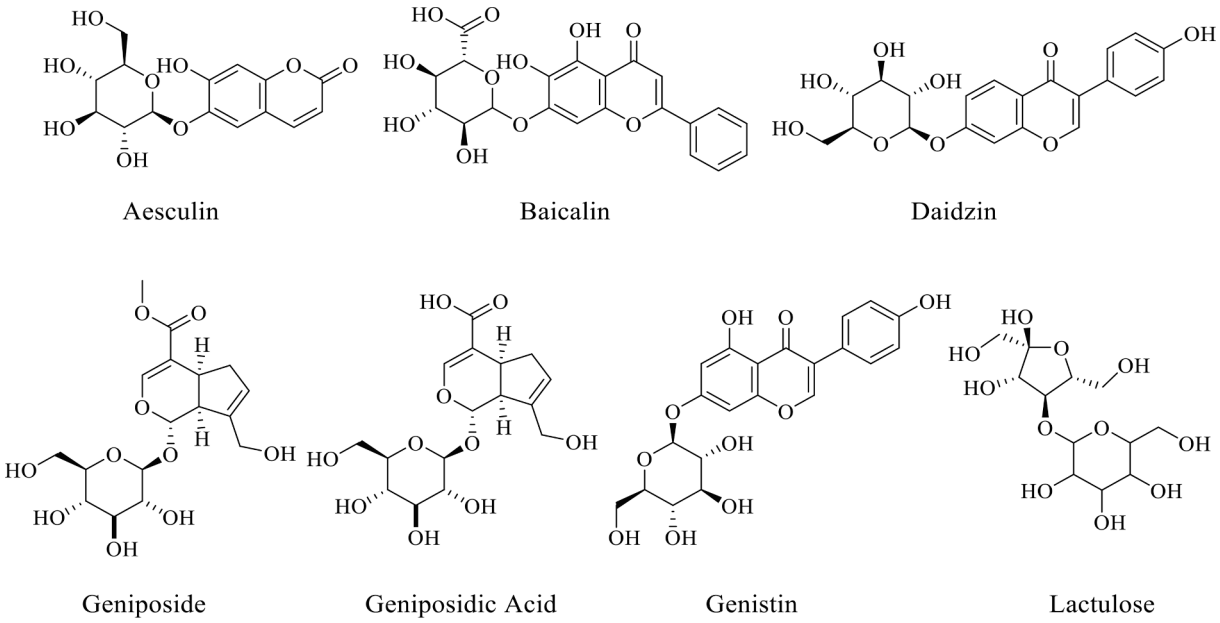

Fig. 4. Ligands obtained from the molecular docking and molecular dynamics simulation screening and tested for their interaction with DS, G4 and iM-DNA by FRET and FID methods.

N, bond length = 3.01 Å). The hexopyranosyloxy group forms two separate hydrogen bonds; iM-Thymine 5-O4 with geniposide O6 (O–H–O, bond length = 2.66 Å), and iM-Cytosine 12-OP2 with geniposide O3 (O–H–O, bond length = 2.75 Å). Geniposidic acid docked in the iM major groove with the carboxylic acid from the cyclopentapyran ring forming interaction with the iM-Cytosine (geniposidic acid O9–H–N4 of iM-Cytosine, bond length = 2.75 Å), and the geniposidic acid hydroxy O10 forming a hydrogen bond with the iM-Cytosine N4 (O10–H–N4, bond length = 3.04 Å) (Fig. SI 19). The geniposidic acid O7 forms hydrogen bond with iM Cytosine 11-N4 and iM-Cytosine 6-N4 with bond length 2.89 and 2.91 Å respectively. The dock score, Glide energy and binding energy of the iM-geniposidic acid complex was −5.03, −34.92 and −19.70 (5.08) (ΔG_{bind} (SD)) kcal/mol respectively, which indicates higher stability of this complex. The docked complex of lactulose shows fewer interactions between the ligand and iM bases. The dock score and Glide energy for this complex was −5.82 and −56.28 Kcal/mol respectively. The iM-Cytosine 10-OP1 formed two interactions with the lactulose O2 and O6 of bond length 2.86 and 2.74 Å respectively. The interaction between lactulose and the iM is less stable and the ligand breaks away from the complex towards the end of the molecular dynamic simulations (Fig. SI 21) leading to high ligand RMSD (SI Table 4).

The iM-ligand complexes gave a clearer picture in MDS with two ligands baicalin and genistin showed low RMSD and matched that with the iM RMSD. The iM-baicalin RMSD remained low throughout the MDS with an average for ligand at 2.84 Å and for the iM at 3.1 Å suggesting a stable complex (Fig. 3). The ligand binds in the major groove with formation of three hydrogen bonds during the docking experiment. After MDS minimisation, simulated annealing and NPT, NVT optimisation steps baicalin retained its bond with the iM-Cytosine11 (Bond length = 2.76 Å). The slight fluctuations in the MDS RMSD does not affect the binding and an example is presented in the Fig. 3(B) where, the ligand achieves several conformations while retaining its hydrogen bond with the iM. The iM-baicalin complex was high ranking complex in the molecular docking simulation with a dock score of −6.50 and Glide binding energy at −51.42 kcal/mol which is highest in case of G4 and iM complexes. However, the binding energy of this complex is to the lower side

(ΔG_{bind} (SD) = -7.78 (4.81)). There was no instance of iM-DNA unwinding during the MDS of the iM-baicalin complex. iM-genistin is another complex which showed better stability compared to other iM-ligand complexes in this study. The average ligand RMSD during MDS for this complex was 3.59 Å and the DNA RMSD was averaged to 4.93 Å. The initial frame of the MD production for iM-genistin show two hydrogen bonds between ligand and the iM-DNA, genistin O9 formed hydrogen bond interaction with the iM-Cytosine-6 N4 (O—H—N bond length = 2.94 Å). The genistin O10 formed interaction with the iM-Adenine-3 OP2 (O–H–O, bond length = 2.69 Å), of these two interactions the first one was retained where as a new interaction was observed between iM-Cytosine11-OP2 and genistin O2 (bond length = 2.61) while the ligand adopted a planer conformation toward the end of the simulation. The binding energy of iM-genistin complex (ΔG_{bind} (SD) = -18.38 (2.61)) was higher than that of baicalin. The iM became unstable leading to unwinding from the 3′end (Fig. SI 20) but not from the 5′ end where the ligand was bound to the base pairs. It is possible that genistin stabilises the iM to a lesser extent compared to the baicalin.

G4-ligand interaction analysis over the time course of the MDS show three ligands with low RMSD, aesculin, baicalin and lactulose displayed average ligand RMSD as 3.18 Å, 2.93 Å and 1.61 Å, respectively. The G4 RMSD was within the range of 1–4 Å for all the complexes with exception of G4-geniposide with slight rise in the RMSD to 6 Å. This might be due to the larger fluctuations in the G4-geniposide complex. The G4-aesculin complex initially showed a lower ligand RMSD for first 30 ns which gradually rose and fluctuated between 6 and 7 Å (Fig. SI22 C and D), the ligand did not form a typical hydrogen bond with the G4 but had an electrostatic interaction and carbohydrate-π stacking between the pyranosyl ring and the G4-Guanine-2. It also shows interactions between aesculin O6 and G4-Adenine-5 O4. The ΔG_{bind} for this complex was −27.7 (3.84) kcal/mol with large contribution from the electrostatic energy component (ΔE_{ELE} = -45.31 (12.4) kcal/mol) (SI Table 2). The G4-baicalin complex had a low RMSD for ligand and the G4 as well, however, baicalin showed several incremental fluctuations between 30 ns with slight rise in RMSD from 2 to 5 Å. This reduced back to the baseline towards the last 25 ns of the MDS. Visual analysis of the complex did not show any hydrogen bond interaction between the ligand and the DNA, but it formed carbohydrate-π stacking between the pyranosyl ring of baicalin and the G4-Guanine-14, this interaction holds up throughout the MDS suggesting for stacking over the bases in G4 as the main mode of interaction and formation of stable complex. The ΔG_{bind} for this complex was −16.91 (3.21) kcal/mol with large contribution from the van der Waals contribution from the MM force field (ΔE_{VDW} = –32.73 (4.42)) and electrostatic energy component (ΔE_{ELE} = -745.20 (14.81) kcal/mol) (SI Table 2). Binding of lactulose with the G4 structure showed a different binding mode compared to geniposide and baicalin. The ligand RMSD for lactulose was not very high, it raised from 2 Å at 10 ns to 3 Å at 90 ns, this is attributed to its conformation change during the MDS. In the initial phase of MDS the lactulose O1 and O3 forms hydrogen bonding with G4-Guanine2 N2 and G4-Thymine 5 N3 of bond length 2.78 and 3.30 Å, respectively. The same interactions were observed towards the end of the MDS with a new bond forming between the lactulose O2 and Adenine7 N6 with bond length 2.88 Å (Fig. SI14). This observation suggests that in case of the G4 DNA these three molecules lactulose, baicalin and geniposide show good interaction and stable complexes.

During the time course of the MD simulations for the ds-ligand complexes (Fig. SI22 A and B, SI table 4), the ligand RMSD for most of the compounds was low. Specifically, it was below 10 Å for most of the ligands except for the lactulose (Avg. ligand RMSD = 33.28 Å) and for the ds-DNA, RMSD was almost below for 6 Å except for lactulose bound ds (Avg. DNA RMSD = 6.60 Å). It was observed that geniposidic acid and genistin had the lowest RMSD, below 5 Å averaging to 3.83 Å and 3.04 Å respectively, throughout the 100 ns of MDS (Table 1). Visual inspection of the complex through the simulation trajectory revealed the retention of a hydrogen bond between the ds-Adenine27-N3 and geniposidic acid-O10 (Fig. SI5). Genistin changed its conformation from a slightly angled pose to planar during the MDS and strengthened its interaction by forming double hydrogen bonds with the minor groove bases i.e., the genistin-O7 formed hydrogen bond interactions with ds-Thymine O2 and ds-Guanine17 N2 of bond length 2.8 Å and 3.02 Å respectively. These two ds-ligand complexes with the highest binding energies (ΔG_{bind}) were ds-geniposidic acid = –23.0 (0.13) kcal/mol and ds-genistin = –22.31(5.11) kcal/mol indicating that the complexes are quite stable.

To support our computational studies, we performed biophysical experiments. Ligand-induced stabilisation was assessed by Förster resonance energy transfer (FRET) melting[40] and relative binding was assessed by fluorescent intercalator displacement (FID) assay.[28,40] We assessed each compound against dsDNA as well as both G4 and iM sequences from the human telomere and the ILPR[21,22].

FRET-based DNA melting experiments are commonly used for assessment of ligand-induced changes in the stability of DNA. FRET melting experiments have advantages over other DNA-based melting experiments in that screening of a large range of conditions/ligand concentrations is possible; lower ligand concentrations are required and it avoids the problem that many ligands absorb in the same region as DNA, which may interfere with the absorbance or ellipticity signal resulting from complex dissociation. Although the technique can give rise to experimental artefacts from inherently fluorescent ligands and compounds which interact with the fluorophores, as the sugar-containing molecules herein are chiral, FRET based melting represents an advantage over circular dichroism melting experiments. We used FRET labelled oligonucleotide probes that were complementary to the sequences employed in the MD simulations (see SI) as well as iM and G4 forming sequences from the ILPR.[23] All sequences were examined at physiological pH (7.4), which reflects physiological conditions, but the iMs from the human telomere and the ILPR are not fully folded at this pH, so the iMs were also tested at pH 5.5, where the structures are fully folded. The results are shown in Table 2 and Fig. SI23.

The FRET results show no compounds that were able to stabilise DNA. This is not unexpected, given tobramycin was not a hit on our original FRET-based screen against i-motif,[39] but was in our FID-based screen.[28] Interestingly, only geniposide (ΔT_m −6.1 ± 1.1 °C for hTeloG) and aesculin (ΔT_m of −7.6 ± 0.0 °C for hTeloG) were found to have any significant effect on the stability of the DNA sequences used in the computational studies. For geniposide, the interaction pattern in the G4-geniposide complex where the pyranosyl ring forms several hydrogen bond interactions like geniposide-O5 and G4-Adenine7-N6 (bond length = 2.87 Å). The geniposide-O3 forms two separate hydrogen bonds with G4-Guanine2-N2 (bond length = 3.25 Å) and G4-Guanine3-O4 (bond length = 2.76 Å) respectively (Fig. SI 11). These interactions could have led to the higher stability of this complex which is reflected in the comparatively better ΔT_m. Although there are thousands of G4-interacting ligands and a large proportion of those have high stabilisation of G4-structures, there are actually few G-quadruplex ligands that have been shown to destabilise the structure: TMPyP4,[35,41] a triarylpyridine derivative,[42] an anthrathiophenedione derivative[43] and a stiff-stilbene derivative.[44] This thus potentially represents another compound to add to that repertoire. Baicalin, geniposidic acid, genistin and lactulose were also found to significantly destabilise the G4 from the ILPR. Only one compound was found to significantly change the stability of iM, aesculin showed a ΔT_m of −6.4 ± 1.5 °C for hTeloC at pH 7.4. At this pH the DNA is not fully folded, but this compound was shown to perturb the structure further. Notably none of the other compounds were found to affect the stability of iM-DNA structures and none of the compounds examined were found to significantly affect the stability of ds-DNA.

Complementary FID assays were performed in analogous sequences for each of the G4 and iM structures (Table 3 and Fig. SI24). The FID assay is a relative binding assay and reflects the displacement of the DNA-binding compound, thiazole orange. This molecule is fluorescent

Table 2
ΔT_m values (in °C) determined by FRET melting at 20 μM (100 eq.) ligand concentration in 10 mM sodium cacodylate and 40 mM NaCl (Mean ± SD, n = 3) One-way ANOVA comparison with Bonferroni post hoc indicates significant difference between control DNA sequences and ligand treatment * $p < 0.05$.

DNA sequence	pH	Aesculin	Baicalin	Daidzin	Geniposide	Geniposidic Acid	Genistin	Lactulose
hTeloG	7.4	−7.6 ± 0.0*	−1.8 ± 3.4	−2.4 ± 1.8	−6.1 ± 1.1*	−4.7 ± 1.8	−6.6 ± 3.0	−4.0 ± 2.2
hTeloC	7.4	−6.4 ± 1.5*	−3.9 ± 2.3	0 ± 0.5	−1.8 ± 1.0	−2.5 ± 1.5	−5.4 ± 3.3	−3.7 ± 2.7
	5.5	−0.4 ± 0.0	−0.4 ± 0.0	−0.3 ± 0.2	−0.5 ± 0.2	−0.5 ± 0.2	−0.8 ± 0.0	−0.4 ± 0.0
ILPRG	7.4	−2.4 ± 0.9	−3.8 ± 0.2*	−1.8 ± 0.6	−2.9 ± 1.0	−3.6 ± 0.5*	−4.7 ± 2.0*	−4.6 ± 0.5*
ILPRC	7.4	−2.6 ± 0.6	−2.6 ± 1.1	−2.4 ± 0.2	−1.8 ± 1.0	−2.2 ± 1.5	−2.4 ± 1.2	−2.7 ± 0.3
	5.5	0.0 ± 0.2	−0.3 ± 0.5	−0.1 ± 0.2	−0.1 ± 0.2	−0.3 ± 0.2	−0.4 ± 0.0	−0.3 ± 0.2
DS	7.4	−2.7 ± 0.4	−2.5 ± 0.6	−2.3 ± 0.2	−3.5 ± 0.8	−4.8 ± 1.1	−3.0 ± 2.0	−2.7 ± 0.9

Table 3
Displacement values (in %) determined by FID using Thiazole Orange at 10 and 100 eq. ligand concentration in 10 mM sodium cacodylate and 40 mM NaCl (Mean ± SD, n = 3). One-way ANOVA comparison with Bonferroni post hoc indicates significant difference between DNA sequences treated with 10 and 100 eq. ligands. * $p < 0.05$, nd = not determined – compound fluorescent. All values are significantly lower than the DNA sequences in the absence of ligand.

DNA	pH	Eq.	Aesculin	Baicalin	Daidzin	Geniposide	Geniposidic Acid	Genistin	Lactulose
hTeloG	7.4	10	nd	24.0 ± 6.4	20.9 ± 7.1	23.1 ± 6.8	23.0 ± 10.6	20.8 ± 6.3	21.4 ± 8.6
		100	nd	46.3 ± 3.5*	29.0 ± 5.6	25.7 ± 3.7	25.2 ± 3.4	28.7 ± 5.4	26.1 ± 6.1
hTeloC	7.4	10	nd	−3.8 ± 1.6	14.9 ± 7.2	−32.0 ± 1.5	2.6 ± 13.4	16.2 ± 10.96	−54.0 ± 2.1
		100	nd	−91.2 ± 58.3*	−154.0 ± 15.1*	−84.1 ± 1.6	−37.8 ± 3.7	−287.7 ± 78.1*	−68.0 ± 10.8*
	5.5	10	nd	12.2 ± 2.0	−4.2 ± 3.3	−9.5 ± 7.3	−10.2 ± 0.4	0.6 ± 5.8	−28.4 ± 28.8
		100	nd	36.5 ± 8.2	20.3 ± 5.1	32.3 ± 8.7*	22.9 ± 7.1*	19.3 ± 3.9	20.2 ± 5.6
ILPRG	7.4	10	nd	4.9 ± 0.8	2.6 ± 2.0	−6.7 ± 1.1	4.1 ± 1.9	1.6 ± 0.1	0.9 ± 0.0
		100	nd	17.7 ± 4.4*	2.4 ± 4.6	1.8 ± 5.5	1.4 ± 4.5	6.4 ± 1.7	0.4 ± 6.7
ILPRC	7.4	10	nd	23.9 ± 1.8	20.0 ± 8.6	21.2 ± 9.1	12.3 ± 9.8	24.2 ± 0.4	26.7 ± 2.7
		100	nd	38.5 ± 0.6	28.9 ± 5.9	31.8 ± 1.7	22.0 ± 6.0	31.8 ± 8.9	34.7 ± 0.3
	5.5	10	nd	9.5 ± 3.7	−2.5 ± 0.2	7.2 ± 3.6	−3.5 ± 0.4	5.3 ± 0.9	−5.0 ± 3.6
		100	nd	35.3 ± 3.0*	22.6 ± 9.3*	27.9 ± 1.7*	25.3 ± 7.1*	23.5 ± 1.0*	23.1 ± 2.8*
DS	7.4	10	nd	8.2 ± 0.6	7.0 ± 6.4	8.5 ± 3.3	6.4 ± 4.0	6.2 ± 2.9	4.3 ± 3.2
		100	nd	25.3 ± 3.5*	10.9 ± 2.7	11.2 ± 3.4	9.6 ± 5.1	14.5 ± 3.4	9.0 ± 3.4

when bound to folded DNA but not when in the unbound state. This loss of fluorescence can be reflected as a percentage displacement. We assessed the ligand-induced displacement of TO at 10 and 100 eq. of compounds. All types of DNA were examined at pH 7.4, but hTeloC and the ILPR are not fully folded at this pH. Some of the displacement values are negative and this can indicate a ligand-induced shift in the equilibrium in solution towards more folded iM present once ligands are added (hence more TO binds and higher fluorescence, rather than less fluorescence) so we also examined the iMs at pH 5.5, to observe displacement of TO from the fully folded structure. Out of all the compounds, the TO displacement was lowest for ds-DNA. For the iM-ligands in the computational studies we identified genistin and baicalin as potential iM-interacting ligands. It is interesting that baicalin shows the highest displacement from hTeloC compared to the other ligands in the study at 12.2 ± 2.0% displacement at 10 eq. and 36.5 ± 8.2% at 100 eq. of ligand. At 10 eq of ligand baicalin displaces TO, whereas none of the other compounds do at this lower ligand concentration. This compound also displaces TO from the iM from the ILPR better than the other compounds in the study. This provides support that baicalin is the best for interacting with iM-DNA out of the compounds examined in this study.

Herein we describe the process of the hybrid of *in silico* and biophysical methods to identify iM-interacting ligands. Computational approaches indicated that baicalin and genoposidic acid had potential to target iM-DNA. Neither of these compounds was found to affect the stability of iM-DNA by FRET melting, but both were found to have relatively good binding affinity for the iM-DNA structures from the human telomere and the ILPR, with baicalin having higher relative binding affinity compared to genoposidic acid. This work has identified new potential sugar-containing natural products as iM-probes and present the starting points for further synthetic development of these as probes for iM-DNA.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This work was funded by Diabetes UK (18/0005820). RVC acknowledges the research computing support by Dr. Md. Ataul Islam, University of Pretoria, and UKZN, SA.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bmcl.2022.128886.

References

1 Yakovchuk P, Protozanova E, Frank-Kamenetskii MD. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Res.* 2006; 34:564–574. https://doi.org/10.1093/NAR/GKJ454.
2 Galtier N, Lobry JR. Relationships Between Genomic G+C Content, RNA Secondary Structures, and Optimal Growth Temperature in Prokaryotes. *J Mol Evol.* 1997;44: 632–636. https://doi.org/10.1007/PL00006186.
3 Musto H, Cacciò S, Rodríguez-Maseda H, Bernardi G. Compositional Constraints in the Extremely GC-poor Genome of Plasmodium falciparum. *Memórias do Instituto Oswaldo Cruz.* 1997;92:835–841. https://doi.org/10.1590/S0074-02761997000600020.
4 Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature.* 2001;409. https://doi.org/10.1038/35057062.
5 Birdsell JA. Integrating Genomics, Bioinformatics, and Classical Genetics to Study the Effects of Recombination on Genome Evolution. *Mol Biol Evol.* 2002;19:1181–1197. https://doi.org/10.1093/OXFORDJOURNALS.MOLBEV.A004176.

6 Day HA, Pavlou P, Waller ZAE. i-Motif DNA: Structure, stability and targeting with ligands. *Bioorg Med Chem*. 2014;22:4407–4418. https://doi.org/10.1016/J.BMC.2014.05.047.
7 Mirkin SM. DNA structures, repeat expansions and human hereditary disorders. *Curr Opin Struct Biol*. 2006;16:351–358. https://doi.org/10.1016/J.SBI.2006.05.004.
8 Lu XJ, Shakked Z, Olson WK. A-form Conformational Motifs in Ligand-bound DNA Structures. *J Mol Biol*. 2000;300:819–840. https://doi.org/10.1006/JMBI.2000.3690.
9 Spiegel J, Adhikari S, Balasubramanian S. The Structure and Function of DNA G-Quadruplexes. *Trends Chem*. 2020;2:123–136. https://doi.org/10.1016/J.TRECHM.2019.07.002.
10 Gehring K, Leroy JL, Guéron M. A tetrameric DNA structure with protonated cytosine-cytosine base pairs. *Nature*. 1993;363:561–565. https://doi.org/10.1038/363561a0.
11 Zeraati M, Langley DB, Schofield P, et al. I-motif DNA structures are formed in the nuclei of human cells. *Nat Chem*. 2018;10:631–637. https://doi.org/10.1038/s41557-018-0046-3.
12 Phan AT, Guéron M, Leroy JL. The solution structure and internal motions of a fragment of the cytidine-rich strand of the human telomere. *J Mol Biol*. 2000;299:123–144. https://doi.org/10.1006/jmbi.2000.3613.
13 Parkinson GN, Lee MPH, Neidle S. Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*. 2002;417:876–880. https://doi.org/10.1038/nature755.
14 King JJ, Irving KL, Evans CW, et al. DNA G-Quadruplex and i-Motif Structure Formation Is Interdependent in Human Cells. *J Am Chem Soc*. 2020;142:20600–20604. https://doi.org/10.1021/JACS.0C11708/SUPPL_FILE/JA0C11708_SI_001.PDF.
15 Martella M, Pichiorri F, Chikhale RV, et al. Motif formation and spontaneous deletions in human cells. *Nucl Acids Res*. 2022;50:3445–3455. https://doi.org/10.1093/NAR/GKAC158.
16 Harris LM, Monsell KR, Noulin F, et al. G-quadruplex DNA motifs in the malaria parasite plasmodium falciparum and their potential as novel antimalarial drug targets. *Antimicrob Agents Chemother*. 2018;62. https://doi.org/10.1128/AAC.01828-17.
17 Belmonte-Reche E, Martínez-García M, Guédin A, et al. G-Quadruplex Identification in the Genome of Protozoan Parasites Points to Naphthalene Diimide Ligands as New Antiparasitic Agents. *J Med Chem*. 2018;61:1231–1240. https://doi.org/10.1021/ACS.JMEDCHEM.7B01672.
18 Ruggiero E, Richter SN. G-quadruplexes and G-quadruplex ligands: targets and tools in antiviral therapy. *Nucleic Acids Res*. 2018;46:3270–3283. https://doi.org/10.1093/NAR/GKY187.
19 Dingley AJ, Peterson RD, Grzesiek S, Feigon J. Characterization of the cation and temperature dependence of DNA quadruplex hydrogen bond properties using high-resolution NMR. *J Am Chem Soc*. 2005;127:14466–14472. https://doi.org/10.1021/JA0540369/ASSET/IMAGES/MEDIUM/JA0540369N00001.GIF.
20 Burge S, Parkinson GN, Hazel P, Todd AK, Neidle S. Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res*. 2006;34:5402–5415. https://doi.org/10.1093/NAR/GKL655.
21 Dhakal S, Lafontaine JL, Yu Z, Koirala D, Mao H. Intramolecular Folding in Human ILPR Fragment with Three C-Rich Repeats. *PLoS One*. 2012;7:e39271.
22 Huppert JL, Balasubramanian S. Prevalence of quadruplexes in the human genome. *Nucleic Acids Res*. 2005;33:2908–2916. https://doi.org/10.1093/NAR/GKI609.
23 Wang J, Chen Y, Ren J, Zhao C, Qu X. G-Quadruplex binding enantiomers show chiral selective interactions with human telomere. *Nucleic Acids Res*. 2014;42:3792. https://doi.org/10.1093/NAR/GKT1354.
24 Yuan DH, Bell GI, Selby MJ, Rutter WJ. The highly polymorphic region near the human insulin gene is composed of simple tandemly repeating sequences. *nature.com*. 1982;295(9):503-517. Accessed May 20, 2022. https://idp.nature.com/authorize/casa?redirect_uri=https://www.nature.com/articles/295031a0&casa_token=qBwszCux4hUAAAAA:ZIB_jtZeUVlYyLAs6a7_q3035Wg8zkhaGPqgMEjxo3NNeJkqjJqVjIN5pZOrs_CePznS8gCq8lqjdbg.
25 Selvam S, Mandal S, Mao H. Quantification of Chemical and Mechanical Effects on the Formation of the G-Quadruplex and i-Motif in Duplex DNA. *Biochemistry*. 2017;56:4616–4625. https://doi.org/10.1021/ACS.BIOCHEM.7B00279/SUPPL_FILE/BI7B00279_SI_001.PDF.
26 Dhakal S, Yu Z, Konik R, Cui Y, Koirala D, Mao H. G-Quadruplex and i-Motif Are Mutually Exclusive in ILPR Double-Stranded DNA. *Biophys J*. 2012;102:2575. https://doi.org/10.1016/J.BPJ.2012.04.024.
27 Day HA, Huguin C, Waller ZAE. Silver cations fold i-motif at neutral pH. *Chem Commun*. 2013;49:7696–7698. https://doi.org/10.1039/C3CC43495H.
28 Sheng Q, Neaverson JC, Mahmoud T, Stevenson CEM, Matthews SE, Waller ZAE. Identification of new DNA i-motif binding ligands through a fluorescent intercalator displacement assay. *Org Biomol Chem*. 2017;15:5669. https://doi.org/10.1039/C7OB00710H.
29 Yuan G, Zhang Q, Zhou J, Li H. Mass spectrometry of G-quadruplex DNA: Formation, recognition, property, conversion, and conformation. *Mass Spectrom Rev*. 2011;30:1121–1142. https://doi.org/10.1002/MAS.20315.
30 Journey SN, Alden SL, Hewitt WM, Peach ML, Nicklaus MC, Schneekloth JS. Probing the hras -1 Y i-motif with small molecules. *Medchemcomm*. 2018;9:2000–2007. https://doi.org/10.1039/C8MD00311D.
31 Dwivedi A, Kumari A, Aarthy M, et al. Spectroscopic and molecular docking studies for the binding and interaction aspects of curcumin-cysteine conjugate and rosmarinic acid with human telomeric G-quadruplex DNA. *Int J Biol Macromol*. 2021;182:1463–1472. https://doi.org/10.1016/J.IJBIOMAC.2021.05.089.
32 Li D, Hou JQ, Long W, Lu YJ, Wong WL, Zhang K. A study on a telo21 G-quadruplex DNA specific binding ligand: enhancing the molecular recognition ability via the amino group interactions †. Published online 2018. 10.1039/c8ra03833c.
33 Verma SD, Pal N, Singh MK, Shweta H, Khan MF, Sen S. Understanding ligand interaction with different structures of G-quadruplex DNA: evidence of kinetically controlled ligand binding and binding-mode assisted quadruplex structure alteration. *Anal Chem*. 2012;84:7218–7226. https://doi.org/10.1021/AC3015998.
34 M G MS, Chikhale R, Nanaware PP, Dalvi S, Venkatraman P. A druggable pocket on PSMD10Gankyrin that can accommodate an interface peptide and doxorubicin. Eur J Pharmacol. 2022;915:174718. doi:10.1016/J.EJPHAR.2021.174718.
35 Haldar S, Zhang Y, Xia Y, et al. Mechanistic Insights into the Ligand-Induced Unfolding of an RNA G-Quadruplex. *J Am Chem Soc*. 2022;144:935–950. https://doi.org/10.1021/JACS.1C11248.
36 Selleckchem.com - Bioactive Compounds Expert (Bioactive Compounds, Compound Libraries). Accessed May 29, 2022. https://www.selleckchem.com/.
37 Drew HR, Wing RM, Takano T, et al. Structure of a B-DNA dodecamer: conformation and dynamics. *Proc Natl Acad Sci USA*. 1981;78:2179–2183. https://doi.org/10.1073/PNAS.78.4.2179.
38 Mixed ab initio QM/MM modeling using frozen orbitals and tests with alanine dipeptide and tetrapeptide - Philipp - 1999 - Journal of Computational Chemistry - Wiley Online Library. Accessed May 29, 2022. https://onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291096-987X%2819991115%2920%3A14%3C1468%3A%3AAID-JCC2%3E3.0.CO%3B2-0.
39 Wright EP, Day HA, Ibrahim AM, et al. Mitoxantrone and Analogues Bind and Stabilize i-Motif Forming DNA Sequences. *Sci Rep*. 2016;6:1–7. https://doi.org/10.1038/srep39456.
40 Abdelhamid MAS, Gates AJ, Waller ZAE. Destabilization of i-Motif DNA at Neutral pH by G-Quadruplex Ligands. *Biochemistry*. 2019;58:245–249. https://doi.org/10.1021/ACS.BIOCHEM.8B00968/SUPPL_FILE/BI8B00968_SI_001.PDF.
41 Weisman-Shomer P, Cohen E, Hershco I, et al. The cationic porphyrin TMPyP4 destabilizes the tetraplex form of the fragile X syndrome expanded sequence d(CGG)n. *Nucleic Acids Res*. 2003;31:3963–3970. https://doi.org/10.1093/NAR/GKG453.
42 Waller ZAE, Sewitz SA, Hsu STD, Balasubramanian S. A small molecule that disrupts G-quadruplex DNA structure and enhances gene expression. *J Am Chem Soc*. 2009;131:12628–12633. https://doi.org/10.1021/JA901892U/SUPPL_FILE/JA901892U_SI_001.PDF.
43 Kaluzhny D, Ilyinsky N, Shchekotikhin A, et al. Disordering of Human Telomeric G-Quadruplex with Novel Antiproliferative Anthrathiophenedione. *PLoS One*. 2011;6:e27151.
44 O Hagan MP, Haldar S, Duchi M, et al. A Photoresponsive Stiff-Stilbene Ligand Fuels the Reversible Unfolding of G-Quadruplex DNA. *Angew Chem Int Ed*. 2019;58:4334–4338. https://doi.org/10.1002/ANIE.201900740.