

# Discovery of novel triple helical DNA intercalators by an integrated virtual and actual screening platform

Patrick A. Holt<sup>1,3</sup>, Patricia Ragazon<sup>1</sup>, Lucjan Strekowski<sup>4</sup>,  
Jonathan B. Chaires<sup>1,2,3</sup> and John O. Trent<sup>1,2,3,\*</sup>

<sup>1</sup>James Graham Brown Cancer Center, <sup>2</sup>Department of Medicine, <sup>3</sup>Department of Biochemistry and Molecular Biology, University of Louisville, 529 S. Jackson Street, Louisville, KY 40202 and <sup>4</sup>Department of Chemistry, Georgia State University, Atlanta, GA 30302, USA

Received November 21, 2008; Revised December 12, 2008; Accepted December 13, 2008

## ABSTRACT

**Virtual Screening is an increasingly attractive way to discover new small molecules with potential medicinal value. We introduce a novel strategy that integrates use of the molecular docking software Surflex with experimental validation by the method of competition dialysis. This integrated approach was used to identify ligands that selectively bind to the triplex DNA poly(dA)-[poly(dT)]<sub>2</sub>. A library containing ~2 million ligands was virtually screened to identify compounds with chemical and structural similarity to a known triplex intercalator, the naphthylquinoline MHQ-12. Further molecular docking studies using compounds with high structural similarity resulted in two compounds that were then demonstrated by competition dialysis to have a superior affinity and selectivity for the triplex nucleic acid than MHQ-12. One of the compounds has a different chemical backbone than MHQ-12, which demonstrates the ability of this strategy to ‘scaffold hop’ and to identify small molecules with novel binding properties. Biophysical characterization of these compounds by circular dichroism and thermal denaturation studies confirmed their binding mode and selectivity. These studies provide a proof-of-principle for our integrated screening strategy, and suggest that this platform may be extended to discover new compounds that target therapeutically relevant nucleic acid morphologies.**

## INTRODUCTION

Nucleic acid sequences that form triple helices have become a source of increasing interest as a way to interfere

with DNA transcription and modulate gene expression (1–3). Several approaches attempt to use triplex nucleic acids to interfere with the transcription of genes, through either inducing the formation of triplex or stabilizing existing triplex nucleic acids. The former approach is the so-called ‘anti-gene’ approach and involves the administration of triplex forming oligonucleotides (TFOs), which are short sequences of nucleic acids that can bind to the major groove of duplex nucleic acids and promote the formation of triplex structures (4,5). TFOs have already been successful in reducing transcription of the c-myc oncogene that is located in the promoter site of genes (4,6). However, there are currently significant challenges associated with the use of TFOs and triplex structures in general. First, TFOs have significantly lower activity in cell-based systems, compared to *in vitro* systems (7). This has been ascribed to many factors including improper cellular localization or degradation of the oligonucleotide (7,8). A second limitation is the inherent low stability of many triplex structures (9,10). The latter limitation is the focus of this work where we demonstrate the use of a novel virtual and actual screening platform for identifying several compounds that can selectively bind to and stabilize a triplex nucleic acid structure. These newly identified small molecules could be used to target triplex structures in several ways. First, the small molecule could stabilize pre-existing triplex structures *in vivo*. The small molecules could be used in an adjuvant setting with TFOs to increase stability, or alternatively the small molecules can be linked to TFOs to enhance the stability of newly formed triplex structures (8). Either of these approaches could be used to control gene expression. These capabilities make these small molecules potentially clinically relevant for treating cancer and other diseases that are closely linked with abnormal gene expression.

Several small molecules are known to intercalate into and stabilize triplex nucleic acids including coralyne, benzo[*e*]pyridoindoles (BePI), benzo[*g*]pyridoindoles (BgPI),

\*To whom correspondence should be addressed. Tel: +1 502 852 2194; Fax: +1 502 852 4311; Email: john.trent@louisville.edu

dibenzophenanthrolines, and anthraquinones (10–14). One of the most selective and extensively studied classes are the naphthylquinolines, which have been shown to intercalate into the TAT DNA triplex, poly(dA)-[poly(dT)]<sub>2</sub> (15–17). Chaires *et al.* performed an extensive study (18) that characterized the selectivity and affinity of 14 naphthylquinoline derivatives. The ligand MHQ-12 emerged from that study as the compound with the highest affinity and greatest selectivity for the poly(dA)-[poly(dT)]<sub>2</sub> triplex. While this approach for the discovery of triplex-selective ligands was successful, it is a laborious and time-consuming process. We propose a novel alternative approach for finding ligands that target a particular structure in which virtual screening is used to identify promising ligand candidates followed by validation using competition dialysis. We demonstrate here that this approach can identify small molecules that intercalate into poly(dA)-[poly(dT)]<sub>2</sub> with higher selectivity and affinity than MHQ-12. A significant result of this approach is that a small molecule with a substantially different molecular scaffold was identified that has superior affinity and selectivity for triplex DNA compared to MHQ-12. This strategy thus provides a new platform for identifying promising small molecule drugs against nucleic acid targets.

Virtual screening using molecular similarity and docking methods is becoming an increasingly important and economical approach to identify small molecule drug candidates (19). While there are numerous studies using such screening methods for targeting proteins, far fewer virtual screening efforts have been performed targeting nucleic acids. The few studies that have been performed targeting nucleic acids have produced promising results. They have shown that screening methods can accurately reproduce crystallographic structures of ligand–nucleic acid complexes using a variety of docking programs including DOCK (20), Autodock (21) and Surflex (22). Our virtual screening approach uses both ligand and structure-based discovery principles to select ligands from a commercially available library that bind to poly(dA)-[poly(dT)]<sub>2</sub> with higher affinity and selectivity than MHQ-12. Initial virtual screening is performed with Surflex-Sim, which is a ligand similarity-based software program that has superior performance compared to most traditional 2D similarity methods (23). This program is an effective tool to rapidly pre-screen large virtual compound libraries to enrich for structurally similar ligands (23). Surflex-Sim maximizes 3D morphological similarity and alignment of a test ligand to the control ligand, which in this work was MHQ-12 (23–25). The quantitative metric that is used for evaluating Surflex-Sim results is the Surflex-Sim score, which embodies an all atom comparison and alignment of the test ligand with the control ligand. The top-ranked Surflex-Sim results were used for structure-based docking studies to dock the ligand to the intercalation site and the three grooves (major–major, major–minor and minor) (18) of the triplex structure using the docking program Surflex-Dock. Surflex-Dock performs docking of test ligands to a ‘protomol’ or idealized representation of the binding site on the nucleic acid target. The ligands are docked to the target and the poses are ranked by a

Surflex Raw Score (SRS) that consists of an affinity score of the ligand for the target (25). This sequential combination of Surflex-Sim followed by Surflex-Dock produced several ligands that had hypothesized higher binding affinity and selectivity for the triplex intercalation site, compared to MHQ-12.

A critically important step after virtual screening is validation by experimental testing of the top candidates. To accomplish this, competition dialysis was employed because of its extensive use to determine the selectivity and affinity of a small molecule for single-stranded, duplex, triplex and quadruplex nucleic acid targets (13,26–36). The advantage of competition dialysis is that it is not limited to the target sequence, or a simple comparison with another form of DNA, but with as many nucleic acid forms as are included in the assay. Competition dialysis involves dialyzing solutions of an array of nucleic acid sequences and structures against a common solution containing a test ligand (26). The solution is allowed to reach equilibrium, and the amount of ligand that is bound to each nucleic acid is measured using either fluorescence or absorbance (26). Comparison of the total and relative amounts of ligand bound to each nucleic acid assesses the affinity and selectivity, respectively, of the ligand for any included nucleic acid. Competition dialysis testing is used here to validate the affinity and selectivity of the top virtual screening hits. Circular dichroism and thermal denaturation were used for further characterization of the triplex binding of the top virtual screening candidate hits.

By using this integrated approach we have identified small molecules that have higher selectivity and affinity for triplex poly(dA)-[poly(dT)]<sub>2</sub> than the original molecule, MHQ-12, and which are among the most selective and tightest triplex binding molecules reported to date.

## MATERIALS AND METHODS

### Virtual library construction

The triplex-selective ligand MHQ-12 was constructed and hydrogen atoms were added using Macromodel (Schrodinger, Inc.). The ligand was energetically minimized using a sequential combination of 2000 iterations of a Steepest Descent algorithm followed by 2000 iterations of a Polak-Ribier Conjugate Gradient algorithm. AMBER ligand atom types were assigned using Sybyl (Tripos, Inc.). The program Antechamber in the software suite Amber (UCSF) was used to calculate AM1-BCC charges for the ligand and to convert to a MOL2 file format. A virtual set of 1.962 million ZINC compounds in MOL2 format were obtained from the ZINC 2007 ‘all-purchasable’ subset of ligands from the University of California San Francisco (37). The first 1.962 million ligands were downloaded out of a total of 2.7 million ligands from the 2007 ZINC ‘all-purchasable’ database. The ‘Reference’ subset of the ‘all-purchasable’ library was used directly as downloaded from the ZINC website, and was not filtered to select for ligands with any particular chemical property. The ligand 3D coordinates were generated by using detailed methods previously described (37).

Other ligand preparation procedures performed were protonation of the ligand based on a 'Reference' pH 7 condition and assignment of partial atomic charges from AMSOL semi-empirical quantum calculations (37). The triplex nucleic acid structure poly(dA)-[poly(dT)]<sub>2</sub> with an intercalation site was constructed using B-type parallel triplex (38) with X-ray structural intercalation site backbone fragments [Protein Data Bank entry 1p20.ent] and minimized holding the heavy atoms fixed.

### Surflex methods

The program Surflex (UCSF) containing both the Surflex-Sim molecular similarity and the Surflex-Dock molecular docking programs was run on 30 AMD Opteron 246 processors (2.0 GHz) running the Linux Red Hat operating system for all virtual screening experiments. Surflex-Sim experiments were performed using the 'align\_list' function to compare the MHQ-12 triplex selective ligand against 1.962 million compounds in the ZINC library. The top 350 ligands, ranked according to the highest Surflex-Sim score, were selected as candidates for Surflex-Dock studies and were extracted as individual MOL2 files from the library files using inhouse PERL scripts. The Surflex-Dock docking algorithm and scoring functions have been described in detail previously (39,40). Briefly, Surflex-Dock operates by a surface shape-based method, aligning each test ligand to a 'protomol.' The protomol consists of a series of molecular fragments that characterize the surface properties of the target active site including steric effects, hydrogen bond acceptor groups and hydrogen bond donor groups (39,40). After alignment of the ligands to the protomol, each pose is scored based on hydrophobic and polar contacts between atoms. The scoring function also includes a term for solvation although this contribution has typically been difficult to incorporate accurately in many molecular docking programs (41). Most importantly for these purposes, Surflex-Dock was previously used successfully to reproduce the crystal structures of several mono and dicationic small molecules bound to either the minor groove of nucleic acids or by intercalation between base pairs (22). This suggests that Surflex-Dock may have particular utility for finding new small molecules that target nucleic acid structures and thus is the basis for its use here. Finally, while many other programs are generally available for molecular docking, most have been used almost exclusively for protein-ligand docking, and have not been validated for nucleic acid-ligand docking. The previous published use of Surflex-Dock for successfully modeling binding of ligands to nucleic acids thus makes this program a rational selection for use in these studies (22). For the Surflex-Dock experiments, four protomols were generated to cover the major-major groove, major-minor groove, minor groove and intercalation sites of the triplex nucleic acid, using the same methods we previously described (22). The 'proto\_bloat' function was set to accommodate all reasonable interactions of the protomols with the triplex target sites. Docking of the ligands to the target was performed using a whole molecule approach, as described previously (22,25,39). The Surflex-Dock experiments

involved docking each of the ligands to all four protomols individually, in separate experiments. Surflex-Dock was operated with parameters 'Multistart 5' and 'Random 5' which we have previously shown returned accurate top-ranked docked poses for a set of small molecules to their respective nucleic acid targets (22). The Surflex-Dock poses were ranked according to the highest Surflex Raw Score. Surflex-Sim and Surflex-Dock poses were visualized using AutoDockTools (The Scripps Research Institute). The properties of compounds 1 and 2 used in the QSAR analysis were generated with QikProp (Schrodinger, Inc.).

### Compounds for biophysical testing

The highest ranked candidates identified by virtual screening were the ligands with ZINC identification numbers 632255 and 4623551, which will be referred to hereafter as compound 1 and compound 2, respectively. Compound 1 (42), is 4-(4-methylpiperazino-2-(2-naphthyl)quinoline and was obtained from Sigma-Aldrich (Milwaukee, WI, USA). Compound 2 is 1-phenyl-4-pyrrolidino-2,3-dihydro-1/H/-pyrrolo[2,3-b/]quinoline and was obtained from Chemical Block (N.D. Zelinsky Institute of Organic Chemistry, Moscow, Russia). As positive controls, the known triplex selective ligands MHQ-15 and OZ-85H were synthesized as previously described (18).

### Competition dialysis method

Competition dialysis experiments were done as previously described (18,27,28,43). The array of oligonucleotides used is given in Supplementary Table S1. The methods used to construct the oligonucleotides are elaborated in detail by Ren and Chaires (26). Briefly, natural DNA including calf thymus, *Micrococcus lysodeikticus* and *Clostridium perfringens* were obtained from Sigma (St Louis, MO, USA) and sonicated, phenol extracted and purified using methods previously described (26,44). The remaining 30-nt long oligonucleotides including poly(dA), poly(dT), poly (dAdT), poly (dAdT)-(dAdT) and poly (dGdC)-(dGdC) were synthesized and obtained from Pharmacia Biotech (Piscataway, NJ, USA). Poly(rA), poly(rU) and quadruplex sequences were obtained from Integrated DNA Technologies (Coralville, IA, USA). The Z-DNA structure was produced by bromination of poly (dGdC)-(dGdC), as previously described (45). Remaining duplex nucleic acids such as poly(rArU) were produced by mixing poly(rA) and poly(rU) in a 1:1 equimolar ratio, heating at 90°C for 2min and subsequently cooling to room temperature. Similarly, triplex nucleic acids were made by mixing poly (dAdT) with poly(dT) in a 1:1 equimolar ratio, heating at 90°C for 2min and subsequently cooling to room temperature. Quadruplex nucleic acid solutions were heated at 90°C, cooled to room temperature and left at 4°C for 48 h prior to use in competition dialysis. All nucleic acid solutions were extensively dialyzed against Na<sub>2</sub>HPO<sub>4</sub> (6 mM), NaH<sub>2</sub>PO<sub>4</sub> (2 mM), NaCl (185 mM), EDTA (0.1 mM), pH 7 using Pierce (Rockford, IL, USA) 3500 Da molecular weight cutoff dialysis cassettes. All nucleic acid structures were characterized by circular dichroism spectra, UV absorbance spectra and thermal denaturation, as

previously described (26). All nucleic acid samples were of identical concentration of 75  $\mu$ M, expressed in terms of monomeric unit (base pairs for duplex DNA, triplets for triplex DNA and tetrads for quadruplex DNA). The monomer concentration scale was used to alleviate possible problems arising from length differences among the polynucleotide and natural DNA samples. At the end of the dialysis equilibration period, ligand concentrations were determined by fluorescence. A volume of 180  $\mu$ l of each sample was carefully transferred into one well of a 96-well microtiter plate (Costar<sup>®</sup> cat 3915; Corning Inc., Corning, NY, USA). To each sample, 20  $\mu$ l of a 10% (w/v) sodium dodecyl sulfate (SDS) stock solution was added to give a final concentration of 1% (w/v) SDS. The technique of adding SDS has been used extensively by many research groups for dissociating ligands from nucleic acid structures (46). Dissociation of the ligand from the DNA is critical to ensure that there are no complexities arising from differences in the optical properties of free and bound ligands. The total ligand concentration ( $C_t$ ) within each dialysis well was determined using a fluorescence standard curve for each tested ligand. Appropriate corrections were made for the small dilution resulting from the addition of the SDS stock solution. The free ligand concentration ( $C_f$ ) was determined from an aliquot of the dialysate solution, which typically did not vary significantly from the initial 1  $\mu$ M concentration. Fluorescence measurements were made using a Safire microplate reader (Tecan US, Durham, NC, USA), with the following parameters: excitation and emission bandwidth, 10 nm, gain: 100. Compound OZ-85H: exc./emission 260/494 nm, compound MHQ-15: exc./emission 260/437 nm, compound 1 exc./emission 260/490 nm, compound 2: exc./emission 348/446 nm. The bound ligand concentration ( $C_b$ ) was then determined by:

$$C_b = C_t - C_f \quad 1$$

Binding constants, specificity sums (SS) and the ratio  $C_b/SS$  were calculated as previously described (46) and are provided in Supplementary Equation (S1).

### CD titration and thermal denaturation methods

CD titrations were done as previously described (47), using a Jasco 810 spectropolarimeter. Instrument settings were: wavelength range (220–500 nm), scan rate (100 nm  $\text{min}^{-1}$ ), averaging time (0.125 s), bandwidth (1 nm), number of scans (2), temperature (20°C). The effect of ligands on the thermal denaturation of triplex DNA was studied using the exact protocol described previously (46).

## RESULTS AND DISCUSSION

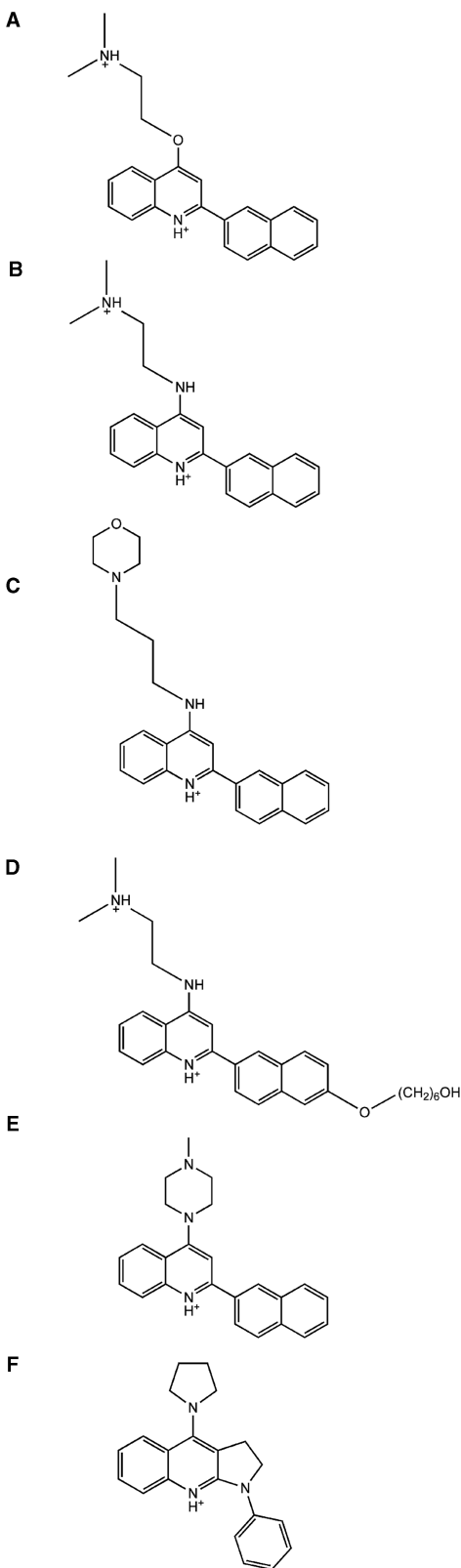
### Virtual screening

The initial step in virtual screening was performing Surflex-Sim to determine which of the ligands in the library were most structurally similar to the known, triplex selective intercalator MHQ-12. Of the approximately 2 million ligands screened for similarity against MHQ-12, 350 ligands had a Surflex-Sim score of greater than 0.70

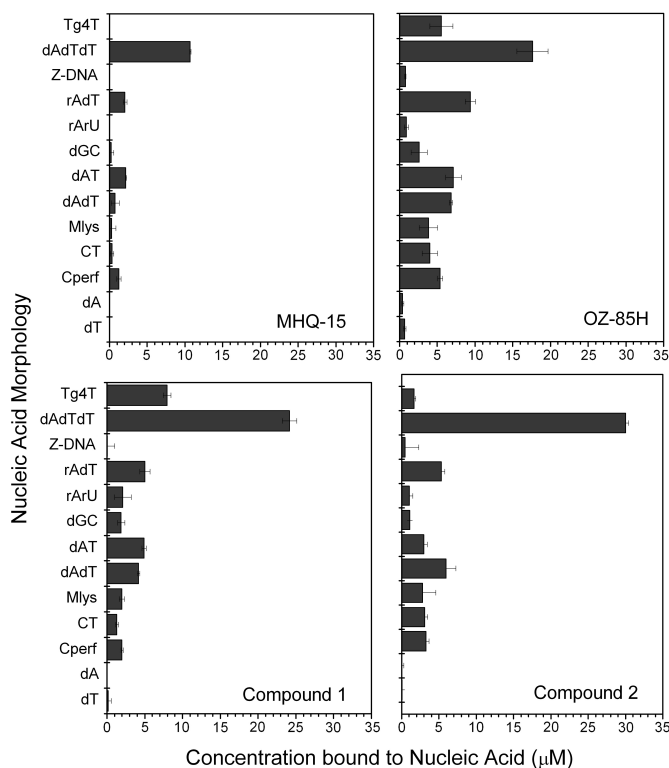
(range 0.875–0.704) and were selected for Surflex-Docking studies. A cutoff Surflex-Sim score of 0.70 was selected based on previous studies which suggested that this is the lowest score where the ligand structure–function relationship is typically maintained (25). The next step in the virtual screening process involved performing Surflex-Dock studies with the top 350 ranked Surflex-Sim ligands using the intercalation site and the three grooves (major–major, major–minor and minor) of the triplex as individual docking targets. Interestingly, MHQ-12 has the top Surflex Raw Score out of all 350 ligands that were docked to the intercalation site, which directly supports the ability of Surflex-Dock to successfully dock and rank a known selective triplex intercalator. We propose a new metric to evaluate the Surflex-Dock results, the ‘Normalized Surflex Raw Score (NSRS)’. The rationale behind the normalization of the Surflex Raw Score is that the score for a ligand binding to a single site on a target measures only the interaction with that one site. However, a ligand may have multiple interaction sites on a particular target. Therefore, for selectivity for a particular mode of binding, it is crucial to determine the binding of the ligand to the site of interest *relative* to the binding to other potential sites on the target. Since ligands interact with nucleic acids typically through either the groove-binding or intercalation, protomols were constructed at the three grooves and the intercalation site (48). Binding of the ligand to the intercalation site *relative* to binding in the three grooves embodies the ‘normalized’ affinity and specificity of the ligand for triplex intercalation. The following metric determines the NSRS for the intercalation site for each of the top 350 Surflex-Dock results:

$$\begin{aligned} \text{NSRS}_{\text{intercalation site}} &= \text{SRS}_{\text{intercalation site}} \\ &\quad - \text{Maximum SRS}_{\text{major–major site, major–minor site, minor site}} \end{aligned} \quad 2$$

Ranking of the 350 intercalation site Surflex-Dock results by NSRS shows that only three ligands have a higher NSRS score than MHQ-12 (NSRS of 6.8) (Figure 1A). The ligands are LS-08 (49) (Figure 1B), compounds 1 (Figure 1E) and 2 (Figure 1F) and have NSRS values of 7.03, 7.34 and 7.39, respectively (Figure 1). Interestingly, LS-08 (Figure 1B) which was identified by our virtual screening methodology, was previously tested by Chairès (46) and shown to be highly triplex selective, which adds validity to our virtual screening approach used to identify triplex selective ligands. Based on the NSRS values, compounds 1 (Figure 1E) and 2 (Figure 1F) were hypothesized to have superior affinity and selectivity for binding to the triplex nucleic acid, and were tested by competition dialysis. Two known triplex selective compounds, MHQ-15 (Figure 1C) and OZ-85H (Figure 1D) served as positive controls, as these compounds have been extensively studied and characterized (18). Biophysical characterization was performed by circular dichroism and thermal denaturation to assess the ability of the compounds to intercalate into the DNA triplex.



**Figure 1.** Chemical structures of the ligands used in virtual screening and competition dialysis experiments. (A) MHQ-12, (B) LS-08, (C) MHQ-15, (D) OZ-85H, (E) compound 1 and (F) compound 2.



**Figure 2.** Competition dialysis results for MHQ-15, OZ-85H, compounds 1 and 2. The concentration of bound ligand to each nucleic acid structure in the array is shown.

### Competition dialysis

The results of the competition dialysis experiments are shown in Figure 2. It is visually apparent that compounds 1 and 2 have a much higher affinity for the TAT triplex than the two positive control reference compounds, MHQ-15 and OZ-85H. The competition dialysis results for MHQ-12 have previously been described in detail (18), and this compound has an SS of 1.32 and a  $C_{\max}/SS$  ratio of 8.93. Determination of the SS (Table 1) for compounds 1 and 2 demonstrates superior triplex selectivity compared to OZ-85H but slightly less selectivity than MHQ-12 and MHQ-15. However, the significantly higher binding affinities of compounds 1 and 2 translate to much higher  $C_{\max}/SS$  values than MHQ-12, MHQ-15 or OZ-85H. The  $C_{\max}/SS$  ratio for compounds 1 and 2 is significant as it suggests that compounds 1 and 2 have a superior combination of binding affinity and selectivity compared to the reference compounds. These results validate the virtual screening approach, and show that the method can be used to identify compounds with high affinity and selectivity for a target nucleic acid, in this case the DNA TAT triplex.

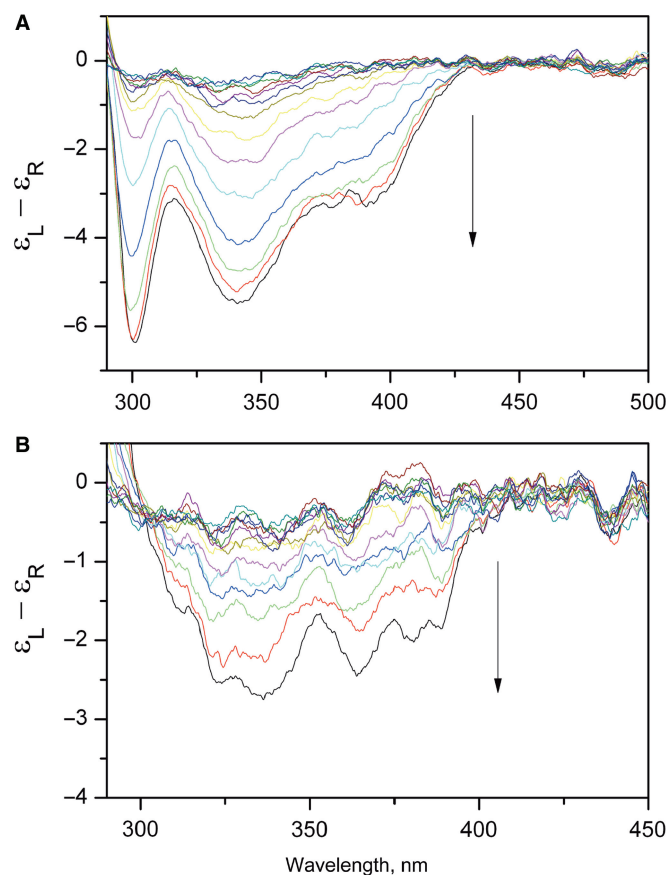
### Circular dichroism

The interaction of compounds 1 and 2 with DNA was studied by circular dichroism (Figure 3). Both compounds show pronounced induced circular dichroism (ICD) in the

**Table 1.** Competition dialysis metric results for the positive controls, MHQ-15, OZ-85H and the virtual screening top results, compounds 1 and 2

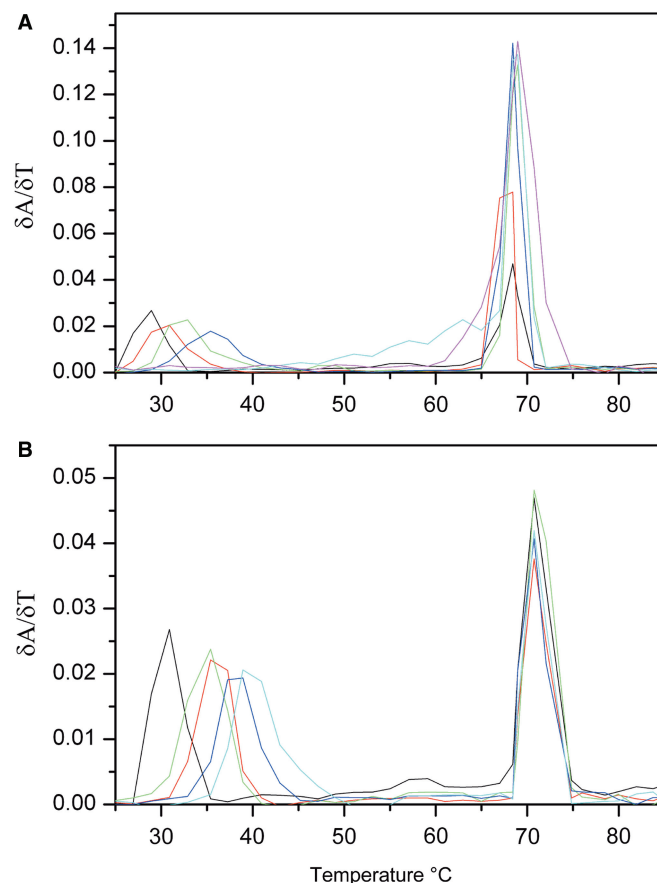
Test ligand	$C_b$ ( $\mu\text{M}$ )	$K_{\text{app}}/10^5$ ( $\text{M}^{-1}$ )	SS	$C_{\text{max}}/\text{SS}$ ( $\mu\text{M}$ )
MHQ-12 <sup>a</sup>	11.8	1.87	1.32	8.93
MHQ-15	10.7	1.7	1.66	6.44
OZ-85H	17.6	3.1	3.69	4.77
Compound 1	24.2	4.8	2.30	10.47
Compound 2	30.0	6.7	1.92	15.63

<sup>a</sup>These data are included from a previous publication (18).



**Figure 3.** Induced circular dichroism results for (A) compound 1 and (B) compound 2. (A) Spectra are shown for a ligand concentration of 45  $\mu\text{M}$  in the presence of triplex DNA ranging from 5  $\mu\text{M}$  to 450  $\mu\text{M}$  triplets. (B) Spectra are shown for a ligand concentration of 22.5  $\mu\text{M}$  in the presence of triplex DNA ranging from 2.25  $\mu\text{M}$  to 225  $\mu\text{M}$  triplets.

presence of triplex DNA. The ICD is in a spectral range where the compounds absorb light but the DNA does not. This ICD is unambiguous proof of the ligand binding to triplex DNA. For both compounds 1 and 2, the ICD is negative in sign, and relatively weak in magnitude. Such behavior is consistent with an intercalative binding mode, although the mode of binding can only be definitively established by high-resolution experimental structural analysis (50).



**Figure 4.** Thermal melting results for (A) compound 1 and (B) compound 2. Derivative melting curves were obtained using 32  $\mu\text{M}$  triplex DNA and ligand concentrations ranging from 0 to 16  $\mu\text{M}$  (A) or 0 to 32  $\mu\text{M}$  (B). The peak near 30°C is for the melting of the third strand, while that near 70°C is for melting of the duplex.

### Thermal denaturation studies

Figure 4 shows the effects of compounds 1 and 2 on the thermal denaturation of the TAT triplex. In the absence of added ligand, two transitions are seen, corresponding to the melting of the third strand ( $\sim 30^\circ\text{C}$ ) and the duplex ( $\sim 70^\circ\text{C}$ ). Titration with both ligands results in a clear elevation of the first transition, indicating stabilization of the triplex. The effect is maximal at saturating concentrations of ligand (1:1, ligand:triplet), where melting of the triplex coalesces with duplex melting. Melting of the triplex is stabilized by  $\sim 40^\circ\text{C}$  indicating tight binding of both compounds. Neither compound 1 nor compound 2 alters the transition temperature of the duplex form to any appreciable extent, an observation that is fully consistent with the weak binding to duplex seen in competition dialysis experiments (Figure 2).

### Validation of QSAR

In the previous study of naphthylquinoline binding to triplex DNA (18), a QSAR was derived from competition

dialysis binding data. The best three-term QSAR to emerge was:

$$\log K_{\text{app}} = 0.00264(\pm 0.00065)\text{SASA} - 0.693(\pm 0.125)\text{EA} \\ - 0.196(\pm 0.02)\text{HB}_a + 4.66(\pm 0.44) \\ N = 14; R = 0.959; \text{RMSE} = 0.130; \\ F = 49.84; P = 0.0001$$

3

In this relationship,  $\log K_{\text{app}}$  is the logarithm of the apparent binding constant (Table 1), SASA is the total solvent accessible surface area in  $\text{\AA}^2$ , EA is electron affinity in eV and  $\text{HB}_a$  is the number of hydrogen bond acceptors. The physical meaning of this is as follows. As SASA increases,  $\log K_{\text{app}}$  increases in magnitude, indicating higher affinity for triplex DNA. Increases in the magnitudes of EA and  $\text{HB}_a$  result in decreasing binding affinity. Increasing the solvent-accessible surface areas of naphthylquinoline compounds results in higher affinity for the triplex. Greater electron affinity and more hydrogen bond acceptors reduce the affinity of naphthylquinolines for triplex DNA.

Binding data obtained for compounds 1 and 2 in this study validate the published QSAR. The molecular descriptors SASA, EA and  $\text{HB}_a$  were calculated using QikProp, and substituted into Equation (3). For compound 1,  $\log K_a = 5.07$  was predicted, compared to a measured value of  $\log K_{\text{obs}} = 5.68$ . For compound 2,  $\log K$ -values of 5.18 and 5.82 were calculated and observed, respectively. The differences in calculated and observed values correspond to a factor of about four in binding constants, an acceptable agreement for predictions from a QSAR.

### Triplex-binding ligands

While these newly reported compounds have some of the tightest binding properties found to date by competition dialysis, it is worth considering limitations to this platform. One consideration is that the array of nucleic acids used in the assay is not all-encompassing and represents only a broad number of structural and sequence variants. Certain morphologies of nucleic acids to which ligands may bind may not be present in the array. To address this, current efforts are underway to expand the assay to include a larger number of nucleic acid solutions. This limitation also highlights however that a key benefit of competition dialysis is the ability to customize any number of targets for the assay. As new potentially therapeutic nucleic acid structures are discovered, they can be added to the assay and screened using this platform. Another consideration from these results is that the assay remains particularly relevant as an *in vitro* test, but may not necessarily reflect *in vivo* conditions and ligand binding behavior. This is particularly true when considering differences in the abundance of various forms of nucleic acids. For example, because of the abundance of duplex DNA, the observed approximately 5-fold difference in binding of compound 2 to the TAT triplex may be insufficient to overcome non-specific binding of the ligand to duplex DNA *in vivo*. However, this may be irrelevant if ligands are discovered that bind with true selectivity and high affinity to only a single nucleic acid structure.

While these are some relevant considerations when using this platform, the approach as demonstrated here is valuable for identifying potential small molecules with desirable binding properties for a single nucleic acid structure.

### CONCLUSION

This work demonstrates a novel strategy for discovering small molecules that can selectively bind to the triplex nucleic acid, poly(dA)-[poly(dT)]<sub>2</sub>. Through the combination of virtual screening by Surfex and experimental validation by competition dialysis, compounds 1 and 2 were discovered. These compounds have the highest overall affinities and selectivities reported for triplex binders as determined by competition dialysis. Further biophysical characterization by circular dichroism and thermal melting confirmed the mechanism of action of these new compounds and verified the predictive nature of the virtual screening methodologies. Several aspects of the virtual screening results are noteworthy. First, the combination of a ligand-based (Surflex-Sim) with a structure-based approach (Surflex-Dock) proved to be a powerful and highly computationally efficient way to identify triplex selective small molecules, as Surflex-Sim is two orders of magnitude faster than Surflex-Dock. Second, our development of the NSRS metric, which can predict a particular mode of binding of triplex selective ligands with both similar and different (scaffold hopping) chemical scaffolds. This is significant as it has the potential to identify new classes of small molecules that may have much higher affinity and selectivity for a given nucleic acid target. Future work will focus on extending this integrated virtual and actual screening platform to target other nucleic acid structures that may hold medicinal value and physiological relevance.

### FUNDING

National Institutes of Health [1R01GM077422]; James Graham Brown Foundation. Funding for open access charge: National Institutes of Health [1R01GM077422].

*Conflict of interest statement.* None declared.

### REFERENCES

- Chin, J.Y., Schleifman, E.B. and Glazer, P.M. (2007) Repair and recombination induced by triple helix DNA. *Front. Biosci.*, **12**, 4288–4297.
- Seidman, M.M., Puri, N., Majumdar, A., Cuenoud, B., Miller, P.S. and Alam, R. (2005) The development of bioactive triple helix-forming oligonucleotides. *Ann. NY Acad. Sci.*, **1058**, 119–127.
- Kalota, A., Shetzline, S.E. and Gewirtz, A.M. (2004) Progress in the development of nucleic acid therapeutics for cancer. *Cancer Biol. Ther.*, **3**, 4–12.
- Seidman, M.M. and Glazer, P.M. (2003) The potential for gene repair via triple helix formation. *J. Clin. Invest.*, **112**, 487–494.
- Buchini, S. and Leumann, C.J. (2003) Recent improvements in antigene technology. *Curr. Opin. Chem. Biol.*, **7**, 717–726.
- Kim, H.G., Reddoch, J.F., Mayfield, C., Ebbinghaus, S., Vigneswaran, N., Thomas, S., Jones, D.E. Jr. and Miller, D.M. (1998)

- Inhibition of transcription of the human c-myc protooncogene by intermolecular triplex. *Biochemistry*, **37**, 2299–2304.
7. Carbone, G.M., McGuffie, E., Napoli, S., Flanagan, C.E., Dembech, C., Negri, U., Arcamone, F., Capobianco, M.L. and Catapano, C.V. (2004) DNA binding and antigene activity of a daunomycin-conjugated triplex-forming oligonucleotide targeting the P2 promoter of the human c-myc gene. *Nucleic Acids Res.*, **32**, 2396–2410.
  8. Lacoste, J., Francois, J.C. and Helene, C. (1997) Triple helix formation with purine-rich phosphorothioate-containing oligonucleotides covalently linked to an acridine derivative. *Nucleic Acids Res.*, **25**, 1991–1998.
  9. Marchand, C., Bailly, C., Nguyen, C.H., Bisagni, E., Garestier, T., Helene, C. and Waring, M.J. (1996) Stabilization of triple helical DNA by a benzopyridoquinoxaline intercalator. *Biochemistry*, **35**, 5022–5032.
  10. Moraru-Allen, A.A., Cassidy, S., Cassio Alvarez, J.L., Fox, K.R., Brown, T. and Lane, A.N. (1997) Coralyne has a preference for intercalation between TA.T triples in intramolecular DNA triple helices. *Nucleic Acids Res.*, **25**, 1890–1896.
  11. Keppler, M., Zegrocka, O., Strekowski, L. and Fox, K.R. (1999) DNA triple helix stabilisation by a naphthylquinoline dimer. *FEBS Lett.*, **447**, 223–226.
  12. Baudoin, O., Marchand, C., Teulade-Fichou, M.-P., Vigneron, J.-P., Sun, J.-S., Garestier, T., Helene, C. and Lehn, J.-M. (1998) Stabilization of DNA triple helices by crescent-shaped dibenzophenanthrolines. *Chem. Eur. J.*, **4**, 1504–1508.
  13. Ren, J., Bailly, C. and Chaires, J.B. (2000) NB-506, an indolocarbazole topoisomerase I inhibitor, binds preferentially to triplex DNA. *FEBS Lett.*, **470**, 355–359.
  14. Strekowski, L., Hojjat, M., Wolinska, E., Parker, A.N., Paliakov, E., Gorecki, T., Tanious, F.A. and Wilson, W.D. (2005) New triple-helix DNA stabilizing agents. *Bioorg. Med. Chem. Lett.*, **15**, 1097–1100.
  15. Cassidy, S.A., Strekowski, L. and Fox, K.R. (1996) DNA sequence specificity of a naphthylquinoline triple helix-binding ligand. *Nucleic Acids Res.*, **24**, 4133–4138.
  16. Keppler, M.D., James, P.L., Neidle, S., Brown, T. and Fox, K.R. (2003) DNA sequence specificity of triplex-binding ligands. *Eur. J. Biochem.*, **270**, 4982–4992.
  17. Strekowski, L., Say, M., Zegrocka, O., Tanious, F.A., Wilson, W.D., Manzel, L. and Macfarlane, D.E. (2003) Bis-4-aminoquinolines: novel triple-helix DNA intercalators and antagonists of immunostimulatory CpG-oligodeoxynucleotides. *Bioorg. Med. Chem.*, **11**, 1079–1085.
  18. Chaires, J.B., Ren, J., Henary, M., Zegrocka, O., Bishop, G.R. and Strekowski, L. (2003) Triplex selective 2-(2-naphthyl)quinoline compounds: origins of affinity and new design principles. *J. Am. Chem. Soc.*, **125**, 7272–7283.
  19. Kapetanovic, I.M. (2008) Computer-aided drug discovery and development (CADD): In silico-chemico-biological approach. *Chem. Biol. Interact.*, **171**, 165–176.
  20. Chen, Q., Shafer, R.H. and Kuntz, I.D. (1997) Structure-based discovery of ligands targeted to the RNA double helix. *Biochemistry*, **36**, 11402–11407.
  21. Evans, D.A. and Neidle, S. (2006) Virtual screening of DNA minor groove binders. *J. Med. Chem.*, **49**, 4232–4238.
  22. Holt, P.A., Chaires, J.B. and Trent, J.O. (2008) Molecular docking of intercalators and groove-binders to nucleic acids using Autodock and Surflex. *J. Chem. Inf. Model*, **48**, 1602–1615.
  23. Jain, A.N. (2004) Ligand-based structural hypotheses for virtual screening. *J. Med. Chem.*, **47**, 947–961.
  24. Cleves, A.E. and Jain, A.N. (2006) Robust ligand-based modeling of the biological targets of known drugs. *J. Med. Chem.*, **49**, 2921–2938.
  25. Cleves, A.E. and Jain, A.N. (2008) Effects of inductive bias on computational evaluations of ligand-based modeling and on drug discovery. *J. Comput. Aided Mol. Des.*, **22**, 147–159.
  26. Ren, J. and Chaires, J.B. (1999) Sequence and structural selectivity of nucleic acid binding ligands. *Biochemistry*, **38**, 16067–16075.
  27. Chaires, J.B. (2005) Competition dialysis: an assay to measure the structural selectivity of drug-nucleic acid interactions. *Curr. Med. Chem. Anticancer Agents*, **5**, 339–352.
  28. Ragazzon, P.A., Garbett, N.C. and Chaires, J.B. (2007) Competition dialysis: a method for the study of structural selective nucleic acid binding. *Methods*, **42**, 173–182.
  29. Ragazzon, P. and Chaires, J.B. (2007) Use of competition dialysis in the discovery of G-quadruplex selective ligands. *Methods*, **43**, 313–323.
  30. Carrasco, C., Rosu, F., Gabelica, V., Houssier, C., De Pauw, E., Garbay-Jauregui, C., Roques, B., Wilson, W.D., Chaires, J.B., Waring, M.J. et al. (2002) Tight binding of the antitumor drug ditercalinium to quadruplex DNA. *Chembiochem.*, **3**, 1235–1241.
  31. Shi, X. and Chaires, J.B. (2006) Sequence- and structural-selective nucleic acid binding revealed by the melting of mixtures. *Nucleic Acids Res.*, **34**, e14.
  32. Murphy, P.M., Phillips, V.A., Jennings, S.A., Garbett, N.C., Chaires, J.B., Jenkins, T.C. and Wheelhouse, R.T. (2003) Biarylpyrimidines: a new class of ligand for high-order DNA recognition. *Chem. Commun.*, **10**, 1160–1161.
  33. Chaires, J.B., Ren, J., Hamelberg, D., Kumar, A., Pandya, V., Boykin, D.W. and Wilson, W.D. (2004) Structural selectivity of aromatic diamidines. *J. Med. Chem.*, **47**, 5729–5742.
  34. Xing, F., Song, G., Ren, J., Chaires, J.B. and Qu, X. (2005) Molecular recognition of nucleic acids: coralyne binds strongly to poly(A). *FEBS Lett.*, **579**, 5035–5039.
  35. Ren, J., Qu, X., Dattagupta, N. and Chaires, J.B. (2001) Molecular recognition of a RNA:DNA hybrid structure. *J. Am. Chem. Soc.*, **123**, 6742–6743.
  36. Arya, D.P., Xue, L. and Tennant, P. (2003) Combining the best in triplex recognition: synthesis and nucleic acid binding of a BQQ-neomycin conjugate. *J. Am. Chem. Soc.*, **125**, 8070–8071.
  37. Irwin, J.J. and Shoichet, B.K. (2005) ZINC—a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model*, **45**, 177–182.
  38. Trent, J.O. (2001) Molecular modeling of drug-DNA complexes: an update. *Methods Enzymol.*, **340**, 290–326.
  39. Jain, A.N. (2003) Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. *J. Med. Chem.*, **46**, 499–511.
  40. Jain, A.N. (2007) Surflex-Dock 2.1: robust performance from ligand energetic modeling, ring flexibility, and knowledge-based search. *J. Comput. Aided Mol. Des.*, **21**, 281–306.
  41. Moitessier, N., Englebienne, P., Lee, D., Lawandi, J. and Corbeil, C.R. (2008) Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. *Br. J. Pharmacol.*, **153**(Suppl 1), S7–S26.
  42. Strekowski, L., Gulevich, Y., Baranowski, T.C., Parker, A.N., Kiselyov, A.S., Lin, S.Y., Tanious, F.A. and Wilson, W.D. (1996) Synthesis and structure-DNA binding relationship analysis of DNA triple-helix specific intercalators. *J. Med. Chem.*, **39**, 3980–3983.
  43. Chaires, J.B. (2005) *Structural Selectivity of Drug-Nucleic Acid Interactions Probed by Competition Dialysis*, Springer-Verlag GMBH & Co., Heidelberg.
  44. Chaires, J.B., Dattagupta, N. and Crothers, D.M. (1982) Studies on interaction of anthracycline antibiotics and deoxyribonucleic acid: equilibrium binding studies on interaction of daunomycin with deoxyribonucleic acid. *Biochemistry*, **21**, 3933–3940.
  45. Moller, A., Nordheim, A., Kozlowski, S.A., Patel, D.J. and Rich, A. (1984) Bromination stabilizes poly(dG-dC) in the Z-DNA form under low-salt conditions. *Biochemistry*, **23**, 54–62.
  46. Chaires, J.B. (2003) A competition dialysis assay for the study of structure-selective ligand binding to nucleic acids. In Beaucage, S.L. (ed), *Current Protocols in Nucleic Acid Chemistry*, John Wiley & Sons, Inc., New York, NY. Chapter 8. Unit 8.3.
  47. Garbett, N.C., Ragazzon, P.A. and Chaires, J.B. (2007) Circular dichroism to determine binding mode and affinity of ligand-DNA interactions. *Nat. Protoc.*, **2**, 3166–3172.
  48. Chaires, J.B. (2006) A thermodynamic signature for drug-DNA binding mode. *Arch. Biochem. Biophys.*, **453**, 26–31.
  49. Wilson, W.D., Zhao, M., Patterson, S.E., Wydra, R.L., Janda, L., Strekowski, L. and Schinazi, R.F. (1992) Design of RNA interactive anti-HIV agents: unfused aromatic intercalators. *Med. Chem. Res.*, **2**, 102–110.
  50. Eriksson, M. and Norden, B. (2001) Linear and circular dichroism of drug-nucleic acid complexes. *Methods Enzymol.*, **340**, 68–98.