

## Accepted Manuscript

Identification of G-quadruplex DNA/RNA binders: Structure-based virtual screening and biophysical characterization

Roberta Rocca, Federica Moraca, Giosuè Costa, Matteo Nadai, Matteo Scalabrin, Carmine Talarico, Simona Distinto, Elias Maccioni, Francesco Ortuso, Anna Artese, Stefano Alcaro, Sara N. Richter

PII: S0304-4165(16)30517-7  
DOI: doi:[10.1016/j.bbagen.2016.12.023](https://doi.org/10.1016/j.bbagen.2016.12.023)  
Reference: BBAGEN 28723

To appear in: *BBA - General Subjects*

Received date: 14 October 2016  
Revised date: 20 December 2016  
Accepted date: 21 December 2016



Please cite this article as: Roberta Rocca, Federica Moraca, Giosuè Costa, Matteo Nadai, Matteo Scalabrin, Carmine Talarico, Simona Distinto, Elias Maccioni, Francesco Ortuso, Anna Artese, Stefano Alcaro, Sara N. Richter, Identification of G-quadruplex DNA/RNA binders: Structure-based virtual screening and biophysical characterization, *BBA - General Subjects* (2016), doi:[10.1016/j.bbagen.2016.12.023](https://doi.org/10.1016/j.bbagen.2016.12.023)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

## Identification of G-quadruplex DNA/RNA binders: Structure-based virtual screening and biophysical characterization

Roberta Rocca<sup>1§</sup>, Federica Moraca<sup>1§</sup>, Giosuè Costa<sup>1\*</sup>, Matteo Nadai<sup>2</sup>, Matteo Scalabrin<sup>2</sup>, Carmine Talarico<sup>1</sup>, Simona Distinto<sup>3</sup>, Elias Maccioni<sup>3</sup>, Francesco Ortuso<sup>1</sup>, Anna Artese<sup>1</sup>, Stefano Alcaro<sup>1\*</sup>, Sara N. Richter<sup>2</sup>

<sup>1</sup>Dipartimento di Scienze della Salute, Università degli Studi "Magna Graecia" di Catanzaro, Campus "Salvatore Venuta", viale Europa, 88100, Catanzaro, Italy.

<sup>2</sup>Department of Molecular Medicine, University of Padua, via Gabelli 63, 35121, Padua, Italy.

<sup>3</sup>Dipartimento di Scienze della Vita e dell'Ambiente, Università degli Studi di Cagliari, via Ospedale 72, Cagliari 09124, Italy.

### *§ Equal Contributors*

*\*Corresponding Authors:* Stefano Alcaro (alcaro@unicz.it) and Giosuè Costa (gcosta@unicz.it), Università degli Studi "Magna Graecia" di Catanzaro, Campus "Salvatore Venuta", viale Europa, 88100, Catanzaro, Italy; *Phone* +39 0961 3694197; *Fax* +39 0961 391270.

**Running Title:** Naphthyridin and chromen derivatives as DNA/TERRA-RNA binders.

**ABSTRACT**

*Background:* Recent findings demonstrated that, in mammalian cells, telomere DNA (Tel) is transcribed into telomeric repeat-containing RNA (TERRA), which is involved in fundamental biological processes, thus representing a promising anticancer target. For this reason, the discovery of dual (as well as selective) Tel/TERRA G-quadruplex (G4) binders could represent an innovative strategy to enhance telomerase inhibition.

*Methods:* Initially, docking simulations of known Tel and TERRA active ligands were performed on the 3D coordinates of bimolecular G4 Tel DNA (Tel<sub>2</sub>) and TERRA (TERRA<sub>2</sub>). Structure-based pharmacophore models were generated on the best complexes and employed for the Virtual Screening of ~257000 natural compounds. The 20 best candidates were submitted to biophysical assays, which included circular dichroism and mass spectrometry at different K<sup>+</sup> concentrations.

*Results:* Three *hits* were here identified and characterized by biophysical assays. Compound **7** acts as dual Tel<sub>2</sub>/TERRA<sub>2</sub> G4-ligand at physiological KCl concentration, while *hits* **15** and **17** show preferential thermal stabilization for Tel<sub>2</sub> DNA. The different molecular recognition against the two targets was also discussed.

*Conclusions:* Our successful results pave the way to further lead optimization to achieve both increased selectivity and stabilizing effect against TERRA and Tel DNA G4s.

*General Significance:* The current study combines for the first time molecular modelling and biophysical assays applied to bimolecular DNA and RNA G4s, leading to the identification of innovative ligand chemical scaffolds with a promising anticancer profile.

**Keywords:** G-quadruplex DNA | TERRA | Virtual Screening | Circular Dichroism | Mass Spectrometry

**Abbreviations:** G4s, G-quadruplex structures; Tel, telomeric DNA G-quadruplex; Tel<sub>2</sub>, bimolecular telomeric DNA G-quadruplex; TERRA<sub>2</sub>, bimolecular telomeric repeat-containing RNA G-quadruplex; PDS, pyridostatin; cPDS, carboxypyridostatin; AR, aromatic ring feature; HY, hydrophobic feature; PI, positive ionisable feature; HBD, H-bond donor feature; HBA, H-bond acceptor feature; CD, circular dichroism; BA, binding affinity.

## 1. INTRODUCTION

Beyond the canonical double helix [1], both DNA and RNA nucleic acids can also adopt non-canonical higher-order structures called G-quadruplexes (G4s). G4s are formed in single-stranded guanine-rich (G-rich) sequences. In these G-rich sequences, four guanines can self-associate into a square-planar platform, called G-quartet, where each guanine is linked through Hoogsteen-type hydrogen bond network and the oxygen O6 atoms are arranged to coordinate metal cations. G-quartets can  $\pi$ -stack upon each other forming four-stranded helical structure known as G4s [2]. In these last years, several studies reported that both DNA and RNA G4s are associated with many important biological processes [3, 4]. Specifically, DNA G4s have been shown to form into biologically significant regions of the genome, such as human telomeres (composed of TTAGGG repeats) [5], oncogene promoter regions [6], immunoglobulin switch regions [7] and ribosomal DNA [8]. By contrast, RNA sequences forming G4s motifs have been retrieved from the human transcriptome, particularly at the 5' untranslated regions (UTRs) of mRNAs, which have been shown to be involved in the post-transcriptional regulation of gene expression [9]. In addition to the 5'-UTRs regions, a recent study has shown that also mammalian telomeres can be transcribed into telomeric repeat-containing RNA (referred to as TERRA) containing UUAGGG repeats forming G4s conformations of heterogeneous length [10]. The direct evidence of the presence of the parallel-stranded TERRA RNA G4s in living cells has been also demonstrated by using a light-switching pyrene probe by Xu and co-workers [11]. Such an evidence provided a useful information to deeper understand the structure and the biological function of TERRA RNA G4. In fact, it has been hypothesized that the presence of TERRA at the chromosome ends may represent a new level of regulation and protection of telomere, suggesting a possible association between TERRA RNA and telomere DNA (Tel), with a possible clinical relevance in the treatment of cancer. All such findings can trace the route for a new anticancer strategy based on ligands that selectively (or not) bind to the RNA G4s. The availability of telomeric RNA G4 experimental structures (NMR and X-Ray) and a deep understanding of their topologies are crucial to design RNA G4s selective ligands. In fact, a little structural diversity has been found, in telomeric RNA G4 with respect to its DNA G4 counterpart. CD, NMR, and X-ray crystallography have shown that G-rich telomeric RNA sequences form bimolecular parallel G4s (TERRA<sub>2</sub>) both under Na<sup>+</sup> and K<sup>+</sup> conditions [12-14]. Recently, Neidle *et al.* have solved the bimolecular crystal structure of a telomere RNA G4 (PDB code: 3MIJ) complexed with two molecules of a triazole-acridine derivative stacked side-by-side onto a single terminal 5' G-quartet face [15]. In a recent work, Collie *et al.* have reported the higher

selectivity of a naphthalene diimide ligand –bearing hydroxyl groups (compound **c** in Fig. 1) in binding bimolecular RNA G4s with respect to BRACO-19, which, in turn, has higher affinity for Tel DNA G4. Such an observation could be due to the presence of the 2'-OH groups in the ribose sugar at the apices of the RNA G4, which may constrict the space available to the ligand side-chains to interact with the loops by reducing the depth and the width of the UUA loops [16]. Therefore, the presence of multiple hydrogen bonds involving the O2' hydroxyl groups of the ribonucleotide sugars in RNA G4s could be an important structural feature to be taken into account when designing selective RNA vs DNA G4-ligands [17]. The small molecule pyridostatin (PDS) (compound **f** in Fig. 1) was proved to be a generic dual bimolecular Tel DNA and TERRA G4 binder, contrary to its derivative carboxypyridostatin (cPDS) that exhibited high molecular selectivity for unimolecular RNA over DNA G4s [18]. Overall, based on such data, telomere TERRA RNA G4s could represent a potential target for anticancer agents directed against telomeres. Therefore, finding novel dual (as well as selective) G4 Tel DNA and TERRA binding ligands could represent an innovative strategy to enhance telomerase inhibition [19].

Herein, to identify novel dual (or selective) Tel DNA and TERRA G4s binders we have employed a receptor-based virtual screening approach based on molecular docking and pharmacophore models targeting the bimolecular DNA G4 (Tel<sub>2</sub>) of the 12-nt telomeric sequence 5'-TAGGGTTAGGGT-3' (PDB code: 3CE5) [20] and its bimolecular RNA counterpart (TERRA<sub>2</sub>) telomeric sequence 12-nt 5'-UAGGGUUAGGGU-3' (PDB code: 2KBP) [13]. The selected best candidates were submitted to biophysical evaluation. Among the 20 tested candidates, *hit 7* (Supplementary Fig. 4), having a 2-(6,7-dimethoxy-8-oxo-8H-benzo[c]indolo[3,2,1-ij][1,5]naphthyridin-12-yl)oxy scaffold, came out as the most interesting Tel<sub>2</sub>/TERRA<sub>2</sub> dual binder, not only from the theoretical approach but also from biophysical data which confirmed its “dual” activity since it showed good stabilization against both Tel<sub>2</sub> and TERRA<sub>2</sub>. In addition, *hits 15* and *17* (Supplementary Fig. 5) have been retrieved as interesting selective Tel<sub>2</sub> binders and further lead optimization studies could be undertaken in order to improve their affinity against Tel<sub>2</sub> or to achieve a certain activity over TERRA<sub>2</sub>. This work paves the way to understand at atomistic level the molecular features needed to obtain Tel/TERRA G4 binders through a Virtual Screening approach.

## 2. MATERIAL AND METHODS

**2.1 Receptors preparation.** The bimolecular human telomeric  $Tel_2$  and  $TERRA_2$  G4s, having PDB codes 3CE5 [20] and 2KBP [13] respectively, were selected to perform the structure-based virtual screening procedure since their sequence and topology were fully comparable. In particular, the PDB model 3CE5 represents the bimolecular X-ray structure (resolution of 2.5 Å) of the telomeric sequence 5'-TAGGGTTAGGGT-3', complexed with the known active ligand BRACO-19 in the presence of physiological concentrations of  $K^+$ . The binding site of the ligand, BRACO-19, derived from the 3,6,9-trisubstituted acridine, is located at the interface of two G4 monomers both with parallel topology, included between a G-tetrad surface and a tetrad TATA. Indeed, the NMR PDB model 2KBP represents the ribonucleic sequence 5'-UAGGGUUAGGGU-3' in the presence of  $K^+$  ion, equivalent to that of the 3CE5 model. In this study, both the X-Ray and NMR models were prepared and refined using the Protein Preparation Wizard implemented in Maestro ver. 9.7 [21] and, in addition, also the corresponding minimized structures were considered. With this aim, 5000 steps of the Polak Ribiere conjugate gradient algorithm [22] and the OPLS 2005 force field with the united atoms notation [23] as implemented in MacroModel version 10.3 [24] were used. A convergence criterion equal to  $0.05 \text{ kJ}\cdot\text{mol}^{-1}\cdot\text{Å}$  was adopted and solvent effects were taken into account by means of the GB/SA implicit water solvation model [25].

**2.2 Actives and Decoys set preparation.** A test set of 9 known active G4 ligands **a-i** (Fig. 1) characterized by different chemical scaffolds, were collected from literature data [26, 27]. They were selected based on their experimental affinity and activity toward bimolecular  $TERRA_2$  and  $Tel_2$  G4s and their selectivity with respect to duplex DNA.

**Please put Figure 1 here**

The decoy set was taken from the Directory of Universal Decoys (DUD) [28], which provided a library of 419 compounds, based on the structural similarity with respect to the active set. Both actives and decoys set were prepared considering the ionization state at physiological pH of 7.4 using the OPLS 2005 force field as implemented in the LigPrep platform ver. 2.9 of Maestro [29].

**2.3 Molecular Docking protocol.** In order to evaluate the accuracy and reliability of our docking protocol, we firstly used the SP (Standard Precision) scoring function of Glide ver. 6.2 software from Schrodinger suite [30] to perform redocking calculations of BRACO-19 against 3CE5 PDB

model, in order to geometrically reproduce the bioactive conformation. The obtained poses with the related RMSD (Root Mean Square Deviation) values are shown in Figure S1 of the Supplementary Materials (Supplementary Fig. 1). BRACO-19 was treated as flexible, while the receptor was kept rigid. The energy grid was built centering the docking box on guanine core and setting its outer box size to 45×45×45 Å. The scaling factor for the receptor Van der Waals radii was set to 1.0 and the number of generated poses was equal to 10. Following this protocol, docking simulations of the 9 actives and 419 decoys were carried out using the same redocking conditions for both Tel<sub>2</sub> and TERRA<sub>2</sub> G4s receptors.

**2.4 Enrichment factor studies.** In order to take into account the effects of receptors flexibility, an ensemble of receptor conformations has been created (Supplementary Table 1). In particular, in the case of the X-Ray model of Tel<sub>2</sub> (PDB ID: 3CE5), both the crystallographic and the minimized structures were considered for each of these different conditions: *i*) in presence of the complexed ligand BRACO-19; *ii*) in absence of the complexed ligand BRACO-19; *iii*) with waters; *iv*) without waters. Indeed, for TERRA<sub>2</sub>, since its NMR model (PDB ID: 2KBP) contains 10 conformers, we decided to consider both experimental and minimized structures. Globally, 26 ensemble structures have been used for enrichment calculations (Supplementary Table 1), in order to select the ensembles of up to 2 receptors, which lead to improved enrichment results of the actives set [31]. The following descriptors have been taken into account: area under the curve (AUC), receiver operating characteristic (ROC), enrichment factor (EF), robust initial enhancement (RIE) and Boltzmann-enhanced discrimination of ROC (BEDROC). Docking calculations were performed using the Glide SP software. Among all the ensemble receptors, only the up to 2 receptors with the best AUC and RIE data were chosen (Figs. 2A and 2B), as able to discriminate the largest number of active compounds and to rank them within the top-score *hits*. Their AUC and RIE values are highlighted in bold in the supplementary Table 1, whereas their % Active compound screen plots are shown in the supplementary Fig. 2.

**Please put Figure 2 here**

**2.5 Structure-based pharmacophore generation.** The 3D Structure-based pharmacophores were developed by means of the software LigandScout ver. 3.1 [32] on the best ranked docked pose of the 9 actives obtained against the previously validated ensemble receptor (Figs. 2A and 2B). From the best docking complexes, the 3D pharmacophore models were automatically generated by LigandScout using predefined features, such as hydrogen bond acceptor (HBA), hydrogen bond

donor (HBD), hydrophobic (HY), negative ionizable (NI), positive ionizable (PI), aromatic ring (AR). Thus, the obtained pharmacophores were further modified increasing or decreasing the radius of tolerance of the previous features, inserting the excluded volumes in order to better describe the shape of the binding cavity and boost the complexity of the pharmacophore [33]. In particular, we manually modified the starting automatic pharmacophores on the basis of the information regarding DNA or RNA G4/ligand interactions reported in literature data and they are described in *Results* paragraph.

**2.6 Structure-based pharmacophore validation.** This procedure was performed using the same active/decoys set already used in the enrichment study. The hypotheses were evaluated on their EF and ROC curve. We used the Enrichment Factor (EF) as validation descriptor [34-35], that was calculated, for all the generated hypotheses, using the following eq. A.1:

$$EF = \frac{TP/n}{A/N} \quad (\text{Eq. A.1})$$

where TP is the number of True Positive, n is the number of active and inactive compounds found from the pharmacophore hypothesis, A is the number of active compounds in the database and N is the database size (total of actives and inactives).

After evaluating the hypotheses based on the EF, the best ones were selected taking into account also the Receiver operating characteristic (ROC) curve and the percentage of the actives found for each hypothesis (Tables 2-3 of the *Results* section). In particular, an ideal ROC curve, where no overlap between the scores of active molecules and decoys exists, proceeds from the origin to the upper-left corner until all the actives are retrieved and the sensitivity (*Se*) reaches the value of 1. Further details are reported in the *Results* paragraph. The validated pharmacophore models are reported in Figs. 3-4.

**2.7 Database preparation and Virtual Screening (VS).** Each of the ten selected and validated pharmacophore models (Figs. 3-4) was used to perform VS separately using a multi-conformational database of natural compounds imported from the ZINC website. It included also the NPD 08 database developed by the University of Innsbruck, for a global number of ~ 257000 natural compounds. A physically relevant ionization state (pH=7.4) was assigned to all compounds and the stereochemistry of chiral centres was determined. Afterwards, the so prepared database was subjected to an energy minimization step using the same protocol adopted for actives and decoys set

[13]. Hence, two databases were created: one including the merged *hits* identified for each pharmacophore model of Tel<sub>2</sub>, the second formed by the merged *hits* found for TERRA<sub>2</sub> RNA G4. The resulting two databases for TERRA<sub>2</sub> and Tel<sub>2</sub> were de-duplicated and further reduced in size by applying the popular empirical Lipinski's rule of five. Molecules that successfully passed this filter were further submitted to docking studies.

**2.8 Similarity clusterization.** To reduce the number of compounds to be tested on Tel<sub>2</sub>, we clustered the obtained *hits* on the basis of their volume overlap, using the average Linkage method. Finally, for each cluster we selected the compound with the best Glide Score (GScore).

**2.9 Selection for biological testing.** After the careful visual inspection analysis of the best docking poses and the evaluation of their commercial availability, the best 20 final *hits* were purchased and submitted to biological tests.

**2.10 Oligonucleotides used in the biophysical analysis.** All desalted oligonucleotides were purchased from Sigma-Aldrich, Milan, Italy (Table 1).

**Please put Table 1 here**

**2.11 Circular dichroism analysis.** CD spectra were recorded on a Chirascan-Plus (Applied Photophysics, Leatherhead, UK) equipped with a Peltier temperature controller using a quartz cell of 5-mm optical path length and scanning a speed of 50 nm/min with a response time of 4 sec over a wavelength range of 230-320 nm. The reported spectrum of each sample represents the average of 2 scans. Observed ellipticities were converted to mean residue ellipticity ( $\theta$ ) = deg x cm<sup>2</sup> x dmol<sup>-1</sup> (Molar Ellipticity). Oligonucleotides were diluted from stock to the final concentration (4-8  $\mu$ M single strand concentration) in Li cacodylate buffer (10 mM, pH 7.4) with 100 mM (or 5 mM or 2 mM) KCl and then annealed by heating at 95 °C for 5 min, gradually cooled to room temperature. Compounds were added at 2x G4 final concentration (8-16  $\mu$ M). CD spectra were recorded after 24 h from 20°C to 95°C, with temperature increase of 5°C. T<sub>m</sub> values were calculated according to the van't Hoff equation, applied for a two state transition from a folded to unfolded state, assuming that the heat capacity of the folded and unfolded states are equal [36].

**2.12 Mass spectrometry analysis.** Oligonucleotides (20  $\mu\text{M}$ ) in 1 mM KCl, 150 mM trimethylammonium acetate pH 7.0 (TMAA) were annealed by heating at 95  $^{\circ}\text{C}$  for 5 min and gradually cooling to room temperature. Afterward isopropanol (IPA) was added to the oligonucleotides (20% of the total volume) before incubating overnight at 4  $^{\circ}\text{C}$  (final buffer composition: 0.8 mM KCl, 120 mM TMAA pH 7.0, 20 % IPA). The folded oligonucleotides were mixed together (Tel<sub>2</sub>/TERRA<sub>2</sub>, Tel<sub>2</sub>/dsDNA, TERRA<sub>2</sub>/dsDNA, each oligo final concentration 2  $\mu\text{M}$ ) and incubated with the tested compound at ratio oligo: compound 1:2 for 1h at 4  $^{\circ}\text{C}$ . Samples were analyzed by direct infusion electrospray ionization (ESI) on a Xevo G2-XS QToF mass spectrometer (Waters, Manchester, UK). The injection was automatically performed by an Agilent 1290 Infinity HPLC (Agilent Technologies, Santa Clara, CA, US) equipped with an auto sampler; the carrying buffer was TMAA 80 mM pH 7.0, 20% IPA. 5  $\mu\text{L}$  of each sample were typically injected for each analysis. The ESI source settings were the following: electrospray capillary voltage set at 1.8 kV, the source and the desolvation temperatures were 45  $^{\circ}\text{C}$  and 65  $^{\circ}\text{C}$  respectively, the sampling cone was set at 75 V. All these parameters ensured minimal fragmentation of the DNAs complexes. The instrument was calibrated using a 2 mg/mL solution of sodium iodide in 50% of IPA. Additionally, the use of the LockSpray during the analysis provided a typical <5 ppm mass accuracy. The internal standard LockSpray consisted in a solution of leu-enkephalin 1  $\mu\text{g}/\text{mL}$  in acetonitrile/water (50:50, v/v) containing 0.1% of formic acid. Binding affinities were calculated for each experiment using the intensity of each species calculated by MassLynx V4.1. The binding affinity was calculated with the following formula:  $[\text{BA} = (\Sigma\text{NAb}/(\Sigma\text{NAf} + \Sigma\text{NAb})) \times 100]$ , where BA is the binding affinity, NAb is the intensity of the bound nucleic acid peaks, NAf is the intensity of the free nucleic acid peaks. NAb comprises DNA/RNA with one or more bound ligands, where present. Signals with charge states, 5<sup>-</sup>, and 4<sup>-</sup> were used. Only dimeric DNA/RNA coordinating two potassium ions or duplex DNA were included in the calculation.

### 3. RESULTS

**3.1 Generation of Tel<sub>2</sub> Structure-based pharmacophore models.** In this study, five different docking complexes for bimolecular DNA G4 (Tel<sub>2</sub>) were selected as input for structure-based pharmacophore generation. Similarly, to the crystallographic pose of BRACO-19, we observed a preference for the bottom position in the recognition of Tel<sub>2</sub> in the complexes of four hypothesis (*D-Hypo1*, *D-Hypo2*, *D-Hypo3*, *D-Hypo4*), in which the planar portion of ligands was sandwiched

between residues DG11 and DT12. In contrast, in the *D-Hypo5* the binding site was identified in the loop of Chain A. Specifically, *D-Hypo1* was generated starting from the docking complex of Naphtalene diimide with a bis-dimethylaminopropyl-hydroxypropylamino side chain (compound **c** in Fig. 1), involved in several hydrogen bonds interactions established with the phosphodiesterasic portion of residues DG11 and DG16 and the carboxyl groups of DT24. The carboxyl group of the aromatic core was found to interact by means of a hydrogen bond with DT24, too. In *D-Hypo2*, Tel<sub>2</sub> was recognized by BRACO-19 (**e**) and it was stabilized by two hydrogen bonds formed between the phosphors of residues DT12 and DG17 with the two pyrrolidine rings of the ligand. In *D-Hypo3*, the 9-Fluoro-2-methyl-5H-thiocromeno[4,3-d] pyrimidine (compound **g** in Fig. 1) interacted with Tel<sub>2</sub> only by means of stacking interactions. *D-Hypo4* described the way in which 2-chloro-6-fluoro-N-(4-fluoro-1-methyl-1H-indazol-3-yl)benzamide (compound **h** in Fig. 1) interacted with Tel<sub>2</sub>, with the establishment of two hydrogen bonds between the ligand amide group and residues DG11 and DT12. In the last *D-Hypo5*, 2-(2-Fluorophenyl)-N-[(5-methyl-2-thienyl)methyl]ethanamine (compound **i** in Fig. 1) was stabilized in the complex through one hydrogen bond between the ligand amine group and the oxygen of the DG5 furanose.

We decided to select all the generated five pharmacophore hypothesis, that are represented in Fig. 3, due to the good EF values for *D-Hypo2*, *D-Hypo3*, *D-Hypo4*, *D-Hypo5* and the high percentage of active compounds and the ROC curve (Supplementary Fig. 3A) for *D-Hypo1*. Details about the features are reported in Table 2 and discussed in the *Discussion* section.

**Please put Figure 3 here**

**Please put Table 2 here**

**3.2 Generation of TERRA<sub>2</sub> Structure-based pharmacophore models.** As concerns the bimolecular telomeric RNA G4 (TERRA<sub>2</sub>), eight pharmacophore hypotheses were generated; their feature details are reported in Table 3 and more thoroughly discussed in *Discussion* section. Also in the TERRA<sub>2</sub>, the bottom position was the preferred binding site, probably because at the top position residues RU1 and RU13 interact with each other, thus limiting the exposure of the planar G-tetrad formed by RG3:RG9:RG15:RG21 residues. Consequently, more favourable interactions between the planar moieties of the ligands are established with RG5, RG11 and RG23 residues. Specifically, the *T-Hypo1* was generated from the docking complex of the Naphtalene diimide with a dimethylamino-propyl side chain (compound **a** in Fig. 1) whose positively charged side chains

engage electrostatic interactions with the phosphodiesterase portion of TERRA<sub>2</sub>. Similar binding modes were observed in the docking complexes of the Naphtalene diimide having the propylamino-pyrrolidin side chains (compound **b** in Fig. 1) and the Naphtalene diimide with the bis-dimethylamino-propyl-hydroxypropyl side chains (compound **c** in Fig. 3), represented, respectively, in *T-Hypo2* and *T-Hypo3*. The only difference of compound **b** was related to the establishment of one hydrogen bond between a pyrrolidine nitrogen with a carboxyl group of residue RU24. In contrast, for compound **c**, the two hydroxyl groups interact by means of two hydrogen bonds with residues RU6 and RU12. *T-Hypo4* describes the complex of Acridine (compound **d** in Fig. 1) with TERRA<sub>2</sub>, in which two electrostatic interactions, between the positively charged nitrogen atom of the ligand side chain and the phosphodiesterase portion of RA8 and RU24 residues, further stabilized the binding mode. In *T-Hypo5* (Supplementary Fig. 4), TERRA<sub>2</sub> was recognized by PDS (Pyridostatin) (compound **f** in Fig. 1), and it was stabilized by three electrostatic interactions among the positively charged nitrogen atom of the moieties and the phosphodiesterase groups and by one hydrogen bond established between a ligand carboxyl group and RU24. The 9-Fluoro-2-methyl-5H-thiocromeno[4,3-d] pyrimidine (compound **g** in Fig. 1), whose binding mode was described in *T-Hypo6*, established two hydrogen bonds with residues RU12 and RG17, contrary to its binding mode on the Tel<sub>2</sub> where it interacted only through stacking interactions. The complex of 2-chloro-6-fluoro-N-(4-fluoro-1-methyl-1H-indazol-3-yl)benzamide (compound **h** in Fig. 1) with TERRA<sub>2</sub> was represented by *T-Hypo7* (Supplementary Fig. 4) and was characterized only by stacking interactions. Finally, *T-Hypo8* (Supplementary Fig. 4) described the binding mode of 2-(2-Fluorophenyl)-N-[(5-methyl-2-thienyl)methyl]ethanamine (compound **i** in Fig. 1) on TERRA<sub>2</sub>, characterized by several stacking interactions and one hydrogen bond with residue RU24.

We decided to select only five pharmacophore hypotheses (Fig. 4). In particular, *T-Hypo1*, *T-Hypo2*, *T-Hypo3* and *T-Hypo4* were selected due to the good EF values, while *T-Hypo6* was chosen taking into account its ROC curve (Supplementary Fig. 3B) and the high percentage of active compounds found.

**Please put Figure 4 here**

**Please put Table 3 here**

**3.3 Database screening.** From the above validation methods, it turned out that five pharmacophore hypotheses for Tel<sub>2</sub> and five for TERRA<sub>2</sub> had a good ability to distinguish the active from decoy sets and to retrieve the greatest number of active compounds. These best pharmacophores were used

separately to screen a database of around ~260000 compounds. The ranked molecules identified for each Tel<sub>2</sub> pharmacophore model were merged, thus obtaining 18725 *hits* for the deoxyribonucleic G4 and the same procedure was performed for *hits* resulting from virtual screening of TERRA<sub>2</sub>, collecting 16578 molecules. Then, Lipinski's rule of five was used to eliminate the *non-drug like* compounds from the *hit* molecules thus obtaining 12212 and 12387 molecules respectively for Tel<sub>2</sub> and TERRA<sub>2</sub> that were further subjected to ensemble docking simulations on their theoretical target.

**3.4 Molecular docking studies.** Molecular docking was performed on Tel<sub>2</sub> and TERRA<sub>2</sub> to allow the investigation of the potential binding orientations of the databases *hit* compounds and select them by their binding affinity score. With this aim, in order to have a reasonable *cut-off*, we used the average GScore obtained from the active compounds against Tel<sub>2</sub> and TERRA<sub>2</sub>, respectively of -9.64 and -8.98 kcal/mol. Thus, the top-ranked 91 *hits* for Tel<sub>2</sub> and 14 *hits* for TERRA<sub>2</sub> were obtained. Afterwards, given the high number of top-ranked *hits*, they were clustered considering their molecular diversity (*see* the “*Similarity clusterization*” paragraph) and finally 32 *hits* were selected for their scaffold diversity. Unfortunately, after investigating their commercial availability, only 20 *hits* were purchased from several chemical suppliers (seven *hits* from TERRA<sub>2</sub> and thirteen from Tel<sub>2</sub>) and submitted for biophysical assays. The 2D structures of the 20 *hits* are shown in supplementary Figs. 5 and 6, while their Glide SP Score (GScore) values are reported in the supplementary Table 2.

**3.5 Circular dichroism screening.** The available *hits* were screened for their ability to stabilize the bimolecular G4 targets by melting circular dichroism spectroscopy (CD). This technique allowed us to discriminate quadruplex topologies and to calculate melting temperatures (T<sub>m</sub>) both in the absence and presence of the compounds.

*Hits* **1-7** and **8-20**, selected by the computational screening as putative binders of TERRA<sub>2</sub> and Tel<sub>2</sub>, respectively, were initially tested against their computationally defined targets (Tables 4-5). *Hit* **7** was the only candidate able to stabilize TERRA<sub>2</sub>, with a  $\Delta T_m$  of 6.8 °C (Table 4), while *hits* **15** and **17** were the best stabilizers of Tel<sub>2</sub>.

**Please put Table 4 here**

The three best active *hits* did not modify the dimeric propeller-like parallel G4 conformation of TERRA<sub>2</sub> [13-14], the CD spectrum of which was characterized by a negative peak at 240 nm and a

positive peak at 265 nm also in the presence of the compounds (Figs. 5 A-B and supplementary Fig. 7). In contrast, they modified the mixed-type conformation of Tel<sub>2</sub>. In particular, the CD spectrum of the oligonucleotide itself showed a negative peak at 240 nm, a positive peak at 265 nm and a shoulder at 295 nm (Fig. 5 C), as the result of two interconverting conformations [37-38]. In the presence of the compounds the equilibrium was shifted towards the parallel-like topology, resulting in the decrease of the shoulder at 295 nm and increase of the dichroic signal at 265 nm, especially for *hits* **15** and **17** (Figs. 5 D-F).

**Please put Figure 5 here**

Because TERRA<sub>2</sub> was much more stable than Tel<sub>2</sub> at 100 mM (T<sub>m</sub> 71.0°C and 52.6, respectively) (Table 5), we reasoned that the stabilization imparted by the compounds would have been naturally more pronounced in the intrinsically less stable oligonucleotide, i.e. Tel<sub>2</sub>. To test this hypothesis, we decreased K<sup>+</sup> concentration to 5 and 2 mM to obtain T<sub>m</sub> values for TERRA<sub>2</sub> similar to those of Tel<sub>2</sub>. TERRA<sub>2</sub> T<sub>m</sub> at 2 mM K<sup>+</sup> (54.3°C) became comparable to Tel<sub>2</sub> T<sub>m</sub> at 100 mM K<sup>+</sup> (52.6°C). Indeed, at lower K<sup>+</sup> concentration, all *hits* were able to stabilize TERRA<sub>2</sub> to a greater extent (Table 5) and at 2 mM K<sup>+</sup> *hits* **15** and **17** stabilized Tel<sub>2</sub> and TERRA<sub>2</sub> by the same ΔT<sub>m</sub>. The one exception was *hit* **7** on TERRA<sub>2</sub> at 2 mM K<sup>+</sup>, conditions at which destabilization rather than stabilization of the G4 structure was obtained. This last data may be an indication that TERRA<sub>2</sub> G4 folding in such low K<sup>+</sup> concentration is so poor that stabilization is detected only by compounds able to induce the G4 conformation. These data indicate that a meaningful direct comparison of the binding efficiency towards oligonucleotides with different T<sub>m</sub> is hardly achievable by CD.

**Please put Table 5 here**

**3.6 Mass Spectrometry (MS) Competition Assay.** To detect which of the two G4 oligonucleotides (TERRA<sub>2</sub> or Tel<sub>2</sub>) was preferentially bound by the selected hits, an electrospray mass spectrometry method was set up. The gentle ionization conditions of electrospray allow to transfer intact biomolecules and non-covalent complexes in the gas phase to obtain a snap-shot of the species present in solution and their relative concentration [39-40]. By the relative intensity of the bound/unbound species, it is possible to gather information on the relative binding affinities and the specificity toward the different investigated oligonucleotides [41-42]. In our experimental procedure, TERRA<sub>2</sub> and Tel<sub>2</sub> were separately folded and mixed just before the addition of the tested

compounds to minimize formation of hybrid DNA/RNA G4s. Surprisingly, the formation of the hybrid species was not observed, even after extended incubation times (i.e. up to 16 h) (Supplementary Fig. 8). Both TERRA<sub>2</sub> and Tel<sub>2</sub> were detected mostly as dimers, as evinced by the coordination of two K<sup>+</sup> ions, which is consistent with the formation of dimeric G4s with three guanine quartets. The monomeric peak was almost absent in TERRA<sub>2</sub> solution, while was detectable, even though at low level, in the Tel<sub>2</sub> sample, confirming the lower stability of the Tel<sub>2</sub> G4. Double-stranded DNA (dsDNA) was also used alongside the G4 folded oligonucleotide to test the *hits*' selectivity towards G4s. DsDNA did not modify G4 formation (Supplementary Fig. 9).

Each compound was competed for the binding to TERRA<sub>2</sub>/Tel<sub>2</sub>, TERRA<sub>2</sub>/dsDNA, Tel<sub>2</sub>/dsDNA (Table 6) (Fig. 6 and supplementary Figs. 10-12). The relative binding affinities (BA) calculated by the intensity of the peaks are reported in Table 6. Interestingly, no adducts to monomeric RNA or DNA were obtained (Fig. 6 and supplementary Figs. 10-12).

*Hit 7* showed a much higher affinity for TERRA<sub>2</sub> compared to Tel<sub>2</sub> and dsDNA (Table 6, Fig. 6 and supplementary Fig. 10); *hit 15* displayed a slightly higher affinity for Tel<sub>2</sub> vs TERRA<sub>2</sub> and low affinity for dsDNA (Supplementary Fig. 11); *hit 17* showed similar BA towards TERRA<sub>2</sub> and Tel<sub>2</sub> and low binding affinity towards dsDNA (Supplementary Fig. 12). The affinity towards dsDNA was similarly low for all compounds (Table 6).

**Please put Table 6 here**

**Please put Figure 6 here**

**3.7 Pharmacophore analysis between Tel<sub>2</sub> vs TERRA<sub>2</sub>.** The interesting results obtained from experimental data against Tel<sub>2</sub> and TERRA<sub>2</sub> for *hits 7, 15* and *17* led us to analyze in detail the differences between the pharmacophore models of TERRA<sub>2</sub> (*T-Hypo*) and Tel<sub>2</sub> (*D-Hypo*). The *D-Hypo2* and the *T-Hypo4*, respectively for Tel<sub>2</sub> and TERRA<sub>2</sub>, were equally able to discover the *hit 7* (Fig. 7 A-B). The two pharmacophore models consist in the same type of features (2 HBD, 1 AR) with very similar interatomic distances (Fig. 7 C-D), thus justifying the dual activity of *7*. Only the greater distance between the two HBD groups is mainly due to the flexibility of the side chain. The positive ionisable feature (PI) does not seem to be discriminant between the two targets.

**Please put Figure 7 here**

*Hit 15* was found only by the *D-Hypo1* consisting of one AR, one HBD and one HBA (Supplementary Fig. 13 A-B). Also in this case the selectivity of this compound towards Tel<sub>2</sub> can be explained by looking at the interatomic distance of the pharmacophore features that are much less spaced compared to those of *T-Hypo4* and *D-Hypo2* (Fig. 7 A-B) and the presence of one HBA.

*Hit 17* was retrieved only by *D-Hypo3* (whose superimposition is shown in supplementary Fig. 14 A-B). The lack of selectivity toward TERRA<sub>2</sub> can be attributed to the absence of any HBD interactions as well as to the different spatial arrangement of the features that are very close to each other.

**3. 8 Docking pose analysis of best hits 7, 15 and 17.** The Glide SP Score (GScore) of *hits 15* and *17* obtained against TERRA<sub>2</sub> fall outside the threshold used to filter docking results (equal to -7.39 and -7.22 kcal/mol, respectively), thus confirming their lower affinity for this target demonstrated also by  $\Delta T_m$  data at 100 mM K<sup>+</sup>. With the aim of clarifying the different affinity of *15* and *17* between TERRA<sub>2</sub> and Tel<sub>2</sub>, we analyzed the relative docking poses and the contacts (Table 7) by means of the Maestro graphical interface (Maestro Graphics User Interface, Schrödinger LLC, New York, NY, USA, 2016) [38, 39].

**Please put Table 7 here**

As previously stated, all the three *hits 7, 15* and *17* favourably interact in the same binding site for both Tel<sub>2</sub> and TERRA<sub>2</sub> located close to the guanine residues G5:G11:G17:G23 which, for Tel<sub>2</sub>, corresponds to the same binding site of BRACO-19 reported in the crystal structure (PDB ID: 3CE5). In addition, *hits 7, 15* and *17* are characterized by the presence of side chains with positively charged nitrogen atoms, able to engage in favorable salt-bridge and Hbond interactions with the phosphate backbone. However, the higher affinity to Tel<sub>2</sub> G4 of *15* and *7* probably results from a sandwiched binding mode between their aromatic core and the thymine nucleobase at positions 12, thus justifying the increasing number of their good and stacking interactions against Tel<sub>2</sub> (Table 7). The only exception is *17*, which does not form any stacking interaction against Tel<sub>2</sub>, but is able to establish the highest number of good contacts with respect to its binding mode against TERRA<sub>2</sub>. In particular, residues RU12 and RU24 in TERRA<sub>2</sub> G4 assume a conformation that does not allow the good accommodation of the ligand aromatic planar portion on the tetrad, since the lack of the

methyl group in the uracil and the resulting reduced steric hindrance of the hydrogen atom make these residues more flexible (Fig. 8 A-F).

**Please put Figure 8 here**

#### 4. DISCUSSION

As widely reported in literature, telomerase activity requires an unfolded single-stranded substrate. Hence, G4 formation directly inhibits telomerase elongation *in vitro* [43] and therefore ligands that selectively bind to and stabilize G4 structures may interfere with telomere conformation and telomere elongation. Since telomerase levels are high in the majority of malignant cells and largely absent in normal cells, these G4 ligands have been proposed as selective anti-cancer agents.

Given the few number of available telomeric G4 RNA selective binders to date [26], our Virtual Screening approach can be considered the first step towards the discovery of novel promising ligands able to stabilize TERRA<sub>2</sub> structures. In particular, our protocol allowed us to identify *hit 7*, characterized by a naphthyridin scaffold with a ((dimethylamino)propyl)acetamide side chain, as a dual Tel<sub>2</sub>/TERRA<sub>2</sub> binder: *hit 7* matches Tel<sub>2</sub> and TERRA<sub>2</sub> pharmacophore features that display very similar interatomic distances and angles, and it exerts good theoretical affinity for both structures, as gathered from docking calculations. *In vitro*, *hit 7* displays thermal stabilization effects on both target nucleic acids even though with a visible preferential stabilization on TERRA<sub>2</sub> as gathered by both CD and MS competition analysis. On the contrary, *Hits 15* and *17*, having respectively a furo-chromen and benzofuran scaffolds, show preferential tendency to stabilize Tel<sub>2</sub> G4 rather than its RNA counterpart. Such a finding probably relates to a reduced  $\pi$ - $\pi$  interactions network of both *hits* due to lower accessibility of the G-tetrad surface linked to the steric hindrance of the terminal uracils RU12 and RU24. Moreover, a poor geometric overlapping between the pharmacophore features of TERRA<sub>2</sub> with both *hits 15* and *17* was observed, thus rationalizing their preferential recognition of Tel<sub>2</sub>. Compound *17* behaviour confirmed also our previous findings on the potential of the furo-chromen scaffold [44]. Optimization on this scaffold is already in progress.

#### 5. CONCLUSIONS

The molecular modelling and biophysical studies reported here allowed us to retrieve three novel promising stabilizers against Tel<sub>2</sub>/TERRA<sub>2</sub> G4s. Specifically, the naphthyridin scaffold is here proposed as the most interesting one against TERRA<sub>2</sub> and could be the starting point for a further lead optimization to enhance its telomeric RNA selectivity. On the other hand, the two derivatives with a furo-chromen and benzofuran scaffolds preferentially recognized Tel<sub>2</sub>: their affinity profile could thus be investigated also against other G4s topologies. *Hit* optimization could lead, in the future, to novel G4-ligand cores with potentially innovative conformational/sequence specificity.

## 6. ACKNOWLEDGMENTS

This work was supported by the Italian Ministry of Education (Funding for Investments of Base Research), [FIRB-IDEAS RBID082ATK]; R.R. was supported by a “HEMMAS” Fellowship [POR Calabria FSE 2007–2013]; G.C. was supported by the UMG grant dedicated to the memory of Anna and Nin Barbieri. The authors also thank the COST action (Multi-target paradigm for innovative ligand identification in the drug discovery process MuTaLig) [CA15135] for the support. The authors would like to thank Dr. Daniela Schuster from the Computer Aided Molecular Design laboratory in Leopold-Franzens-University of Innsbruck for having provided the NPD 08 database.

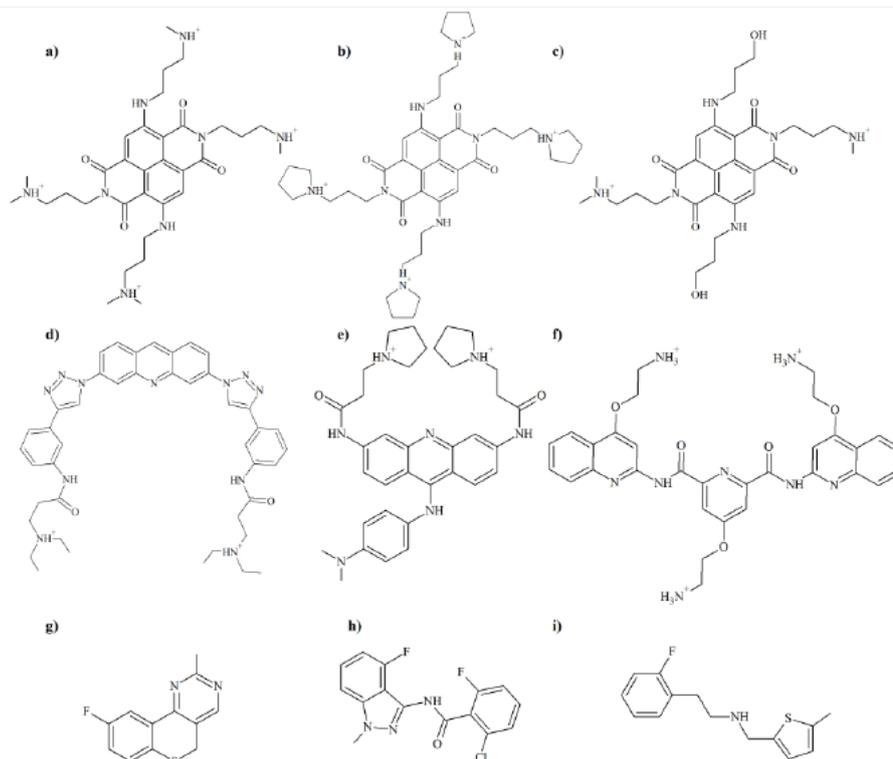
## 7. REFERENCES

- [1] J.D. Watson and F.H.C. Crick, A Structure for Deoxyribose Nucleic Acid, *Nature* 171 (1953) 737-738.
- [2] L. Parrotta, F. Ortuso, F. Moraca, R. Rocca, G. Costa, S. Alcaro, A. Artese, Targeting unimolecular G-quadruplex nucleic acids: a new paradigm for the drug discovery?, *Expert Opin. Drug Discov.* 9 (2014) 1167-87.
- [3] G. Biffi, M. Di Antonio, D. Tannahill, S. Balasubramanian, Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells, *Nat. Chem.*, 6 (2014) 75–80.
- [4] G. Biffi, D. Tannahill, J. McCafferty, S. Balasubramanian, Quantitative visualization of DNA G-quadruplex structures in human cells, *Nat. Chem.*, 5 (2013) 182–186.
- [5] E. Henderson, C.C. Hardin, S.K. Walk, et al. Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine-guanine base pairs, *Cell*, 51 (1987) 899-908.

- [6] Y. Qin, L.H. Hurley, Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions, *Biochimie*, 90 (2008) 1149-71.
- [7] N. Maizels, Dynamic roles for G4 DNA in the biology of eukaryotic cells, *Nat. Struct. Mol. Biol.* 13 (2006) 1055-9.
- [8] D. Drygin, A. Siddiqui-Jain, S. O'Brien, M. Schwaebe, A. Lin, J. Bliesath, C.B. Ho, C. Proffitt, K. Trent, J.P. Whitten, J.K. Lim, D. Von Hoff, K. Anderes, W.G. Rice, Anticancer activity of CX-3543: a direct inhibitor of rRNA biogenesis. *Cancer Res.* 69 (2009) 7653-61.
- [9] A. Bugaut, S. Balasubramanian, 5'-UTR RNA G-quadruplexes: translation regulation and targeting, *Nucleic Acids Res.* 40, (2012) 4727-4741.
- [10] C.M. Azzalin, P. Reichenbach, L. Khoriantseva, E. Giulotto, J. Lingner, Telomeric repeat-containing RNA and RNA surveillance factors at mammalian chromosome ends, *Science*, 318 (2007) 798-801.
- [11] Y. Xu, Y. Suzuki, K. Ito, M. Komiyama, Telomeric repeat-containing RNA structure in living cells, *Proc. Natl. Acad. Sci. U S A.* 107 (2010) 14579-84.
- [12] Y. Xu, K. Kaminaga, M. Komiyama, G-Quadruplex Formation by Human Telomeric Repeats-Containing RNA in Na<sup>+</sup> Solution, *J. Am. Chem. Soc.* 130 (2008) 11179-84.
- [13] H. Martadinata, A.T. Phan, Structure of Propeller-Type Parallel-Stranded RNA G-Quadruplexes, Formed by Human Telomeric RNA Sequences in K<sup>+</sup> Solution. *J. Am. Chem. Soc.* 131 (2009) 2570-8.
- [14] G.W. Collie, S.M. Haider, S. Neidle, G.N. Parkinson, A crystallographic and modelling study of a human telomeric RNA (TERRA) quadruplex. *Nucleic Acids Res.* 38 (2010) 5569-80.
- [15] G.W. Collie, S. Sparapani, G.N. Parkinson, S. Neidle, Structural basis of telomeric RNA quadruplex-acridine ligand recognition. *J. Am. Chem. Soc.* 133 (2011) 2721-8.
- [16] G. Collie, A.P. Reszka, S.M. Haider, V. Gabelica, G.N. Parkinson, S. Neidle, Selectivity in small molecule binding to human telomeric RNA and DNA quadruplexes. *Chem. Commun.* 48 (2009) 7482-4.
- [17] Y. Xu, M. Komiyama, Structure, function and targeting of human telomere RNA, *Methods*, 57 (2012) 100-105.
- [18] M. Di Antonio, G. Biffi, A. Mariani, E.A. Raiber, R. Rodriguez, S. Balasubramanian, Selective RNA versus DNA G-quadruplex targeting by in situ click chemistry. *Angew. Chem. Int. Ed. Engl.* 51 (2012) 11073-8.
- [19] Y. Xu, Human Telomere RNA: A Potential Target for Ligand Recognition, *Curr. Pharm. Des.* 18 (2012) 2096-2101.
- [20] N.H. Campbell, G.N. Parkinson, A.P. Reszka, S. Neidle, Structural basis of DNA quadruplex recognition by an acridine drug. *J. Am. Chem. Soc.* 130 (2008) 6722-6724.
- [21] Schrödinger Suite 2014-1 Protein Preparation Wizard; Epik version 2.7, Schrödinger, LLC, New York, NY, 2013; Impact version 6.2, Schrödinger, LLC, New York, NY, 2014; Prime version 3.5, Schrödinger, LLC, New York, NY, 2014.

- [22] E. Polak, G. Ribiere, Note sur la convergence de methodes de directions conjuguées. *Rev. Fr. Inf. Rech. Oper.* 16 (1969) 35-43.
- [23] G.A. Kaminski, R.A. Friesner, J. Tirado-Rives, W.L. Jorgensen, Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides, *J. Phys. Chem. B*, 105 (2001) 6474-6487.
- [24] Schrödinger Suite; Schrodinger: New York, 2014.
- [25] C. Still, A. Tempczyk, R. Hawley, T. Hendrickson, Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* 112 (1990) 6127-6129.
- [26] M. Garavís, B. López-Méndez, A. Somoza, J. Oyarzabal, C. Dalvit, A. Villasante, R. Campos-Olivas, C. González, Discovery of Selective Ligands for Telomeric RNA G-quadruplexes (TERRA) through 19F-NMR Based Fragment Screening. *ACS Chem. Biol.* 9 (2014) 1559-1566.
- [27] G.W. Collie, G.N. Parkinson, The application of DNA and RNA G-quadruplexes to therapeutic medicines. *Chem. Soc. Rev.* 40 (2011) 5867-5892.
- [28] M.M. Mysinger, M. Carchia, J.J. Irwin, B.K. Shoichet, Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *J. Med. Chem.* 55 (2012) 6582-6594.
- [29] LigPrep, version 2.5, Schrödinger, LLC, New York, NY, 2012.
- [30] Glide, version 6.2, Schrödinger, LLC, New York, NY, 2014.
- [31] S. Rao, P.C. Sanschagrin, J.R. Greenwood, M.P. Repasky, W. Sherman, R. Farid, Improving database enrichment through ensemble docking. *J. Comput. Aided Mol. Des.* 22 (2008) 621-7.
- [32] G. Wolber, T. Langer, LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *J. Chem. Inf. Model.* 45 (2005) 160–169.
- [33] R. Rocca, G. Costa, A. Artese, L. Parrotta, F. Ortuso, E. Maccioni, O. Pinato, M.L. Greco, C. Sissi, S. Alcaro, S. Distinto, F. Moraca, Hit Identification of a Novel Dual Binder for h-telo/c-myc G-Quadruplex by a Combination of Pharmacophore Structure-Based Virtual Screening and Docking Refinement. *ChemMedChem* (2016) doi: 10.1002/cmdc.201600053.
- [34] D. Schuster, P. Markt, U. Grienke, J. Mihaly-Bison, M. Binder, S.M. Noha, J.M. Rollinger, H. Stuppner, V.N. Bochkov, G. Wolber, *Bioorg Med Chem.* 19 (2011) 7168–7180.
- [35] J. Kirchmair, P. Markt, S. Distinto, G. Wolber, T. Langer, Evaluation of the performance of 3D virtual screening protocols: RMSD comparisons, enrichment assessments, and decoy selection—What can we learn from earlier mistakes? *J. Comput. Aided Mol. Des.* 22 (2008) 213-228.
- [36] N.J. Greenfield, Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nat. Protoc.* 1 (2006), 2527-2535.
- [37] A.T. Phan, D.J. Patel, Two-repeat human telomeric d(TAGGGTTAGGGT) sequence forms interconverting parallel and antiparallel G-quadruplexes in solution: distinct topologies, thermodynamic properties, and folding/unfolding kinetics. *J. Am. Chem. Soc.* 125 (2003), 15021-15027.

- [38] G.N. Parkinson, M.P. Lee, S. Neidle. Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*, 417 (2002), 876-880.
- [39] F. Rosu, V. Gabelica, C. Houssier, E. De Pauw, Determination of affinity, stoichiometry and sequence selectivity of minor groove binder complexes with double-stranded oligodeoxynucleotides by electrospray ionization mass spectrometry, *Nucleic Acids Res.* 30 (2002) e82.
- [40] E.B. Erba, R. Zenobi, Mass spectrometric studies of dissociation constants of noncovalent complexes, *Annual Reports Section "C" (Physical Chemistry)*. 107 (2011) 199. doi:10.1039/c1pc90006d.
- [41] K.B Turner, Inhibitory effects of archetypical nucleic acid ligands on the interactions of HIV-1 nucleocapsid protein with elements of  $\omega$ -RNA, *Nucleic Acids Research*. 34 (2006) 1305-1316. doi:10.1093/nar/gk1004.
- [42] G. Yuan, Q. Zhang, J. Zhou, H. Li, Mass spectrometry of G-quadruplex DNA: Formation, recognition, property, conversion and condensation, *Mass Spectrometry Reviews*. 30 (2011) 1121-1142. doi:10.1002/mas.20315.
- [43] A. M. Zahler, J. R. Williamson, T. R. Cech, D. M. Prescott, Inhibition of telomerase by G-quartet DNA structures. *Nature* 350 (1991) 718-720.
- [44] S. Alcaro, C. Musetti, S. Distinto, M. Casatti, G. Zagotto, A. Artese, L. Parrotta, F. Moraca, G. Costa, F. Ortuso, E. Maccioni, C. Sissi, Identification and characterization of new DNA G-quadruplex binders selected by a combination of ligand and structure-based virtual screening approaches, *J. of Med. Chem.*, 56 (2013) 843-855.



**Fig. 1.** 2D structures of 9 known active G4 ligands selected on the basis of their affinity and selectivity toward bimolecular RNA (TERRA<sub>2</sub>) and DNA G4s (Tel<sub>2</sub>) *versus* duplex DNA: **a)** Naphtalene diimide with a dimethylamino-propyl side chains, **b)** Naphtalene diimide with a propylamino-pyrrolidin side chain, **c)** Naphtalene diimide with a bis dimethylamino-hydroxypropylamino side chain, **d)** Acridine-N,N0-((1,10-(acridine-3,6-diyl)bis(1H-1,2,3-triazole-4,1-diyl))bis(3,1-phenylene))bis(2-(diethylamino)acetamide, **e)** BRACO-19, **f)** PDS (Pyridostatin), **g)** 9-Fluoro-2-methyl-5H-thiocromeno[4,3-d] pyrimidine, **h)** 2-chloro-6-fluoro-N-(4-fluoro-1-methyl-1H-indazol-3-yl)benzamide, **i)** 2-(2-Fluorophenyl)-N-[(5-methyl-2-thienyl)methyl]ethanamine.

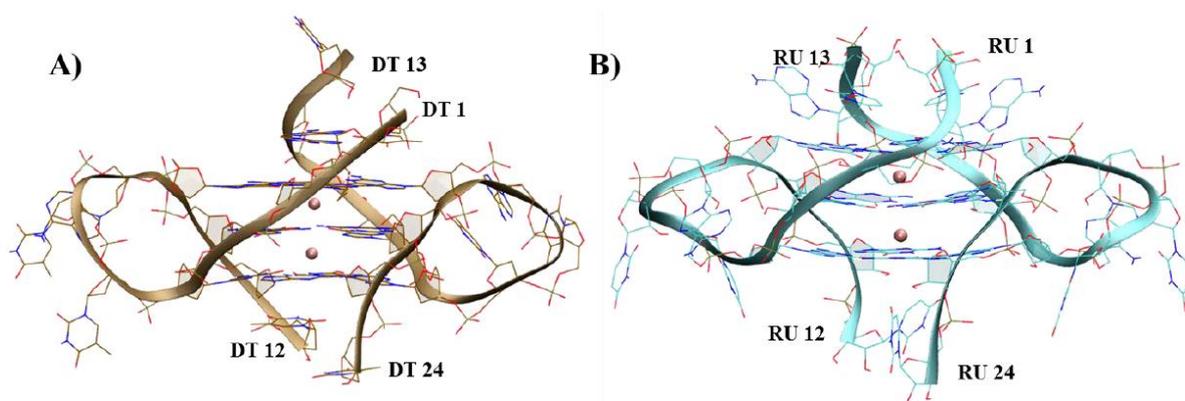
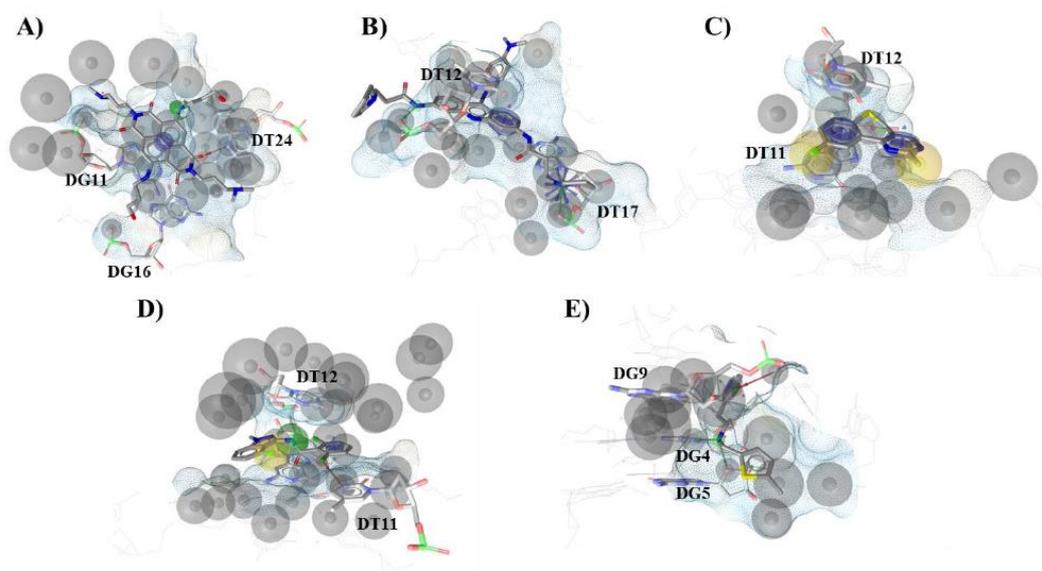


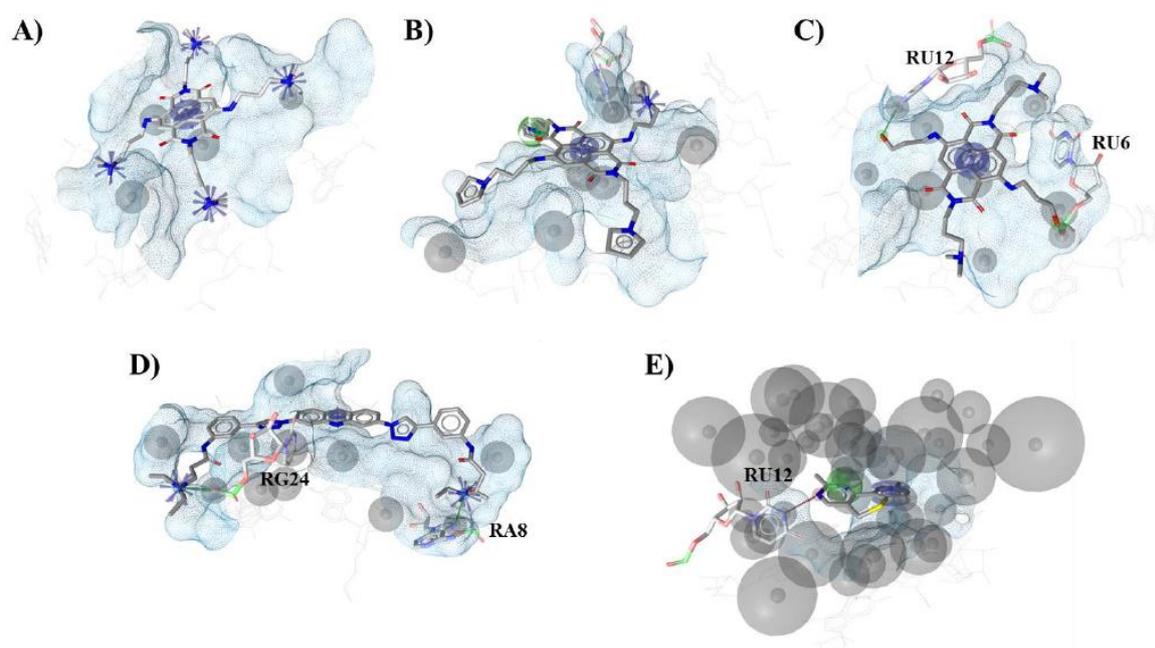
Fig. 2. 3D structures of the receptor models selected from the enrichment study. (A) Tel<sub>2</sub> and (B) TERRA<sub>2</sub>.

ACCEPTED MANUSCRIPT



**Fig. 3.** Structure-based pharmacophore models generated by LigandScout ver. 3.1 based on G4-DNA complexes after the validation step. **(A)** *D-Hypo1*, **(B)** *D-Hypo2*, **(C)** *D-Hypo3*, **(D)** *D-Hypo4*, **(E)** *D-Hypo5* are related to compounds **c**, **e**, **g**, **h** and **i**, respectively. Ligands are represented as grey carbon sticks. Chemical features are color-coded: blue stars represent the positive ionisable (PI) feature; green arrows the hydrogen bond donors (HBD); blue circles the aromatic features (AR). The excluded volumes are reported as dark-grey spheres.

ACCEPTED



**Fig. 4.** Structure-based pharmacophore models generated by LigandScout ver. 3.1 based on  $TERRA_2$  complexes after the validation step. (A) *T-Hypo1*, (B) *T-Hypo2*, (C) *T-Hypo3*, (D) *T-Hypo4*, (E) *T-Hypo6* are related to compounds **a**, **b**, **c**, **d** and **g**, respectively. Ligands are represented as grey carbon sticks. Chemical features are color-coded: blue stars represent the positive ionisable (PI) feature; green arrows the hydrogen bond donors (HBD); blue circles the aromatic features (AR). The excluded volumes are reported as dark-grey spheres.

ACCEPTED

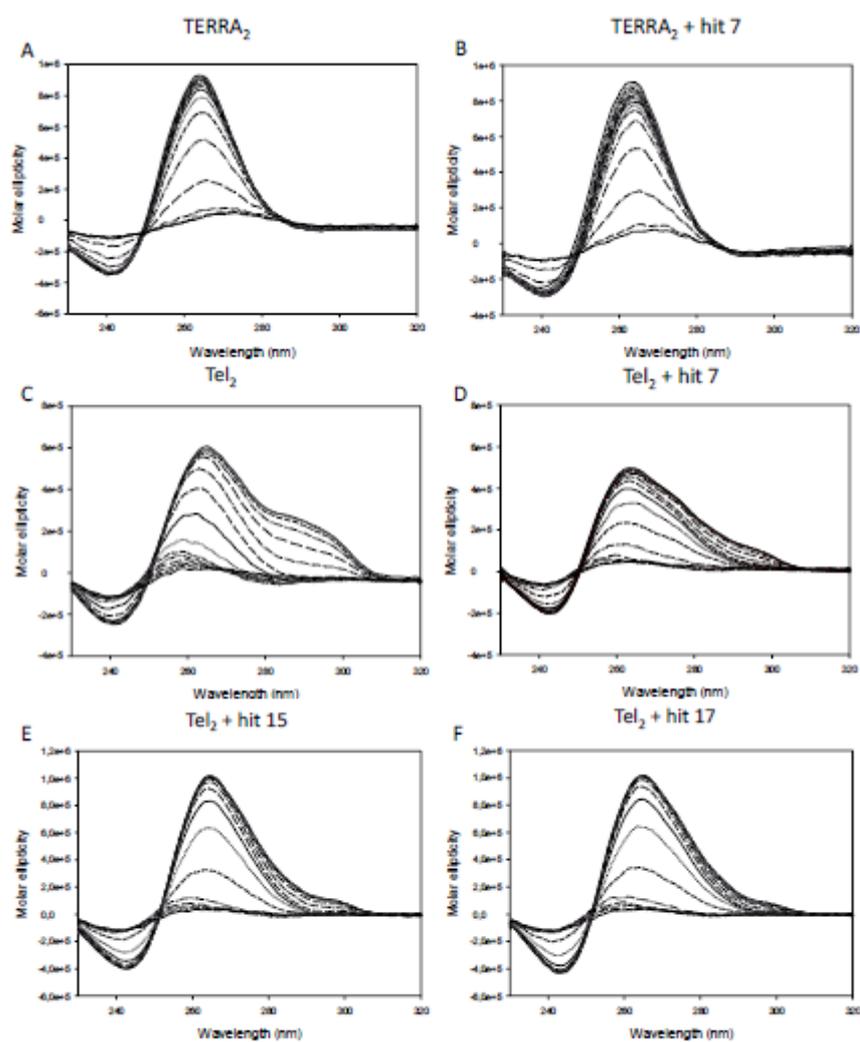
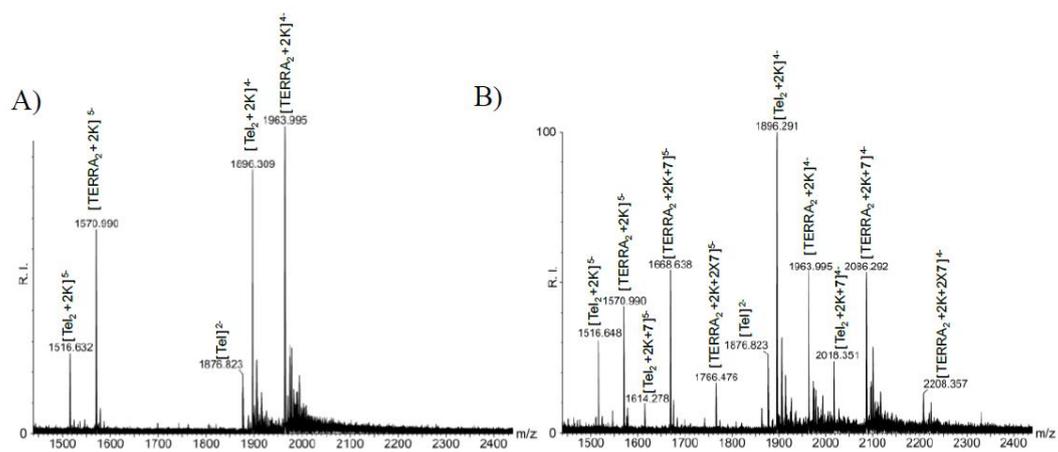


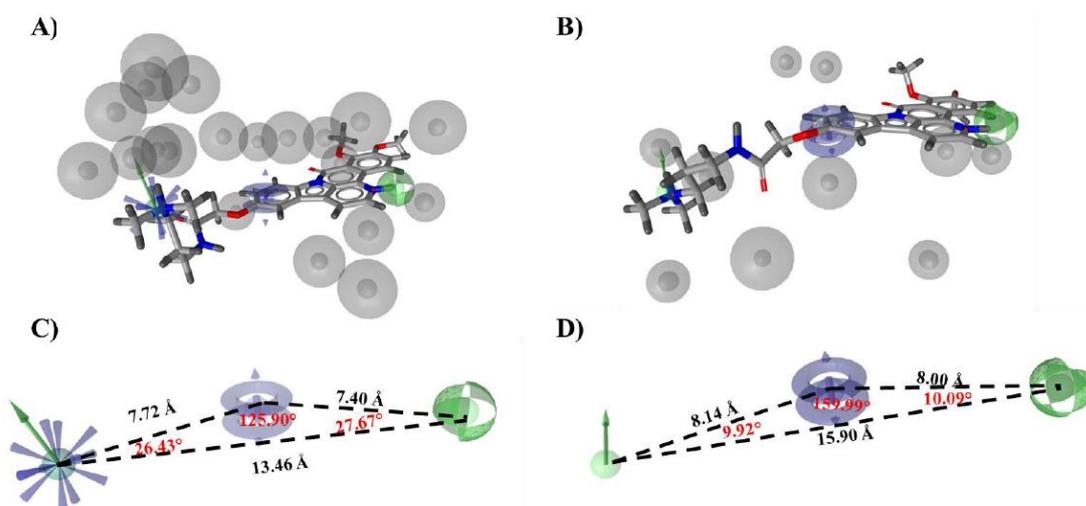
Fig. 5. CD thermal unfolding spectra of the nucleic acid G4s alone and in the presence of the selected hits. A) TERRA<sub>2</sub> RNA G4 (4 μM single strand concentration) in 100 mM K<sup>+</sup>, B) TERRA<sub>2</sub> RNA G4 in the presence of *hit 7* (16 μM) as representative compound. C) Tel<sub>2</sub> DNA G4 (8 μM single strand concentration) in 100 mM K<sup>+</sup>, D-F) Tel<sub>2</sub> DNA G4 in the presence of the indicated compounds (16 μM).





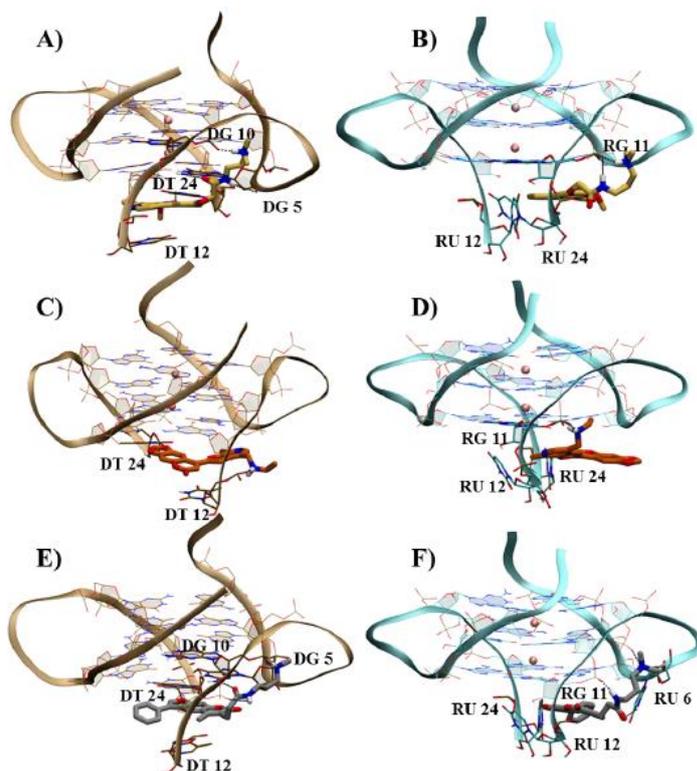
**Fig. 6.** MS spectra of  $TERRA_2$  and  $Tel_2$  separately folded and mixed in 0.8 mM KCl, 120 mM TMAA pH 7.0 and 20 % of IPA in the absence (A) and presence (B) of *hit 7*.

ACCEPTED MAN



**Fig. 7.** Details about the two pharmacophore models retrieving the dual *hit 7*. A) 3D view of *D-Hypo2* and B) *T-Hypo4*, superimposed with the best *hit 7*. Focus on the pharmacophore features respectively for C) *D-Hypo2* and D) *T-Hypo4*, with the relative distances and angles.

ACCEPTED M



**Fig. 8.** Docking pose analysis of the best *hits* of **7**, **15** and **17**. *Hit 7* in complex with (A) Tel<sub>2</sub> and (B) TERRA<sub>2</sub> G4s. Ligand is depicted as yellow carbon sticks. In panels (C) and (D) is reported the binding mode of **15** in complex with Tel<sub>2</sub> and TERRA<sub>2</sub> G4s, respectively. Ligand is depicted as orange carbon sticks. Finally, in panels (E) and (F) binding mode of **17** in complex with Tel<sub>2</sub> and TERRA<sub>2</sub> G4s, respectively. Ligand is depicted as silver carbon sticks. The nucleic acids are shown as brown and cyan ribbons for Tel<sub>2</sub> and TERRA<sub>2</sub> G4s, respectively, while K<sup>+</sup> ions are represented as pink spheres. Ligand/G4s hydrogen bonds are shown as dashed black lines. These binding modes derived from the docking simulation performed at the end of the Virtual Screening.

**Table 1.** Oligonucleotides used in this study.

<b>Name</b>	<b>Sequence 5'-3'</b>
Tel <sub>2</sub>	d(TAGGGTTAGGGT)
TERRA <sub>2</sub>	r(UAGGGUUAGGGU)
dsDNA forward	d(CGGCATGCGTGGCT)
dsDNA reverse	d(AGCCACGCATGCCG)

ACCEPTED MANU

**Table 2.** List of the validated hypotheses of the best docking complexes against Tel<sub>2</sub> with their related chemical features and EF value. Chemical features description: HY=hydrophobic, AR= aromatic ring, PI= positive ionizable, HBD=H bond donor, HBA= H bond acceptor.

Pharmacophore Hypothesis	Bound Actives	n° of features	Chemical features	EF	Actives found (%) (n=5)	Decoys found (%) (n=419)
<i>D-Hypo1</i>	<b>c</b>	3	1AR, 1 HBD <sub>s</sub> <sup>*</sup> , 1 HBA	38.8	60	0.78
<i>D-Hypo2</i>	<b>e</b>	4	1AR <sup>**</sup> , 2HBD <sub>v</sub> , 1PI	77.6	40	0
<i>D-Hypo3</i>	<b>g</b>	4	2AR <sup>**</sup> , 2HY	77.6	20	0
<i>D-Hypo4</i>	<b>h</b>	4	1AR, 1HBD <sub>s</sub> , 1HBA, 1HY	58.2	20	0
<i>D-Hypo5</i>	<b>i</b>	4	1AR <sup>***</sup> , 1HBD <sub>v</sub> , 1HBA	77.6	40	0

\*Tolerance radius decreased of 0.45 Å

\*\*Tolerance radius increased of 0.30 Å

\*\*\*Tolerance radius increased of 0.15 Å

<sub>s</sub> and <sub>v</sub> indicate, respectively, HBD features as spheres or vector

ACCEPTED M

**Table 3.** List of the validated hypotheses for each best docking complex of TERRA<sub>2</sub> active compounds with their related chemical features and EF values. Chemical features description: HY=hydrophobic, AR= aromatic ring, PI= positive ionisable, HBD=H bond donor, HBA= H bond acceptor.

Pharmacophore Hypothesis	Bound Actives	n° of features	Chemical features	EF	Actives found (%) (n=8)	Decoys found (%) (n=419)
<i>T-Hypo1</i>	<b>a</b>	5	1AR <sup>*</sup> , 4PI	48.9	12.5	0
<i>T-Hypo2</i>	<b>b</b>	3	1AR <sup>**</sup> , 1HBD <sub>s</sub> , 1PI	48.9	37.5	0
<i>T-Hypo3</i>	<b>c</b>	3	1AR <sup>*</sup> , 2HBD <sub>v</sub>	48.9	37.5	0
<i>T-Hypo4</i>	<b>d</b>	5	1AR, 2PI, 2HBD <sub>v</sub>	48.9	12.5	0
<i>T-Hypo5</i>	<b>f</b>	3	1AR <sup>**</sup> , 1HBA <sub>s</sub> <sup>***</sup> , 1HY <sup>****</sup>	29.3	37.5	0.52
<i>T-Hypo6</i>	<b>g</b>	3	1AR <sup>*****</sup> , 1HBD <sub>s</sub> , 1HBA <sub>v</sub>	31.1	87.5	1.04
<i>T-Hypo7</i>	<b>h</b>	3	1AR, 2HY <sup>*****</sup>	16.3	37.5	1.56
<i>T-Hypo8</i>	<b>i</b>	4	2HY, 1HY <sup>*****</sup> , 1HBD	24.4	25	0.52

<sup>\*</sup>Tolerance radius increased of 0.45 Å

<sup>\*\*</sup>Tolerance radius increased of 0.30 Å

<sup>\*\*\*</sup>Tolerance radius decreased of 0.30 Å

<sup>\*\*\*\*</sup>Tolerance radius increased of 0.60 Å

<sup>\*\*\*\*\*</sup>Tolerance radius decreased of 0.15 Å

<sub>s</sub> and <sub>v</sub> indicate, respectively, HBD/HBA features as spheres or vector

ACCEPTED

**Table 4.**  $\Delta T_m$  of TERRA<sub>2</sub> RNA G4 (4  $\mu$ M single strand concentration) and Tel<sub>2</sub> DNA G4 (8  $\mu$ M single strand concentration) in 100 mM K<sup>+</sup>, in the presence of the compounds (8  $\mu$ M and 16  $\mu$ M, respectively), measured by melting CD spectroscopy.  $\Delta T_m$  values are reported in Celsius degrees (°C).

TERRA <sub>2</sub>		Tel <sub>2</sub>	
<i>Hit #</i>	$\Delta T_m$ (°C)	<i>Hit #</i>	$\Delta T_m$ (°C)
<b>1</b>	< 1.0	<b>7</b>	8.3 ± 0.3
<b>2</b>	< 1.0	<b>8</b>	< 1.0
<b>3</b>	< 1.0	<b>9</b>	< 1.0
<b>4</b>	< 1.0	<b>10</b>	< 1.0
<b>5</b>	< 1.0	<b>11</b>	< 1.0
<b>6</b>	< 1.0	<b>12</b>	< 1.0
<b>7</b>	6.8 ± 0.4	<b>13</b>	< 1.0
		<b>14</b>	< 1.0
		<b>15</b>	9.0 ± 0.3
		<b>16</b>	< 1.0
		<b>17</b>	9.1 ± 0.2
		<b>18</b>	1.8 ± 0.2
		<b>19</b>	< 1.0
		<b>20</b>	< 1.0

ACCEPTED

**Table 5.**  $\Delta T_m$  ( $^{\circ}\text{C}$ ) of  $\text{Tel}_2$  DNA G4 and  $\text{TERRA}_2$  RNA G4 in the presence of the selected hits at 100 mM  $\text{K}^+$  and at lower  $\text{K}^+$  concentrations. The symbol \* indicates destabilization of the G4. NA stands for nucleic acid.

NA	<b>Tel<sub>2</sub> DNA G4</b>		<b>TERRA<sub>2</sub> RNA G4</b>					
<b>K<sup>+</sup> (mM)</b>	<b>100</b>		<b>100</b>		<b>5</b>		<b>2</b>	
	<b>T<sub>m</sub></b>	<b><math>\Delta T_m</math></b>	<b>T<sub>m</sub></b>	<b><math>\Delta T_m</math></b>	<b>T<sub>m</sub></b>	<b><math>\Delta T_m</math></b>	<b>T<sub>m</sub></b>	<b><math>\Delta T_m</math></b>
<b>free NA</b>	52.6± 0.1		71.0± 0.1		60.5± 0.1		54.3± 0.2	
<i>hit 7</i>		8.3± 0.3		6.8± 0.4		10.8± 0.3		*
<i>hit 15</i>		9.0± 0.3		2.8± 0.3		5.8± 0.4		9.2± 0.3
<i>hit 17</i>		9.1± 0.2		2.1± 0.2		5.6± 0.4		9.0± 0.5

ACCEPTED MANUSCRIPT

**Table 6.** Relative binding affinities (BA) of the selected *hits* to TERRA<sub>2</sub>, Tel<sub>2</sub> and dsDNA.

	BA								
	<i>hit 7</i>			<i>hit 15</i>			<i>hit 17</i>		
Competing G4	TERRA <sub>2</sub>	Tel <sub>2</sub>	dsDNA	TERRA <sub>2</sub>	Tel <sub>2</sub>	dsDNA	TERRA <sub>2</sub>	Tel <sub>2</sub>	dsDNA
TERRA <sub>2</sub> /Tel <sub>2</sub>	62.6	23.3		45.2	57.9		51.3	47.0	
Tel <sub>2</sub> /dsDNA		25.8	29.8		59.8	29.6		53.2	27.4
TERRA <sub>2</sub> /dsDNA	82.1		33.6	47.7		31.1	57.3		34.2

ACCEPTED MANUSCRIPT

**Table 7.** Contact analysis of the best tested *hits* **7**, **15** and **17** against Tel<sub>2</sub> and TERRA<sub>2</sub> G4s.

<i>Hit</i>	Tel <sub>2</sub>			TERRA <sub>2</sub>		
	Hbonds	Stacking	Good	Hbonds	Stacking	Good
<b>7</b>	3	8	330	3	5	278
<b>15</b>	1	6	290	1	3	293
<b>17</b>	3	0	314	2	2	246

ACCEPTED MANUSCRIPT

### Highlights

- TERRA G4, involved in crucial biological processes, could be an anticancer target.
- First virtual screening performed on TERRA G4 structures.
- We identified three novel promising stabilizers against Tel<sub>2</sub>/TERRA<sub>2</sub> G4s.
- The naphthyridin scaffold is proposed as the most interesting one against TERRA<sub>2</sub>.
- A furo-chromen and a benzofuran derivatives preferentially recognized Tel<sub>2</sub>.

ACCEPTED MANUSCRIPT