

Scientific paper

Prediction of Hit-to-Lead Ligand Molecule Interaction with G-Quadruplex DNA from c-Myc Oncogene Promoter Region

Petar M. Mitrasinovic*

Center for Biophysical and Chemical Research, Belgrade Institute of Science and Technology, 11060 Belgrade, Serbia

* Corresponding author: E-mail: pmitrasinovic.ist-belgrade.edu.rs@tech-center.com

Received: 03-02-2019

Abstract

Targeting guanine (G)-rich DNA sequences, folded into non-canonical G-quadruplex (G4) structures, by small ligand molecules is a potential strategy for gene therapy of cancer disease. BRACO-19 has been recently established as a unique (thermodynamically favorable and highly selective) binder, being involved in the external stacking mode of interaction with a G4-DNA formed in the c-Myc oncogene promoter region (P. M. Mitrasinovic, *Croat. Chem. Acta* **2019**, 92, 43–57). Herein, hit-to-lead ligands are identified using high-throughput virtual screening (HTVS). Search of the Kyoto Encyclopedia of Genes and Genomes (KEGG) databases is performed using the key pharmacophore features of BRACO-19. At the very outset, out of a total of 29,009 entries, 95 hits are extracted and evaluated by docking them in the binding sites of G4. Then, 22 hits are chosen by observing the binding free energies. Consequently, 3 hit-to-lead candidates are selected on the basis of structural criteria. Finally, a lead candidate structure is proposed using analog design and considering both the structural and physicochemical requirements for optimal biological activity and a variety of pharmacological standpoints. Implications of the present study for experimental research are discussed.

Keywords: Anti-cancer drug design; BRACO-19; c-Myc oncogene promoter; G-quadruplex DNA; hit-to-lead ligand

1. Introduction

In addition to forming various canonical duplex structures, highly dynamical DNA macromolecules are able to fold into non-canonical structures, including hairpin, triplex, G-quadruplex, and i-motif. G-quadruplexes (or G-tetraplexes) are secondary structures that form within guanine rich strands of regulatory genomic regions (human telomeres, oncogene promoter regions, immunoglobulin switch regions, ribosomal DNA, some regions of RNA). Even though G4s associate with various conformations and folding energies, and their thermodynamic stabilities are comparable to those of duplex structures, the function of G4s *in vivo* is not fully understood. G4s are hypothesized to participate in important biological phenomena, including telomere maintenance, end-capping and protection, chromosome stability, gene expression, viral integration, and recombination.^{1,2} A relevant consequence of G-quadruplex formation in telomeric DNA is the inhibition of telomere elongation by telomerase in can-

cer cells.^{3,4} An increasing number of identified G4-binding proteins means that protein/G4 interactions are associated with important cellular events. Use of small molecules for targeting G4 in order to disrupt protein/G4 recognition emerges as a potential strategy for directing anti-cancer therapy.⁵

G4 structures primarily consist of two or more stacked G-tetrads (or G-quartets) assembled either from a single strand of DNA in an intramolecular (backfolded) way or from two-, three-, or four DNA strands in an intermolecular way. Every single G-tetrad contains four G-G base pairs (bps) linked by Hoogsteen hydrogen bonds. G4s are more compact structures than duplex DNAs and display well-defined binding sites (external stacking, intercalating, and groove/loop).^{2,6} Small ligand molecules are expected to be complementary in shape and charge to the biological target. The question of finding ligands that conform to the structural and physicochemical requirements for optimal biological activity is of current relevance. This work is, to some extent, imagined to contribute to the bet-

ter formulation of a daunting challenge – how to design small molecules that can bind selectively to each of the many possible G4 structures.

A G-rich element of repeated sequences with three or four guanine residues (between –137 and –115 bp upstream of the P1 promoter in the c-Myc oncogene) can fold in an intramolecular G4 structure (Figure 1) in order to suppress c-Myc transcription in a silenced form.⁷ This element is a potential target for down-regulation of c-Myc overexpression in tumor cells.^{8,9} The dynamics of noncovalent interaction between structurally diversified ligand molecules (with a pronounced propensity for the receptor)⁹ and the G4 was investigated in a systematic fashion.¹⁰ Among the highest affinity ligands, BRACO-19 (Figure 2), a pure G-quartet binder, was established as a unique – thermodynamically favorable ligand, increasing conformational flexibility of the G4 structure through its stacking mode of interaction.¹⁰ By using the pharmacophore features of BRACO-19 (Figure 2), that is, the structural features of the ligand that are recognized at a receptor site and responsible for the ligand's biological activity, a subtle *in silico* protocol followed by analog design is employed in this work, with the ultimate goal to determine lead candidate structure.

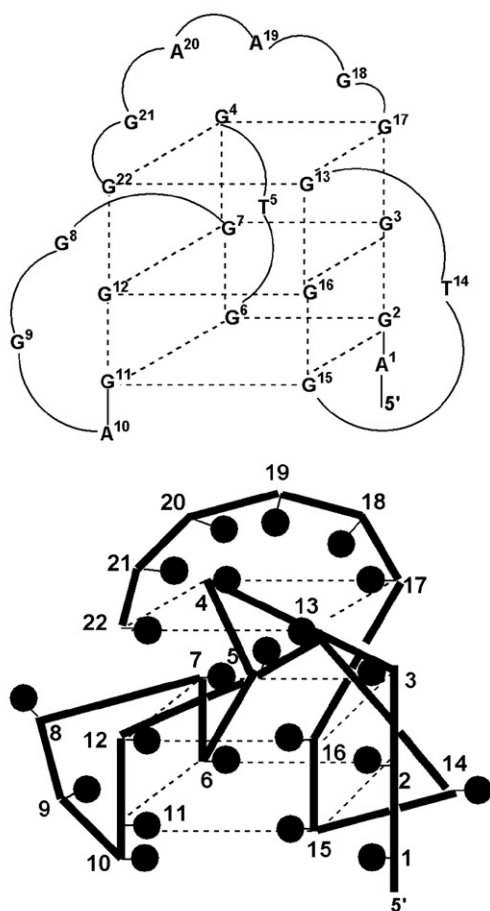


Figure 1. Assembly and topology of G-tetrads in G-quadruplex structure.

2. Methods

Experimental structure of the monomeric parallel-stranded G-quadruplex (Figure 1) was retrieved from the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB) in order to obtain the initial coordinates of the target atoms (PDB ID: 2A5P).¹¹

The term “pharmacophore” means a spatial arrangement of the essential features of an interaction.^{12–14} These features of BRACO-19 (Figure 2) were identified using the interface Pharmit – an online, interactive environment for exploration of chemical space.¹⁵ Features supported by the web server include hydrogen bond acceptors and donors, negative and positive charges, aromatics, and hydrophobic features. For a provided ligand structure as a PDB input, the algorithm searches for these features using tolerance spheres. Structural parts of a compound match if they can be positioned in such a way that their corresponding features are located within these spheres. Some features can have additional constraints, such as size (number of atoms) for hydrophobic features and direction for hydrogen bonds and aromatics.¹⁵

Hits were generated by searching the Kyoto Encyclopedia of Genes and Genomes (KEGG) databases COMPOUND and DRUG. This search was based upon both the key pharmacophore features of BRACO-19 and SIMCOMP (SIMilar COMPOund) – a graph-based method that is implemented in the KEGG system for searching and comparing chemical structures in the databases.^{16–18} SIMCOMP provides the atom alignments between two chemical compound graphs and calculates the similarity of two chemical compounds by counting the number of matched atoms in those atom alignments. For all calculations in SIMCOMP, the Global Search was performed and the KEGG Atom Types were chosen as a representation of atoms in order to detect biochemically meaningful features. The KEGG Atom Types are based on the chemical concept of functional groups and 68 atom types (vertex types) are defined for carbon, nitrogen, oxygen, and other atomic species with different environments.^{16–18} All the other default options were exploited.

Virtual screening was performed for small molecules (hits) collected from KEGG against G-quadruplex DNA (PDB ID: 2A5P). The Windows platform-based graphical interface Raccoon was used for preparing and automating the AutoDock virtual screening.¹⁹

Flexible docking of each ligand (hit) in the receptor was performed by AutoDock 4.2.^{20,21} Noteworthy is to see into why the particular method was chosen. Docking problem is an exhaustive search problem that includes many degrees of freedom. It means that the use of efficient docking algorithms is critical for finding optimal ligand/receptor configuration and for predicting accurate binding free energy without fetching formal statistical mechanics methods. A fundamental idea underlying AutoDock 4.2 is to calculate the total ligand/receptor binding free energy

by summing distinct, physically interpretable contributions.²² Scoring functions are calibrated using multivariate regression analysis of a set of ligand/target receptor complexes with respect to experimentally determined structures and binding affinities. The final form of a scoring function depends on the size and quality of the training set. Since scoring functions are derived from diverse ligand/receptor complexes, possible applications are not restricted to a particular set of ligands or a specific receptor. An average level of thermochemical accuracy of 2 kcal mol⁻¹ in binding affinity predictions makes empirical scoring acceptable for the structure-based drug (or ligand) design.^{23–32} The distinct energetic terms considered throughout this work account for the hydrogen bonding, the van der Waals (vdW) interactions, the electrostatic interactions, the desolvation-mediated ligand/receptor binding, the total internal energy, the torsional potential, and the unbound system's energy respectively.²⁰ Entropy of ligand interaction is reflected through the loss of degrees of freedom upon binding and is included via the torsional potential being proportional to the number of torsions (sp³ bonds) in the ligand. The Lamarckian Genetic Algorithm in combination with a grid-based energy evaluation method was employed to calculate grid maps, while atomic potential grid map was computed by AutoGrid4 with a 0.536 Å spacing in a 65 Å × 65 Å × 65 Å (1 Å = 10⁻¹⁰m) box centered on the macromolecule. All the other default options were chosen in the AutoDockTools4 for preparing the systems for runs.²¹ The lowest energy and physically meaningful (in terms of the spatial orientation of a ligand with respect to the compact binding sites of G4) conformations were extracted from docking experiments. A hit affinity for the receptor was estimated by the total binding

free energy ($\Delta G_{\text{binding}}$) or the dissociation constant (K_d), taking into account the relation $\Delta G_{\text{binding}} = RT \ln(K_d)$ ($R = 1.9872 \text{ kcal K}^{-1} \text{ mol}^{-1}$ – the gas constant, $T = 300 \text{ K}$ – the absolute temperature).

3. Results and Discussion

Ligand π - π stacking at the end of G-quadruplex can be considered as a preferred mode of interaction according to experimental^{9,33} and theoretical^{5,10} results. Although nonspecific ligand-groove/loop binding is not inherently stable due to its dependence on a particular topology of the groove/loop,^{5,9} the groove/loop is of interest for the structure-based drug design. This recognition motif is a viable site for blocking the interactions between G4 and its binding proteins in aqueous solution.⁵ Search for ligands that satisfy the structural and physicochemical requirements for optimal biological activity is currently needed.

From a rigorous biophysical standpoint, the dynamics of interaction of structurally different ligand molecules with the G4 (Figure 1) was recently explored and characterized in a systematic fashion.¹⁰ As a consequence, the highest affinity ligands, being involved in external stacking and groove binding, were observed respectively. Interestingly, BRACO-19 (Figure 2) – a pure G-quartet binder was established to be a unique (thermodynamically favorable) ligand in terms of increasing conformational flexibility of the receptor upon external stacking.¹⁰ The pharmacophore of BRACO-19, which can be defined as a set of structural features in the ligand that is recognized at a receptor site and is responsible for the ligand's biological activity, is a starting point of the present work. The particular set of structural features consists of: i) three aromatic and hydrophobic rings making the core scaffold, ii) two peripheral and hydrophobic hexagons being symmetrically attached to the core scaffold through adequate linkers, iii) three hydrogen-bond donors (HBDs), iv) five hydrogen-bond acceptors (HBAs), and v) a side chain that contains an aromatic and hydrophobic hexagon as well as a hydrophobic region on top of it (Figure 2).

KEGG database search generated hits using the pharmacophore of BRACO-19 as a template. Out of a total of 29,009 entries, 95 hits were extracted, 21 from the database Compound (Table S1, Supplementary Material) and 74 from the database Drug (Table S2, Supplementary Material). Conformations obtained by docking the hit structures in the compact binding sites of G4 were scored and affinities for G4 were evaluated.

The potency of a substance (the concentration required to achieve a defined biological effect) must be significant in order to identify a hit-to-lead. The particular concentration is in the micromolar (10⁻⁶ M) range for a hit and in the nanomolar (10⁻⁹ M) range for a lead candidate.^{34,35} Affinity issue can be conveniently seen through the total binding free energy ($\Delta G_{\text{binding}}$) or the dissociation

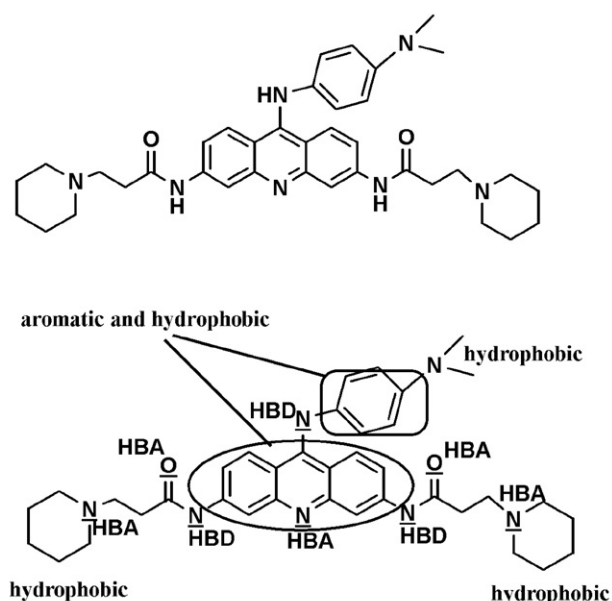


Figure 2. Structure of BRACO-19 and its pharmacophore features. Each underlined atom is hydrogen bond donor (HBD) or hydrogen bond acceptor (HBA).

constant (K_d), taking into account the relation $\Delta G_{\text{binding}} = RT \ln(K_d)$ ($R=1.9872 \text{ kcal K}^{-1} \text{ mol}^{-1}$ – the gas constant, $T=300 \text{ K}$ – the absolute temperature). The values of $\Delta G_{\text{binding}}$ and K_d for the BRACO-19:G4 complex, $-6.77 \text{ kcal mol}^{-1}$ and $12.01 \mu\text{M}$ (the footnote a of Table 1), are the references for

selecting a hit that is supposed to have a higher affinity for the receptor. Thus, twenty two hits (out of the previous ninety five), which satisfy this criterion, are selected (Figure 3) and their affinity-based ranking is summarized in Table 1.

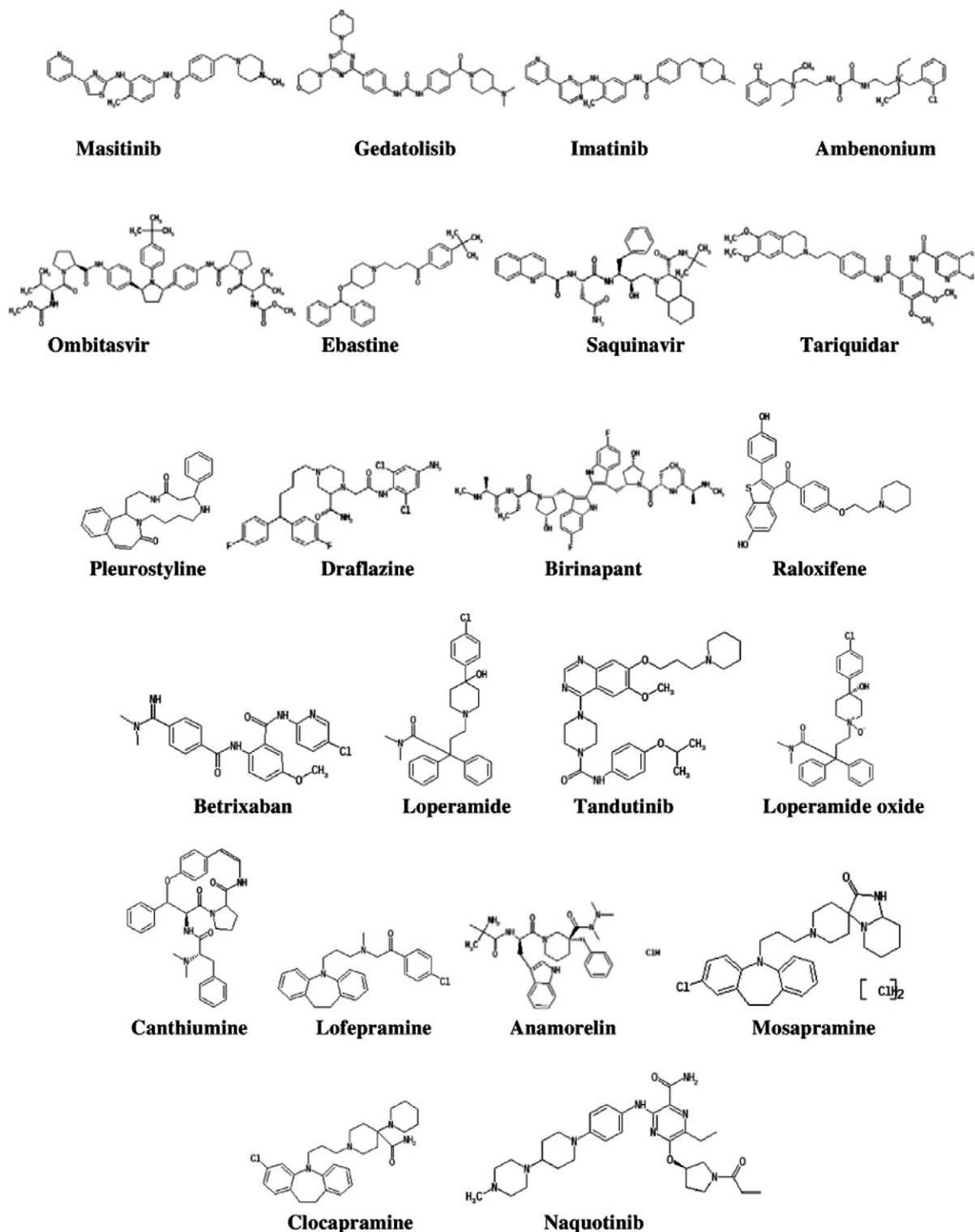


Figure 3. Structures of hit ligand molecules with the most pronounced affinity for the G4-DNA target.

Table 1. Ranking of hits according to the predicted binding free energy

Entry	Ligand name ^(a)	$\Delta G_{\text{binding}}^{\text{(b)}}$ (kcal mol ⁻¹)	Dissociation constant ^(c) K_d (μM)	$\Delta G_{\text{intermolecular}}^{\text{(d)}}$ (kcal mol ⁻¹)
D10229	Masitinib	-11.85	0.0021	-13.04
D10635	Gedatolisib	-10.50	0.022	-11.69
D08066	Imatinib	-10.36	0.025	-11.56
D01001	Ambenonium	-9.88	0.061	-11.07
D10576	Ombitasvir	-9.82	0.064	-11.90
D01478	Ebastine	-9.48	0.11	-10.97
D00429	Saquinavir	-9.45	0.12	-10.35
D06008	Tariquidar	-9.38	0.13	-10.87
C10612	Pleurostyline	-9.27	0.16	-9.27
D03906	Draflazine	-9.18	0.19	-9.77
D10417	Birinapant	-9.17	0.19	-10.66
C07228	Raloxifene	-8.86	0.32	-10.05
D08873	Betrixaban	-8.84	0.33	-9.73
D08144	Loperamide	-8.68	0.43	-9.57
D06005	Tandutinib	-8.52	0.57	-10.01
D07113	Loperamide oxide	-8.45	0.64	-9.05
C10000	Canthiumine	-8.45	0.64	-8.75
D08856	Anamorelin	-8.32	0.80	-9.21
D08140	Lofepamine	-8.29	0.84	-9.18
D01548	Mosapramine	-8.23	0.92	-8.53
D07718	Clocapramine	-8.21	0.95	-8.21
D10958	Naquotinib	-8.20	0.97	-9.99

^(a) Reference values for the BRACO-19:G4 complex are -6.77 kcal mol⁻¹, 12.01 μM , and -9.99 kcal mol⁻¹ respectively.

^(b) The AutoDock 4.2 score: $\Delta G_{\text{binding}} = E_{\text{vdW}} + E_{\text{Hbond}} + E_{\text{desolvation}} + E_{\text{electrostatic}} + E_{\text{internal}} + E_{\text{torsional}} - E_{\text{unbound}}$

^(c) $\Delta G_{\text{binding}} = RT \ln(K_d)$, R – the gas constant (1.9872 kcal K⁻¹ mol⁻¹), T – the absolute temperature (300 K), 1 $\mu\text{M} = 10^{-6}$ M.

^(d) $\Delta G_{\text{intermolecular}} = E_{\text{vdW}} + E_{\text{Hbond}} + E_{\text{desolvation}} + E_{\text{electrostatic}}$. The intermolecular energy represents the largest (most negative) contribution to the stability (binding free energy) of the complexes and does not conform to the trend displayed by the values of both $\Delta G_{\text{binding}}$ and K_d .

The intermolecular energy (IE), also known as the interaction energy, is a key part of the enthalpy of formation of a molecular complex. In practical energetic analyses, the intermolecular energy is viewed as the largest contribution to the stability (total binding free energy) of a complex. The IE is defined as the sum of distinct energetic terms that account for the van der Waals (vdW) interactions, the hydrogen bonding, the desolvation-mediated receptor-ligand binding, and the electrostatic interactions respectively (the footnote d of Table 1). A numerical inspection of the values given in the last column of Table 1 shows that the key (negative) contribution to $\Delta G_{\text{binding}}$ comes from the IE and that the trend of IE values does not conform to the trend displayed by the values of $\Delta G_{\text{binding}}$ and K_d separately. The IE of the BRACO-19:G4 complex, -9.99 kcal mol⁻¹, is the reference (the footnote a of Table 1). In comparison to the reference, the hits can be divided into two subgroups: the first one with eleven hits having the IE that is more negative than the refer and the second one with the remaining hits. Thus, the members of the first subgroup are: Masitinib, Gedatolisib, Imatinib, Ambenonium, Ombitasvir, Ebastine, Saquinavir, Tariquidar, Birinapant, Raloxifene, and Tandutinib (Table 1). In other words, as far as affinity issue is concerned, hit-to-lead candidates belong to this subgroup of hits.

In order to further filter out hit-to-lead ligands, it is necessary to invoke some structural arguments. The template structure of BRACO-19 mainly takes part in π - π stacking with the G²G⁶G¹¹G¹⁵ tetrad by way of its core aromatic scaffold and, therefore, BRACO-19 is considered to be a G-quartet-binding ligand, even though it is involved in several, additional electrostatic interactions with the residues A¹, G⁶, G¹¹, and G¹⁵ by way of its side chains.¹⁰ Since the interaction energy of BRACO-19 is rooted in π - π stacking, any hit-to-lead candidate with a more negative interaction energy is expected to be both primarily associated with external stacking of its core scaffold and more prone than BRACO-19 to electrostatic interactions via its side chain configurations. Fact that the structure of BRACO-19 contains four aromatic and hydrophobic rings (Figure 2) is employed to recruit hit-to-lead candidates from the first subgroup of hits. An inspection of the eleven hit structures (Figure 3) illustrates that the structures of Masitinib, Imatinib and Raloxifene only have four aromatic and hydrophobic rings (Figure 4).

Knowing the structures of hit-to-lead candidates (Figure 4), the question to be raised is: what is a relevant structural basis upon which a lead candidate should rely? Besides observing individual structural and functional features of every single hit-to-lead candidate, a postulate of

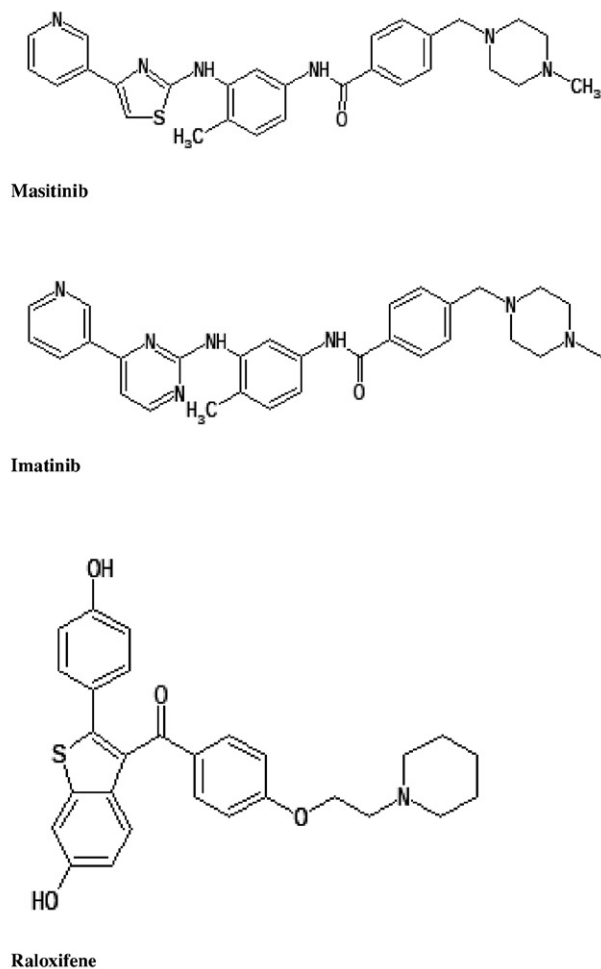


Figure 4. Structures of hit-to-lead candidates.

outstanding importance is to maintain the structural similarity between a lead candidate and a template structure (BRACO-19) as much as possible.

In contrast to Imatinib, noteworthy is that Masitinib and Raloxifene contain thiophene – a five-membered, sulfur-containing heteroaromatic ring that is often a building block in drugs (Figure 4). Metabolism of thiophene can cause formation of reactive metabolites that may be responsible for drug-induced liver damage. Even though its presence in drugs does not necessarily result in toxic effects, thiophene is seen as a kind of structural alert. For example, tienilic acid – a thiophene-based drug was removed from the market after being both associated with severe cases of immune hepatitis and in use for only several months.³⁶ BRACO-19 does not contain a thiophene moiety (Figure 2). These observations substantiate the choice of Imatinib as a favorable hit-to-lead candidate. This choice is agreeable with the experimentally detected ability of Imatinib to downregulate telomerase activity and to inhibit proliferation in telomerase-expressing cell lines by targeting various cellular components.³⁷

The structural alterations of Imatinib (Figure 4) were needed in order to proceed to the proposal of lead candi-

date (Figure 5). These steps were guided by the structural and physicochemical requirements for optimal biological activity. To make the core scaffold composed of three fused aromatic rings (as that of BRACO-19, Figure 2), a carbon atom of the bottom methyl group (Figure 4) is replaced by nitrogen and an adjacent N atom is replaced by a C atom. The newly introduced N and C atoms are then bonded and the closure of the intermediate ring is achieved (Figure 5). Also, nitrogen on top of the left-hand side ring is substituted by a C atom and carbon in the central ring is replaced by an N atom (Figure 5). As for the template structure of BRACO-19 (Figure 2), two peripheral and hydrophobic hexagons that are symmetrically attached to the core scaffold through adequate linkers were shown to additionally stabilize an external stacking conformation in the stable regime of molecular dynamics simulation.¹⁰ To mimic this functionality of BRACO-19, a copy of the right-hand side chain of Imatinib (Figure 4) is introduced (Figure 5) by replacing an aromatic ring (Figure 4) that is attached to the left-hand side of the core scaffold. The side chain of BRACO-19, which arises from the middle ring of the core scaffold and contains an aromatic six-membered ring (Figure 2), was found not to interact with the receptor, but its primary role in the binding conformation was to reduce deviations (or distortions) of the stacking portion of the BRACO-19 structure from horizontal planarity.¹⁰ To additionally maintain a clear resemblance of lead candidate to BRACO-19, the particular side chain (as is – without any change) is attached to the intermediate ring of the core scaffold of the modified Imatinib (Figure 5).

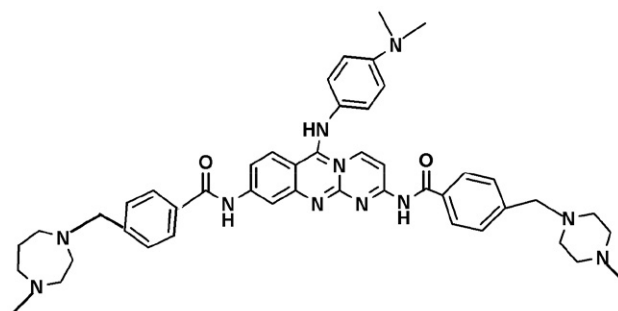


Figure 5. Proposal of lead candidate structure.

As discussed so far, virtual screening resulted in heat-to-lead candidate (Imatinib), while a transition process, from Imatinib to lead candidate, was treated as analog design. The overall protocol was initially imagined to suggest lead candidate that is able to functionally outperform BRACO-19 in binding to the G-quadruplex. To examine the extent of success of this undertaking, proposed lead candidate is docked in the target and obtained binding conformation is contrasted to that of BRACO-19 quantitatively and qualitatively. The values of $\Delta G_{\text{binding}}$ and K_d for the lead candidate:G4 complex ($-11.29 \text{ kcal mol}^{-1}$ and $0.0053 \text{ }\mu\text{M}$) relative to those for the BRACO-19:G4

complex ($-6.77 \text{ kcal mol}^{-1}$ and $12.01 \text{ }\mu\text{M}$) indicate a stronger affinity of the lead candidate for the receptor in comparison to the reference. A K_d value of 5.3 nM conforms to the requirement of being in the nanomolar (10^{-9} M) range that is relevant for a lead. The intermolecular energy of the lead candidate is estimated to be $-13.68 \text{ kcal mol}^{-1}$, which is more negative both than any corresponding value for a hit (Table 1) and than the reference value for BRACO-19 ($-9.99 \text{ kcal mol}^{-1}$). To rationalize the origin of the more pronounced complex stability, the mode of interaction between lead candidate and the G4 is observed with respect to the mode of interaction between BRACO-19 and the G4 (Figure 6). The simultaneous external stacking and groove binding of the lead candidate has visible stabilizing advantages over the external stacking of BRACO-19 in forming a complex with the receptor. In this light, it is important to note the extent of conformational flexibility of the core structures of the lead candidate and BRACO-19 (Figure 6). Flexible core scaffold of the lead candidate is an advantage relative to the rigid one of BRACO-19. The conformational flexibility of small molecules proved to be more

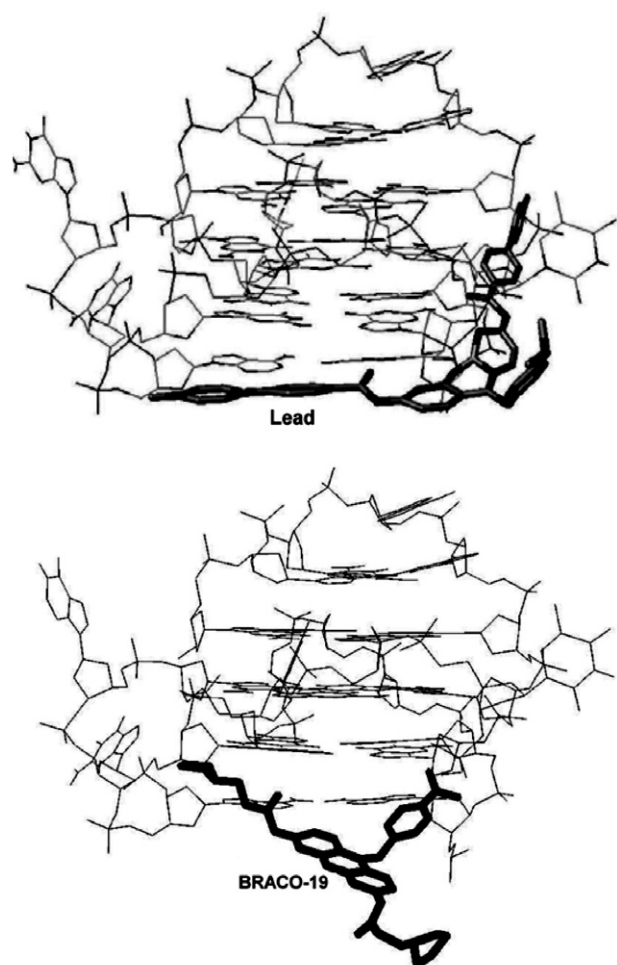


Figure 6. Interaction of lead candidate with G4 through external stacking and groove binding simultaneously versus external stacking of BRACO-19 with G4.

preferable compared to locking the ligands in a presumed bioactive G4 conformation.³⁸

The structural design of optimal groove/loop binders is a challenge, as this mode of interaction is nonspecific and dependent on the particular topology of groove/loop residues. A pure G-quartet-binding mode is hypothesized to be more stable than a multiple-binding mode – two ligands that are involved in external stacking and loop binding respectively.⁵ A likely reason for this is that a groove/loop-binding ligand induces loop rearrangement and perturbations to the interactions between the side chains of the other G-quartet-binding ligand and the loop/groove of G-quadruplex. There are indications that a multiple-binding mode increases conformational rigidity of G-quadruplex and decreases conformational flexibility of both G-quartets and backbone.⁵ It means that such a mode of interaction, which includes two ligands, would be thermodynamically unfavorable. The present proposal of lead candidate structure, being a G-quartet and groove/loop binder at the same time, is inclined to bypass this kind of glitch. This was accomplished using the following reasoning. Knowing that DNA-groove/ligand recognition is mainly driven by charge-induced phenomena,^{39–41} the lead candidate was made more prone than BRACO-19 to electrostatic interactions. While the core aromatic scaffold of BRACO-19 only has one hydrogen-bond acceptor (an N atom, Figure 2), the core aromatic scaffold of lead candidate has three hydrogen-bond acceptors (three N atoms, Figure 5). While two symmetric side chains of BRACO-19 have four HBAs (two N and two O atoms) and two HBDs (two NH groups, Figure 2), two symmetric side chains of lead candidate have six HBAs (four N and two O atoms) and two HBDs (two NH groups, Figure 5). To better conceive this aspect, the mode of interaction of the lead candidate with the receptor is illustrated in Figure 7 containing a molecular surface plot with standard atom colors. The structural basis is an advance in the development of effective ligand molecules that are able to block the interactions of G4 with proteins having G4-groove/loop as binding site. Taking into account both this point and the way in which

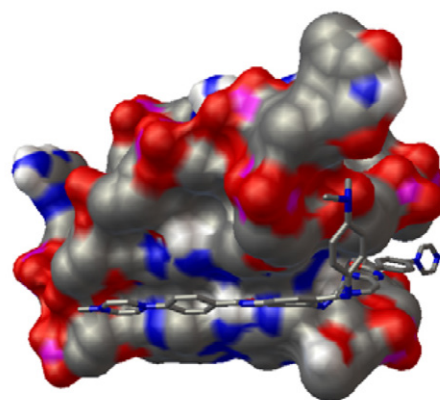


Figure 7. Lead candidate is proposed to interact through external stacking and groove binding with G4 simultaneously.

Table 2. Drug likeness of lead candidate according to Lipinski's Rule of Five^(a)

Ligand	Number of H-bond donors	Number of H-bond acceptors	Molecular weight (D)	Log P ^(b)
Lead	3	10	702.02	0
BRACO-19	3	6	550.01	2.35

^(a) The Rule of Five got its name from cut-off values that are five or a multiple of five. The rule states that poor absorption or permeation is more likely when: (i) a compound has more than 5 H-bond donors (sum of OHs and NHs), (ii) there are more than 10 H-bond acceptors (sum of Ns and Os), (iii) the molecular weight is over 500 Dalton and (iv) the Log P is over 5 (or MLog P is over 4.15).⁴³

^(b) Overall hydrophobicity is measured by the partition coefficient P . P is the water-octanol partition coefficient and is a measure of the equilibrium concentration of solute in octanol divided by the concentration of the same species in water. Log P is a measure of hydrophilicity/phobicity of a compound.

Table 3. Predictors of oral bioavailability of lead candidate according to Veber's rules^(a)

Ligand	Number of torsions	Polar surface area (Å ²)	Number of H-bond donors	Number of H-bond acceptors
Lead	11	60.74	3	10
BRACO-19	13	9.72	3	6

^(a) Based on measurements in rats for over 1,100 drug candidates, compounds that meet the following criteria may be associated with good oral bioavailability: (i) molecular flexibility reflected through 10 or fewer rotatable bonds – torsions, (ii) polar surface area equal to or less than 140 Å², and (iii) a total number of hydrogen-bond donors and acceptors equal to or less than 12.⁴⁴

the lead candidate structure was developed via analog design on top of HTVS, the lead candidate and BRACO-19 can be observed neither like clear structural nor like clear functional analogs. They should rather be placed in between structural and functional analogs.

Virtual screening is an effective method for reducing the initial number of potential candidates. A small molecule with binding affinity to increase the conformational flexibility of an apo (ligand-free) G4 through π - π stacking at the end of G4 can be conceivable as a unique, specific pharmacophore for designing novel lead candidate compounds by high-throughput virtual screening.^{6,10} The lead candidate (Figure 5), designed by this approach and aimed to target the c-Myc promoter G4 through external stacking and groove binding simultaneously, would have useful implications for overcoming the challenge of designing specific groove/loop binders. The challenge stems from the dependence of the groove/loop interaction mode on the particular arrangement of residues. Even though the pure quartet-binding mode is more stable than the groove/loop binding mode, groove/loop is a viable binding site that is of interest for the structure-based drug design. The use of grooves/loops offers distinct environments in order to gain specificity among many types of G4s by way of subtle variations of G4 topologies, groove widths, and loop sequences without affecting binding affinity.⁴²

The correlation between the structure of a drug candidate and its oral absorption is an important point of consideration when attempting to design novel anti-cancer therapeutics. Empirical recommendations predict drug likeness on the basis of the molecular structure of drug candidate and represent useful guide in drug design process. Lipinski's rule of five (see the footnotes of Table 2)⁴³

and Veber's rules (see the footnote of Table 3)⁴⁴ are used to evaluate the lead candidate with respect to BRACO-19. A close inspection of the data for the lead candidate reveals that molecular weight (MW) is only out of an expected range in Table 2, while both the number of torsions and the total number of hydrogen-bond donors and acceptors are essentially lined up with the upper bounds of suggested ranges in Table 3. The MW of lead candidate is not likely to affect its good absorption as the MW of the reference (BRACO-19), being a highly selective G4-binder that is widely available on the market, is out of range as well (Table 2). An important predictor is P – an indicator of hydrophobicity and hydrophilicity. A zero value of log P in Table 2 means that the lead candidate is equally hydrophobic and hydrophilic. This is well-correlated with the polar surface area of the lead candidate (60.74 Å² in Table 3) that is roughly in middle of the expected range. In contrast, a value of 2.35 for log P in Table 2 means that BRACO-19 is substantially (about 224 times) more hydrophobic than hydrophilic, so that its polar surface area (9.72 Å² in Table 3) is in a close vicinity of the lower bound of the expected range. These predictions substantiate our fundamental idea to design a lead molecule that is remarkably more susceptible to charge-induced interactions with the receptor (or more specific) than BRACO-19. Further investigations correlating oral bioavailability of the particular molecule in humans and simple molecular property-based rules may be required.⁴⁵

In creating a synthetic route for the development of a ligand molecule, it is necessary to create a molecular entity in which functional groups are correctly positioned in three-dimensional space; this will enable the creation of functional biophoric fragments such as the pharmacophore

ore. The lead candidate, proposed herein, does not have chiral centers (Figure 5) and it may be eventually synthesized using the small libraries of already-prepared (e.g. by means of a split-mix approach) structural fragments (analogues),⁴⁶ even though a potential disadvantage of synthetic libraries is their limited structural diversity. Its atomic composition (Figure 5), presumably, does not interfere with serious side effects. Future research is supposed to see into both pharmacokinetic/dynamic and toxicity profiles *in vitro/in vivo*. The need to know potential targets at the whole genome level means that global genome transcriptome profiling may help in the determination of which genes are affected by a rationally designed G4-interactive small molecule.⁴⁷ Consequently, the selectivity and potency of a new G4-preferred compound can be explored using *in vitro* cell assays and *in vivo* models. We believe that this report will inspire modern organic chemists and pharmacists to face new interesting challenges of vital importance with vigor.

4. Conclusions

It has been shown that high-throughput virtual screening in combination with analog design may be an efficient tool for predicting the chemical structure of lead candidate aimed at guiding further steps in a drug design and development process.

A substantial propensity of the lead candidate to stabilize G-quadruplex DNA from the c-Myc oncogene promoter region has been predicted by satisfying the structural and physicochemical requirements for optimal biological activity – by binding to the target through external stacking and groove binding simultaneously. This work has somewhat contributed to the better formulation of a daunting challenge – how to design small molecules that can bind selectively to each of the many possible G4 structures.

It is believed that the present *in silico* study has provided a fruitful ground for the upcoming investigations of pharmacokinetics/pharmacodynamics and toxicity properties *in vitro/in vivo*.

Supplementary Material

Results of the high-throughput virtual screening for hits against G-quadruplex DNA (PDB ID: 2A5P).

5. References

1. E. Ruggiero, S. N. Richter, *Nucl. Acids Res.* **2018**, *46*, 3270–3283. DOI:10.1093/nar/gky187
2. N. W. Luedtke, *Chimia* **2009**, *63*, 134–139. DOI:10.2533/chimia.2009.134
3. R. J. Harrison, J. Cuesta, G. Chessari, M. A. Read, S. K. Basra, A. P. Reszka, J. Morrell, S. M. Gowan, C. M. Incles, F. A. Tanius, W. D. Wilson, L. R. Kelland, S. Neidle, *J. Med. Chem.* **2003**, *46*, 4463–4476. DOI:10.1021/jm0308693
4. J. H. Tan, T. M. Ou, J. Q. Hou, Y. J. Lu, S. L. Huang, J. Y. Wu, Z. S. Huang, K. Y. Wong, L. Q. Gu, *J. Med. Chem.* **2009**, *52*, 2825–2835. DOI:10.1021/jm801600m
5. J. Q. Hou, S. B. Chen, J. H. Tan, H. B. Luo, D. Li, L. Q. Gu, Z. H. Huang, *J. Comput. Aided Mol. Des.* **2012**, *26*, 1355–1368. DOI:10.1007/s10822-012-9619-1
6. P. M. Mitrasinovic, *J. Biomol. Struct. Dyn.* **2018**, *36*, 2292–2302. DOI:10.1080/07391102.2017.1358670
7. D. J. Cashman, R. Buscaglia, M. W. Freyer, J. Dettler, L. H. Hurley, E. A. Lewis, *J. Mol. Model.* **2008**, *14*, 93–101. DOI:10.1007/s00894-007-0254-z
8. J. Dai, M. Carver, L. H. Hurley, D. Yang, *J. Am. Chem. Soc.* **2011**, *133*, 17673–17680. DOI:10.1021/ja205646q
9. B. J. Chen, Y. L. Wu, Y. Tanaka, W. Zhang, *Int. J. Biol. Sci.* **2014**, *10*, 1084–1096. DOI:10.7150/ijbs.10190
10. P. M. Mitrasinovic, *Croat. Chem. Acta* **2019**, *92(1)*, 43–57. DOI:10.5562/cca3456
11. A. T. Phan, V. Kuryavyi, H. Y. Gaw, D. J. Patel, *Nat. Chem. Biol.* **2005**, *1*, 167–173. DOI:10.1038/nchembio723
12. D. R. Koes, N. A. Pabon, X. Deng, M. A. Phillips, C. J. Camacho, *PLoS ONE* **2015**, *10(8)*, e0134697. DOI:10.1371/journal.pone.0134697
13. S. Y. Yang, *Drug Discov. Today* **2010**, *15*, 444–450. DOI:10.1016/j.drudis.2010.03.013
14. A. R. Leach, V. J. Gillet, R. A. Lewis, R. Taylor, *J. Med. Chem.* **2010**, *53*, 539–558. DOI:10.1021/jm900817u
15. J. Sunseri, D. R. Koes, *Nucl. Acids Res.* **2016**, *44*, W442–W448. DOI:10.1093/nar/gkw287
16. M. Hattori, N. Tanaka, M. Kanehisa, S. Goto, *Nucleic Acids Res.* **2010**, *38*, W652–W656. DOI:10.1093/nar/gkq367
17. M. Hattori, Y. Okuno, S. Goto, M. Kanehisa, *J. Am. Chem. Soc.* **2003**, *125*, 11853–11865. DOI:10.1021/jp972635z
18. M. Hattori, Y. Okuno, S. Goto, M. Kanehisa, *Genome Inform.* **2003**, *14*, 144–153. DOI:10.11234/gi1990.14.144
19. S. Forli, R. Huey, M. E. Pique, M. F. Sanner, D. S. Goodsell, A. J. Olson, *Nat. Protoc.* **2016**, *11(5)*, 905–919. DOI:10.1038/nprot.2016.051
20. G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew, A. J. Olson, *J. Comput. Chem.* **1998**, *19*, 1639–1662. DOI:10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B
21. G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell, A. J. Olson, *J. Comput. Chem.* **2009**, *30*, 2785–2791. DOI:10.1002/jcc.21256
22. H. J. Böhm, *J. Comput. Aided Mol. Des.* **1994**, *8*, 243–256. DOI:10.1007/BF00126743
23. P. M. Mitrasinovic, *J. Comp. Chem.* **2001**, *22*, 1387–1395. DOI:10.1002/jcc.1097
24. A. D. Becke, *J. Chem. Phys.* **2002**, *117*, 6935–6938. DOI:10.1063/1.1503772
25. P. M. Mitrasinovic, *Chem. Phys.* **2003**, *286*, 1–13. DOI:10.1016/S0301-0104(02)00902-3

26. A. Pavlov, P. M. Mitrasinovic, *Curr. Org. Chem.* **2010**, *14*, 129–138. DOI:10.2174/138527210790069866
27. S. Cosconati, S. Forli, A. L. Perryman, R. Harris, D. S. Goodsell, A. J. Olson, *Expert Opin. Drug Discov.* **2010**, *5*(6), 597–607. DOI:10.1517/17460441.2010.484460
28. B. V. Babu, N. K. Konduru, W. Nakanishi, S. Hayashi, P. M. Mitrasinovic, N. Ahmed, *Anti-Cancer Agents Med. Chem.* **2013**, *13*, 307–332. DOI:10.2174/1871520611313020017
29. P. M. Mitrasinovic, *Curr. Drug. Targ.* **2013**, *14*, 817–829. DOI:10.2174/1389450111314070009
30. P. M. Mitrasinovic, *Med. Chem.* **2014**, *10*, 252–270. DOI:10.2174/157340641003140304143442
31. P. M. Mitrasinovic, *Med. Chem.* **2014**, *10*, 46–58. DOI:10.2174/157340641001131226122124
32. N. Deng, L. Wickstrom, P. Cieplak, C. Lin, D. Yang, *J. Phys. Chem. B* **2017**, *121*(46), 10484–10497. DOI:10.1021/acs.jpcc.7b09406
33. S. Neidle, *Curr. Opin. Struct. Biol.* **2009**, *19*, 239–250. DOI:10.1016/j.sbi.2009.04.001
34. R. Deprez-Poulain, B. Deprez, *Curr. Top. Med. Chem.* **2004**, *4*, 569–580. DOI:10.2174/1568026043451168
35. M. Keseru, G. M. Makara, *Drug Discov. Today* **2006**, *11*, 741–748. DOI:10.1016/j.drudis.2006.06.016
36. D. Gramec, L. P. Masic, M. S. Dolenc, *Chem. Res. Toxicol.* **2014**, *27*, 1344–1358. DOI:10.1021/tx500134g
37. O. Uziel, E. Fenig, J. Nordenberg, E. Beery, H. Reshef, J. Sandbank, M. Birenbaum, M. Bakhanashvili, R. Yerushalmi, D. Luria, M. Lahav, *Br. J. Cancer* **2005**, *92*, 1881–1891. DOI:10.1038/sj.bjc.6602592
38. B. Prasad, J. Jamroskovic, S. Bhowmik, R. Kumar, T. Romell, N. Sabouri, E. Chorell, *Chem. Eur. J.* **2018**, *24*, 7926–7938. DOI:10.1002/chem.201800078
39. S. Neidle, *Nat. Prod. Rep.* **2001**, *18*, 291–309. DOI:10.1039/a705982e
40. P. Pandya, M. M. Islam, G. S. Kumar, B. Jayaram, S. Kumar, *J. Chem. Sci.* **2010**, *122*, 247–257. DOI:10.1007/s12039-010-0029-4
41. P. M. Mitrasinovic, *J. Chem. Inf. Model.* **2015**, *55*, 421–433. DOI:10.1021/ci5006965
42. S. Asamitsu, T. Bando, H. Sugiyama, *Chem. Eur. J.* **2019**, *25*, 417–430. DOI:10.1002/chem.201802691
43. C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney, *Adv. Drug Delivery Rev.* **2001**, *46*, 3–26. DOI:10.1016/S0169-409X(00)00129-0
44. D. F. Veber, S. R. Johnson, H. Y. Cheng, B. R. Smith, K. W. Ward, K. D. Kopple, *J. Med. Chem.* **2002**, *45*, 2615–2623. DOI:10.1021/jm020017n
45. T. Hou, J. Wang, W. Zhang, X. Xu, *J. Chem. Inf. Model.* **2007**, *47*, 460–463. DOI:10.1021/ci6003515
46. M. Tassinari, A. Lena, E. Butovskaya, V. Pirota, M. Nadai, M. Freccero, F. Doria, S. N. Richter, *Molecules* **2018**, *23*, 1874. DOI:10.3390/molecules23081874
47. C. Marchetti, K. G. Zyner, S. A. Ohnmacht, M. Robson, S. M. Haider, J. P. Morton, G. Marsico, T. Vo, S. Laughlin-Toth, A. A. Ahmed, G. Di Vita, I. Pazitna, M. Gunaratnam, R. J. Besser, A. C. G. Andrade, S. Diocou, J. A. Pike, D. Tannahill, R. B. Pedley, T. R. J. Evans, W. D. Wilson, S. Balasubramanian, S. Neidle, *J. Med. Chem.* **2018**, *61*, 2500–2517. DOI:10.1021/acs.jmedchem.7b01781

Povzetek

Ciljanje z gvaninom (G)-bogatih zaporedij DNA, zvitih v ne-kanonične strukture G-kvadrupleksov, z nizkomolekularnimi ligandi, predstavlja potencialno strategijo za gensko terapijo rakavih bolezni. BRACO-19 je bil nedavno potrjen kot edinstven (temodinamsko favoriziran in visoko selektiven) vezalec, ki je vključen v zunanji nalagalni način interakcije z G4-DNA, ki jo tvori promotorska regija c_myc onkogene (P. M. Mitrasinovic, *Croat. Chem. Acta* **2019**, *92*, 43–57). V tem članku smo identificirali spojine vodnice z uporabo visoko-zmogljivostnega virtualnega rešetanja. Izvedeno je bilo iskanje v podatkovni bazi Kyoto Encyclopedia of Genes and Genomes (KEGG) z uporabo ključnih farmakofornih značilnosti BRACO-19. Med skupno 29,009 vnosov je bilo izbranih 22 zadetkov z opazovanjem proste vezavne energije. Posledično so bile določene tri kandidatne spojine vodnice na podlagi strukturnih kriterijev. Na koncu je predlagana vodilna kandidatna struktura z uporabo analognega načrtovanja in ob upoštevanju fizikalno-kemijskih zahtev za optimalno biološko aktivnost in nabora farmakoloških vidikov. Obravnavan je tudi pomen študije za eksperimentalne raziskave.



Except when otherwise noted, articles in this journal are published under the terms and conditions of the Creative Commons Attribution 4.0 International License