Published in final edited form as:

Biochimie. 2018 September ; 152: 134–148. doi:10.1016/j.biochi.2018.06.024.

G-quadruplex Virtual Drug Screening: A Review

Robert C. Monsen^{#b} and John O. Trent^{#a,b,c,†}

^a)James Graham Brown Cancer Center, University of Louisville, Louisville, KY 40206 USA

^b)Department of Biochemistry and Molecular Biology, University of Louisville, Louisville, KY 40206 USA

^c)Department of Medicine, University of Louisville, Louisville, KY 40206 USA

[#] These authors contributed equally to this work.

Abstract

Over the past two decades biologists and bioinformaticians have unearthed substantial evidence supporting a role for G-quadruplexes as important mediators of biological processes. This includes telomere damage signaling, transcriptional activity, and splicing. Both their structural heterogeneity and their abundance in oncogene promoters makes them ideal targets for drug discovery. Currently, there are hundreds of deposited DNA and RNA quadruplex atomic structures which have allowed researchers to begin using *in silico* drug screening approaches to develop novel stabilizing ligands. Here we provide a review of the past decade of G-quadruplex virtual drug discovery approaches and campaigns. With this we introduce relevant virtual screening platforms followed by a discussion of best practices to assist future G4 VS campaigns.

Keywords

G-quadruplex; virtual screening; docking; drug discovery

1 Introduction

G-quadruplexes (G4s) are secondary structures which occur in both DNA and RNA under physiologically relevant conditions [2]. G4s contain 2 or more stacks of 4 coplanar guanine residues stabilized via Hoogsteen hydrogen bonding. The stacking interaction is also facilitated by monovalent cations, such as sodium and potassium, as well as π -stacking of the purine bases (Figures 1 and 15) [3]. Although it is unclear what promotes G4 formation *in vivo*, they are increasingly implicated in important biological events such as telomere maintenance, transcription regulation, mRNA translation, and replication [2, 4-9]. More

[†]Corresponding author: John O. Trent (john.trent@louisville.edu).

Publisher's Disclaimer: 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 citable 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.

Conflicts of Interest

The authors report no conflicts of interest.

recently, chromatin immunoprecipitation and high through-put sequencing analyses have provided *in vivo* evidence for the presence of ~9,000 non-telomeric G-quadruplexes that reside in nucleosome-depleted promoter regions, confirming many of the previously proposed regulatory G4s [10, 11]. Thus, G-quadruplexes appear to be excellent targets for anti-cancer therapeutics [7].

Currently there are greater than 1,000 characterized G-quadruplex stabilizing ligands which have been discovered through virtual screening (VS), traditional high-throughput screening (HTS), and plenty of serendipity (see: <http://g4ldb.org/> for a listing of many verified G4 ligands). Although many of these compounds (TMPyP4, pyridostatin, telomestatin, BRACO-19, etc.) bind with high affinity to G4s, it is often by an end-pasting mechanism and, therefore, non-specific. Furthermore, these compounds commonly do not possess drug-like properties, e.g. they do not pass Lipinski's rule of five [12], nor do they have documented ADMET (absorption, distribution, metabolism, excretion, and toxicity) profiles [13]. Extensive work has gone into modifying general end-pasting drug scaffolds such as porphyrins [14-17], phenanthrolines [18-21], anthracenes [22, 23], naphthalenes [24], and quinolones [25, 26] (Figure 2A-E). Only Quarfloxin (CX-3543) (Figure 3A), an end-paster, has progressed to clinical trials [27]. Alternative G4 drug discovery strategies have focused on developing ligands that target the loops and grooves. An example groove-binding ligand is distamycin A (Figure 3B) which was shown by Randazzo et al. to interact with the grooves of the parallel tetramolecular quadruplex [d(TGGGGT)₄] by ¹H-NMR (proton nuclear magnetic resonance spectroscopy) studies [28, 29]. Unfortunately, most G4 groove-binding ligands have poor selectivity over dsDNA (double-stranded DNA), which was the case for distamycin A and netropsin [30] (Figure 3C). To address this selectivity problem, many researchers have turned to VS drug discovery methods.

VS strategies have been building momentum in G4 drug discovery both as a low-cost enrichment step and as a lead development step in the discovery pipeline, which our laboratory has previously discussed [31]. Whereas traditional HTS methods rely on obtaining and screening hundreds or thousands of compounds from curated libraries, VS simply requires knowledge of known ligand structures (for similarity and pharmacophore searches) or a receptor structure to which a library of virtual compounds can be docked. These methods are known as ligand-based or receptor-based drug discovery, respectively. Ligand-based methods use an identified set of known active ligands to search a database for compounds which have similar properties. These techniques operate under the assumption that ligands with a similar 2D or 3D structure will offer similar interactions with their targets. Conversely, receptor-based methods screen virtual libraries against a target structure, and so require an X-ray crystallographic, NMR, or homology-derived 3D atomistic model of the target. These coordinate files can be downloaded from databases such as the Protein Data Bank (PDB) (>133,000) or the Nucleic Acid Database (NAD) (>900), which are continuously being updated with new structures.

VS platforms have been extensively used in ligand discovery [32, 33], however, until now there has not been an assessment of strategies specifically targeting G4s. Here we briefly discuss some of the common screening strategies, such as docking and pharmacophore screening, as well as relevant aspects including: library preparation, scoring, and analysis.

This is followed with a commentary on suggested best practices for *in silico* G4 drug discovery based on the authors' own experience and knowledge gleaned from successful campaigns.

2 Pharmacophore & Similarity Based Screening

Ligand-based methods such as pharmacophore and similarity search platforms are widely used and often integrated into a VS docking campaign pre- and/or post-docking (Section 4). The term 'pharmacophore' as we use it refers to an abstract, 3D physio-chemical representation of the chemical moieties necessary for ligand-receptor interaction. Pharmacophore screens use multiple ligands of the same binding site to derive an ensemble of chemical features necessary for an ideal interaction (i.e. hydrogen bond donor/acceptor, aromatic ring elements, cations, anions, etc.). The resulting model is known as a hypothesis. These hypotheses, which are 3D chemical descriptors, are then used to screen a virtual library to find "pharmacophore-similars" that satisfy the hypothesis [34]. The result is a list of compounds which are ranked for their probability of favorable interactions based on their physical and chemical similarity to the initial query structure. Various pharmacophore search platforms are available such as Pharmer (ZINC) [35], Discovery Studio's 3D-QSAR module (Accelrys) [36], LigandScout (Inteligand) [37], MOE (Chemical Computing Group) [38], Phase (Schrödinger) [39], SYBYL-X2.1.1 (Certera) [40], and Pharao (Silicos) [41].

An example of a successful G4 pharmacophore screening campaign comes from Chen et al. [42] in which the authors used Discovery Studio's 3D-QSAR pharmacophore generation module to construct a model based on acridine derivatives. By weighting hydrophobic interactions higher than aromatic interactions in the hypothesis the authors enriched for compounds with scaffolds unlike the acridines. This was achieved by screening their own in-house library. The resulting compound was a triaryl-substituted imidazole derivative (Figure 4A) which had a K_d of 0.5 μ M against a human telomere G4 and displayed selectivity over dsDNA based on circular dichroism (CD) and fluorescence melting experiments. Interestingly, this compound is very similar to the triarylpyridines discovered previously [43] (Figure 4B).

The second and most rapid ligand-based strategy is known as a structural similarity search. These platforms require only knowledge of active ligand's chemical composition (i.e. a chemical structure). In the past, this approach utilized a rigid-body alignment approach using 2D (2-dimensional) and 3D chemical fingerprints to align and rank each molecule. This was enhanced with the advent of semi-flexible and flexible superposition algorithms that allow for a more comprehensive search in 3D space by ranking each molecule based on the volume overlap within the query structure. See [44] for more in-depth discussion.

The structural similarity software vROCS (OpenEye) [45] has been utilized by Musumeci et al. [46] to screen the Maybridge [47] HitFinder database (~14,400 compounds) using Distamycin A (Figure 3B) as a query. Using the Tanimoto coefficient (Section 6.4) and vROCS's colour scoring (atom/feature similarity) criteria the authors discovered a set of novel G-quadruplex groove-binding ligands (Figure 5A-C). These ligands bound with higher affinity to the grooves of human telomeric quadruplexes over dsDNA (detected by UV-Vis,

fluorescence, and oligo affinity support analysis [46]) but had no observable melting temperature (T_m) shift. It was also shown that 3 of the 7 compounds induced a DNA damage response at the telomeres, further confirming their G4 binding activity. While this isn't the first reported campaign using vROCS in G4 drug discovery [48] it is a proof-of-concept that this relatively straight forward lead-discovery approach can enrich for novel scaffolds which interact in a favorable manner.

3 Libraries

Arguably the most important consideration in virtual screening methodologies is the selection of compound library. VS libraries contain hundreds to millions of virtual compounds that will inevitably dictate the scaffold diversity of resultant hits. The benefit of using large, diverse libraries is the expanded chemical search space. Fortunately, there are many large libraries available: MayBridge [47], AnalytiCon [49], ZINC [50], ChemDiv [51], SPECS [52], Mcule [53], eMolecules [54], PubChem [55], Life Chemicals [56], ChemBridge [57]. Some databases, such as the ZINC database, offer sub-libraries for a more tailored search (e.g. lead-like, fragment-like, drug-like, and natural products) which often contain readily purchasable or synthesized compounds. Conversely, some researchers choose to develop their own curated libraries [42, 58] which can be beneficial when the user has limited computing resources available.

The biggest challenge is finding the optimal balance among speed, accuracy, and library composition. A library of ~1 million compounds docked to a single receptor will take as little as weeks to as much as a year of computing time with a single workstation using a rigorous algorithm. Many researchers have circumvented this by reducing libraries to smaller, more manageable subsets which only contain compounds that conform to a predefined criterion. Specifically, using a shape-based or pharmacophore search on a library one can significantly reduce the size and enrich for chemical moieties that are well suited to the system of interest (see [42, 46, 48, 59-61]). However, limiting searches to a pre-defined chemical search space introduces significant bias and is bound to limit compound diversity.

Alternatively, increased computing power by use of a research cluster or computing grid can greatly reduce the computational time required for a screening campaign of >1 million compounds [62]. The authors have had success using grid computing which can dock as many as ~25 million compounds in just a few days to a single receptor site [31]. While grid computing is becoming more commonplace in research institutes, not everyone has access to large scale grids. Therefore, care must be taken if curating a library to be docked at a smaller scale. Hand picking small subsets of compounds can lead to significant bias (Section 6), or worse, no enrichment of meaningful hits [63].

4 Docking

Docking has been in use since the early 80's and has gained traction commensurate to the number of published protein and nucleic acid structures since [64]. In general, docking seeks to use the physical and chemical information provided by an atomistic receptor to dock whole or fragmented molecules from a library and rank them using a scoring function. Each

docking platform has its own algorithm as well as flavor of scoring function, which has made cross-platform comparisons difficult [65-67]. The lack of convergence onto any one platform is likely due to the unique features inherent to each, such as: cost, speed, scoring terms, ease-of-use, scalability, receptor flexibility, ligand flexibility, and the option of implementing molecular dynamics (MD) force fields.

There are multiple docking platforms suitable for use with nucleic acid receptors. These include: DOCK v4-6 (UCSF) [68-70], AutoDock (Scripps) [71], AutoDock Vina (Scripps) [72], GOLD (Cambridge Crystallographic Data Centre) [73], Surflex-DOCK (BioPharmics) [74], Glide (Schrödinger) [75], and ICM (Molsoft) [76]. Many of these software have been compared elsewhere in the context of protein docking [77]. The author has also compared two of these platforms, Surflex-DOCK and AutoDock, in the context of nucleic acids [78]. Both platforms performed equally well with Surflex being slightly faster and more easily scalable. DOCK, AutoDock, AutoDock Vina, and GOLD are all freely available to academic institutions. Each docking platform varies with respect to sampling algorithms and scoring functions.

A sampling algorithm is a systematic way to sample from a population of possible molecular conformations and binding modes without exhausting all possibilities. The primary hurdle in docking is the vast number of potential docked positions for a given set of molecules. Minimizing the computational time necessary for each docking run is of prime importance for high-throughput screening. Strategies to minimize computational time include: library optimization (reduced size, generation of tautomers, protonation, filtering), robust computational infrastructure (computing grids), and selection of the appropriate sampling algorithm(s). Such sampling algorithms include: geometric matching algorithms (GM), incremental construction methods (IC), Monte Carlo (MC) searches, genetic algorithms (GA), and molecular dynamics (MD) (see ref. [79] for an overview). Table 1 lists the various algorithms employed by the docking platforms discussed here.

4.1 Geometric Matching

The first sampling algorithm, which is used in versions 4-6 of DOCK [68], shares characteristics with similarity or pharmacophore search algorithms. DOCK (v4-6) uses a geometric matching (GM) algorithm to place fragments into the receptor. In this algorithm the receptor is treated as a rigid object in which flexible ligands are docked. The receptor is defined by a set of overlapping spheres, while each ligand is defined by rigid segments whose conformation can be optimized within the user-defined binding site. The ligand 'flexibility' comes from an anchor-and-grow algorithm which uses the molecule's rotatable bonds to partition it into rigid segments. Initial docked segments are deemed 'anchors', and from these anchors the remainder of the molecule is appended, followed by optimization and scoring [69]. By mapping each molecule into the active site of a receptor, GMs have the advantage of being very rapid techniques and well suited for large database screening [69, 77]. The recent release of DOCK v6 has added features that allow much more versatility to both the docking and scoring functions. Most importantly it has incorporated MD simulation capabilities [70], validated using a set of RNA-ligand complexes.

Park and Kang used DOCK (v5.4) in conjunction with the UNITY-3D pharmacophore platform to identify three compounds (Figure 6A-C) that stabilize the *c-myc* G-quadruplex [59]. The authors filtered 560,000 compounds from publicly accessible databases, ChemDiv [51] and SPECS [52], based on a query generated in UNITY-3D. The resulting set of compounds were energy minimized and then docked into an NMR-derived *c-myc* quadruplex (PDB ID: 2A5R). The authors optimized this receptor by changing inosine bases back to their original wild type bases followed by short energy minimization. After docking and scoring, each compound was re-scored using a Generalized Born solvent accessible surface area (GBSA) scoring function to account for solvation. Interestingly, the top three compounds showed little or no thermal stabilization based on Förster resonance energy transfer (FRET) screening, but were diverse in structure, and had polymerase stalling ability as well as *in vivo* activity in Ramos, CA46, and HeLa cell lines.

4.2 Incremental Construction

The incremental construction (IC) docking approach fragments the ligand where it has rotatable bonds and systematically docks each fragment. This allows for very rapid flexible ligand docking and is employed by DOCK (v4+) and Surflex-Dock. The Surflex-Dock approach, which is an adaptation of the Hammerhead docking procedure [80], places head fragments from each ligand into the receptor site and aligns them to 'probe' atoms. These probes are predefined idealized representations of favorable interactions. After placement, each head fragment is scored, and the top scoring fragments are retained. It next aligns the tail fragments to the head fragments and adjacent probes and scores them. In this way there is a drastic reduction in computation time by only following up with a small portion of possible conformations [78].

As an example of successful IC docking, Hou et al. identified a novel *c-myc* stabilizing ligand with the Surflex-Dock platform [81]. Using the NMR derived quadruplex (PDB ID: 1XAV) the authors docked 28,530 compounds from the ChemBridge [57] database, which was filtered for compounds containing 3 aromatic rings. These compounds were subsequently redocked to a duplex DNA structure (PDB ID: 1Z3F) and scored based on intercalation. A third round of docking and scoring was performed on each compound, this time in the groove(s) of a duplex DNA (PDB ID: 1K2Z). Compounds with a score ratio of >1.0 (G4 score/ dsDNA score) and >1.1 (G4 score/ Groove score) were chosen. Although the resultant top hit, a pyrrolopyrazine derivative (Figure 7A) was less effective than the control compound SYUIQ-5 (Figure 7B) in luciferase assays, it was much more selective for the G-quadruplex over dsDNA as determined by surface plasmon resonance assays. Furthermore, there are no similar reported scaffolds in the G4 ligand database, indicating that this is a novel quadruplex stabilizing ligand.

4.3 Stochastic Sampling

ICM, AutoDock Vina, AutoDock, Glide, Surflex-Dock, and GOLD are platforms that incorporate the stochastic algorithms: Monte Carlo (MC) and genetic algorithms (GA). Stochastic sampling algorithms iteratively generate new molecular conformations to be placed and scored using random movements (MC) or 'mutations' and 'selection' (GA). Although stochastic algorithms can be computationally more expensive than GM or IC

methods alone [78], they have traditionally out-performed in reproducing poses of ligands co-crystallized with their receptors [77].

4.3.1 Monte Carlo—In MC algorithms, each ligand's initial conformation is altered through random steps of bond rotation, rigid-body translation, or rotation, and subsequently scored until a pre-defined number of steps have been reached. At each step, the score is assessed based on steric conflict which is followed by an empirical potential calculation (Section 5). If this new step has improved the score sufficiently, the molecule's configuration will be saved and used in another iteration of random conformational sampling [82]. MC sampling is used in ICM, Glide, AutoDock Vina, and earlier versions of AutoDock.

ICM has previously been shown to perform exceedingly well at reproducing the correctly docked conformation of ligands to protein receptors over DOCK, AutoDock 3.0, and GOLD [83]. In 2010 Lee et al. [84] used ICM-Pro to screen a natural products database (AnalytiCon [49]) of 20,000 compounds against the *c-myc* nuclear hypersensitivity element III₁ (NHE III₁) G4 which was modified from a human telomeric quadruplex structure (PDB ID: 1KF1). Testing the top 5 scoring compounds in polymerase stop assays resulted in the discovery of fonsecin B (Figure 8A), a naphthopyrone pigment, which at the time of discovery was a novel scaffold. Later, Chan et al. [85], using the same modified receptor (PDB ID: 1KF1) and docking platform (ICM-Pro), screened 3,000 compounds from a library of FDA approved drugs. This led to the identification of methylene blue (Figure 8B), a phenothiazinium derivative. Methylene blue is already known to be a dsDNA intercalator and is likely a G4 end-paster. Thus, the authors modified this scaffold with side chains to improve selectivity. Interestingly, one derivative (Figure 8C) had higher affinity for the *c-myc* quadruplex both *in vivo* (luciferase assays, MTT proliferation assays) as well as *in vitro* (fluorescent intercalator displacement assay, PCR stop assay, mass spec, UV-vis). The Ma group later applied the same approach (ICM-Pro targeting *c-myc* PDB ID: 1KF1) [86] to screen a natural product-like database of 20,000 compounds to identify potential groove-binding scaffolds by limiting their search space to the grooves. This screen resulted in a compound containing carbamide, diphenyl ether, and tetracyclic moieties (Figure 8D), which is a unique G4 scaffold. NMR titration and re-docking were then used to show that the ligand is a *de facto* groove-binder. Whether this ligand is specific for G-quadruplexes over dsDNA has yet to be determined.

Another MC software which offers speed, a user-friendly interface, and great reproducibility of co-crystallized conformations [87] is AutoDock Vina. Vina was used by Alcaro et al. [60] in 2013 as a final step in their screening pipeline where they discovered a psoralen derivative (Figure 9). Psoralens have long been known as DNA intercalators; however, this is the first reported instance of psoralens as G-quadruplex stabilizers. Their initial screening began with the ZINC library of >2.7 million compounds, which were filtered down to ~4,000 compounds using 7 query structures in shape-based ROCS (Rapid Overlay of Chemical Structures [88]) and 2D fingerprint filter MACCS (Molecular ACCess System – MDL Information Systems inc.). This was followed by the removal of inorganic components, adjusting pH to 7.4, energy minimizing the structures, and finally, removing compounds that have a similarity of less than 0.7 Tanimoto coefficient (Section 6.4). Altogether, ~7,000 compounds were docked in AutoDock Vina using ensemble docking against the human

telomere quadruplexes (PDB ID: 143D, 1KF1, 2HY9, and 2JPZ). This resulted in 904 compounds which were clustering to obtain 28 compounds for testing, resulting in the psoralen.

4.3.2 Genetic Algorithm—GA sampling uses a molecule's location, orientation, and conformation to specify the state of a random population of individuals. These individuals have a genotype (the ligand's states) and a phenotype (atomic coordinates) and the ligand's overall fitness is equivalent to its interaction energy with the receptor. Individuals from the initial top scoring population are iteratively “mated” and have offspring which gain random mutations (state changes) as well as inherit genes (states) from both parents, known as “crossover”. Selection of each offspring for subsequent mating occurs based on the individual's fitness score [89]. Algorithms such as these can be computationally expensive, but have traditionally performed well at reproducing known ligand orientations in active sites [77]. GOLD and AutoDock (v3.0+) use this type of sampling method.

Kaserer and colleagues used GOLD [48] in parallel with the structural similarity search ROCS [45] and the pharmacophore search LigandScout [37] to find consensus hits between the three techniques. The pharmacophore models were generated based on the human telomere quadruplexes in complex with naphthalene diimide derivative BMSG-SH-3 (PDB ID: 3SC8), naphthalene diimide derivative MM41 (PDB ID: 3UYH), and berberine (PDB ID: 3R6R). Overall, they found 252 unique hits from the <http://Specs.net> [52] database. Next, using vROCS they selected the 9 best-performing shape-based models which were based on queries derived from co-crystallized ligands (PDB IDs: 3UYH, 3SC8, 3R6R) or from the energy-minimized ligand. Using an Implicit Mills-Dean force field, with additional weighting for aromatic interactions, they found 2620 hits. Last, the authors selected the human telomere quadruplex (PDB ID: 3CE5) to directly dock the Specs library. From this screen, the top 10 ranked molecules were selected. In total from the three techniques, 5 consensus compounds and 30 other top scoring compounds were tested, plus some derivatives. Overall, they found 14 ligands (Figure 10) that were active and had affinities that compared well with other contemporary VS screening approaches [42, 60, 90, 91]. This tour de force campaign demonstrated that a combined approach with cross validation can significantly enrich for *real* hits, although it did not produce much scaffold diversity.

Autodock (v4.2) remains a powerful tool in identifying G4 groove-binding ligands. In 2009 Cosconati et al. [92] used Autodock to screen the Life Chemicals database of ~6,000 compounds against the tetramolecular, parallel G4 sequence [d(TGGGGT)]₄ (PDB ID: 1S45) using a grid enveloping just one of the identical grooves. The compounds were scored and selected based on visual inspection. Specifically, compounds unable to form H-bonds with guanine bases or establish electrostatic interactions with the backbone phosphates were removed. Thirty top-scoring compounds were selected and used in NMR titrations which resulted in an impressive 6 out of 30 interacting as groove-binders. This was followed up with a more in-depth investigation by Trotta et al. [93] showing that 3 of these compounds (Figure 11A-C) bind with higher affinity to the grooves of [d(TGGGGT)]₄ than distamycin A using ITC (isothermal titration calorimetry) and NMR. Similarly, Di Leva [94] used Autodock to screen ~19,000 compounds from the ChemDiv database against the 24nt human telomere quadruplex (PDB ID: 2GKU). Out of the 18 compounds tested, one (Figure

11D) showed significant thermal stabilization and appeared to interact as a groove-binder based on NMR and re-docking experiments. The identified benzylpiperidine-containing compound was also shown to cause telomere damage in three cancer lines (HeLa, U2OS, HT29) but not a normal fibroblast line (BJ-hTERT). Subsequently, Amato et al. [95] used Autodock to screen ~59,000 compounds from the Molecule database against a G-triplex structure (PDB ID: 2MKM), an apparent intermediate state in the G-quadruplex folding pathway [96]. 15 compounds were selected for purchase but only 1 (Figure 11E) had significant stabilizing ability. Although this compound did not distinguish G-quadruplex from G-triplex, it did have selectivity for the higher order structures over dsDNA. Thus, these studies are undeniably a testament to Autodock's ability to successfully enrich for compounds which target the grooves of G-quadruplexes.

4.3.3 Molecular Dynamics Combined Approach—MD is primarily associated with simulations of molecular and macromolecular systems but is also applied to other modeling techniques such as docking. MD has long been used as a method to simulate structural changes and molecular interactions at the resolution of atoms using force fields [97]. Force fields are the equations which are solved to determine the potential of a given system and are necessary to determine the force acting on each atom. Once a force is determined, Newton's laws of motion can dictate the new atomic position [98]. The forces, therefore, must consider each atom's charge, bond length, and angle relative to all other atoms in each system. Thus, docking that utilizes MD allows for the ultimate amount of flexibility of ligand and receptor, resulting in efficient local optimization of docked ligands [77, 79]. Unfortunately, this comes with a high computational cost [99], and so is typically only used as a post-docking refinement step or in estimations of binding free energy [100].

Glide incorporates both MC and MD in its algorithm and performs well relative to other flexible algorithms (GOLD, ICM) in protein docking [77]. Glide (grid-based ligand docking with energetics) docks in essentially two stages: (1) each ligand is passed through hierarchical filters which evaluate spatial fit and complementarity of ligand-receptor interactions and, (2) poses that pass the initial screen are subjected to MD minimization based on the OPLS-AA force field (optimized potentials for liquid simulations – all atom force field) [101]. Kar and colleagues [61] applied Glide (v5.7) SP (standard precision) mode to dock 14,400 molecules from the Maybridge [47] database, followed by re-docking with the more extensive XP (extra precision) mode to the human telomere quadruplex (PDB ID: 2ld8). A docking site was not selected, rather, the authors constructed a grid encompassing the entire quadruplex. Two G4 ligands (Figure 12A, B) were selected from this screen and were shown to have moderately low affinities (K_d of 31 and 137 μ M), as measured by fluorescence titrations, but had selectivity over GC-rich dsDNA.

In a mixed pharmacophore/docking approach, Rocca et al. [102] used the pharmacophore screen LigandScout [37] to generate hypotheses based on 9 ligands known to bind DNA and RNA G4s. The ligand conformations for hypothesis derivation were extracted from top-ranked docked positions in the human telomere or TERRA (TElomeric Repeat-containing RNA) quadruplexes (PDB IDs: 3CE5, 2KBP). 257,000 natural product compounds from the ZINC [50] database were minimized and ionized to pH 7.4 in Maestro's Ligprep module (Schrödinger) [103] before being subjected to pharmacophore screening. The compounds

were subsequently filtered based on Lipinski's rule of five. The resulting compounds (~12,000) were clustered and then subjected to Glide's ensemble docking and scoring. Testing of the top 20 scored compounds resulted in 1 ligand (Figure 13) which showed interaction *in vitro* as determined by CD, FRET melting, and mass spectrometry. However, the compound was reportedly a naphthyridine derivative, a class which has previously been reported to interact with telomeric G4s and inhibit telomerase [104].

Bhat and colleagues [105] have put forward a robust VS workflow to derive novel ligands targeting the *c-myc* NHEIII₁ quadruplex (PDB ID: 2A5P). The steps are as follows: (1) the Maybridge database [47] (~55,000 compounds) was imported into Maestro's Ligprep [103] program, which generates all protonation states, conformations, and tautomeric structures for a given pH (~1.5 million compounds); (2) multiple stages of refinement for conformational restraints, conformational groups, and Lipinski's rule of five (~88,000 compounds); (3) Glide docking and re-docking to the 5' end of the quadruplex using all three modes: HTVS (high-throughput virtual screening), SP, and XP. This campaign resulted in three compounds which were chosen for testing, and one, a carbamoylpiperidinium-containing compound (Figure 14), stabilized the *c-myc* G4 by an end-pasting mechanism. A biological response was also observed in cells by luciferase expression assays as well as the induction of apoptosis selectively in T47D cancer cells, but not normal NKE cells.

Stand-alone MD simulations have also been used to study the interactions of known ligands with their receptors. These simulations have inherent advantage over traditional docking in that they can explicitly model solvent contributions. Not only does MD allow for calculation of relative binding free energies but it can also estimate the k_{on} and k_{off} rate constants. The latter has been difficult to assess due to the long timescale simulations needed for the ligand to come back 'on' to the receptor. This has been addressed with biased force fields in what is known as funnel-metadynamics [106]. This technique has been applied by Moraca et al. [100] to accurately calculate the free energy of binding of the ligand berberine to the human telomere sequence (PDB ID: 3R6R). Steady state fluorescence measurements were made to determine the actual free energy of $G = -9.8$ kcal/mol which compares well with the calculated $G = -10.3$ kcal/mol. Techniques such as this will likely play a major role in virtual lead development in the future.

5 Scoring Functions

Docking algorithms attempt to find 'solutions' to the orientation and ranking of ligand-receptor interactions. In doing so they must have a way to order the thousands or millions of complexes. This is achieved by scoring, which approximates the binding affinity (G_{bind}). Relative binding free energies can be approximated by free energy perturbation methods using molecular dynamics simulations [99]; however, these methods are far too computationally expensive for routine docking, and so more approximate solutions have been devised.

The first type of free energy approximation is the "empirical" [107] scoring function, which is an additive equation derived from each of the different modes of interaction of the system

[101, 108]. As implied by the name, empirical score values are derived from a set of known ligand-receptor complexes. As an example (as adapted from [101]):

$$\Delta G_{bind} = \Delta G_{hb} + \Delta G_{ionic} + \Delta G_{rot} + \Delta G_{vdw} \quad (1)$$

where G_{bind} would be the total docking score based on the additive scores from H-bonds (*hb*), ionic interactions (*ionic*), rotational constraints of constituent groups (*rot*), and Van der Waals (*VDW*) interactions. These terms can also be modified by the user with weighting to favor or disfavor interactions depending on the system in question. Similarly, there are modifier (or “penalty”) terms which can be applied to disfavor improper H-bond angles, distance restraints, hydrophobic interactions, and torsions. Autodock 4, DOCK v4-6, GOLD, Surflex-Dock, and Autodock Vina use empirical scoring terms.

There are also force-field (FF) based scoring functions. These functions implement current molecular mechanics (MM) force fields (e.g. AMBER, CHARMM) to estimate enthalpy of binding from VDW and electrostatic interactions, strain energies, and solvation effects. The latter is typically estimated by calculating the desolvation energy using MM/PBSA (Poisson-Boltzmann surface area) or MM/GBSA (generalized Born surface area) methods. However, MM/PBSA and MM/GBSA are too computationally expensive to be used in high throughput screening [109]. FF scoring is achieved by pair-wise evaluation of each non-bonded interaction, with the following general format (example taken from Autodock v4.2’s manual [110]):

$$\Delta G = \left(V_{bound}^{L-L} - V_{unbound}^{L-L} \right) + \left(V_{bound}^{P-P} - V_{unbound}^{P-P} \right) + \left(V_{bound}^{P-L} - V_{unbound}^{P-L} + \Delta S_{conf} \right) \quad (2)$$

where L is the ligand, P is the receptor, and V is the calculated potential term from MD force fields. Eq. (2) shows the 6 pair-wise evaluations and entropy term to account for any changes in conformational entropy. The force field potentials used here are comparable to that used in the Amber, CHARMM, or GROMACS force fields but can be modified by the user if desired. Glide, ICM, and early versions of DOCK and Autodock use FF based scoring functions with empirical weighting.

6 Discussion

Virtual screening approaches in the discovery of new G-quadruplex ligands have clearly shown promise. Higher throughput computational screens are allowing for more comprehensive searches of chemical space. As workstation computing power increases and more researchers gain access to resources such as computing clusters, we expect to see the number of successful VS screening campaigns increase. While some campaigns described here have proven the utility of virtual drug discovery methods, the methodologies and pitfalls are worth discussing.

6.1 Receptors

G-quadruplex receptors can be downloaded from one of the various databases (Protein Data Bank, Nucleic Acid Data Bank, etc.) or modeled. NMR solution structures should be used if available or the quadruplex modeled based on similar NMR coordinates. X-ray crystallography derived structures are prone to conformational bias and may be artificial due to the packing environment [111]. Thus, X-ray structures require pre-treatment with modeling techniques, such as short MD simulations with energy-minimizations [111]. Often, modified bases such as inosine are used to select for a single conformation in NMR or X-ray crystallographic techniques. These should be modified, as described above [59], by swapping the inosine with their natural residue and carefully energy-minimizing the structure prior to docking. A second major concern is loop flexibility. Loops have inherently high mobility, and this is rarely accounted for in traditional rigid receptor docking. This can be addressed with short MD simulations to allow the loops to search conformational space. Multiple conformations from MD or NMR PDB files can then be used in ensemble docking [60, 102], which is an excellent tactic for highly flexible receptors. These approaches are sufficient for most single G4 systems such as the *c-myc* or 22-24 nucleotide long telomere G-quadruplexes. However, targeting non-canonical G4s [112], or large, multi-G4 complexes [113, 114] with unknown structure can be challenging.

Multi-G4 systems have the interesting characteristic of large loop domains which span G4-G4 junctions. In theory, these loop-G4 pockets (Figure 15) can potentially serve as highly specific binding sites, much like that of enzyme substrate pockets. Unfortunately, determination of large nucleic acid secondary structures with traditional techniques is difficult. Nucleic acids rarely exist as a homogenous population in solution, making them difficult to characterize. Base substitution of non-tetrad guanine with inosine has often been used in NMR experiments to elucidate small G4 structures, but larger systems are not amenable due to spectral overlap and low proton density. Large DNA and RNA complexes are also difficult to crystallize. Even when crystallization is achieved, the packing environment can promote the formation of unrepresentative structures or features that are absent or only present in a small minority of the molecules in solution [111]. Thus, low resolution techniques (small angle X-ray/neutron scattering, analytical ultracentrifugation, dynamic light scattering, and CD spectroscopy) paired with MD simulations are now commonly used to develop structures currently unobtainable with traditional techniques [112, 113, 115, 116].

6.2 Libraries

When docking any library, no matter the size, there will always be “top scoring” compounds. Thus, many campaigns often use theoretical validation, such as re-docking, complimentary screening, or receiver operating characteristic [117] analyses for assessing the ability of their screen to detect real hits [48, 59, 61, 85, 118]. This is particularly important when working with small databases and can help to minimize false-positives. Conversely, docking known ligands as positive controls or simply increasing the library size can help to identify *real* high scoring compounds. Keep in mind that the use of in-house curated libraries, small drug libraries, or filtered libraries can impose serious limitations on the potential for identifying unique scaffolds. In fact, this appears to be the case for some of the campaigns mentioned

here [42, 60, 81, 85, 92, 102]. Three of these reports limited their chemical search space by filtering larger databases and one searched a database with only 3,000 compounds. In every instance the resulting hits were already known to bind nucleic acids. Conversely, focused libraries can be useful when searching for lead compounds based on validated hits, which was the case for Musumeci et al. [46] who used a similarity search to enrich for groove-binders based on a well characterized groove-binding ligand.

Before using a library, the compounds typically require optimization. Fortunately, many of the virtual screening databases have pre-optimized ligands for screening. Optimization ensures that each ligand has been desalted, neutralized, energy-minimized, and correctly protonated before docking [12, 32, 119]. This can be achieved using programs such as Maestro's Ligprep module [103]. Similarly, when ligands are optimized for a screen this should be reported, along with other relevant information such as the version of the database, the type of database, total number of compounds docked, purchased, tested, and validated as hits. This will allow for a more comprehensive comparison of G4 docking techniques.

6.3 Screening

The choice of screening approach(s) used is highly dependent on the user's intentions. Lead optimization strategies should include pharmacophore or similarity screens of large databases [46, 48], potentially followed with docking and/or MD minimizations and GBSA scoring. This will minimize time by selecting for ligands which closely resemble the validated scaffolds while also allowing for solvated and flexible receptor-ligand interactions to determine improvement of binding scores [100]. However, this approach may not be so useful in *de novo* drug discovery campaigns where it is more advantageous to screen a large diverse library. As mentioned above, filtering and reducing your library places an inherent bias on the size of the chemical space that will be evaluated, leading to redundancy in scaffolds (see Figures 4 & 10). Ideally one should select as large a database as possible and screen with a rapid, flexible ligand screening platform, such as Surflex-dock or Autodock (which has now been surpassed in speed and user friendliness with the release of Autodock Vina), followed with extensive re-docking, consensus docking, or MD simulations with MM/PBSA or MM/GBSA calculations.

A second consideration is the definition of the site to be docked. Most docking platforms have features to allow for ligand-based docking site generation. Conversely, in *de novo* discovery (which is often the case of groove-binders) there is usually no defined site, and therefore a site must be chosen by the user. This is done by defining a 3D grid about the putative ligand binding site (Glide, Autodock, Autodock Vina, DOCK, and ICM), by generation of a space filling protomol (which is a pre-computed representation of an ideal ligand [120])(Surflex-Dock), or simply by defining a bound ligand or set of residues (Surflex-dock, GOLD). Drawing on the authors' own experience, the docking site should be as small and focused as possible. Unfortunately, most ligands bound to G4s in crystal structures are cationic, polycyclic, and highly conjugated end-pasters bound to the 5' or 3' tetrad faces. It follows that using these complexes as the basis for docking will undoubtedly

result in top ranked compounds with similar features [48, 59, 84, 85] and, thus, not useful in the discovery of groove-binders or loop-interacting ligands.

6.4 Screening Analyses

Regardless of screening strategy, the user will likely generate more compounds than can feasibly be tested. If only the highest-ranking molecules are to be purchased, visual inspection is recommended. Although tedious, this process appears to increase enrichment in real groove-binders [92, 95] by removing erroneous ‘false-positives’, high steric clashing compounds, and molecules with poor hydrogen bonding interactions. This can also help rule out compounds docked into unintentional sites because of poorly defined docking sites or grids. Additionally, post-docking clustering based on molecular similarity criterion can reduce the redundancy in large screens and help inform purchasing decisions for diverse scaffolds [60, 92, 95, 102]. This is commonly done using similarity coefficients. Tanimoto, Dice, and Cosine similarity coefficients are numerical values computed from molecular attributes which are commonly used in clustering analyses [101]. It is also apparent that selectivity can be enriched for by re-docking the top-ranking compounds to potential off-targets, such as dsDNA [81], and others have reported enrichment from cross-platform consensus scoring [48].

7 Concluding Remarks

We present here a comprehensive overview of G4 virtual screening methodologies, along with suggestions to help guide future campaigns. These reports have shown that proper receptor optimization, large screening libraries, and appropriate downstream analyses of hits can result in great enrichment for novel G4 ligands. Conversely, we find that filtered libraries impose a major limitation on ligand diversity. Furthermore, there is a fundamental deficiency in reporting relevant information regarding VS campaigns, such as: library sizes, library preparation (optimizing, filtering, tautomer generation), and contents (fragment-like, drug like, natural products), total purchased vs. tested compounds, receptor preparation (protonation, modified bases, energy minimizations, MD), and downstream analysis (clustering, visual inspections, re-docking). This information is critical for evaluating G4 virtual drug discovery strategies.

There are potentially hundreds or thousands of G-quadruplexes which form within promoters, telomeres, RNA transcripts, and even LINEs and SINES [9, 11, 121, 122]. As articulated previously [27], G-quadruplexes are easily targetable with heterocyclic aromatic compounds because of the common tetrad face. Selectivity, then, must come from groove-interacting ligands or by end-pasting molecules with “built-in” selectivity for loops around the 5’ or 3’ interface. This is best achieved using massive, un-filtered libraries targeted at small pockets in and around the loops and grooves (Figure 15). The authors have recently used Surflex-Dock version 2.1 to screen the ZINC [50] drug-like libraries (versions 2014 and 2016) for a total of ~45 million compounds docked to multiple residue-defined loop/groove pockets of a modeled hTERT G-quadruplex (modeled using guanine stacks from the parallel *c-myc* G4, PDB ID: 1XAV) [113]. The quadruplex was subjected to MD simulations and stripped of waters and ions before docking. Docking was carried out using a computing

grid known as the DatascanGrid [123] which utilizes computers across schools in Kentucky. Ligands were used as-is from the ZINC drug-like database. Purchased compounds were chosen by hierarchical clustering of the top 6,000 molecules using Tanimoto similarity coefficients. From this analysis, 69 compounds were selected and screened using FRET, CD, ITC, fluorescent intercalator displacement assays, and analytical ultracentrifugation. The initial FRET screen resulted in ~33/69 G4 interacting compounds. The top 3 were further characterized, resulting in 2 potent groove or loop interacting ligands (unpublished) which are currently undergoing optimization and lead development.

Virtual screening of G-quadruplexes and other higher order nucleic acid structures is still in its infancy. As noted here, few VS platforms have been used in G4 drug discovery and even fewer have been used extensively enough with nucleic acids as to permit cross-platform comparisons. Furthermore, like protein systems, nucleic acids remain sensitive (if not more so) to the limitations of VS technologies. As mentioned previously [33, 124] receptor flexibility remains difficult to address in a high-throughput manner, and so G4 loops remain a challenge to target. Similarly, while docking algorithms can be very reproducible and rapid, there remains a dire need for accurate, robust scoring approximations [124]. Fortunately, the predictions [125] of hit enrichment from high performance computing and large libraries were correct. So, while the world awaits breakthroughs in scoring, receptor flexibility, and machine learning [33, 126], it might be valuable to seek out your nearest computing cluster to carry out your G4 screening.

Acknowledgements

The authors would like to thank Lynn DeLeeuw and Robert D. Gray for reviewing this manuscript prior to submission.

Molecular graphics and analyses were performed with the UCSF Chimera package. Chimera is developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIGMS P41-GM103311).

Funding

This work was supported by the National Institute of Health (grants GM077422 and P30GM106396).

Abbreviations:

G4	G-quadruplex
VS	virtual screening
HTS	high-throughput screening
¹H-NMR	proton nuclear magnetic resonance spectroscopy
dsDNA	double-stranded DNA
CD	circular dichroism
T_m	melting temperature
MD	molecular dynamics

GM	geometric matching
IC	incremental construction
MC	Monte Carlo
GA	genetic algorithms
GBSA	generalized Born solvent accessible surface area
PBSA	Poisson-Boltzmann solvent accessible surface area
PDB	protein data bank
FF	force-field
MM	molecular mechanics
FRET	Förster resonance energy transfer
ITC	isothermal titration calorimetry
VDW	van der Waals

References

- [1]. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* 2004;25(13):1605–1612. Epub 2004/07/21. doi: 10.1002/jcc.20084. PubMed PMID: . [PubMed: 15264254]
- [2]. Rhodes D, Lipps HJ G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res* 2015;43(18):8627–8637. doi: 10.1093/nar/gkv862. PubMed PMID: ; PMCID: PMC4605312. [PubMed: 26350216]
- [3]. Lane AN, Chaires JB, Gray RD, Trent JO Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res* 2008;36(17):5482–5515. doi: 10.1093/nar/gkn517. PubMed PMID: ; PMCID: PMC2553573. [PubMed: 18718931]
- [4]. Schiavone D, Guilbaud G, Murat P, Papadopoulou C, Sarkies P, Prioleau MN, Balasubramanian S, Sale JE Determinants of G quadruplex-induced epigenetic instability in REV1-deficient cells. *EMBO J*. 2014;33(21):2507–2520. doi: 10.15252/embj.201488398. PubMed PMID: ; PMCID: PMC4282387. [PubMed: 25190518]
- [5]. Bugaut A, Balasubramanian S 5'-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic Acids Res* 2012;40(11):4727–4741. doi: 10.1093/nar/gks068. PubMed PMID: ; PMCID: PMC3367173. [PubMed: 22351747]
- [6]. Takahama K, Oyoshi T Specific binding of modified RGG domain in TLS/FUS to G-quadruplex RNA: tyrosines in RGG domain recognize 2'-OH of the riboses of loops in G-quadruplex. *J Am Chem Soc* 2013;135(48):18016–18019. doi: 10.1021/ja4086929. PubMed PMID: . [PubMed: 24251952]
- [7]. Hansel-Hertsch R, Di Antonio M, Balasubramanian S DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential. *Nat Rev Mol Cell Biol* 2017;18(5):279–284. doi: 10.1038/nrm.2017.3. PubMed PMID: . [PubMed: 28225080]
- [8]. Guilbaud G, Murat P, Recolin B, Campbell BC, Maiter A, Sale JE, Balasubramanian S Local epigenetic reprogramming induced by G-quadruplex ligands. *Nat Chem* 2017;9(11):1110–1117. doi: 10.1038/nchem.2828. PubMed PMID: ; PMCID: PMC5669467. [PubMed: 29064488]

- [9]. Bochman ML, Paeschke K, Zakian VA DNA secondary structures: stability and function of G-quadruplex structures. *Nat Rev Genet* 2012;13(11):770–780. doi: 10.1038/nrg3296. PubMed PMID: ; PMCID: PMC3725559. [PubMed: 23032257]
- [10]. Chambers VS, Marsico G, Boutell JM, Di Antonio M, Smith GP, Balasubramanian S High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat Biotechnol* 2015;33(8):877–881. doi: 10.1038/nbt.3295. PubMed PMID: . [PubMed: 26192317]
- [11]. Hansel-Hertsch R, Beraldi D, Lensing SV, Marsico G, Zyner K, Parry A, Di Antonio M, Pike J, Kimura H, Narita M, Tannahill D, Balasubramanian S G-quadruplex structures mark human regulatory chromatin. *Nat Genet* 2016;48(10):1267–1272. doi: 10.1038/ng.3662. PubMed PMID: . [PubMed: 27618450]
- [12]. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 2001;46(1-3):3–26. PubMed PMID: . [PubMed: 11259830]
- [13]. Neidle S Human telomeric G-quadruplex: the current status of telomeric G-quadruplexes as therapeutic targets in human cancer. *FEBS J.* 2010;277(5):1118–1125. doi: 10.1111/j.1742-4658.2009.07463.x. PubMed PMID: . [PubMed: 19951354]
- [14]. Shi DF, Wheelhouse RT, Sun D, Hurley LH Quadruplex-interactive agents as telomerase inhibitors: synthesis of porphyrins and structure-activity relationship for the inhibition of telomerase. *J Med Chem* 2001;44(26):4509–4523. PubMed PMID: . [PubMed: 11741471]
- [15]. Izbicka E, Wheelhouse RT, Raymond E, Davidson KK, Lawrence RA, Sun D, Windle BE, Hurley LH, Von Hoff DD Effects of cationic porphyrins as G-quadruplex interactive agents in human tumor cells. *Cancer Res* 1999;59(3):639–644. PubMed PMID: . [PubMed: 9973212]
- [16]. Wei C, Han G, Jia G, Zhou J, Li C Study on the interaction of porphyrin with G-quadruplex DNAs. *Biophys Chem* 2008;137(1):19–23. doi: 10.1016/j.bpc.2008.06.006. PubMed PMID: . [PubMed: 18599180]
- [17]. Maraval A, Franco S, Vialas C, Pratviel G, Blasco MA, Meunier B Porphyrin-aminoquinoline conjugates as telomerase inhibitors. *Org Biomol Chem* 2003;1(6):921–927. PubMed PMID: . [PubMed: 12929629]
- [18]. Kiełtyka R, Fakhoury J, Moitessier N, Sleiman HF Platinum phenanthroimidazole complexes as G-quadruplex DNA selective binders. *Chemistry.* 2008;14(4):1145–1154. doi: 10.1002/chem.200700783. PubMed PMID: . [PubMed: 18022967]
- [19]. Pierce SE, Kiełtyka R, Sleiman HF, Brodbelt JS Evaluation of binding selectivities and affinities of platinum-based quadruplex interactive complexes by electrospray ionization mass spectrometry. *Biopolymers.* 2009;91(4):233–243. doi: 10.1002/bip.21130. PubMed PMID: ; PMCID: PMC2725001. [PubMed: 19117031]
- [20]. Zhang H, Xiang J, Hu H, Liu Y, Yang F, Shen G, Tang Y, Chen C Selective recognition of specific G-quadruplex vs. duplex DNA by a phenanthroline derivative. *Int J Biol Macromol* 2015;78:149–156. doi: 10.1016/j.ijbiomac.2015.03.034. PubMed PMID: . [PubMed: 25841379]
- [21]. Mergny JL, Lacroix L, Teulade-Fichou MP, Hounsou C, Guittat L, Hoarau M, Arimondo PB, Vigneron JP, Lehn JM, Riou JF, Garestier T, Helene C Telomerase inhibitors based on quadruplex ligands selected by a fluorescence assay. *Proc Natl Acad Sci U S A.* 2001 ;98(6):3062–3067. doi: 10.1073/pnas.051620698. PubMed PMID: ; PMCID: PMC30607. [PubMed: 11248032]
- [22]. Gama S, Rodrigues I, Mendes F, Santos IC, Gabano E, Klejevska B, Gonzalez-Garcia J, Ravera M, Vilar R, Paulo A Anthracene-terpyridine metal complexes as new G-quadruplex DNA binders. *J Inorg Biochem* 2016;160:275–286. doi: 10.1016/j.jinorgbio.2016.04.002. PubMed PMID: . [PubMed: 27267415]
- [23]. Perry PJ, Gowan SM, Reszka AP, Polucci P, Jenkins TC, Kelland LR, Neidle S 1,4- and 2,6-disubstituted amidoanthracene-9,10-dione derivatives as inhibitors of human telomerase. *J Med Chem* 1998;41(17):3253–3260. doi: 10.1021/jm9801105. PubMed PMID: . [PubMed: 9703471]
- [24]. Hampel SM, Sidibe A, Gunaratnam M, Riou JF, Neidle S Tetrasubstituted naphthalene diimide ligands with selectivity for telomeric G-quadruplexes and cancer cells. *Bioorg Med Chem Lett* 2010;20(22):6459–6463. doi: 10.1016/j.bmcl.2010.09.066. PubMed PMID: . [PubMed: 20932753]

- [25]. Paritala H, Firestone SM. Benzo(h)quinoline derivatives as G-quadruplex binding agents. *Bioorg Med Chem Lett* 2009;19(6):1584–1587. doi: 10.1016/j.bmcl.2009.02.016. PubMed PMID: . [PubMed: 19243943]
- [26]. Zeng DY, Kuang GT, Wang SK, Peng W, Lin SL, Zhang Q, Su XX, Hu MH, Wang H, Tan JH, Huang ZS, Gu LQ, Ou TM Discovery of Novel 11-Triazole Substituted Benzofuro[3,2-b]quinolone Derivatives as c-myc G-Quadruplex Specific Stabilizers via Click Chemistry. *J Med Chem* 2017;60(13):5407–5423. doi: 10.1021/acs.jmedchem.7b00016. PubMed PMID: . [PubMed: 28514170]
- [27]. Balasubramanian S, Hurley LH, Neidle S. Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nat Rev Drug Discov*. 2011;10(4):261–275. doi: 10.1038/nrd3428. PubMed PMID: ; PMCID: PMC3119469. [PubMed: 21455236]
- [28]. Martino L, Virno A, Pagano B, Virgilio A, Di Micco S, Galeone A, Giancola C, Bifulco G, Mayol L, Randazzo A Structural and thermodynamic studies of the interaction of distamycin A with the parallel quadruplex structure [d(TGGGGT)]₄. *J Am Chem Soc* 2007; 129(51): 16048–16056. doi: 10.1021/ja075710k. PubMed PMID: . [PubMed: 18052170]
- [29]. Pagano B, Fotticchia I, De Tito S, Mattia CA, Mayol L, Novellino E, Randazzo A, Giancola C Selective Binding of Distamycin A Derivative to G-Quadruplex Structure [d(TGGGGT)]₄(4). *J Nucleic Acids*. 2010;2010. doi: 10.4061/2010/247137. PubMed PMID: ; PMCID: PMC2915651. [PubMed: 20725616]
- [30]. Randazzo A, Galeone A, Esposito V, Varra M, Mayol L Interaction of distamycin A and netropsin with quadruplex and duplex structures: a comparative 1H-NMR study. *Nucleosides Nucleotides Nucleic Acids*. 2002;21(8–9):535–545. doi: 10.1081/NCN-120015067. PubMed PMID: . [PubMed: 12484449]
- [31]. Holt PA, Buscaglia R, Trent JO, Chaires JB A Discovery Funnel for Nucleic Acid Binding Drug Candidates. *Drug Dev Res* 2011;72(2):178–186. doi: 10.1002/ddr.20414. PubMed PMID: ; PMCID: PMC3090163. [PubMed: 21566705]
- [32]. Shoichet BK Virtual screening of chemical libraries. *Nature*. 2004;432(7019):862–865. doi: 10.1038/nature03197. PubMed PMID: ; PMCID: PMC1360234. [PubMed: 15602552]
- [33]. Lionta E, Spyrou G, Vassilatis DK, Cournia Z Structure-based virtual screening for drug discovery: principles, applications and recent advances. *Curr Top Med Chem* 2014;14(16):1923–1938. PubMed PMID: ; PMCID: PMC4443793. [PubMed: 25262799]
- [34]. Horvath D Pharmacophore-based virtual screening. *Methods Mol Biol* 2011;672:261–298. doi: 10.1007/978-1-60761-839-3_11. PubMed PMID: . [PubMed: 20838973]
- [35]. Koes DR, Camacho CJ Pharmer: efficient and exact pharmacophore search. *J Chem Inf Model*. 2011;51(6):1307–1314. doi: 10.1021/ci200097m. PubMed PMID: ; PMCID: PMC3124593. [PubMed: 21604800]
- [36]. Accelrys Discovery Studio. [cited 2017 16 Sept]. Available from: <http://accelrys.com/products/discovery-studio>.
- [37]. Wolber G, Langer T LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *J Chem Inf Model*. 2005;45(1):160–169. doi: 10.1021/ci049885e. PubMed PMID: . [PubMed: 15667141]
- [38]. Labute P, Williams C, Feher M, Sourial E, Schmidt JM Flexible alignment of small molecules. *J Med Chem* 2001;44(10):1483–1490. PubMed PMID: . [PubMed: 11334559]
- [39]. Dixon SL, Smondyrev AM, Rao SN PHASE: a novel approach to pharmacophore modeling and 3D database searching. *Chem Biol Drug Des* 2006;67(5):370–372. doi: 10.1111/j.1747-0285.2006.00384.x. PubMed PMID: . [PubMed: 16784462]
- [40]. SYBYL-X Suite. [cited 2017 16 Sept]. Available from: <http://tripos.com>.
- [41]. Taminiau J, Thijs G, De Winter H Pharao: pharmacophore alignment and optimization. *J Mol Graph Model*. 2008;27(2):161–169. doi: 10.1016/j.jmgm.2008.04.003. PubMed PMID: . [PubMed: 18485770]
- [42]. Chen SB, Tan JH, Ou TM, Huang SL, An LK, Luo HB, Li D, Gu LQ, Huang ZS Pharmacophore-based discovery of triaryl-substituted imidazole as new telomeric G-quadruplex ligand. *Bioorg Med Chem Lett* 2011;21(3):1004–1009. doi: 10.1016/j.bmcl.2010.12.019. PubMed PMID: . [PubMed: 21211968]

- [43]. Waller ZA, Shirude PS, Rodriguez R, Balasubramanian S Triarylpyridines: a versatile small molecule scaffold for G-quadruplex recognition. *Chem Commun (Camb)*. 2008(12): 1467–1469. doi: 10.1039/b718854d. PubMed PMID: ; PMCID: PMC2527513. [PubMed: 18338058]
- [44]. Lemmen C, Lengauer T Computational methods for the structural alignment of molecules. *J Comput Aided Mol Des* 2000;14(3):215–232. PubMed PMID: . [PubMed: 10756477]
- [45]. vROCS. [cited 2017 16 Sept]. Available from: <https://docs.eyesopen.com/rocs/>.
- [46]. Musumeci D, Amato J, Zizza P, Platella C, Cosconati S, Cingolani C, Biroccio A, Novellino E, Randazzo A, Giancola C, Pagano B, Montesarchio D Tandem application of ligand-based virtual screening and G4-OAS assay to identify novel G-quadruplex-targeting chemotypes. *Biochim Biophys Acta* 2017;1861(5 Pt B):1341–1352. doi: 10.1016/j.bbagen.2017.01.024. PubMed PMID: . [PubMed: 28130159]
- [47]. Maybridge. [cited 2018 03/24]. Available from: <http://www.maybridge.com>.
- [48]. Kaserer T, Rigo R, Schuster P, Alcaro S, Sissi C, Schuster D Optimized Virtual Screening Workflow for the Identification of Novel G-Quadruplex Ligands. *J Chem Inf Model*. 2016;56(3): 484–500. doi: 10.1021/acs.jcim.5b00658. PubMed PMID: . [PubMed: 26841201]
- [49]. AnalytiCon. [cited 2018 03/24]. Available from: <https://ac-discovery.com/screening-libraries/>.
- [50]. ZINC. [cited 2018 03/24]. Available from: <http://zinc.docking.org/>.
- [51]. ChemDiv. [cited 2018 03/24]. Available from: <http://www.chemdiv.com/services-menu/screening-libraries/>.
- [52]. SPECS. [cited 2018 03/24]. Available from: <http://www.specs.net>.
- [53]. Mcule. [cited 2018 03/24]. Available from: <https://mcule.com/database/>.
- [54]. eMolecules. [cited 2018 03/24]. Available from: <https://www.emolecules.com/info/plus/download-database>.
- [55]. PubChem [cited 2018 03/24]. Available from: <https://pubchem.ncbi.nlm.nih.gov/>.
- [56]. L. Chemicals. [cited 2018 03/24]. Available from: <http://www.lifechemicals.com/>.
- [57]. ChemBridge. ChemBridge [04/14/18]. Available from: <http://www.chembridge.com/screeninglibraries/>.
- [58]. Shan C, Lin J, Hou JQ, Liu HY, Chen SB, Chen AC, Ou TM, Tan JH, Li D, Gu LQ, Huang ZS Chemical intervention of the NM23-H2 transcriptional programme on c-MYC via a novel small molecule. *Nucleic Acids Res* 2015;43(14):6677–6691. doi: 10.1093/nar/gkv641. PubMed PMID: ; PMCID: PMC4538829. [PubMed: 26117539]
- [59]. Kang HJ, Park HJ In silico identification of novel ligands for G-quadruplex in the c-MYC promoter. *J Comput Aided Mol Des* 2015;29(4):339–348. doi: 10.1007/s10822-014-9826-z. PubMed PMID: . [PubMed: 25527072]
- [60]. Alcaro S, Musetti C, Distinto S, Casatti M, Zagotto G, Artese A, Parrotta L, Moraca F, Costa G, Ortuso F, Maccioni E, Sissi C Identification and characterization of new DNA G-quadruplex binders selected by a combination of ligand and structure-based virtual screening approaches. *J Med Chem* 2013;56(3):843–855. doi: 10.1021/jm3013486. PubMed PMID: . [PubMed: 23294188]
- [61]. Kar RK, Suryadevara P, Jana J, Bhunia A, Chatterjee S Novel G-quadruplex stabilizing agents: in-silico approach and dynamics. *J Biomol Struct Dyn* 2013;31(12):1497–1518. doi: 10.1080/07391102.2012.742246. PubMed PMID: . [PubMed: 23244447]
- [62]. Levesque MJ, Ichikawa K, Date S, Haga JH Design of a grid service-based platform for in silico protein-ligand screenings. *Comput Methods Programs Biomed* 2009;93(1):73–82. doi: 10.1016/j.cmpb.2008.07.005. PubMed PMID: ; PMCID: PMC2665129. [PubMed: 18771812]
- [63]. Schneider G Virtual screening: an endless staircase? *Nat Rev Drug Discov* 2010;9(4):273–276. Epub 2010/04/02. doi: 10.1038/nrd3139. PubMed PMID: . [PubMed: 20357802]
- [64]. Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE A geometric approach to macromolecule-ligand interactions. *J Mol Biol* 1982;161(2):269–288. PubMed PMID: . [PubMed: 7154081]
- [65]. Hevener KE, Zhao W, Ball DM, Babaoglu K, Qi J, White SW, Lee RE Validation of molecular docking programs for virtual screening against dihydropteroate synthase. *J Chem Inf Model*.

2009;49(2):444–460. doi: 10.1021/ci800293n. PubMed PMID: ; PMCID: PMC2788795. [PubMed: 19434845]

- [66]. Warren GL, Andrews CW, Capelli AM, Clarke B, LaLonde J, Lambert MH, Lindvall M, Nevins N, Semus SF, Senger S, Tedesco G, Wall ID, Woolven JM, Peishoff CE, Head MS A critical assessment of docking programs and scoring functions. *J Med Chem* 2006;49(20):5912–5931. doi: 10.1021/jm050362n. PubMed PMID: . [PubMed: 17004707]
- [67]. Kontoyianni M, McClellan LM, Sokol GS Evaluation of docking performance: comparative data on docking algorithms. *J Med Chem* 2004;47(3):558–565. doi: 10.1021/jm0302997. PubMed PMID: . [PubMed: 14736237]
- [68]. DoCk. [cited 2017 16 Sept]. Available from: <http://dock.compbio.ucsf.edu/>.
- [69]. Ewing TJ, Makino S, Skillman AG, Kuntz ID DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J Comput Aided Mol Des* 2001;15(5):411–428. PubMed PMID: . [PubMed: 11394736]
- [70]. Lang PT, Brozell SR, Mukherjee S, Pettersen EF, Meng EC, Thomas V, Rizzo RC, Case DA, James TL, Kuntz ID DOCK 6: combining techniques to model RNA-small molecule complexes. *RNA*. 2009;15(6):1219–1230. doi: 10.1261/rna.1563609. PubMed PMID: ; PMCID: PMC2685511. [PubMed: 19369428]
- [71]. AutoDock. [cited 2017 16 Sept]. Available from: <http://autodock.scripps.edu/>.
- [72]. Trott O, Olson AJ AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 2010;31(2):455–461. doi: 10.1002/jcc.21334. PubMed PMID: ; PMCID: PMC3041641. [PubMed: 19499576]
- [73]. GOLD. [cited 2017 16 Sept]. Available from: <https://www.ccdc.cam.ac.uk/solutions/csd-discovery/components/gold/>.
- [74]. Jain AN Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. *J Med Chem* 2003;46(4):499–511. doi: 10.1021/jm020406h. PubMed PMID: . [PubMed: 12570372]
- [75]. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin PS Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* 2004;47(7):1739–1749. doi: 10.1021/jm0306430. PubMed PMID: . [PubMed: 15027865]
- [76]. Neves MA, Totrov M, Abagyan R Docking and scoring with ICM: the benchmarking results and strategies for improvement. *J Comput Aided Mol Des* 2012;26(6):675–686. doi: 10.1007/s10822-012-9547-0. PubMed PMID: ; PMCID: PMC3398187. [PubMed: 22569591]
- [77]. Pagadala NS, Syed K, Tuszynski J Software for molecular docking: a review. *Biophys Rev* 2017;9(2):91–102. doi: 10.1007/s12551-016-0247-1. PubMed PMID: ; PMCID: PMC5425816. [PubMed: 28510083]
- [78]. Holt PA, Chaires JB, Trent JO Molecular docking of intercalators and groove-binders to nucleic acids using Autodock and Surflex. *J Chem Inf Model*. 2008;48(8):1602–1615. doi: 10.1021/ci800063v. PubMed PMID: ; PMCID: PMC2755229. [PubMed: 18642866]
- [79]. Meng XY, Zhang HX, Mezei M, Cui M Molecular docking: a powerful approach for structure-based drug discovery. *Curr Comput Aided Drug Des* 2011;7(2):146–157. PubMed PMID: ; PMCID: PMC3151162. [PubMed: 21534921]
- [80]. Welch W, Ruppert J, Jain AN Hammerhead: fast, fully automated docking of flexible ligands to protein binding sites. *Chem Biol* 1996;3(6):449–462. PubMed PMID: . [PubMed: 8807875]
- [81]. Hou JQ, Chen SB, Zan LP, Ou TM, Tan JH, Luyt LG, Huang ZS Identification of a selective G-quadruplex DNA binder using a multistep virtual screening approach. *Chem Commun (Camb)*. 2015;51(1):198–201. doi: 10.1039/c4cc06951j. PubMed PMID: . [PubMed: 25388435]
- [82]. Hart TN, Read RJ A multiple-start Monte Carlo docking method. *Proteins*. 1992; 13(3):206–222. doi: 10.1002/prot.340130304. PubMed PMID: . [PubMed: 1603810]
- [83]. Bursulaya BD, Totrov M, Abagyan R, Brooks CL, 3rd. Comparative study of several algorithms for flexible ligand docking. *J Comput Aided Mol Des* 2003;17(11):755–763. PubMed PMID: . [PubMed: 15072435]
- [84]. Lee HM, Chan DS, Yang F, Lam HY, Yan SC, Che CM, Ma DL, Leung CH Identification of natural product fonsecin B as a stabilizing ligand of c-myc G-quadruplex DNA by high-

- throughput virtual screening. *Chem Commun (Camb)*. 2010;46(26):4680–4682. doi: 10.1039/b926359d. PubMed PMID: . [PubMed: 20383387]
- [85]. Chan DS, Yang H, Kwan MH, Cheng Z, Lee P, Bai LP, Jiang ZH, Wong CY, Fong WF, Leung CH, Ma DL Structure-based optimization of FDA-approved drug methylene blue as a c-myc G-quadruplex DNA stabilizer. *Biochimie* 2011;93(6):1055–1064. doi: 10.1016/j.biochi.2011.02.013. PubMed PMID: . [PubMed: 21377506]
- [86]. Ma DL, Chan DS, Fu WC, He HZ, Yang H, Yan SC, Leung CH Discovery of a natural product-like c-myc G-quadruplex DNA groove-binder by molecular docking. *PLoS One*. 2012;7(8):e43278. doi: 10.1371/journal.pone.0043278. PubMed PMID: ; PMCID: PMC3422278. [PubMed: 22912844]
- [87]. Castro-Alvarez A, Costa AM, Vilarrosa J The Performance of Several Docking Programs at Reproducing Protein-Macrolide-Like Crystal Structures. *Molecules*. 2017;22(1). doi: 10.3390/molecules22010136. PubMed PMID: . [PubMed: 28106755]
- [88]. Hawkins PC, Skillman AG, Nicholls A Comparison of shape-matching and docking as virtual screening tools. *J Med Chem* 2007;50(1):74–82. Epub 2007/01/05. doi: 10.1021/jm0603365. PubMed PMID: . [PubMed: 17201411]
- [89]. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry*. 1998;19(14):1639–1662. doi: 10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B.
- [90]. Castillo-Gonzalez D, Mergny JL, De Rache A, Perez-Machado G, Cabrera-Perez MA, Nicolotti O, Introcaso A, Mangiatordi GF, Guedin A, Bourdoncle A, Garrigues T, Pallardo F, Cordeiro MN, Paz YMC, Tejera E, Borges F, Cruz-Monteaugudo M Harmonization of QSAR Best Practices and Molecular Docking Provides an Efficient Virtual Screening Tool for Discovering New G-Quadruplex Ligands. *J Chem Inf Model*. 2015;55(10):2094–2110. doi: 10.1021/acs.jcim.5b00415. PubMed PMID: . [PubMed: 26355653]
- [91]. Ma DL, Lai TS, Chan FY, Chung WH, Abagyan R, Leung YC, Wong KY Discovery of a drug-like G-quadruplex binding ligand by high-throughput docking. *ChemMedChem* 2008;3(6):881–884. doi: 10.1002/cmdc.200700342. PubMed PMID: . [PubMed: 18383062]
- [92]. Cosconati S, Marinelli L, Trotta R, Virno A, Mayol L, Novellino E, Olson AJ, Randazzo A Tandem application of virtual screening and NMR experiments in the discovery of brand new DNA quadruplex groove binders. *J Am Chem Soc* 2009;131(45):16336–16337. doi: 10.1021/ja9063662. PubMed PMID: . [PubMed: 19856964]
- [93]. Trotta R, De Tito S, Lauri I, La Pietra V, Marinelli L, Cosconati S, Martino L, Conte MR, Mayol L, Novellino E, Randazzo A A more detailed picture of the interactions between virtual screening-derived hits and the DNA G-quadruplex: NMR, molecular modelling and ITC studies. *Biochimie* 2011;93(8):1280–1287. doi: 10.1016/j.biochi.2011.05.021. PubMed PMID: . [PubMed: 21658428]
- [94]. Di Leva FS, Zizza P, Cingolani C, D'Angelo C, Pagano B, Amato J, Salvati E, Sissi C, Pinato O, Marinelli L, Cavalli A, Cosconati S, Novellino E, Randazzo A, Biroccio A Exploring the chemical space of G-quadruplex binders: discovery of a novel chemotype targeting the human telomeric sequence. *J Med Chem* 2013;56(23):9646–9654. doi: 10.1021/jm401185b. PubMed PMID: . [PubMed: 24256368]
- [95]. Amato J, Pagano A, Cosconati S, Amendola G, Fotticchia I, Iaccarino N, Marinello J, De Magis A, Capranico G, Novellino E, Pagano B, Randazzo A Discovery of the first dual G-triplex/G-quadruplex stabilizing compound: a new opportunity in the targeting of G-rich DNA structures? *Biochim Biophys Acta* 2017;1861(5 Pt B):1271–1280. doi: 10.1016/j.bbagen.2016.11.008. PubMed PMID: . [PubMed: 27836755]
- [96]. Gray RD, Trent JO, Chaires JB Folding and unfolding pathways of the human telomeric G-quadruplex. *J Mol Biol* 2014;426(8):1629–1650. doi: 10.1016/j.jmb.2014.01.009. PubMed PMID: ; PMCID: PMC3969789. [PubMed: 24487181]
- [97]. Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, Profeta S, Weiner P A new force field for molecular mechanical simulation of nucleic acids and proteins. *Journal of the American Chemical Society*. 1984;106(3):765–784. doi: 10.1021/ja00315a051.

- [98]. Hospital A, Goni JR, Orozco M, Gelpi JL Molecular dynamics simulations: advances and applications. *Adv Appl Bioinform Chem* 2015;8:37–47. doi: 10.2147/AABC.S70333. PubMed PMID: ; PMID: PMC4655909. [PubMed: 26604800]
- [99]. Leach AR, Molecular Dynamics Simulation Methods, Molecular Modelling: Principles and Applications, Pearson/Prentice Hall 2001, pp. 353–406.
- [100]. Moraca F, Amato J, Ortuso F, Artese A, Pagano B, Novellino E, Alcaro S, Parrinello M, Limongelli V Ligand binding to telomeric G-quadruplex DNA investigated by funnel-metadynamics simulations. *Proc Natl Acad Sci U S A*. 2017;114(11):E2136–E2145. doi: 10.1073/pnas.1612627114. PubMed PMID: ; PMID: PMC5358390. [PubMed: 28232513]
- [101]. Leach AR, The Use of Molecular Modelling and Chemoinformatics to Discover and Design New Molecules, Molecular Modelling: Principles and Applications, Pearson/Prentice Hall 2001, pp. 641–726.
- [102]. Rocca R, Moraca F, Costa G, Nadai M, Scalabrin M, Talarico C, Distinto S, Maccioni E, Ortuso F, Artese A, Alcaro S, Richter SN Identification of G-quadruplex DNA/RNA binders: Structure-based virtual screening and biophysical characterization. *Biochim Biophys Acta* 2017;1861(5 Pt B):1329–1340. doi: 10.1016/j.bbagen.2016.12.023. PubMed PMID: . [PubMed: 28025082]
- [103]. Schrodinger, LigPrep, Schrodinger, Schrodinger, LLC, New York, NY, 2018.
- [104]. Nakatani K, Hagihara S, Sando S, Sakamoto S, Yamaguchi K, Maesawa C, Saito I Induction of a remarkable conformational change in a human telomeric sequence by the binding of naphthyridine dimer: inhibition of the elongation of a telomeric repeat by telomerase. *J Am Chem Soc* 2003;125(3):662–666. Epub 2003/01/16. doi: 10.1021/ja027055g. PubMed PMID: . [PubMed: 12526665]
- [105]. Bhat J, Mondal S, Sengupta P, Chatterjee S In Silico Screening and Binding Characterization of Small Molecules toward a G-Quadruplex Structure Formed in the Promoter Region of c-MYC Oncogene. *ACS Omega*. 2017;2(8):4382–4397. doi: 10.1021/acsomega.6b00531. [PubMed: 30023722]
- [106]. Limongelli V, Bonomi M, Parrinello M Funnel metadynamics as accurate binding free-energy method. *Proc Natl Acad Sci U S A*. 2013;110(16):6358–6363. Epub 2013/04/05. doi: 10.1073/pnas.1303186110. PubMed PMID: ; PMID: PMC3631651. [PubMed: 23553839]
- [107]. Liu J, Wang R Classification of current scoring functions. *J Chem Inf Model*. 2015;55(3):475–482. doi: 10.1021/ci500731a. PubMed PMID: . [PubMed: 25647463]
- [108]. Fenu L A, Lewis R A, Good, C. A, M. B, Essex, W. J, Scoring Functions: From Free-energies of Binding to Enrichment in Virtual Screening, Structure-Based Drug Discovery, Springer 2007, pp. 226–236.
- [109]. Sottriffer C, Matter H, Classes of Scoring Functions, Virtual Screening: Principles, Challenges, and Practical Guidelines, John Wiley & Sons 2011, pp. 48.
- [110]. Garrett M DSG. Morris, Pique Michael E., Lindstrom William “Lindy”, Huey Ruth, Stefano, Forli WEH, Halliday Scott, Belew Rik and Olson Arthur J.. AutoDock Version 4.2.2012:8–9.
- [111]. Li J, Correia JJ, Wang L, Trent JO, Chaires JB Not so crystal clear: the structure of the human telomere G-quadruplex in solution differs from that present in a crystal. *Nucleic Acids Res* 2005;33(14):4649–4659. doi: 10.1093/nar/gki782. PubMed PMID: ; PMID: PMC1187823. [PubMed: 16106044]
- [112]. Meier M, Moya-Torres A, Krahn NJ, McDougall MD, Orriss GL, McRae EKS, Booy EP, McEleney K, Patel TR, McKenna SA, Stetefeld J Structure and hydrodynamics of a DNA G-quadruplex with a cytosine bulge. *Nucleic Acids Res* 2018. Epub 2018/05/03. doi: 10.1093/nar/gky307. PubMed PMID: . [PubMed: 29718405]
- [113]. Chaires JB, Trent JO, Gray RD, Dean WL, Buscaglia R, Thomas SD, Miller DM An improved model for the hTERT promoter quadruplex. *PLoS One*. 2014;9(12):e115580. doi: 10.1371/journal.pone.0115580. PubMed PMID: ; PMID: PMC4272262. [PubMed: 25526084]
- [114]. Petraccone L, Spink C, Trent JO, Garbett NC, Mekmaysy CS, Giancola C, Chaires JB Structure and stability of higher-order human telomeric quadruplexes. *J Am Chem Soc* 2011;133(51):20951–20961. doi: 10.1021/ja209192a. PubMed PMID: ; PMID: PMC3244555. [PubMed: 22082001]

- [115]. Garbett NC, Mekmaysy CS, Chaires JB Sedimentation velocity ultracentrifugation analysis for hydrodynamic characterization of G-quadruplex structures. *Methods Mol Biol* 2010;608:97–120. Epub 2009/12/17. doi: 10.1007/978-1-59745-363-9_7. PubMed PMID: ; PMCID: PMC3008627. [PubMed: 20012418]
- [116]. Del Villar-Guerra R, Trent JO, Chaires JB G-Quadruplex Secondary Structure Obtained from Circular Dichroism Spectroscopy. *Angew Chem Int Ed Engl* 2017. Epub 2017/10/28. doi: 10.1002/anie.201709184. PubMed PMID: ; PMCID: PMC5920796. [PubMed: 29076232]
- [117]. Triballeau N, Acher F, Brabet I, Pin JP, Bertrand HO Virtual screening workflow development guided by the "receiver operating characteristic" curve approach. Application to high-throughput docking on metabotropic glutamate receptor subtype 4. *J Med Chem* 2005;48(7):2534–2547. Epub 2005/04/02. doi: 10.1021/jm049092j. PubMed PMID: . [PubMed: 15801843]
- [118]. Rocca R, Moraca F, Costa G, Alcaro S, Distinto S, Maccioni E, Ortuso F, Artese A, Parrotta L Structure-based virtual screening of novel natural alkaloid derivatives as potential binders of h-telo and c-myc DNA G-quadruplex conformations. *Molecules*. 2014;20(1):206–223. doi: 10.3390/molecules20010206. PubMed PMID: . [PubMed: 25547724]
- [119]. Polgar T, Magyar C, Simon I, Keseru GM Impact of ligand protonation on virtual screening against beta-secretase (BACE1). *J Chem Inf Model*. 2007;47(6):2366–2373. doi: 10.1021/ci700223p. PubMed PMID: . [PubMed: 17944457]
- [120]. Jain AN Surflex-Dock 2.1: robust performance from ligand energetic modeling, ring flexibility, and knowledge-based search. *J Comput Aided Mol Des* 2007;21(5):281–306. doi: 10.1007/s10822-007-9114-2. PubMed PMID: . [PubMed: 17387436]
- [121]. Rigo R, Palumbo M, Sissi C G-quadruplexes in human promoters: A challenge for therapeutic applications. *Biochim Biophys Acta* 2017;1861(5 Pt B):1399–1413. doi: 10.1016/j.bbagen.2016.12.024. PubMed PMID: . [PubMed: 28025083]
- [122]. Sahakyan AB, Murat P, Mayer C, Balasubramanian S G-quadruplex structures within the 3' UTR of LINE-1 elements stimulate retrotransposition. *Nat Struct Mol Biol* 2017;24(3):243–247. doi: 10.1038/nsmb.3367. PubMed PMID: . [PubMed: 28134931]
- [123]. DataseamGrid. DataseamGrid [04/15/18]. Available from: <http://www.kydataseam.com/>.
- [124]. Heikamp K, Bajorath J The future of virtual compound screening. *Chem Biol Drug Des* 2013;81(1):33–40. Epub 2012/12/21. doi: 10.1111/cbdd.12054. PubMed PMID: . [PubMed: 23253129]
- [125]. Yuriev E Challenges and advances in structure-based virtual screening. *Future Med Chem* 2014;6(1):5–7. Epub 2013/12/24. doi: 10.4155/fmc.13.186. PubMed PMID: . [PubMed: 24358942]
- [126]. Chen H, Engkvist O, Wang Y, Olivecrona M, Blaschke T The rise of deep learning in drug discovery. *Drug Discov Today*. 2018;23(6):1241–1250. Epub 2018/01/26. doi: 10.1016/j.drudis.2018.01.039. PubMed PMID: . [PubMed: 29366762]

Highlights

- Introduction to G-quadruplex *in silico* screening methodologies.
- Review of G-quadruplex drug scaffolds derived from *in silico* approaches.
- Integration of best practices for G-quadruplex *in silico* docking strategies.

Figure 1. G-quadruplex structure.

(A) Orientation of guanines in a G-quadruplex quartet. (B) Monovalent cations often occupy the middle of two quartets, helping to stabilize the partial negative charge shared among the O_6 oxygen of adjacent quartets. Phosphate backbone is shown as vertical black lines.

Figure 2. Common G-quadruplex "end-pasting" molecular scaffolds found in the literature. (A) porphyrin, (B) phenanthroline, (C) anthracene, (D) naphthalene, (E) quinoline.

Figure 3.
Structures of (A) Quarfloxacin, (B) Distamycin A, and (C) Netropsin.

Figure 4.
Structures of (A) a triarylpyridine and (B) a triaryl imidazole.

Figure 5.
Structures of the human telomere groove-binding ligands discovered by Musumeci et al.

Figure 6.
Structures of the reported *c-myc* quadruplex stabilizing compounds from Kang et al.

Figure 7. Structures of (A) a pyrrolopyrazine compound which stabilizes the *c-myc* quadruplex discovered by Hou et al. and (B) SYUIQ-5.

Figure 8.

Structures of (A) fonsecin B, (B) methylene blue, and the *c-myc* quadruplex stabilizing compounds (C) a methylene blue derivative discovered by Chan et al., and (D) a carbamide containing compound discovered by Ma et al.

Figure 9.
The psoralen derivative discovered by Alcaro et al. which stabilized the human telomere quadruplexes.

Figure 10.
Structures of the 14 compounds discovered by Kaserer et al. using a multi-platform consensus approach to target the human telomere quadruplexes.

Figure 11.

Structures of (A-C) the parallel groove-binders discovered by Trotta et al. and (D) the human telomere interacting groove-binder discovered by Di Leva. (E) The dual G-quadruplex/G-triplex stabilizing compound discovered by Amato et al.

Figure 12.
Structures of the two telomere interacting compounds discovered by Kar et al. with moderate selectivity for G4s over dsDNA.

Figure 13.
Structure of the naphthyridine compound discovered by Rocca et al. and shown to stabilize both RNA and DNA G-quadruplexes.

Figure 14.
Structure of the carbamoylpiperidinium containing compound discovered by Bhat et al. and shown to stabilize the *c-myc* G4 by an end-pasting mechanism.

Figure 15.

(Left) hTERT core promoter G-quadruplex model created by Chaires and Trent et al. Phosphate backbone is shown in tan, nucleotides in blue, and potassium in purple. (Right) Surface representation showing a large binding pocket (dark area inside of yellow dashed oval) at the junction between the first and second G4s of the hTERT G-quadruplex. Images were rendered in Chimera v1.12 [1].

Table 1.

Docking platform and algorithm presented and discussed in this review.

Docking Platform	Algorithm
Glide	OPLS-AA force field optimization with Monte-Carlo refinement (FFMC/MD)
ICM	Flexible Monte-Carlo (MC)
AutoDock Vina	Flexible Monte-Carlo (MC)
DOCK V4-V6	Geometric matching and Incremental (GM/IC/MD)
AutoDock	Flexible Genetic Algorithm (GA)
GOLD	Flexible Genetic Algorithm (GA)
Surflex-DOCK	Hammerhead fragment-based algorithm and Genetic algorithm (IC/GA)