We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Chapter 6

Structure-Based Approaches Targeting Oncogene Promoter G-Quadruplexes

Dik-Lung Ma, Victor Pui-Yan Ma, Ka-Ho Leung, Hai-Jing Zhong, Hong-Zhang He, Daniel Shiu-Hin Chan and Chung-Hang Leung

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/54809

1. Introduction

The genetic information stored in DNA can be transcribed and translated into functional proteins with various biological roles, and the control of gene expression and cell division is tightly controlled under normal physiological conditions. However, genetic mutations arising during DNA replication can trigger uncontrolled cell growth, leading to the development of various types of cancers (Croce 2008). The cellular transformational events associated with cancer have been linked with mutations in particular genes, termed proto-oncogenes. These genes are necessary for the normal development and differentiation of cells, but when mutated into oncogenes they can lead to the overexpression of proteins involved in signal transduction and mitosis, ultimately resulting in cancer development. Blocking oncogenic translation using siRNAs has attracted intense attention in the literature (Heidenreich 2009; Ventura et al. 2009), but inhibiting oncogenic transcription through targeting DNA itself has been less explored.

While DNA is a well-established biomolecular target for anti-cancer therapy, most DNA-binding drugs such as cisplatin (Alderden et al. 2006) and its analogues interact with DNA non-selectively, resulting in adverse side effects (Jung et al. 2007). Consequently, this has driven interest in the targeting of unusual, non-canonical structures in DNA, in order to achieve selectivity for particular (onco)genes while potentially reducing adverse side effects. One such DNA structure that has attracted significant attention in the recent literature as an anti-cancer target is the G-quadruplex. While G-quadruplexes were initially regarded as somewhat of a structural curiosity when they were first discovered, accumulating evidence over the past decade have suggested that these non-canonical DNA structures may play important roles in modulating various biological processes (Lipps et al. 2009).

INTECH
open science | open minds

© 2012 Leung et al., licensee InTech. This is an open access chapter distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

G-quadruplexes are four-stranded guanine-rich DNA structures that were first found at the ends of eukaryotic telomeres, and the role of telomeric G-quadruplexes for inhibiting telomerase activity has been intensely studied since the early 1990s (Blackburn 1991). Human telomeric DNA is usually 4–14 kilobases long, and is comprised of TTAGGG tandem repeats. Up-regulated telomerase activity in cancer cells maintains the length of telomeres after cell division, conferring immortality. Hurley and co-workers demonstrated that the activity of telomerase can be inhibited by small molecule-induced stabilization of telomeric G-quadruplex (Wheelhouse et al. 1998).

A few years later, Hurley and co-workers reported the seminal discovery of a potential G-quadruplex structure in the nuclease hypersensitive element III_1 ($NHEIII_1$) of the promoter region of the c-*myc* oncogene, and they further demonstrated that the transcriptional repression of c-*myc* can be achieved by induction of putative G-quadruplex formation by a small molecule (Siddiqui-Jain et al. 2002). Evidently, c-*myc* transcription was inhibited by the putative formation of the G-quadruplex structure in the promoter region, thus suppressing oncogenic expression. Later, other studies identified the presence of G-quadruplex-forming sequences in the promoter regions of other oncogenes such as c-*kit (*Rankin et al. 2005*)*, KRAS (Cogoi et al. 2006), *bcl*-2 (Dai et al. 2006) and VEGF (Jiang et al. 1991). In 2007, Huppert and Balasubramanian conducted a large-scale bioinformatics analysis throughout the human genome, and found that G-quadruplex-forming sequences are enriched in the promoter regions of genes, and that >40% of annotated genes bears at least one potential G-quadruplex sequence within 1kb of the transcription start site (Huppert et al. 2007). Recent evidence has suggested that G-quadruplexes may exist *in vivo* and may play putative roles in various biological processes, such as the regulation of gene expression (Dexheimer et al. 2009; Gonzalez et al. 2009). Consequently, targeting the oncogenic G-quadruplexes using small molecules has emerged as an alternative strategy for the potential treatment of cancers (Balasubramanian et al. 2009).

Since the discovery of the first c-*myc* G-quadruplex stabilizer TMPyP4 and inhibitor of c-*myc* oncogenic expression by Hurley and co-workers, many other c-*myc* interactive small molecule ligands have been identified. For example, cationic porphyrins (Grand et al. 2002), quindoline and berberine derivatives (Ou et al. 2007; Lu et al. 2008; Ma et al. 2008), and trisubstituted isoalloxazines (Bejugam et al. 2007) have been demonstrated to interfere with the oncogenic transcription *in vitro*. Quarfloxin, developed by Cylene Pharmaceuticals, entered clinical trials due to its ability to interact with G-quadruplexes *in vivo* (Duan et al. 2001). Interestingly, quarfloxin concentrates in the nucleus and disrupts the G-quadruplex-nucleolin interaction, leading to the redistribution of nucleolin in the nucleoplasm which ultimately triggers the apoptosis and inhibition of cancer cell growth.

With advances in computer processing power and in the development of algorithms for molecular stimulation and docking, the use of high-throughput virtual screening for drug discovery has become increasingly popular (McInnes 2007). The rapid screening of a large chemical library using computational programs can efficiently weed out non-binding ligands *in silico*, thus dramatically reducing the number of compounds to be tested *in vitro*. While the use of computer-aided virtual screening for discovering enzyme antagonists has

been widely employed, the use of computational analysis for identifying G-quadruplex ligands has been comparatively less explored (Ma et al. 2012). In this chapter, we first describe the general structure of G-quadruplexes and their involvement in transcriptional events, particularly those relevant to oncogenic expression. We then discuss the use of *in silico* methods to identify small molecule ligands of oncogenic promoter G-quadruplexes, and identify features or limitations of each method. Finally, we highlight recent, representative, examples of promoter G-quadruplex targeting by small molecules discovered using *in silico* methods.

2. General structure of the G-quadruplex and its involvement in transcriptional events

G-quadruplexes are constructed from stacks of G-tetrads, which consist of four guanine bases aligned in a co-planar arrangement stabilized by Hoogsteen hydrogen-bonding and monovalent cations (e.g. K^+ and Na^+) in the central cavity (Figure 1) (Mergny et al. 1998; Parkinson et al. 2002; Huppert et al. 2007). G-quadruplexes exhibit a high degree of structural polymorphism, contributing to the wide variety of distinct G-quadruplex topologies that differ in strand orientation, loop size, surface and groove dimensions (Burge et al. 2006). Consequently, G-quadruplexes formed from different DNA sequences may exhibit unique structural features that can be specifically targeted by small molecule ligands (Monchaud et al. 2008).

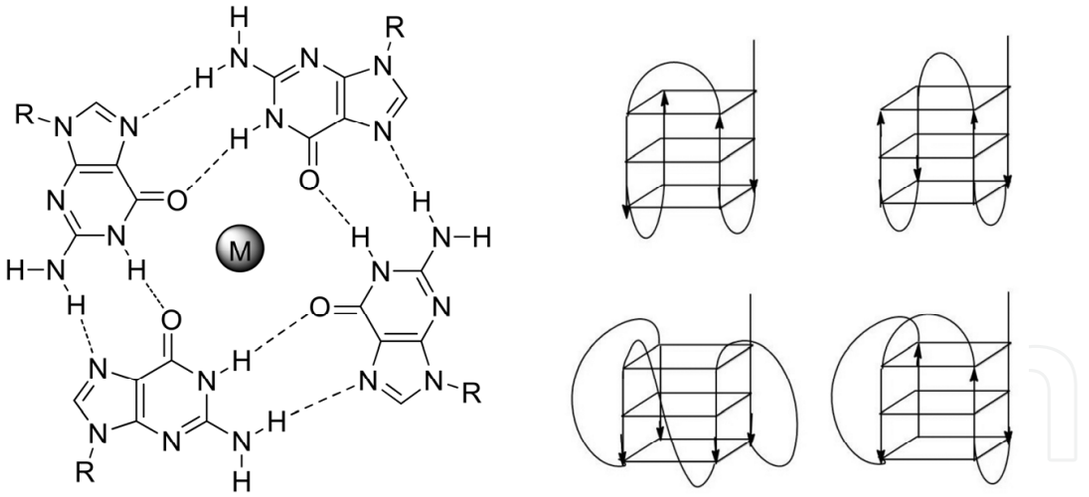


Figure 1. The structure of a G-tetrad stabilized by Hoogsteen hydrogen bonding and a monovalent cation resided in the central channel (left). Some possible topologies for an intramolecular G-quadruplex (right).

As previously mentioned, the occurrence of G-quadruplex-forming regions in the promoter region of oncogenes offers an alternative therapeutic avenue for the treatment of cancer. The induction of the G-quadruplex structure in the promoter region of the target gene could inhibit transcription of the oncogene, thus suppressing the production of the resultant oncoprotein. The potential to repress oncogenic expression by G-quadruplex formation can

be illuminated by considering the history of the efforts targeted against well-studied oncogene c-*myc*.

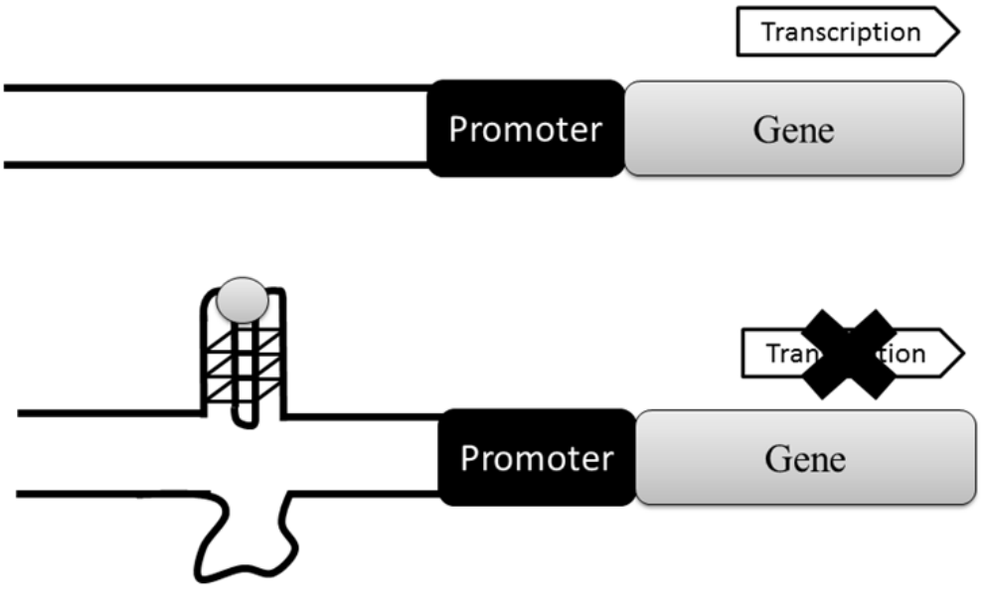


Figure 2. Transcription regulation of oncogenes by promoter G-quadruplex formation mediated by small molecules (circle).

MYC protein is a transcription factor that controls cell proliferation, differentiation and apoptosis (Marcu et al. 1992), and its cellular level is strictly regulated in normal cells. Mutation of c-*myc* and the overexpression of the MYC protein are observed in around 80% of solid tumors, including cervical carcinoma, myeloid leukemias and osteosarcomas (Lutz et al. 2002; Meyer et al. 2008; Wierstra et al. 2008). Accumulating evidence has revealed that the c-*myc* promoter region plays a pivotal role in the regulation of c-*myc* transcriptional activity. In particular, the nuclear hypersensitivity element III_1 (NHE III_1), a 27 bp guanine rich sequence located upstream of the c-*myc* protein, has been reported to control around 90% of c-*myc* transcription (Davis et al. 1989). *In vitro* experiments suggested that this sequence is able to fold into an intramolecular parallel G-quadruplex with predominant 1:2:1 and 2:1:1 loop topologies (Seenisamy et al. 2004). Hurley and co-workers showed the basal transcription activity of c-*myc* can be significantly enhanced by destabilizing the c-*myc* G-quadruplex through a guanine-to-thymine mutation in the quadruplex-forming sequence (Siddiqui-Jain et al. 2002). In the same report, they demonstrated the suppression of oncogenic c-*myc* transcription activity by a cationic porphyrin that can stabilize the G-quadruplex structure (Siddiqui-Jain et al. 2002). These results demonstrated that the c-*myc* promoter G-quadruplex may act as a regulator of oncogenic transcription, and that small molecule stabilizers of the G-quadruplex could potentially down-regulate the expression of oncogenes (Figure 2).

These promoter G-quadruplex-stabilizing ligands have potential advantages as alternative anti-cancer compounds compared to conventional protein or enzyme inhibitors (Balasubramanian et al. 2011). Firstly, since the availability of G-quadruplexes in cells is

generally limited, a lower concentration of inhibitor could theoretically be used to achieve the desired biological effect. Secondly, due to the unique structural diversity of G-quadruplex motifs, superior selectivity towards a particular G-quadruplex may be potentially achieved by the rational design and modification of the lead compound. Thirdly, a number of oncogenes such as c-*kit*, BRAF and c-*myc*, which have been reported to contain G-quadruplex-forming motifs in their promoter regions, encode kinase or protein products that have been clinically validated as targets for the treatment of cancer. However, a number of issues remain for the development of effective promoter G-quadruplex ligands for the treatment of human diseases. These include acquiring more detailed and comprehensive structural information on the relevant topologies of G-quadruplexes in living systems, as well as developing ligands with sufficient G-quadruplex selectivity and affinity for potential *in vivo* application. Leading experts Balasubramanian, Hurley and Neidle have recently reviewed the targeting of oncogenic promoter G-quadruplexes as a potential anti-cancer strategy (Balasubramanian et al. 2011).

3. *In silico* methods in drug discovery

Virtual screening techniques have recently emerged as a complementary technique to traditional high-throughput screening technologies employed in the pharmaceutical industry (Shoichet 2004; Ghosh et al. 2006; Cavasotto et al. 2007). Using computer-aided methodologies, large numbers of compounds can be rapidly screened in order to efficiently eliminate non-binding compounds *in silico*, thus dramatically reducing the costs associated with preliminary testing in a drug discovery project. However, while the application of *in silico* techniques for discovering enzyme inhibitors has been well-established, the targeting of DNA structures using virtual screening has been comparatively less explored. Broadly speaking, virtual screening can be sub-divided into pharmacophore modelling and molecular docking. A representative list of commercially available molecular docking softwares for both pharmacophore modelling and molecular docking (receptor-ligand modelling) is given in Table 1.

Pharmacophore modelling can be further classified into structure-based and ligand-based methods. In structure-based pharmacophore modelling, the structure of receptor must be first determined using techniques such as X-ray crystallography and nuclear magnetic resonance (NMR). Alternatively, if the structure of particular target is not known, a model can be constructed by homology with closely-related structures. In general, a structure containing the biomolecular target complexed with its ligand is advantageous for virtual screening since the key features of the interaction between the ligand and the binding pocket can be directly examined. Some commercially available computational software programs such as LIGANDSCOUT (Wolber et al. 2004) and POCKET v.2 (Chen et al. 2006) are able to analyse the binding interaction and calculate the relevant contributions of each feature to the specificity and inhibitory potency of the ligand. Ligand–target interactions can include hydrogen bonding, ionic interactions and hydrophobic interactions, and this information can be harnessed to generate a three-dimensional (3D) pharmacophore model.

Company/Institution	Software	Uses
Accelrys	Discovery Studio, Insight II	Pharmacophore modelling, receptor-ligand docking, de novo drug design, molecular stimulation
MolSoft	ICM-Pro	Pharmacophore modelling, receptor-ligand docking, 3D QSAR model constructions
Tripos	Sybyl	Receptor-ligand docking, chemoinformatics, 3D QSAR model constructions
Scripps Research Institute	Autodock	Receptor-ligand docking
Schrodinger	Phase	Receptor-ligand docking, pharmacophore modelling

Table 1. Examples of commercially available drug discovery softwares.

In contrast, a prior knowledge of the biomolecular target is not needed in ligand based pharmacophore modelling, but instead a library of compounds with known potencies towards the biomolecular target is required for the construction of a training set. *In silico* techniques are then employed to generate a 3D pharmacophore that bears the representative electronic and steric features of the compounds from the training set. To obtain a reliable 3D pharmacophore, the training set should include structurally diverse compounds with *in vitro* potencies spanning a few orders of magnitude.

To confirm the validity of the 3D pharmacophore generated from either structure-based or ligand-based pharmacophore modelling, cost analysis techniques can be carried out based on statistical calculations in order to generate the "best" hypothetical structure. The validated pharmacophore is then subjected to virtual screening from chemical libraries to identify molecules that possess similar steric and electronic features with the pharmacophore. However, a drawback of pharmacophore modelling is that since the affinity calculation only involves the matching of geometry and functional groups of the potential ligand with the 3D pharmacophore, the screening process will tend to reveal ligands that structurally and electronically resemble the training set of compounds, rather than uncovering novel hit scaffolds.

On the other hand, molecular docking represents a totally different approach for virtual screening of bioactive compounds. Molecular docking involves stimulating the interactions between biomolecules and the ligands by computational algorithms. Molecular modelling has been gaining in popularity due to the increasing availability of biomolecular structures determined by either X-ray crystallography or NMR. In addition, advances in computational power and the continual development of more refined docking algorithms help to mitigate the relatively high computational strain demanded by molecular docking. In molecular docking, knowledge of the 3D biomolecular structure is essential, with or without the binding ligand. As previously described, the use of a biomolecular structure co-crystallized

with a ligand is preferred as the binding pocket of the ligand can be easily identified and the subsequent docking analysis can then be restricted to the areas around the binding pocket in order to avoid wastage of computational resources and to eliminate false positives that interact outside of the binding site.

After completion of a virtual screening campaign, the resulting hit list of compounds can be subjected to experimental assays for hit validation (Figure 3). Alternatively, the hit structures can be used to construct analogues that can be screened *in silico* to potentially generate more potent ligands before chemical synthesis and biological testing.

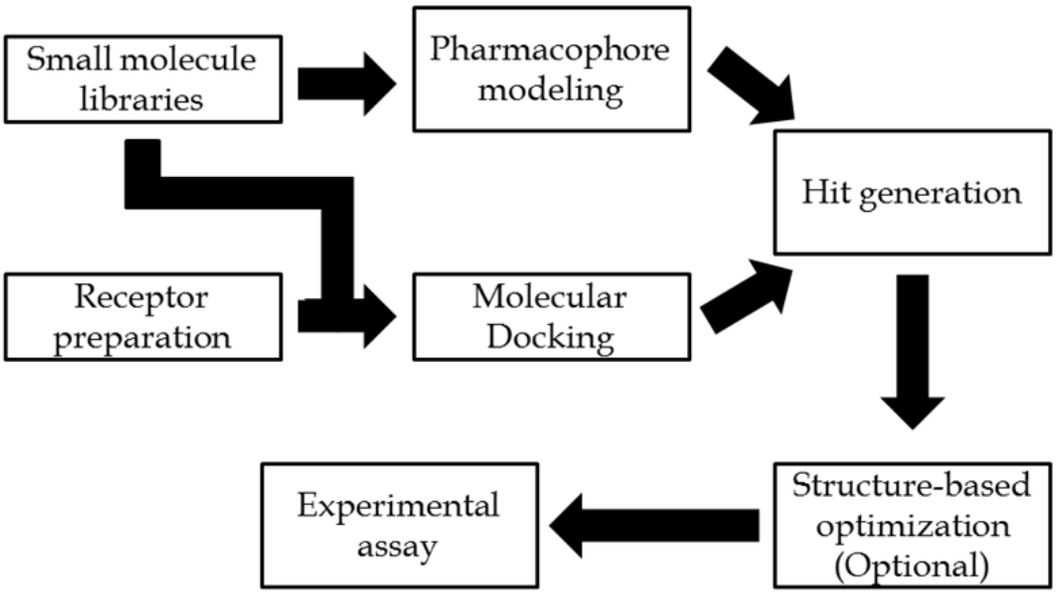


Figure 3. Pipeline of computer-aided drug discovery and lead optimization processes.

4. Molecular docking to discover promoter G-quadruplex stabilizing ligands

In order to drive the development of more potent and selective ligands targeting promoter G-quadruplexes, it is important to understand the detailed interactions between the G-quadruplex and the ligand at the molecular scale. Molecular modelling can provide a tool for visualizing the three-dimensional interactions of the G-quadruplex-ligand complex in order to better understand the structural or functional features required for effective binding. Compared to pharmacophore-based methods, molecular docking can potentially make more effective use of the structural information of the receptor for the discovery of novel G-quadruplex-targeting compounds. In particular, high-quality structural data on the distinctive features of different promoter G-quadruplexes may aid the design and optimization of bioactive ligands that are able to discriminate between related G-quadruplex topologies. In this section, we give a general overview for the *in silico* structure-based discovery of oncogenic promoter G-quadruplex stabilizing ligands.

Computer-aided high-throughput molecular docking and hit validation usually involves three stages (Tang et al. 2006). The first stage is the construction and preparation/selection of the chemical library, and the preparation of the biomolecular model for molecular docking. The second stage is the docking of the individual compounds of the chemical library against the biomolecule, followed by score calculation. In the third stage, the high-scoring compounds can be selected for *in vitro* biological assays to validate their activities towards the biomolecular target.

4.1. Selection of chemical library

A poorly-designed chemical library can result a high rate of false positives, or otherwise poor-quality hits. Therefore, the careful selection of a chemical library containing members possessing favourable pharmacokinetic properties (absorption, distribution, metabolism, excretion, and toxicity; ADMET) or structural diversity could improve the hit rate of a single docking campaign. Today, most chemical libraries are focused in some way by applying a manually selected pre-filter. For example, the Lipinski rule-of-five is a common filter that represents a collection of structural properties correlated with desirable solubility and bioavailability of small molecules (Lipinski et al. 2001). Screening compounds libraries with a pre-filter reduces the likelihood of identifying hit compounds with undesirable ADMET properties, therefore minimizing any loss of investment in chemical synthesis or biological assays. Two types of chemical libraries commonly chosen for virtual screening campaigns are drug/drug-like databases and natural product libraries.

Name	Company	Size of library	URL
ChEMBL	European Bioinformatics Institute (EBI)	>1.1 million	https://www.ebi.ac.uk/chembl/
ZINC	Bioinfomatics and Chemical Informatics Research Center (BCIRC)	>21 million	http://zinc.docking.org/
IBS Database	InterBioScreen Ltd	>45000	http://www.ibscreen.com/natural.shtml
NatDiverse	Analyticon Discovery	>20000	http://www.ac-discovery.com
Super Drug Database	Humboldt-University	~3000	http://bioinf.charite.de/superdrug/
DrugBank	University of Alberta	6711	http://www.drugbank.ca/

Table 2. Examples of commercially available chemical libraries used in computer-aided drug discovery.

Approved drugs usually have favourable or validated pharmacokinetic properties and toxicological profiles, which can improve the hit rate of the screening campaign, and could allow promising hit compounds to potentially bypass early-stage testing, thus streamlining the hit-to-lead optimization process. However, the use of an existing drug library for virtual

screening cannot uncover novel bioactive compounds against the biological target. On the other hand, natural products represent the largest class of compounds in the chemical world. The interactions of natural products with biomolecules have been refined throughout evolutionary timescales, and these unique interactions can be harnessed by medicinal chemists to discover potential drugs. Since most natural products do not strictly adhere to Lipinski rule-of-fives, the virtual screening of natural product libraries can yield novel bioactive scaffolds that could not be obtained from drug-like or combinatorial libraries. Examples of commercially available drug databases and natural product libraries that can be used in high-throughput virtual screening are shown in Table 2.

4.2. Receptor preparation

To construct the receptor model for molecular docking, the atomic coordinates of G-quadruplex solved by the X-ray crystallography and NMR studies with or without bound ligand can usually be retrieved from the Protein Data Bank (Berman et al. 2000) or Nucleic Acid Database (Berman et al. 1996). Generally, structural data obtained from X-ray crystallography is considered more advantageous compared to those from solution NMR studies, as more detailed structural information can be obtained at the atomic scale. For G-quadruplexes lacking hard structural data, a model can be constructed by homology by modification of known, related G-quadruplex structures determined by X-ray crystallography. Commercially available software such as Discovery Studio (Accelrys Inc.) or ICM-Pro (Molsoft) can perform modification of the G-quadruplex conformation or topology through the addition or deletion of nucleobases, addition of monovalent cations in the central ion channel, or modification of the loop length and/or addition of nucleotides in the loop region (Lee et al. 2010).

4.3. G-quadruplex flexibility

The receptor model prepared can then be subjected to local energy minimization to generate the most suitable conformer for subsequent molecular docking analysis. While the small molecule ligands are usually assumed to be flexible so that the binding geometry of the ligand can be corrected predicted, the target is usually assumed to be mostly rigid, as the explicit treatment of receptor flexibility in the docking calculations would be too computationally expensive. Several approaches have been proposed to account for receptor flexibility in virtual screening campaigns. In the case of the G-quadruplex, the flexibility of the loop regions could be important especially for G-quadruplex groove-binding ligands.

An early approach tackling the problem of receptor flexibility was the “soft-docking” method (Jiang et al. 1991). In this approach, the compounds need not fit perfectly to the binding pocket of receptor and a certain degree of steric crash is allowed. During the docking process, the ligand and the receptor adjust their conformations continuously in order to achieve the most suitable conformation with maximum interaction. However, this method only utilizes a single receptor conformation, and thus the choice of receptor model for docking is of the utmost importance.

An alternative strategy that may be useful in G-quadruplex ligand discovery is the use of multiple receptor conformations (MRC) to probe the receptor flexibility (Totrov et al. 2008). This could involve a combination of multiple structures experimentally determined by X-ray crystallography or NMR, or could be generated by molecular stimulation (MD). By considering the different receptor features from multiple conformations, a more representative receptor conformation could be generated for virtual screening. Some modern docking algorithms are able to explicitly model receptor flexibility, but this is usually constrained to the ligand binding domain in order to conserve computing resources. A more thorough discussion of the common approaches used to model receptor flexibility can be found in review articles by Kavraki and co-worker (Teodoro et al. 2003), and Durrant and co-worker (Durrant et al. 2010).

4.4. Global energy optimization

The compounds from the chemical libraries are docked to the receptor structure individually. Generally, assigning the docking site across the entire G-quadruplex structure yields end-stacking compounds as the highest-scoring hits. For discovering groove-binders, which typically display weaker binding affinities, the search area for docking can be limited to the groove or loop regions of the G-quadruplex. Once the compound has been docked into the receptor, most computer algorithms will perform global energy optimization of the small molecule inside the binding pocket to find the most favourable orientation of the small molecule (Abagyan et al. 1994). For example, ICM-Pro (Molsoft) docking software (Abagyan et al. 1997) includes the following steps for global energy optimization:

1. A random conformational change of free variables according to a predefined continuous probability distribution.
2. A local energy minimization of analytical differentiable terms.
3. A calculation of the complete energy including non-differentiable terms such us entropy and solvation energy.
4. An acceptance or rejection of the total energy based on the Metropolis criterion and a return to the first step.

4.5. Score assignment

After the global energy optimization, score assignment is then performed to rank the compounds according to their predicted binding affinities. The score is a qualitative parameter that reflects the binding strength of the compound to the receptor and is composed a collection of factors such as hydrophobic interactions, van der Waal interactions, hydrogen bonding, and electrostatic interactions. However, the accuracy of the docking score will necessarily be limited by the assumptions and approximations of the scoring function. Other factors which may not be explicitly predicted by the computational algorithms, such as solvent environment and binding pocket availability, could also influence the actual binding affinity of the ligand.

Different docking programs may employ different scoring functions, which are generally classified into the following types: 1) force-field functions; 2) knowledge-based scoring functions; and 3) empirical scoring functions (Kitchen et al. 2004). These scoring functions perform calculations that involve different parameters such as statistical potential and weighted interaction terms to rank the apparent potency of the compounds. To improve the accuracy of the scoring assignment, the consensus scoring approach has been investigated. This strategy involves the combination of the weighted scores obtained for a single ligand from different score functions, to improve the hit rate of a docking campaign (Charifson et al. 1999; Clark et al. 2002; Baber et al. 2005; Yang et al. 2005).

5. Structure-based lead optimization

In the conventional drug discovery, validation of a screening hit by *in vitro* assays is usually followed by the synthesis of a range of structurally related analogues in order to optimize the binding and selectivity of the ligand towards the target. However, this approach necessarily entails a significant investment into manpower and materials, and can be very time-consuming. An alternative strategy utilizes the principles of computer-aided structure-based design in order to achieve the more efficient allocation of resources towards analogues with higher predicted binding affinities. By analysis of the receptor-ligand complex determined using X-ray crystallography or molecular modelling, a library of derivatives can be generated *in silico* that retain the important features of hit ligand that contribute to high binding affinity. This focused library can then undergo a second round of molecular docking procedure to identify the most promising derivatives for synthesis and evaluation.. The application of *in silico* structure-based optimization has also been applied for the development of oncogenic promoter G-quadruplex stabilizing ligands.

6. Discovery of oncogenic promoter G-quadruplex-stabilizing ligands using structure-based approaches

The use of *in silico* virtual screening to discover promoter G-quadruplex-stabilizing ligands has only been recently reported. Tang and co-workers utilized ligand-based pharmacophore modelling techniques to identify two non-planar alkaloids as groove binders of the parallel G-quadruplex (Li et al. 2009). In their report, the representative pharmacophore was constructed using the CATALYST software package (version 4.11, Accelrys Inc.) (Nicklaus et al. 1997). A total of 38 1,4-disubstituted anthraquinone derivatives comprised the training set, with IC_{50} values against rat glioma C6 cells spanning three orders of magnitudes (from 0.07 mM to 103 mM). Ten hypothetical models were constructed using the HypoGen hypothesis process, with the best pharmacophore containing one hydrogen bond receptor, one hydrogen bond donor, one positive ionizable group and two hydrophobic sites. The best pharmacophore model was selected for virtual screening and was mapped against a natural product database containing *ca.* 10,000 compounds derived from Chinese herbal medicines. A total of 176 hit compounds were identified with a diversity of scaffolds different to those of the training set, and 20 compounds were chosen for further evaluation

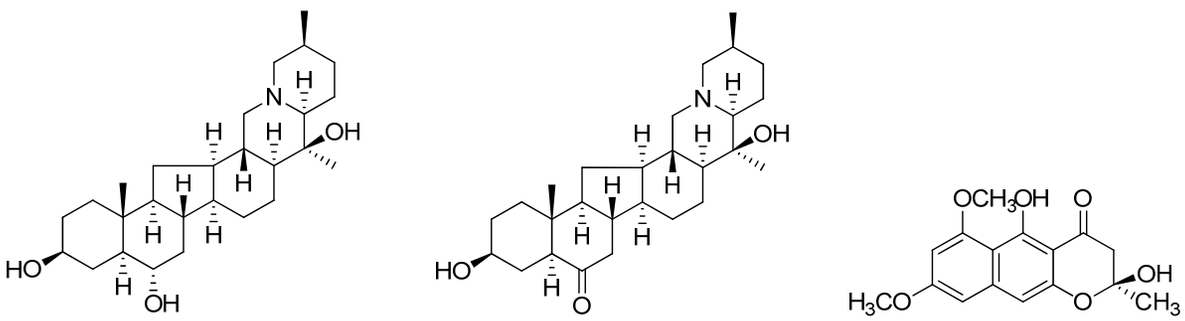
based on compound availability. Intriguingly, the hit compounds included two neutral non-planar compounds, peimine (**1**) and peimimine (**2**). In UV melting experiments, peimine (**2**) and peimimine (**3**) were found to stabilize the tetramolecular G-quadruplex motif with significant increases in T_m. Further experiments indicated that both compounds were selective for parallel G-quadruplexes, and did not stabilize other G-quadruplex topologies or duplex DNA. Circular dichroism (CD) experiments found that compound **1** was able to enhance the characteristic parallel G-quadruplex CD signal at 262 nm of all the parallel G-quadruplexes examined, including c-*kit* oncogenic promoter G-quadruplex. The study from Tang and co-workers demonstrated the feasibility of employing ligand-based pharmacophore modelling to identity novel oncogenic promoter G-quadruplex-stabilizing compounds. However, further research would be required to fully characterize the possible biological effects of the compound in living cells.

In 2010, our group has employed high-throughput virtual screening techniques to identify fonsecin B (**3**) as a c-*myc* G-quadruplex stabilizer (Lee et al. 2010). Since no X-ray structure of the c-*myc* G-quadruplex was available, a molecular model of the predominant 1:2:1 loop isomer of c-*myc* G-quadruplex was constructed using the X-ray crystal structure of the related intramolecular human telomeric G-quadruplex. The model was built by the insertion or deletion of nucleobases and modification of the loop size to correspond to the 1:2:1 loop isomer of the c-*myc* G-quadruplex (Ou et al. 2007) using ICM-Pro (Molsoft). After the preparation of the receptor model, over 20,000 compounds from a natural product library were docked against the molecular model using the Molsoft ICM-Pro (3.6.1 d) docking protocol. Since most G-quadruplex stabilizing ligands possess a large polyaromatic scaffold for end-stacking, the docking area to the termini of the G-quadruplex was restricted to avoid the wastage of computational time. From the results of the virtual screening campaign, four hits were identified and tested in a preliminary *in vitro* PCR stop assay to assess their abilities to stabilize the c-*myc* G-quadruplex, and fonsecin B (**3**) emerged as the top candidate.

A variety of experiments were performed to analyze the interaction and selectivity of fonsecin B towards the c-*myc* G-quadruplex. UV-visible absorption spectroscopy revealed that compound **3** displayed 5.5-fold and 16.5-fold higher binding affinities for the c-*myc* G-quadruplex over duplex and single-stranded DNA, respectively. We then performed a detailed molecular modelling experiment in order to investigate the binding mode of the compound to the c-*myc* G-quadruplex. The modelling results revealed that **3** was stacked against the 3′-terminal of G-quadruplex with a binding energy of –48 kcal mol^{-1}. The phenolic and carbonyl oxygen atoms were predicted to orientate towards the central ionic channel, where the two oxygen atoms could possibly be stabilized by electrostatic interactions with the potassium ion. By comparison, intercalation of **3** into the G-quadruplex was calculated to be extremely unfavourable, with a binding energy of *ca.* 25 kcal mol^{-1}. PCR stop assays showed that **3** was able to stabilize the c-*myc* G-quadruplex with the similar potency to the well-known G-quadruplex ligand TmPyP4.

Apart from the high-throughput virtual screening of chemical libraries, structure-based optimization by *in silico* approaches have also been employed to improve the potency of the

lead compounds to a particular oncogenic promoter G-quadruplex target. In 2009, Che and co-workers developed a series of Pt(II) complexes as c-*myc* G-quadruplex stabilizing ligands using an *in silico* structure-based optimization strategy (Wu et al. 2009). Among a series of Pt(II)-salphen complexes tested in preliminary *in vitro* assays, complex **4** was found to be most potent and was chosen for *in silico* structural modification. Over 60 derivatives of complex **4** were designed that contained side chains with various lengths and functional groups to interact with the grooves of the G-quadruplex, and these compounds were docked to the c-*myc* G-quadruplex using the ICM program. In the molecular docking analysis, the highest scoring compound **5** was found to bind more favorably to the c-*myc* G-quadruplex compared to the parent complex **4** due to the additional interactions between the side chains of **5** with the G-quadruplex grooves regions. Compound **5** was then synthesized for biological evaluation, and the PCR stop assay results showed **5** could stabilize the formation of the c-*myc* G-quadruplex with an IC_{50} value of 4.4 μM, which was an order lower than that of parent compound **4**. In this report, Che and co-workers successfully demonstrate the use of structure-based optimization of a Pt(II)-salphen complex to devise a more promising scaffold for stabilization of the c-*myc* oncogenic promoter G-quadruplex.



1 **2** **3**

Figure 4. Structures of promoter G-quadruplex-targeting compounds discovered via high-throughput virtual screening.

Later, the Che group reported another successful application of computer-based lead optimization of Pt(II) metal complexes to discover efficient c-*myc* G-quadruplex stabilizing ligands (Wang et al. 2010). Based on hit complex **6**, over 550 derivatives were designed by attaching side chains of various lengths and functionality to the parent scaffold, the library of compounds were rapidly screened *in silico.* Three of the highest scoring complexes **7–9** were then synthesized and subjected to comprehensive *in vitro* assays to evaluate their ability to stabilize the c-*myc* G-quadruplex. In the UV-Vis absorption experiments, all three complexes showed at least 10-fold higher binding affinities towards the *c-myc* G-quadruplex over duplex DNA. Furthermore, the complexes increased the T_m of the c-*myc* G-quadruplex by over 9 °C, and displayed improved potency at stabilizing the c-*myc* G-quadruplex in the PCR stop assay when compared to the parent compound. Subsequent reverse transcriptase PCR (RT-PCR) experiments showed that the mRNA level of the c-*myc* gene could be significantly diminished in the presence of complexes **7–9**, suggesting that these compounds could be used as suppressors of oncogenic expressing in living cells. This report by Che and

co-workers again demonstrated the feasibility of *in silico* structure-based lead optimization of metal complexes, and suggested that the use of a larger chemical library of derivatives could generate a larger diversity of hits with potentially improved potencies.

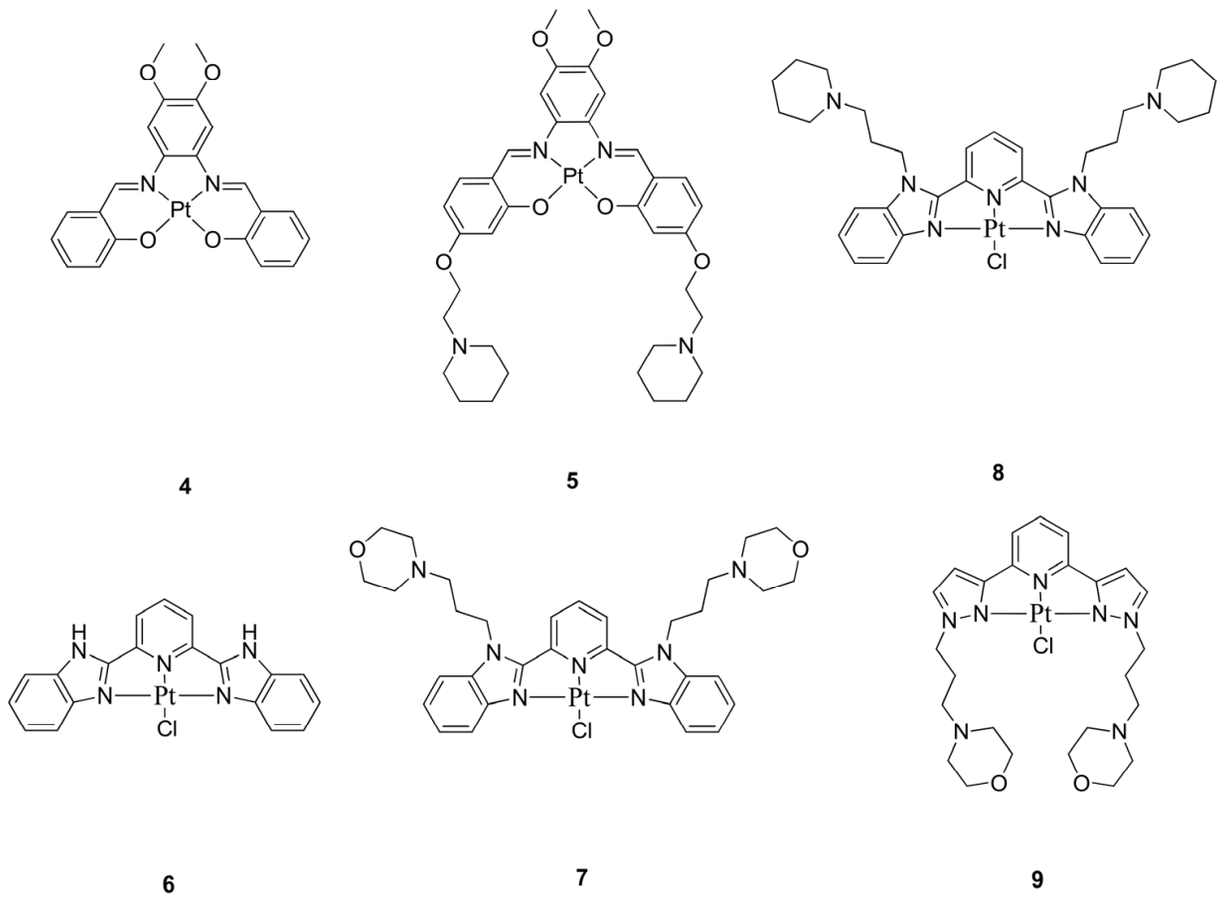


Figure 5. Chemical structures of the platinum(II) complexes discovered through *in silico* structure-based optimization as c-*myc* oncogenic G-quadruplex stabilizing ligands.

Our group has recently reported the structural-based optimization of FDA-approved drug methylene blue (MB) to generate more potent analogues as c-*myc* G-quadruplex stabilizers (Chan et al. 2011). Over 3,000 FDA-approved drugs were screened *in silico* against the 1:2:1 loop isomer model of the c-*myc* G-quadruplex developed by our group, and MB emerged as the top candidate. Although the MB is a well-known DNA intercalator and has been previously reported to bind the G-quadruplex, its application as a c-*myc* oncogenic promoter G-quadruplex stabilizer was first discovered by our group. 50 MB derivatives were designed *in silico* and were docked against the c-*myc* G-quadruplex using ICM-Pro software. Compounds **10a–c** bearing a bromophenyl pendant linked by an aliphatic side chain showed the greatest binding energy from the virtual screening, and they were synthesized for biological evaluation. In the fluorescence intercalator displacement (FID) assay, compound **10b** was found to effectively displace thiazole orange (TO) from the c-*myc* G-quadruplex with a DC_{50} value of 0.75 μM, while compounds **10a** and **10c** displayed higher DC_{50} values of *ca.* 6 and 2 μM, respectively. Furthermore, compound **10b** could inhibit *Taq*

polymerase mediated-extension of the c-*myc* sequence through induction of the G-quadruplex structure in the PCR stop assay with superior potency compared to the parent compound MB. Detailed molecular docking analysis revealed that compound **10b** was predicted to form strong end-stacking interactions with the terminal of c-*myc* G-quadruplex with groove interactions, whereas the parent compound MB was predicted to interact with the G-quadruplex *via* a mostly intercalative mode. In living cells, compound **10b** was shown to be effectively down-regulate the c-*myc* promoter activity with an IC_{50} value of *ca.* 1 μM as revealed by a luciferase assay. The increased activity of the **10b** compared to MB against c-*myc* promoter activity could be potentially attributed, at least in part, to the stabilization of c-*myc* G-quadruplex structure. This report demonstrated the structure-based lead optimization approach effectively generate novel analogues of existing drug as oncogenic G-quadruplex stabilizing ligands.

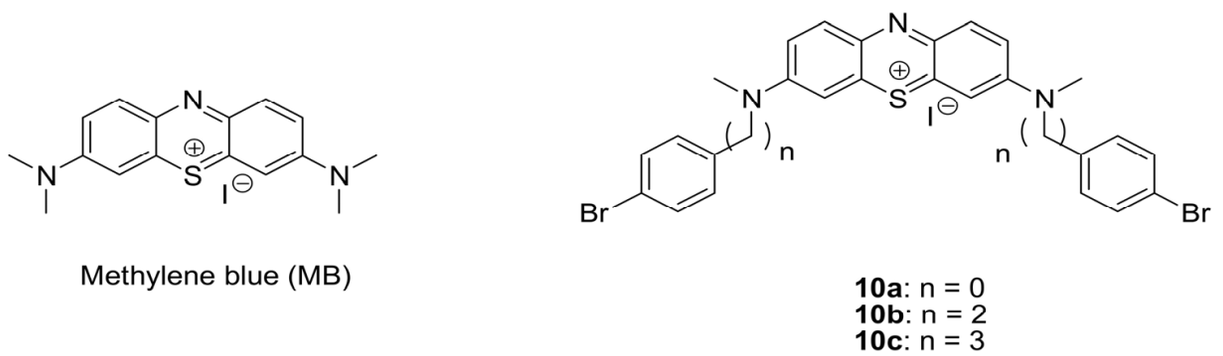


Figure 6. Chemical structures of the FDA-approved drug methylene blue and its analogues designed by a structural-based optimization strategy as c-*myc* oncogenic G-quadruplex stabilizing ligands.

7. Conclusion

The identification of oncogenes involved in the progression of various types of tumours has stimulated the development of various anti-cancer strategies targeting oncogenic expression. The discovery of G-quadruplex motifs in the promoter regions of oncogenes and the elucidation of their putative roles in the regulation of oncogenic transcription has opened a new potential therapeutic avenue for the treatment of cancer. However, it should be noted that the application of G-quadruplex-stabilizing ligands for the modulation of oncogenic activity in living systems is still in its infancy. Most promoter quadruplex ligands discovered thus far have not yet progressed past pre-clinical investigation. To advance further, several important criteria have to be addressed. These include the bioavailability of G-quadruplex-binding compounds as well their conformational rigidity and promiscuity for other physiological targets. In particular, the action of the lead candidates against the large number of other gene promoters and G-quadruplex structures that are likely to be present in normal cells should be rigorously assessed. These factors would aid in the determination of the permissible dosage and therapeutic window of the G-quadruplex-targeting compounds for the potential treatment of cancer. With continual advances in computational technologies and modelling techniques, as well as the concurrent development of more

focused yet diverse chemical libraries, we envisage that the discovery and investigation of novel promoter G-quadruplex-stabilizing ligands would continue to thrive in the near future. Furthermore, *in silico* hit-to-lead optimization allows the chemical space around hit compounds to be explored without necessitating the actual synthesis of analogue molecules, thus significantly reducing expenses associated with materials and manpower.

Author details

Dik-Lung Ma*, Victor Pui-Yan Ma, Ka-Ho Leung, Hong-Zhang He
and Daniel Shiu-Hin Chan
Department of Chemistry, Hong Kong Baptist University, Kowloon Tong, Hong Kong

Hai-Jing Zhong and Chung-Hang Leung*
State Key Laboratory of Quality Research in Chinese Medicine, Institute of Chinese, Medical Science, University of Macau, Macao, China

Acknowledgement

This work is supported by Hong Kong Baptist University (FRG2/11-12/009), Environment and Conservation Fund (ECF Project 3/2010), Centre for Cancer and Inflammation Research, School of Chinese Medicine (CCIR-SCM, HKBU), the Health and Medical Research Fund (HMRF/11101212), the Research Grants Council (HKBU/201811 and HKBU/204612), the Science and Technology Development Fund, Macao SAR (001/2012/A) and the University of Macau (SRG013-ICMS12-LCH, MYRG091(Y1-L2)-ICMS12-LCH and MYRG121(Y1-L2)-ICMS12-LCH).

8. References

Abagyan, R. and M. Totrov (1994). "Biased Probability Monte Carlo Conformational Searches and Electrostatic Calculations for Peptides and Proteins." *J Mol Biol* 235: 983-1002.

Abagyan, R. and M. Totrov (1997). "Flexible protein-ligand docking by global energy optimization in internal coordinates." *Proteins* 29: 215-220.

Alderden, R. A., M. D. Hall and T. W. Hambley (2006). "The Discovery and Development of Cisplatin." *J Chem Educ* 83: 728.

Baber, J. C., W. A. Shirley, Y. Gao and M. Feher (2005). "The Use of Consensus Scoring in Ligand-Based Virtual Screening." *J Chem Inf Model* 46: 277-288.

Balasubramanian, S., L. H. Hurley and S. Neidle (2011). "Targeting G-quadruplexes in gene promoters: a novel anticancer strategy?" *Nat Rev Drug Discov* 10: 261-275.

Balasubramanian, S. and S. Neidle (2009). "G-quadruplex nucleic acids as therapeutic targets." *Curr Opin Chem Biol* 13: 345-353.

* Corresponding Authors

Bejugam, M., S. Sewitz, P. S. Shirude, R. Rodriguez, R. Shahid and S. Balasubramanian (2007). "Trisubstituted Isoalloxazines as a New Class of G-Quadruplex Binding Ligands: Small Molecule Regulation of c-kit Oncogene Expression." *J Am Chem Soc* 129: 12926-12927.

Berman, H. M., A. Gelbin and J. Westbrook (1996). "Nucleic acid crystallography: A view from the nucleic acid database." *Prog Biophys Mol Biol* 66: 255-288.

Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, et al. (2000). "The Protein Data Bank." *Nucleic Acids Res* 28: 235-242.

Blackburn, E. H. (1991). "Structure and function of telomeres." *Nature* 350: 569-573.

Burge, S., G. N. Parkinson, P. Hazel, A. K. Todd and S. Neidle (2006). "Quadruplex DNA: sequence, topology and structure." *Nucleic Acids Res* 34: 5402-5415.

Cavasotto, C. and A. Orry (2007). "Ligand docking and structure-based virtual screening in drug discovery." *Curr Top Med Chem* 7: 1006-1014.

Chan, D. S.-H., H. Yang, M. H.-T. Kwan, Z. Cheng, P. Lee, L.-P. Bai, et al. (2011). "Structure-based optimization of FDA-approved drug methylene blue as a c-myc G-quadruplex DNA stabilizer." *Biochimie* 93: 1055-1064.

Charifson, P. S., J. J. Corkery, M. A. Murcko and W. P. Walters (1999). "Consensus Scoring: A Method for Obtaining Improved Hit Rates from Docking Databases of Three-Dimensional Structures into Proteins." *J Med Chem* 42: 5100-5109.

Chen, J. and L. Lai (2006). "Pocket v.2: Further Developments on Receptor-Based Pharmacophore Modeling." *J Chem Inf Model* 46: 2684-2691.

Clark, R. D., A. Strizhev, J. M. Leonard, J. F. Blake and J. B. Matthew (2002). "Consensus scoring for ligand/protein interactions." *J Mol Graph Model* 20: 281-295.

Cogoi, S. and L. E. Xodo (2006). "G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription." *Nucleic Acids Res* 34: 2536-2549.

Croce, C. M. (2008). "Oncogenes and Cancer." *New Engl J Med* 358: 502-511.

Dai, J., D. Chen, R. A. Jones, L. H. Hurley and D. Yang (2006). "NMR solution structure of the major G-quadruplex structure formed in the human BCL2 promoter region." *Nucleic Acids Res* 34: 5133-5144.

Davis, T. L., A. B. Firulli and A. J. Kinniburgh (1989). "Ribonucleoprotein and protein factors bind to an H-DNA-forming c-myc DNA element: possible regulators of the c-myc gene." *Proc Natl Acad Sci USA* 86: 9682-9686.

Dexheimer, T. S., S. S. Carey, S. Zuohe, V. M. Gokhale, X. Hu, L. B. Murata, et al. (2009). "NM23-H2 may play an indirect role in transcriptional activation of c-myc gene expression but does not cleave the nuclease hypersensitive element III1." *Mol Cancer Ther* 8: 1363-1377.

Duan, W., A. Rangan, H. Vankayalapati, M.-Y. Kim, Q. Zeng, D. Sun, et al. (2001). "Design and Synthesis of Fluoroquinophenoxazines That Interact with Human Telomeric G-Quadruplexes and Their Biological Effects1." *Mol Cancer Ther* 1: 103-120.

Durrant, J. D. and J. A. McCammon (2010). "Computer-aided drug-discovery techniques that account for receptor flexibility." *Curr Opin Pharmacol* 10: 770-774.

Ghosh, S., A. Nie, J. An and Z. Huang (2006). "Structure-based virtual screening of chemical libraries for drug discovery." *Curr Opin Chem Biol* 10: 194-202.

Gonzalez, V., K. Guo, L. Hurley and D. Sun (2009). "Identification and Characterization of Nucleolin as a c-myc G-quadruplex-binding Protein." *J Biol Chem* 284: 23622-23635.

Grand, C. L., H. Han, R. M. Muñoz, S. Weitman, D. D. Von Hoff, L. H. Hurley, et al. (2002). "The Cationic Porphyrin TMPyP4 Down-Regulates c-MYC and Human Telomerase Reverse Transcriptase Expression and Inhibits Tumor Growth in Vivo 1 This research was supported by grants from the NIH and the Arizona Disease Control Research Commission.1." *Mol Cancer Ther* 1: 565-573.

Heidenreich, O. (2009). Targeting Oncogenes with siRNAs. siRNA and miRNA Gene Silencing. M. Sioud, Humana Press. 487: 221-242.

Huppert, J. L. and S. Balasubramanian (2007). "G-quadruplexes in promoters throughout the human genome." *Nucleic Acids Res* 35: 406-413.

Jiang, F. and S.-H. Kim (1991). "“Soft docking”: Matching of molecular surface cubes." *J Mol Biol* 219: 79-102.

Jung, Y. and S. J. Lippard (2007). "Direct Cellular Responses to Platinum-Induced DNA Damage." *Chem Rev* 107: 1387-1407.

Kitchen, D. B., H. Decornez, J. R. Furr and J. Bajorath (2004). "Docking and scoring in virtual screening for drug discovery: methods and applications." *Nat Rev Drug Discov* 3: 935-949.

Lee, H.-M., D. S.-H. Chan, F. Yang, H.-Y. Lam, S.-C. Yan, C.-M. Che, et al. (2010). "Identification of natural product Fonsecin B as a stabilizing ligand of c-myc G-quadruplex DNA by high-throughput virtual screening." *Chem Commun* 46: 4680-4682.

Li, Q., J. Xiang, X. Li, L. Chen, X. Xu, Y. Tang, et al. (2009). "Stabilizing parallel G-quadruplex DNA by a new class of ligands: Two non-planar alkaloids through interaction in lateral grooves." *Biochimie* 91: 811-819.

Lipinski, C. A., F. Lombardo, B. W. Dominy and P. J. Feeney (2001). "Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings." *Adv Drug Deliver Rev* 46: 3-26.

Lipps, H. J. and D. Rhodes (2009). "G-quadruplex structures: in vivo evidence and function." *Trends Cell Biol* 19: 414-422.

Lu, Y.-J., T.-M. Ou, J.-H. Tan, J.-Q. Hou, W.-Y. Shao, D. Peng, et al. (2008). "5-N-Methylated Quindoline Derivatives as Telomeric G-Quadruplex Stabilizing Ligands: Effects of 5-N Positive Charge on Quadruplex Binding Affinity and Cell Proliferation." *J Med Chem* 51: 6381-6392.

Lutz, W., J. Leon and M. Eilers (2002). "Contributions of Myc to tumorigenesis." *BBA-Rev Cancer* 1602: 61-71.

Ma, D.-L., V. P.-Y. Ma, D. S.-H. Chan, K.-H. Leung, H.-J. Zhong and C.-H. Leung (2012). "In silico screening of quadruplex-binding ligands." *Methods* 57: 106-114.

Ma, Y., T.-M. Ou, J.-Q. Hou, Y.-J. Lu, J.-H. Tan, L.-Q. Gu, et al. (2008). "9-N-Substituted berberine derivatives: Stabilization of G-quadruplex DNA and down-regulation of oncogene c-myc." *Bioorg Med Chem* 16: 7582-7591.

Marcu, K. B., S. A. Bossone and A. J. Patel (1992). "myc Function and Regulation." *Annu Rev Biochem* 61: 809-858.

McInnes, C. (2007). "Virtual screening strategies in drug discovery." *Curr Opin Chem Biol* 11: 494-502.

Mergny, J.-L. and C. Helene (1998). "G-quadruplex DNA: A target for drug design." *Nat Med* 4: 1366-1367.

Meyer, N. and L. Z. Penn (2008). "Reflecting on 25 years with MYC." *Nat Rev Cancer* 8: 976-990.

Monchaud, D. and M.-P. Teulade-Fichou (2008). "A hitchhiker's guide to G-quadruplex ligands." *Org Biomol Chem* 6: 627-636.

Nicklaus, M. C., N. Neamati, H. Hong, A. Mazumder, S. Sunder, J. Chen, et al. (1997). "HIV-1 Integrase Pharmacophore: Discovery of Inhibitors through Three-Dimensional Database Searching†." *J Med Chem* 40: 920-929.

Ou, T.-M., Y.-J. Lu, C. Zhang, Z.-S. Huang, X.-D. Wang, J.-H. Tan, et al. (2007). "Stabilization of G-Quadruplex DNA and Down-Regulation of Oncogene c-myc by Quindoline Derivatives." *J Med Chem* 50: 1465-1474.

Parkinson, G. N., M. P. H. Lee and S. Neidle (2002). "Crystal structure of parallel quadruplexes from human telomeric DNA." *Nature* 417: 876-880.

Rankin, S., A. P. Reszka, J. Huppert, M. Zloh, G. N. Parkinson, A. K. Todd, et al. (2005). "Putative DNA Quadruplex Formation within the Human c-kit Oncogene." *J Am Chem Soc* 127: 10584-10589.

Seenisamy, J., E. M. Rezler, T. J. Powell, D. Tye, V. Gokhale, C. S. Joshi, et al. (2004). "The Dynamic Character of the G-Quadruplex Element in the c-MYC Promoter and Modification by TMPyP4." *J Am Chem Soc* 126: 8702-8709.

Shoichet, B. K. (2004). "Virtual screening of chemical libraries." *Nature* 432: 862-865.

Siddiqui-Jain, A., C. L. Grand, D. J. Bearss and L. H. Hurley (2002). "Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription." *Proc Natl Acad Sci USA* 99: 11593-11598.

Tang, Y., W. Zhu, K. Chen and H. Jiang (2006). "New technologies in computer-aided drug design: Toward target identification and new chemical entity discovery." *Drug Discov Today Technol* 3: 307-313.

Teodoro, M. L. and L. E. Kavraki (2003). "Conformational flexibility models for the receptor in structure based drug design." *Curr Pharm Des* 9: 1635-1648.

Totrov, M. and R. Abagyan (2008). "Flexible ligand docking to multiple receptor conformations: a practical alternative." *Curr Opin Struct Biol* 18: 178-184.

Ventura, A. and T. Jacks (2009). "MicroRNAs and Cancer: Short RNAs Go a Long Way." *Cell* 136: 586-591.

Wang, P., C.-H. Leung, D.-L. Ma, S.-C. Yan and C.-M. Che (2010). "Structure-Based Design of Platinum(II) Complexes as c-myc Oncogene Down-Regulators and Luminescent Probes for G-Quadruplex DNA." *Chem Eur J* 16: 6900-6911.

Wheelhouse, R. T., D. Sun, H. Han, F. X. Han and L. H. Hurley (1998). "Cationic Porphyrins as Telomerase Inhibitors: the Interaction of Tetra-(N-methyl-4-pyridyl)porphine with Quadruplex DNA." *J Am Chem Soc* 120: 3261-3262.

Wierstra, I. and J. Alves (2008). The c-myc Promoter: Still MysterY and Challenge. Advances in Cancer Research. F. V. W. George and K. George, Academic Press. Volume 99: 113-333.

Wolber, G. and T. Langer (2004). "LigandScout: 3-D Pharmacophores Derived from Protein-Bound Ligands and Their Use as Virtual Screening Filters." *J Chem Inf Model* 45: 160-169.

Wu, P., D.-L. Ma, C.-H. Leung, S.-C. Yan, N. Zhu, R. Abagyan, et al. (2009). "Stabilization of G-Quadruplex DNA with Platinum(II) Schiff Base Complexes: Luminescent Probe and Down-Regulation of c-myc Oncogene Expression." *Chem Eur J* 15: 13008-13021.

Yang, J.-M., Y.-F. Chen, T.-W. Shen, B. S. Kristal and D. F. Hsu (2005). "Consensus Scoring Criteria for Improving Enrichment in Virtual Screening." *J Chem Inf Model* 45: 1134-1146.

IntechOpen