Published in final edited form as:
Exp Mol Pathol. 2009 June ; 86(3): 141–150. doi:10.1016/j.yexmp.2009.01.011.

Structure-based Drug Design: From Nucleic Acid to Membrane Protein Targets

Magdalena M. Dailey[1], **Chayanendu Hait**[2], **Patrick A. Holt**[3], **Jon M. Maguire**[2], **Jason B. Meier**[3], **M. Clarke Miller**[2], **Luigi Petraccone**[4], and **John O. Trent**[1,2,3,5]

[1] Department of Chemistry, Belknap Research Building, University of Louisville, Louisville, KY 40292

[2] James Graham Brown Cancer Center, 529 South Jackson Street, Louisville KY 40202

[3] Department of Biochemistry and Molecular Biology, University of Louisville, Louisville, Kentucky 40292

[4] Dept. Chimica “P. Corradini”, University of Naples “Federico II”, 80122 Naples, Italy

[5] Department of Medicine, University of Louisville, Louisville KY 40202

Abstract

The *in silico* methods for drug discovery are becoming increasingly powerful and useful. That, in combination with increasing computer processor power, in our case using a novel distributed computing grid, has enabled us to greatly enhance our virtual screening efforts. Herein we review some of these efforts using both receptor and ligand-based virtual screening, with the goal of finding new anticancer agents. In particular, nucleic acids are a neglected set of targets, especially the different morphologies of duplex, triplex, and quadruplex DNA, many of which have increasing biological relevance. We also review examples of molecular modeling to understand receptors and using virtual screening against G-protein coupled receptor membrane proteins.

Keywords

Virtual screening; drug discovery; membrane protein; G-protein coupled receptor; telomere; quadruplex; DNA

1.0 Introduction

There have been significant advances in the prevention and treatment of cancer in the last decade; however, the fact remains that the lifetime risk of developing a malignancy is still very high. In the United States it is one chance in three for women and one chance in two for men (www.cancer.org). One hurdle that new treatments face is the lengthy time for the necessary clinical trials to be completed before these new agents are openly available. There are several issues with current treatments, some of which are based on decades old drugs.

Corresponding author: J.O. Trent, James Graham Brown Cancer Center, 529 South Jackson Street, Louisville KY 40202, john.trent@louisville.edu, 502 852 2194.

Conflict of Interest: The authors declare that there are no conflicts of interest.

Most common chemotherapy agents are selective toxins and are more active on rapidly proliferating cells, including many normal epithelial and hematopoietic cells. This leads to extensive nonspecific side effects. The fundamental problem is that the drugs are not effective on all forms of cancer, in part because cancer is not a single disease. Also, the basic mechanisms of action of these agents on particular molecular targets are poorly understood, or they are simply not directed to a particular target. Nevertheless, many drugs in clinical use do relatively more good than harm.

The James Graham Brown Cancer Center has made translation of its own basic science to develop new therapies a major focus of its research; we have accordingly established numerous collaborations between the Molecular Targets and Structural/Computational Biology groups to facilitate this central goal.

2.0 Can We Discover And Develop New Anticancer Therapeutics?

There are several examples mentioned in this issue of anti-cancer agents that originated from the Brown Cancer Center and are in clinical trials. The initial drug with which we were involved was the precursor to AS1411 (www.antisoma.com, and see Bates *et al.* in this issue), a first in class anticancer agent discovered and initially developed by Paula J. Bates, Donald M. Miller and John O. Trent (Bates *et al.*, 1999; Dapic *et al.*, 2003; Dapic *et al.*, 2002; Girvan *et al.*, 2006; Teng *et al.*, 2007; Xu *et al.*, 2001). The Phase I clinical trial of AS1411 started in 2003 at the Brown Cancer Center for solid tumors. Antisoma started two Phase II trials in acute myeloid leukemia (AML), and renal cell carcinoma in 2008.

3.0 Can We Find Additional New Potential Anticancer Agents?

After the initial success of AS1411, we turned to discovering small molecule inhibitors, initially against the AS1411 target. This review details one of our approaches using *in silico* prescreening (computational virtual screening) for the enrichment of candidates to be experimentally evaluated. The two major advantages of this drug discovery approach is the time to discover lead templates (we use the nomenclature “lead template” because the initial hits will need to be subsequently optimized), and the costs are both significantly lower than traditional methods. It has been estimated that 10% of current drugs have been found using such techniques and that may increase to 20% by 2010 (Kapetanovic, 2008). It should be noted that no matter how you discover a lead template the same issues of optimization, ADME-toxicity, and safety apply.

3.1 Initial Success With Virtual Screening

3.1.1. Nucleolin—Our initial foray into virtual screening was to target the protein nucleolin, the presumptive target of AS1411. Paula Bates demonstrated (Dapic *et al.*, 2003) a strong correlation of the activity of G-rich oligonucleotides, like AS1411, with binding to nucleolin. We developed a homology model, using Modeller (Sali *et al.*, 1993), of human nucleolin based on the NMR-derived structure of the hamster ortholog (Allain *et al.*, 2000) and docked one theoretical model of AS1411 onto nucleolin using GRAMM (Vakser, 1997). We then used that binding site to target nucleolin using the Ludi (Accelrys) virtual screening program to process the iResearch library (Chemnavigator). Using EMSA’s and MTT assays, we have found two potential nucleolin inhibitors (unpublished work). With the use of GRID computer system (see later) and Surflex-dock (Jain, 2003) we were able to search over 3,500,000 small molecules that were commercially available using the ZINC library (Irwin *et al.*, 2005). We chose 32 compounds with the highest Surflex-Dock docking score to be tested in molecular biology experiments. In MTT antiproliferative assays we found six of these compounds significantly inhibited the growth of lung (A549), prostate (U937) and cervical (HeLa) cancer cells.

***3.1.2. PFKFB3 (*6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase*)*—**In collaboration with the Chesney group at the Brown Cancer Center we targeted PFKFB3, which is described in detail in the chapter by Yalcin et al., in this issue. One of the best-characterized activities of cancer cells is the preferential use of glycolysis for energy production. This phenomenon, known as the Warburg effect (Warburg, 1956; Warburg *et al.*, 1924), allows cancer cells to produce ATP and NADH by lactic acid fermentation, even under normoxic conditions. One potential mechanism explaining the Warburg effect in cancer cells involves the ras signaling pathway producing increased levels of fructose 2,6 bisphosphate (F26B) (Kole *et al.*, 1991; Mazurek *et al.*, 2001; Ramanathan *et al.*, 2005). F26BP interaction with PFK1 shifts PFK1 into its high affinity state for F6P, allowing glycolysis to move forward past the irreversible PFK1 step (Van Schaftingen E *et al.*, 1981). F2,6BP levels in the cell are regulated by the bifunctional enzyme 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase (PFKFBP), a four gene family (PFKFB1–4) which can convert F6P to F2,6BP or catalyze the reverse reaction and convert F2,6BP to F6P (Okar *et al.*, 1999). Telang *et al*. sequentially deleted PFKFBPases in ras-transformed mouse lung fibroblasts and identified PFKFB3 as the isoform most closely linked with ras activation of the glycolytic pathway in oncogenic cells (Telang *et al.*, 2006). PFKFB3 is an inducible PFKFBPase characterized by upregulation in inflammation and hypoxia (Chesney *et al.*, 1999), as well as overexpression in several cancer types. PFKFB3's strong role in the ras-mediated glycolytic activity of cancer cells, broad expression pattern, and the inducible nature of the target made this enzyme a very likely target for new chemotherapy agents.

Antineoplastic agents targeting PFKFB3 were generated using a molecular modeling and virtual screening approach. Homology modeling was used to generate a structure for PFKFB3 based on the X-ray structure of rat testes PFKFB4 (PDB code 1BIF). Clustal W was used to align the PFKFB4 and PFKFB3 sequences and 4 homology models were generated from the aligned PFKFB3 sequence using Modeller (Sali *et al.*, 1993). The structure that best reproduced the PFKFB3 binding site (Chesney *et al.*, 1999) was used for molecular docking.

The PFKFB3 model was read into InsightII (Accelrys) and residues essential to the active site were correlated with the aligned sequence. Three of these residues, Arg66, Tyr161, and Thr94, were selected as centroid targets for virtual screening using the Ludi (Accelrys) virtual screening program to process the chemnavigator iResearch library (www.chemnavigator.com). Compounds that Ludi scored above 500 were analyzed by visual inspection in the PFKFB3 active site, the 200 highest scoring molecules were identified for purchase using Scifinder Scholar (www.cas.org), and the top 45 compounds were selected for experimental assays. Thirteen of these molecules were purchased commercially and tested for activity. One compound, 3-(3-pyridinyl)-1-(4-pyridinyl)-2-propen-1-one (3PO) was found to suppress glycolysis and to be cytostatic to cancer cells (Clem *et al.*, 2008). 3PO was found to inhibit recombinant PFKFB3, decrease glucose uptake and depress the concentrations of several mediators of glycolysis in cancer cells such as F2,6BP, lactate, ATP, NAD^+ and NADH. 3PO reduced the proliferation of several cancer cell lines with IC_{50} values ranging from 1.4–24 μM. 3PO also was selectively cytostatic to ras-transformed cell lines and inhibited tumor growth.

In the case of 3PO, examination of the PFKFB3 substrate-binding site revealed a long, tunnel-like pocket with separate regions corresponding to different aspects of the PFKFB3/F6P interaction and regions that were not involved with the interaction at all (Figure 1, left panel). So, rather than expanding the search space to encompass the full active site, as we did with MIF (see below), we needed to narrow the search space for PFKFB3. To do this, we used the active residues in the F6P binding site that were well characterized by Bertrand *et al*.'s mutagenesis study of the kinase domain of PFKFB proteins (Bertrand *et al.*, 1998).

We developed two different targets, one that consisted of the F6P binding area, and one that contained the F6P binding area and the catalytic regions that bound magnesium and ATP. We had our greatest success targeting the F6P region alone. When we examined our initial docked compounds we found that the highest scored compounds were interacting primarily with the F6P binding residues. Furthermore, targeting the F6P binding site alone produced better docking scores. With PFKFB3, docking success was obtained by restricting the targeted region to the F6P binding site, rather than the entire active site. A different approach proved necessary for docking success with the next target.

3.1.3. MIF (Migratory Inhibition Factor)—In collaboration with the Mitchell group at the Brown Cancer Center we targeted macrophage migration inhibitory factor (MIF), which is a pleiotropic enzyme with a wide variety of functions in the areas of metabolism, immunity, and the cell cycle. MIF mediates immune effects by overriding the immunosuppressive effects of glucocorticoids, which produces pathology in the form of septic shock, arthritis, and glomerulonephritis (Hoi *et al.*, 2007). MIF also affects cell growth, apoptosis and angiogenesis (Hagemann *et al.*, 2007). MIF expression is increased in malignant and metastatic tumors of the breast, prostate (Meyer-Siegler *et al.*, 1998), colon (Wilson *et al.*, 2005), brain (Markert *et al.*, 2001), skin (Shimizu *et al.*, 1999) and lung (Coleman *et al.*, 2008). MIF also has a dopachrome tautomerase activity (Lubetsky *et al.*, 2002). While this activity has no known physiological function, it is useful in that it provides a fast and easy mechanism for assessing enzymatic inhibition.

A new antineoplastic compound targeting MIF, 4-iodo-6-phenylpyrimidine (4-IPP) was discovered (Winner *et al.*, 2008). This compound was identified through a virtual screening program similar to that conducted for PFKFB3. In this case, the MIF model was constructed using the MIF crystal structure (PDB id: 1MIF). Examination of the MIF active site showed that the side chain of the A2 methionine formed a protrusion at the base of the MIF pocket, while the catalytic A1 proline was situated on the side of the pocket and closer to the rim. We felt that targeting this protrusion would allow the docking program to more completely explore the docking area. Ludi (Accelrys) was used to dock a library of 343,802 compounds from the ACD (MDL) library into the MIF active site. Of the top 100 compounds, 76 were available commercially, and 41 of these compounds were found to be soluble at 100 μmol/L. Nine of these compounds were found to be inhibitory at concentrations of 50 μMol/L or less. The activities of these compounds were compared to an existing MIF inhibitor, ISO-1 (Al-Abed *et al.*, 2005), and one compound, 4-IPP, was found to have an IC_{50} that was an order of magnitude less than the IC_{50} of ISO-1. Further investigation of 4-IPP demonstrated that this compound covalently modifies the catalytic proline on MIF, producing inhibition of MIF dependent migration and anchorage-independent growth (Figure 1, right panel). This compound is currently being optimized and activity is down to ~30 nM.

One of the major lessons in the identification 4-IPP has been that screening efforts have a better chance at success with well-characterized and understood targets. By fully examining the MIF active site, we could ascertain that targeting the active N-terminal catalytic A1 proline would result in exploration of only one side of the active pocket, which would neglect several potential stabilizing interactions with residues on the other side of the pocket and put undue restriction on the amount of space for the docked molecules. By targeting the A2 methionine residue side chain, the full pocket was explored, and docking scores for compounds targeting this residue increased. As 4-IPP demonstrated, these compounds could still interact with the catalytic proline, but also retained several stabilizing interactions with residues throughout the active pocket. Thus in the case of MIF, the success of the virtual screen was in being able to make use of the entire active site, instead of restricting the targeted region to the reactive N-terminal A2 proline.

3.2. Building On Initial Success: From Desktop To Grid Computing

It rapidly became apparent that these methods should be expanded in the types of software used, the number of compounds in the screening libraries, and the number of targets screened. To do this, substantial computational time would be required that was not available at that time. We therefore formed a partnership with the Kentucky Dataseam Initiative (www.kydataseam.com), and started developing and using Grid computing for virtual screening. Dataseam, a Kentucky-based not-for-profit company, built and maintains a large managed computing grid that uses desktop computers in K-12 schools across the Commonwealth of Kentucky. The company also provides ongoing workshops and training to improve the educational use of the technology in the classrooms.

Grid computing is a form of distributed computing whereby a cluster of loosely-coupled networked computers act in concert to perform very large tasks. This is particularly suited to computational problems that can use these “embarrassingly parallel” systems, such as virtual screening, as each “docking” is independent. This technology has been applied to computationally-intensive virtual screening efforts, such as the Screensaver Project (Richards, 2002), or FightingAids@home (fightaidsathome.scripps.edu). In a unique way we have tapped into the computational processing power of thousands of Apple computers in schools of various school districts across the Commonwealth of Kentucky when they are not being used. The Xgrid package which is an integral part of the Mac OS X operating system makes it easy to aggregate the desktop power into a powerful computational grid.

At present we have around 6000 individual agents attached to the fully functional grid. This enables ~450 CPU years a month of research computing. At any time 99–100% of the GRID is utilized to run virtual screening. The advantage of Mac OS X is that it is Unix-based and virtual screening programs like Autodock (Morris *et al.*, 1998), Surflex-Dock (Jain, 2003), and DOCK (http://dock.compbio.ucsf.edu/DOCK_6/index.htm) can run on this grid environment. Submission of jobs takes place from a master controller that sends necessary information to the sub-controller sets. In turn the sub-controller sets are individually attached to numerous agent computers. These agents individually execute jobs and the results are brought back to the master controller. In essence, the “GRID” is a tightly integrated collection of subgrids.

Several critical issues had to be resolved for this to be a production GRID due to the unique environment, although these issues are common in volunteer or donated distributed computing systems. The first is that these computers primary use is as a student resource, which cannot be affected in any way. The second issue is network bandwidth usage. Since this is running on top of existing networks and their normal traffic, we cannot affect its primary purpose. We optimized for bandwidth usage and frequency and quantity of traffic by parsing the millions of compounds of a library into smaller libraries and running thousands of jobs. The third issue is automation. While Xgrid performs well in submitting a particular job to an agent, you cannot (reasonably) manually do this for 5,000 jobs for each target. To facilitate this, we have scripted for intelligent job submission (both single subgrid and spreading the job to the available processors on any or multiple subgrids), status reporting, error reporting, resubmission of failed jobs, and result retrieval. We also have automated the analysis of the entire virtual screening run. This is essential, as with a single target as much as 150 GB of data can be returned. The Perl scripts we developed are extremely useful as they work both on the grid and on our in house 440 processor IBM server.

3.3 Virtual Screening Process

For successful receptor-based virtual screening using our approach there must be a three-dimensional structure or homologous structure available for the region of interest. Our procedure for virtual screening is the following: 1) Understanding the structure of the target, i.e. is it in the active or inactive state, is it the correct state to target?; 2) Cleaning the target structure from X-ray crystallography/NMR/homology modeling. X-ray crystal structures are not necessarily at energy minima and assigning hydrogen atoms must be accompanied by minimization/regularization of the structure to ensure good stereochemistry; 3) Virtual screening. We most commonly use Autodock, Surflex-Dock, and DOCK on our current libraries: 2008 ZINC Total 8,490,191 compounds, 2008 ZINC Drug-Like: 5,348,215 compounds, 2008 iResearch Sourceable 5,363,141 compounds, 2008 Pubchem: 19,327,825 compounds; 4) Ranking, Absorption, Distribution, Metabolism, Excretion (ADME) filters based on solubility, logP, and other common criteria, and clustering for representative family members; 5) Purchase “hits”. Not all ”available” chemicals are actually available; 6) Biological testing. There must be a protein assay, functional or binding, and cell-based models, and ideally animal models available; 7) Similarity searches (“poor man’s QSAR”).

Most of the well-established software programs perform adequate docking, but the ranking methods can cause great variation in ranking. Our “poor mans QSAR” consists of a derivative search of the databases, or exemplifying a particular cluster, and retesting those compounds after purchase. This procedure has been used on over 30 targets, and to date the success rate, as defined as biological activity of less than 10 μM, is on the order of 10–30%. It is noteworthy that using only one virtual screening software is not recommended as no one piece of software is universally applicable to every system. A similar approach can be used for similarity searching or “scaffold hopping” if you start with a small molecule of interest. An additional step of docking into the known target site, if available, is also added.

3.4. Understanding The Target

For *in silico* projects it is important to understand the potential therapeutic target. Establishing that a target is appropriate requires more than performing a quick search for a PDB file, downloading a structure, and stripping out any water included in the file (this is usually relevant only for x-ray diffraction structures). The target and its biological role must be well understood to generate any meaningful results. No matter if one is experimenting with a complex DNA structure or a soluble protein, careful selection of the target is required for success. This necessity pushes to the forefront in the move from *in silico* model to the bench top. Even if there is good interaction between a potential drug and the selected target *in silico,* that interaction must be confirmed by *in vitro* techniques (such as biophysical and molecular biology methods) and eventually *in vivo* testing. *In vitro* confirmation of computer results usually requires that the selected target be isolated and purified. An excellent example of ambiguity in a target molecule can be found in the study of the human telomere and quadruplex DNA (described below).

3.4.1 Telomerase Inhibition Via Human Telomere Stabilization: A Difficult Target—Telomeres are specialized species-specific DNA sequences (the human sequence is $d(GGGTTA)_n$) that cap the ends of eukaryotic chromosomes and are thought to contribute to genetic stability by preventing the ends of the chromosome from being eroded away during replication, ultimately leading to chromosome fusion. This is because of the end replication problem and the mechanism of DNA copying, which necessarily results in end shortening on each round of replication. The human telomere is 5–8 thousand base pairs long with a single stranded 3′ overhang of 100 to 200 bases (Wright *et al.*, 1997). This G-rich, single-stranded overhang can adopt complex structures, i.e. a G-quadruplex. A G-quadruplex is made up of stacked G-quartets in a square planer array stabilized by Hoogsten

hydrogen bonding. Telomere length is maintained by the activity of the enzyme telomerase, which is active in stem cells and embryonic cells but not in adult terminally differentiated cells. However, immortalization by re-expression of the telomerase genes is a hallmark of cancers. Telomerase activation has been found to be involved in greater than 90% of all cancers (Shay *et al.*, 1997). Formation of G-quadruplexes has been shown to decrease the activity of telomerase. Thus the human telomere sequence is an attractive target for therapeutic strategies using small molecules to stabilize these complex structures, limit telomerase function, and therefore restrict growth of cancer cells (Hahn *et al.*, 1999).

In addition to quadruplex formation in the human telomeric sequence, other areas of the genome have been identified where putative quadruplex forming sequences occur (Huppert *et al.*, 2005; 2007). The first of these was the discovery of a quadruplex forming sequence in the promoter region of the proto-oncogene c-myc. It was shown that this quadruplex was stabilized by a cationic porphyrin, TMPyP4, which suppressed c-myc transcriptional activation (Siddiqui-Jain *et al.*, 2002; Simonsson *et al.*, 1998). In the years since the c-myc discovery many more genes have been found to have potential G-quadruplexes in their promoter regions, such as the proto-oncogenes c-kit, bcl-2, and VEGF, and a recent genome-wide search has revealed more than 375,000 potential G-quadruplex forming sequences in the human genome (Dai *et al.*, 2006; Fernando *et al.*, 2006; Huppert *et al.*, 2005, 2007; Sun *et al.*, 2005). Researchers are now beginning to uncover putative G-quadruplex forming sequences in prokaryotes creating the possibility of a new class of antibiotic agents (Rawal *et al.*, 2006).

In addition to the tremendous potential for quadruplex/drug interactions to regulate gene expression in cancer or as a target for novel antibiotics in prokaryotes, quadruplexes themselves can act as drugs. A notable example is the previously mentioned AS1411 (Ireson *et al.*, 2006), which is described in detail by Bates *et al.* in this issue. Quadruplexes also have been shown to have antiviral activity and have been demonstrated to be effective against HIV-1 *in vivo* (Bishop *et al.*, 1996; Suzuki *et al.*, 2002). Mainly due to their ability to recognize both nucleic acids and proteins with a high degree of specificity and because of their stability and nuclease resistance quadruplexes are rapidly becoming attractive targets for development of novel therapeutics and preclinical studies are underway for several other quadruplex-based therapeutics (Cogoi *et al.*, 2006; Qi *et al.*, 2006).

When selecting a quadruplex target, the question becomes: where to start? There are several structures associated with the human telomere sequence, but which ones are biologically relevant? Various quadruplex structures formed by the human telomere and other sequences, with and without drug-like compounds, have been solved yielding nearly 100 crystal and NMR structures. While this seems like an embarrassment of riches, these structures represent only a small number of all possible structures. When one considers all of the theoretical 26 possible looping topologies and 8 possible tetrad arrangements available based on the possible glycosyl bond angles there are over 200 possible unimolecular structures for the human telomere sequence $d(GGGTTA)_3GGG$ (Webba da Silva, 2007). This number does not include the possibilities for bimolecular or tetramolecular quadruplex formation.

Additionally, an argument can be made that the available crystal structures do not reflect the species actually present in solution. Aside from the classic crystal vs. solution phase argument, it is possible that any crystallized version of the human telomere has been selected by the crystallization conditions and may not accurately represent a realistic structure (Lane *et al.*, 2008; Li *et al.*, 2005). Even if we ignore this possibility, the available crystal and NMR structures may not be reasonable representations of reality as we are

currently unable to assess how these reported structures relate to the form(s) present under conditions found in the cell.

Normally, overcoming these problems would require a thorough understanding and rigorous characterization of the system one wishes to study. Unfortunately, for the quadruplexes formed by the human telomere DNA, it is difficult to apply standard separation and biophysical techniques and expect reasonable results because different coexisting folds of the same oligonucleotide would have nearly identical physical properties. Instead, the various structural configurations have been "isolated" via artificial direction to a particular form by modifying the sequence of the quadruplex-forming DNA (Parkinson *et al.*, 2002; Luu *et al.*, 2006; Phan *et al.*, 2007; Dai, Carver *et al.*, 2007; Dai, Punchihewa *et al.*, 2007). These modifications themselves throw into question the validity of the results. As with any line of inquiry, when our understanding of these systems improves we can expect greater yields from bench top and *in silico* efforts.

3.4.2 Combination of Theoretical and Experimental Techniques in Structural Studies—As there is ambiguity about the traditional structural techniques dealing with human telomeric DNA quadruplex, we, in collaboration with the Chaires group at the Brown Cancer Center, tested the feasibility of linking computational and experimental methodologies to address these complex structural problems (Li *et al.*, 2005). Structural knowledge of telomeric DNA is critical for the understanding of telomere biological function and for the utilization of telomeric DNA as target in chemotherapy. However, determination of the particular structure adopted by the human telomeric DNA in physiological conditions is a not trivial problem due to its extreme structural polymorphism. To address this problem we developed a new approach combining several experimental and molecular modeling techniques. The key point was to estimate, from the possible models of the human quadruplex structures, some physical properties that could be compared to the experimentally determined values. This comparison critically tested the validity of the structural models and allowed us to distinguish between alternate conformational forms. We started from the high-resolution structures reported for short (22–26 nt) segments of the human telomeric DNA under a variety of solution conditions. In sodium solution, an antiparallel "basket" structure forms with two lateral loops and one diagonal loop connecting three stacked quartets. We initially focused on the expected hydrodynamic properties of the different structures calculated by means of the HYDROPRO program. This software allows the calculation of several hydrodynamic properties of macromolecules from their known atomic level structures by use of the "bead" models. To obtain more reliable values, molecular dynamics simulations refined the telomeric DNA structures and the trajectories were used as input for the HYDROPRO program thus obtaining sedimentation coefficient for distribution in each model.

These results confirmed the feasibility and validity of linking computed structures with experimental hydrodynamic values as a method to discriminate between possible quadruplex structures in solution. Another useful property that can be calculated from the models using the NACESS software is the Solvent Accessibility Surface Area (SASA) of the adenine residues. This property was experimentally estimated by quantitative fluorescence studies using strategic and systematic single-substitution of 2-aminopurine for adenine bases. The comparison of the computed SASA values with the experimental fluorescence quenching result allowed us to further discriminate between possible conformations. This mixed experimental and computational approach can be used to explore complex structural problems that can be solved using the standard NMR or crystallographic methods.

3.5. DNA: An Overlooked Target

DNA is an attractive target for a number of reasons. Genes are present in a small number compared to mRNA and proteins, are not turned over and are at the start of the amplification cascade. However, the vast majority of virtual screening efforts have typically focused on protein targets, presumably because of the knowledge base and large repository of crystal structures of protein targets. Targeting of nucleic acids has been largely ignored, perhaps due to poor understanding of the heterogeneity and various morphologies of nucleic acids. However, with advances in the knowledge of the structure and function of duplex, triplex and quadruplex structures of nucleic acids, nucleic acids are becoming attractive targets for small molecule development. This is particularly important since certain morphologies of nucleic acids may hold medicinal value such as triplex and quadruplex nucleic acid structures which have been strongly associated with gene modulation and anti-cancer activity, respectively. For these reasons, interest in virtual screening of small molecules against nucleic acid targets is likely to become increasingly popular in the quest for the development of new drugs.

3.5.1 Can Virtual Screening Be Successfully Used to Target DNA?—We focus here on two aspects of virtual screening of nucleic acid targets. First and of primary importance is determining whether current virtual screening software can be used successfully to dock small molecules to known nucleic acid targets. Since most of the virtual screening software was developed to target proteins, it is unclear if the software also can be used to target nucleic acids. Our recent report (Holt *et al*., 2008) demonstrated that Surflex-Dock and Autodock can be optimized to successfully reproduce ligand-nucleic acid crystal structure complexes. In this study, distamycin and pentamidine, two ligands that are known to bind to the minor groove of DNA, and daunorubicin and ellipticine, which are known to intercalate between base pairs in DNA, are docked accurately to their nucleic acid targets. This study validated the use of Surflex-Dock and Autodock for targeting nucleic acids that, surprisingly, in the case of Surflex-Dock had not been done. The second objective is to present an example of a ligand-based approach. We have used virtual screening software for the identification of selective, high-affinity ligands for triplex nucleic acids. A known high affinity ligand was used as a basis for identifying ligands with similar structural features using Surflex-Sim. These ligands were then docked to the triplex nucleic acid poly(dA)-$[poly(dT)]_2$ using Surflex-Dock. The top-ranking ligands identified by virtual screening were tested empirically by Competition Dialysis to verify the predicted selectivity and affinity of these ligands for poly(dA)-$[poly(dT)]_2$ (manuscript in preparation).

In-Silico Virtual Library Preparation: For the Surflex-Dock and Autodock validation study, ligand nucleic-acid complexes for the minor groove binders, distamycin and pentamidine, and the intercalators, daunorubicin and ellipticine were obtained from the Protein Data Bank, with the following identification codes, 2dnd, 1d64, 152d, and 1z3f, respectively. For the triple helical ligand identification study, a triplex-selective ligand was constructed and served as the initial basis for Surflex-Sim experiments. The triplex nucleic acid structure poly(dA)-$[poly(dT)]_2$ with an intercalation site was constructed and used for Surflex-Dock experiments.

In-Silico Virtual Screening Methods: For the Surflex-Dock and Autodock validation study, Surflex-Dock version 2.11 and Autodock version 4.0 were compiled for Macintosh OS X PowerMac G5 and Linux workstations. The Surflex-Dock “Multistart 5” and “Random 5” options were investigated because these parameters were thought to play a role in the accuracy and ranking of poses generated by these programs. The “Multistart 5” option initiates docking of a ligand to a target from 5 different orientations around the target. The “Random 5” option randomizes the X, Y, Z coordinates of the ligand with respect to the

crystal structure prior to docking. These parameters may effect the docking of a ligand to the target in the case that the initial starting position is energetically unfavorable. The Autodock parameters that were tested were the number of energy evaluations performed prior to determining the best dock as well as total number of docks performed. The number of energy evaluations was varied as 200,000 (2E5), 2,000,000 (2E6) or 20,000,000 (2E7) while the number of docks was varied by 5, 10 or 20, to determine whether this could impact ligand docking performance. To determine the accuracy of the dockings, the Root Mean Square Deviation (rmsd) was calculated between the top ranked docked pose and the crystal structure using the Surflex-Dock rmsd-scoring calculator. A rmsd level of significance of 2 angstroms was designated to compare the accuracy of docking to data in the reported literature, as this is typically a threshold that is considered an accurate dock. For the triplex selective ligand identification study, a combination of Surflex-Sim and Surflex-Dock was used. Surflex-Sim initially was used to find structurally similar ligands to a known triplex selective ligand from a commercially available ZINC database of 1.96 million compounds. Surflex-Dock was subsequently used to determine how well the top ranking ligands fit into the triplex intercalation site (manuscript in preparation).

3.5.2. Competition Dialysis Methods—A critical component in virtual screening is the experimental validation or testing of predictions. The Competition Dialysis Method has been utilized and described (Chaires, 2003, 2005a, Chaires, b; Ragazzon *et al.*, 2007; Shi *et al.*, 2006) and gives binding and affinity of a ligand over many different DNA sequences and morphologies. Briefly, a 0.2 mL volume of a 75 μM solution of each nucleic acid is dialyzed against a solution of test ligand at 1 μM. The concentration of nucleic acid is expressed in terms of monomeric unit, with base pairs for duplex DNA, triplets for triplex DNA and tetrads for quadruplex DNA. After reaching equilibrium, the ligand bound in each dialysis well is dissociated using 20 μL of 10% (w/v) SDS and the total ligand is quantified spectrophotometrically.

3.5.3. Surflex-Dock and Autodock Accurately Reproduce and Rank the Crystallographic Structures of Nucleic Acid Ligand Complexes—The docking performance of Autodock and Surflex-Dock for daunorubicin, distamycin, ellipticine and pentamidine can be assessed by docking accuracy and ranking of the poses. Docking accuracy determines if the crystal pose can be successfully reproduced and is determined by calculating the rmsd for each of the docked poses compared to the crystal pose, irrespective of ranking. The ranking performance determines if the lowest rmsd pose is ranked as the top pose returned by the docking program. With respect to docking accuracy, both Autodock and Surflex-Dock are able to dock daunorubicin and distamycin to their crystal structure targets within a resolution of 2 angstroms. Interestingly, ellipticine has a higher rmsd value dock of approximately 6 angstroms, but this appears to be due to several factors including not only the overall marginal accuracy of docking to the target but also symmetry of the target nucleic acid as ellipticine is able to dock into the intercalation side from either side of the target. For pentamidine, Autodock docks the ligand in the correct orientation relative to the crystal structure ligand while Surflex-Dock docks the ligand in an inverted orientation relative to the crystal structure. However, due to the molecular symmetry of pentamidine, the Surflex-Dock dock is actually a far better dock than the rmsd initially predicts. These results bring to light the importance of considering both target symmetry, in the case of ellipticine, and ligand symmetry, in the case of pentamidine, when considering the quality of the dock. Target symmetry can be taken into account by flipping and superposition of the target and docked ligand onto the crystal pose while internal ligand symmetry also can be accounted for by an additional Actual rmsd ISO Surflex-Dock function. After accounting for these points of symmetry, the docking performance of ellipticine and pentamidine is significantly improved. The ability to rank the multiple docked poses is critical, particularly

in virtual screening applications where, due to the large number of test ligands, typically only the top-ranked pose is considered. Both Autodock and Surflex-Dock appear to successfully rank the poses, with the top-ranked pose typically having low rmsd values compared to other docked poses (Figure 2). These rankings are again improved substantially for ellipticine and pentamidine when accounting for target and ligand symmetry. A complete discussion of all of the findings of these docking studies is beyond the scope of this review, but for additional detail, see Holt *et al.*, 2008. It should be noted that certain software parameterizations for Surflex-Dock and Autodock appear to optimally balance docking accuracy and ranking and are recommended for virtual screening applications. Specifically, Surflex-Dock performed optimally at a software parameterization of either “Multistart 5” only or “Multistart 5” and “Random 5” while Autodock performed best under conditions of “2E7 energy evaluations” and “5 docks.” Overall, these docking studies successfully validated the use of both Surflex-Dock and Autodock for targeting ligands to nucleic acids.

3.5.4. Virtual Screening and Competition Dialysis Successfully Identifies New Triplex Selective Intercalators—With successful completion of the validation studies, the next question was first, whether these molecular docking tools are capable of finding new ligands that bind to a known target and, second, if the predictive nature of virtual screening software can be validated by Competition Dialysis. A triple helical DNA structure, poly(dA)-$[poly(dT)]_2$ was selected as a basis for these experiments. A ligand that has a previously demonstrated high affinity and selectivity for this triplex structure (Cassidy *et al.*, 1996) was selected as the ligand structure for Surflex-Sim experiments. The commercially available component of the ZINC database of chemical compounds was screened to select potential ligands with similar structural features to the initial control ligand. The results of this initial screen yielded several hundred compounds that were then docked using Surflex-Dock into the triplex intercalation site to check for their fit into the binding pocket. Two top-ranked ligands were chosen that were predicted to have high affinity and selectivity for the triplex intercalation site. These compounds were tested by Competition Dialysis to determine binding affinity and selectivity for poly(dA)-$[poly(dT)]_2$. The ligands identified by virtual screening appear to bind with much higher affinity (amount of ligand bound to the triplex structure) and comparable selectivity (amount of ligand bound to triplex structure compared to other structures) for the triplex nucleic acid, particularly when compared to another known selective triplex binding ligand (Figure 3). This study demonstrates the power of virtual screening for the discovery of new ligands that can bind to nucleic acids. Importantly, this approach can be extended to find ligands that target other therapeutically relevant morphologies of nucleic acids that may be associated with disease pathologies.

3.6. Virtual Screening on Membrane Proteins

Understanding the target structure is critical in targeting membrane proteins. However, for G-protein coupled receptors (GPCRs) the severe limitation is the paucity of such structures, with only two Class A GPCRs, represented by the beta2 adrenergic receptor (Cherezov *et al.*, 2007) and rhodopsin (Palczewski *et al.*, 2000), being currently available for homology modeling. Also, these homology models have to be relaxed in the appropriate environment in molecular dynamics simulations. It is also critical that techniques such as specific site mutagenesis is used to validate the theoretical predictions as the protein target are usually unavailable in large purified quantities to perform biophysical or direct binding assays.

3.6.1. BLT1—Leukotriene B4 (LTB4) mediates a variety of inflammatory diseases such as asthma, arthritis, atherosclerosis, and cancer through activation of the G-protein-coupled receptor, BLT1. We undertook a solvated lipid bilayer molecular dynamics approach in collaboration with the Bodduluri group at the Brown Cancer Center to propose a possible

binding site of LTB4 (Basu *et al.*, 2007). From this approach it was possible to predict residues that were involved in binding in the extracellular loop R156, and in the transmembrane domains III (H94A and Y102A), V (E185A), and VI (N241A). From the ligand-free and ligand-bound states, we observed an activation core comprising of Asp-64, displaying multiple dynamic interactions with Asn-36, Ser-100, and Asn-281 and a triad of serines, Ser-276, Ser-277, and Ser-278. Thus, as we now have a predictive model, virtual screening efforts are underway and we have found one antagonist at the 40 μM level.

3.6.2. Targeting The CXCR4:G-Protein Interface—CXCR4 is a seven transmembrane G-protein coupled receptor. It is widely expressed on leukocytes, and serves to regulate leukocyte hematopoiesis and trafficking (Gulino, 2003) as well as anchoring developing cells in their proper areas in the bone marrow (Sugiyama *et al.*, 2006). CXCR4 is most widely known as the coreceptor for T-tropic strains of the human immunodeficiency virus (HIV) (Murdoch, 2000). CXCR4 also plays an important role in development as CXCR4 deletions are embryonically lethal, with embryos displaying defects in neuronal and cardiac development (Bagri *et al.*, 2002; Lu *et al.*, 2002; McGrath *et al.*, 1999) as well as neovascularization (Lima-e-Silva *et al.*, 2007). CXCR4 heterozygous mutation can lead to WHIM (warts, hypogammaglobulinemia, infections and myelokathexis) syndrome (Gulino, 2003). CXCR4, and its ligand CXCL12, also play a role in the progression of cancer. CXCR4 plays a role in both angiogenesis (Chu *et al.*, 2008) and metastasis (Yasuoka *et al.*, 2008) in several tumor types, including basal cell carcinoma (Chu *et al.*, 2008), thyroid cancer (Yasuoka *et al.*, 2008), squamous cell carcinoma (Oliveira-Neto *et al.*, 2008), renal cell carcinoma (Reckamp *et al.*, 2008), hepatocellular carcinoma (Li *et al.*, 2007), breast (Zlotnik, 2004), lung (Zlotnik, 2004), and prostate cancers (Zlotnik, 2004). The role of CXCR4 in angiogenesis and metastasis of cancer cells suggests that this protein would make an effective target for anti-neoplastic agents.

We have targeted two separate regions of CXCR4 for drug discovery, the extracellular loop regions and the intracellular loop region. The intracellular loop region is discussed here. Using the molecular docking program Surflex-Dock we have virtually screened ~3,600,000 compounds targeting the G-protein interaction surface on the intracellular loops in an extension of our earlier solvated lipid bilayer molecular dynamics study of CXCR4 (Trent *et al.*, 2003). To test compounds identified in this virtual screen we have developed cell lines expressing human CXCR4 with a GFP tag. These cell lines were developed from parental 300.19 and RBL-2H3 cells, which were transfected by electroporation. Stable transfectants were selected by geneticin and single cell clones were developed with matched levels of hCXCR4-GFP surface expression by immunofluorescent sorting. These cell lines are used to test CXCR4 activity in two primary assays, intracellular calcium mobilization and an *in vitro* chemotaxis assay. In the intracellular calcium mobilization assay, cells were treated with an initial screening concentration of 50 μM of inhibitor, with AMD3100 used as a positive control, and DMSO alone as a negative control. Treated cells then were stimulated with CXCL12 (Peprotech) and the median fluorescence ratio of Indo-1 acetoxymethyl ester was measured as a ratio of emission at 390 and 490 nm. In the *in vitro* chemotaxis assay, a two-chamber boyden assay was conducted. In the lower chamber, chemotaxis buffer was treated with CXCL12 and a screening concentration of 50 μM of inhibitor, or AMD3100 or DMSO alone. Cells were placed in the upper chambers and, following a 3 hour incubation, cells that migrated into the lower chamber were counted.

Analysis of these compounds is ongoing, but initial results of compound screening look promising. Of twenty nine compounds tested using intracellular calcium mobilization, nine compounds produced at least 40% reduction in calcium mobilization at 50 μM, with two compounds producing reduced mobilization at 1 μM. Seventeen of twenty nine compounds produced at least 40% reduction in chemotaxis at 50 μM, with seven compounds producing

reduced chemotaxis at 1 μM. Four compounds produced at least 60% reductions in both assays. Further investigation will be necessary to understand the significance of these results, but the fact that we are seeing ~40% hit rates for intracellular calcium mobilization, and ~59% hit rates for chemotaxis is promising, even if it is at the higher concentration.

4.0. Conclusions

We have developed a platform to perform rapid virtual screening on a wide range of target types. This has been successfully applied to different DNA morphologies, soluble proteins, and membrane proteins. The inherent understanding of the molecular structure and nature of the target is a critical component for success, as is having the appropriate assays to validate the *in silico* predictions. The James Graham Brown Cancer Center is a highly collaborative setting in which the Molecular Targets Group extensively interacts with the Structural Biology Group enabling the ability to go from new target to new lead template rapidly.

Acknowledgments

This work has been supported by National Institutes of Health grant 1R01CA113735-02, National Center for Research Resources grant 1P20RR018733, and National Institutes of Health grant RO1GM077422-01, and the KY Lung Cancer Research Program. Thanks to Andrew Lane for critical reading of this review.

Abbreviations

rmsd	root mean squared deviation
GPCR	G-protein coupled receptor
ADME	Absorption, Distribution, Metabolism, Excretion
MTT	cell proliferation assay (3-(4,5-dimethylthiazolyl-2)-2,5-diphenyltetrazolium bromide)
EMSA	Electrophoretic Mobility Shift Assay
MIF	Migratory Inhibition Factor
PFKFB3	6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase

References

Al-Abed Y, Dabideen D, Aljabari B, Valster A, Messmer D, Ochani M, Tanovic M, Ochani K, Bacher M, Nicoletti F, Metz C, Pavlov VA, Miller EJ, Tracey KJ. ISO-1 binding to the tautomerase active site of MIF inhibits its pro-inflammatory activity and increases survival in severe sepsis. J Biol Chem. 2005; 280(44):36541–36544. [PubMed: 16115897]

Allain FH, Gilbert DE, Bouvet P, Feigon J. Solution structure of the two N-terminal RNA-binding domains of nucleolin and NMR study of the interaction with its RNA target. J Mol Biol. 2000; 303(2):227–241. [PubMed: 11023788]

Bagri A, Gurney T, He X, Zou Y, Littman D, Tessier-Lavigne M, Pleasure S. The chemokine SDF1 regulates migration of dentate granule cells. Development. 2002; 129(18):4249–4260. [PubMed: 12183377]

Basu S, Jala VR, Mathis S, Rajagopal ST, Del Prete A, Maturu P, Trent JO, Haribabu B. Critical role for polar residues in coupling leukotriene B4 binding to signal transduction in BLT1. J Biol Chem. 2007; 282(13):10005–10017. [PubMed: 17237498]

Bates PJ, Kahlon JB, Thomas SD, Trent JO, Miller DM. Antiproliferative activity of G-rich oligonucleotides correlates with protein binding. J Biol Chem. 1999; 274(37):26369–26377. [PubMed: 10473594]

Bertrand L, Vertommen D, Freeman PM, Wouters J, Depiereux E, Di-Pietro A, Hue L, Rider MH. Mutagenesis of the fructose-6-phosphate-binding site in the 2-kinase domain of 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase. Eur J Biochem. 1998; 254(3):490–496. [PubMed: 9688258]

Bishop JS, Guy-Caffey JK, Ojwang JO, Smith SR, Hogan ME, Cossum PA, Rando RF, Chaudhary N. Intramolecular G-quartet motifs confer nuclease resistance to a potent anti-HIV oligonucleotide. J Biol Chem. 1996; 271(10):5698–5703. [PubMed: 8621435]

Cassidy SA, Strekowski L, Fox KR. DNA sequence specificity of a naphthylquinoline triple helix-binding ligand. Nucleic Acids Res. 1996; 24(21):4133–4138. [PubMed: 8932362]

Chaires JB. A competition dialysis assay for the study of structure-selective ligand binding to nucleic acids. Curr Protoc Nucleic Acid Chem. 2003; Chapter 8(Unit 8 3)

Chaires JB. Competition dialysis: an assay to measure the structural selectivity of drug-nucleic acid interactions. Curr Med Chem Anticancer Agents. 2005a; 5(4):339–352. [PubMed: 16101486]

Chaires, JB. Structural Selectivity of Drug-Nucleic Acid Interactions Probed by Competition Dialysis. Heidleberg: Springer-Verlag GMBH & Co; 2005b.

Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SG, Thian FS, Kobilka TS, Choi HJ, Kuhn P, Weis WI, Kobilka BK, Stevens RC. High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. Science. 2007; 318(5854):1258–1265. [PubMed: 17962520]

Chesney J, MitchelL R, Benigni F, Bacher M, Spiegel L, Al-Abed Y, Han J, Metz C, Bucala R. An inducible gene product for 6-phosphofructo-2-kinase with an AU-rich instability element: role in tumor cell glycolysis and the Warburg effect. Proc Natl Acad Sci USA. 1999; 96(6):3047–3052. [PubMed: 10077634]

Chu C, Cha S, Lin W, Lu P, Tan C, Chang C, Lin B, Jee S, Kuo M. Stromal-cell-derived factor-1{alpha} (SDF-1{alpha}/CXCL12)-enhanced angiogenesis of human basal cell carcinoma cells involves ERK1/2-NF-{kappa} B/interleukin-6 pathway. Carcinogenesis. 2008 October 9. Epub ahead of print.

Clem B, Telang S, Clem A, Yalcin A, Meier J, Simmons A, Rasku MA, Arumugam S, Dean WL, Eaton JW, Lane AN, Trent JO, Chesney J. Small-molecule inhibition of 6-phosphofructo-2-kinase activity suppresses glycolytic flux and tumor growth. Mol Cancer Ther. 2008; 7(1):110–120. [PubMed: 18202014]

Cogoi S, Xodo LE. G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. Nucleic Acids Res. 2006; 34(9):2536–2549. [PubMed: 16687659]

Coleman AM, Rendon BE, Zhao M, Qian MW, Bucala R, Xin D, Mitchell RA. Cooperative regulation of non-small cell lung carcinoma angiogenic potential by macrophage migration inhibitory factor and its homolog, D-dopachrome tautomerase. J Immunol. 2008; 181(4):2330–2337. [PubMed: 18684922]

Dai J, Chen D, Jones RA, Hurley LH, Yang D. NMR solution structure of the major G-quadruplex structure formed in the human BCL2 promoter region. Nucleic Acids Res. 2006; 34(18):5133–5144. [PubMed: 16998187]

Dapic V, Abdomerovic V, Marrington R, Peberdy J, Rodger A, Trent JO, Bates PJ. Biophysical and biological properties of quadruplex oligodeoxyribonucleotides. Nucleic Acids Res. 2003; 31(8): 2097–2107. [PubMed: 12682360]

Dapic V, Bates PJ, Trent JO, Rodger A, Thomas SD, Miller DM. Antiproliferative activity of G-quartet-forming oligonucleotides with backbone and sugar modifications. Biochemistry. 2002; 41(11):3676–3685. [PubMed: 11888284]

Fernando H, Reszka AP, Huppert J, Ladame S, Rankin S, Venkitaraman AR, Neidle S, Balasubramanian S. A conserved quadruplex motif located in a transcription activation site of the human c-kit oncogene. Biochemistry. 2006; 45(25):7854–7860. [PubMed: 16784237]

Girvan AC, Teng Y, Casson LK, Thomas SD, Juliger S, Ball MW, Klein JB, Pierce WM Jr, Barve SS, Bates PJ. AGRO100 inhibits activation of nuclear factor-kappaB (NF-kappaB) by forming a complex with NF-kappaB essential modulator (NEMO) and nucleolin. Mol Cancer Ther. 2006; 5(7):1790–1799. [PubMed: 16891465]

Gulino A. WHIM syndrome: a genetic disorder of leukocyte trafficking. Curr Opin Allergy Clin Immunol. 2003; 3(6):443–450. [PubMed: 14612668]

Hagemann T, Robinson S, Thompson R, Charles K, Kulbe H, Balkwill F. Ovarian cancer cell-derived migration inhibitory factor enhances tumor growth, progression, and angiogenesis. Mol Cancer Ther. 2007; 6(7):1993–2002. [PubMed: 17620429]

Hahn WC, Stewart SA, Brooks MW, York SG, Eaton E, Kurachi A, Beijersbergen RL, Knoll JH, Meyerson M, Weinberg RA. Inhibition of telomerase limits the growth of human cancer cells. Nat Med. 1999; 5(10):1164–1170. [PubMed: 10502820]

Hoi A, Iskander M, Morand E. Macrophage migration inhibitory factor: a therapeutic target across inflammatory diseases. Inflamm Allergy Drug Targets. 2007; 6(3):183–190. [PubMed: 17897055]

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. [PubMed: 18642866]

Huppert JL, Balasubramanian S. Prevalence of quadruplexes in the human genome. Nucleic Acids Res. 2005; 33(9):2908–2916. [PubMed: 15914667]

Huppert JL, Balasubramanian S. G-quadruplexes in promoters throughout the human genome. Nucleic Acids Res. 2007; 35(2):406–413. [PubMed: 17169996]

Ireson CR, Kelland LR. Discovery and development of anticancer aptamers. Mol Cancer Ther. 2006; 5(12):2957–2962. [PubMed: 17172400]

Irwin JJ, Shoichet BK. ZINC--a free database of commercially available compounds for virtual screening. J Chem Inf Model. 2005; 45(1):177–182. [PubMed: 15667143]

Jain AN. Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. J Med Chem. 2003; 46(4):499–511. [PubMed: 12570372]

Kapetanovic IM. Computer-aided drug discovery and development (CADDD): In silico-chemico-biological approach. Chemico-Biological Interactions. 2008; 171(2):165–176. [PubMed: 17229415]

Kole H, Resnick RJ, Doren MV, Racker E. Regulation of 6-phosphofructo-1-kinase activity in ras-transformed rat-1 fibroblasts. Arch Biochem Biophys. 1991; 286(2):586–589. [PubMed: 1832835]

Lane AN, Chaires JB, Gray RD, Trent JO. Stability and kinetics of G-quadruplex structures. Nucleic Acids Res. 2008; 36(17):5482–5515. [PubMed: 18718931]

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. [PubMed: 16106044]

Li W, Gomez E, Zhang Z. Immunohistochemical expression of stromal cell-derived factor-1 (SDF-1) and CXCR4 ligand receptor system in hepatocellular carcinoma. J Exp Clin Cancer Res. 2007; 26(4):527–533. [PubMed: 18365549]

Lima-e-Silva R, Shen J, Hackett S, Kachi S, Akiyama H, Kiuchi K, Yokoi K, Hatara M, Lauer T, Aslam S, Gong Y, Xiao W, Khu N, Thut C, Campochiaro P. The SDF-1/CXCR4 ligand/receptor pair is an important contributor to several types of ocular neovascularization. FASEB J. 2007; 21(12):3219–3230. [PubMed: 17522382]

Lu M, Grove E, Miller R. Abnormal development of the hippocampal dentate gyrus in mice lacking the CXCR4 chemokine receptor. Proc Natl Acad Sci USA. 2002; 99(10):7090–7095. [PubMed: 11983855]

Lubetsky JB, Dios A, Han J, Aljabari B, Ruzsicska B, Mitchell R, Lolis E, Al-Abed Y. The tautomerase active site of macrophage migration inhibitory factor is a potential target for discovery of novel anti-inflammatory agents. J Biol Chem. 2002; 277(28):24976–24982. [PubMed: 11997397]

Markert JM, Fuller CM, Gillespie GY, Bubien JK, McLean LA, Hong RL, Lee K, Gullans SR, Mapstone TB, Benos DJ. Differential gene expression profiling in human brain tumors. Physiol Genomics. 2001; 5(1):21–33. [PubMed: 11161003]

Mazurek S, Zwerschke W, Jansen-Dürr P, Eigenbrodt E. Metabolic cooperation between different oncogenes during cell transformation: interaction between activated ras and HPV-16 E7. Oncogene. 2001; 20(47):6891–6898. [PubMed: 11687968]

McGrath K, Koniski A, Maltby K, McGann J, Palis J. Embryonic expression and function of the chemokine SDF-1 and its receptor, CXCR4. Dev Biol. 1999; 213(2):442–456. [PubMed: 10479460]

Meyer-Siegler K, Fattor RA, Hudson PB. Expression of macrophage migration inhibitory factor in the human prostate. Diagn Mol Pathol. 1998; 7(1):44–50. [PubMed: 9646034]

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. J Computational Chem. 1998; 19(14):1639–1662.

Murdoch C. CXCR4: chemokine receptor extraordinaire. Immunol Rev. 2000; 177:175–184. [PubMed: 11138774]

Okar D, Lange A. Fructose-2,6-bisphosphate and control of carbohydrate metabolism in eukaryotes. Biofactors. 1999; 10:1–14. [PubMed: 10475585]

Oliveira-Neto H, Silva E, Leles C, Mendonça E, Alencar-Rde C, Silva T, Batista A. Involvement of CXCL12 and CXCR4 in lymph node metastases and development of oral squamous cell carcinomas. Tumour Biol. 2008; 29(4):262–271. [PubMed: 18781098]

Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE, Yamamoto M, Miyano M. Crystal structure of rhodopsin: A G protein-coupled receptor. Science. 2000; 289(5480):739–745. [PubMed: 10926528]

Qi H, Lin CP, Fu X, Wood LM, Liu AA, Tsai YC, Chen Y, Barbieri CM, Pilch DS, Liu LF. G-quadruplexes induce apoptosis in tumor cells. Cancer Res. 2006; 66(24):11808–11816. [PubMed: 17178877]

Ragazzon PA, Garbett NC, Chaires JB. Competition dialysis: a method for the study of structural selective nucleic acid binding. Methods. 2007; 42(2):173–182. [PubMed: 17472899]

Ramanathan A, Wang C, Schreiber S. Perturbational profiling of a cell-line model of tumorigenesis by using metabolic measurements. Proc Natl Acad Sci USA. 2005; 102(17):5992–5997. [PubMed: 15840712]

Rawal P, Kummarasetti VB, Ravindran J, Kumar N, Halder K, Sharma R, Mukerji M, Das SK, Chowdhury S. Genome-wide prediction of G4 DNA as regulatory motifs: role in Escherichia coli global regulation. Genome Res. 2006; 16(5):644–655. [PubMed: 16651665]

Reckamp K, Strieter R, Figlin R. Chemokines as therapeutic targets in renal cell carcinoma. Expert Rev Anticancer Ther. 2008; 8(6):887–893. [PubMed: 18533798]

Richards WG. Virtual screening using grid computing: the screensaver project. Nat Rev Drug Discov. 2002; 1(7):551–555. [PubMed: 12120261]

Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol. 1993; 234(3):779–815. [PubMed: 8254673]

Shay JW, Bacchetti S. A survey of telomerase activity in human cancer. Eur J Cancer. 1997; 33(5): 787–791. [PubMed: 9282118]

Shi X, Chaires JB. Sequence- and structural-selective nucleic acid binding revealed by the melting of mixtures. Nucleic Acids Res. 2006; 34(2):e14. [PubMed: 16432258]

Shimizu T, Abe R, Nakamura H, Ohkawara A, Suzuki M, Nishihira J. High expression of macrophage migration inhibitory factor in human melanoma cells and its role in tumor cell growth and angiogenesis. Biochem Biophys Res Commun. 1999; 264(3):751–758. [PubMed: 10544003]

Siddiqui-Jain A, Grand CL, Bearss DJ, Hurley LH. 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. 2002; 99(18):11593–11598. [PubMed: 12195017]

Simonsson T, Pecinka P, Kubista M. DNA tetraplex formation in the control region of c-myc. Nucleic Acids Res. 1998; 26(5):1167–1172. [PubMed: 9469822]

Sugiyama T, Kohara H, Noda M, Nagasawa T. Maintenance of the hematopoietic stem cell pool by CXCL12-CXCR4 chemokine signaling in bone marrow stromal cell niches. Immunity. 2006; 25(6):977–988. [PubMed: 17174120]

Sun D, Guo K, Rusche JJ, Hurley LH. Facilitation of a structural transition in the polypurine/polypyrimidine tract within the proximal promoter region of the human VEGF gene by the presence of potassium and G-quadruplex-interactive agents. Nucleic Acids Res. 2005; 33(18): 6070–6080. [PubMed: 16239639]

Suzuki J, Miyano-Kurosaki N, Kuwasaki T, Takeuchi H, Kawai G, Takaku H. Inhibition of human immunodeficiency virus type 1 activity in vitro by a new self-stabilized oligonucleotide with guanosine-thymidine quadruplex motifs. J Virol. 2002; 76(6):3015–3022. [PubMed: 11861867]

Telang S, Yalcin A, Clem A, Bucala R, Lane AN, Eaton JW, Chesney J. Ras transformation requires metabolic control by 6-phosphofructo-2-kinase. Oncogene. 2006; 25(55):7225–7234. [PubMed: 16715124]

Teng Y, Girvan AC, Casson LK, Pierce WM Jr, Qian M, Thomas SD, Bates PJ. AS1411 alters the localization of a complex containing protein arginine methyltransferase 5 and nucleolin. Cancer Res. 2007; 67(21):10491–10500. [PubMed: 17974993]

Trent JO, Wang ZX, Murray JL, Shao W, Tamamura H, Fujii N, Peiper SC. Lipid bilayer simulations of CXCR4 with inverse agonists and weak partial agonists. J Biol Chem. 2003; 278(47):47136–47144. [PubMed: 12958314]

Vakser IA. Evaluation of GRAMM low-resolution docking methodology on the hemagglutinin-antibody complex. Proteins Suppl. 1997; 1:226–230.

Van Schaftingen E, Hers HG. Inhibition of fructose-1,6-bisphosphatase by fructose 2,6-biphosphate. Proc Natl Acad Sci U S A. 1981; 78(5):2861–2863. [PubMed: 6265919]

Warburg O. On the origin of cancer cells. Science. 1956; 123:309–314. [PubMed: 13298683]

Warburg O, Posener K, Negelein E. On the metabolism of cencer cells. Biochem Z. 1924; 152:319–344.

Webba da Silva M. Geometric formalism for DNA quadruplex folding. Chemistry. 2007; 13(35): 9738–9745. [PubMed: 17972263]

Wilson JM, Coletta PL, Cuthbert RJ, Scott N, MacLennan K, Hawcroft G, Leng L, Lubetsky JB, Jin KK, Lolis E, Medina F, Brieva JA, Poulsom R, Markham AF, Bucala R, Hull MA. Macrophage migration inhibitory factor promotes intestinal tumorigenesis. Gastroenterology. 2005; 129(5): 1485–1503. [PubMed: 16285950]

Winner M, Meier J, Zierow S, Rendon BE, Crichlow GV, Riggs R, Bucala R, Leng L, Smith N, Lolis E, Trent JO, Mitchell RA. A novel, macrophage migration inhibitory factor suicide substrate inhibits motility and growth of lung cancer cells. Cancer Res. 2008; 68(18):7253–7257. [PubMed: 18794110]

Wright WE, Tesmer VM, Huffman KE, Levene SD, Shay JW. Normal human chromosomes have long G-rich telomeric overhangs at one end. Genes Dev. 1997; 11(21):2801–2809. [PubMed: 9353250]

Xu X, Hamhouyia F, Thomas SD, Burke TJ, Girvan AC, McGregor WG, Trent JO, Miller DM, Bates PJ. Inhibition of DNA replication and induction of S phase cell cycle arrest by G-rich oligonucleotides. J Biol Chem. 2001; 276(46):43221–43230. [PubMed: 11555643]

Yasuoka H, Kodama R, Hirokawa M, Takamura Y, Miyauchi A, Sanke T, Nakamura Y. CXCR4 expression in papillary thyroid carcinoma: induction by nitric oxide and correlation with lymph node metastasis. BMC Cancer. 2008; 8(1):340. [PubMed: 19025611]

Zlotnik A. Chemokines in neoplastic progression. Semin Cancer Biol. 2004; 14(3):181–185. [PubMed: 15246053]

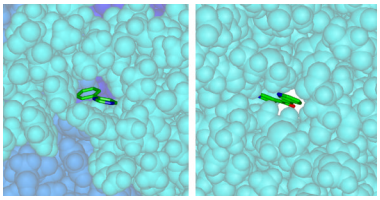

Figure 1.

3PO and 4-IPP in their respective target sites. Left) 3PO in the theoretical docking pose from virtual screening. Right) 4-IPP covalently binds with the N-terminal proline as shown in this X-ray crystal structure (Winner et al., 2008)

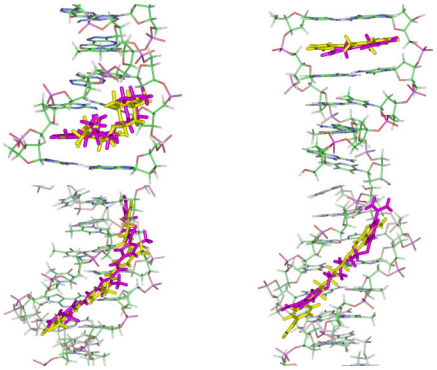

Figure 2.
Comparison of Surflex-Dock poses and X-ray crystallographic pose for A) daunorubicin, B) ellipticine, C) distamycin, and D) pentamidine. Crystallographic pose is in yellow with the Surflex-Dock pose in magenta.

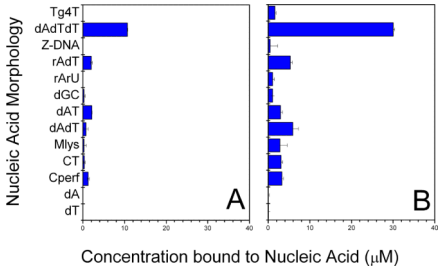

Figure 3.
Competition dialysis analysis of A) MH15 and B) a compound found in the similarity search.