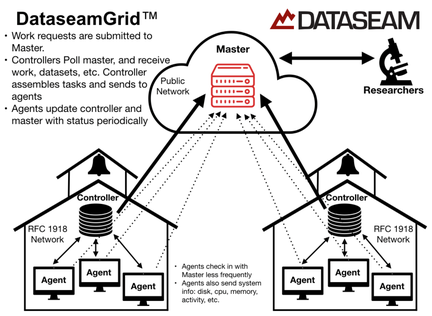
Structure-based Drug Discovery: Computational Virtual Screening

Robert C. Monsen[1], Lynn Deleeuw[3], Jon Maguire[3], William L. Dean[1,3], Robert D. Gray[3], Jonathan B. Chaires[1,2,3], John O. Trent[1,2,3]

Departments of [1]Biochemistry & Molecular Genetics, [2]Medicine, [3]James Graham Brown Cancer Center
University of Louisville, Louisville, KY USA

Computational Resources



Figures 1 & 2. (Left) Overview of the DataseamGrid computing architecture and information flow, and (Top) Kentucky map showing counties with facilities participating in the DataseamGrid (shown in red).

The **DataseamGrid™** Architecture arranges computing nodes, called Agents, in a hierarchical fashion, and utilizes special service nodes in each district called Controllers, which are a dedicated resource. Both Agents and Controllers reside in the school districts inside their private (RFC 1918) network. On the public internet the primary grid host, called Master, cannot initiate communications through the various district firewalls, and thus relies on polling requests from the Controllers (and Agents) to distribute work, collection results, and maintain metadata concerning the overall status of the system. The DataseamGrid was initially developed as a primarily government funded replacement for Apple Inc.'s xGrid, which had an analogous architecture.

Job Submission is by researchers that submit versioned datasets to the Master, which synchronizes with each active Controller. Work request submission then simply references the versioned datasets, and the runnable package with all the required components is assembled on the Controller, which itself is responsible for scheduling the task and pushing to the Agent. **Figure 1** provides a basic representation.

The benefits of the DataseamGrid architecture:

- Reduced network traffic across a district's link when moving large datasets.
- Insulation of computational activity from temporary network segmentation events at the district level.
- Allows for tracking of participation by district.
- Provides redundant agent tracking metrics, reducing chances of over- or under-allocation of work to a particular district controller.
- Single point of contact for work submission and management, which simplifies use.
- Ease of scalability.

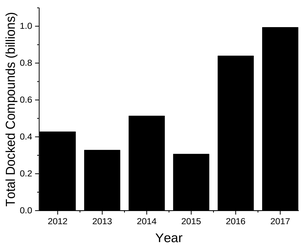


Figure 3. Number of docking calculations performed per year using the DataseamGrid.

Q-serve is our in-house queueing and analysis pipeline for job submission to the DataseamGrid, built by Jon Maguire. Q-serve has three components to allow for a fully automated job submission: a queue file, configuration file, and a master script. The **Queue file** contains: target file (receptor: protein or nucleic acid 3D structure), protomol (negative image of receptor docking site), library of compounds, and the type of job we want to run. The **Configuration file** contains variables which are necessary for the master script, such as what utilization to maintain and the location of other scripts. The **Master script** runs periodically using crontab and initiates a series of checks (if it is already running, utilization level, etc.) to see if it is under the utilization threshold and, if so, submits another job from the queue. This system runs 24/7/365 and has allowed us to reach >1 billion docking computations per year (**Figure 3**). The following is a simplified diagram of the Q-serve implementation:

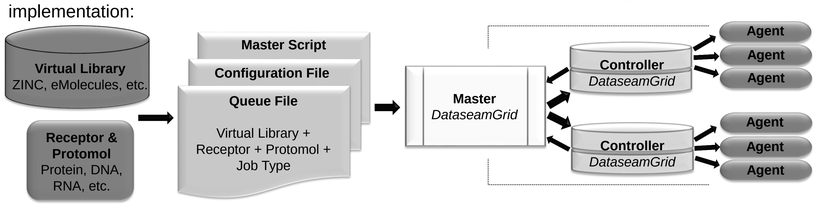


Figure 4. Graphical representation of the Q-serve to DataseamGrid pipeline.

Acknowledgments

- Dept. Biochemistry & Molecular Genetics, UofL, Louisville KY USA
- Kentucky Lung Cancer Research Program

Virtual Screening

Goal: To use the vast computational power of the KY DataseamGrid to rapidly enrich for lead molecules in the design of novel therapeutics.

Method: Using Surflex-Dock[2], a flexible ligand docking software, on the DataseamGrid we can rapidly screen libraries containing thousands to millions of compounds against protein and nucleic acid targets that are implicated in disease. A total of 28,000 dedicated virtual processor cores (4,000 UofL, 24,000 Dataseam) allows for computation time equivalent to >2,300 CPU years per month. Routine screening libraries include the ZINC Drug-like library (24,877,119 compounds), eMolecules (7,016,711 compounds), Molplex (2,100,000 compounds), PubChem (43,010,233). Using the pipeline outlined in **Figure 4** we can rapidly screen and enrich for novel scaffolds that can be further refined as lead hits.

Results: Identification of more than 38 experimentally validated inhibitors or activators (modulators of protein function) of proteins implicated in various diseases (**Table 1**) from virtual screening with hit rates from 10-66%. This has resulted in >45 research disclosures as well as 19 separate patents (**Table 2**), with two drugs making it into clinical trials thus far.

Virtual Screening Targets	
PFKFB3, PFKFB4	N-terminal methyltransferases
Choline Kinase	Venezuelan equine encephalitis virus
Macrophage Migration Inhibitor Factor	Growth Factor Independent 1 (GFI1)
Chemokine Receptor 4 (CXCR4)	TNF Receptor-associated Factor 6
Nucleolin	Iron Regulatory Protein 1
Arylamine N-acetyltransferases (NAT1, NAT2)	Sphingosine Kinase 1
APC2, APC11	E3 Ubiquitin-protein Ligase (UBR5)
Aminoacyltransferase	Chemokine CXCL11
Leukotriene B4 Receptors (BLT1, BLT 2)	Nuclear Import Receptor (Kpnβ1)
COP9 Signalosome Complex Subunit 5 (CSN5)	Transaldolase
CaMKK2	Galectin-1
RAS	Bcl-2
RAL	Glutaminase
Interleukin Receptor 6 (IL-6)	IκB kinase 2 (IKK2)
c-Myc	RNA Helicase A
hTERT	Lactate Dehydrogenase A
Pyruvate Carboxylase	DNA Repair Protein (REV1)
D-Dopachrome Tautomerase	SOX9
Phosphoserine Aminotransferase	BAX

Patents
Bates, P. J.; Miller, D. M.; Trent, J. O.; Xu, X. ; Method for the diagnosis and prognosis of malignant diseases. US20050053607A1. 2005.
Chand, P.; Chesney, J. A.; Clem, B. F.; Tapolsky, G. H.; Telang, S.; Trent, J. O.; Small-molecule choline kinase inhibitors as anti-cancer therapeutics. US20110257211A1. 2011.
Chand, P.; Chesney, J. A.; Clem, B. F.; Tapolsky, G. H.; Telang, S.; Trent, J. O.; Small molecule inhibitors of 6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3 and glycolytic flux and their methods of use as anti-cancer therapeutics. WO2011103557A1. 2011.
Clem, B.; Telang, S.; Trent, J.; Chesney, J.; Chand, P.; Tapolsky, G.; Small molecule choline kinase inhibitors, screening assays, and methods for safe and effective treatment of neoplastic disorders. US20110212994A1. 2011.
Cunningham, A.; Trent, J. O.; Computer-based hybrid fragment-ligand model for classifying chemical compounds. WO2012031215A1. 2012.
Ghim, S.-J.; Jenson, A. B.; Trent, J. O.; Use of VP2 proteins and virus-like particles in vaccines for preventing and treating erythrovirus or parvovirus B19 infections. WO2015138424A1. 2015.
Leaner, V. D.; Van der Watt, P. J.; Trent, J. O.; Small molecule inhibitors for cancer therapy. WO2016020892A1. 2016.
Li, C.; Trent, J. O.; Bax-activating cancer therapeutics. US20150190379A1. 2015.
Miller, D. M.; Bates, P. J.; Trent, J. O.; Antiproliferative activity of G-rich oligonucleotides and method of using same to bind to nucleolin. WO2000061597A1. 2000.
Miller, D. M.; Bates, P. J.; Trent, J. O.; Xu, X.; Non-antisense guanosine-rich nucleolin-binding oligonucleotides (GROs) for the diagnosis and prognosis of malignant diseases. US20030194754A1. 2003.
Mitchell, R. A.; Trent, J. O.; Chand, P.; Tapolsky, G. H.; Preparation of iodopyrimidine derivatives for use in the treatment of macrophage migration inhibitory factor (MIF)-implicated diseases. WO2011038234A2. 2011.
Mitchell, R. A.; Trent, J. O.; Meier, J. B.; Macrophage migration inhibitory factor antagonists and methods of using same. WO2007140263A2. 2007.
States, J. C.; Taylor, B. F.; Trent, J. O.; Methods of treating cancer by modulating anaphase-promoting complex/cyclosome (APC/C) expression. WO2012149266A1. 2012.
Tapolsky, G.; Chand, P.; Trent, J. O.; Telang, S.; Clem, B. F.; Chesney, J. A.; Family of 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase 3 (PFKFB3) inhibitors with anti-neoplastic activities. US20120177749A1. 2012.
Tezel, T. H.; Kaplan, H. J.; Mitchell, R. A.; Trent, J. O.; Methods and compositions for modulating ocular damage with a modulator of a migration inhibitory factor (MIF). WO2011146824A1. 2011.
Trent, J. O.; Bates, P. J.; Miller, D. M.; Antiproliferative activity of G-rich oligonucleotides and method of using same to bind to nucleolin. US20080318890A1. 2008.
Trent, J. O.; Bates, P. J.; Miller, D. M.; Antiproliferative, nucleolin-binding activity of g-rich oligonucleotides for cancer therapy. US20080318888A1. 2008.
Trent, J. O.; Bates, P. J.; Miller, D. M.; Antiproliferative activity of G-rich oligonucleotides and method of using same to bind to nucleolin. US20110178161A1. 2011.
Trent, J. O.; Meier, J. B.; Napier, K. B.; Compounds targeting CXCR4 or a G protein for treating disease, for administering, and for pharmaceutical compositions. WO2011127333A2. 2011.

Tables 1 & 2. (Top) List of successfully targeted proteins using our massively parallel screening approach. (Right) List of patents generated as a result of successful virtual screening campaigns.

Methods

- **Circular Dichroism (CD)**: Spectra were obtained in TBAP (200mM KCl, 1mM EDTA, 10mM tetrabutylammonium phosphate, pH 7) at 20.0°C in a JASCO-710 spectropolarimeter for annealed segments and the full, SEC purified hTERT sequence to Identify conformation and number of stacked G-quartets. Oligos were purchased from IDT and Eurofins. Parameters were 20.0°C cell temperature, scan range of 450nm (or less) to 220nm, 1nm Pitch, 4s response, and averaging of 4 scans. Spectra were normalized to concentration using the following formula: $\Delta\varepsilon = \text{mdeg}/(32{,}980 * C * L)$.
- **Analytical Ultracentrifugation (AUC):** Sedimentation velocity experiments were carried out on a Beckman Coulter ProteomeLab XL-A centrifuge in TBAP buffer at 20.0°C with a run speed of 50k rpm in continuous mode. Analysis was performed in Sedfit using a continuous distribution C(s) model (non-interacting species) with 100 scans using wavelengths appropriate for each species (260nm for DNA, 310-320nm for compounds).
- **Electron Microscopy (OpNS-EM):** hTERT quadruplex DNA was annealed in Tris buffer (100mM KCl) and purified by SEC followed by concentration to ~100uM in Amicon 3K MWCO concentrators. Samples were then sent to Creative Biostructure for analysis.
- **Molecular Dynamics (MD):** The G-quadruplex structure was created as outlined previously[1] and MD simulations were run using the Amber suite of tools on a workstation equipped with two GeForce GTX Titan GPUs. Analysis and clustering of the 100ns trajectories was performed using the cpptraj module of AmberTools. Calculations of hydrodynamic properties were done using the program HYDROPRO 10[3].
- **Virtual Drug Screening:** Docking was performed using Surflex-Dock 2.11[2] on the Kentucky DataseamGrid. Over 45 million compounds were docked at 12 different locations of hTERT from the 2014 and 2016 drug-like libraries from the ZINC database as-is. Protomols were generated for 12 sites positioned in the loop structures based on the G-quartet residues. Sixty-nine compounds were purchased from MolPort based on centroids after clustering using a Tanimoto similarity of 0.9 in Schrödinger's Canvas application.
- **Fluorescent Screening:** FRET-pair labelled oligonucleotides were annealed in BPEK buffer (6mM Na_2HPO_4, 2mM NaH_2PO_4, 1mM Na_2EDTA, 185mM KCl, pH 7.2) and mixed in 96-well plates at a 100:1 ratio of compound to quadruplex and fluorescent emission was measured at 520nm from 20.2 to 99.0°C in an AB Applied Biosystems Step One RT-PCR. Melt temperatures were determined from the peak of the 1st derivative of the normalized fluorescence emission curves.
- **Hydrodynamic Calculations:** Hydrodynamic dimensions were predicted using HYDFIT[4] based on experimentally derived values.

References

1. Chaires, J. B., Trent, J. O., Gray, R. D., Dean, W. L., Buscaglia, R., Thomas, S. D., and Miller, D. M. (2014) An improved model for the hTERT promoter quadruplex. PLoS One 9, e115580
2. Jain, A. N. (2003) Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. J Med Chem 46, 499-511
3. A. Ortega, D. Amoros and J. García de la Torre, "Prediction of hydrodynamic and other solution properties of rigid proteins from atomic and residue-level models", Biophysical Journal 101, 892-898 (2011)
4. D. Amorós, A. Ortega, S. E. Harding and J. García de la Torre. "Multi-scale calculation and global-fit analysis of hydrodynamic properties of biological macromolecules: determination of the overall conformation of antibody IgG molecules". European Biophysics Journal 39, 361-370 (2010)

hTERT G4 Drug Discovery

Background: The enzyme telomerase is responsible for maintenance of telomeric DNA, which resides at the ends of chromosomes to protect them from degradation due to multiple rounds of replication. Human telomerase reverse transcriptase (hTERT), which is not typically present in normal somatic cells, has been found to be up-regulated in >85% of cancers and allows for their unlimited growth potential. The hTERT core promoter spans from -180 to +1 bases upstream of the transcriptional start site. This region contains many runs of guanine tracts which may enable the formation of multiple G-quadruplexes (G4s) which are four-stranded DNA structures[1]. Recent investigations found that >70% of glioblastomas and melanomas that overexpress telomerase contain mutations in this region and are often the result of point mutations within the putative quadruplex G-tracts. Therefore, the hTERT G-quadruplex is an excellent target for the repression of telomerase. Here we show the results of molecular dynamic and modeling simulations, as well as *in silico* drug screening aimed at selectively targeting the hTERT quadruplex.

Results 1: Biophysical analyses, molecular dynamics, and EM imaging of the hTERT full core promoter sequence suggest it forms three contiguous, stacked, parallel quadruplexes (**Figures 5 & 6**).

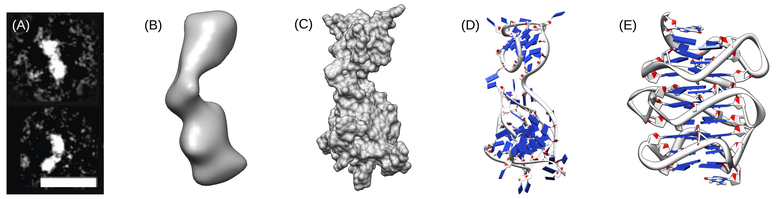


Figure 5. (A) Negative Staining Electron Micrograph (OpNS-EM) of hTERT quadruplex sequence in Tris buffer (100mM KCl) (scale bar = 100Å). (B) Single particle reconstruction of EM. (C) Space-filling model of hTERT G4 without 5' quadruplex stacking (0.96 correlation coefficient when compared to 5b). (D) MD derived structure of hTERT G4 without 5' quadruplex stacking. (E) MD derived hTERT G4 with all three quadruplexes stacked (with extraneous bases removed for clarity).

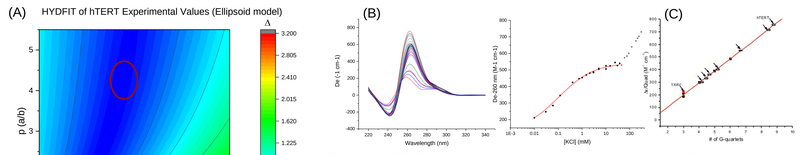
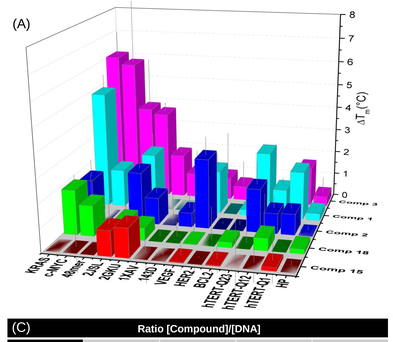
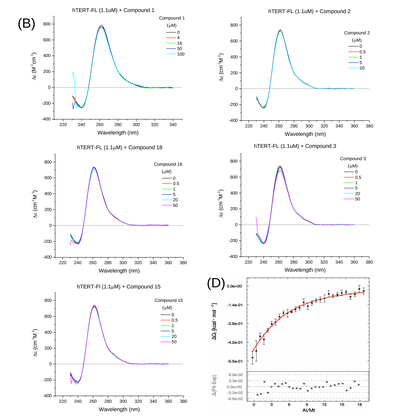


Figure 6. (A) Plot of axial ratio (p=a/b) vs. longest axis (a) predicted from hydrodynamic measurements in 200mM KCl TBAP buffer, indicating a structure of no more than ~55 angstrom length. (B) Circular dichroism spectra showing hTERT's dependence on [K+] and the increase in intensity at 260nm. (C) Relationship between molar absorptivity at 260nm and the number of contiguous G-quartets.

Results 2: The full length, contiguously stacked hTERT quadruplex (**Figure 5e**) was prepared as a receptor and submitted to the Q-serve/DataseamGrid pipeline where roughly 45 million compounds from the ZINC database were docked to a total of 12 sites (protomols) generated around the loops and grooves (~500 million docking calculations). The top 12,000 hits were clustered and sorted to remove redundancies and 69 were selected based on drug-likeness and visual inspection. Among the 69 tested, 5 have so far have been confirmed as hTERT G4 interacting ligands and are currently being characterized for their mode of binding.





(C) Ratio [Compound]/[DNA]				
Sequence	**Comp 2**	**Comp 3**	**Comp 15**	**Comp 18**
Q12	1.4	3.1	1.9	1.1
Q23	1.7	3.8	2	1.7
hTERT-FL	2.6	8.6	3.6	2.1

Figure 7. Results of initial drug screening. (A) FRET melting temperatures for various G4s in the presence of saturating compounds, demonstrating selective stabilization. (B) CD titrations of hTERT-FL G4 with 0-100 μM of compounds 1, 2, 3, 15, and 18 showing no ICD or change in overall structure with the exception of compound 3. (C) Table of stoichiometries of binding from AUC measurements where quadruplexes 1, 2, and 3 of the full length hTERT were used for screening purposes (i.e. Q12 contains the first two G4s and Q23 the last two G4s). (D) Representative ITC titration curve for compound 1 against the hTERT partial sequence Q23.