Figure 1. Overview of the DatasetGrid computing architecture and information flow. (Top) Kentucky map showing counties with facilities participating in the DatasetGrid (shaded in gray). The DatasetGrid 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 (PIG) 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, collect results, and maintain metadata concerning the state of the system. The DatasetGrid was initially developed as a primarily government-funded replacement for Apple Inc.'s grid, which had an analogous architecture.

Job Submission is handled by passing jobs to the service nodes called Agents, which synchronize with each active Controller. Work request submission then simply references the versioned datasets, and the runnable package with the required components is assembled on the Controller, which itself is responsible for scheduling the task and passing the Agent. Figure 1 provides a basic implementation of the DatasetGrid architecture:

- Reduced network traffic across a district's link when moving large datasets.
- Insulation of computational activity from temporary network segmentation events at the school 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.
- Supports multiple Agents for work submission and management, which simplifies use.
- Ease of scalability.

G-service is our in-house querying and analysis pipeline for job submission to the DatasetGrid, built by Jon Maguire. G-service has three components that allow for a fully automated job submission: a queue file, configuration file, and a master script. The Queue file contains: target file (protein, or nucleic acid 3D structure), protein sequence (or PDB receptor docking site), list 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 cron 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 G-service implementation:

![Diagram](image)

Figure 4. Graphical representation of the G-service in DatasetGrid pipeline.


Figure 5. (A) Fluorescence Emission Spectra of UBR5(ΔC1) and UBR5(ΔC2) in the presence of hTERT-FL (1.1 uM) + Compound 3 (100 μM). (B) Single point reactivity of APC2, APC11, and APC10 with hTERT-FL (1.1 uM) + Compound 3 (100 μM). (C) Mapping model of hTERT-FL (1.1 uM) + Compound 3 (100 μM) with three quadruplexes docked (with extraneous bases removed for clarity).

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

Methods

hTERT quadruplexes were characterized using circular dichroism (CD), UV-Vis, and Optical Emission Spectroscopy (OES), with the resonance energy transfer (RET) technique to determine the binding constant (Kd) of the compounds.

1. Circular Dichroism (CD): hTERT quadruplexes were determined in 20 mM Tris-HCl, 150 mM KCl, 10 mM L-glutathione, pH 7.5, at 20°C in a JASCO-710 spectrophotometer for annealed samples and the full, SSC purified hTERT sequence in identity contact with filter affinity matrix. The melting temperature (Tm) temperatures were determined from the figure 3.

2. Analytical Ultracentrifugation (AUC): Sedimentation velocity experiments were carried out on an Beckman Coulter ProteomLab XL-A analytical ultracentrifuge in 20 mM Tris-HCl, 150 mM KCl, pH 7.5, at 20°C in 10-mm clear analytical cells. Sedimentation coefficients and diffusion coefficients were determined using SEDFIT software. AUC experiments were performed in 20 mM Tris-HCl, 150 mM KCl, pH 7.5, and purified by SEC following anion exchange chromatography and negatively charged oligonucleotides.

3. Molecular Dynamics (MD): The molecular structure used were calculated using the protein LDB900. Evolution of the hTERT quadruplex was performed using the program LDB900 (Borgstrom et al., 2012).

4. Virtual Docking: Docking was performed using Surflex-Dock 2.19 on the Kentucky DatasetGrid. Over 45 million compounds were docked at 12 different locations of hTERT from the 2014 and 2016 drug libraries from the ZINC database as well as obtained from SixPaths.org. The best scoring docked hits were displayed using PyMol (DeLano, W.L., 2002). The PyMol program was developed by DeLano Scientific, LLC and runs on all common operating systems, including Windows, Mac OS, and Linux.

5. Quantification of Hit (H): Hit compounds were selected based on a 10-fold enrichment of the hit compounds over the background compounds.

6. Fluorescence Emission Spectroscopy: The emission spectra were recorded using a JASCO FP-6500 fluorescence spectrophotometer. Emission spectra were recorded from 20 to 30°C and 5°C in continuous mode. The excitation wavelength was 280 nm and the emission wavelength was from 300 to 400 nm.

7. Optical Emission Spectroscopy (OES): The optical emission spectroscopy was performed using a Jobin Yvon H2000 spectrophotometer. The excitation wavelength was 280 nm and the emission wavelength was from 300 to 400 nm.

8. Resonance Energy Transfer (RET): The resonance energy transfer (RET) technique was applied to determine the binding constant (Kd) of the compounds.

9. Results: The full length, contiguously structured hTERT quadruplex (Figure 5) was prepared as an agent and submitted to the G-service/DatasetGrid pipeline where roughly 45 million compounds from the ZINC database were docked to a total of 12 sites (proteins) 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 been confirmed as hTERT interacting ligands and are currently being characterized for their mode of binding.

Acknowledgments:

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