OpenEye Scientific Software: Drug Discovery – The Best in Class Tools for Competitive Advantage

Table of Contents

January 2005

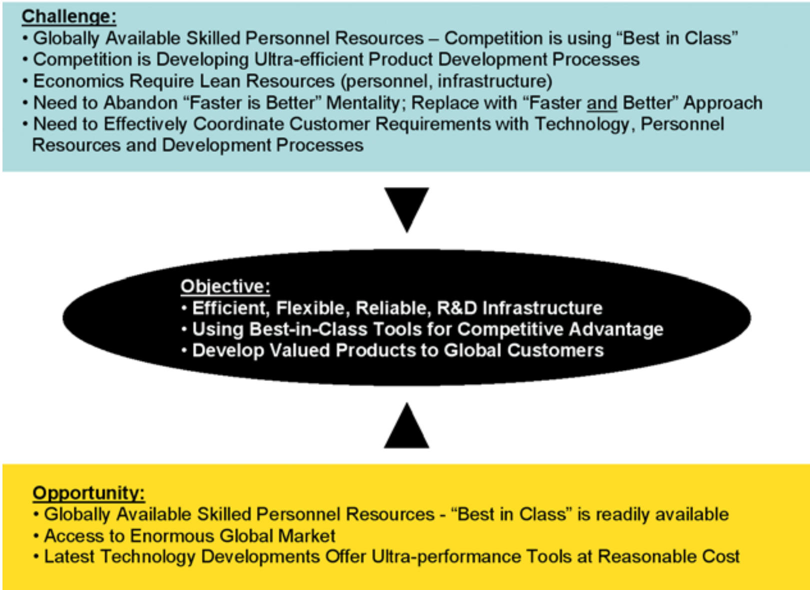
Global Technology Development: The Challenges and Opportunities

The late 20th century technological and industrial efficiencies have resulted in an integrated global environment where R&D, manufacturing, distribution, finance, and sales can maximize regional efficiencies and productivity. This "tightly integrated global market" offers significant advantages to manufacturers of all types with an enormous customer base and a product development infrastructure which allows management teams to quickly adjust global resources based on market requirements and regional capabilities. Clearly, real-time worldwide asset management would not be possible without advanced technologies and, together with resource management opportunities, R&D and manufacturing organizations have the greatest potential for significant improvement in productivity. However, since this environment has taken the overall efficiency of global product development to a level of ultra-efficiency, great pressures have been placed on R&D teams to maximize their productivity.

To manage these economic and productivity pressures, scientific R&D organizations are re-evaluating the processes and tools they rely on to get their work done more efficiently. In drug discovery and development, for example, companies can now spend over $1 billion and 10 years to bring a new drug into the market[1]. In the recent past, a common way to decrease drug development R&D costs was to use inexpensive mass-marketed technologies wherever possible. In the field of high-performance computing, this resulted in a focus on "microprocessor performance" and large clusters built by connecting off-the-shelf computers with high performance networks. While this strategy has proven very successful for some workloads, many organizations now realize that to truly improve the product development process and their global competitiveness, they need a comprehensive computing solution in which the appropriate microprocessors are combined with a high performance system infrastructure and tuned software.

[1] See: "The Soaring Cost of Drug Discovery," in Bio-It World, December 11, 2003.

Forces Affecting Competitive Global Product Development

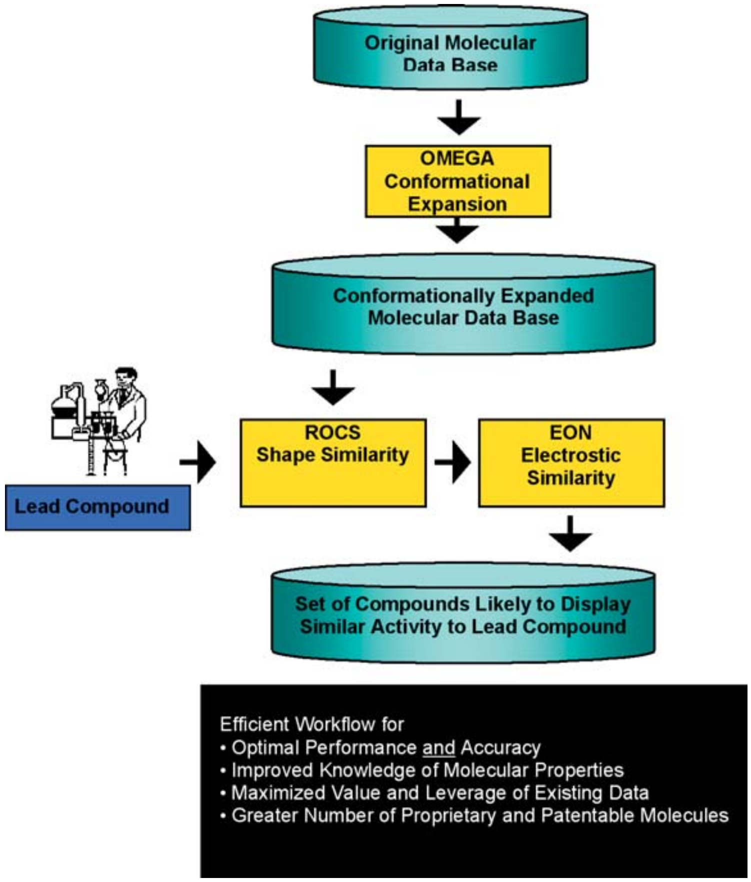


Drug Discovery: The Best in Class Tools for Competitive Advantage

In the domain of product development for technology-based industries, improving the process of drug discovery remains one of the greatest challenges in the business community. Clearly, significant improvements in the early discovery phases of the development process must be achieved to decrease both the 10-year-cycle time and $1 billion required to successfully launch a drug. In the past decade High Throughput Screening (HTS) technologies have leveraged robotics and miniaturized assays to automate the identification of proprietary lead molecules and subsequently validate their effects on drug targets. These technologies have improved the efficiency and dependability of lead identification and testing and have become part of the standard discovery process. Most drug discovery organizations now depend upon this relatively new infrastructure for process improvement and to remain competitive.

Although HTS is a widely applicable technology, scientists have been challenged to further leverage the initial lead molecules by using them as a template molecule so that a number of proprietary molecules with similar activity but better drug profile can be developed. This expansion of lead molecules can be achieved in silico through virtual high throughput screening (vHTS) technologies. Effective virtual screening of compounds can significantly reduce the cost and time involved in drug discovery by improving the quality of new structures chemists synthesize, while at the same time reducing the number of structures synthesized for each project. Since the list of synthetically tractable compounds can be unlimited, vHTS can direct combinatorial library design and eliminate unpromising compounds before they are synthesized.

Effective implementation of vHTS is dependent upon the coordination of sophisticated software with an effective and reliable platform for running the software. OpenEye Scientific Software products OMEGA, ROCS and EON are designed to rapidly, reliably, and accurately quantify the shape and electrostatics of a molecule for the purposes of lead discovery. The software is based on the principle that chemical interactions are governed by shape and electrostatics, and it is designed for scientific rigor, speed, and scalability, thus enabling the screening of millions of compounds in a reasonable time frame, rather than mere thousands.



Conformational expansion is typically the first step in the vHTS workflow. Designed for use with the large libraries required for computer-aided drug design, OMEGA generates multi-conformer structure databases with unprecedented speed and reliability. Conformational expansion of drug-like molecules can be performed in fractions of a second, yielding a throughput of hundreds of thousands of compounds per processor per day. Although similar programs are available for generating conformational expansions, “OMEGA is three orders of magnitude faster than comparably accurate products on the market,” says George Vacek, Ph.D., OpenEye’s Vice President of Business. Researchers at Vertex Pharmaceuticals have concluded “OMEGA achieves the best balance between speed and performance among the programs examined, and it appears to be the most suitable tool for conformational analysis on large chemical databases.”[2]

Next in the vHTS workflow, the hundreds of millions of conformational structures need to be reduced to a tractable set by screening for specific properties. In the case of ROCS and EON, the library is screened for shape and electrostatics similarity to a lead compound, which is known to be active against a protein target. ROCS, a molecular shape comparison program which determines a Tanimoto measure of the shape similarity as its metric for evaluation, is capable of processing 600-800 comparisons each second making it possible to rapidly search multiconformer representations of large libraries. Similarly, EON can be used to screen for electrostatic similarity; EON compares through-space electrostatics near molecules and determines rigorous Tanimoto measures between the fields. Typically, EON is used to analyze the output structures from ROCS, thus producing compounds with both similar shape and similar electrostatics to the known active compound. Thus the combination of OMEGA, ROCS, and EON can enhance the drug discovery process through the rapid identification of molecules expected to demonstrate similar activity.

[2] See: "Conformational Analysis of Drug-Like Molecules Bound to Proteins: An Extensive Study of Ligand Reorganization upon Binding," E. Perola and P.S. Charifson, J. Med. Chem. 47 (2004) 2499-2510.

SGI® Altix®, Intel® Itanium® 2 processors, and OMEGA, ROCS, and EON: A Winning Combination for Production and Reliability

vHTS is extremely computationally intensive and, to achieve a competitive advantage by obtaining the greatest value from this OpenEye series of applications, it is critical that a flexible and effective high-performance computational platform is chosen. Beyond simply increasing the speed of data analysis, the computing infrastructure must offer scientists the ability to ask new questions of the data and, ultimately, bring their research forward into new areas. However, while intellectual freedom is necessary to obtain a competitive advantage, today's reality is that computing solutions are cost-constrained to systems based on high-volume microprocessors combined with the now-standard open source Linux® operating system. Despite these restrictions, there are numerous choices and a careful assessment of various parameters should be considered before choosing a computing system. These include:

- The number of users and size of their problems
- Size of data and how data is accessed (e.g. disk access vs. resident in memory)
- Type of computational power needed:
  - Number of processors required to process jobs
  - Type of math needed in calculations (e.g. integer vs. floating point math)
  - 32-bit vs. 64-bit
  - Speed of connection between components (e.g., processors, memory, visualization, other peripherals)
- Requirement for visualization
- Availability of software to solve real problems on the system of choice

Based on the Intel® Itanium® 2 microprocessor, the SGI Altix system is an excellent platform for running the OpenEye applications. It satisfies the requirements for a standard microprocessor-based system running Linux and offers unique capabilities that will support state-of-the-art R&D.

The Itanium 2 microprocessor is based on the Explicitly Parallel Instruction Computing or EPIC architecture. Since the instruction set is explicitly parallel, the architecture attains high levels of parallelism inside the processor (so called "instruction level parallelism"). Leveraging the EPIC architecture of the Itanium 2 processor, Altix adds parallel system architecture, a massive set of internal resources, and high bandwidth connection to I/O and memory, delivering industry leading performance to scientific computing applications. This means more jobs can be run and more users serviced in less time. "The Itanium 2 processor's EPIC architecture is perfect for the calculations involved in vHTS software" said Tim Mattson Ph.D., Global Manager of Life Sciences at Intel. "Additionally, the Altix interconnect lets applications scale up to large numbers of processors. And with the large shared address space provided by the NUMA architecture in the Altix system, programming is dramatically simplified."

SGI Altix family of servers

In addition to the raw microprocessor performance offered by the Itanium 2 processor, 64-bit Altix offers scientists a virtually unlimited memory capability in a tightly coupled system ultra-high speed communication environment that connects the system's microprocessors, memory, visualization graphic processor units (GPUs) and other peripherals. With this capability, the scalable Altix shared-memory platform enables research teams to manage CPU and memory resources more effectively so the most urgent jobs complete more quickly. This ability to dynamically allocate resources enables scientists to be more flexible and hence make more effective use of their server. Using these tools, a mix of jobs having differing requirements for processing and memory resources can easily be managed since they can independently draw from the pooled

Altix resources according to their needs. This job management scenario is typical in drug discovery and tools that can easily address this problem provide R&D teams significant productivity advantages.

> ***"OMEGA, ROCS, and EON are very robust on Altix."*** *Ryszard Czerminski, Ph.D., Senior Scientist, Novartis Institutes for Biomedical Research.*

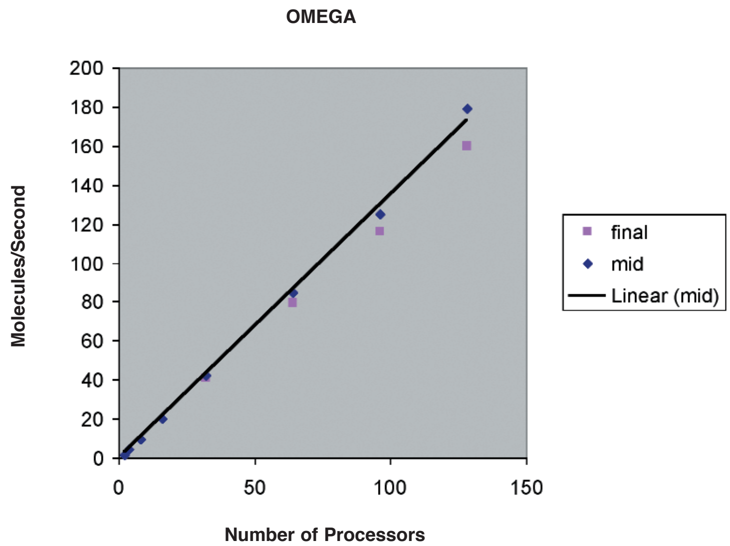
Ultimately, the highly integrated Itanium 2-based Altix server, together with the available programming tools needed for software to utilize these capabilities, provides optimal performance for OpenEye's OMEGA, ROCS, and EON applications. At Novartis, OpenEye tools run on the SGI Altix system to provide a significantly improved usage of computational resources compared to their standard Linux cluster. "OMEGA, ROCS, and EON are very robust on Altix. We have been using it for weeks without having a problem," says Ryszard Czerminski, Ph.D., Senior Scientist, Novartis Institutes for Biomedical Research.

OMEGA, ROCS and EON: Performance on SGI Altix

The OpenEye OMEGA and ROCS performance tests on Altix demonstrate dramatic results. Each of the performance tests were performed on a dedicated Altix® 3000 consisting of 128 Itanium 2 processors with 448 GB of shared memory. Regardless of the number of processors being used, in each case, the listed job was the only process running on the system.

OMEGA PERFORMANCE. The initial OMEGA test was performed with the default parameters (message compression on, message-size = 3) using one master and a varying numbers of compute slaves. Results are shown for the average molecules/sec after completion of the run ("final") and the average molecules/sec after steady-state is reached in the middle of the run ("mid"). Data for lower processor counts were only collected for the middle of the tests because full analysis would take too long. The results of this test demonstrate the near-linear speedup of OMEGA on Altix (Figure1). The "final" numbers fall off relative to the "mid" rates because some nodes finish their tasks earlier than other nodes.

Figure 1: OMEGA test demonstrating the near-linear speedup of OMEGA on Altix



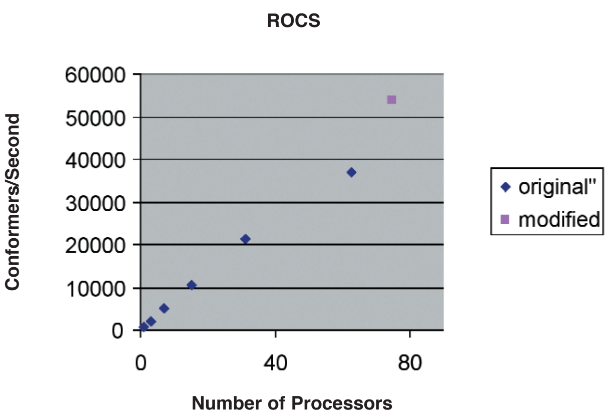
OMEGA's default value for the varying number of molecules to be passed in a message had been determined to be 3 and was based on previous development on a 40 processor cluster. However, on Altix with 128 processors, the optimum value for this parameter is between 8 and 12. When set in this range, maximum efficiency of 111% and 94.5% were achieved for mid-run and final throughput, respectively. The >100% efficiency is probably due to reduction of over-head costs that affect the master, but not the slaves, and due to variance in the processing time of each message.

ROCS PERFORMANCE. Initial testing of ROCS demonstrated an over-taxed master when more than 40 slaves were operating with a message size of 20. However, when the handling of the input format was adjusted and the message size was set to 4, a throughput of 54,037 conformers/second was achieved with 75 slaves (Figure 2). This represents an outstanding 98.7% efficiency relative to the single-slave mode. More importantly, this allowed rapidly processing a 1.1 million

"This outstanding performance allows researchers to complete this analysis in less than 30 minutes on Altix." *Bob Tolbert, Ph.D., OpenEye's Vice President of Development*

compound data-base, containing almost 80 million conformers. According to Bob Tolbert, Ph.D., OpenEye's Vice President of Development, "This outstanding performance allows researchers to complete this analysis in less than 30 minutes on Altix. Now researchers can explore their data in 'real time' rather than waste their time waiting for the same old day-long analyses."

Figure 2. ROCS test demonstrating an outstanding 98.7% efficiency relative to the single-slave mode



Lessons Learned

Pharmaceutical companies using OpenEye applications including OMEGA, ROCS, and EON on the SGI Altix platform with Intel Itanium 2 processors are finding that:

- The shared-memory architecture and the high speed system component communication in the SGI Altix platform enables users of OpenEye applications for drug discovery to extract maximum performance from the Intel Itanium 2 processor and dramatically improve application performance.
- The SGI Altix platform combined with OpenEye's OMEGA, ROCS and EON permits a faster analysis of molecules, their structures and their conformers enabling less costly drug discovery and a competitive advantage.
- SGI Altix and Itanium 2 microarchitectures give drug discovery research organizations a clear roadmap for cost-effective productivity and growth.

3754 [02.11.2005]

J14783