

High-Throughput Screening in Silico

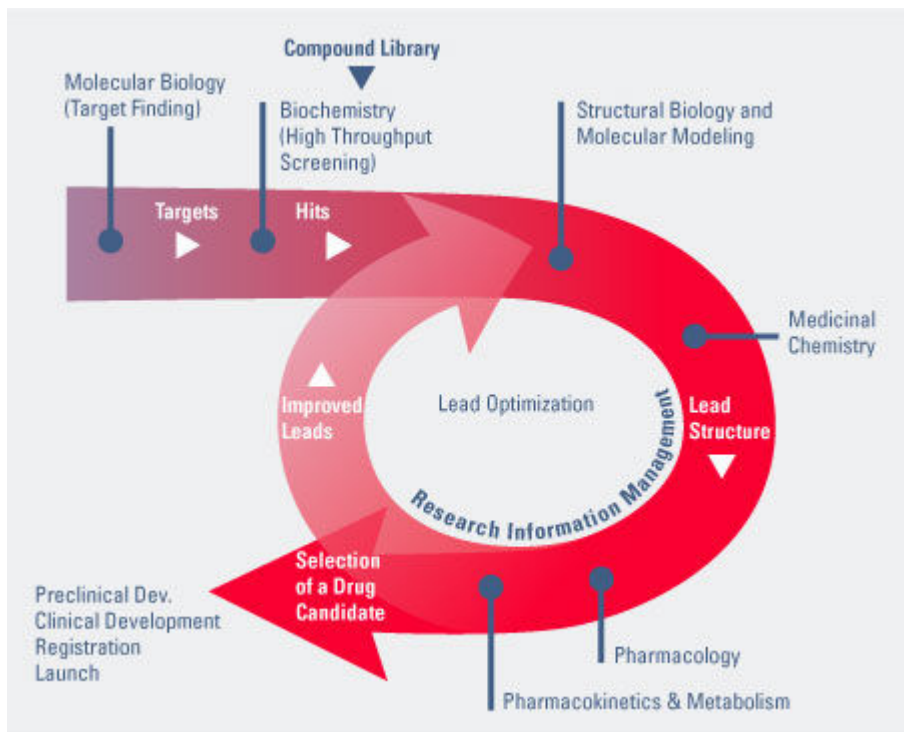
“The costs of discovering and developing new drugs are increasing faster than our research budgets,” states Steven Galson, MD, MPH Acting Director, FDA Center for Drug Evaluation and Research.

Discovering and bringing one new drug to the public typically costs a pharmaceutical or biotechnology company nearly \$900 million and takes an average of 10 to 12 years. These costs are driving companies to seek more efficient methods of testing compounds in the area of absorption, distribution, metabolism, excretion and toxicity. The major goal of these tests is to identify the problematic compounds early in the drug discovery process, which increases the success rate of the compounds that do make it into the later, more expensive stages of the process.

The Drug Discovery Process

The pharmaceutical industry is increasingly operating in a world where medicines have to add real value in an environment where costs are under constant pressure. This is the background to the drivers that are causing the evolution of the drug discovery process.

The following sequence of research activities begins the process that results in development of new medicines: ¹



Target identification. Drugs usually act on either cellular or genetic chemicals in the body, known as targets, which are believed to be associated with the disease. Scientists use a variety of techniques to identify and isolate a target and learn more about its functions and how these influence disease. Compounds are then identified that have various interactions with drug targets helpful in the treatment of a specific disease. High-throughput screening (HTS) is used during this process to rapidly assess the activity of a large number of compounds or extracts on a given target. Today, HTS commonly involves semi-automation or full automation for liquid handling, sample preparation, running of the actual assays, as well as data analysis. HTS laboratories frequently employ robots and the latest detection technologies for assay readouts.

Target prioritization/validation. To select targets most likely to be useful in the development of new treatments for disease, researchers analyze and compare each drug target to others based on their association with a specific disease and their ability to regulate biological and chemical compounds in the body. Tests are conducted to confirm that interactions with the drug target are associated with a desired change in the

behavior of diseased cells. Research scientists can then identify compounds that have an effect on the target selected.

Lead identification. A lead compound or substance is one that is believed to have potential to treat disease. Laboratory scientists can compare known substances with new compounds to determine their likelihood of success. Leads are sometimes developed as collections, or libraries, of individual molecules that possess properties needed in a new drug. Testing is then done on each of these molecules to confirm its effect on the drug target.

Lead optimization. Lead optimization compares the properties of various lead compounds and provides information to help pharmaceutical and biotechnology companies select the compound or compounds with the greatest potential to be developed into safe and effective medicines. Often during this same stage of development, lead prioritization studies are conducted in living organisms (in vivo) and in cells in the test tube (in vitro) to compare various lead compounds and how they are metabolized and affect the body.²

In the future, this process will have to be more efficient and quicker to deliver a higher percentage of pipeline molecules to the market. It is anticipated that screening millions of compounds per target will become a gold standard for the major pharmaceutical companies³.

Aetion Technologies offers HTS in silico, or via computer simulation, to automatically test millions of compound collections for potentially active compounds ('hits') in order to allow further development of compounds for pre-clinical testing ('leads'). The survivors of in silico screening undergo in vitro screening or specialized testing, and the handful of survivors of in vitro screening undergoes in vivo testing.

Why is *in silico* screening needed?

Traditionally, drugs were discovered by testing compounds synthesized in time-consuming multi-step processes against a battery of *in vivo* biological screens. Promising compounds were then further studied in development, where their pharmacokinetic properties, metabolism and potential toxicity were investigated. Adverse findings were often made at this stage, with the result that the project would be halted or restarted to find another clinical candidate — an unacceptable burden on the research and development budget of any pharmaceutical company.⁴

Today, this paradigm has been re-worked in several ways. The testing of drug metabolism, pharmacokinetics and toxicity is today done much earlier; that is, before a decision is taken to evaluate a compound in the clinic. *In silico* screening makes it possible to synthesize large series of closely related libraries of chemicals using the same chemical reaction and appropriate reagents. Such libraries are then run through the HTS to find hits around which further, more focused, series are designed and synthesized in a next round. As the capacity for biological screening and chemical synthesis have dramatically increased, so has the demands for large quantities of early information on absorption, distribution, metabolism, excretion (ADME) and toxicity data (together called ADMET data).⁵

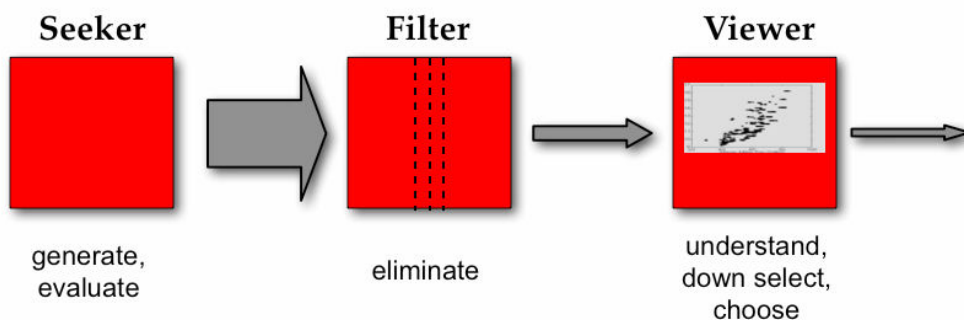
In silico screening serves two key aims — first, at the design stage of new compounds and compound libraries so as to reduce the risk of late-stage attrition; and second, to optimize the screening and testing by looking at only the most promising compounds.⁶

A Solution for HTS in Silico

Aetion holds exclusive licenses to basic patents for a potent family of advanced methods for multiple-criteria decision making that are especially well suited to screening vast numbers of candidate molecules and reducing the set to a relatively few.

Aetion's offering is broken down into three functional components, the Seeker, the Filter, and the Viewer, as shown in the following figure. The "Seeker" generates alternatives by rule-governed composition of components and evaluates them according to multiple criteria. Then, the "Filter" removes the dominated alternatives, and the efficient frontier is viewed interactively in the "Viewer" by means of cross-linked diagrams wherein the

same alternatives are identically colored in each diagram, enabling the comparison of alternatives from multiple perspectives.⁷



The Seeker – The Seeker generates alternatives and evaluates them according to multiple criteria using simulations and/or other computational methods. The seeker generates a large number of candidate small molecules, reaching toward the best molecules by a process of selective breeding. Multi-criterial analysis, spread over several computers by the filter, allows all these molecules to be rapidly evaluated for their potential to interact with various proteins. The number of molecules evaluated is limited mainly by the number of computers available, the time available, and the time it takes to evaluate. If needed, the Seeker can use simulated evolution to generate alternatives by mutation and breeding. By considering millions of alternatives through this simulated evolution, users can effectively search among *billions of alternatives*.

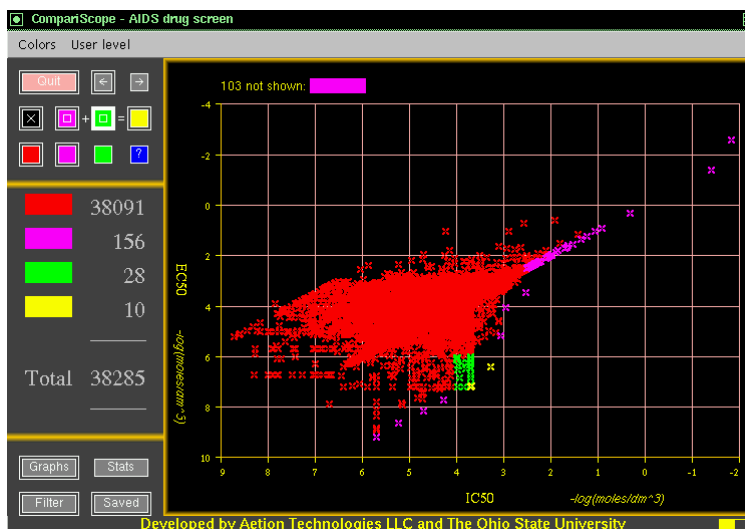
The filter and the viewer allow the user to explore and to trade off the predicted behaviors of the resulting inhibitors in selecting a subset that are worthy of further analysis.

Filter - The total number of alternatives might be quite large, so it is desirable to use the computer to select a relatively smaller number of alternatives that are worth further examination. The Filter can employ a method based on a dominance rule so that filtering is lossless, i.e., so that there are guarantees that no good alternatives are excluded. In other words, an alternative *A dominates* another alternative *B* if *A* is superior or equal to *B* with respect to every criterion of evaluation and distinctly superior with respect to at least one criterion. The alternatives generated by the seeker are systematically

compared, and any candidate that is dominated by any other candidate is removed. The survivors comprise the *Pareto subset*. No member of the Pareto subset dominates any other element in the subset. Thus, deciding among alternatives in the Pareto subset is always a matter of tradeoffs.

Viewer – The Viewer enables decision-makers to perform interactive visual trade-off analyses of the elements of the Pareto set and to narrow to a subset for further exploration. The Viewer is used to explore the surviving design alternatives, and to visualize the trade-offs between pairs of evaluation criteria. The Viewer enables the user to visualize performance characteristics and other attributes of sets of alternatives using various types of interlinked displays, including *trade-off diagrams*, which are two-dimensional, scatter plots with points that represent alternatives and axes that represent criteria. The user can graphically select a subset of the alternatives and explore their ranges of values in multiple dimensions. In this way, the users gain an understanding of the decision space and can act to successively narrow the focus to reduce the number of alternatives under consideration. The user is able to zoom in on subsets with desirable tradeoff characteristics, and reduce the number of alternatives toward making a final selection. The Viewer is graphical, mouse driven, and browser-embeddable so it can be easily deployed over the Internet or over an intranet.

An example of a trade-off diagram below demonstrates how one might see the results when screening potential AIDS drugs for good candidates. The results in green show the drug candidates with most promise to researchers.



Advantages of the SFV Solution

The SFV solution is used to enrich libraries with molecules that have preferred properties or, equally important, to eliminate compounds that have characteristics that are clearly incompatible with the discovery requirements.

Some of the advantages of *in silico* methods in general are the reduction in the number of molecules made and tested through database searching to find inhibitors or substrates, increased speed of experiments through reliable prediction of most pharmaceutical properties from molecule structure alone, and ultimately reductions in animal and reagent use. Specifically, the Aetion SFV solution:

- Empowers decision makers to explore many more possibilities - orders of magnitude of more possibilities - than would otherwise be possible
- Discovers a greatly reduced number of good alternatives within massive, even infinite, search spaces.
- Allows the user to continue to filter the survivors with comparative tradeoff studies.
- Guarantees losslessness.
- Gives people the ability to quickly see exactly what the range of options is and how different factors interact. This allows alternatives to be chosen in an informed, natural way.

Case Study

Aetion's current project, in partnership with Columbus Children's Research Institute, entitled "Fast Docking with Fewer Degrees of Freedom, Combined with Intelligent Optimization of Small Molecules" is expected to demonstrate the effectiveness of a relatively rapid way of testing small molecules for activity against a target protein in simulation.

Background: Molecular docking represents one of the growing applications wherein molecular modeling techniques are used to predict how any macromolecules (typically a protein) interact with other molecules (may be other proteins, nucleic acids, or small drug-like molecules). The ability of a protein to interact with small molecules governs a significant part of the protein's dynamics, which may enhance or inhibit its biological function. This plays an important role in the rational design of drugs.

Docking is most pertinent to the field of drug design—most drugs are small molecules, and using a computational approach allows researchers to quickly screen large databases of potential drugs against protein targets. Traditional discovery of drug candidates occurs by chance or through painstaking work in the lab. For example, virtual screening and related combinatorial chemistry techniques are particularly important in searching for new antibiotics as strains of resistant bacteria increasingly appear due to overuse of antibiotics.⁸

Aetion's Approach to Molecule Docking: Aetion uses a revolutionary FlatWorld approach to molecule docking. The electrostatic surface of the protein is unrolled onto a two-dimensional map, with parts being replicated at various parts of the map depending on the protein's surface features (clefts, etc.). Because a small molecule only interacts with a small part of the protein surface at a time, via weak short-range interactions, this two-dimensional view of the protein surface is still tolerably accurate, and reduces the degrees of freedom such that interaction with active sites can be found much more rapidly. The company's two-dimensional approach allows screening of molecules at significantly higher rates than traditional docking software, and can be used as an initial screen where the better molecules are passed to a fuller three-dimensional in silico treatment that considers approach trajectory.

Using classical computational chemistry techniques, the time required for docking small molecules to large molecules (i.e. repressor molecules to enzymes) is dominated by electrostatic force calculations. This computation is expensive due to both the unavoidable mathematical complexity of the calculation, and the rotational and translational degrees of freedom that must be afforded the system to produce a realistic simulation of the small molecule's trajectory as it approaches, and ultimately docks with the large molecule. These factors conspire to produce a system where many time-intensive instructions must be executed at each time point of a docking simulation, and where the expensive results can neither be pre-computed, nor cached for re-use, due to the relative movement of the molecules between time points.

Despite the necessity for numerous, expensive calculations for physically correct simulations of docking, Aetion proposes that computational screens will accept or reject potential binding candidates from a library with fewer time-consuming calculations. Classical computational techniques simulate a large number of trajectories for each small-molecule candidate, each starting from a randomized location, and randomized translational and rotational velocities. Candidate docks are predicted for small molecules that consistently gravitate to one particular region of the target molecule, over a significant fraction of the test runs. Regardless of trajectory behaviors, however, the success of a dock between molecules is ultimately determined by the local electrostatic field presented by each molecule, the size and shape of the small molecule, and the local topography of the large molecule at the binding site. A library of small-molecule candidates can therefore be screened for potential binding activity, without the expense of calculating trajectories. To be accepted for further consideration, a small molecule must fit – in a "puzzle piece" context (both physically and electrostatically) on the target molecule. Since only a small subset of possible candidate ligands will have appropriate characteristics for docking, eliminating the ones that do not, and thereby eliminating the expense of trajectory calculations for them, affords a considerable speedup in candidate ligand screening.

Using an extension of the FlatWorld surface mapping, we produce topologically planar descriptions of macromolecular and candidate ligand surfaces. These descriptions capture local topography and electrostatic configuration as projected on a plane, for all

aspects presented by the molecules simultaneously. Significantly, the descriptions are two-dimensional in character and stored in memory. This allows detection of small-molecule motifs that fit a macromolecular surface by examination of the macromolecular FlatWorld map for regions that correspond (conversely) to the map of the small molecule's surface. Since the trajectory that leads to a dock is not of concern in this model, the relative motion of the macromolecule and potential ligand need not be considered. This reduces the setup degrees of freedom to only two-dimensional translations, and one dimensional rotations. The docking procedure itself is likewise reduced to only immediate goodness of fit in the initial position and orientation. In addition to the significant speed improvements afforded by elimination of trajectory calculations in screening candidate docks, encoding the macromolecular surface in an appropriate spatial database can appreciably improve the speed of the two-dimensional translational search.

FlatWorld surface maps are constructed by projecting the surface properties of a molecule onto a close-fitting regular solid, and storage of the resultant projection in a two-dimensional map representing that solid. For rapid screening purposes, the stored information includes electrostatic field, hydrophathy, and the distance between the regular solid and the actual molecular surface. This projection results in a planar height-field representation of the surface as a discretized grid, with relevant binding information stored in each grid cell. The molecule is effectively "skinned" and the unrolled skin is used as a surface map. A simple RNA molecule can be skinned effectively by projection onto a cylinder. Complex proteins require projection onto other regular solids, such as spheres. In the case of multi-domain proteins that cannot be conveniently fit with a single regular solid, regions of the molecule can be projected individually, and a composite planar map constructed from the union of the separate segment maps. To reduce the effect of distortions in the projection and the possibility of congruent features on the molecular surface being divided by skinning seams, some regions of the surface can be represented in multiple locations on the two-dimensional map. This increases the effective surface area of the macromolecule under consideration, but should only be necessary when the binding/activity properties of the macromolecule are completely unknown, and when the binding/catalytic sites cannot be mapped to optimal regions of the simple projection. In no case does duplicate projection of features increase the time-cost of the calculation by more than a linear factor.

Take Aways

- The costs and lead times needed to develop new drugs are driving companies to seek more efficient methods of testing compounds. The major goal of these tests is to identify the problematic compounds early in the drug discovery process, which increases the success rate of the compounds that do make it into the later, more expensive stages of the process.
- The entire process of drug discovery is critically dependent upon the ability to reduce the screening efforts to identify better lead compounds more quickly and efficiently. Aetion uniquely enables high-throughput screening *in silico*, or via computer simulation, to automatically test millions of compound collections for potentially active compounds ('hits') in order to allow further development of compounds for pre-clinical testing ('leads'). The survivors of *in silico* screening undergo *in vitro* screening, and the handful of survivors of *in vitro* screening undergoes *in vivo* testing.
- Aetion holds exclusive licenses to basic patents for a potent family of advanced methods for multiple-criteria decision making that are especially well suited to screening vast numbers of candidate molecules and reducing the set to a relatively few.
- Advantages of *in silico* methods in general are the reduction in the number of molecules made and tested through database searching to find inhibitors or substrates, increased speed of experiments through reliable prediction of most pharmaceutical properties from molecule structure alone, and ultimately reductions in animal and reagent use.

Aetion Contact Information

Info@Aetion.com

(614) 340-1835

(614) 487-3704 (fax)

Aetion Technologies LLC

1275 Kinnear Road

Columbus, OH 43212-1155

¹ Emiliangelo Ratti† and David Trist, “Continuing evolution of the drug discovery process in the pharmaceutical industry,” *Pure Appl. Chem.*, Vol. 73, No. 1, pp. 67–75, 2001.

² Pharmaceutical Product Development Web site, www.ppd.com, updated March 2007.

³ Jürgen Bajorath, “Integration of Virtual and High Throughput Screening,” *Department of Computer-Aided Drug Discovery, Albany Molecular Research, Inc. (AMRI), Bothell Research Center (AMRI-BRC), University of Washington, Nov. 2002.*

⁴ Han van de Waterbeemd* and Eric Gifford, “ADMET in Silico Modeling: Towards Prediction Paradise?” Nature Publishing Group, March 2003.

⁵ Ibid.

⁶ Ibid.

⁷ M. Carroll, J. Josephson, J. Russell, Tradeoffs on the Efficient Frontier of Network Disruption Attacks, This research was prepared under sponsorship from the U.S. Naval Surface Warfare Center under contract N00178-02-C-3063, through participation in the Advanced Decision Architectures Collaborative Technology Alliance sponsored by the U.S. Army Research Laboratory under Cooperative Agreement DAAD19-01-2-0009, and by federal flow-thru by the Department of Defense under contract FA8652-03-3-0005 (as a subcontract from Wright State University and Wright Brothers Institute).

⁸ Referenced from Wikipedia Encyclopedia at http://en.wikipedia.org/wiki/Molecular_docking.