

Benchware HTS DataMiner

Support for HTS Data Analysis



Benchware® HTS DataMiner provides researchers with a high-capacity chemical spreadsheet environment and specialized modules to address specific research requirements. Chemical structures and chemical/biological data can be imported, joined from multiple sources, manipulated, and analyzed in an intuitive spreadsheet environment.

The base functionality of Benchware HTS DataMiner can be extended through two additional modules:

HTS Module - complete tools for the analysis of HTS and vHTS data through data mining algorithms (SAR Rules), and visualization and exploration of entire SAR landscapes (SAR Maps) including definition, management, and prioritization of chemical series for further research

HQSAR Module - rapid, automated QSAR technique for creation of predictive models and prioritization of chemical series from HTS data analysis

Base Functionality

The base functionality of Benchware HTS DataMiner provides the chemical spreadsheet foundation for further modules of the Benchware HTS DataMiner system. The high-capacity, chemically aware spreadsheet processes molecular structures, enabling scientists to work interactively with large datasets such as those derived from High-Throughput Screening (HTS) and virtual High-Throughput Screening (vHTS) experiments. Chemical series creation, manipulation, and analysis tools allow users to deal simply with subsets of molecules within a broader chemical and biological context.

A wide variety of molecular descriptors can be computed and relationships within chemical-biological datasets investigated. Substructure search capabilities combined with maximal common substructure determination allow users to locate compounds containing a particular scaffold or undesirable molecular fragments. Researchers can view, compute properties, and analyze datasets of molecules as either whole molecules or as scaffolds and R-groups in an easily generated SAR table.

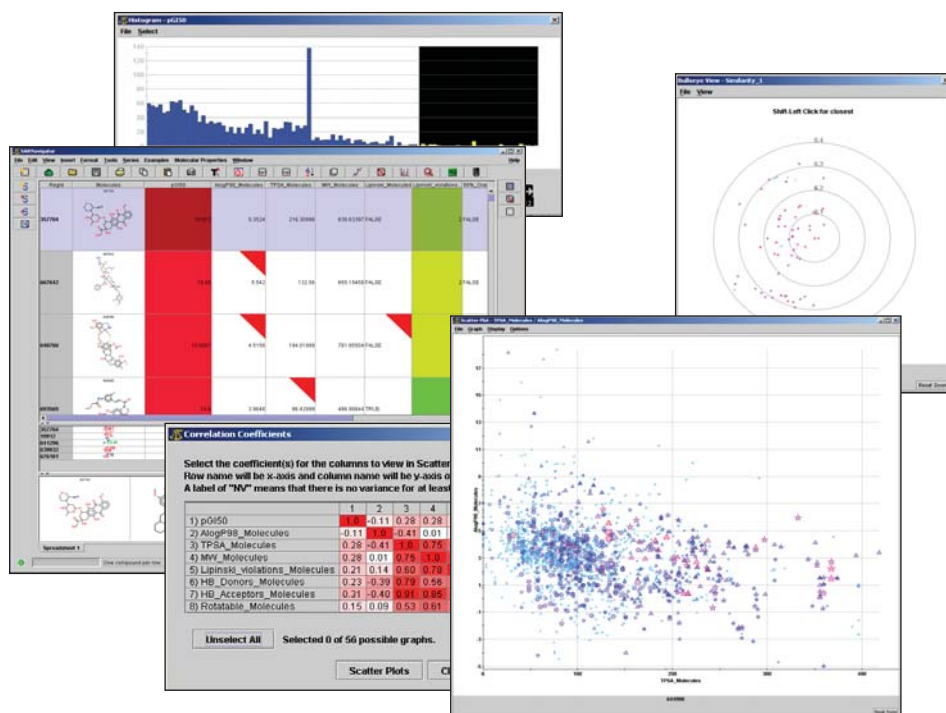
Base functionality can be utilized for decision support in medicinal chemistry, enabling researchers to understand relationships between biological activity and chemical structures. The base

functionality of Benchware HTS DataMiner can be extended through the addition of other modules or through customization.

HTS Module

The HTS module of Benchware HTS DataMiner provides the functionality and workflow required to analyze the large

results datasets produced from HTS and vHTS experiments. The HTS functionality has been designed and developed in conjunction with HTS data analysis groups at major pharmaceutical companies to address bottlenecks and challenges in HTS analysis procedures.



Data Exploration Capabilities

Datasets can be explored in Benchware HTS DataMiner using spreadsheet, scatter plots, histograms, correlation coefficients, and advanced visualization such as similarity maps and bull's-eye plots for high-throughput screening data analysis.

Benefits

- Rapid identification of structurally similar, biologically active molecules from HTS data
- Comprehensive set of tools for the management, manipulation, and prioritization of chemical series for further research
- Prioritization of chemical series and prediction of biological activities based on statistically robust, rapidly calculable QSAR models

HQSAR Module

The HQSAR module of Benchware HTS DataMiner adds the ability to generate quantitative structure-activity relationship models to Benchware HTS DataMiner. Users can predict molecular properties or biological activity and highlight the molecular fragments important for that activity. Users can create and use their own QSAR models or share models with other Benchware HTS DataMiner users. The ability to create an explanatory HQSAR model on series of compounds derived from HTS screening results can be helpful in the prioritization of those series.

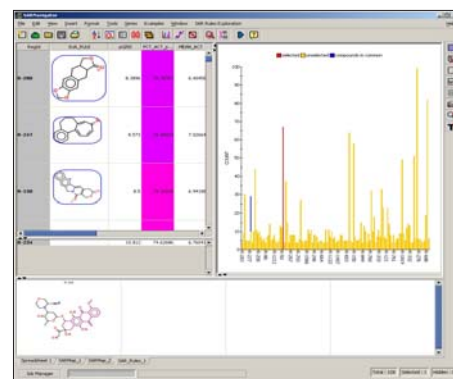
SAR Maps

SAR Maps allow researchers to view and explore entire SAR landscapes from large-scale HTS, uHTS, and vHTS experiments incorporating both hit and similar non-hit compounds. SAR landscapes are viewed as 2D plots where compounds or clusters of compounds are plotted such that similar structures are in close proximity. Clustering, drill-down, and projection

tools allow researchers to explore all areas of chemical information within a screening deck where a biological signal has been detected. Visualization of screening data through SAR Maps enables researchers to see the interrelationships of different chemical classes and allows simple, interactive identification and definition of chemical series for further research.

SAR Rules

SAR Rules technology allows researchers to rapidly mine HTS and vHTS data for groups of structurally related compounds that are highly enriched in hits. SAR Rules depict groups of connected molecular fragments that are found prevalently in screening hits and the compounds that contained those structural signatures. SAR Rules are generated through the Structural Units Analysis (SUA) algorithm and provide an automated method to determine groups of compounds that are the basis for chemical series. Users can control the number of compounds that a rule must contain as well as the nature and number



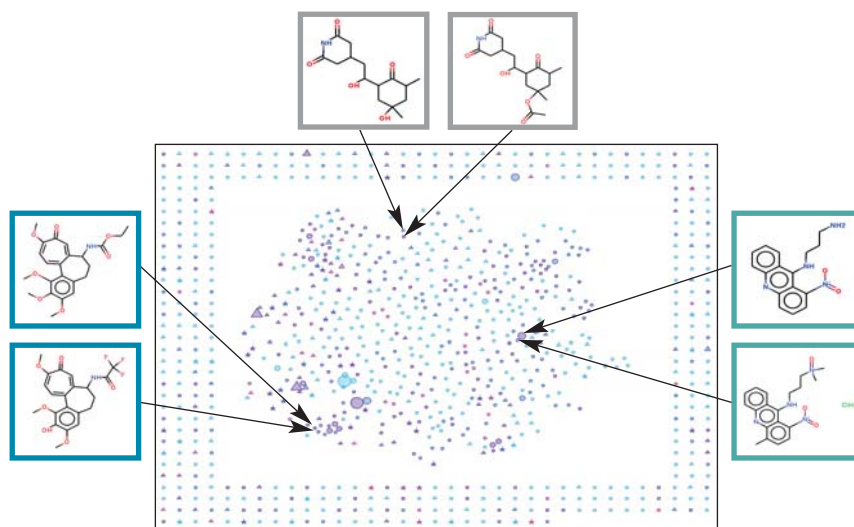
SAR Rules

SAR Rules data mining application within SARNavigator automatically determines scaffolds and substructures that elicit high biological activity. The sets of compounds found by SAR Rules provide an excellent starting point for the definition of chemical series for further research. SAR Rules take into account both active and inactive compounds and can be used as a navigational aid with SAR Maps.

of structural units that are examined. SAR Rules can be employed alone to determine chemical series or as a navigational aid to SAR Map data projections. Use of SAR Rules can also be used to analyze smaller groups of compounds such as project data to determine structural factors that lead to improved chemical and biological properties.

Series management tools in Benchware HTS DataMiner allow interesting chemical series to be captured, manipulated, and expanded once discovered. Defined chemical series can be prioritized for further research using molecular property profiles, SAR, or QSAR information. The combination of SAR Maps and SAR Rules provides Benchware HTS DataMiner with the most complete set of HTS analysis functions available in the industry. Using Benchware HTS DataMiner for HTS data analysis ensures the best possible decision about which compounds tested in HTS are worthy of secondary screening and further chemical exploration.

The results of HTS data analysis can be shared with other Benchware HTS DataMiner users allowing collaborative



SAR Landscape View

The HTS module employs SAR landscape views to allow users to visualize the entire biological signal found in the high-throughput screening deck. SAR landscape views are similarity plots where structurally similar compounds or clusters are plotted close in space. Biological activity, ADME, toxicology, selectivity, and other available data can be mapped onto plotted points guiding users to groups of chemical structures which possess the properties required to warrant further investigation. The ability to generate property statistics for clusters for landscape views, which include both hit and non-hit compounds, allows users to readily locate likely false negatives, false positives, and chemical series with highly sensitive SAR.

Advantages

- Chemically aware spreadsheet environment with modular access to advanced tools ensures a consistent software research environment through HTS data analysis and beyond.
- Straightforward access to SAR views of data and a rapid QSAR technique ensure that datasets can be interactively examined for determinants of biological activity.

analysis. Chemical series defined in this way become initial SAR datasets to guide chemical synthesis decisions after secondary screening has confirmed biological activity.

Base Module Features

- High-capacity chemical spreadsheet
- Structure property calculation
- Chemical series management tools
- Substructure searching
- Creation of SAR tables through deconvolution of molecular structure into scaffold and R-groups
- Graphing with property mapping and data analysis tools

HTS Module Features

- SAR Maps - entire SAR landscape view of screening data through novel PCA/NLM procedure; macro recording and replay capabilities
- SAR Rules - rapid, flexible, automated identification of structurally related chemical series
- Interactive exploration tools for the analysis of large HTS datasets for use with both SAR Maps and SAR Rules
- Clustering, cluster analysis, and drill-down capabilities

Chemical series expansion capabilities to locate unscreened compounds suitable for inclusion in chemical series

HQSAR Module Features

- Rapid determination of predictive QSAR models
- Minimal requirement for user input
- Color coding of molecular structures to highlight important fragments

Validation

The base and HTS functionalities have been developed in conjunction with HTS data analysis groups at major pharmaceutical companies where they

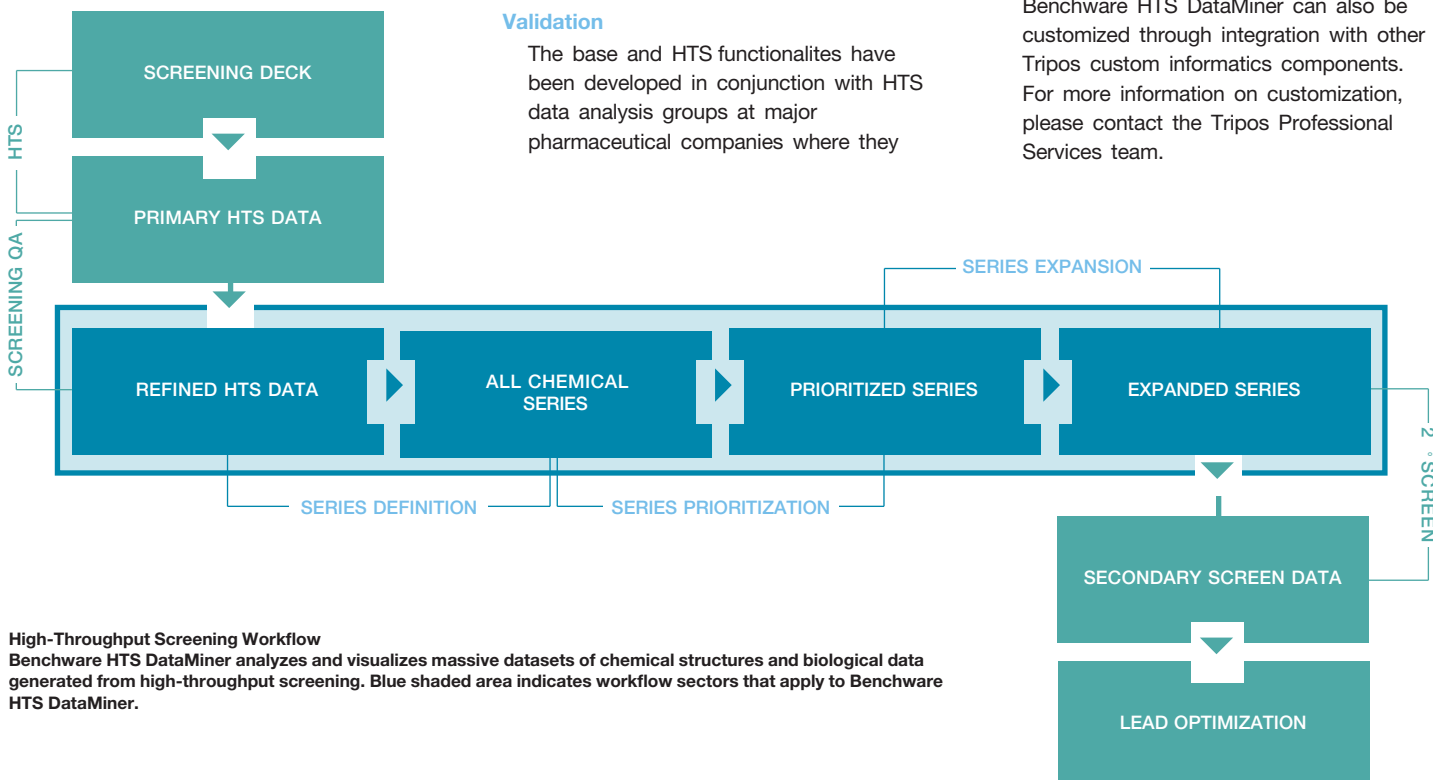
have been successfully employed to identify high-quality chemical series from HTS experiments for secondary screening and further chemical exploration. The HQSAR module utilizes the same underlying base code as the HQSAR module in the SYBYL® environment which has been extensively tested, validated, used, and published.

Customization

Benchware HTS DataMiner can be customized by Tripos or the user to address specific research goals and requirements. Two forms of customization are available:

- Jython scripting - for minor modifications, such as the computation of a new molecular property or column type, or implementation of new functionality
- Java services - for major customization, such as introduction of new data analysis modules

Benchware HTS DataMiner can also be customized through integration with other Tripos custom informatics components. For more information on customization, please contact the Tripos Professional Services team.



High-Throughput Screening Workflow

Benchware HTS DataMiner analyzes and visualizes massive datasets of chemical structures and biological data generated from high-throughput screening. Blue shaded area indicates workflow sectors that apply to Benchware HTS DataMiner.

Advantages *(continued)*

- HTS data analysis tools and processes allow the definition and prioritization of chemical series from views of the entire SAR landscape of an HTS screen including information from both hit and non-hit compounds.
- Customization capabilities allow ready incorporation of new science and workflow enhancements.

Hardware Requirements

Minimum system requirements are designed to allow complete functionality on collections of tens of thousands of compounds, and most functionality on collections of hundreds of thousands of compounds. Additional memory and disk space will provide maximum performance and the ability to process larger compound collections.*

Operating System

- Recommended: Windows® 2000 Service Pack 3 or Windows® XP
- Minimum: Windows® 2000 Service Pack 3 or Windows® XP

RAM

- Recommended: 512 MB
- Minimum: 384 MB

CPU Speed

- Recommended: 2.0 GHz
- Minimum: 1.0 GHz

Free Disk Space

- Recommended: 2 GB
- Minimum: 500 MB

*Using Benchware HTS DataMiner with less than minimum system configuration will result in degraded software speed and performance.

Maximum Number of Compounds

- Recommended: Allows the use of all features on up to 15,000 compounds and Base module functionality on 200,000 - 300,000 compounds
- Minimum: Allows the use of all features on up to 2,000 compounds and the use of Base module features on up to 10,000 compounds

Benchware HTS DataMiner Analysis Process:

Datasets can be analyzed in a variety of ways using Benchware HTS DataMiner. For high-throughput screening data analysis a process has been developed which allows researchers to explore the entire chemical information content of a screening deck where a biological signal has been found. By examining all compounds found to hit in a particular screen as well as similar compounds which were not hits, researchers can see all pertinent screening information with the minimum of computational overhead.

