



Technical Manual

Chalice™ Analyzer
Overview

Contents

1	Analytical Methods	2
1.1	Measuring biological activity	2
1.2	Identifying replicate dose series and dose matrices	2
1.3	Merging replicate dose series	3
1.4	Merging replicate dose matrices	3
1.5	Sigmoidal response models for Dose Response data	3
1.6	Combination effect reference models	4
1.7	Scoring synergistic interaction	4
1.8	Displaying potency shifting using isobolograms	5
1.9	Scoring potency shifting using combination index	5
1.10	References	5
2	Replicate Data	7
2.1	Identifying Replicate Observation Blocks	7
2.2	Determining Consensus Observation Block Sampling	7
2.3	Producing Consensus Observation Block Output	7
3	Combination Interaction Models	8
3.1	Built-In Models	8
4	Combination Interaction Analytical Output	10
5	Chalice Viewer Overview	11
5.1	Querying Data	11
6	Analysis Options	12
6.1	Data Aggregation Method	12
6.2	Replicate Data	12
6.3	Consensus Method	12
6.4	Consensus Aggregation Ext	12
6.5	Test Substance Orientation	13
6.6	Analytical Output	13
7	Specifying Data Characteristics	14
7.1	Data source Query	14
8	Using the Matrix view	17
9	Filtering	18
10	Importing Data	21
11	Exporting and Reporting	22
11.1	Exporting Tabular Datasets	22
11.2	Creating PDF Reports	22

1 Analytical Methods

1.1 Measuring biological activity

The measure of effect in our screening is usually the inhibition of cellular response relative to the untreated level (vehicle alone). For untreated vehicle and treated levels V and T , we calculate a fractional inhibition $I = 1 - T/V$. The inhibition ranges from 0% at the untreated level to 100% when $T = 0$. Inhibition levels are negative for agents that actually increase levels. Other effect measures, such as an activity ratio $r = T/V$ may be more appropriate for some assays. When activity ratios (e.g., fold increase over stimulated control) are being used, the effect can be measured using an induction $I = \ln(T/V)$. With this definition, all effect expressions are the same as for inhibition.

For oncology screens with time zero reference growth levels, we use a "Growth Inhibition" (GI) based on the measures proposed for the NCI60 screens [Shoemaker'06]. We define $GI = 1 - [T < V ? (T - V_0)/V_0 : (T - V_0)/(V - V_0)]$, using the conditional syntax [condition ? result.if.true : result.if.false], and where V_0 is the vehicle-treated level at time zero. The descriptive levels for this measure are "silent" (GI=0), "active" (GI=0.5), "static" (GI=1), "toxic" (GI=1.5), and "lethal" (GI=2). As defined for the NCI60 screens [Shoemaker'06], the GI50, TGI, and LC50 are the concentrations that achieve active, static, and toxic levels respectively. Some treatments which actually increase growth will have $T > V$, with a corresponding "agonizing" reference level (GI= 0.5).

Error estimates for these measures result from comparing each treated level T to the median untreated level $V \pm \sigma V$, determined for each plate by finding the median level among the untreated control wells arranged across the plate. For example, applying error propagation $\sigma^2(f(x_i)) = \sigma^2(x_i) (df/dx_i)^2$ to the expression for fractional inhibition I , the estimated standard error is $\sigma I \sim (\sigma V/V) \sqrt{1-I}$. The error estimates are further increased to account for variations between replicate combination blocks as well as a minimum assumed fractional uncertainty of $\sigma_{min} \sim 3\%$. Thus for inhibition, the standard error estimate becomes $\sigma I \sim \sqrt{(\sigma V/V)^2 (1-I) + \sigma_{rep}^2 + \sigma_{min}^2}$.

We use medians rather than averages to reduce the effect of occasional outliers on the consensus. Note that while medians are robust to outliers, they are more sensitive to statistical noise, yielding ~30% larger deviations for normally-distributed data. Standard deviations are estimated from the median absolute deviation (MAD), where for a normal distribution, the sample deviation $\sigma_{dat} \sim 1.5 \text{ MAD}$ [Filliben'05]. The error for the median is then $\sigma_{med} \sim \sigma_{dat}/\sqrt{N}$, given N data values.

1.2 Identifying replicate dose series and dose matrices

Our cHTS screening produces phenotypic measurements for single agents or fixed ratio combinations as a series of concentrations (doses) or for pairwise combinations as dose matrices which contain all mixtures of two single agents each sampled at a series of concentrations, including zero [Borisy '03; Lehár '09]. The information from replicate doses can be summarized by generating a consensus block of data (either dose series or dose matrix) out of the individual blocks from the replicates.

In the Chalice Analyzer, single agent dose series are identified by specifying the desired chemical's name or C number, along with other search criteria specifying the assay details or other properties of the data set. The Analyzer will then collect all examples of the requested chemical that were tested as single agents and will either present them as separate blocks or merge them according to instructions from the user.

Combination data are identified by specifying pairs of chemical names or C numbers, along with other search criteria. Each combination dose matrix contains internal copies of the single agent curves which are used as the reference for combination effects. The properties of these single agent dose series are used to guide the merging process between replicate data blocks.

1.3 Merging replicate dose series

The single agent curve data are recorded at different doses (chemical concentrations) and arranged into a series with increasing dose. To merge dose series from replicate data sets, these series are resampled to generate a consensus dose series which will be used for calculating the median dose response values. Two algorithms are currently employed, basic and periodic (Fourier) resampling. For both algorithms, the data from all individual dose series are collected into a single data set and sorted by dose.

In basic resampling, a sliding window of a specified width in log-concentration (logarithm of the tested concentration) space is moved through the tested doses. Starting with the lowest nonzero tested concentration in the series, all data points within the specified window are merged, and the lowest tested data point outside the window serves as the starting dose for the next window. The window is advanced iteratively through the dose series until all data points have been merged, and for each such window the average log-concentration within the window is defined as the consensus dose. These consensus doses are then reconverted to concentrations through an inverse logarithm operation.

In periodic resampling, a Fourier algorithm is employed to define a regular series of doses with a fixed dilution factor covering the tested doses, phased to best align the consensus doses with the original sampling. For a set of N doses x covering M distinct log-concentrations (logarithm of the tested concentration), with δx and Δx as the smallest and largest intervals between the nonzero sampled doses, a frequency interval δf is set to be $1/(2\Delta x)$. Then for all frequencies f covering the range from $2M\delta f$ to $1/(\delta x)$ in steps of δf , the Fourier coefficients $A(f) = \sum_i \cos(2\pi f x_i)$ and $B(f) = \sum_i \sin(2\pi f x_i)$ are summed across all N doses x_i , and the peak frequency f_{peak} is identified with the largest Fourier power A^2+B^2 . The period $P_{\text{peak}} = 1/f_{\text{peak}}$ and optimal phase $\phi_{\text{peak}} = P_{\text{peak}} \text{atan2}(B_{\text{peak}}, A_{\text{peak}})/2\pi$ are calculated, and optimal doses $\xi_k = \phi_{\text{peak}} + kP_{\text{peak}}$ are chosen for all integers k yielding ξ_k within the tested dose range. All data within a dosing window of $\pm P_{\text{peak}}/2$ from each ξ_k are then merged and assigned a consensus dose ξ_k , which are then reconverted to concentration through an inverse logarithm operation. Dosing windows in the periodic series that contain no data are assigned null response values.

After resampling, all data points within each consensus window are merged by taking the median of the corresponding response values, and assigning the consensus dose for that window. Error estimates are calculated for each consensus response value following the error propagation methods described in "Measuring biological activity" above.

1.4 Merging replicate dose matrices

Pairwise combinations are organized into dose matrices which contain all mixtures of two single agents each sampled at a series of concentrations, including zero [Borisy '03; Lehár '09]. When many such dose matrix blocks are collected for merging, the consensus block needs to account for the possibility of different single agent sampling between data blocks.

The process for merging dose matrices is as follows: (1) first, the single agents are aligned (put on the same vertical or horizontal axis) across all of the data blocks; (2) next, each single agent's dose series is merged according to the single agent merging algorithms described above, yielding a separate consensus dose series for each single agent; (3) finally, all the combination points within each two-dimensional window defined by each of the single agents' concentration windows are merged and assigned consensus concentration values corresponding to the single agent consensus doses. All data points within a dosing window are merged by taking the median data value with its associated uncertainty as described above in "Measuring biological activity". For those dosing points defined by the single agent sampling windows that contain no combination data, null response values are assigned.

1.5 Sigmoidal response models for Dose Response data

The single agent activity is characterized by fitting a sigmoidal function of the form $I = I_{\text{max}}C\alpha/[C\alpha+EC50\alpha]$. Here, C is the concentration, $EC50$ is the $EC50$, and α is the sigmoidicity. We also use these parameters to calculate an inhibitory concentration IC_{cut} , where the fitted curve crosses a chosen effect level I_{cut} . The fitting is performed using a parameter search algorithm based on cost minimization. The cost function is a chi-squared between the fitted sigmoid and the

measured data points with standard errors, with the last data point having double weight to ensure that compounds with activity only at the highest concentration are represented. Additional penalties are applied to restrict parameter ranges. Using fractional inhibitions, for example, hard limits on I_{max} are enforced by multiplying the cost function by $\Delta 10$, where Δ is $(I_{max} - I_{lim})/0.001$, where I_{lim} is the upper or lower limit allowable for that measure. For EC_{50} , a softer multiplicative penalty of $\Delta 2$ is applied, where Δ is the logarithm of the ratio between the fitted EC_{50} and the largest or smallest nonzero concentration in the data series. The sigmoidicity values are restricted between 1 and 10 using a penalty of $\Delta 2$, where Δ is the distance from α outside the nearest limit. The curve fits were optimized using simplex cost minimization [Nelder&Mead'65; Press '97], with the starting simplex vectors connecting the center of each parameter's range to its upper limit.

1.6 Combination effect reference models

Combination effects can be most readily characterized by comparing each data point's inhibition to that of a combination reference model that was derived from the single agent curves [Greco '95]. Three models are generally used: (1) The highest single agent model $I_{HSA}(CX, CY) = \max(I_X, I_Y)$ is a simple reference model, where CX, Y are the concentrations of the X and Y compound, and I_X, Y are the inhibitions of the single agents at CX, Y ; (2) Bliss independence [Bliss'39] $I_{Bliss}(CX, CY) = I_X + I_Y - I_X I_Y$ represents the statistical expectation for independent competing inhibitors; and (3) Loewe additivity [Loewe'28], where $I_{Loewe}(CX, CY)$ is the inhibition that satisfies $(CX/IC_x) + (CY/IC_y) = 1$, and $IC_{x,y}$ are the inhibitory concentrations at I_{Loewe} for the single agent curves. Loewe additivity is the generally accepted reference for synergy [Greco '95], as it represents the combination response generated if X and Y are the same compound. Both I_{HSA} and I_{Bliss} are easily calculated from I_X, Y , but determining I_{Loewe} requires interpolation and numerical root finding [Berenbaum'85].

1.7 Scoring synergistic interaction

To select promising candidates from large combination screens, it is most useful to design a numerical score that captures the qualities of desirable combinations, which can be used to prioritize the manual selection process. These scores should be designed for each screen to capture the experimental objectives, but there are usually three important considerations: (1) significant synergy over the Loewe additive model; (2) substantial activity where the synergy occurs; and (3) sufficient potency shifting.

Activity over Loewe additivity is most easily calculated using a simple volume score, where $V_{Loewe} = \log f_X \log f_Y \sum (I_{data} - I_{Loewe})$, summed over all non-single agent concentration pairs and where $\log f_X, Y$ are the natural logarithm of the dilution factors used for each single agent. This effectively calculates a volume between the measured and Loewe additive response surfaces, corrected for varying dilution factors. This volume score emphasizes the overall synergistic or antagonistic effect of the combination, thus minimizing the effects of outlying data spikes and identifying combinations with a robust synergy across a wide range of concentrations and at high effect levels. V_{Loewe} is positive for mostly synergistic combinations and negative for antagonism. The uncertainty σV can be calculated based on the measured errors σI and standard error propagation.

We usually use a "Synergy Score" $S = f_{cov} \ln f_X \ln f_Y \sum \max(0, I_{data}) \max(0, I_{data} - I_{Loewe})$, which is a positive-gated, inhibition-weighted volume over Loewe additivity. This provides an additional prioritization favoring combinations whose synergy occurs at high effect levels, ignoring antagonistic portions of the response surface. Here f_X, Y are the dilution factors used for each single agent and the coverage factor f_{cov} accounts for missing data, scaling the score up by the ratio of total/tested combination dose matrix points. S is always positive, and its uncertainty σS can be calculated based on the measured errors σI and standard error propagation. An alternative to the synergy score is the "Hit Score" $H = f_{COV} \log f_X \log f_Y \sum \max(0, I_{data}) \max(0, I_{data} - I_{HSA})$, which refers to the HSA model. The key distinctions between S and H lie in the different underlying models and also in how the single agents are used in the model calculations. In the Chalice Analyzer, the HSA model is calculated directly from the single agent responses at corresponding concentrations, while the Loewe additive model is derived from the sigmoidal fits to the single agent response curves.

To prioritize hits, we generally use distributions of our preferred score (e.g., S) and its error to define an appropriate selection cut-off. For example, combinations with $S > 3\sigma S$ are "individually significant" at ~99% confidence, assuming normal errors. To estimate systematic experimental errors that are not tested by replicate plates, we sometimes use the distribution of synergy scores for any drug-with-itself combinations acquired during the experiment to determine a plausible range for non-detections. Alternatively, the score distribution for the whole experiment could be used to identify outliers at a chosen confidence level.

1.8 Displaying potency shifting using isobolograms

Potency shifting is usually shown using an isobologram [Greco '95] which shows how much less drug is required in combination to achieve a desired effect level, when compared to the single agent doses needed to reach that effect.

The choice of effect level for the isobologram display and combination index calculations is somewhat arbitrary, and can either be manually or automatically selected in the Analyzer. The automatic iso-level selection algorithm finds the observed I_{data} with the largest $I_{data} - I_{Loewe}$, excluding those points with I_{data} exceeding the lesser single agent's I_{max} . This exclusion is applied to ensure that the isobologram reflects the best synergy at levels covered by both single agents. The chosen iso-level is then floored at a 20% inhibition level, beneath which inhibitions often are unreliable, and rounded to the nearest 5%.

Having selected an isobologram level I_{cut} , the isobologram is drawn by identifying the locus of concentrations that correspond to crossing the chosen iso-level. In the Analyzer, this is done by finding the crossing point for each single agent concentration in a dose matrix across the concentrations of the other single agent. Practically, first each vertical concentration C_Y is held fixed while a bisection algorithm [Press '97] is used to identify the horizontal concentration C_X in combination with that vertical dose that gives the chosen effect level in the response surface $Z(C_X, C_Y)$. These concentrations are then connected by linear interpolation to generate the isobologram display. The uncertainty for each crossing point is estimated from the response errors using bisection to find the concentrations where $Z - \sigma Z(C_X, C_Y)$ and $Z + \sigma Z(C_X, C_Y)$ cross I_{cut} .

1.9 Scoring potency shifting using combination index

Potency shifting is scored a combination index [Chou&Talalay'84] CI. For a chosen iso-effect level I_{cut} , $CI = (C_X/ECX)^I + (C_Y/ECY)^I$, where $(C_X/ECX)^I$ for a particular data point is the ratio of the X compound's measured concentration to its effective concentration at the chosen inhibition level. The CI is a rough estimate of how much drug was needed in combination relative to the single agent doses required to achieve the chosen effect level. CI values in the range of 0.5-0.7 are typical for in vitro measurements of current clinical combinations [Greco '95]. The combination index error σCI is calculated using standard error propagation through the CI calculation based on the isobologram errors derived as described above in "Displaying potency shifting using isobolograms".

In the Chalice Analyzer, the best CI is reported from the many combination index values calculated for each I_{cut} crossing concentration. Among all the measured CI values, the one with the largest signal-to-noise level $(1-CI)/\sigma CI$ is reported as the best combination index.

1.10 References

Berenbaum (1985). *The expected effect of a combination of agents: the general solution. J Theor Biol* 114(3): 413-31. #4021503

Bliss (1939). *The toxicity of poisons applied jointly. Ann. Appl. Biol.* 26: 585-615.

Borisy, Elliott, Hurst, Lee, Lehár, Price, Serbedzija, Zimmermann, Foley, Stockwell, Keith (2003). *Systematic discovery of multicomponent therapeutics. PNAS* 100(13): 7977-82. #12799470

Chou, Talalay (1984). Quantitative analysis of dose-effect relationships: the combined effects of multiple drugs or enzyme inhibitors. Adv Enzyme Regul 22: 27-55. #6382953

Filliben. (2005, 2005). Exploratory Data Analysis. Engineering Statistics Handbook Retrieved 10/14, 2007, from <http://www.itl.nist.gov/div898/handbook/eda/section3/eda356.htm>.

Greco, Bravo, Parsons (1995). The search for synergy: a critical review from a response surface perspective. Pharmacol Rev 47(2): 331-85. #7568331

Lehár, Krueger, Avery, Heilbut, Johansen, Price, Rickles, Short, Staunton, Jin, Lee, Zimmermann, Borisy (2009). Synergistic drug combinations tend to improve therapeutically relevant selectivity. Nat Biotechnol 27(7): 659-66. #19581876

Loewe (1928). Die quantitativen Probleme der Pharmakologie. Ergebn. Physiol. 27: 47-187.

Nelder, Mead (1965). A Simplex Method for Function Minimization. The Computer Journal 7(4): 308-313.

Press, Teukolsky, Vetterling, Flannery (1997). Numerical Recipes in C: the Art of Scientific Computing, 2nd Ed. Cambridge, CUP.

Shoemaker (2006). The NCI60 human tumour cell line anticancer drug screen. Nat Rev Cancer 6(10): 813-23. #16990858

2 Replicate Data

Replicate data can be analyzed individually or aggregated to form a consensus representation of the individual replicate results. The top level analysis option controlling the handling of biological replicates is the "Replicate Data" option. There are two values that can be selected for this option:

Individual - Selecting this value will cause replicate data to be processed individually, and each item will appear in the output.

Consensus - Selecting this value will cause replicate data to be merged according to the selected resampling method and selected data aggregation mode, and the consensus item will appear in the output.

2.1 Identifying Replicate Observation Blocks

Individual observation blocks can be aggregated during the analytical processing to form groups of replicates that are subjected to data merging operations. The groupings are established by the following rules for matching characteristics:

1. The individual observation blocks have the same test substance mixtures present in the same degree of dimensionality.
2. The individual observation blocks share background treatment (if present) for additional substances not varying on concentration.
3. The individual observation blocks share the same protocol name, and the values of the established principal protocol parameters are matching.

2.2 Determining Consensus Observation Block Sampling

In the case that the consensus data is requested and replicate data exists for aggregation, the first step in handling each set of replicates is to determine the appropriate concentration sampling for the consensus output. The analysis engine imposes no limitations on the merging of replicate data with different concentration sampling. Replicate data can be merged for all cases of heterogeneous sampling amongst replicates. It is important to understand the different resampling algorithms and how they affect the construction of consensus views, including the case where all of the replicates have identical sampling.

The analysis engine provides three resampling algorithms which are selected using the "Resample Method" analysis option. The following table describes the resampling algorithms. The "Resample Method" analysis option is only available if the value "Consensus" has been selected for the analysis option "Replicate Data".

Name	Purpose	Description
Regular	Produces consensus output with regular log-linear dilution steps.	This algorithm performs a Fourier signal analysis of the replicates to determine the dominant sampling frequency to be used in constructing the consensus output. The selected frequency is gated by the "Minimum Sample (fold)" analysis option, which is only available if "Regular" has been selected for the "Resample Method" analysis option.
Basic	Produces consensus output with irregular, but bracketed dilution steps.	This algorithm arranges all of the distinct concentrations present from all replicates and performs a top down bracketing of concentrations according to the analysis option "Threshold (fold)", which is only available if "Basic" has been selected for the "Resample Method" analysis option.
None	Produces consensus output with irregular unbracketed dilution steps.	This algorithm preserves every distinct concentration present in the replicates, no matter how trivial the fold difference between any set of concentrations is.

2.3 Producing Consensus Observation Block Output

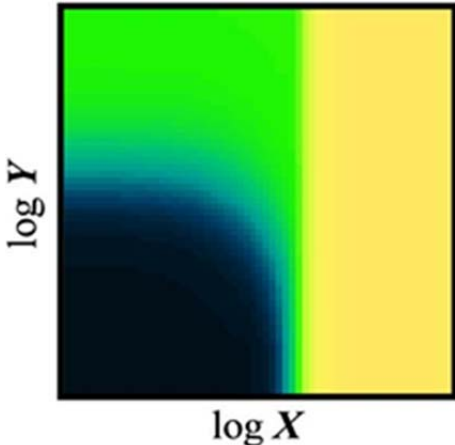
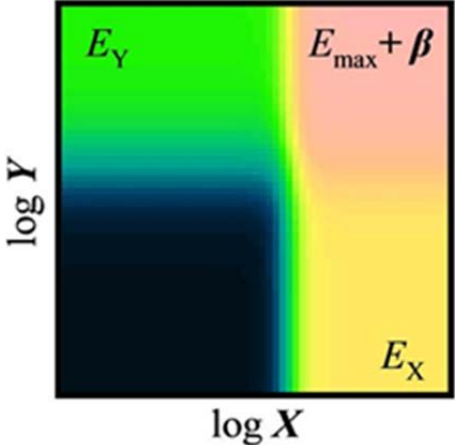
Once the resampling has taken place, and the extents of the consensus data item have been established, each location in the consensus output is compared to the available data in the replicates, and any data points that belong in that location are selected and aggregated as either the mean or the median according to the value of the "Data Aggregation Method" analysis option.

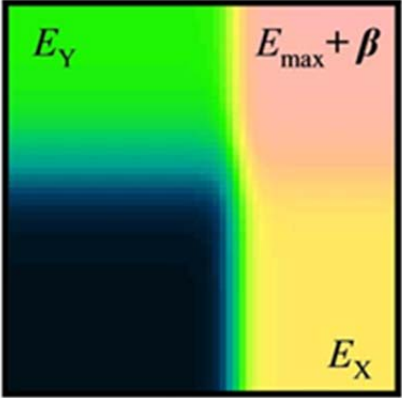
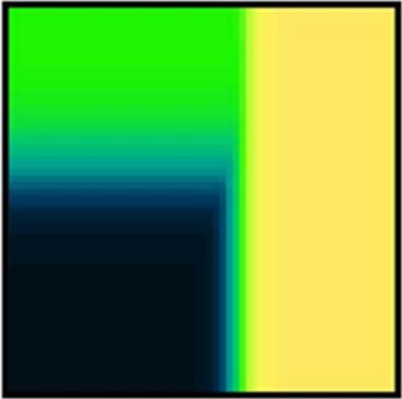
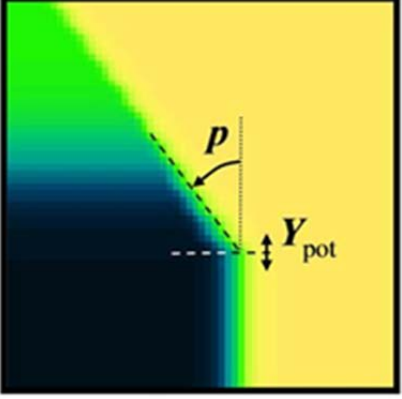
3 Combination Interaction Models

The Analyzer Server has an extensible catalog of combination interaction models. Each defined combination interaction model is instantiated within the context of the experimentally observed data which is the subject of the requested analysis. The following tables summarize the built-in combination interaction models and the analytical output associated with each model.

3.1 Built-In Models

An unlimited number of combination interaction models can be defined using the Chalice Analyzer SDK, the default built-in models are summarized in the following table.

Prototype	Name	Description
	Loewe Additivity (ADD)	<p>The most widely used combination reference is Loewe additivity^{1,2}, or "dose additivity" which describes the trade-off in potency between two agents when both sides of a dose matrix contain the same compound.</p>
	Bliss Independence (BLISS)	<p>The Bliss independence model¹ corresponds to a multiplicative effect in growth measures, and is the preferred reference for synergy in some contexts², especially genetics³.</p>

	<p>Bliss Boosting (BOOST)</p>	<p>Bliss boosting¹, which models boosts in efficacy at high concentrations different from what the single agents can achieve, is adapted the Bliss independence model² that corresponds to a multiplicative effect in growth measures.</p>
	<p>Highest Single Agent (HSA)</p>	<p>The Highest Single Agent (HSA) model, or Gaddum's non-interaction reference (Berenbaum 1989), is based on the intuition that if a combination's effect exceeds those of its constituents, there must be some interaction.</p>
	<p>Potentiation (POT)</p>	<p>The power-law "potentiation" model describes strong shifts in potency¹.</p>

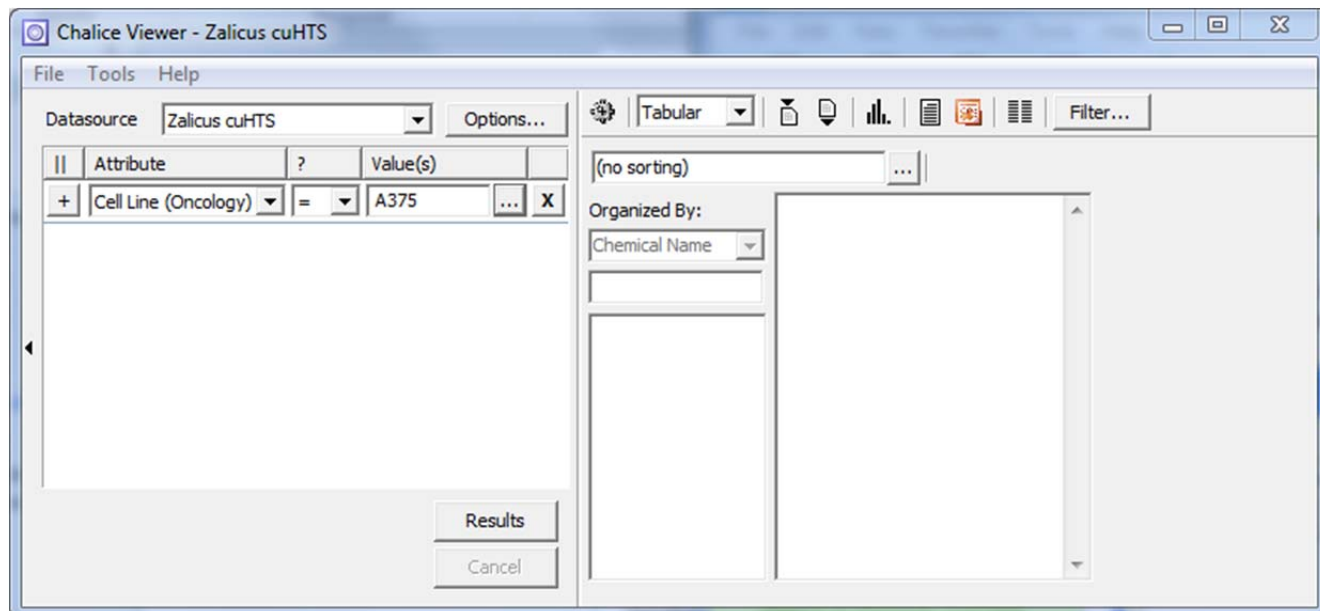
4 Combination Interaction Analytical Output

The Analyzer Server produces a defined set of individual analytical outputs for each defined combination interaction model. The analytical outputs are named according to the model abbreviation and the analytical output name. The following table summarizes the combination model analytical outputs.

Name	Description
Model	<p>The model attribute is a visual representation of the instance of the model for the data item that is the subject of the analysis.</p> <p>The model attribute is used to evaluate the model effect predicted for each dose and ratio of the test substances present in the data item that is the subject of the analysis. Each location in the display matrix is labeled with the numerical effect value for the model, and a color that corresponds to the intensity of the value.</p>
Excess	<p>The excess attribute is a visual representation of the deviation between the observed data and the instance of the model for the data item that is the subject of the analysis.</p> <p>The excess attribute is used to evaluate the magnitude of the observed data that exceeds the model effect predicted for each dose and ratio of the test substances present in the data item that is the subject of the analysis. Each location in the display matrix is labeled with the numerical excess value, and a color that corresponds to the intensity of the excess.</p>
Volume	<p>The volume is the sum of the individual deviations between the observed data and the instance of the model for the data item that is the subject of the analysis.</p> <p>The volume attribute is used to assess the overall magnitude of the combination interaction excess. The volume attribute can be used to compare different data items to assess their relative combination interaction magnitude.</p>
Volume Error	<p>The volume error is the standard deviation for the individual deviations between the observed data and the instance of the model for the data item that is the subject of the analysis.</p> <p>The volume error is used to assess the magnitude of the uncertainty in the given volume value.</p>
Average	<p>The average is the mean of the individual deviations between the observed data and the instance of the model for the data item that is the subject of the analysis.</p> <p>The average attribute is used to assess the breadth of the combination interaction excess compared to the overall volume. The volume attribute can be used to compare different data items to assess their relative combination interaction magnitude.</p>
Average Error	<p>The average error is the sample variance divided by the number of samples for the individual deviations between the observed data and the instance of the model for the data item that is the subject of the analysis.</p>
Parameters	<p>The parameters are the computed parameters, if any, for the instance of the model for the data item that is the subject of the analysis.</p>
Log Chi Squared	<p>The log chi squared is the base 10 logarithm of the chi squared for the comparison of the observed data to the instance of the model.</p>

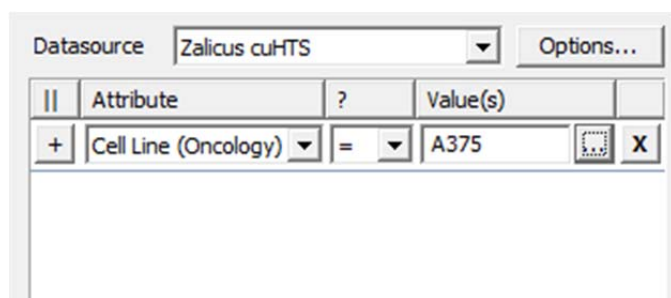
5 Chalice Viewer Overview

The Chalice Viewer is desktop software designed to work with the Chalice Analysis Server software to provide data query, review, manipulation, and reporting functions. The Chalice Viewer stores analyzed data in individual data files which can be used collaboratively.



5.1 Querying Data

Querying of experimental data is completed using data query panel.



Step 1 - Select a data source

Using the data source pick list allows for selection of experimental data from different sources. The default data source will be selected automatically. Consult the Administrator for additional information regarding the available data sources.

Step 2 - Select analytical options

A number of analysis options are available to provide control over how experimental data will be measured, aggregated, and analyzed when the data query is executed. For more information about analytical options see [Analysis Options](#).

Step 3 - Specify data characteristics

Specify which data is to be included in the query data results using available data characteristics. For more information about how to specify data characteristics see [Specifying Data Characteristics](#). Consult the Administrator for additional information regarding available characteristics for a given data source.

Step 4 - Execute the query

Execute the query by actuating the "Results" button. Once actuated, a progress dialog will be displayed to indicate overall progress toward delivering the results. Use the "Cancel" button to discontinue the results generation prior to completion.

6 Analysis Options

6.1 Data Aggregation Method

The value selected for the Data Aggregation Method option determines how aggregate signal levels and consensus data views will be calculated, using means or medians.

Individual - When Individual is selected for the Replicate Data option, each individual Observation Block item will be subjected to the requested analytical procedures and will appear in the resulting analytical output as a result item.

Consensus - When Consensus is selected for the Replicate Data option, replicate Observation Block items will be merged into a single consensus Observation Block representation, and the consensus Observation Block item will be subjected to the requested analytical procedures and will appear in the resulting analytical output as a result item.

For more information regarding replicate data [see replicate data.]

6.2 Replicate Data

The value selected for the Replicate Data option determines how replicate Observation Block items will be handled.

For more information regarding replicate Observation Block determination see [identification of replicate data].

Individual - When Individual is selected for the Replicate Data option, each individual Observation Block item will be subjected to the requested analytical procedures and will appear in the resulting analytical output as a result item.

Consensus - When Consensus is selected for the Replicate Data option, replicate Observation Block items will be merged into a single consensus Observation Block representation, and the consensus Observation Block item will be subjected to the requested analytical procedures and will appear in the resulting analytical output as a result item.

For more information regarding merging see [merging of replicate data].

6.3 Consensus Method

The selected algorithm is used to define the consensus sampling for the Observation Block that will be the merged representation Observation Block replicate items.

Basic Resampling - The Basic Resampling algorithm provides a consensus sampling that represents each distinct concentration value in any of the individual replicate items for each test substance. Distinct concentration values are rounded to the nearest value within the precision specified for the option Resample Precision.

For a more detailed description of the Basic Resampling algorithm, see [Basic Resampling Algorithm]

Regular Resampling - The Regular Resampling algorithm provides a consensus sampling that is the fixed dilution factor which is the best match for the concentrations represented by the individual replicate items for each test substance. The minimum acceptable sampling is controlled by the Minimum Sample parameter of the algorithm.

For a more detailed description of the Fourier Resampling algorithm, see [Regular Resampling Algorithm].

6.4 Consensus Aggregation Ext

When the Replicate Data option has the value Consensus, the value of Consensus Aggregation Ext defines additional protocol parameters that will be used to determine replicate Observation Block items. The value of Consensus Aggregation Ext is a semicolon delimited list of parameter names defined by the Data source and Protocol. For more information regarding Data source see [Data sources].

For more information regarding Protocol and Protocol Parameter see [Protocols].

6.5 Test Substance Orientation

Test substances and/or substance mixtures are oriented canonically in multi-dimensional Observation Block items to ensure consistent presentation.

- **Default** - Default canonical orientation is used to produce analytical output.
- **Reversed** - Reverse canonical orientation is used to produce analytical output.

6.6 Analytical Output

The value selected for the Analytical Output option determines the scope of analytical procedures to which the Observation Block items are subjected.

For more information see the Analytical Output and Analysis Methods handbook.

- **Minimal** - The minimal analytical output is created for each Observation Block item.
- **Full** - The full analytical output is created for each Observation Block item.

7 Specifying Data Characteristics

7.1 Data source Query

The data source query panel is used to specify the characteristics of observation blocks to be subjected to analysis and output in a dataset.

Individual characteristics of observation blocks are specified by selecting an attribute, an operator, and one or more values which are to be tested using the selected operator. Multiple values specified for an individual characteristic are tested serially, and all observations blocks which match based on the attribute, operator, and any individual value are considered to match the specified characteristic.

	Attribute	?	Value(s)	
+	Substance(s) ▾	* ▾	Aspirin	X

The reserved keyword, "with", can be used to search multiple substances for combination results: "Aspirin with Acarbose", for example. Note that there is an implied wildcard (*) following each term so searching for "Aspirin with Acarbose" is analogous to searching for "Asp with Aca" except that the latter would also return results for Asparaginase with Acamprostate.

Multiple characteristics can be specified (using the [+] button) and are considered in parallel to determine the subset of observation blocks that match all specified characteristics.

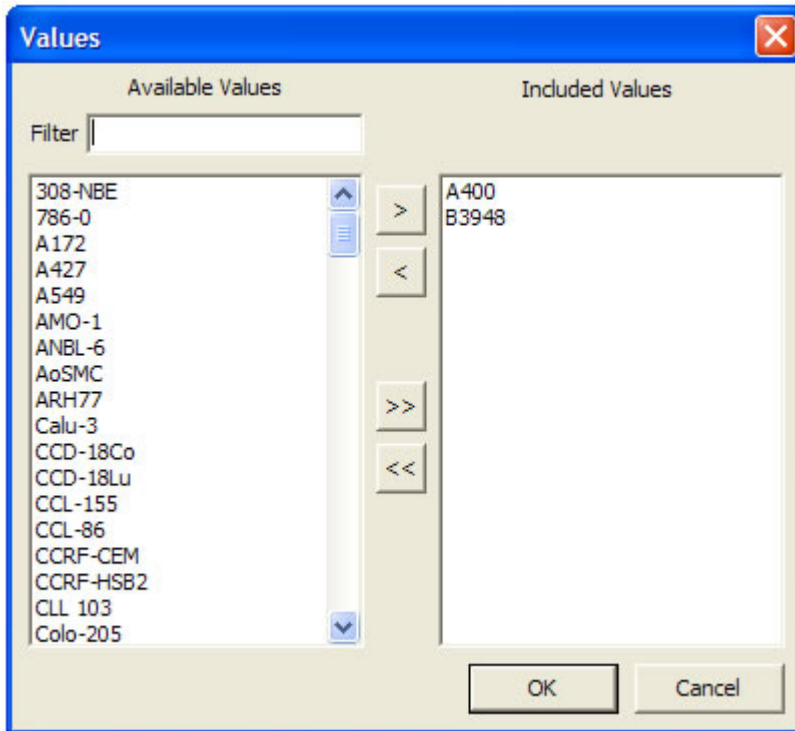
	Attribute	?	Value(s)	
+	Substance(s) ▾	* ▾	Aspirin	X
	Cell Line (Onc ▾	= ▾	A400; B3948	... X

Multiple groupings of characteristics can be specified by invoking the [||] button in the leftmost column header to create additional groupings. Each grouping of characteristics is considered serially, and all observation blocks matching the characteristics of any group are subjected to analysis. Groupings are portrayed using alternating background color.

	Attribute	?	Value(s)	
+	Substance(s) ▾	* ▾	Aspirin	X
	Cell Line (Onc ▾	= ▾	A400; B3948	... X
+	Substance(s) ▾	* ▾	Prednisolone	X
	Cell Line (Onc ▾	= ▾	A384; C388	... X

Any individual characteristic can be removed by invoking the [X] button at the rightmost position of each line.

The data source provides enumerations for some attributes. If an attribute is enumerated, then the query panel will display the [...] button to the right of the value(s) editor. Invoking the [...] button will cause the value selector form to be displayed.




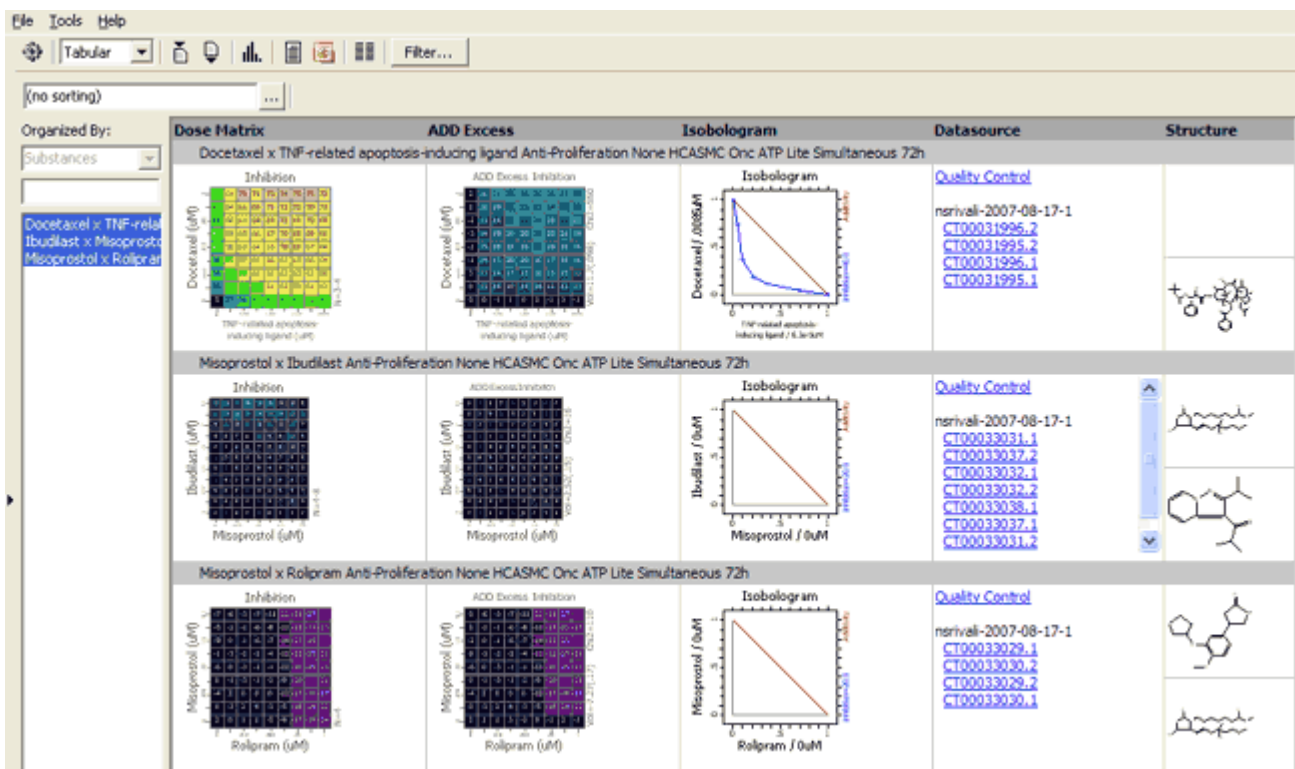
The value selector form is used to select one or more values to be included for the individual characteristic. Available values are shown in the listing on the left. Included values are shown in the listing on the right. Both listing support the selection of multiple items using the <ctrl> and <shift> keys. With items selected in the available values listing, the [>] button will cause those values to be represented in the included values listing. With items selected in the included values listing, the [<] button will cause them to be excluded from the included values listing. The [>>] button will cause all values from the available values listing to be included in the included values listing. The [<<] button will cause all values to be excluded from the included values listing.

The filter field is used to refine the list of available values based on one or more keywords or phrases. When keywords are present, separated by whitespace, the available values list is restricted to only values that contain all of the keywords in the filter field. Keywords containing whitespace may be entered by surrounding the keyword with the quotation character.

Reviewing Data

Using the Tabular View

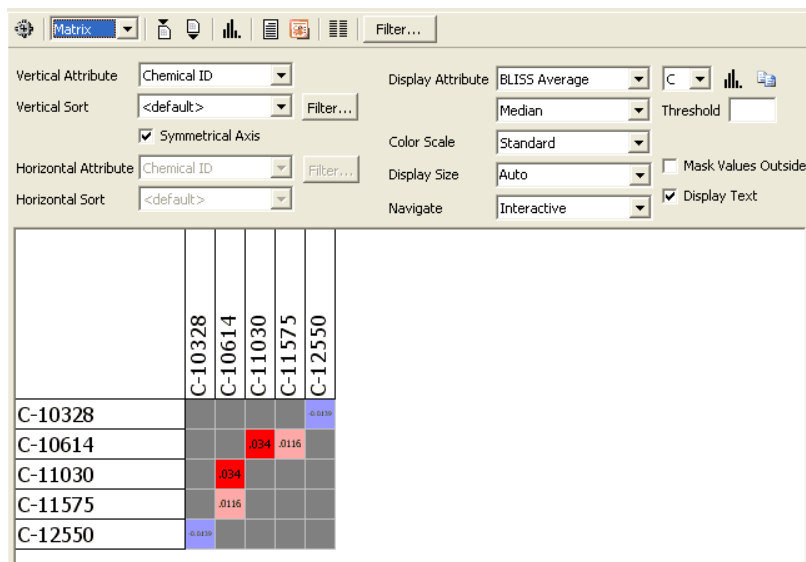
While [the Matrix view](#) offers a more global view of your result data, the tabular view is an Excel-like view with each selected item from the result pane corresponding to a row in the detail pane. The columns in the detail pane correspond to attributes of each item. Use the Select Columns button  to show and hide these various detail attributes.




The details pane is not sorted by default but click the ellipsis button to the right of the "(no sorting)" input to specify your sorting criteria. You might also want to use the filter button to [filter your results](#).

8 Using the Matrix view

The Matrix view provides a more global view of your result data, in contrast to [the Tabular view](#). Placing various attributes of your data on each axis provides numerous ways to shape your data. You can quickly look across your entire data set or [use filters](#) to selectively pare down your results.




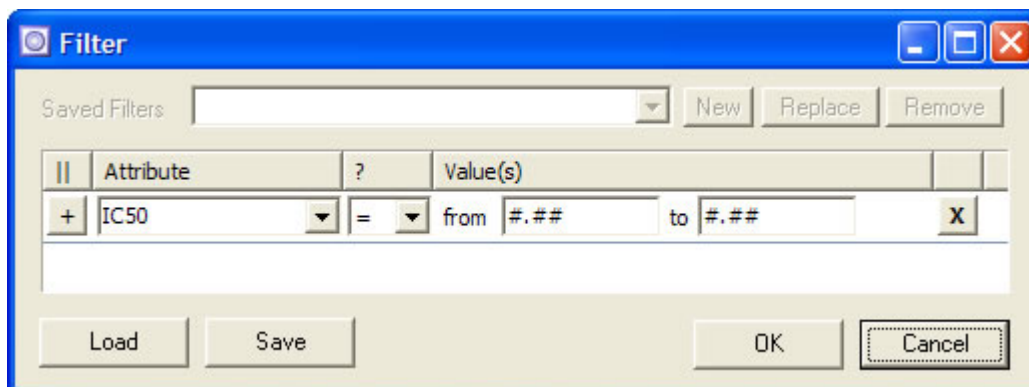
Once you have assigned, and optionally sorted, axis attributes, choose a measure using the display attribute dropdown. Further limit values by using a threshold that applies to your chosen display attribute.

As elsewhere, the histogram button  will allow for further analysis and paring down of your dataset.

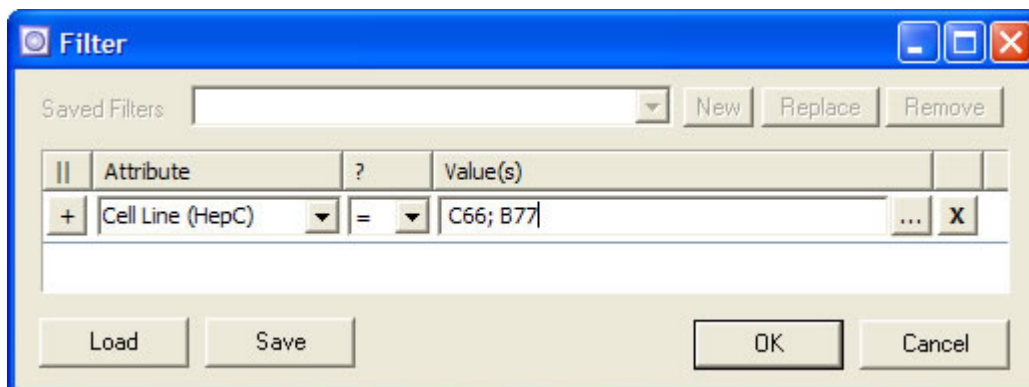
By choosing "Interactive" as the Navigate option, double-clicking individual data points in the matrix will pull up the tabular view of that data.

9 Filtering

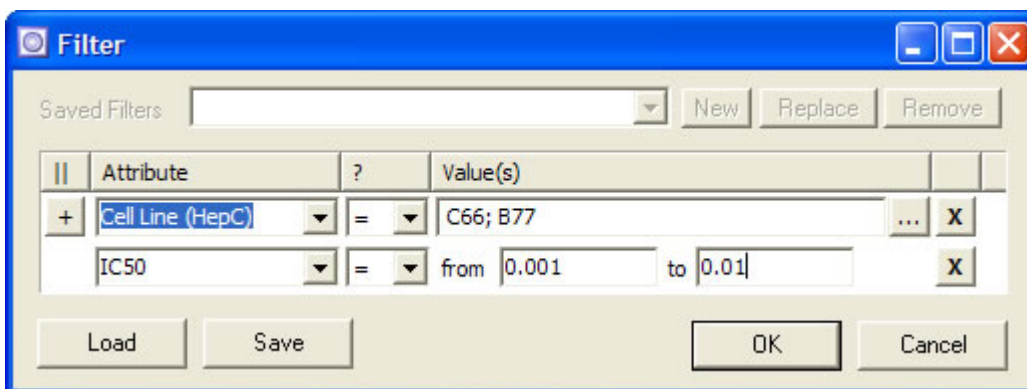
Filtering is performed by invoking the "Filter..." button  in the main toolbar, and specifying the filter specification in the filter dialog. The filter dialog is used to specify the characteristics of individual result items which are used to determine which items are active in the displays. Filters can be saved and reloaded at a later time using the "Load" and "Save" buttons in the lower left hand corner of the dialog.



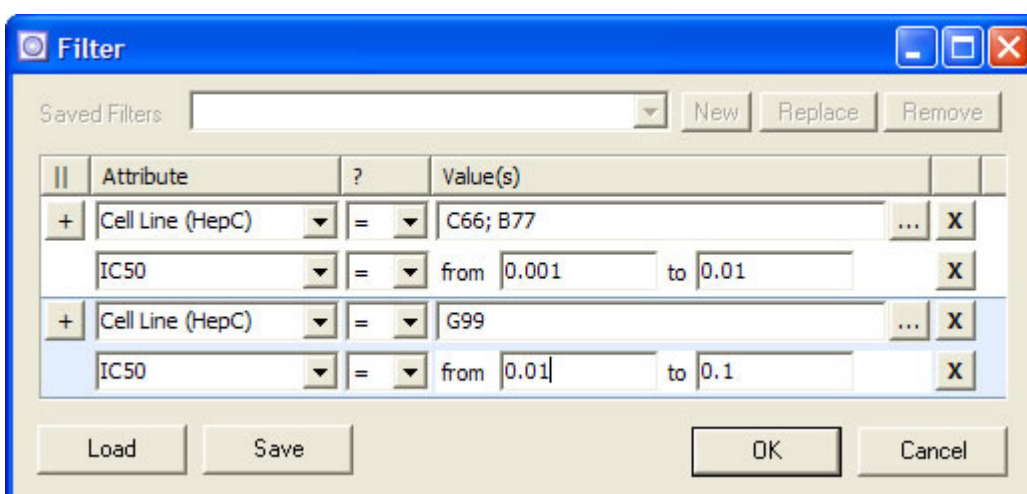
Individual characteristics of result items are specified by selecting an attribute, an operator, and one or more values which are to be tested using the selected operator. Multiple values specified for an individual text characteristic are tested serially, and all result items which match based on the attribute, operator, and any individual value are considered to match the specified characteristic. Numeric characteristics are specified as a range defined by two input values. Specifying both values will include all items inclusive in the range. Specifying only the lower limit or the upper limit will include all items above or below the limit inclusive, respectively.



Multiple characteristics can be specified (using the [+] button) and are considered in parallel to determine the subset of result items that match all specified characteristics.

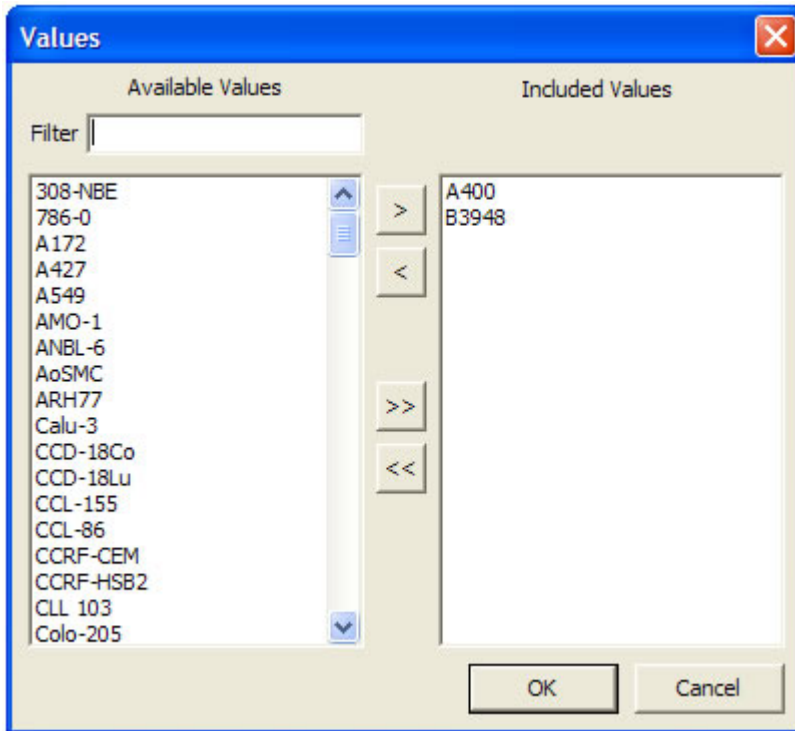


Multiple groupings of characteristics can be specified by invoking the [||] button in the leftmost column header to create additional groupings. Each grouping of characteristics is considered serially, and all result items matching the characteristics of any group are made active. Groupings are portrayed using alternating background color.



Any individual characteristic can be removed by invoking the [X] button at the rightmost position of each line.


The filtering mechanism provides enumerations for some attributes. If an attribute is enumerated, then the query panel will display the [...] button to the right of the value(s) editor. Invoking the [...] button will cause the value selector form to be displayed.



The value selector form is used to select one or more values to be included for the individual characteristic. Available values are shown in the listing on the left. Included values are shown in the listing on the right. Both listing support the selection of multiple items using the <ctrl> and <shift> keys. With items selected in the available values listing, the [>] button will cause those values to be represented in the included values listing. With items selected in the included values listing, the [<] button will cause them to be excluded from the included values listing. The [>>] button will cause all values from the available values listing to be included in the included values listing. The [<<] button will cause all values to be excluded from the included values listing.

The filter field is used to refine the list of available values based on one or more keywords or phrases. When keywords are present, separated by whitespace, the available values list is restricted to only values that contain all of the keywords in the filter field. Keywords containing whitespace may be entered by surrounding the keyword with the quotation character.

10 Importing Data

Importing of additional data values is performed by actuating the import button  from the main toolbar, and selecting the tab delimited file that contains the additional data values to be imported. Importing of additional data is performed subsequent to querying of data from the data source.


The import files follow a standard tab delimited format. The structure of the file is one or more columns of data, where the first column represents an existing attribute to be used as a key, and subsequent columns represent attributes to be appended to the data set.

As an example, the following import file specifies additional test substance attributes named "Att1", "Att2", and "Att3" to be appended to the existing data set. The attribute "Chemical ID" is used as the key attribute.

Chemical ID	Att1	Att2	Att3
C-10234	A	E	I
C-10235	B	F	J
C-10236	C	G	K
C-10237	D	H	L

11 Exporting and Reporting

11.1 Exporting Tabular Datasets

The active data set can be exported to a tab delimited file for use in external analysis programs. The exported data set is created by actuating the export button  and selecting a location and name for the exported file.

11.2 Creating PDF Reports

PDF reports can be generated for individual result items. Each result item forms the basis for a single report page. The PDF report can be single a combined report for multiple data items or alternatively multiple individual files for each data item.

Step 1 - Select items to include in the report. Available options are "Entire Dataset", "Active Filter", and "Current Selection".

Step 2 - Refine included items. The items in the current selection can be further refined for inclusion in the report by modifying the state of the checkbox associated with each item in the list.

Step 3 - Choose an output mode.

- **Single combined file** - When this option is selected, each of the items becomes a row and each of the attributes becomes a column. The result is saved to a tab delimited text file. In this mode, multiple attributes can be selected but any attribute that represents an "Image" or a "Preformatted Text" value cannot be output into the text file. Instead, the value "(drawing)" or "(text)" is placed into the output "cell". A file naming dialog will be displayed when Ok is actuated, and all report pages will be created in the individual file.
- **Individual file per item** - The purpose of this option is to allow the mass creation of exported image or preformatted text files as individual files. Conceptually, this is like exporting a whole column of data and creating an individual file for each cell. This mode is available only when exactly one attribute is selected and the selected attribute is an image or preformatted text type. When Ok is actuated, a file location dialog will be displayed and the individual report files will be created according to the file naming convention specified. The file naming convention interprets text within the square braces to mean a variable attribute of the individual data items. All other text is interpreted literally. To control the naming of the output files, a filename prototype is specified that can contain variable values that can reference other attributes of the exported item. For example, you can export all isobolograms and name the image files according to the "Title" attribute that is associated with the item to which the isobologram belongs.

Report Options ✖

Include the following items in the report:

- Entire Dataset (9284 items)
- Active Filter (undefined)
- Current Selection

Chemical 0001 x Chemical 0001 HepC Replicon HuH7 Luc 4000 48h None

Create the following report output:

- Single combined report
- Individual report per item

File naming