PepSlide® Analyzer 1.4

User's Quick Start Guide

03 June 2013 - Rev 10

PepSlide® Analyzer is only intended for research and not intended or approved for diagnosis of disease in humans or animals.

SICASYS Software GmbH

Im Neuenheimer Feld 583

D-69120 Heidelberg

Germany

Phone	+49 (62 21) 7 28 50 40
Fax	+49 (62 21) 7 28 48 94
Email	info@sicasys.de
Web	www.sicasys.de

PepSlide® Analyzer (PSA) facilitates microarray data analysis and is particularly suited for peptide arrays and peptide microarrays. Discovery of meaningful information is supported with statistical data mining tools. In addition, the software enables analysis of peptide libraries.

This Quick Start Guide describes basic commands of PSA to enable immediate access to the functionalities of the software solution. More details are found in the User's Guide.

1.   Preparation

Installation

PSA is natively supported on both Windows and Mac OS X platforms, thus the software is able to run with high performance and native look. Installation of PSA requires rights of a system administrator.

For Windows users, PSA works on Windows XP, Windows 7, and Windows 8. Simply run the PSA setup. If the current Windows account is not an administrator, the user will be asked to input an administrative account and its password.

On Mac OS X platforms, the software runs on Mac OS X 10.7 and 10.8. Unzip the package and double-click on the .pkg file to launch the installer. During the installation the user will be prompted to provide a system administrator's account and password. Upon completion, PSA is installed in the */Applications/PepSlide/Analyzer* folder.

Analysis of Peptide Microarray Data

A PSA analysis project, including the input data and analysis results, can be saved and reopened with the *Project* menu. To analyze a peptide microarray, two input data are required:

- Scanned images of the microarray of TIFF format.
- An array design file with spot annotations of either GAL or PSF format.

2.   Quantification of Microarray Data

Input Data

- Images: Choose the *Images > Open Image* menu. For grayscale images, select either the *Red* channel or the *Green* channel to display the signal of each image.
- Array design: Choose the *Arrays > Open Array* menu.

The *Imaging* sheet consists of a control panel and a graphical canvas. After being loaded, the images and the array are shown in the canvas. To obtain an appropriate view, the user can use the *Zoom In* and *Zoom Out* buttons on the toolbar or select a predefined zoom level in the *Zoom* combo-box. Alternatively, an arbitrary value can be entered directly into this combo-box. The panel displays properties of the image and selected objects in the array, in addition to options for image processing and data quantification tasks.

If the array file is of PSF format, spots are grouped into an object called *Spot Family (SF)*. A SF is either a *peptide library* or a *control*. To view the information of an SF, first activate the *Spot Family* tab in the *Design Information* section of the *Imaging* sheet, and then click on the SF in the canvas.

In the case the array file is of GAL format, the software shows properties of a GAL block in the *Block* tab, which is located next to the *Spot Family* tab. When the *Block* tab is activated, the user can select a block, align it by moving and rotating, and quantify its spots.

The information about a peptide spot, such as its unique position on the array and the sequence of amino acids, can be viewed similarly: First activate the *Spot* tab, and then click on that spot. Within each spot, the *spotted region* is bounded by a *dashed rectangle* or *circle* depending on the spot's specification. During the array alignment, the user will reallocate the array such that the spotted regions as close to the spots in the images as possible.

Array Alignment

Before aligning the array, the user may need to adjust the images' contrast and brightness to make the spots more visible. These utilities are available in the *Imaging* sheet > *Image Processing* section > *Pixels* tab.

For aligning spots in the array with their counterpart in the microarray image, first select the spot family (SF) containing the spots and then adjust the SF's position. Please note that one can only move or rotate SFs, not spots. Therefore, the *Spot Family* tab in the *Design Information* section of the *Imaging* sheet must be activated.

- Selecting SFs
  - Use Ctrl-A to select all SFs in the array.
  - To select individual SFs, click on the respective SF in the canvas. Hold down the Ctrl key for selecting multiple SFs simultaneously.
- Moving SFs: Click on the selection and drag it to the desired position.
- Rotating SFs: Choose the *Imaging* sheet > *Image Processing* section > *Rotation* tab.
  - Image: Flip and rotate images at angles of 90°, 180°, or 270°.
  - Array: Rotate the selected SFs at an arbitrary angle.

When a GAL file is loaded, blocks can be selected, moved, and rotated like SFs. Please remember to activate the *Block* tab before using functions on blocks.

The aligned array file can be saved with Ctrl-S or *the Arrays > Save* menu. It can be saved to another array file using the *Arrays > Save Array As* menu.

Adjustment of Spot and Grid Size

In addition to the reallocation of SFs or blocks, the user may need to adjust the spots' diameter or the distance between them, so that the spots in the array fit their counterpart in the image(s). These parameters can be changed in the *Imaging* sheet > *Design Information* section > *Configuration* tab.

Quantification of Microarray Data

- Right-click at a point within the image but outside any SF or block. Choose *Image Processing (entire array)* to quantify all spots of the microarray.
- To quantify selected SFs or blocks, right-click on the selection and choose *Image Processing*.

For each spot of the array, the median and the mean of its *raw*, *background*, and *foreground* values in each channel (Red and/or Green) are calculated. *Raw* represents the intensity value of the spot's signal. *Background* is the estimated value of the signal caused by non-specific binding. The value of interest is *foreground*; it is computed by subtracting the background from the raw value.

Furthermore, when a peptide is replicated the software also provides the intensity value of the peptide calculated by aggregating signals from its replicated spots. If e.g. P1 and P2 are replicas of a peptide P, then the results include the intensity values for P, in addition to those for P1 and P2.

In the *Imaging* sheet, the user can view the spots and their quantified data at the same time. Selecting a spot in the array highlights its quantified results in the table and vice versa. On the other hand, the *Imaging Results* sheet shows only the quantified data. The user can export the quantified results to a CSV file for further analysis.

Considering Change of Intensity Values during Quantification

During the array alignment, the user may have adjusted the contrast and/or the brightness of the images to make the spots visible. This changes the images' signal and alters the analysis results. It is recommended that *the original image data be used for the quantification.* Therefore, by default *the quantification procedure uses the original image's signal*, i.e. it *excludes* changes made to the image's signal due to inversion, noise filtering, or the adjustment of contrast and/or brightness.

This option can be intentionally changed as follows. In the *Imaging* sheet > *Image Processing* section > *Quantification* tab, select the *Included option* below the *Change of the images' intensity value* label.

To use the original image data for the data quantitation again, select the *Excluded option.* Please note that when the this option is chosen, the *Spot Image* widget in the *Imaging*, *Imaging Results*, and *Peptide Library Analysis* sheets shows the spot's counterpart image with the original image's signal.

Spot Detection Methods

In the *Imaging* sheet > *Image Processing* section > *Quantification* tab, the *Spot Detection Method* defines how the raw and background values of a spot are calculated.

- *Fixed-Spot*: The software uses the signal in the spotted region to compute the raw value. The background value is calculated from the signal of the remaining region within the spot. This is the default method. It is recommended if the spots look consistent throughout the scanned images and match well with the spotted regions in the array.
- *Other methods*: The software combines the known information about the spotted region with the detected signals within each spot. It then flexibly separates the region consisting of signals for the raw value from the one for the background value.

The default method for calculating the background value is called *Local Method*; it uses only the spot's signal. By means of background controls, the user can have background values using the signal from other regions of the image. Background correction methods can be selected in the *Imaging* sheet > *Image Processing* section > *Quantification* tab.

Background Correction using Background Controls

Instead of having the background values automatically calculated as explained above, the user can use background controls to explicitly specify image regions from which the background values are calculated. One may choose to have a *global background value for the entire array* using a background control as follows.

- Right-click at a point in the image where the signal may represent the background for the spots. Choose *Create Background Control*. Specify the shape and the number of spots.
- Perform the data quantification. All spots in the array now have the same background value which is obtained from the background control. We call that an "*association*" between the SFs and the background control.

In addition to a global background value for the entire array, one can flexibly choose *an individual background value for a spot family*. This is done by creating a background control just for the SF and establishing the association between them.

- Right-click on the SF to popup the context menu. Choose *Background Control > Remove Association* to release the SF from the first background control.
- Right-click on the SF again and choose *Background Control > Establish Association*. Click on the background control with which you want to associate this SF.

For GAL files, having a global background value or an individual background value for a *block* can be done similarly; the user creates a background control for the blocks and associates them.

Please note that the background correction method is automatically set to *Background Controls* after the creation of the first background control. In the *Imaging* sheet > *Image Processing* section > *Quantification* tab, the user can switch back to use the *Local Method*.

Customization of Quantified Data

Click the *Configurations* menu and choose *Imaging Result Values* to show the setup dialog. Here, the user can add or remove parameters to be calculated during the data quantification.

3.   Peptide Library Analysis

PepSlide® Analyzer facilitates analysis of a peptide library using its design information. Take an overlapping peptide library for example; the design information includes the protein, the shift step, and the peptide length. The software supports analysis of *random*, *overlapping*, and *substitution* (also called *permutation* or *positional scanning*) peptide libraries.

If the array file is of PSF format, simply select the spot families, right-click on the selection, and choose *Peptide Library Analysis*. The results are shown in the *Peptide Library Analysis* sheet.

Analysis of peptide library cannot be performed with GAL files in the first place. However, the user can easily specify the design of a peptide library for GAL blocks (e.g. the protein, the shift step, and the peptide length of an overlapping peptide library), as detailed in Section 4. The analysis can then be preceded as with SFs. The distinction between a block (contains only spots and their annotation) and an SF (a peptide library and may consist of multiple blocks) enables a convenient use of GAL files.

Analysis of Overlapping Peptide Libraries

This analysis aims at selecting peptide spots with strong signals. It is also able to detect consensus motifs shared by some overlapping peptides.

A two-dimensional chart in the center of the sheet depicts the spots. Their coordinates can be selected among the quantified intensity values. By moving the lower and upper threshold bars, the chart facilitates the selection of spots whose intensity values being within the two thresholds. Selected spots are then populated into the table below the chart. Here, overlapping peptides sharing a consensus motif are highlighted with an individual color. Right-click on the highlight and choose *Show Consensus Motif* to view the information.

By clicking on a spot in the chart or selecting the corresponding row in the table, the user can view the spot information such as the peptide and its counterpart image. The information is shown in the *Spot Details* widget located on the right of the sheet. The chart, the table, and the *Spot Details* widget thus provide the user with different angles to look at the data of the peptide library.

The user can export the data in the table to a CSV file. The chart can also be saved to an image file.

Analysis of Random Peptide Libraries

This analysis facilitates browsing and selecting peptide spots with strong signals for further optimization. Except for the detection of consensus motifs, the user interface and functions in this analysis are the same as those in the analysis for overlapping peptide libraries.

Analysis of Substitution Peptide Libraries

Peptides of a *substitution* peptide library are created by replacing residues of a known peptide, the so-called *parent peptide*, with some other amino acids. The purpose of this analysis is thus to examine all variants of the parent peptide and select those of the user's interest.

All substitutions and their intensity value are simultaneously shown in a three-dimensional chart. Here, the (P)-Axis lists residues of the parent peptide and the (S)-Axis contains amino acids for replacing, thus a "point" in the (PS)-plane features a substitution. The user can evaluate different intensity values of the substitutions represented by the corresponding bar along the Z-Axis. To change the perspective of the 3D chart, hold-down the left-mouse and move the mouse.

Spots are selected by entering the lower and/or upper threshold values in the *Spots of Interest* section in the left panel. The selection will be populated into the table below the chart. Similar to other analyses, the user can view information about a spot in the *Spot Details* widget. Selecting a spot in the table highlights its version in the chart and vice versa.

For quality control, the user can choose to display only spots having the same amino acid sequence as the parent peptide. Since these spots are the same peptide, their intensity values should not be considerably different. This option is found in the *Spots with Parent Peptide* section in the left panel.

The chart can be saved to an image file. In addition, its data can be exported to a CSV file. The user can also save the selected data in the table to a CSV file.

4.   Peptide Library Analysis for GAL Files

The software supports a convenient way to work with GAL files, even for the analysis of peptide library. Whether the peptides of a *peptide library (PL)* are spotted into one block or different blocks, the user can easily reconstruct the PL and then analyze it as mentioned in section 3. The reconstruction can be done in two steps:

- Specify a peptide library design for the blocks. This creates a PL consisting of these blocks.
- If the GAL file does not contain the peptide in the spots' ID, import the peptide mapping file.

Each step is described in detail below. For reading convenience, we assume that an overlapping PL is being reconstructed. The procedure is the same for substitution and random peptide libraries.

Specify a Peptide Library Design for GAL Blocks

- Hold down the Ctrl key while clicking on the blocks to select them simultaneously.
- Right-click on the selection and choose *Type > Overlapping Peptide Library*.
- Enter the peptide library design such as the protein, the shift step, and the peptide length.

An overlapping peptide library consisting of the blocks has been created. The user can activate the *Spot Family* tab and review the information of the PL.

Import the Peptide Mapping File

This step is only necessary if the ID of the spots in the blocks does not contain the peptide yet.

- Activate the *Spot Family* tab to work on peptide libraries.
- In the canvas, select the PL by clicking on one of its blocks. Right-click on the selection and choose *Import Peptide Mapping File* from the context menu.
- Specify the linker which will be excluded from the peptides. Leave it blank if no linker is used.
- Select the file which maps the peptides of the PL to their spot ID in the blocks.

The peptide mapping file is a Tab-delimited text file or a CSV file. It has two columns; the first one contains the ID of the peptides in the blocks and the second column contains the peptides. The first line of the file is only for annotation and will be skipped. An example of the file's content is shown below. In this example, the linker is Cx.

Spot_ID	Peptide
1	CxMGPGIAAVLLSLAVA
2	CxGIAAVLLSLAVALAR
3	CxAVLLSLAVALARVPA
…	
128	CxLLDEGPGPARRGAPY

After importing the peptide mapping file, activate the *Spot* tab in the *Design Information* section. The user can view the peptide in the *ID* field when clicking on a spot in the canvas. The overlapping PL has been reconstructed. The user can now analyze the newly constructed PL as described in Section 3.

Please note that *saving the GAL file also keeps the information about the peptide library design*.

5.   Statistical Data Mining Tools

Input Datasets for Data Mining Tools

In both Principal Component Analysis and Hierarchical Clustering sheets, click the *Load Data* button to select an input dataset. It can be:

- a list of PepSlide® Analyzer files (*.psa);
- a list of GenePix Result files (*.gpr);
- a CSV file exported from a dataset.

The PSA and GPR files must contain the quantified microarray data for the same set of *features* (for peptide microarrays a feature can be a peptide). In other words, they all based on a *template array file*. In the dataset, the *samples* (*experiments*) are named after the files' name. The features' value is selected among the list of quantified intensity values in the *Evaluation Parameter* section.

After loading, the dataset contains only the quantified values selected from the input files and thus has small size. If the user plans to analyze this dataset again, export it to a CSV file. It is much faster to load this CSV file than to read many PSA or GPR files in each analysis.

Principal Component Analysis (PCA)

Click the *Analysis* menu and choose *Principal Component Analysis* to activate the PCA sheet.

- Load the input dataset and specify the evaluation parameter if necessary.
- Choose to reduce the dimension of features or samples.
- Click the *Start PCA* button.

The projection of the input dataset to the *three* features or samples that most probably describe the relationship between the data is depicted in the chart. It is three-dimensional initially and can be switched to two-dimensional views.

The input dataset, the eigenvalues, and the eigenvectors are shown in the table. They can be exported to a CSV file. In addition, PCA charts can be saved to an image file using the context menu.

Hierarchical Clustering (HC)

Click the *Analysis* menu and choose *Hierarchical Clustering* to activate the HC sheet.

- Load the input dataset and specify the evaluation parameter if necessary.
- Choose to construct the clustering tree for features, or samples, or both.
- Select the desired *Distance Metric* and *Algorithm Type*.
- Click the *Start HC* button to begin analysis.

The clustering tree(s) are then constructed. The values in the dataset, each represents a feature in one sample, are graphically represented by means of a heat map.

The clustering trees with the heat map can be saved to an image file using the context menu.

K-means Clustering

After analyzing *overlapping* or *random* peptide libraries, the user can use K-means Clustering to classify peptides into many groups based on their quantified intensity values.

- Click the *Analysis* menu and choose *K-means Clustering* to activate the setup dialog.
- Enter the desired number of clusters and choose their color if necessary. Click OK; the chart then shows spots highlighted in their cluster's color. The chart can be saved to an image file.
- In the *Clusters* section in the left panel, select the cluster of interest. The table then shows only spots of that cluster. Data in the table can be exported to a CSV file.