Unified Software Solution for Efficient SPR Data Analysis in Drug Research

Göran Dahl1*, Stephan Steigele2*, Per Hillertz3, Anna Tigerström1, Anders Egnéus3, Alexander Mehrle2, Martin Ginkel2, Fredrik Edfeldt1, Geoff Holdgate4, Nichole O’Connell5, Bernd Kappler2, Annette Brodte2, Philip B. Rawlins4, Gareth Davies6, Eva-Lotta Westberg3, Rutger H. A. Folmer1, and Stephan Heyse2

Abstract
Surface plasmon resonance (SPR) is a powerful method for obtaining detailed molecular interaction parameters. Modern instrumentation with its increased throughput has enabled routine screening by SPR in hit-to-lead and lead optimization programs, and SPR has become a mainstream drug discovery technology. However, the processing and reporting of SPR data in drug discovery are typically performed manually, which is both time-consuming and tedious. Here, we present the workflow concept, design and experiences with a software module relying on a single, browser-based software platform for the processing, analysis, and reporting of SPR data. The efficiency of this concept lies in the immediate availability of end results: data are processed and analyzed upon loading the raw data file, allowing the user to immediately quality control the results. Once completed, the user can automatically report those results to data repositories for corporate access and quickly generate printed reports or documents. The software module has resulted in a very efficient and effective workflow through saved time and improved quality control. We discuss these benefits and show how this process defines a new benchmark in the drug discovery industry for the handling, interpretation, visualization, and sharing of SPR data.

Keywords
automation or robotics, database and data management, label-free technologies, pharmacology, ligand binding, receptor binding, general pharmaceutical process

Introduction
Surface plasmon resonance (SPR) is a powerful biophysical method for measuring molecular interactions. The versatility of the technology means that it can be used both as a screening platform to identify binders and for detailed kinetic characterization of binding interactions. This has resulted in SPR being used widely in the drug discovery community.1-4 With the increased interest in the binding kinetics of drugs5,6 to better understand the mode of binding and possibly predict therapeutic effectiveness, the demand for SPR data is rising. In order to accommodate this increased demand, AstraZeneca explored how it could simplify and optimize its SPR workflow.

For this streamlining exercise, instrumentation and analysis software were scrutinized to identify opportunities for optimization. Currently, the small-molecule drug discovery performed by Discovery Sciences at AstraZeneca uses the General Electric Healthcare (GEHC) Biacore 4000, S200, T200, and 3000 SPR instruments. Two major areas for improvement in the Biacore software were readily identified. First, each SPR instrument comes with its own control and data evaluation software, which are aimed at the primary uses of the particular platform, such
as compound screening (Biacore 4000) or affinity and kinetic characterization (Biacore 3000, S200, and T200). As a consequence, data generated on one platform cannot be easily analyzed with the evaluation software from another. Second, the existing evaluation software is poorly interfaced with the overall reporting structure within most pharmaceutical companies, where data have to be deposited into data repositories and electronic lab books in customizable formats.

A list of features desired in an SPR data analysis software was created (Table 1). A number of SPR analysis solutions can do some of the things listed, but no software existed that could do all of them. Developing a new software at AstraZeneca, including all the desired functionality, was neither desirable nor practical. Coinciding with this process, AstraZeneca had started to use Genedata Screener as its software of choice for screening data evaluation purposes.7 Screener is a software platform for processing and analyzing data from numerous different sources, and it has the ability to directly report results into Excel or pdf documents and, an important benchmark for this project, fully automatically report data into corresponding repositories.8 Hence, it was logical at this time to reach out to Genedata and investigate the possibility to use their software for SPR data analysis. After some initial discussions, it was decided to codevelop a module in Screener capable of handling SPR data and fulfilling the requirements as listed in Table 1.

We required two key features for the SPR Screener module:

1. A unifying software platform for analysis and reporting of all current and future SPR data
2. Direct reporting of processed data to the corporate database

These two requirements would enable the analysis and reporting of all SPR data, irrespective of data origin (instrument) or data type (screening results vs. detailed kinetic characterization). This would also remove any potential variability from the use of different evaluation software. In addition, there was an automatic reporting procedure to capture both end results (such as on- or off-rate constants, and Rmax and Kd values) and sensorgram data (raw images) into the corporate database for more effective sharing with project teams. With this, we aim to replace the manual copy and pasting of data and/or images between different software packages and workstations (Fig. 1).

### Methods

Setting up SPR experiments destined to be analyzed with Screener for SPR requires the same amount of input from the user as if the data were to be analyzed with the legacy software. Hence, there is no difference in the method files. In order to be analyzed, SPR data have to be processed in several steps: raw data parsing, preprocessing, postprocessing, and end result calculation. In order to meet the requirements listed in Table 1, the following functionalities were implemented in the Genedata Screener SPR module.

### Raw Data and Protocol Parsing

Most raw data from various biophysical binding assay platforms (e.g., GE Biacore, Pall Octet, or Creoptix) and different experimental setups have a similar structure. A cycle is the basic unit of data representation. It is a sequence of either one compound injection with wash-off or multiple injections of increasing concentration of the analyte. An individual capture step can also be part of each cycle, such as...
as capturing a protein to an immobilized antibody. Besides the \( x \) coordinate (time) and \( y \) coordinate (response signal) that comprise the actual sensorgram, each cycle also contains information on the flow cell used, an analyte identifier (i.e., compound ID), the analyte concentration(s), the analyte molecular weight, and the injection start and stop time points.

SPR data require extensive annotation for processing, with the main requirements being the specific experimental protocol and timing information. When parsing the data, each flow cell/spot is represented as a “well” on a single “virtual plate” in Screener. This results in 1 × 4 well virtual plates for Biacore 3000, T200, and S200 data (one flow cell with four different channels) and 4 × 5 well plates for Biacore 4000 data (four flow cells with five spots in each). Each cycle is represented in a separate virtual plate, and each measured sensorgram is represented as a curve associated with one well.

The following properties are set automatically during data loading if present in the binary instrument output file or in the Biacore 4000 database, respectively: numbering of cycles, sample types, reference cell per cycle, spot type (like “Ligand” or “Reference” plus optional suffix), original plate barcode, original well position, ligand name, the well index of the well that should be used for reference subtraction of this well, molecular weight (in daltons), and injection start and stop times.

**Preprocessing Methods**

Preprocessing methods are required to align and present sensorgrams in an analyzable and meaningful way. The user can monitor every preprocessing step and observe the effect of preprocessing on the sensorgrams with a single mouse click. It is also possible to adjust the settings to control how the preprocessing is performed.

- **Baseline adjustment**: The baseline adjustment method aligns traces from all wells to a common baseline of \( y = 0 \) prior to the start of the first injection.
- **Time alignment**: All injection start time points from different flow cells and spots are aligned by using the first derivative of each trace to identify the precise injection time from the data.
- **Reference spot subtraction**: The signal from a control spot or channel is subtracted from the signal on the active surface. Any spot or channel can be used as a reference, a treatment that is only supported in the current Biacore 3000 software.
- **Blank subtraction**: The signal from a buffer or DMSO control injection is subtracted from the signal on the active surface.

**Postprocessing Methods**

The postprocessing methods can be enabled or disabled if required and consist of procedures to further correct and normalize already processed data.

- **Solvent correction**: Solvent correction accounts for differences in bulk shifts between samples due to differences in solvent content (e.g., DMSO) and exclusion of volume. Cycles consisting of sample buffer with varying concentrations of the solvent are used to obtain a calibration curve, which can then be used to correct for variations in solvent content.
Molecular weight adjustment: Molecular weight adjustment normalizes end point data by accounting for the differences in molecular weight between different compounds.

Surface activity adjustment: Surface activity adjustment corrects for systematic signal decrease from a loss of binding capacity of the surface during the course of the experiment. It uses signals from either a repeated positive or negative control or both to adjust for this decrease.

Result Calculation Methods

The result calculation methods include procedures for fitting both steady-state and kinetic models to obtain values for equilibrium dissociation constants and/or kinetic rate constants.

Steady-state fit: The steady-state fit is used to determine equilibrium dissociation constants. A report point value is taken from each sensorgram within a compound titration and plotted versus the compound concentration. A Langmuir binding isotherm is fitted to the resulting saturation curve.

Global binding fit: The global binding fit is a collection of kinetic binding models that can be simultaneously fitted to the association and dissociation data from a titration series of a compound. These models are the 1:1 binding model (Langmuir binding isotherm), a two-state binding model, a heterogeneous ligand model, a bivalent analyte model, a 1:1 dissociation model, and a double exponential decay model.

Additional terms that can be parametrized in the models include refractive index change (when significant bulk shifts remain in the data) and mass transfer limitation (when there is a risk of analyte depletion).14

All reported results comprise the individual fit parameters and their standard errors.

Results and Discussion

Importing Data from Various Instruments

A generic data import application programming interface (API) allows importing data from various instruments by providing a fixed interface between external and internal raw and metadata and their mappings to distinct measurement cycles. Similar to instrument providers’ software, SPR result files are opened within Screener from seconds to a few minutes, where loading time scales linearly with the size of the data set and the complexity of the analyses performed.

The wide range of experimental protocols used in SPR, from single concentration screening to detailed kinetic characterization, is captured by experiment templates. A set of basic templates consists of (1) single concentration screening, (2) concentration–response with steady-state fit, and (3) concentration–response with steady-state and kinetic fit. These basic templates serve to quickly perform the required preprocessing and result calculations needed. The user can then adapt them further as needed to include project-specific processing methods and calculations and save them as a new template tailored for that particular project for reusing. Analyzing a new data set in the same way will subsequently be very fast as such experimental templates fully automate processing—once data are loaded, all results are already computed and the user can focus immediately on the quality of the data and analysis.

Data Visualization and Analysis Capabilities

Inspection of the sensorgrams is an important aspect of data analysis. A sensorgram contains a lot of information, and visual inspection is required in certain cases to increase the quality of the experimental results. For example, artifacts/irregularities, nonspecific binding, curvature, goodness of fit, baseline drifts, and other effects are often easily identified by visual inspection and can be highly informative on the nature of the biophysical interactions occurring at the sensor surface.

Several additional features facilitate the process of SPR data analysis and its quality control further. For example, any measured or calculated result can be plotted versus another by means of one- or two-dimensional plots with conditional coloring based on various conditions (e.g., cycle types or ligand names), providing almost endless ways to visualize the data. Filtering analytes by one-to-many SPR result parameters enables narrowing down large screens to the most interesting analytes very quickly.

Interactive quality control of irregular or bad-looking data works by means of masking, flagging, or removing data points (sensorgrams) per single compound concentration or for complete titration series. The immanent injection spikes and similar small irregularities can be removed by smoothing, resulting in more visually appealing sensorgrams without changing critical information. Furthermore, results from different SPR runs can be opened jointly in the same session.

Enabling situation-specific data visualizations is crucial, for example, to visually check the preprocessing requirements to show concentration-series sensorgram overlays per compound next to each other in the sequence of preprocessing (e.g., raw, time-aligned, baseline-aligned, reference-subtracted, and blank-subtracted sensorgrams). This SPR module facilitates different ways of viewing the data to support the scientist. Situation specific arrangements are generated as one-click actions: windows, plots, and tables can be moved, resized, and arranged. Once a view layout appears...
mature and suitable for a particular application, it can be saved and reused by others. Figures 2 and 3 show two examples of such data views for a screening and kinetics experiment, respectively.

**Reporting SPR Results**

Streamlining the process of data reporting was one of the major drivers for developing a novel software approach for SPR data analysis. There are many factors that help in decision making, but clearly, the possibility to process, filter, and view any results side by side provides a rapid and more effective means to decide which data should be reported. In our SPR module, the user can choose which data to report from any of the different analyses made and also incorporate comments or remarks with the data.

Once automatically reported to our corporate database, SPR results are placed in the context of other experimental information, and then gain their real value for discovery programs. In our lead discovery programs, such SPR results are viewed side by side with other data available in the database, like IC$_{50}$ values, cell data, or in vivo results. This orthogonal analysis across multiple modalities is crucial for most project decisions, for example, on compound progression or follow-up experiments. Also, results from any SPR experiments processed in Screener can be directly viewed side by side with results from other experiments also analyzed in the Screener cross-assay profiling module. This makes it possible to look at sensorgrams and IC$_{50}$ curves side by side, something that is typically not possible in corporate databases, as these are often restricted to numerical values.

Besides these two ways of reporting and visualizing the data, the SPR module also allows the generation of Excel or pdf reports, containing both the sensorgrams and derived results. This gives a fast means to share these reports in
Working with Screener for SPR: A Year of Experience

During the first year of production usage of the Genedata Screener platform with the SPR module at AstraZeneca,16 we applied the software in more than 20 different projects, analyzing the data from tens of thousands of interactions. The key improvements over the legacy solution were as follows:

1. Time saving. The combination of loading data from any experiment on any computer (within a web browser), using experimental templates, the adjustable user interface, data filter functionalities, and automatic data reporting, resulted in an estimated overall 50% time reduction. Time is spent on quality controlling the results instead of processing data (Fig. 1).

2. Increased consistency and visibility. Since all data are analyzed in the same way, the result consistency is improved. Results from different analysis approaches can be compared side by side, and it is possible to set up guidelines or rules for how data are analyzed. The reports, originally thought to be used primarily for electronic lab books, have also been very positively received by the project teams. Visual exemplification of, for example, different kinetic profiles is considered much more illustrative than reviewing only values in a table. An example of such a report can be seen in the supplementary material.

3. Reduced training and infrastructure burden. The quick training of new SPR users is important for organizations with frequent, almost naturally occurring personnel changes, since they only need to learn one evaluation software, as well as the easier maintenance of IT infrastructure, since updates to the SPR module are applied on a server environment and become available to all SPR users immediately on any computer within the network.

The greatest benefits of the SPR module are seen when running larger sets of compounds and when data are analyzed and reported in a similar way every time. An example is when SPR is used for fragment screening or high-throughput screening follow-up, requiring the processing of thousands of compounds per week, followed by hundreds of compounds in dose–response measurements.

When very small sets of analytes are being tested, or when data are not going to be reported or presented directly (such as during assay development or for quick yes/no...
experiments), Screener for SPR offers no considerable advantage over the legacy software.

Furthermore, Screener for SPR is not a stand-alone solution but a module to the Screener platform, most often used in a setting where it supports and integrates analysis of data from multiple/diverse platforms. Hence, its implementation is most valuable in an organization with a critical mass of different screening activities, like medium to larger biotech or pharma companies, to fully benefit from all advantages as illustrated in this paper.

**Future Use**

The SPR module was developed with AstraZeneca R&D in the context of evaluating the binding of small molecules measured in Biacore SPR systems. However, the module is also used for the characterization of large-molecule binding, and will be further enhanced in this direction by future developments like epitope binning. Other improvements to the module include support for new Biacore instruments (Biacore 8K) and the possibility to analyze data generated on other vendors’ instruments that share a data structure similar to that of the Biacore systems.

The ambition for the joint development of the SPR module was to combine the analytical SPR methods previously used at AstraZeneca with the workflow logic, scalability, and systems integration of the Genedata Screener platform to provide a more effective and efficient workflow. The resulting module fulfills the criteria as initially defined (Table 1), and the newly implemented workflow is both simpler and faster (Fig. 1B). Within the global Discovery Sciences department at AstraZeneca, the module is now used for all SPR-dependent projects, resulting in large time savings and an increase in analysis consistency. Genedata Screener for SPR is a fully enabled software solution that has a wide utility for SPR users across a number of different applications.

**Declaration of Conflicting Interests**

The authors declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: All authors are employed by AstraZeneca R&D or Genedata AG.

**Funding**

The authors received no financial support for the research, authorship, and/or publication of this article.

**References**

1. Holdgate, G. A.; Anderson, M.; Edfeldt, F.; et al. Affinity-Based, Biophysical Methods to Detect and Analyze Ligand Binding to Recombinant Proteins: Matching High Information Content with High Throughput. *J. Struct. Biol.* **2010**, *172*, 142–157.

2. Holdgate, G.; Geschwindner, S.; Breeze, A.; et al. Biophysical Methods in Drug Discovery from Small Molecule to Pharmaceutical. *Methods Mol. Biol.* **2013**, *1008*, 327–355.

3. Folmer, R. H. Integrating Biophysics with HTS-Driven Drug Discovery Projects. *Drug Discov. Today* **2016**, *21*, 491–498.

4. Edfeldt, F. N.; Folmer, R. H.; Breeze, A. L. Fragment Screening to Predict Druggability (Ligandability) and Lead Discovery Success. *Drug Discov. Today* **2011**, *16*, 284–287.

5. Dahl, G.; Akerud, T. Pharmacokinetics and the Drug–Target Residence Time Concept. *Drug Discov. Today* **2013**, *18*, 697–707.

6. de Witte, W. E.; Wong, Y. C.; Nederpelt, I.; et al. Mechanistic Models Enable the Rational Use of In Vitro Drug-Target Binding Kinetics for Better Drug Effects in Patients. *Expert Opin. Drug Discov.* **2016**, *11*, 45–63.

7. Genedata. Screening Platform. https://www.genedata.com/news-events/press-releases/detail/news/genedata-announces-astazeneca-standardizes-on-genedata-screener-as-its-global-screening-platform/ (accessed May 31, 2016).

8. Genedata. Genedata Screener. https://www.genedata.com/products/screener/ (accessed May 31, 2016).

9. Frostell-Karlsson, Å.; Remaeus, A.; Roos, H.; et al. Biosensor Analysis of the Interaction between Immobilized Human Serum Albumin and Drug Compounds for Prediction of Human Serum Albumin Binding Levels. *J. Med. Chem.* **2000**, *43*, 1986–1992.

10. Langmuir, I. The Adsorption of Gases on Plane Surfaces of Glass, Mica and Platinum. *J. Am. Chem. Soc.* **1918**, *40*, 1361–1403.

11. Morton, T. A.; Myszka, D. G.; Chaiken, I. M. Interpreting Complex Binding Kinetics from Optical Biosensors: A Comparison of Analysis by Linearization, the Integrated Rate Equation, and Numerical Integration. *Anal. Biochem.* **1995**, *227*, 176–185.

12. Karlsson, R.; Michaelsson, A.; Mattsson, L. Kinetic Analysis of Monoclonal Antibody-Antigen Interactions with a New Biosensor Based Analytical System. *J. Immunol. Methods* **1991**, *145*, 229–240.

13. Oshannessy, D. J.; Brighamburke, M.; Soneson, K.; et al. Determination of Rate and Equilibrium Binding Constants for Macromolecular Interactions Using Surface Plasmon Resonance: Use of Nonlinear Least Squares Analysis Methods. *Anal. Biochem.* **1993**, *212*, 457–468.

14. Myszka, D. G.; He, X.; Dembo, M.; et al. Extending the Range of Rate Constants Available from BIACORE: Interpreting Mass Transport-Influenced Binding Data. *Biophys. J.* **1998**, *75*, 583–594.

15. Genedata. Genedata Screener for Cross-Assay Processing and Compound Profiling. https://www.genedata.com/products/screener/profiling/ (accessed May 9, 2016).

16. Genedata. New Genedata Screener for Surface Plasmon Resonance Co-Developed with AstraZeneca. Announced at 2014 Bio-IT World Conference & Expo. https://www.genedata.com/news-events/press-releases/detail/news/new-genedata-screener-for-surface-plasmon-resonance-co-developed-with-astazeneca-announced-at-2014/ (accessed May 31, 2016).
The following highly-rated podium presentations were delivered at SLAS2016. Stream them at your convenience as another benefit of being a dues-paying member of SLAS.

- **Scott A. Busby**, Novartis Institutes for Biomedical Research, HT-MALDI-MS as a Complete Label-Free Drug Discovery Platform: From Target Characterization Through High-Throughput Screening to Hit Follow Up

- **Jason Cassaday**, Merck, Development of a Platform to Enable Fully Automated Cross-Titration Experiments

- **Tiffany J. Chen**, Cytobank, Inc. and Stanford University, Identifying Druggable Cells: Automated Methods for High-Content Single-Cell Screening

- **Louis Cohen**, Icahn School of Medicine at Mount Sinai, High-Throughput Screening of Metagenomic DNA Libraries

- **Edward Rebar**, Sangamo BioSciences, Genome Engineering with Zinc Finger Nucleases

- **Sindy KY. Tang**, Stanford University, Droplet Microfluidics: Amphiphilic Nanoparticles as Droplet Stabilizers for High-Fidelity and Ultrahigh-Throughput Droplet Assays

- **Peter Tonge**, Stony Brook University, Drug-Target Residence Time: Target Engagement, Target Vulnerability and Predictions of in Vivo Drug Activity

VISIT: [www.slas.org/events/webinars/](http://www.slas.org/events/webinars/) to access these presentations.
The following highly-rated podium presentations were delivered at SLAS2016 this January. Stream them at your convenience as another benefit of being a dues-paying member of SLAS.

- **Scott A. Busby**, Novartis Institutes for Biomedical Research. *HT-MALDI-MS as a Complete Label-Free Drug Discovery Platform: From Target Characterization Through High-Throughput Screening to Hit Follow Up*

- **Jason Cassaday**, Merck. *Development of a Platform to Enable Fully Automated Cross-Titration Experiments*

- **Tiffany J. Chen**, Cytobank, Inc. and Stanford University. *Identifying Druggable Cells: Automated Methods for High-Content Single-Cell Screening*

- **Louis Cohen**, Icahn School of Medicine at Mount Sinai. *High-Throughput Screening of Metagenomic DNA Libraries*

- **Edward Rebar**, Sangamo BioSciences. *Genome Engineering with Zinc Finger Nucleases*

- **Sindy KY. Tang**, Stanford University. *Droplet Microfluidics: Amphiphilic Nanoparticles as Droplet Stabilizers for High-Fidelity and Ultrahigh-Throughput Droplet Assays*

- **Peter Tonge**, Stony Brook University. *Drug-Target Residence Time: Target Engagement, Target Vulnerability and Predictions of in Vivo Drug Activity*

**VISIT: www.slas.org/events/webinars/ to access these presentations.**