

ChemBioNews.com 14.4

BioSeek & CambridgeSoft Infrastructure for Drug Discovery with Integrative Biology Approach

[Evangelos Hytopoulos, Ph.D.](#), [Leon Kao](#), [Lilly Xia](#)

Dr. Hytopoulos, the Director of Scientific and Corporate Informatics, leads the development of information systems, database and analytical approaches for the BioMap™ technology at BioSeek. He has over 12 years of experience in developing and implementin

BIOSEEK IS A HUMAN SYSTEMS biology company focused on the discovery and development of novel anti-inflammatory drugs. The company has developed a unique primary cell-based profiling platform (BioMAP® Systems), which by capturing the complexity of human disease biology in a novel, and highly efficient approach, enables rapid identification and optimization of compounds that are more likely to be safe and effective in patients.

Technology Platform

In BioMAP Systems, combinations of human primary cell types are simultaneously activated, replicating complex cell and pathway interactions normally found in disease physiology. Compounds are profiled by measuring the levels of optimized sets of protein readouts representing the disease biology of interest. Depending on the mechanism of action, each compound will induce changes in expression of protein readouts, giving a specific BioMAP profile. Because of the multitude of pathway and cell interactions that are captured, each BioMAP System is capable of detecting a broad range of compound classes. Combining data from multiple systems gives the technology the unprecedented power to discriminate among closely related compounds. The profiles are stored in the company's database, and analyzed using BioSeek's proprietary statistical and modeling algorithms.

In the discovery process, profiles of newly identified leads are compared to BioMAP profiles of known therapeutics in the database to identify compounds that exhibit desired therapeutic activity and minimal potential for side effects. Even though highly complex, BioMAP Systems are automated and of sufficient throughput to provide rapid discovery of high-quality drug candidates.

Challenge

BioSeek had internally developed a Web-based application for a) assay layout design (template creation), b) storage of raw data values and c) query and export of experimental values. The Web application acts as the interface to an Oracle database that was originally storing only the raw data from the lab. The data represent plate reader measurements (multiple lengths) for each of the wells of a microplate. The layout of the plate depends on the type of experiment, i.e., assay development, research, partner compound screening, etc.. Each well can contain multiple experimental conditions, test agents and potentially different assay readouts (e.g proteins, etc.). The application is able to accommodate any possible combination of assay condition, test agent, assay readout and layout combinations per well.

The original design was biased towards providing an interface that would be able to accommodate maximum flexibility for designing complex development assays. The application had to manage multiple repeats on the same plate (well repeats) as well as multiple repeats across plates. As the company grew and new analysis methodologies were introduced, data transformation and normalization needed to be standardized and presented to the user in the form of BioMAP profiles. New features were constantly added to support the increasing requirements of research. Our group faced a major challenge when high throughput screening was implemented. Three major bottlenecks in the workflow became evident:

With increasing number of assay systems and experimental measurements added to the screening panel (over 14 HTS systems and 120+ readouts) data integrity constraints were slowing down database performance.

Visualization of the profiles for multiple systems simultaneously was not cleanly supported from the database schema.

Design of new HTS templates was cumbersome for large experiments due to the original design of the application that favored flexibility (complex layouts) over ease of designing HTS layouts (i.e. multiple plates with same compound layout but different experimental conditions and readouts).

Addressing the Problem

The solution to the problem required a radical redesign of the whole approach. The original design had grown out of the concept of storing and visualizing raw experimental values and it was not able to adjust to the changing requirements of our data analysis and visualization. A team consisting of the Scientific Computing, Biology and HTS members examined the current and future requirements and prepared a document outlining the features of the new application.

Given the time, cost and functionality requirements it was decided to explore vendor solutions along with an in-house developed solution. The team developed a set of requirements and "nice-to-have" features and invited five vendors to present their solutions. The decision proceeded in three stages. The first stage involved presentations and demonstrations by the vendors to the evaluating team. A matrix was created and its vendor was scored against every requirement based on the demonstrations. Vendors that were clearly lacking any of the requirements were eliminated from consideration.

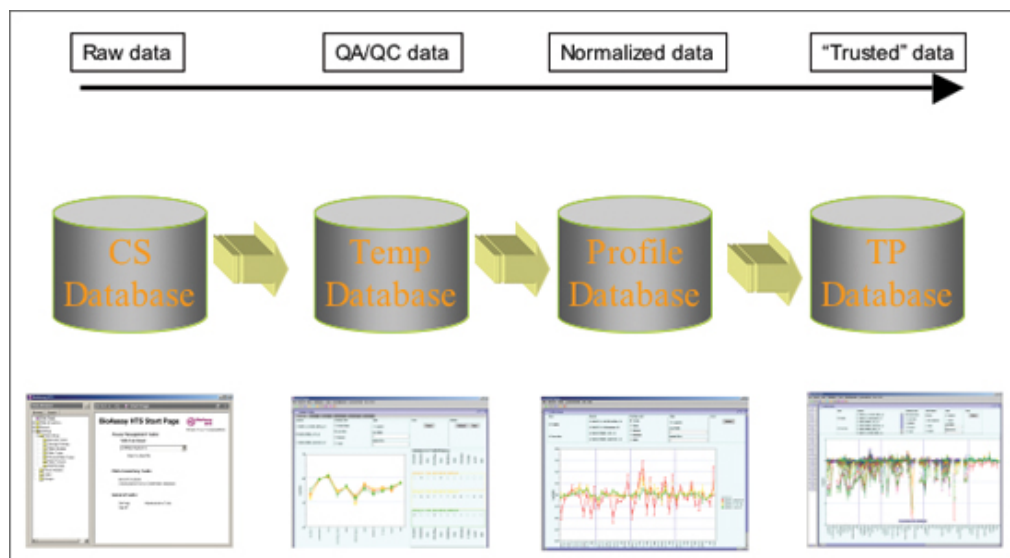


Figure 1: Database architecture for BioMAP data.

Selecting the Solution

The next step involved a demonstration of the solution using a set of our data as the starting point and a given presentation style as the output. Using the matrix that was developed, vendor solutions were judged for ease of use, ease of template creation, data transformation and presentation ability, speed and of course cost. The solutions presented to us had, as expected, strengths and weaknesses but a few of them seemed to better fit our needs. Using our ranking system, we selected Cambridge-Soft's BioAssay as our lead candidate for final consideration.

Our internal development for data analysis and visualization proceeded rapidly while the external solution investigation was taking place. By the time we selected BioAssay as our final candidate, an early version of the visualization software and analysis platform was prototyped and tested inside the company. The application was not at that moment accessing the database, but was working out of Excel files. This was sufficient for the users to give us feedback. The challenge was to decide how to split the database development burden between internal development and BioAssay.

Setting up a Successful Pilot

This question was a highly technical challenge and could not be answered by a test case scenario. Our estimates for a fully internally developed solution, given our resources, was for a nine month development time to an alpha version. CambridgeSoft's solution had to cut this time in half to make it financially competitive. The team decided to proceed with a pilot project, whose aim was to test the ability of BioAssay to integrate easily and seamlessly with our internally developed platform that was already in progress. In collaboration with the CambridgeSoft team, we developed and agreed upon a pilot proposal focused on two primary goals:

- ease of designing HTS templates for big experiments
- transforming raw data into BioMAP profiles

The pilot program was expected to be completed within a month and provide working versions of HTS templates as well as a roadmap for the integration between the BioAssay schema and our internally designed schema for dealing with BioMAP profiles. Based on the experience that we gained by examining many solutions, we concluded that it would be a more flexible solution to separate the raw data database from the profiles database. This proved to be a key decision in the success of the project. BioAssay is used to upload and store the data for each assay point where the views produced by the program are used as the access point for our application in transforming the data into the profile data.

At the end of the pilot, a number of HTS experiment templates were available for use, as well as a first version of the integration between the BioAssay views and our application. Key users actively participating in the pilot were able to extend these templates to new layouts with the help of our IT personnel, and our developer (Leon Kao) very quickly completed the integration between the two databases using a home grown View Management System. Some VB programming and a control vocabulary for compounds and genes allowed BioAssay templates also to integrate directly the existing internal database of chemicals and genes.

The solution was first deployed to the HTS group and proved to be a big success. The Biology group followed and currently the system is fully functional and used by everybody in the company. Data from the old database deemed to be important were also transformed and moved to the new solution.

Figure 1, on the previous page, shows a high level abstraction of the current work flow. HTS data (raw data) obtained in the lab are uploaded into BioAssay and stored. Raw data are transformed to profiles and "pushed" into the "Temp Profiles" database. These profiles are then subjected to a QA/QC process. Any problems identified with the data are flagged and the users can reject profiles that do not meet quality criteria. This step also helps the scientists to identify potential problems with the assays and/or execution of an experiment and take the appropriate steps to correct it.

Once the experiment has been deemed successful and the problem data have been flagged, the profiles are "pushed" into the "Profiles" database. These profiles contain normalized data that as a group describe the unique activity of the test agent (compound, genetic perturbation or combination). The "Profiles" database contains all the normalized data that have ever been run through our HT screening. These profiles can be retrieved and visualized through BioMAPViewer™, the home grown application for query, analysis and visualization of our proprietary data. This application, developed in JAVA, allows the user to retrieve profiles at several levels of

granularity. Data can be grouped under Client, Project, Experiment and/or Repeat categories.

Although every test agent is repeated multiple times on a plate, our internal statistical analysis also dictates the number of experiment repeats that we need to perform to obtain data with required confidence limits.

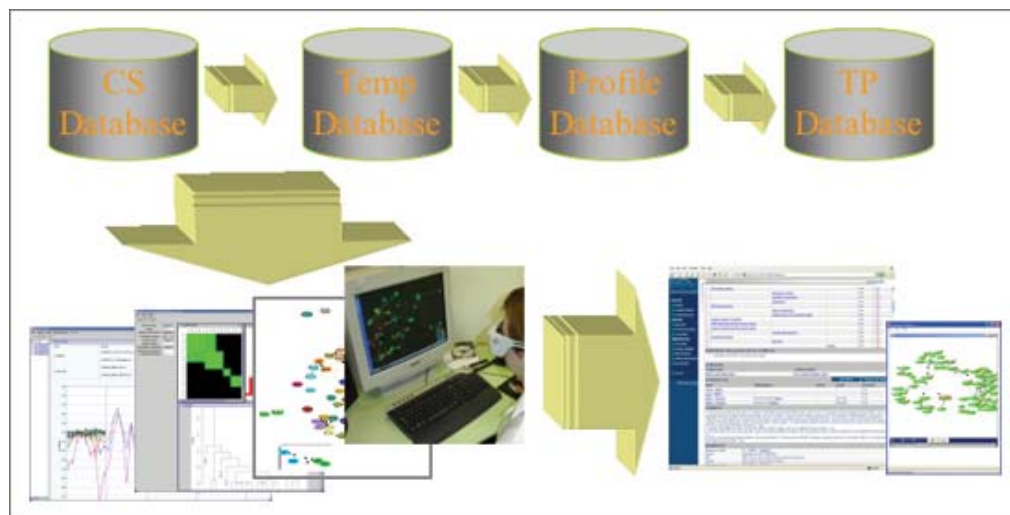


Figure 2: IT infrastructure for data capture analysis and visualization of BioMAP profiles.


The profiles from multiple repeats are then used to create a "trusted profile". The "trusted profiles" are updated every night and are stored in a new database ("Trusted Profiles" database). Since these profiles represent an activity profile of a specific agent obtained from multiple experimental repeats and characterized by specific confidence level, the notion of repeats and experiments ceases to be of importance. Thus the "Trusted Profile" database provides grouping of the profiles under a Client category that needs to be maintained so confidentiality can be assured.

The trusted profiles are available for searches and analysis using our proprietary algorithms as well as commercially available analysis packages (Figure 2). Scientists can identify compounds with similar functional profiles, create networks of genes and drugs and annotate them with useful information like class and target information. The information is also visualized using 2D and/or 3D visualization techniques. Knowledge acquired from our analysis is integrated into a knowledge management system called Cognia Molecular (Cognia Inc.). Researchers can use all the information available (proprietary and public sources) to generate new hypotheses that can be tested in the lab.

Ongoing development of the system guarantees that the corporate goals are fully supported through a very efficient and nimble IT infrastructure.

Summary

In summary, CambridgeSoft's products were instrumental in providing a powerful and flexible environment for supporting drug discovery and collaborations. The combination of a carefully designed evaluation process and the excellent collaboration with CambridgeSoft allowed BioSeek to develop and deploy a production-ready data management system in record time. We believe that this combined effort represents an example of a very successful project. Given the time and resource constraints that IT organizations operate under, well designed collaborations between a vendor and a customer are necessary for continuous success of both parties.

For more information about CambridgeSoft Enterprise solutions, please send email to solutions@cambridgesoft.com, or call 1 617 588-9300 in America, in Europe call 00 800 875 20000, or in Japan call 0120-731-800. 

CambridgeSoft.com

ChemBioNews.com

ChemFinder.com

SciStore.com

All trademarks are the property of their respective holders.
©2007 CambridgeSoft Corporation. All Rights Reserved. [Privacy Statement](#)