Infosys®
POWERED BY INTELLECT
DRIVEN BY VALUES

Life Sciences
CASE STUDY

High Throughput Screening for Small Molecules Binding to Target Active Site

Predictive identification of ligands through data mining

Summary

Identification of a new ligand that binds to an active site in a protein is an experimentally and computationally challenging process of screening through a large database of chemical compounds. It is envisaged that the process of predictive identification of ligands would help to quickly shortlist potential compounds, thereby reducing the time for lead identification. Ligand hits are arrived by searching for the best scoring ligands based on force-field values of the constituent atoms present in the ligand molecule.

Background

The ligand molecule binds to the active site of a protein, subject to various physical, chemical and conformational properties. Ligands are classified based on their binding to the active site of a family of proteins. For identifying a new ligand that binds to the protein active site, a researcher has to potentially screen through a huge database of chemical compounds. This means the compounds that have to be selected should be screened by techniques like visualization and then experimented upon for determining the binding capability of the compound. This is a time consuming process and the probability of failure is very high.

The Solution

This problem can be addressed by following a process of predictive identification of potential ligands through a high throughput matching and analysis process.

Enormous amounts of data is available in the literature regarding families of small molecules (chemicals/ligands) that bind to a particular target. The literature also contains qualitative & quantitative data about their systematic SAR (Structure-Activity Relationship). The knowledge hidden in such data can be used as a predictive aid in the ligand identification process.

The ligand molecules map into the active site of the protein, subject to various physical, chemical and conformational properties. The atomic level mapping of properties such as hybridization, nature of the chemical atom, partial charge and *van der Waals* radius are derived from the standard libraries for force-field values of the constituent atoms of the ligand molecule.

Patterns can be identified within the family of ligands in terms of these properties. Profiles are generated using a suitable modeling technique (an appropriately chosen statistical method) for the compounds of the family. The profile is then used to search the chemical database. The resultant compounds are ranked based on a scoring technique and given to the user. These are the best matching compounds from the chemical database that have a high probability to bind with the active site in the protein.

The MDL's ACD (Available Chemicals Directory) - a structure-searchable database, provides information for over 350,000 chemicals and over

About Infosys

Infosys, a world leader in consulting and information technology services, partners with Global 2000 companies to provide business consulting, systems integration, application development and product engineering services.

Infosys enables its clients to fully leverage technology for business transformation. Clients leverage Infosys' Global Delivery Model to achieve higher quality, rapid time-to-market and high-RoI solutions.

Life Sciences Practice

The Life Sciences Practice helps companies in pharmaceutical, biotechnology and related industries leverage information technology to drive productivity across their business processes. Through custom designed solutions, the LS Practice partners clients in realizing value from technology investments in discovery, development, manufacturing, sales and marketing, core IT and other enabling functions.

The Practice brings together specialists in software engineering, quality processes, quantitative methods, drug development, manufacturing, regulation, life science informatics and related areas. A key thrust is to deploy in-depth domain expertise on concept-to-implementation engagements, through both in-house capabilities and alliances.

Solution Portfolio

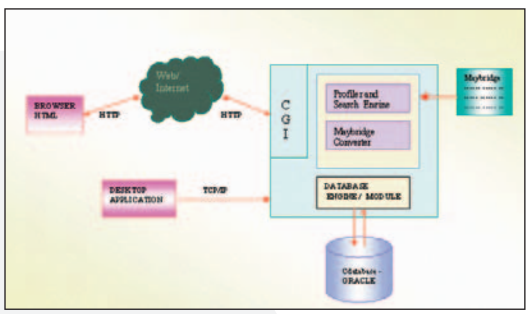
The solution offerings of the Life Sciences Practice include the following

- Life science informatics
- Technology consulting
- Business process consulting
- Application development and life cycle support
- Software product co-development
- Data integration and analytics
- Regulatory compliance
- Knowledge management and collaborative R&D
- Document / content management
- Package implementation
- Pre-clinical and clinical trial management
- Laboratory management
- Manufacturing solutions
- Sales force automation and CRM solutions

one million research-grade and bulk chemical products. Similarly, the drug discovery compounds producer, Maybridge's database of organic collection consists of over 56,000 organic compounds for HTS (High Throughput Screening) lead generation for lead optimization. At such volumes, there is a need for screening and selecting compounds based on valid criteria to determine the list of compounds with a high feasibility of binding.

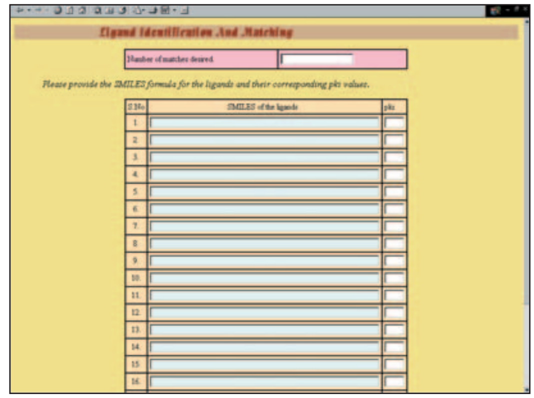
Predictive identification of the ligands helps the researcher to quickly shortlist potential compounds, thereby reducing the time consumed during the lead identification process.

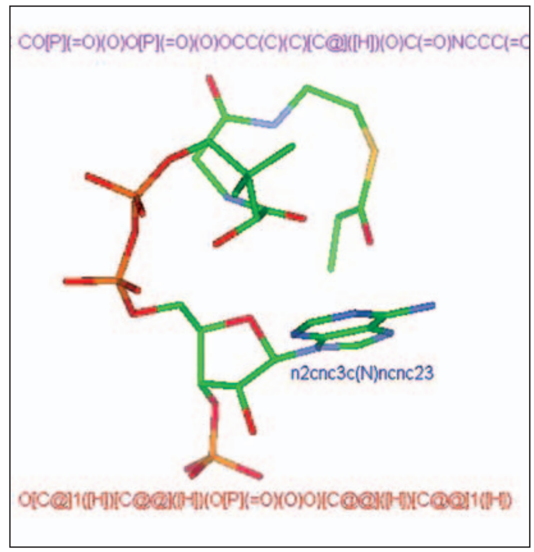
Architecture Overview and Technologies



Platform	:	Sun Solaris 5.8
Web server	:	Apache 1.3.19
Scripting	:	Perl 5.0
Database	:	Oracle 8i

Solution Snapshots





Assimilating chemical moieties from SMILES format

For more information about Infosys' Life Sciences Practice, please contact:

US (East)

Sriram Chandrasekaran
sriram_c@infosys.com

Ashu Tandon
ashu_tandon@infosys.com

US (West)

Sanjay Deshpande
sanjay_deshpande@infosys.com

Europe

Sudarshan Roongta
sudarshan_roongta@infosys.com

Asia Pacific

Thillai Rajan Annamalai
thillai_annamalai@infosys.com