



Bilateral Research and Industrial Development Enhancing
and Integrating Grid Enabled Technologies

**Transcontinental Workflows for Simulation, Engineering
and Design on Heterogeneous Grid Platforms**

DRUG DISCOVERY IN BRIDGE



BRIDGE has received research funding of the European Commission under the FP6 Information Society Technologies Programme (IST), contract number IST-2006-045609.

Maximum community contribution to project: 1.7 Mio Euro

Project start: 1 January 2007

Duration: 24 months

Partners involved: 13

Type of project: Specific Targeted Research Project

Project website:

<http://www.bridge-grid.eu>

The project is co-ordinated by the Fraunhofer Institute SCAI in Sankt Augustin, Germany.



Cover: Different ligands (small molecules) docked in the binding pocket of a protein involved in the metabolism of a malaria virus.

About BRIDGE

BRIDGE addresses bilateral research and industrial development between EU and China for enhancing and integrating Grid-enabled technologies. The project has attracted major industrial and academic partners from Europe and China. The consortium consists of 13 partners, seven from Europe and six from China. Among them are industrial partners, research institutions and universities.

BRIDGE aims at demonstrating the benefits of Grid technology for international cooperation, in particular between Europe and China. BRIDGE addresses technical issues, which result from the far distance of the collaboration partners as well as from the conflicting goal of intense collaboration and protection of intellectual property rights.

BRIDGE develops an interoperability interface between heterogeneous infrastructures based in Europe (GRIA) and in China (GOS). It allows in particular the execution of distributed workflows and access to distributed data repositories. Remote access to specific analysis services will allow their controlled usage for product and process development.

Virtualization of Drug Discovery in BRIDGE

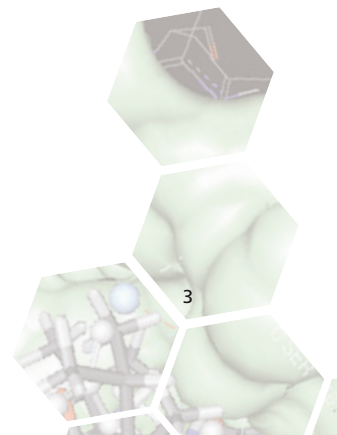
Our ability to find possible drugs by screening a virtual library with billions of compounds is hindered by the task's intrinsic difficulty, as it often involves enormous computational demands and the management and analysis of large amounts of data in a collaborative environment. The goal of the task is to find small molecules that can modulate the actions of proteins and the interactions between them. The key challenge is transforming the available data into useful knowledge for drug discovery.

As a »proof of concept« the BRIDGE Pharma partners have chosen a Grid-based Virtual Screening application integrating different docking tools as scientific experiment for identifying novel potential lead candidate drugs against Avian Flu and Malaria.

With Pharma research involving many stages ranging from target identification to selection through selection and optimization to pre-clinical and clinical trials it can take years to traverse the drug discovery pipeline and costs can rise into billions of dollars. Here Grid computing offers an opportunity to address the main computing challenge such as how to find the drug candidate in less time and with less money. Both at algorithmic and infrastructure level BRIDGE will help to access the Chinese and vice versa the European market for many existing products and services based on Grid technology.

Collaboration Requirements

Drug discovery is a collaborative activity. Traversing the whole drug discovery pipeline requires a continuous flow of information and feedback between all stages. Researchers working on any one stage require access to new results, methods and information published by other groups.



Virtual Screening and Protein Docking

Protein docking aims at the positioning of small molecules (potential drugs) into the binding site of a target protein (associated with a specific disease). For this, one has to screen a large number of chemical compounds/ligands against target receptors.

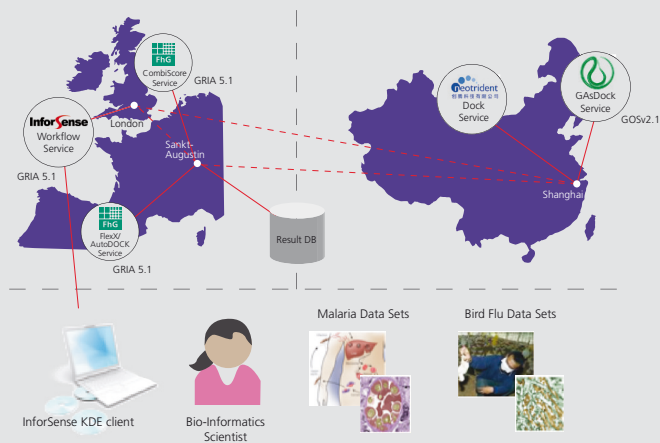
The most important application of docking techniques in Pharma research is virtual screening of large databases of available chemicals to select likely drug candidates.

Existing protein docking tools (such as FlexX, DOCK by and AutoDock, GAsDock...) operate on different principles... and the docking results vary depending on the specific tool or algorithm used.

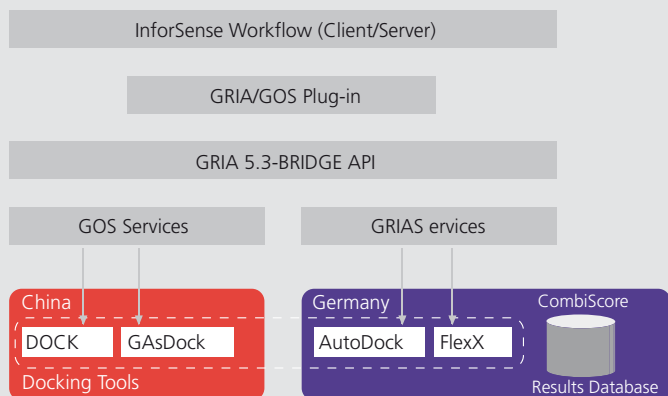
Virtual Drug Discovery

Starting point for the BRIDGE Pharma Application were the interconnected infrastructure and middleware already developed in the EU-funded SIMDAT project. BRIDGE extends the SIMDAT Pharma data integration to compute intensive applications for high throughput virtual screening.

The ultimate goal is to formalize the scientific application scenario and evaluate the scientific data and results via deploying an integrated platform for Grid-based composition of Virtual Screening tools, with the four docking tools (FlexX, AutoDock, DOCK and GAsDock) integrated.

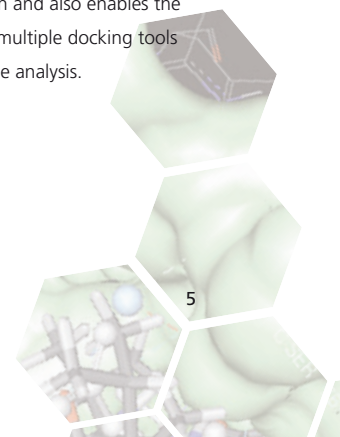


The DockFlow concept is based on enabling end users to seamlessly access and integrate multiple protein docking tools within the same analysis. A key challenge addressed within DockFlow is enabling the interoperability between protein docking tools executing remotely on HPC resources using a heterogeneous grid fabric. Within BRIDGE a prototype of such a platform has been deployed using a mix of four protein docking tools; with two protein docking tools (FlexX and AutoDock) executing in Europe on European grid middleware and two tools (DOCK and GAsDock) executing in China on Chinese middleware.



DockFlow

DockFlow aims at constructing an interoperable Grid-based virtual screening platform based on integration of four protein docking tools executing on a mix of European and Chinese grid middleware. This is based on accessing and executing two tools (FlexX and Autdock) in Europe on European grid middleware (GRIA) and accessing executing two tools (DOCK and GAsDock) in China on Chinese middleware (GOS). Access to the individual tools is made available through a workflow interface that hides the details of the underlying Grid implementation and also enables the integration of multiple docking tools within the same analysis.



Application scenario set-up

The current set-up consists of an Oracle Database hosted at Fraunhofer SCAL in Germany that can be accessed by other partners in the project.

The production cluster at Fraunhofer is used to host the GRIA services, namely AutoDock and FlexX. Sun Grid Engine is used to schedule the execution of the docking tools amongst the clusters along with a License Server to allow execution of FlexX.

A Grid-based Virtual Screening Platform for Scientists

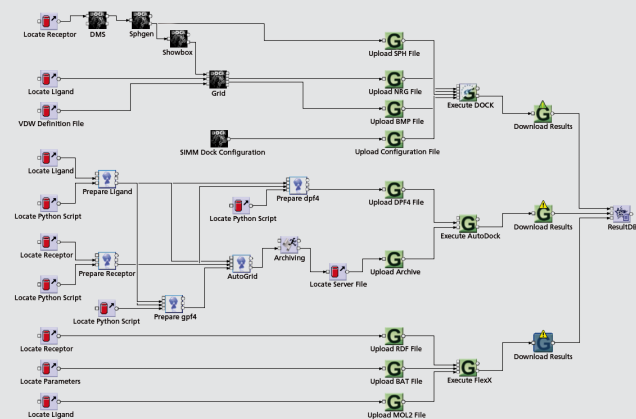
To realise complex virtual screening workflows the different docking tools (FlexX, AutoDOCK, DOCK, GAsDock) are assembled in workflows and seamlessly integrated. Two data sets were chosen for the docking experiments: The Malaria data set is provided by the European, the Bird Flu data sets by the Chinese partners. The evaluation will include natural product compound development.

Target	PDB	FlexX	Autodock	Dock	GasDock	Total
Avian flu	2hu0	9,357	9,384	9,989	81	9,990
1	2hu4	9,359	9,833	7,623	81	9,990
Malaria	1q4j_a	51	51	42	51	51
2	1q4j_b	51	51	38	51	51

BRIDGE test scenarios

In each case, the workflows start by accessing target and ligand databases, selecting the required records and applying pre-processing to them to put them in formats suitable for input to the protein docking tools, which typically are executed using remote high-performance resources. The required protein docking tool is then invoked using its native interface and the results collected and the output scored, visualized and analyzed by the researchers and stored in a database.

Interoperable Workflow Solutions



The DockFlow workflow implemented as an InforSense workflow.

For each docking tool, the workflow orchestrates the pre-processing steps and invocation of the remote applications before collecting the results in a single database and running the Combscore scoring tool. The docking tools are accessed by components built using the GRIA/GOS Adaptive API. This allows seamless integration with other InforSense components which are used to handle the pre-processing stages, such as the Command Line components used for invoking local Python scripts and bespoke components that wrap applications that prepare the data for the docking tools.

Workflow implementation

AutoDock and FlexX docking workflows are based on deploying them as a GRIA service. They were integrated within the InforSense workflow server using the standard InforSense plug-in and GRIA components enabling the service to be discovered and integrated into the InforSense platform without any issues.

Both DOCK and GAsDock share the same invocation interfaces and pre-processing steps. For the first prototype, the implementation of the DOCK and GAsDock workflows is based on separating preprocessing and execution steps. The actual DOCK/GAsDock execution is handled as a GOS service with a GRIA wrapper and is run in China.

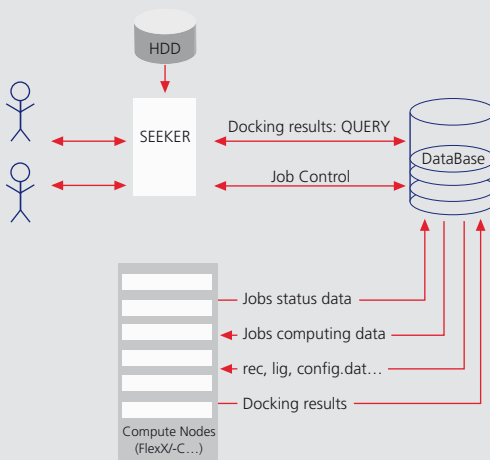
Docking Database (DDB)

A key pre-requisite for enabling the use of multiple docking tools within the same analysis is the ability to combine the results of different docking tools and to have them available at the same place. In DockFlow the docking results of DOCK, GAsDock, FlexX and AutoDock are therefore saved in one Result Database. The Docking Database (DDB) is an integrated virtual docking environment developed by BioSolveIT that allows a detailed analysis of massive amounts of docking results. The aim is to provide easy access to the wealth of data of a virtual screening run that is otherwise buried in piles of log-files that nobody ever looks at. In addition to storing the input and output data of a protein docking runs, it comes with a variety of analyzing tools like the definition of new scoring functions or the application of several filters.

DBB: An Integrated Virtual Docking Environment

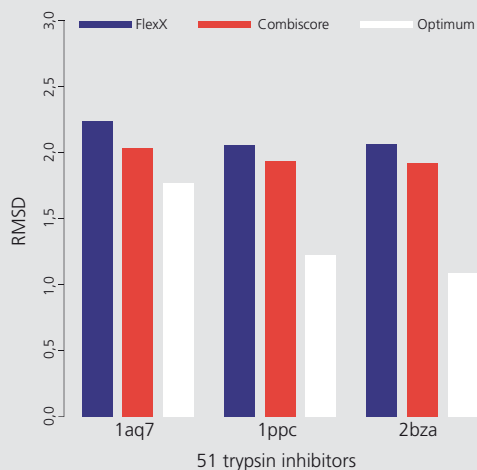
The results of the docking tools are stored and integrated using an extended version of the BioSolveIT database (DDB).

The comparison of the output results of multiple tools is enabled through the implementation of a novel scoring tool, Combiscore. The development of such a combined scoring scheme (Combiscore) enhances significantly the impact of the DockFlow activity as it would neutralize the limitations inherent to independently running docking programs.



BioSolveIT (DDB)

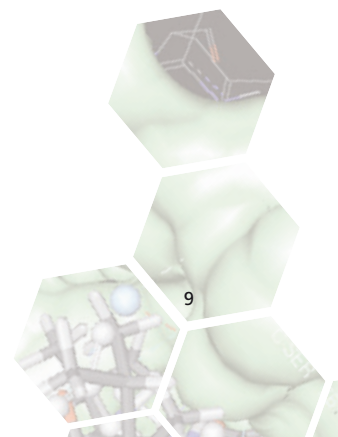
The ultimate test of docking workflows is prospective prediction of novel ligands. A pragmatic approach for routine testing of algorithmic developments is to use experimentally-observed poses of selected ligands and enrichment of known actives as performance evaluation criteria.



Median RMSD (root mean square deviation) between docked structure and native one for the 51 trypsin inhibitors obtained by FlexX alone (blue) and Combiscore (red). For comparison the minimal found RMSD is also shown (white).

Evaluation of Docking Accuracy

Traditionally, the docking accuracy is evaluated by comparing the docked structure of the ligand to the one obtained from the experimentally determined crystal structure. The lower the RMSD between these two structures is, the better the docking accuracy.

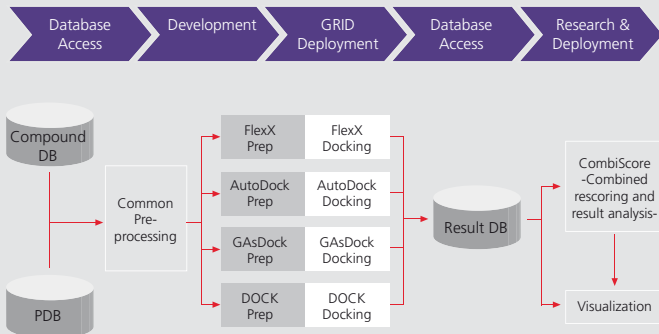


Achievements

The developments of BRIDGE in the pharmaceutical virtual drug discovery have shown how PharmaGrids can open new perspectives for cheaper and faster drug development by virtualizing large parts of the drug development process, starting from target identification and validation via compound screening and optimization towards the prediction of ADME (absorption, distribution, metabolism and excretion) and toxicity properties.

More Efficient Identification of Novel Potential Lead Candidate Drugs

In DockFlow the four different docking tools FlexX, AutoDock, DOCK and GAsDock were successfully integrated. The tools were executed as remote docking services at geographically distributed locations in Europe and China from a common workflow. Multiple algorithms for solving the same problem in a branched workflow were used to address the problem of diversity of results produced by different docking tools. The storage of the outcomes of these tools in a common docking database supports a comparison of the results and a faster and more convenient way of result analysis in such large scale collaborative screening experiments by providing a common platform with sorting, filtering and visualizing facilities.



Within BRIDGE a number of valuable resources in terms of tools and data sets has been created. The Pharma tools and resources will be used directly by the project partners, e.g. within the WISDOM project consortium in Europe and the Drug Discovery Grid consortium in China.

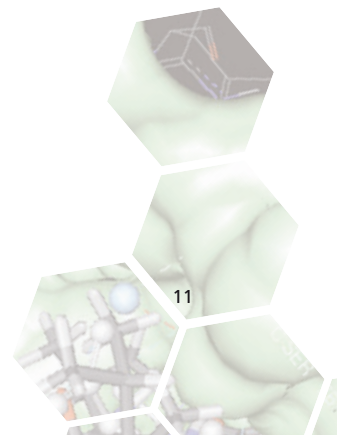
Furthermore, the BRIDGE Pharma partners have successfully engaged collaborators from other parts of the world in using the developed infrastructure, including researchers in Egypt, which is one of the countries that have been affected by the Bird Flu epidemic (Up to April 2009, Egypt had 26 human fatalities from Bird Flu, the third highest fatality rate in the world at that date). The deployment at Nile University in Cairo, Egypt included installation of the workflow tools and portal front ends, which enable Egyptian researchers to access services in Europe and China. Furthermore, the deployment included deploying Dock and Auto-Dock GRIA services in Egypt, which can be used by European and Chinese Researchers.

Achievements

Within BRIDGE the problem of analyzing the data produced in a distributed virtual screening system has been successfully addressed.

The application of different algorithmic approaches allows the selection of the most reliable solution.

It has successfully been demonstrated on a test set that the combinatorial approach Combiscore improves docking accuracy compared to the use of a single tool.



CONTACT

Fraunhofer Institute for
Algorithms and Scientific
Computing SCAI

Yvonne Havertz and Gilbert Kalb
Schloss Birlinghoven
53754 Sankt Augustin
Germany

Phone +49 2241 14-2244
www.bridge-grid.eu

Pharma Partners

InforSense Ltd.

- Workpackage leader
- Workflow tools
- Workflow-based interoperability

CNIC

(with support from Beihang University & IT Innovation)

- GOS infrastructure and interfaces
- GRIA infrastructure and interfaces

Neotrident

- Coordinating case studies
- Validation and usability

Fraunhofer SCAI

- Algorithmic interoperability
- Scientific Case Study (Malaria)
- Support for Autodock & FlexX
- Subcontractor: BioSolveIT

SIMM

- Scientific Case study (Bird Flu)
- Support for Dock and GAsDock

