



## Lifecycle centred Drug Design and Implications for Discovery Outcomes

### The *in-silico* Experimental Lifecycle.

Workflows and workflow development are central to the *in-silico* research process undertaken by Chimatica Ltd. drug design scientists, but are also part of a wider experimental method. Workflows, and the resources they process, exist in a wider context of scientific data, scientific protocol and study management, all of which draw upon and contribute to an accumulated pool of knowledge, knowhow, results and interpretations that can be shared between scientists inside and outside an organisation.

One scheme describes the whole experimental *in-silico* lifecycle. During design a workflow may be tentatively prototyped and evolved; during publication a data result is likely to be analysed.

### myGrid: A framework for supporting *in silico* experiments.

The myGrid project has been running since late 2001. Originally it was a UK EPSRC-funded e-Science pilot project made up of a consortium of UK Universities and institutes, and supported by nine industrial partners including GSK and IBM.

Originally developed to support *in-silico* experiments in biology, myGrid has researched and developed open source high-level service-based middleware, drawing on innovative technologies from the semantic web community. This has developed into a comprehensive loosely-coupled toolkit of core components for forming, executing, managing and sharing data intensive

discovery experiments. The components are intended to be adopted in a “pick and mix” way by developers and tool builders to produce end applications.

Scientists and service providers develop and run experiments via the Taverna workbench which allows the assembling, adapting and running of workflows. Workflows that execute remote or local web services and Java applications are the chief mechanism for forming experiments. Any web service can be incorporated – there is no restriction on the type of science.

The project’s Scufll workflow language is a user-oriented abstraction over general graph languages hiding the details of service invocation and control flow. Users can also configure support for fault management and service failover. The software can be freely downloaded and has been used for building discovery workflows for investigations into Williams-Beuren Syndrome and Grave’s Disease by collaborating life scientists.

### Realising a Lifecycle centred Drug Design Platform

Chimatica Ltd. has adopted the myGrid toolkit and Taverna workbench. In order to achieve the computational power required to tackle large-scale drug design challenges, such as multiple receptor-ligand docking. The toolkit was securely linked to processor farms in the UK (NW-GRID) and the US through additional grid middleware. Docking algorithms and internal databases were wrapped and exposed as services. The

services and data were composed into workflows.

A workflow was developed that chains together incremental precisions of Schrodinger's Glide. Using a variant of just this single workflow, they were able to simultaneously train and screen drug-like and lead-like sets totalling over one million molecules against all available X-ray structures of twenty-five clinical targets in a matter of months.

As a direct result of these knowledge management systems, we are able to update the collections "on the fly" as new chemotypes become available in the global pool and are transferred to our central repository. Duplicate screening events are excluded and only new ligands are screened and the sets are updated automatically. Additional scoring functions can be added or removed as required. These screening sets are now being used as starting points for discovery research by partners and clients.

## Conclusions

This project has demonstrated that our scientists have:

**A.** Overcome the conventional problems of chaining together internal and

external database searches, algorithms from a variety of vendors, an array of analytical tools through systematic application of experimental protocols captured as workflows.

**B.** A means of composing and orchestrating any combination of algorithms and data at any required scale.

**C.** Removed the need for complex scripts to overcome incompatibilities, or by manually cutting and pasting between Web/ GUI interfaces.

**D.** A platform that collects and records information systematically and easily during workflow enactment – elements that are vital for validation and verification of outcomes, building client project expertise and the knowledge base through re-use and iteration throughout the discovery timeline.

**E.** Systems that support the scientific process of managing, sharing and reusing the results, their provenance, and the methods used to generate them.

**F.** Spent more time in analysis than performing the workload of the experiment, adding to the quality of the results interpretation.

The image displays a workflow diagram and a screenshot of the Chimica Grid interface. The workflow diagram on the left shows a sequence of steps: 'Worknode' leading to 'Glide docking VS', 'VS\_SP', 'Glide docking SP', 'SP\_SP', 'Glide docking XP', and 'Output File'. The screenshot on the right shows the Chimica Grid web interface with a table of jobs and a workflow diagram below it.

JOB ID	START TIME	FINISH TIME	STATUS	DESCRIPTION	USER	GROUP
1	2006-12-12 10:00:00	2006-12-12 10:00:00	Completed	Job 1	user	group
2	2006-12-12 10:00:00	2006-12-12 10:00:00	Completed	Job 2	user	group
3	2006-12-12 10:00:00	2006-12-12 10:00:00	Completed	Job 3	user	group
4	2006-12-12 10:00:00	2006-12-12 10:00:00	Completed	Job 4	user	group
5	2006-12-12 10:00:00	2006-12-12 10:00:00	Completed	Job 5	user	group
6	2006-12-12 10:00:00	2006-12-12 10:00:00	Completed	Job 6	user	group
7	2006-12-12 10:00:00	2006-12-12 10:00:00	Completed	Job 7	user	group
8	2006-12-12 10:00:00	2006-12-12 10:00:00	Completed	Job 8	user	group
9	2006-12-12 10:00:00	2006-12-12 10:00:00	Completed	Job 9	user	group
10	2006-12-12 10:00:00	2006-12-12 10:00:00	Completed	Job 10	user	group