



## SHARE

### USE CASE SCENARIO FOR INNOVATIVE MEDICINE

---

Document ID:	<b>SHARE-D5.1b_v1.doc</b>
Date:	<b>25/01/07</b>
Authors:	<b>V. Breton, N. Jacq, I. Blanquer, V. Hernandez</b>
Activity:	<b>WP5: Applications</b>
Document status:	<b>FINAL</b>
Document link:	<b><a href="http://eu-egee.org/deliverables.html">http://eu-egee.org/deliverables.html</a></b>
Confidentiality:	<b>Public</b>
Keywords:	

---

**Abstract:** This document presents an analysis of the status of research and development of advanced ICT on Innovative Medicine. The document outlines the structure of a general use-case representative of the subject and which can be tackled from a HealthGrid approach. The document discusses the medical problem, the data sources and providers, the current procedure and its limitation and outlines the technologies and requirements needed for achieving the desired aims, and the benefits expected.



### Document Log

Issue	Date	Comment	Author
0.9	19/1/07	First release	Vincent Breton, Nicolas Jacq
1	31/1/07	Editing, formatting and insertion of common sections, executive summary.	I. Blanquer, V. Hernández

### Document Change Record

Issue	Item	Reason for Change
0.9	19/1/07	Writing of the first version
1	1/2/07	Adaptation to format. Insertion of Executive summary, abstract, conclusions and use case sections



---

## CONTENT

<b>1. INTRODUCTION.....</b>	<b>4</b>
1.1. PURPOSE.....	4
1.2. APPLICATION AREA.....	4
1.3. REFERENCES.....	4
1.4. TERMINOLOGY.....	5
<b>2. EXECUTIVE SUMMARY.....</b>	<b>6</b>
<b>3. THE USE CASE SCENARIOS.....</b>	<b>7</b>
<b>4. USE CASE 1: INNOVATIVE MEDICINE .....</b>	<b>8</b>
4.1. MEDICAL PROBLEM.....	8
4.2. USERS.....	10
4.3. DATA AND DATA PROVIDERS .....	11
4.4. CURRENT PROCEDURE.....	11
4.5. LIMITATIONS .....	12
4.6. AIM AND BENEFITS.....	12
4.6.1. <i>Grids offer unprecedented opportunities for resource sharing and collaboration.....</i>	<i>13</i>
4.6.2. <i>Grids open exciting perspectives for handling information flows.....</i>	<i>13</i>
4.6.3. <i>Grids provide the resources to speed up the execution of time-consuming software.....</i>	<i>13</i>
4.7. TECHNICAL REQUIREMENTS.....	14
4.8. OTHER CONSTRAINTS .....	15
<b>5. CONCLUSION .....</b>	<b>16</b>

## 1. INTRODUCTION

### 1.1. PURPOSE

The purpose of the document is to present an analysis of the status, needs, requirements and trends in the area of Innovative Medicine. This document presents a general use-case scenario in Innovative Medicine that reflects a wide application community and discusses the suitability and current trends in using Grids. This document has a corresponding version for Epidemiology identified under the code D51.a. This document is self-contained.

This document is intended for both users (academic or private Healthcare centres and Molecular Biology and Biochemistry researchers) and developers (technologists and legal experts) to define the major hurdles that must be undertaken for the adoption of Grids. More basic information about HealthGrids, is provided in the deliverable D3.1 “Healthgrids Framework” and on the HealthGrid White Paper [3].

This document is written in preparation for a roadmap towards the adoption of Grid technology in Health and Life Sciences. The merged analysis of this document along with D51.b and deliverables D3.2, D3.3 (for the current status and bottlenecks of technological and security aspects) and D4.2, D4.3 (for the current status and bottlenecks of ethical, legal, social and economical aspects) will result on the final roadmap.

All deliverables can be downloaded from the SHARE web-site at [www.eu-share.org](http://www.eu-share.org).

### 1.2. APPLICATION AREA

The document is intended for internal and external use. It will be used as a dissemination tool for the Share project.

### 1.3. REFERENCES

Most of the content of the document was extracted from Nicolas Jacq PhD thesis [2] and from the HealthGrid white paper [3].

- [1] L-M. Birkholtz, O. Bastien, G. Wells, D. Grando, F. Joubert, V. Kasam, M. Zimmermann, P. Ortet, N. Jacq, N. Saidani, S. Hofmann-Apitius, M. Hofmann-Apitius, V. Breton, A.I Louw and E. Marechal, Integration and mining of malaria molecular, functional and pharmacological data: how far are we from a chemogenomic knowledge space?, *Malaria Journal*, Vol. 5, (2006) 110.
- [2] N. Jacq, <http://members.healthgrid.org/~njacq/>
- [3] V. Breton, K. Dean and T. Solomonides, editors on behalf of the Healthgrid White Paper collaboration, “The Healthgrid White Paper”, Proceedings of Healthgrid conference, IOS Press, Vol 112, 2005
- [4] Doman, T.N., et al., Molecular docking and high throughput screening for novel inhibitors of protein tyrosine phosphatase 1B. *J. Med. Chem.* 45 2213–2221 (2002).

## 1.4. TERMINOLOGY

### Glossary

Term	Definition
Innovative Medicine	A treatment or therapy of empirical benefit that is yet outside the mainstream of conventional medicine
QSAR	Quantitative Structure Activity Relationship is a quantitative correlation process of chemical structure with well-defined methods.
Drug Discovery	Drug discovery is the process by which drugs are discovered and/or designed.
Screening	Large-scale study that covers a whole population.
Target	A cellular or genetic molecule which is believed to be associated with a desired change in the behavior of diseased cells and on which drugs usually act
In Silico	Computer-based simulation of a process.
Docking	A research technique for predicting whether one molecule will bind to another, usually a protein



## 2. EXECUTIVE SUMMARY

This deliverable presents a use case concerning Innovative Medicine and analyses the current status and trends that could be related to the introduction of HealthGrid Technologies. The Pharmaceutical R&D enterprise presents unique challenges for Information Technologists and Computer Scientists. The diversity and complexity of the information required to arrive at well-founded decisions based on both scientific and business criteria is remarkable and well-recognised in the industry.

In this use case, the main users (academic or private Molecular Biology and Biochemistry researchers, pharmaceutical companies), gather the data from different medical and biological sources (such as molecular, cellular, tissue and population as well as from molecular biology and chemistry), integrate them and execute advanced *in-silico* simulation. Drug discovery is the process by which drugs are discovered and/or designed. Drug candidates are inputs to the drug development process.

Recent progress in genomics, transcriptomics, proteomics, high throughput screening, combinatorial chemistry, molecular biology and pharmacogenomics has radically changed the traditional physiology-based approach to drug discovery where the organism is seen as a black box. *In silico* drug discovery contributes to increasing biological system knowledge. The efficiency gains of such an integrated knowledge system could correspond to save 35% costs, or about US\$300 million, and 15% time, or two years of development time per drug.

Current efforts within the pharmaceutical industry are directed at reducing the time and costs for drug development.

The limited adoption of virtual screening comes also from the wrong structures of some protein models, the inaccuracy of scoring functions, the lack of flexibility of the target protein in docking methods, the lack of an automated virtual screening pipeline and the high computation time required screening millions of compounds. In this sense, and despite of the difficulties, applying *in silico* techniques before, after and in parallel with high throughput screening improves the overall process quality in terms of information content and success rate.

These problems can be tackled using Grid Technologies. On a grid, data can be stored anywhere and still be transparently accessed by any authorized user. The computing resources of a grid are also shared and can be mobilized on demand, mapping complex workflows. Moreover, the intrinsic capability of Grids of structuring the sharing of data and resources and the collaboration is also crucial.

So, from the analysis of the problems in Innovative Medicine, Grid constitutes an enabling technology that can solve many technological issues and foster scientific development especially on Drug Discovery.

### 3. THE USE CASE SCENARIOS

A use-case scenario represents a significant example in the application area related. This example must be generic enough to be representative but specific enough to enable a clear and accurate analysis. The use-case must be related to a real application case, or at least to a reasonable foreseeable application. It does not need to involve totally computers currently and must be sensible for the use of Grids. A use-case mainly describes how a set of actors solve a problem using some data and following a specified procedure requiring some technical means, which produce some benefits and have some limitations. These what, how, who and where must be clearly defined. The purpose of defining and analysing the use case scenario is to cross-validate the requirements and needs identified in the technical roadmaps.

The use-case scenarios must consider all the actors, data, processes, limitations and benefits, and present them in a clear way. Thus, for each case the following data are selected:

- **Medical Problem.** This should describe the medical problem at wide scope, indicating the main implications in health management and outlining its importance.
- **Users.** The use-case could involve different users depending on the interest for the results. Medical users, industry users or researchers could share the same use-case but considering different targets or input sources. The maturity in the use of ICT, the size of the collective and their relevance should be outlined to assess the importance of the use-case.
- **Aim and Benefits.** The benefits will be clearly identified for all the users involved, as well as the side-effects. Desired aim and potential non-reached benefits should also be outlined.
- **Data and Data Providers.** The data sources that are necessary for the use-case should be identified. Important factors in this part are the availability of the data, the restriction in its usage, privacy issues, representativeness and coverage, quality of the data and ICT maturity.
- **Current Procedure.** The means used to solve the medical problem must be described. This means will surely involve computer-based processing, but possibly not at large or for a reduced part of the problem. Actors involved in the process and identified in the “users” section must be mentioned.
- **Limitations.** The current procedure will be bound to some restrictions and limitations that reduce its impact and relevance. These limitations can be technical (lack of performance or resources) or scientific (lack of knowledge) due to the procedure itself.
- **Technical Requirements.** The technical requirements needed for the current procedure should be identified. Technical requirements to achieve the expected aim should be also identified if possible.
- **Other Constraints.** Along with the technical constraints, there could be legal, ethical or economical constraints that must be faced to achieve the aim.

The use-case scenarios selected are epidemiology and innovative medicine. The case of epidemiology is described in this deliverable.

## 4. USE CASE 1: INNOVATIVE MEDICINE

This section describes, according to the definition of a use-case provided before, the use case scenario 1, related to innovative medicine.

### 4.1. MEDICAL PROBLEM

The Pharmaceutical R&D enterprise presents unique challenges for Information Technologists and Computer Scientists [3]. The diversity and complexity of the information required to arrive at well-founded decisions based on both scientific and business criteria is remarkable and well-recognised in the industry. The decisions can form the basis for multi-year multi-person multi-millions of Euro investments and can create new scientific territory and intellectual property. Thus all aspects of managing, sharing and understanding this information is critical to the R&D process and subject to substantial investment and exploration of new informatics approaches.

Pharma R&D information includes large variety of scientific data as well as sources of critical organisational information such as project and financial management data and competitor intelligence information. This data takes some fairly unique formats as well. Examples are images, models, sequences, full text scientific reports, records of prescriptions and physician encounter re-imbursements. These sources of information consist of internal proprietary, external commercial and open-source data.

The problems range from knowledge-representation and integration, to distributed systems search and access control, to data mining and knowledge management, to real-time modelling and simulations, to algorithm development and computational complexity.

Grid technology holds out the promise of more effective means to manage information and enhance knowledge-based processes in just the sort of environment that is well established in Pharma R&D.

A pharmaceutical Grid should be a shared *in silico* resource to guarantee and preserve knowledge in the areas of discovery, development, manufacturing, marketing and sales of new drug therapies and cover three dimensions:

- a resource that provides extremely large CPU power to perform computing intense tasks in a transparent way by means of an automated job submission and distribution facility
- a resource that provides transparent and secure access to storage and archiving of large amounts of data in an automated and self-organized mode
- a resource that connects, analyses and structures data and information in a transparent mode according to pre-defined rules (science or business process based)

Pharmaceutical Grids open the perspective of cheaper and faster drug development. Pharmaceutical Grids should enable parallel processes in drug development, away from the traditional approach where target discovery, target validation, lead discovery, lead optimization and transition to development take on average 12 years. These parallel processes would take advantage of *in silico* science platforms for target identification and validation, compounds screening and optimization, clinical trials simulation for detection of deficiencies in drug absorption, distribution, metabolism and elimination.

For competitive and intellectual property protection reasons, pharmaceutical Grids will predominantly be private enterprise-wide internal Grids with strict control and standards. At least this will likely be the case in the near-term as more and more R&D organisations explore and become comfortable with this technology and its potential.

To reduce the scope of the use case described in this document, we are going to focus on drug discovery.

### In silico drug discovery

Drug discovery is the process by which drugs are discovered and/or designed [2]. Drug candidates are inputs to the drug development process. Drug development manages preclinical safety studies and clinical phases, with clinical trials. Registration and delivery are the last steps of the full process. Reducing the research timeline in the discovery stage and having enhanced information about the leads are key priorities for pharmaceutical companies worldwide. Collaborations with academic laboratories and small biotechnology or pharmaceutical companies are crucial, mainly in exploratory research, then in the chemistry stage and progressively less during the drug development phases

The drug discovery goal is to find new molecules that bind with specific macromolecules, known to play a key role in the disease evolution, in a manner that changes their function for the benefit of life.

Recent progress in genomics, transcriptomics, proteomics, high throughput screening, combinatorial chemistry, molecular biology and pharmacogenomics has radically changed the traditional physiology-based approach to drug discovery where the organism is seen as a black box. The approach is now to understand how disease and infection are controlled at the molecular and physiological level and to target specific entities based on this knowledge.

In silico drug discovery is one of the most promising strategies to speed-up the drug discovery process. It is important to know and control the *in silico* process, that is described below.

Figure 1 shows the different phases of a drug discovery process with their approximate duration, their success rate and the corresponding in silico contributions.

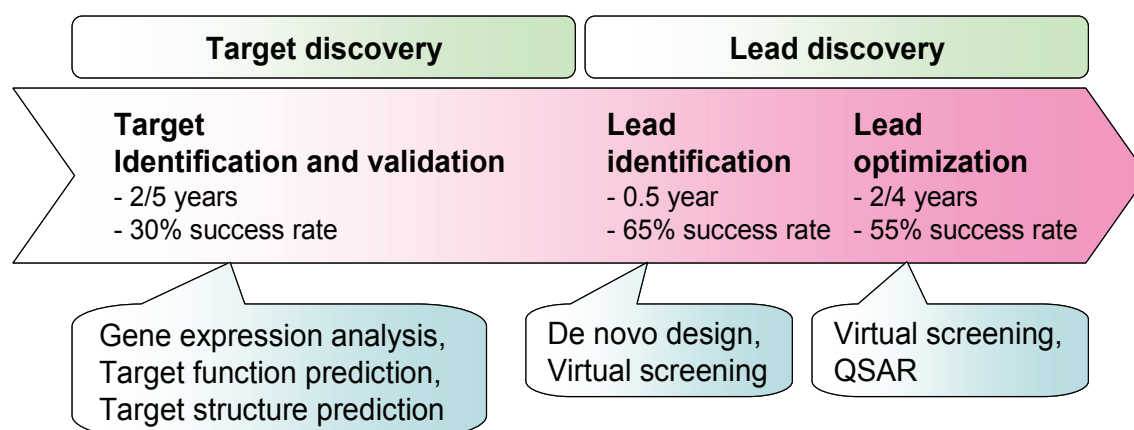


Figure 1: Representation of the different phases of the drug discovery process with their duration, their success rate and the corresponding *in silico* contributions.



A target is a cellular or genetic molecule which is believed to be associated with a desired change in the behavior of diseased cells and on which drugs usually act. The target identification and validation aims to isolate and select it. *In silico* drug discovery contributes to the target discovery by gene expression analysis, target function prediction and target three-dimensional (3D) structure prediction for post-processing.

To identify a lead compound, a substance affecting the target selected in a drug-like way, two different *in silico* pipelines can be used which speed up the process and reduce costs avoiding useless *in vitro* tests: the *de novo* design and virtual screening. *De novo* design builds iteratively a compound from the structure of a protein active site. Virtual screening selects *in silico* the best compound from a molecule database.

Lead optimization addresses the development from the most promising lead compounds to a safe and effective drug. Instead of expensive and longer *in vitro* and *in vivo* tests, evaluation of the basic chemical properties can be achieved by virtual screening and Quantitative Structure Activity Relationship (QSAR). QSAR is a quantitative correlation process of chemical structure with well-defined methods, such as optimization for pharmaceutical properties (Absorption, Distribution, Metabolism, Excretion and Toxicity (ADMET)) or efficacy against the target organism.

*In silico* drug discovery contributes to increasing biological system knowledge, to managing data in a collaboration space, to speeding up analysis and consequently increasing the low success rate of the traditional “wet” approach. The efficiency gains of such an integrated knowledge system could correspond to save 35% costs, or about US\$300 million, and 15% time, or two years of development time per drug.

Nevertheless, in spite of increasing levels of investment in *in silico* techniques, there is a steady decline in the number of new molecules that enter clinical development and reach the market. Many factors have changed over the past 10 years, particularly the domination of the target-based drug discovery paradigm, favoring screening and rational drug discovery programs. A new approach aims to integrate rational drug discovery with a strong physiology and disease focus.

## 4.2. USERS

In this section, we are going to describe the users involved in the drug discovery process. These users can belong to academic or private structures. Their affiliation may impact the level of confidentiality of the data they produce and exchange on the grid.

We identified the following actors:

- Healthcare centres collecting epidemiological information on the disease
- Molecular biology and biochemistry public research labs producing new molecular data on the relevant organisms and able to identify new targets out of their research activities
- Bioinformatics labs producing new tools to enable biological research on the relevant organisms
- biochemical research laboratories able to prepare targets for virtual screening and to analyze the results
- chemo-informatics and chemo-genomics labs able to provide new tools for target



identification, hit selection and biochemical data analysis

- biological laboratories able to do *in vitro* and *in vivo* testing of the best hits coming out of virtual screening
- Information technology developers to develop knowledge management tools
- computing and computer science centres

Later on in the drug development process, healthcare centres will be involved in clinical tests. Finally, several organizations will be involved in selling and distributing the new drugs (pharmaceutical labs, healthcare administrations, healthcare professionals and non profit organizations,...)

### 4.3. DATA AND DATA PROVIDERS

Data involved in the drug discovery process come from multiple sources:

- Medical data include molecular data (ex. genomics, proteomics), cellular data (ex. pathways), tissue data (ex. cancer types, wound healing, personal data (ex. EHR), population ( ex. epidemiology)
- molecular biology data on the relevant organisms (virus, parasites, bacteria) are produced at different expression levels (genomics, proteomics, transcriptomics, metabolomics) and stored in databases. Data providers are molecular biology research labs.
- Chemical data include potentially active compounds stored in private or public databases.

Tools are also needed to handle all these data. These tools include bioinformatics algorithms, databases, chemo-informatics and knowledge management software.

### 4.4. CURRENT PROCEDURE

The traditional approach to drug discovery includes target discovery, target validation, lead discovery, lead optimization and transition and relies on expensive and long *in vitro* and *in vivo* tests. At the end, the drug candidates are inputs to the drug development process. Drug development manages preclinical safety studies and clinical phases, with clinical trials. Registration and delivery are the last steps of the full process. Drug discovery and development represent a complex multi-phase (12-15 years) and multi million-dollar process (at least \$800 million).

There are currently more than 200 major pharmaceutical companies. As in some other industries, economic pressures are forcing pharmaceutical companies toward greater efficiency. Only 1 New Chemical Entity (drug candidate obtained by the drug discovery process) in 10,000 becomes a product after preclinical evaluations and clinical trials. Biopharmaceutical properties such as oral bioavailability and formulation issues are responsible for about 39% of failures, whereas toxicity constitutes about 21%. These factors are as important as lack of efficacy, which is responsible for about 29% of failures. Consequently biopharmaceutical properties and toxicity factors must be taken account as soon as possible in the drug discovery process.



Reducing the research timeline in the discovery stage and having enhanced information about the leads are key priorities for pharmaceutical companies worldwide. Collaborations with academic laboratories and small biotechnology or pharmaceutical companies are crucial, mainly in exploratory research, then in the chemistry stage and progressively less during the drug development phases.

#### 4.5. LIMITATIONS

Current efforts within the pharmaceutical industry are directed at reducing the time and costs for drug development. Within many groups, virtual screening remains marginal. The reason for this is that they have invested in facilities for high throughput screening. Sophisticated logistics and dedicated manpower allow 1 million compounds to be screened in less than two months. Consequently, the use of *in silico* screening to reduce the number of compounds docked *in vitro* is still limited. The limited adoption of virtual screening comes also from the wrong structures of some protein models, the inaccuracy of scoring functions, the lack of flexibility of the target protein in docking methods, the lack of an automated virtual screening pipeline and the high computation time required screening millions of compounds.

Despite these difficulties, applying *in silico* techniques before, after and in parallel with high throughput screening improves the overall process quality in terms of information content and success rate. For instance, Doman [4] compared the traditional and virtual screening performances for the protein tyrosin phosphatase 1B. The virtual screening hit rate reached 34.8% while the hit rate of high throughput screening was only 0.021%. Selected molecules had better IC<sub>50</sub> and drug-like quality.

#### 4.6. AIM AND BENEFITS

Developing *in silico* drug discovery for neglected and emerging infectious diseases requires a robust infrastructure and relevant services to support distributed resource sharing. Resources are here defined as: computing, storage and network; as well as data, knowledge, software and workflow; but also instruments and sensors; and finally people and organization.

The grid is a new Information Technology that can provide these resources. A grid is the combination of networked resources and the corresponding management software, which provides services for the user. Grid technology provides the collaborative Information Technology environment to enable the combination between life science research, field work, health systems, pharmaceutical industries and infrastructure. It proposes a new paradigm for the collection and analysis of distributed information where data are no longer centralized in one single repository. On a grid, data can be stored anywhere and still be transparently accessed by any authorized user. The computing resources of a grid are also shared and can be mobilized on demand.

The grid added value in the development of *in silico* drug discovery for neglected and emerging infectious diseases has multiple dimensions:

- grids offer unprecedented opportunities for resource sharing and collaboration;
- grids open exciting perspectives for handling information flows;
- grids provide the resources to speed up the execution of time-consuming software.

#### **4.6.1. Grids offer unprecedented opportunities for resource sharing and collaboration.**

Grid allows

- the sharing of resources in a cross-organizational collaboration space between the pharmaceutical industry and academic research institutions, and between developed and least developed countries,
- the creation of as a virtual laboratory for the different actors, increasing cooperation and communication between partners,
- the mobilizing of resources routinely or in an emergency,
- the sharing of diverse, complex, large and distributed information for collaborative exploration and mutual benefit,
- the use of new Information Technology such as large databases or time-consuming software,
- the optimal exploitation of resources by taking advantage of spare computing cycles or by maximizing the use of high performance computing platforms usage,
- the reduction of hardware costs.

#### **4.6.2. Grids open exciting perspectives for handling information flows.**

Grid allows

- the deployment of services for healthcare and research centres in endemic regions,
- the deployment of infrastructures to collect data and improve disease surveillance and monitoring,
- the building of knowledge space with genomics and medical information (epidemiology, status of clinical tests, drug resistances, etc.),
- access to relevant data, periodically updated data bases and publications,
- the federation of regional or international databases for disease study and monitoring of vector control, clinical trials and drug delivery,
- the provision of transparent and secure access to storage and the archiving of large amounts of data in an automated and self-organized fashion,
- connection, analysis and structuring of data and information in a transparent mode according to pre-defined rules (science or business process based).

#### **4.6.3. Grids provide the resources to speed up the execution of time-consuming software.**

Grid allows

- access to large computing resources for *in silico* drug discovery, data analysis and mathematical modelling,
- the application of high performance computing to new areas,
- the production of additional or more accurate analyses,
- the facilitation of the exchange of tools and workflows between scientists,
- the performance of computing intense tasks in a transparent way by means of an

automated job submission and distribution facility,

- access to services and resources 24 hours a day,
- the running of the same job on many platforms across different sites,
- access to computing resources by a single efficient path.

Grids are unique tools for collecting and sharing information, networking experts, mobilizing resources routinely or in an emergency. Grid is thus an appropriate environment to develop *in silico* drug discovery.

Figure 2 extracted from [1] illustrate the organization and treatment of genomic, post-genomic and chemical information for the prediction and characterization of target and drug candidates for malaria on a HealthGrid. The same picture holds for any other disease.

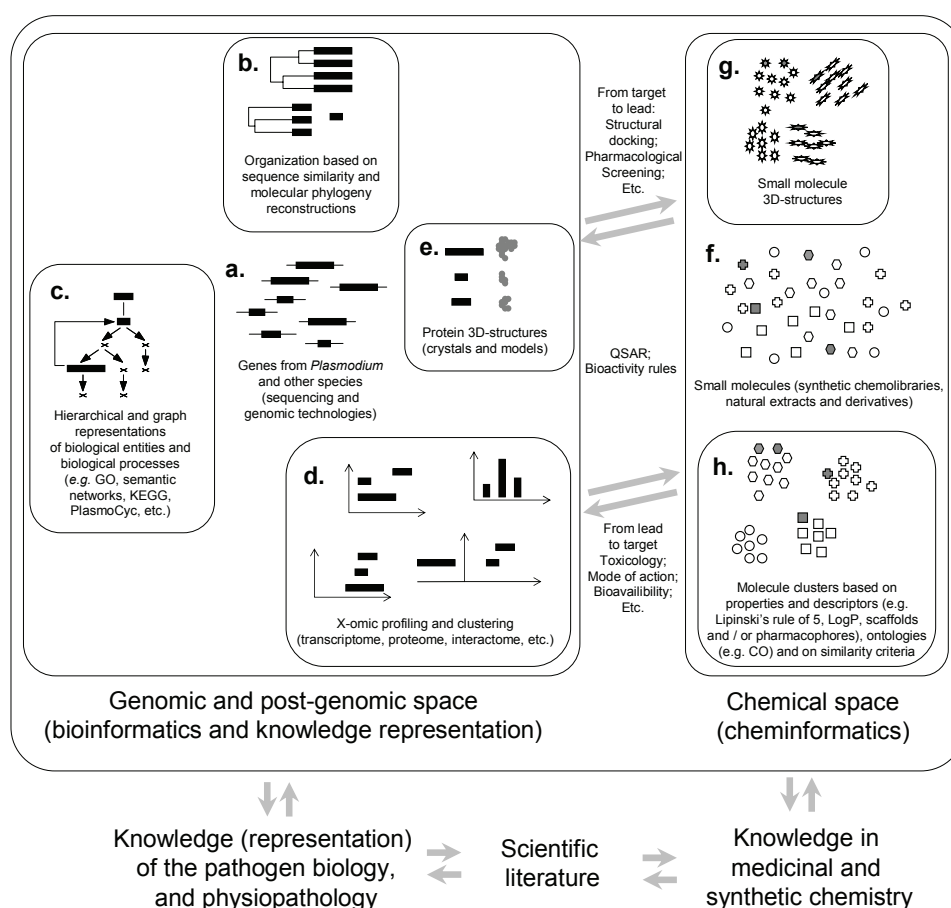


Figure 2 from [1] : organization and treatment of genomic, post-genomic and chemical information for the prediction and characterization of malaria target and drug candidates.

## 4.7. TECHNICAL REQUIREMENTS

Reducing the research time and cost in the discovery stage and enhancing information about the leads are key priorities for pharmaceutical companies worldwide. To achieve this goal, *in silico* drug discovery must meet the following requirements:

- The *in silico* drug discovery process includes the management of a large variety and quantity of scientific data. For example: images, sequences, models, databases. Data integration is thus a challenge to increase knowledge discovery but also to ease the complex workflow. This implies data format standardization, dataflow definition in a distributed system, infrastructure and software providers for data storage, services for data and meta-data registration, data manipulation and database updates.
- The *in silico* drug discovery process also includes the management of a large variety and quantity of software. Software integration is another challenge to build efficient and complex workflows and to ease data management and data mining. Software can be provided in a distributed environment such as a web server on the Internet. Different experts are absolutely necessary to maintain and update software and workflows to propose new methods or pipelines, to use remote services, exploit outputs, and finally to propose compounds for assay. A software workflow will assist the scientist and the decision-maker in organizing their work in a flexible manner, and in delivering the information and knowledge to the organization.
- Deploying intensive computing is a challenge for *in silico* drug discovery. For instance, computing 1 million docking probabilities or modelling 1,000 compounds on one target protein requires in the order of a few TFlops during one day. Very large computing resources are also needed to describe accurately protein structure models by computational methods based on all-atom physics-based force fields including implicit solution. Computing power is also required for bioinformatics resource centres where server access is saturated by the large number of short tasks requested by users.
- Joining the new Information Technologies with life science to enable *in silico* drug discovery requires strong remote collaboration between different public and private experts when addressing neglected and emerging infectious diseases. It also involves strong sharing of resources: data and knowledge, software and workflow, and infrastructures such as computing, storage and networks. The collaboration space needs experts to maintain the resources. Having tools and data accessible to everyone in collaboration requires intuitive interfaces that need to be maintained. These interfaces reduce the development time of new methods. They also help the integration of data and software from *in silico* drug discovery but also from experimental processes.
- Security is a key challenge for pharmaceutical industries but also for academic institutes in most cases. Effective protection of intellectual properties and sensitive information requires, for instance, authentication of users from different institutions, mechanisms for management of user accounts and privileges and support for resource owners to implement and enforce access control policies.

In summary, the main requirements to develop *in silico* drug discovery are data and software integration, intensive computing deployment, remote collaboration and resources sharing, and of course security. Thus, there is a need for a powerful and secured environment sharing and integrating remote resources such as tools, data, computing and storage.

#### 4.8. OTHER CONSTRAINTS

One of the main challenges for pharmaceutical laboratories to share information with public labs is the level of security offered by the grid.



## 5. CONCLUSION

In conclusion, this document has defined the Innovative Medicine case study to be considered for the analysis with respect to the Grid technical requirements identified. The main conclusions of this analysis will validate the trends defined on deliverables D3.2, D3.3 (for the current status and bottlenecks of technological and security aspects) and D4.2, D4.3 (for the current status and bottlenecks of ethical, legal, social and economical aspects).

The technical, ethical, legal, social and economical aims that are outlined on these deliverables should be sufficient to cover the needs and limitations of the application areas defined in this deliverable and deliverable D5.1.a (Use Case Scenario for Epidemiology), and should consider:

- Data management. Tools are needed to handle large-scale heterogeneous data, providing an efficient way to semantically organised the knowledge. Infrastructures and healthcare services must be deployed to securely gather this distributed information.
- Security. Higher levels of privacy management must be implemented due to the impact on economics of the research activities.
- Processing. Large scale processing requires reliability and high-performance, with strong support to complex workflows and transparent management of jobs.

From the consolidation of deliverables D3.2, D3.3, D4.2, D4.3, D5.1a and D5.1a, the final roadmap will be prepared.