Drug discovery on the grid: the WISDOM initiative… and beyond

V. Breton

CNRS-IN2P3, LPC Clermont-Ferrand

Credit: A. Da Costa, V. Kasam

Table of content

- Goals
- Materials and methods
- Results
- Issues encountered
- Perspectives

Introduction

- **WISDOM stands for World-wide In Silico Docking On Malaria**
- **Main objective: use grids to foster R&D on neglected diseases**
- **Driving idea: grid-enabled drug discovery should enable non for profit public – private partnership at reduced cost**

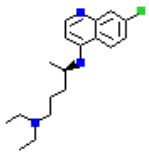
Grid-enabled virtual screening

Millions of potential drugs to test against interesting proteins!

~~High Throughput Screening~~
~~1-10$/compound, several hours~~

Too costly for neglected disease!

Compounds:

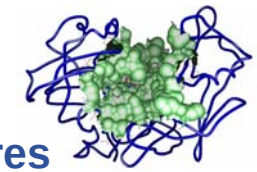


ZINC: 4.3M

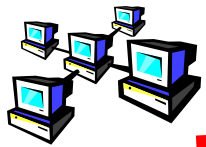
Chembridge: 500 000

Molecular docking (FlexX, Autodock)
~1 to 15 minutes

Targets:



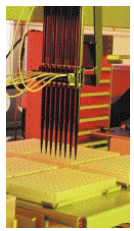
PDB: 3D structures



Data challenge on EGEE
~ 2 to 30 days on ~5000 computers

Cheap and fast!

Selection of the best hits



Hits screening using assays performed on living cells

Leads

Clinical testing

Drug

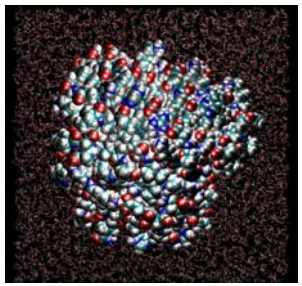
Materials and methods: virtual screening workflow

FLEXX
AUTODOCK

Millions

Molecular docking

AMBER

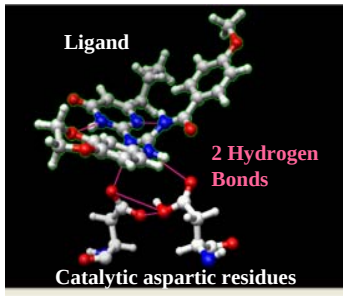


5000

Molecular dynamics

Re-ranking
MMPBSA-GBSA

CHIMERA



180

Complex visualization

30

***In vitro* tests**

WET LABORATORY

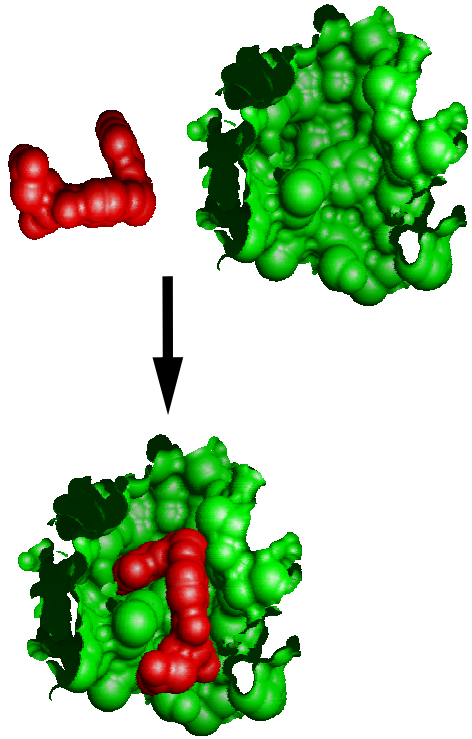
10

***In vivo* tests**

Credit: A. Da Costa

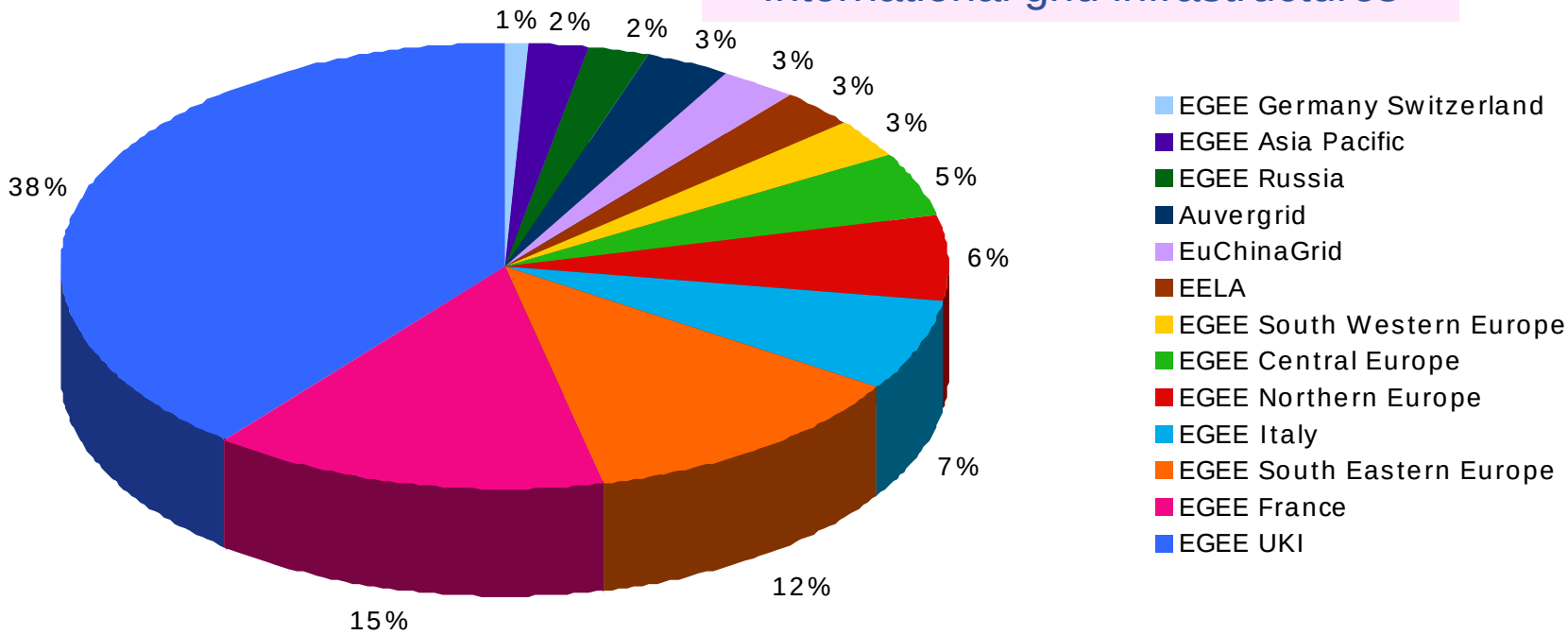
Docking

- **Targets : structures publicly available on Protein Data Base**
  - Re-refinement of the structures on the Grid (Embrace - EGEE)
- **Docking software packages**
  - Open source: Autodock
  - Licensed: FlexX
- **Public database of drug-like ligands**
  - Chembridge (~ 500.000 compounds)
  - Zinc (> 4 Millions)



Grid performances for docking

Significant contributions from several International grid infrastructures



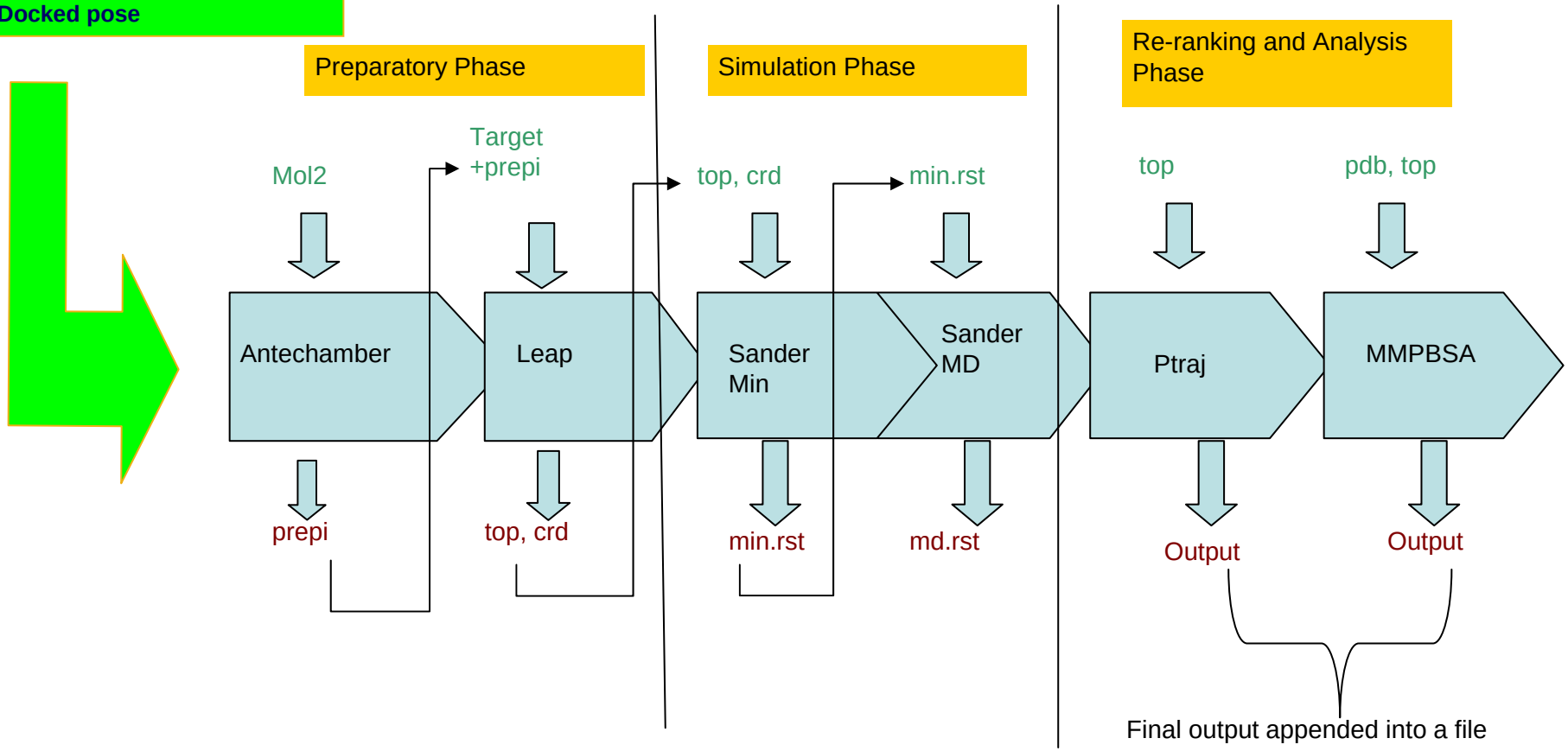
Over 420 CPU years in 10 weeks to dock 4 malaria targets in 2006
A record throughput of 100.000 docked compounds per hour

MD refinement using Amber

Best hits from docking step based on:

- **Docking energy**
- **Docked pose**

For one complete simulation, all necessary steps are embedded in one single script.



☐A. Ferrari, G. Degliesposti, M. Sgobba, G. Rastelli. Validation of an automated procedure for the prediction of relative free energies of binding on a set of aldose reductase inhibitors. Bioorganic & Medicinal Chemistry. 2007. In Press.

Grid Performances for MD

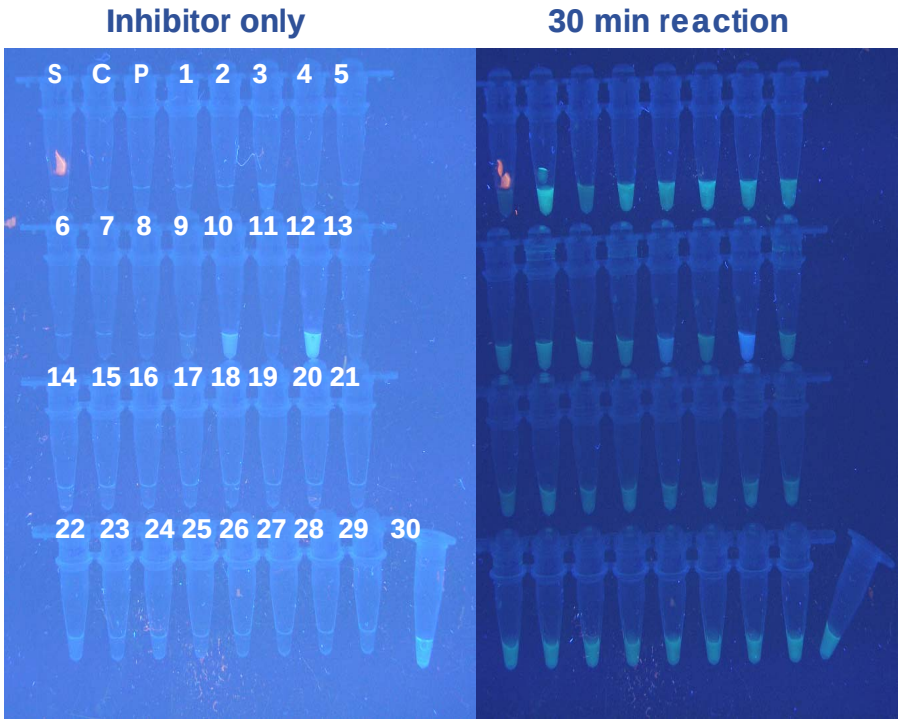
25, 000 compounds:
- Plasmepsin: 5000 compounds
- Pf-DHFR: 15,000 compounds
- Pf-GST: 5000 compounds

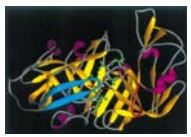
Number of Jobs	500
Total Number of compounds simulated	25000
Estimated duration on 1 CPU	347 days
Duration on the grid	25 days
Maximum number of concurrent running jobs	90
Number of computing elements used	1
Average duration of a job	16.6 hours

Virtual screening on the grid: deployment status

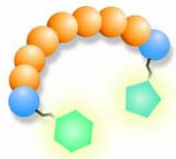
Dates	Target (s)	CPU consumed EGEE AuverGrid	Data produced	Specific features	Status
Summer 2005	Malaria: plasmepines	80 years	1TB	First data challenge	In vitro tests In vivo tests
Spring 2006	Avian flu: Neuraminidase N1	100 years*	800 GB	<45 days needed for preparation	In vitro tests
Winter 2006	Malaria: GST, DHFR, Tubulin	400 years	1,6TB	> 100.000 dockings / hr	Under analysis
Fall 2007	Avian flu: Neuraminidase N1	Estimated 100 CPU years*	Estimated 800 GB*	Joint deployment on CNGrid	First stage under analysis
Spring 2008	Diabetes: amylase	Estimated 120 CPU years	Estimated 800 GB	New production environment	Under way
Spring 2008	*Malaria: DHPS*	*To be estimated*	*To be estimated*	*Joint deployment on desktop grid*	*In preparation*

Post-MD: *in vitro* biological validation





Plasmepsin



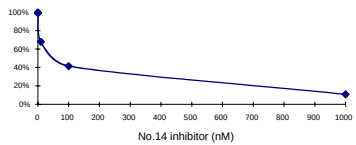
FRET substrate

30 compounds

Compound	IC_{50} FRET (nM)[a]
1	154.91
2	146.03
3	209.90
4	101.32
5	174.97
6	91.32
7	86.61
8	98.72
9	248.84
10	241.51
11	107.80
12	127.31
13	155.54
14	72.17
15	123.93
16	85.48
17	73.65
18	82.59
19	74.56
20	72.24
21	71.24
22	163.50
23	99.56
24	115.62
25	88.23
26	94.42
27	75.62
28	100.40
29	114.84
30	246.37

- **All 30 compounds show inhibition activity for nM concentration**
- **6 compounds have IC50 under 80 nM**

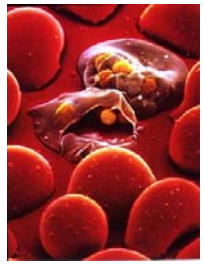
Fluorescence measurement -> IC50



Post-MD: *in vivo* biological validation

- **Impact on *Plasmodium falciparum* growth**

Compound	IC_{50} (M)
1	>8,3E-06
2	>>2,5E-05
3	>8,3E-06
4	>8,3E-06
5	4,63E-06
6	>8,3E-06
7	>>2,5E-05
8	3,76E-06
9	>8,3E-06
10	>>2,5E-05

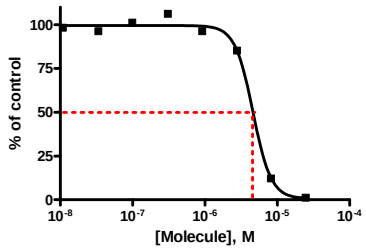


P. falciparum infected red blood cells

+

10 compounds

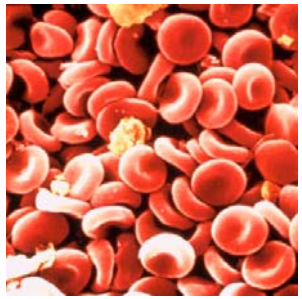
Radioactivity incorporation in parasite -> IC50



2 compounds show IC50 in micromolar range

Post-MD: *in vivo* biological validation

- Cytotoxicity on Human Cell model



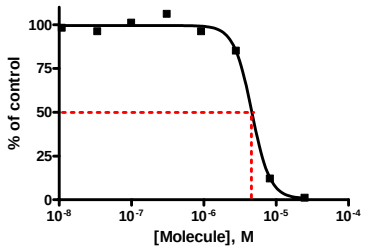
Human cells

+

10 compounds

Results not yet known

Radioactivity incorporation on Human Cells -> IC50

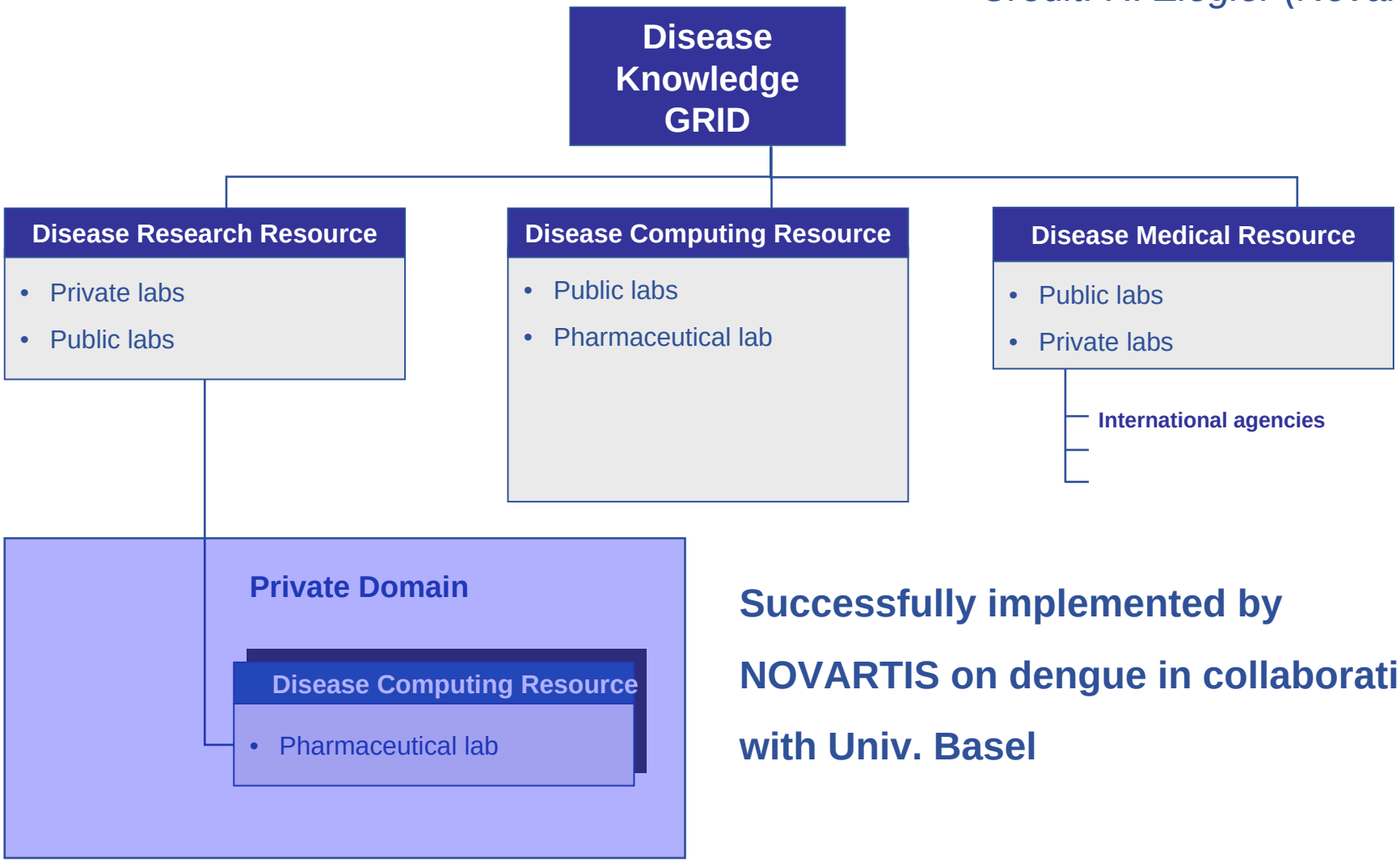


Issue: deployment of licensed software on the grid

- **Most docking and Molecular Dynamics software packages are distributed under commercial licenses**
- **Solutions found on a case by case basis**
- **FlexX: licenses made freely available by BioSolveIT**
  - License server deployed at Fraunhofer Institute
  - One token per grid node running FlexX
  - Up to 5000 concurrent tokens used
- **Amber: gentleman agreement**
  - Users: Amber available only to grid users from institutes owning an Amber license
  - Resources
    - First step: Amber deployed only on clusters owning an Amber license
    - Second step: Amber freely deployed on the grid

Issue: private - public collaboration model

Credit: R. Ziegler (Novartis)



Successfully implemented by NOVARTIS on dengue in collaboration with Univ. Basel

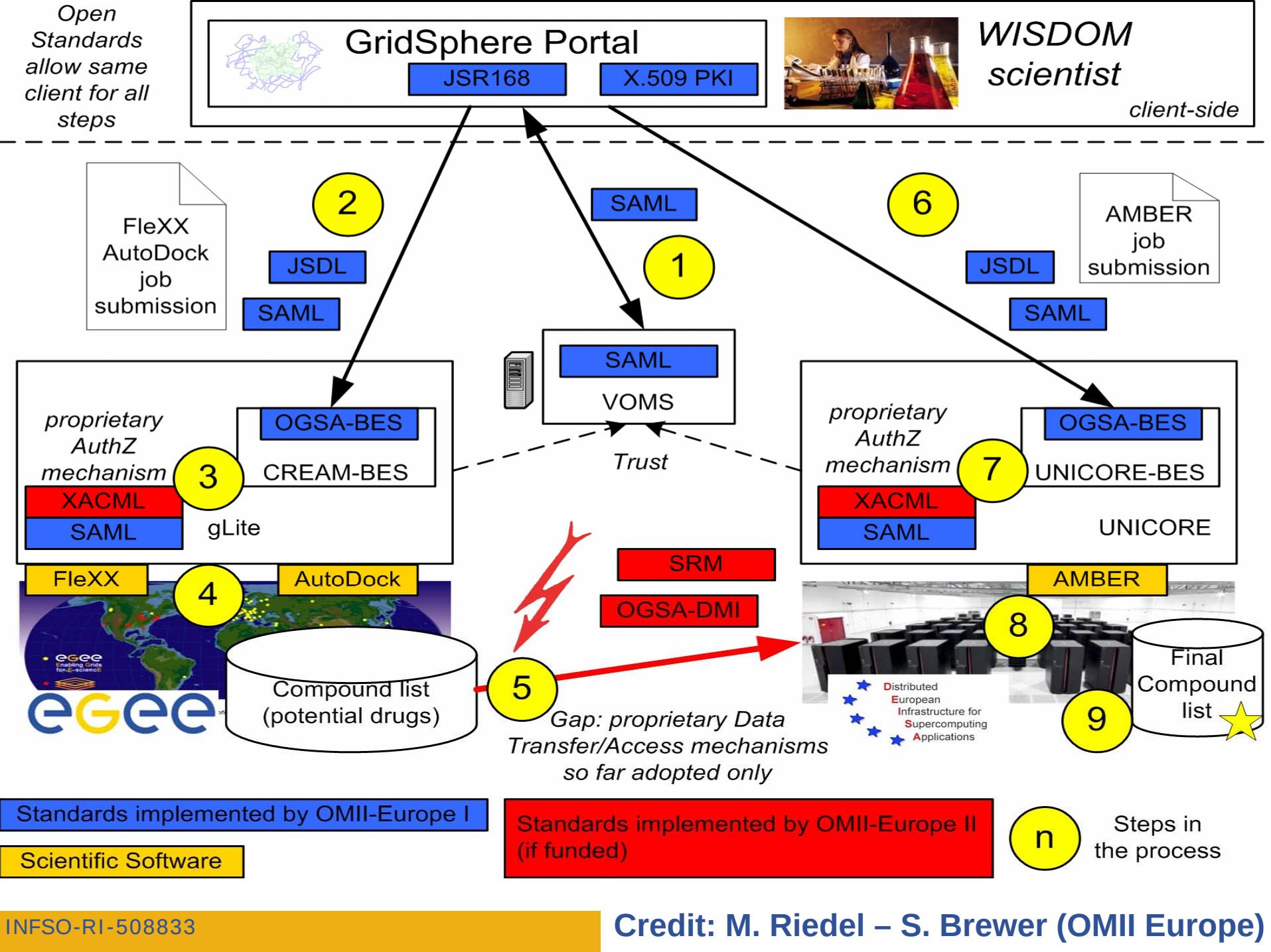
Issue: intellectual property

- **Molecules selected *in silico* belong to public databases**
  - Molecules patented for their antimalarial activity

- **Who claims IP?**
  - Many actors: computing centres, WISDOM collaboration partners, wet labs
  - Agreed statement
    - § all information including analysis of potential hits are made publicly available. If a group takes screening information and synthesizes the physical compound and tests it extensively in the wet-lab, it might establish IP on their side and can establish claims on the physical compound and its behavior in biological assays as long as they cite the source for their initial analysis correctly.

Issue: access to resources

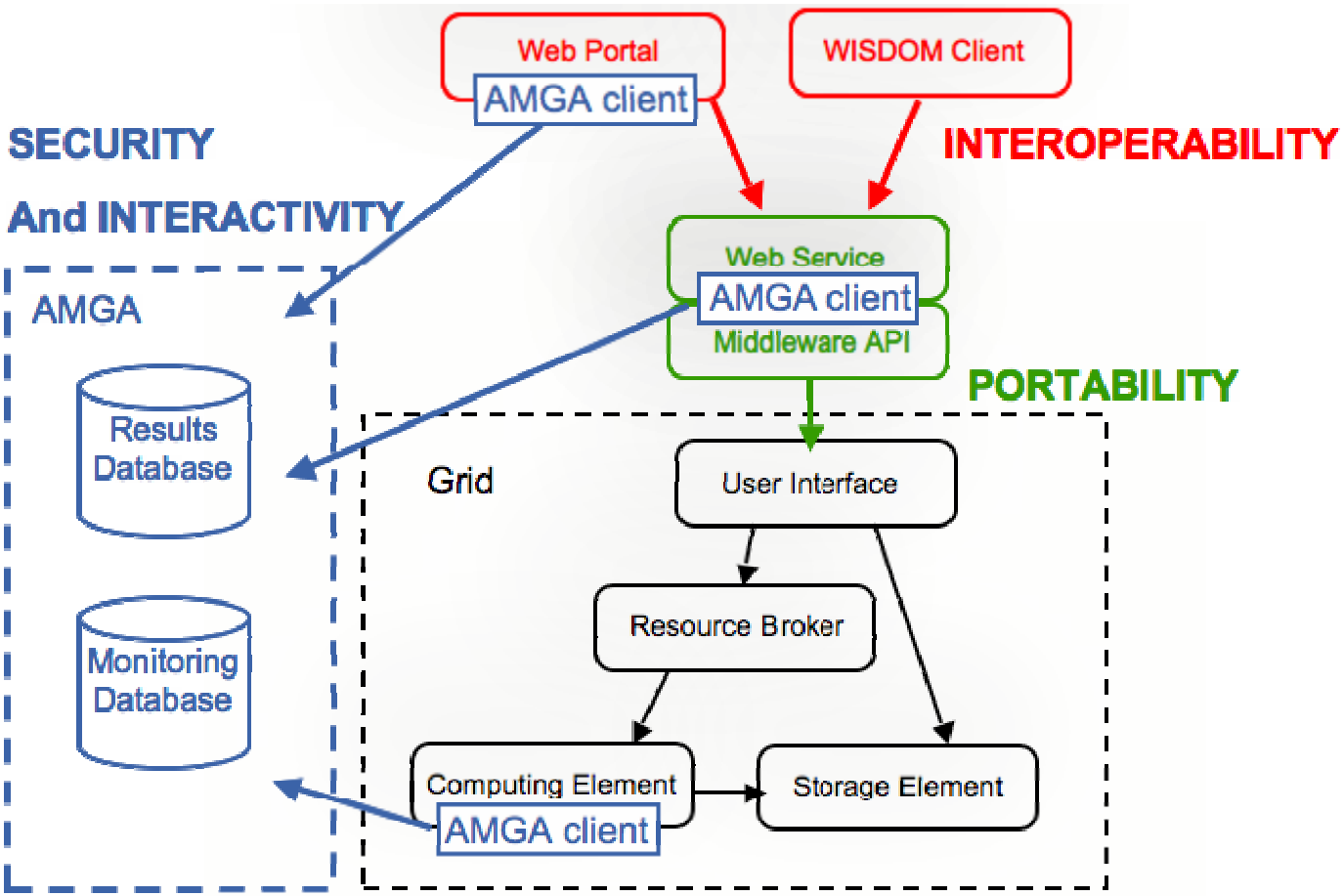
- **Docking one biological target on 1 million drug-like compounds requires between 2 and 30 CPU years depending on the software**
- **Requests received from research groups around the globe for docking targets related to numerous diseases**
  - AIDS, Avian Flu, Diabetes, Malaria, Schistosomiasis, Tuberculosis
- **Need for more resources**
- **Interoperability -> Access to multiple grids**
  - Interoperability of desktop grids and EGEE for molecular docking (EDGES)
  - Implementation of open standards to access EGEE and DEISA for molecular dynamics (OMII-Europe)



Credit: M. Riedel – S. Brewer (OMII Europe)

Issue: security

- **All biomedical data (input, output) are stored in a metadata catalogue with fine grain access control**



WISDOM Production Environment

Perspectives

- **Innovative Medicines Initiative (European Commission)**
  - Goal: revitalize the biopharmaceutical research and development
  - (R&D) environment for Europe
- **Strategic Research Agenda**
  - Identification of bottlenecks in the pharmaceutical R&D process
  - Recommendations to improve competitiveness
- **Key action: improve knowledge management**
  - Develop enhanced knowledge representation models and data exchange standards for complex systems,
  - Build a core reference database of validated experimental data extracted from the literature,
  - Design standards for and build an expert tool to allow the federation of local databases in a secured environment.

Pharmaceutical R&D: requirements on knowledge management

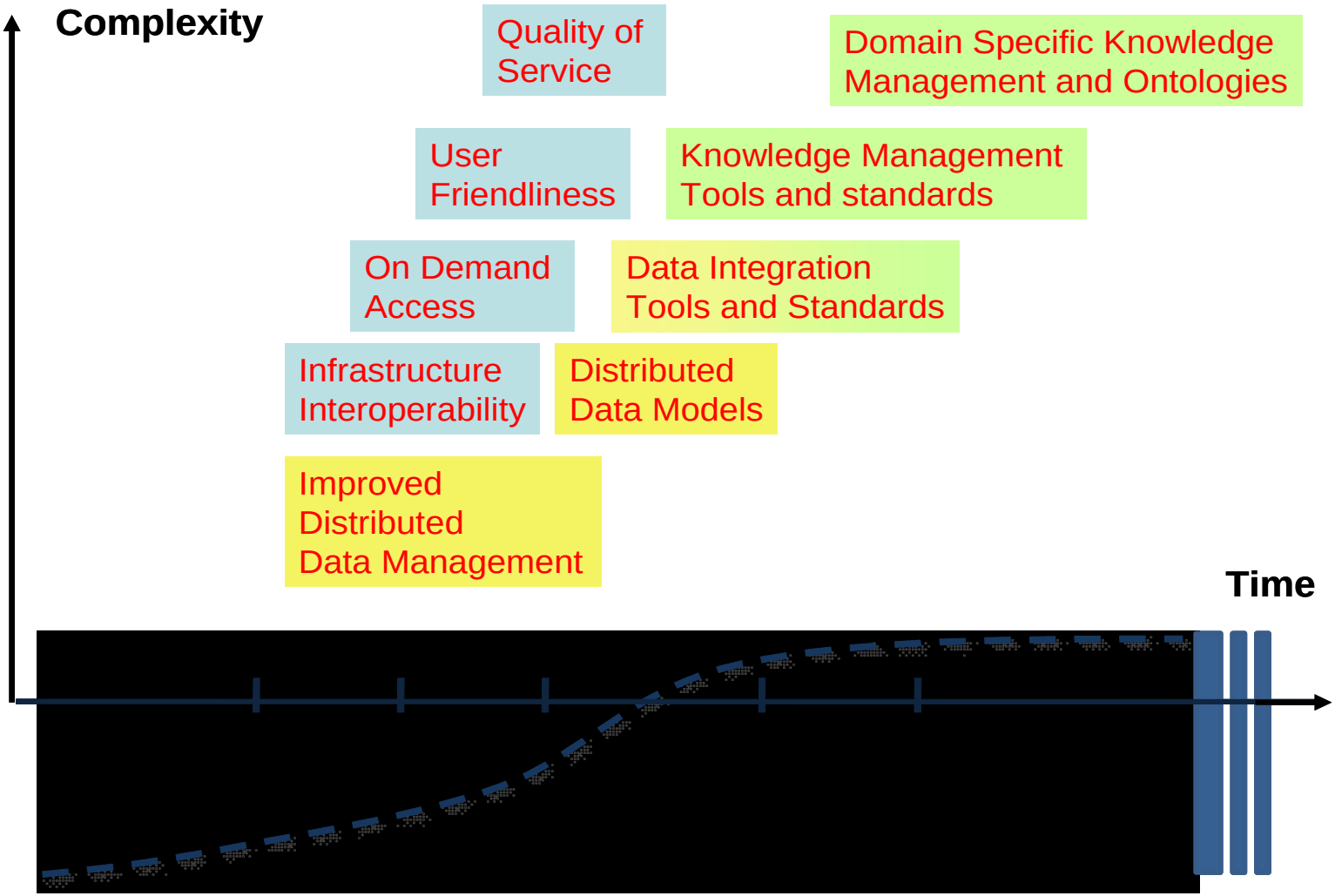
- Capacity to search, query, extract, integrate and share data in a scientifically and semantically consistent manner across heterogeneous sources (public and proprietary) ranging from chemical structures and “omics” to clinical trial data,
- Capacity to integrate and share scientific tools (*e.g.*, modelling, simulation) as modules in a generic framework and apply them to relevant dynamic data sets,
- Expressive data representation and exchange standards,
- Dynamic and customizable configuration of applications,
- Encapsulation of validated physiological models, when applicable,
- Flexible, secure (covering all aspects of data protection encountered in a biomedical context), and scalable IT infrastructure.

Recommended actions to address technical gaps

- Develop a strategy to identify the areas of interest to all stakeholders,
- Provide mechanisms for data federation across heterogeneous data sources,
- Provide a flexible and secure collaborative environment serving all stakeholders,
- Provide standards and mechanisms for consistent data integration and data sharing,
- Provide standards and mechanisms for consistent integration of complex scientific tools and computational models,
- Insure interoperability of computing services across organizations,
- Develop broad and generic research projects for bridging gaps in current technologies.

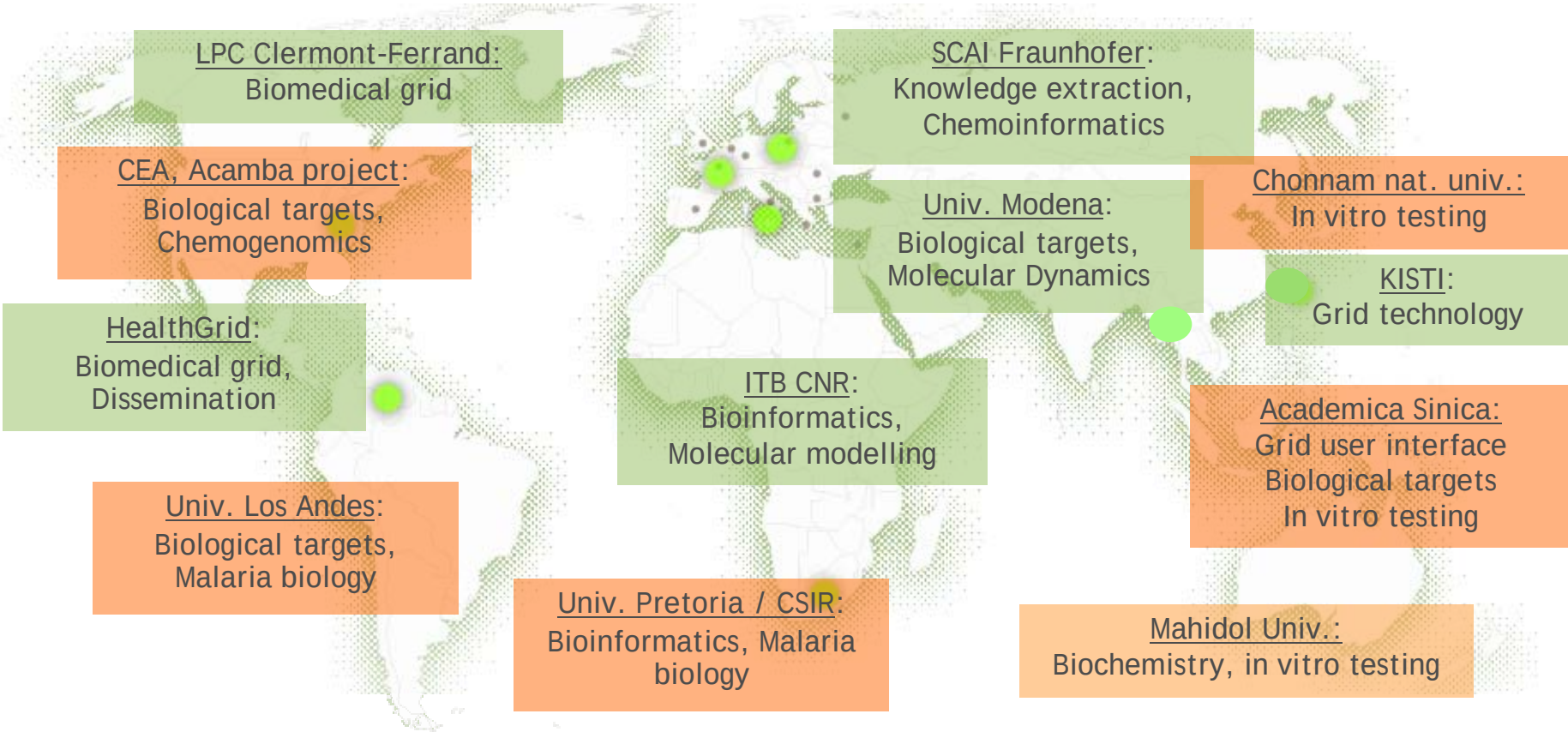
Roadmap to knowledge management

SHARE roadmap
http://share.healthgrid.org



Conclusion

- **In silico drug discovery routinely deployed on grid infrastructures**
  - Number of issues already addressed
- **Next step: improve knowledge management to enable pharmaceutical R&D environment**



Acknowledgements

Academia Sinica, Taiwan
Hurng-Chun LEE, Simon C. LIN (Grid Computing Center)
Ying-Ta WU, Chon-Chen LEE (Genomic Research Center)

HealthGrid
Nicolas SPALINGER, Nicolas JACQ,

SCAI-Fraunhofer Institute, Germany
Martin HOFMANN, Vinod KASAM

Modena University, Italy
Giulio RASTELLI, Gianluca DEGLIESPOSTI

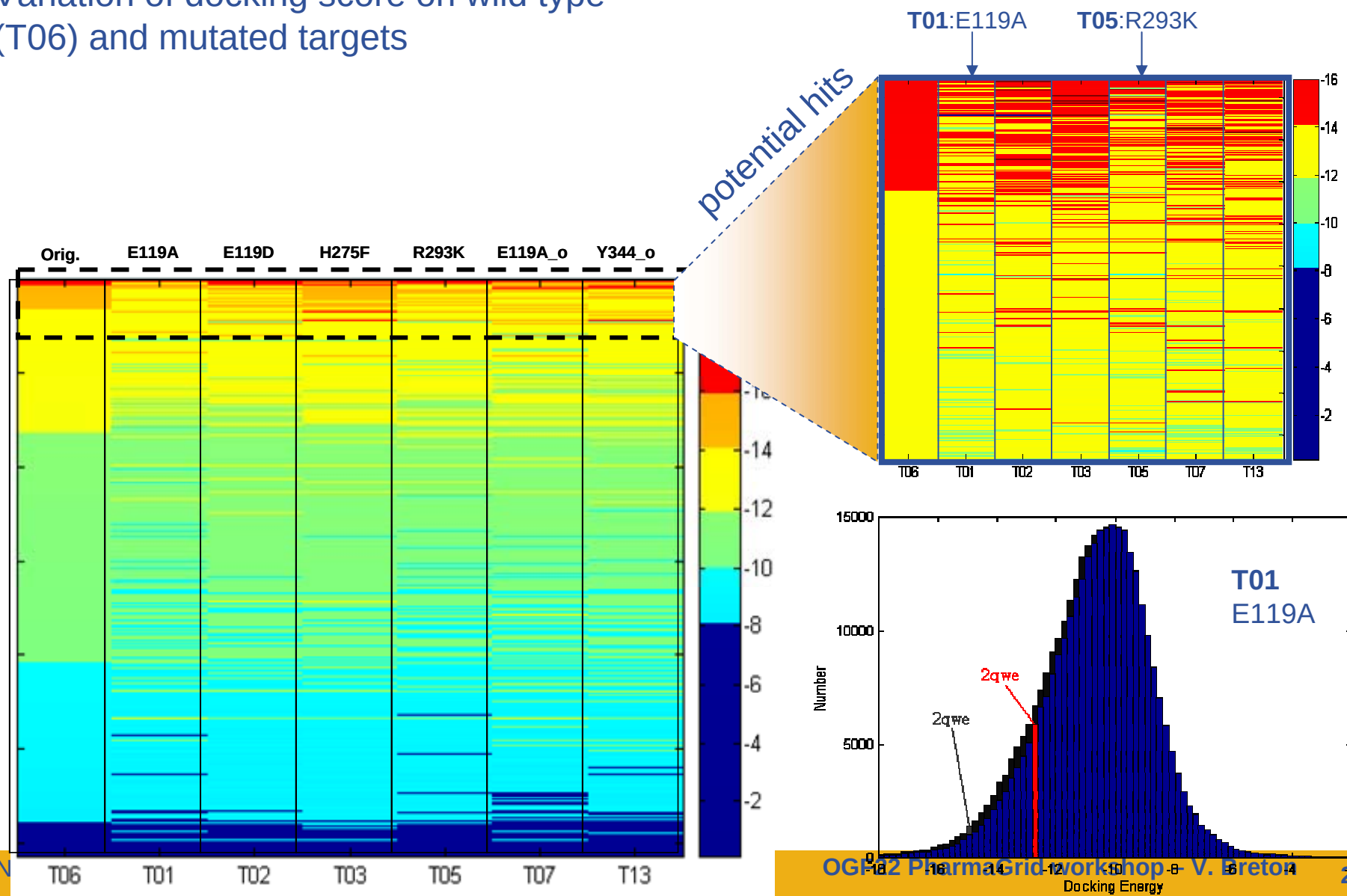
Chonnam National University, Korea
Doman KIM, Young-Min KIM (Neuraminidases), Hee-Kyoung KANG (Plasmepsin)

ITB-CNR
Luciano MILANESI, Pasqualina D'URSI, Gabrielle TROMBETTI

CNRS-IN2P3-LPC, Clermont-Fd, France
Jean SALZEMANN, Ana DA COSTA,
Vincent BLOCH, Yannick LEGRE

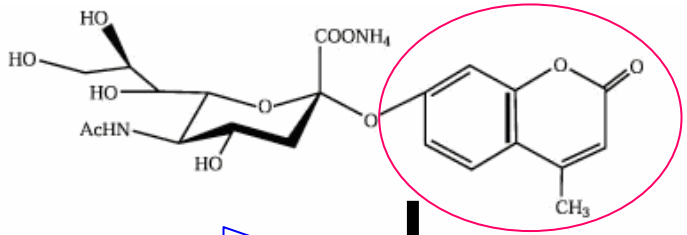
s do impact inhibitory

Variation of docking score on wild type (T06) and mutated targets



In vitro tests at Chonnam National University

4-Methylumbeliferyl-*N*-acetyl-α-*D*-neuramininic acid ammonium salt [4MU-NANA]; Substrate



First screening

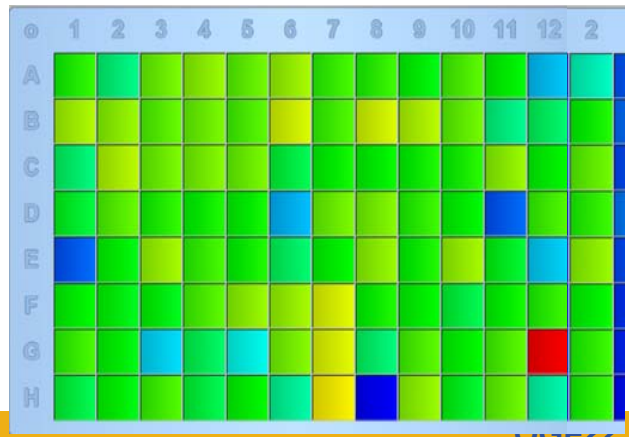
(200 nmol)

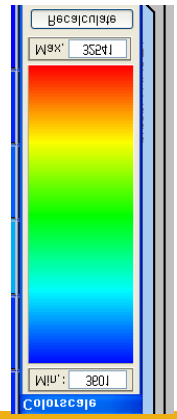
Recombinant Neuraminidase

Second screening
(2 nmol)

Kinetic study

Spectrofluorometric detector RF-551
362 nm excitation and **448** nm emission wavelengths



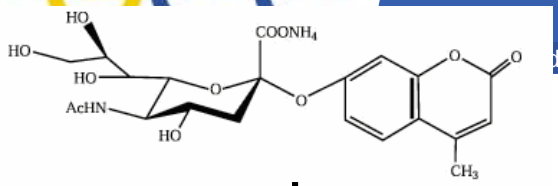


Red

Inhibition

Blue

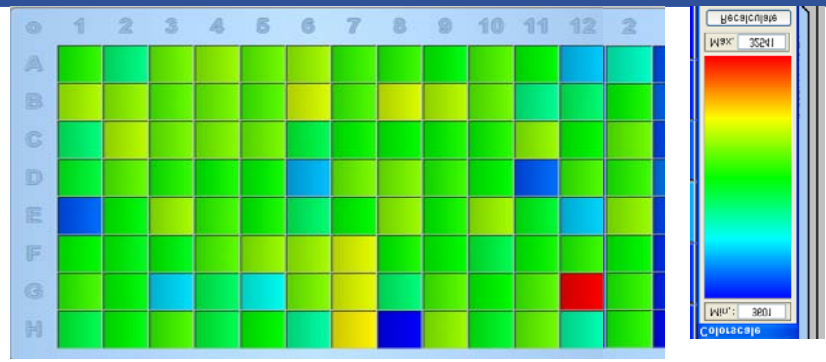
Results on 308 compounds tested in vitro

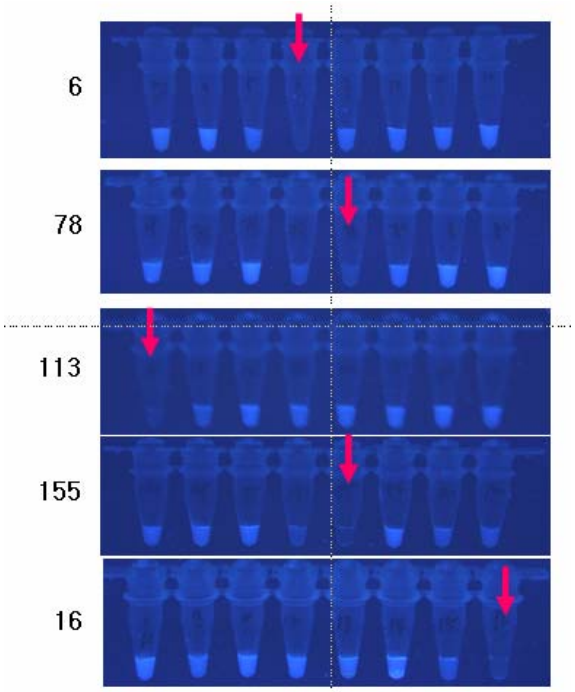


4MU-NANA
: 20 µM/RM

Neuraminidase
: 10 mU/reaction

Measure at excitation 362 nm and emission at 448 nm





On UV

Rank	Compounds	Relative activity of Neu1
1	113	67
2	16	72
3	6	73
4	155	74
5	78	78
⋮	⋮	⋮
63	Tamiflu	100
⋮	⋮	⋮