

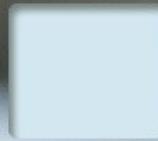


Premières rencontres scientifiques France Grilles

- Quand ? le 19 Septembre 2011
- Où ? à la Cité des Congrès de Lyon
- Remise du 1er prix France Grilles pour la meilleure contribution scientifique
- Appel à contributions en ligne le 31 Mai sur <http://france-grilles.fr>
- Opportunité de présenter les grilles régionales sur un grand stand pendant le technical forum EGI
- Pour en savoir plus, contacter G. Romier



CENTRE NATIONAL
DE LA RECHERCHE
SCIENTIFIQUE



Grilles en sciences du vivant

Vincent Breton

Lille, 17 Mai 2011



Life sciences activities on the French NGI

- Early adoption of the grid paradigm: 2001
- Topics addressed:
 - bioinformatics (phylogenetics, proteomics)
 - Structural biology
 - In silico drug discovery
 - Epidemiology
 - Medical physics
 - Medical imaging (T. Glatard)
 - Neurosciences (T. Glatard)

Life sciences applications on regional grid initiatives

AUVERGRID:~1554 cores/271TB

CIMENT: 2200 cores/290TB

GRIF: ~1500 cores/350TB

Grille Aquitaine: ~200 cores/8TB

Lille: 324 cores

MSFG (Montpellier Sud France Grilles):
~104 cores /10TB

Strasbourg Grand Est:
~1200 cores/550TB

Tidra: ~10000 cores/50PB





L'activité sur la grille en France

Clermont-Ferrand: **LPC, LIMOS**

Grenoble: **TIMC, INSERM 438,
RMN Bioclinique, LECA,
LPMMC, IN, LPSC, LSP, SIMAP**

Lyon: **CREATIS, BBE, IBCP**

Montpellier: **LIRMM**

Nice: **I3S**

Strasbourg: **IPHC**

**GRISBI: Dedicated
infrastructure for bioinformatics**

**DECRYPTTHON: Grid to help
cure Muscular Dystrophy**



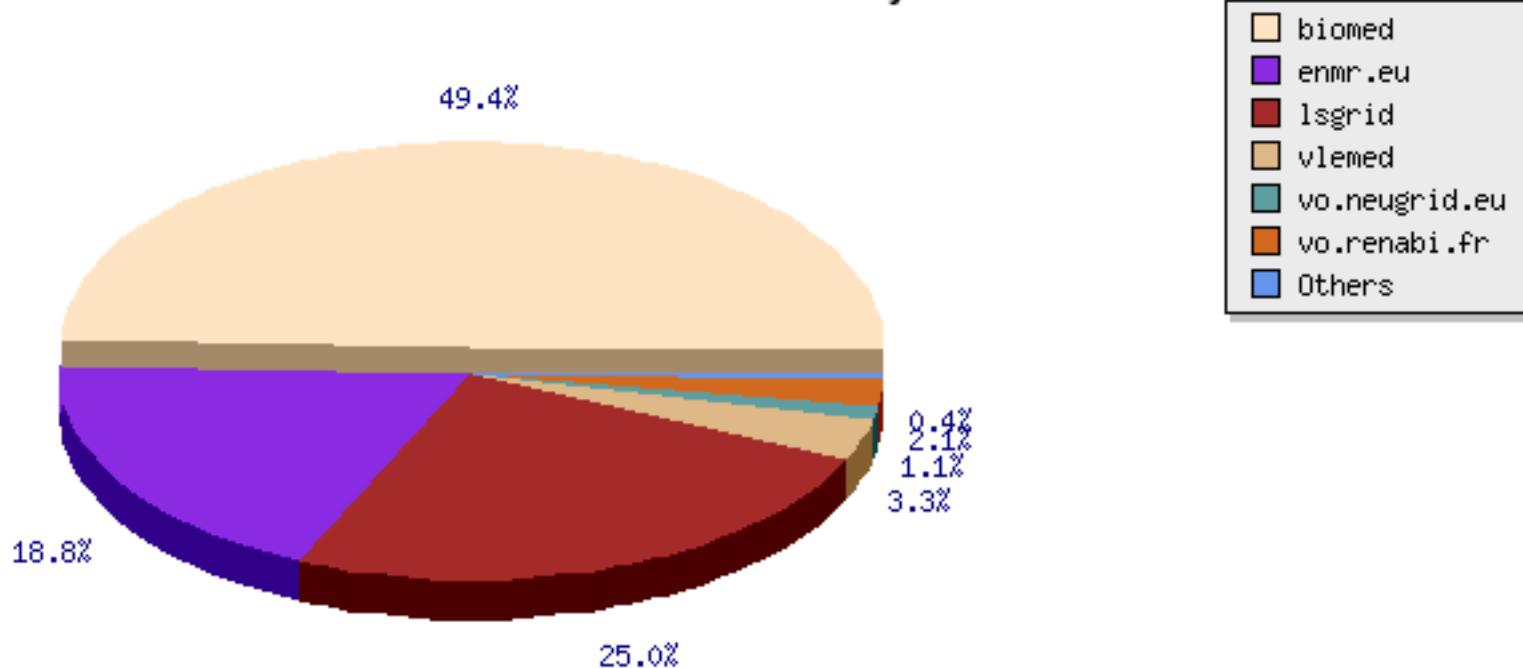
A very active user community

- 30% of the biomed VO members
- The largest users

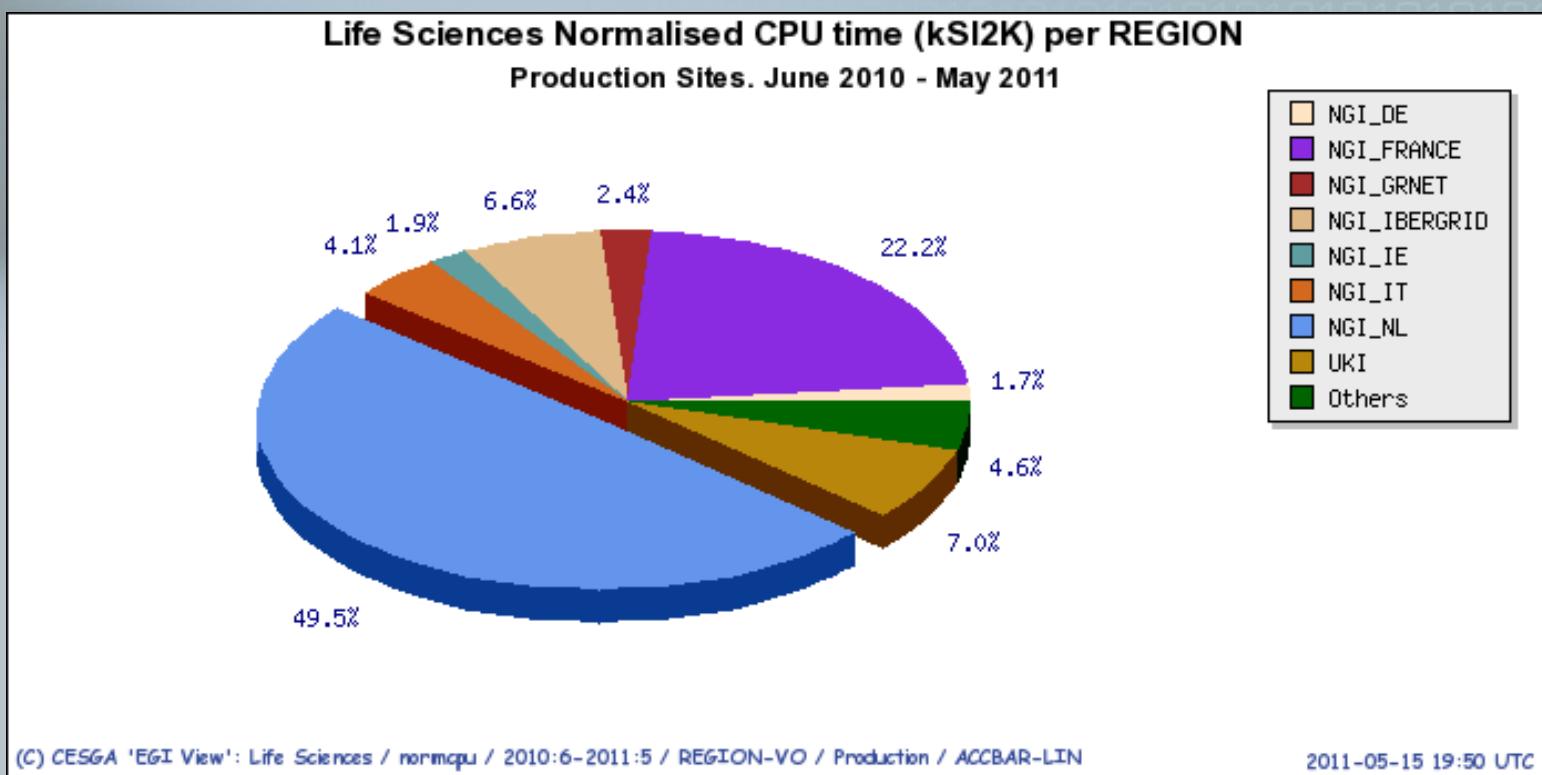
Biomed	CPU elapsed	njobs
France	4005953	1372514
Hungary	309528	134265
Italy	199571	96349
South Korea	40371	12992

Life Sciences Normalised CPU time (kSI2K) per VO

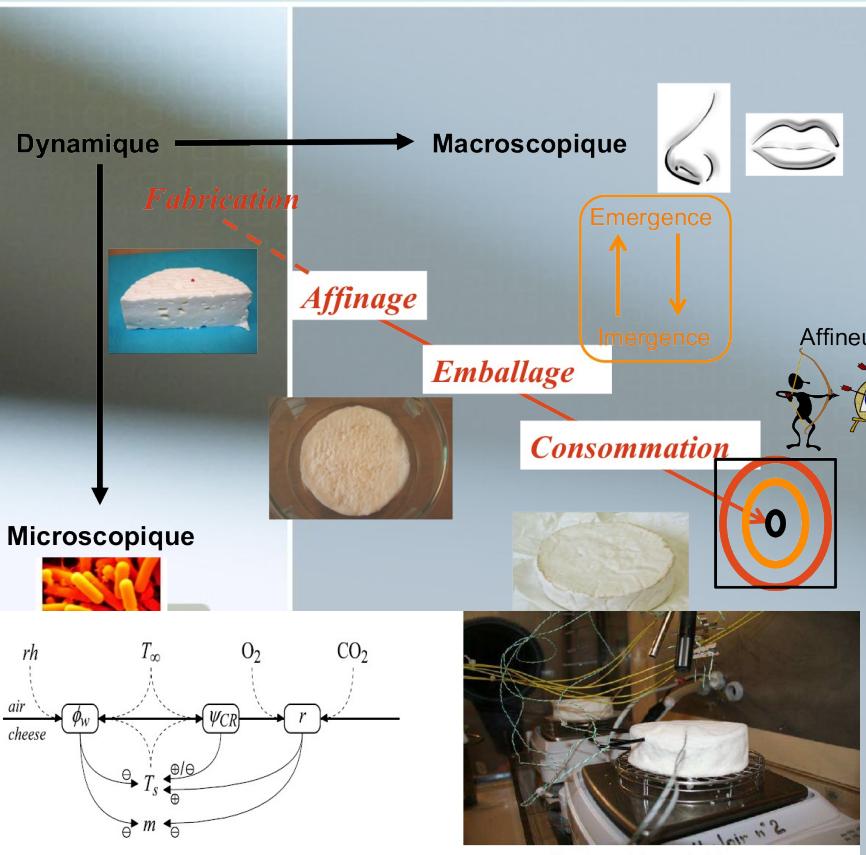
Production Sites. June 2010 - May 2011



France Grilles is the second largest contributor of resources to life sciences on EGI



Complex sciences: the example of cheese refining



1,5 months of simulation done in 2 days
on 1600 CPUs to define the best
environment for camembert refinement



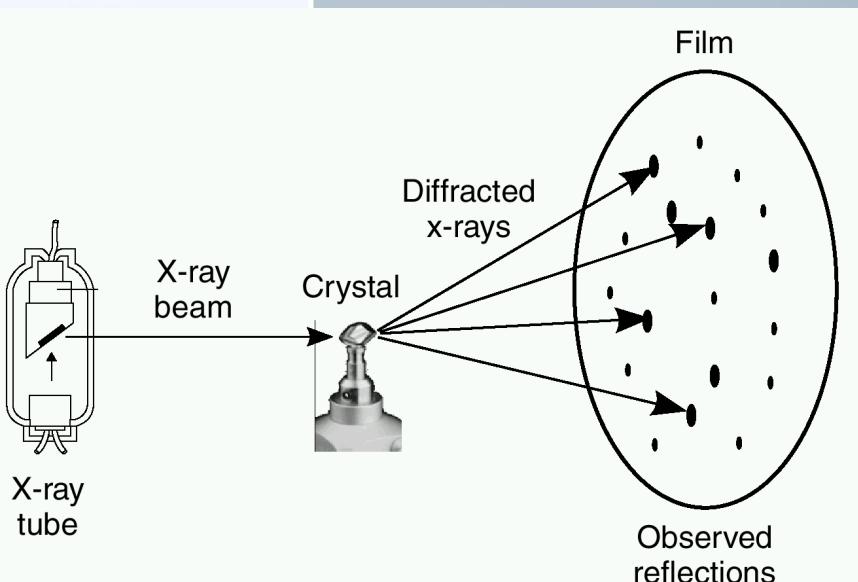
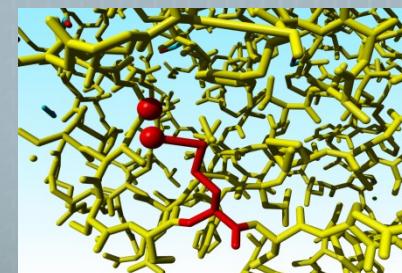
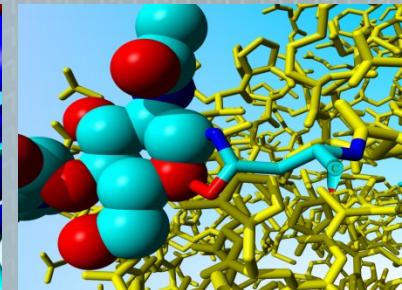
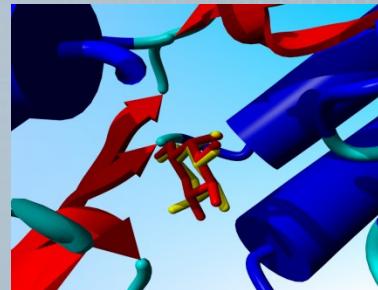
- Evaporation flux: $\phi_w = k(a_{ws}P_{sv}(T_s) - rhP_{sv}(T_\infty))$.
- Radiative and convective flux: $\psi_{cr} = h(T_s - T_\infty) + \epsilon\sigma(T_s^4 - T_\infty^4)$.
- Respiration activity: $r = \left(\frac{r_{O_2} + r_{CO_2}}{2}\right)$.

$$(S) \begin{cases} \frac{T_s}{dt} = \frac{s}{mC}(-\psi_{cr} - \lambda\phi_w + \alpha r) \\ \frac{dm}{dt} = -s(\phi_w + w_cr) \end{cases}$$

Credit: CEMAGREF - INRA

Structural biology: recalculating protein 3D structures in PDB

- The PDB data base gathers publicly available 3D protein structures
 - Full of bugs
- Goal: redo the structures by recalculating the diffraction patterns



PDB-files	42.752
X-ray structures	36.124
Successfully recalculated	~36.000
Improved R-free	12.500/17000
CPU time estimate	21.7 CPU years
Real time estimate	1 month on Embrace VO on EGEE

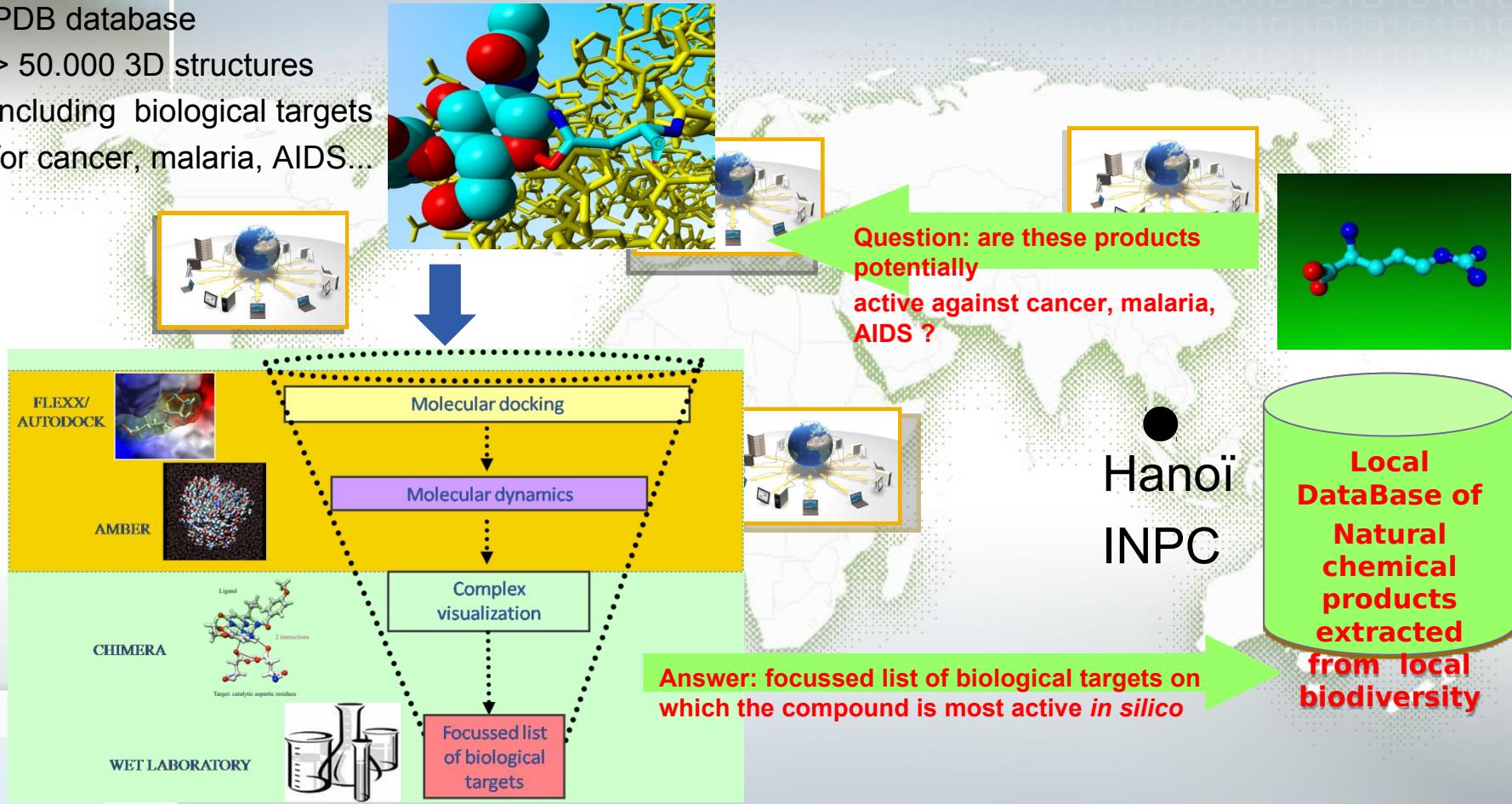
R.P Joosten et al, Journal of Applied Crystallography, (2009) 42, 1-9



In silico drug discovery

Docking compounds coming from biodiversity (CNRS, IFI, INPC, IOIT)

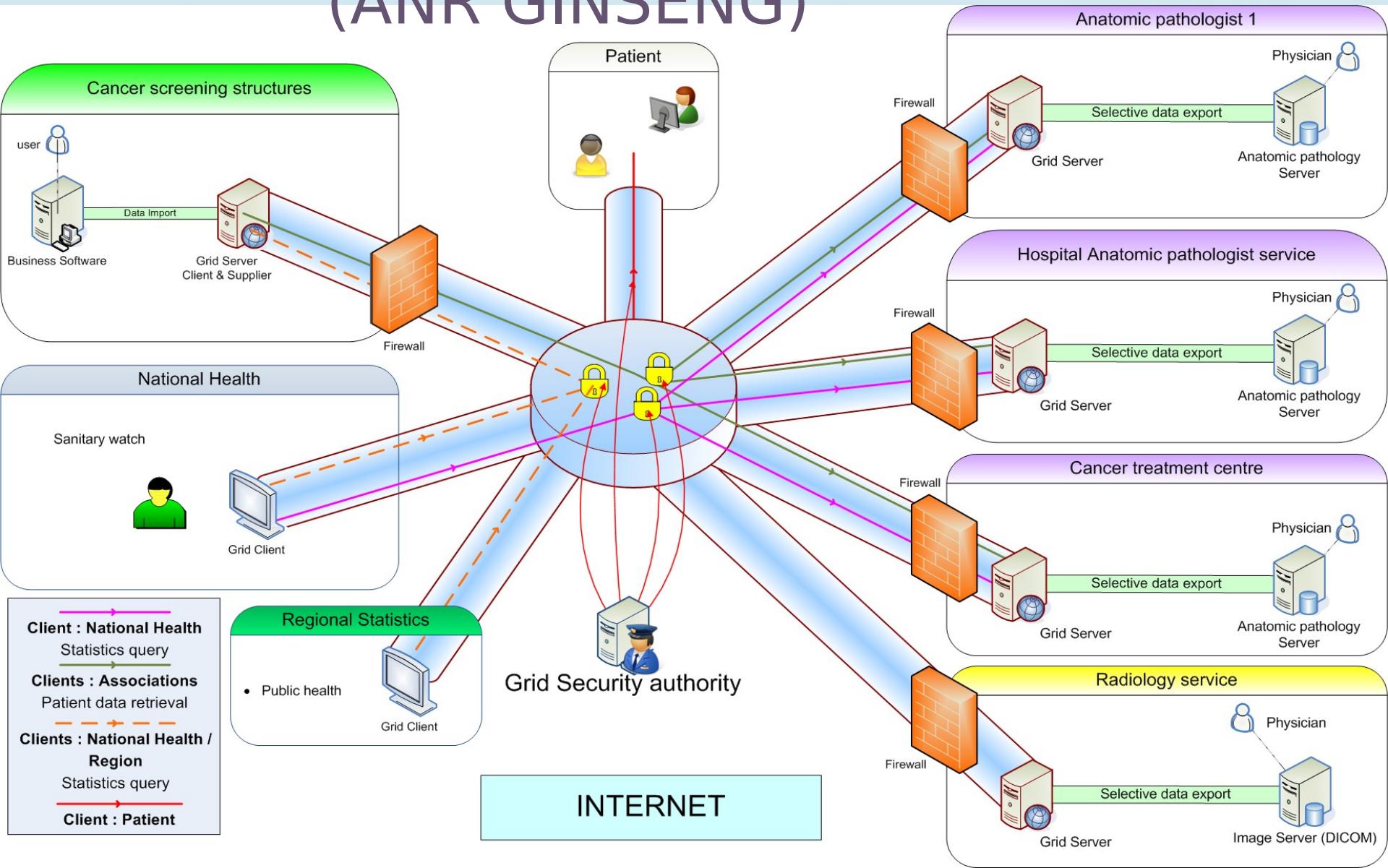
PDB database
> 50.000 3D structures
including biological targets
for cancer, malaria, AIDS...





Epidemiology

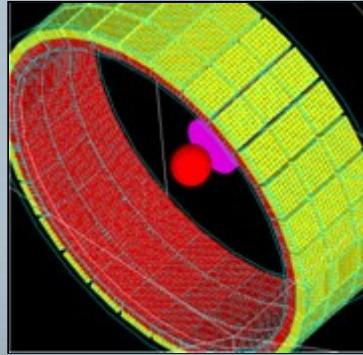
Cancer surveillance network (ANR GINSENG)



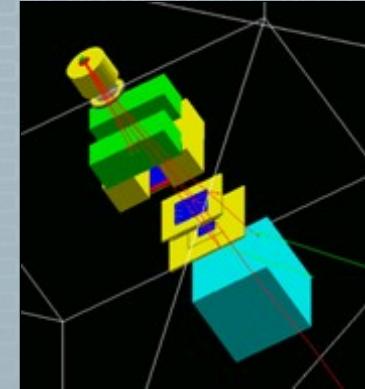


Medical physics GateLab (ANR VIP)

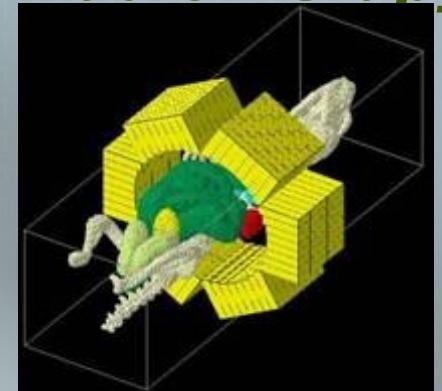
- User interface for launching Gate on distributed environments
- Execution on GPUs, CPUs, clusters
- Current functionalities
 - Parses simulation (mac) file
 - Finds local inputs files and copies them on the grid
 - Submits simulation
 - Keeps track of simulation history
 - Allows to choose the Gate executable among a list of releases
 - Allows to split the simulation automatically in a number of jobs depending on an estimation of the total CPU time
 - Stop and merge (new feature)



PET camera



Radiotherapy



Small animal imaging

G A T E

SIMULATIONS OF
PRECLINICAL AND
CLINICAL SCANS IN
EMISSION TOMOGRAPHY



A new field: NGS (Next Generation Sequencing)

- NGS technology massively parallelises nucleotide sequencing procedures, making the sequencing of genomes and of transcriptomes much faster and cheaper than ever before.
- The new technology is, however, posing massive (bio-) informatics challenges which require new ways of thinking and novel solutions

Credit: E/ Bongkam-Rudloff
SeqAhead Cost Action



Challenges

Credit: E/ Bongkam-Rudloff
SeqAhead Cost Action

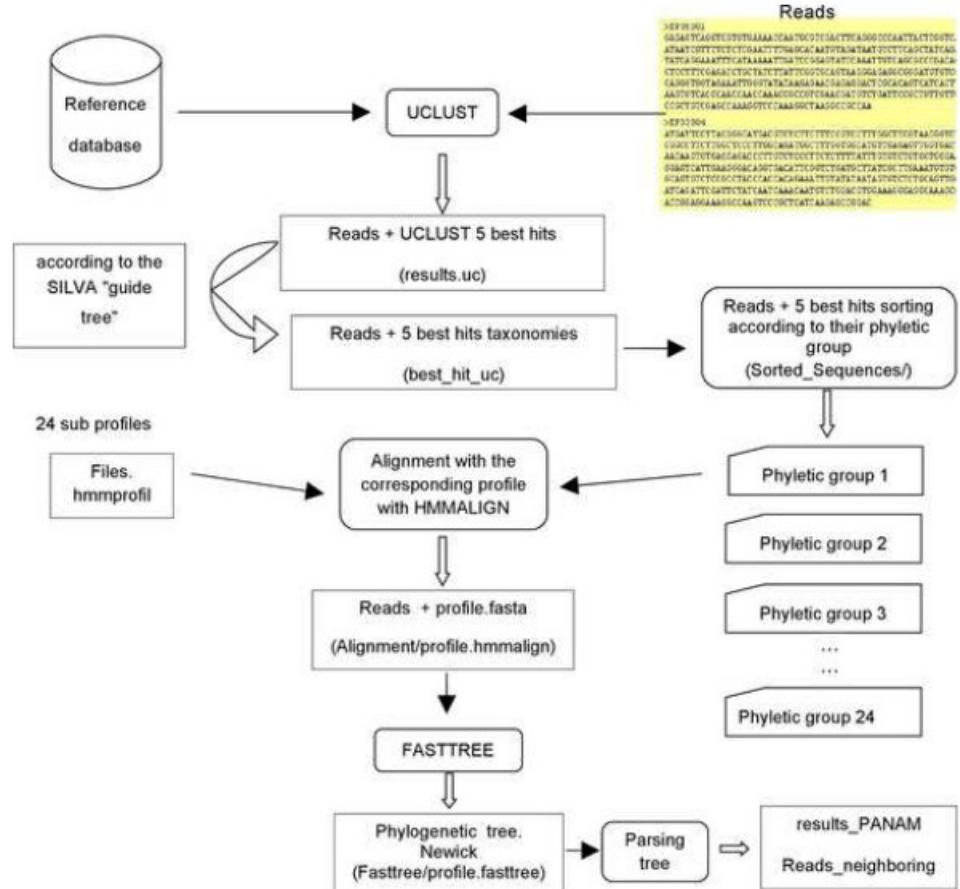
- Data analysis

- This is the most diverse & most challenging NGS area. Many techniques are available to answer biological questions; the number & types of experiment, & the research fields to which they can be applied, are too numerous to list here

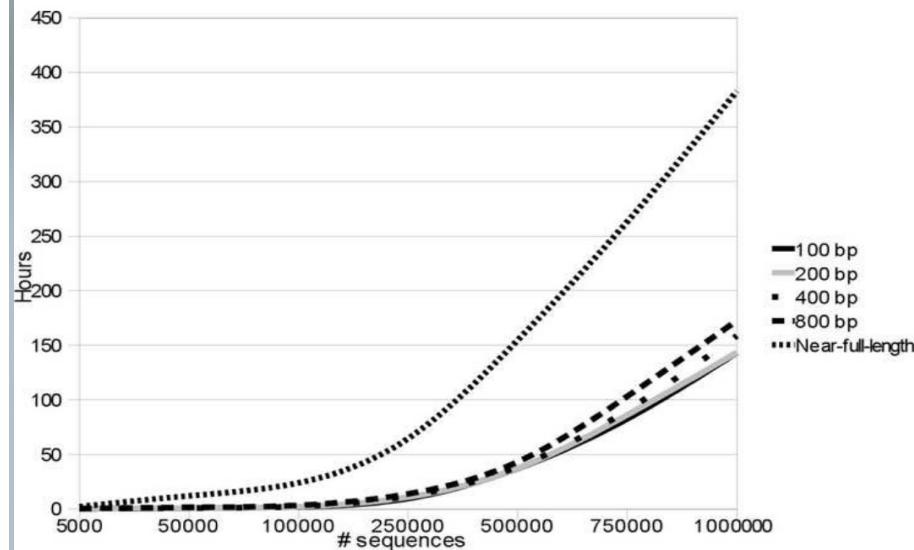
- Data storage

- The vast amount of data arriving daily at computing centres creates completely new challenges in hardware (e.g., new data-storage facilities, large bandwidth for data transfer) and software (data security, algorithms for data quality control, analysis)

Example: ePANAM



- Molecular ecology: sequencing of microorganisms directly from environmental samples
- Goal: study the variety of diverse ecosystems
- Method: NGS +



Credit: G. Bronner, D. Debroas,
N. Taib (LMGE)



Conclusion

- Activité intense et variée sur la grille en sciences du vivant
 - Importante production scientifique
 - Impact au niveau européen
- Perspectives
 - Poursuite du développement au niveau local
 - Poursuite du soutien aux organisations virtuelles internationales (biomed)
 - Forte implication dans la Life Science Virtual Research Community (T. Glatard)