Bioinformatics Applications in EU-EGEE project

Christophe Blanchet

Institut de Biologie et Chimie des Protéines
CNRS IBCP, University Lyon1, Lyon cedex 07, France

christophe.blanchet@ibcp.fr
http://gbio.ibcp.fr/cblanchet

EU-EGEE User Tutorial, Marseille, 4 octobre, 2006

Outline

- **EU-EGEE and Biomedical Applications**
- **GPS@: ex. of bioinformatics application**
  - Biological database
  - Legacy bioinformatics tools
  - Protein sequence analysis
- **Virtualization of biological data on EGEE**
  - EGEE-enabling Parrot
  - Local copy vs I/O access forward
  - Application to grid Web portal

The EU EGEE grid project

egee
Enabling Grids
for E-sciencE

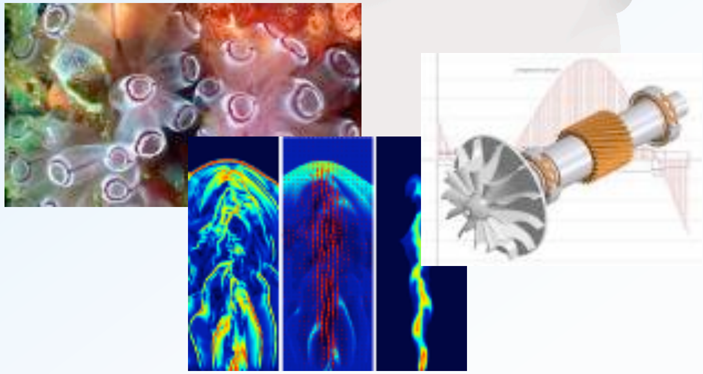
Information Society
and Media

- **EGEE**
  - 1 April 2004 – 31 March 2006
  - 71 partners in 27 countries, federated in regional Grids
- **EGEE-II**
  - 1 April 2006 – 31 March 2008
  - 91 partners in 32 countries
  - 13 Federations
- **Objectives**
  - Large-scale, production-quality infrastructure for e-Science
  - Improving and maintaining “gLite” Grid middleware
  - Attracting new resources and users from industry as well as science

Asia
Central Europe
CERN
France
Germany/Switzerland
Italy
Networking

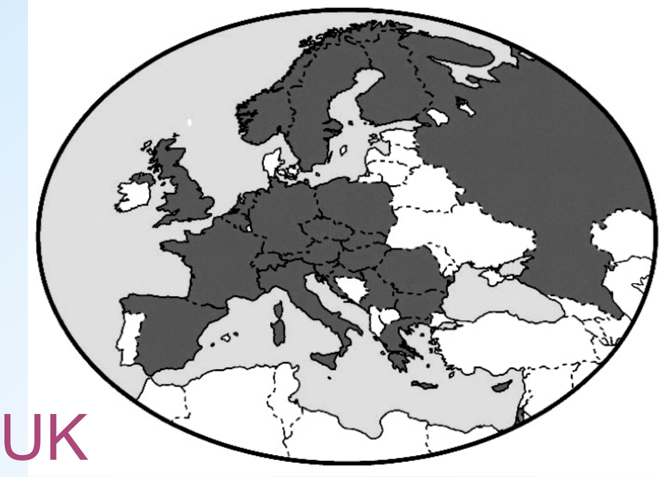
Northern Europe
Russia
South-Eastern Europe
South-Western Europe
UK/Ireland
USA

Size of the infrastructure (Sept. 2006):
192 sites in 40 countries
~25 000 CPU
~ 3 PB disk, + tape MSS



NA4 Activity

- **Application Identification and Support (NA4)**
  - 25 countries, 40 partners, 280+ participants, 1000s of users
- **Support the large and diverse EGEE user community:**
  - **Promote dialog**: Users' Forums & EGEE Conferences
  - **Technical Aid**: Porting code, procedural issues
  - **Liaison**: Software and operational requirements
- **Need active participation:**
  - **Feedback**: Infrastructure, configuration, and middleware
  - **Resources**: Hardware and human
- **Users' Forum**
  - In conjunction with OGF Manchester, UK
  - OGF — May 7-9
  - EGEE Users' Forum — May 9-11



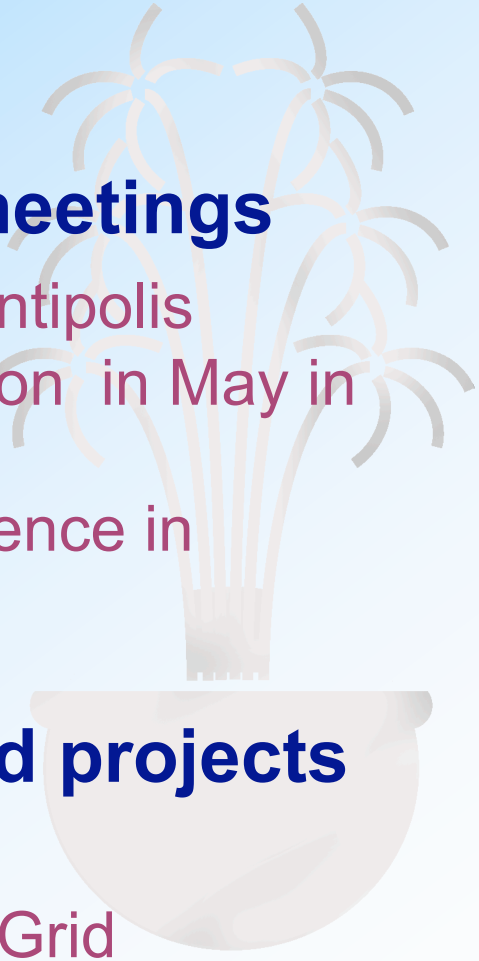
Source: Dr C. Loomis
NA4 Leader

Biomed VO Status

- **Three active subgroups**
  - Medical imaging (J. Montagnat)
  - Bioinformatics (C. Blanchet)
  - Drug discovery (V.Breton)

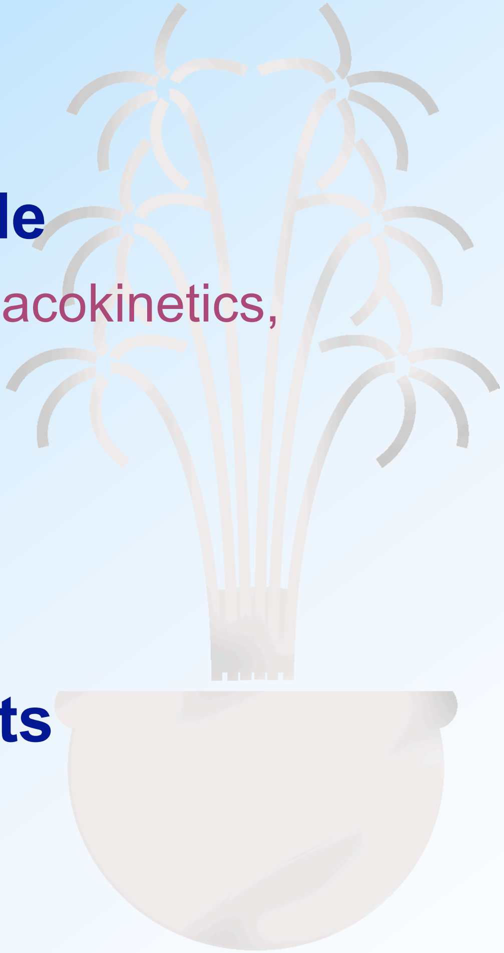
- **The three subgroups have separate meetings**
  - Medical imaging meeting in July in Sophia-Antipolis
  - Bioinformatics meeting on database replication in May in Pisa
  - Drug discovery meeting at Healthgrid conference in Valencia in June

- **Active relationships with EGEE related projects and other EC projects**
  - BioinfoGRID, EELA, EUChinaGRID, EUMedGrid
  - Embrace



Medical imaging

- Services are available on EGEE for the medical imaging community
  - Medical Data Management
  - Workflow engines: Moteur, DAGMAN
  - Portals: P-GRADE, GENIUS
- Several applications are in production mode
  - Bronze standard, GATE, 3D MRI simulation, pharmacokinetics, GPTM3D, Clinical Decision Support System
- New applications are under development
  - SEE++ strabismus surgery planning
  - SPM based early diagnosis of Alzheimer
  - FreeSurfer-based brain image analysis
- Emerging collaboration with related projects
  - Health-e-child



Source: Dr J. Montagnat
NA4 Med. Imaging Leader

Bioinformatics Activity

- **10 Bioinformatics Applications**
  - In production: **Splatche**
  - Prototype: **GPS@, bioDCV, Dengue Docking**
  - Porting: **Large Scale Pathway, BiG, 3DEM**, …
- **Key activities:**
  - Data Virtualization: Enabling legacy bioinformatics applications
    - *with grid and secure data access (**EncFile, GFAL, Perroquet**)*
    - *with large-scale data capability (**3DEM**)*
  - Grid-enabling bioinformatics tools with special requirements:
    - *short job (**GPS@**),*
    - *large job, workflow (**Large Scale Pathway, Splatche, BiG,**)*
  - End-user interfaces: providing biologists with Web portal, Web services (**BiG, GPS@, bioDCV**)
- **Collaboration with related projects: NoE EMBRACE, EELA, BIOINFOGRID, SwissBioGrid.**
- **Contact: Christophe.Blanchet@ibcp.fr**

Next event: EGEE User Forum, Manchester, UK, 9-11 april 2007

Drug discovery

- **First WISDOM data challenge**
  - Results analyzed
  - Further processing using Molecular Dynamics explored within BioinfoGRID

- **Avian flu data challenge**
  - Results under analysis
  - Need for a second data challenge on a newly published protein structure

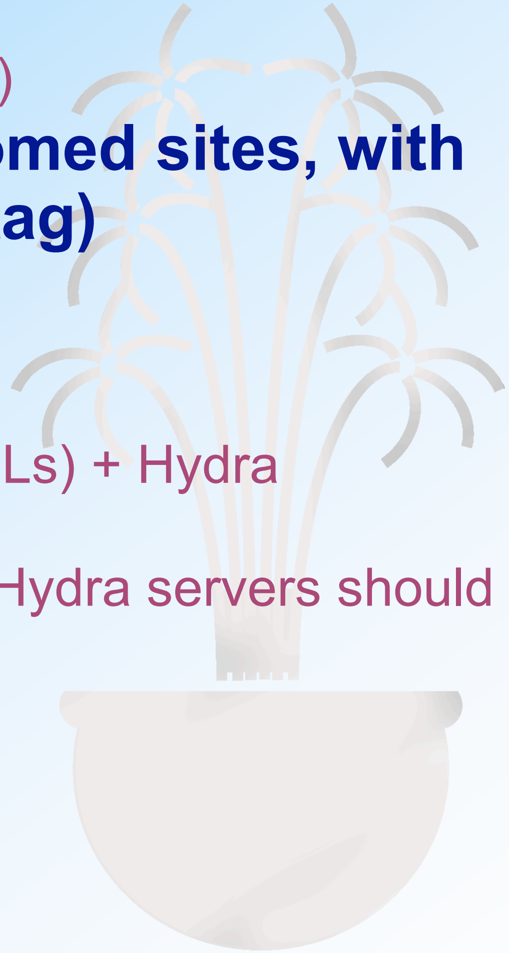
- **Second WISDOM data challenge (Oct 1st – Dec 15th)**
  - Focus on malaria (4 targets)
  - 5 infrastructures are contributing: Auvergrid, EGEE, EELA, EUChinaGRID, EUMedGRID
  - 2 other EC projects involved: BioinfoGRID, Embrace

Source: Dr V. Breton
NA4 Drug Discov. Leader

Current major issues

- Short Jobs (<5 min): SDJ workgroup
  - SDJ WG has defined some CE setup rules to decrease grid middleware overhead to ~2 min
  - But only one site (LAL) is enabled (at least publishing it!)

⇒ **deploying SDJ recommendations on other biomed sites, with adequate publication (CE named with « sdj » tag)**



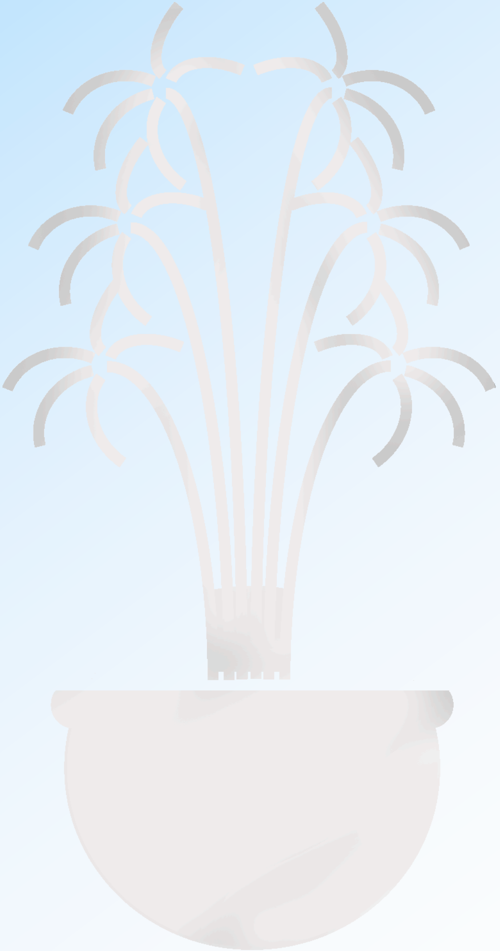
- Data confidentiality
  - Data security addressed through gLiteIO + Fireman (ACLs) + Hydra (encryption)
  - Only clients available in gLite3.0: gLiteIO, Fireman and Hydra servers should be installed by the users
  - Limited security through GFAL + LFC
- Data management
  - No tool available in gLite to allow database integration

EMBRACE (EU FP6)

- « A European Model for Bioinformatics Research and Community Education »
- Goals:
  - simplify and standardize the way in which biological information is served to the researchers who use it.
  - Integrating biological data and bioinformatics tools in grid
- Network of Excellence (2005-2010)
  - From Feb 1st, 2005
  - partners: EBI (PI), EMBL, SIB, CNRS, MPI_MG, INRA, ITB CNR, CNB, ...
- Funded by the European Union (EU-FP6, LHSG-CT-2004-512092)
  - EMBRACE uses a test problem driven development method. The services will be developed through a set of test problems, which will use tasks from real biological research, designed to stretch the system in critical ways

GRID: a Challenge in Bioinformatics

- **Very different applications …**
  - Different requirements and priorities
  - Different resources involved
    - *Hardware*
    - *Human*
  - Different scientific communities adressed
    - *But all are biologists*
    - *Don’t care of the « infra »-structure*
- **… but some common requirements**
  - Data
    - *Deploying updatable databases*
    - *Security of biological data (medical or industrial)*
  - Tools
    - *Integrating numerous, complex programs*
    - *Legacy application*
    - *Portal and user interfaces*
    - *MPI parallel applications*

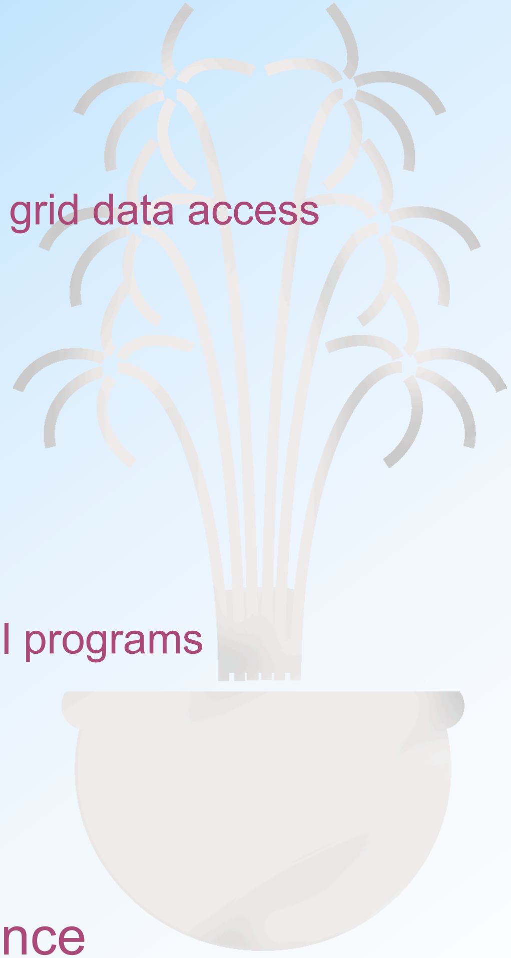


Bioinformatics Applications

GPS@	CNRS IBCP	Christophe Blanchet (IBCP) Christophe.Blanchet@ibcp.fr	Prototype	http://gpsa-pbil.ibcp.fr/
SPLATCHE	External	Dr. Nicolas Ray nicolas.ray@zoo.unibe.ch	Production	**http://cmpg.unibe.ch/software/splatche/**
Large-scale Pathway Analysis	CNRS IBCP	Ralf Herwig herwig@molgen.mpg.de	Porting	
bioDCV	INFN, ICTP (E-GRID)	Cesare Furlanello furlan@itc.it	Prototype	**http://biodcv.itc.it/**
Phylojava	CNRS	Manolo Gouy mgouy@biomserv.univ-lyon1.fr Alexandre Dehne Garcia dehneg@prabi.fr	Porting	**http://pbil.univ-lyon1.fr/software/phylojava/phylojava.html**
BiG	UPV	Ignacio Blanquer iblanque@dsic.upv.es	Porting	
Superlink-online	TAU	Prof. David Horn horn@post.tau.ac.il Mark Silberstein marks@techunix.technion.ac.il	Feasibility	**http://bioinfo.cs.technion.ac.il/superlink-online/**
3DEM	CNB/CSIC	Jose-Maria Carazo carazo@cnb.uam.es	Porting	**http://3dem.ucsd.edu/**
CAST	UCY	George Tsouloupas (UCY) georget@ucy.ac.cy Maria Poveda (UCY) mpoveda@cs.ucy.ac.cy	Feasibility	
Dengue Docking Project	CSCS	Michael Podvinec (Biozentrum Basel, CH)	Prototype	

Next Meeting of EGEE-Bioinformatics

- **Agenda**
  - Opening and status of EGEE project
  - Applications status and feed-back
  - Key themes:
    - *Biological data on EGEE, access and security.*
      - Data Virtualization: Enabling bioinformatics applications with grid data access
        - SE DPM ? GFAL ? Parrot/Perroquet ? Fuse ?
      - Security: Working on security issues about biological data
        - MDM ? EncFile ?
    - *Computation management*
      - Portal and user interfaces
      - Workload mgmt: short job (SDJ), prioritized job, pilot job
      - Complex job: parallel job (MPI) , application built with several programs
  - AOB
  - Conclusions
- **Location and date**
  - Institute of Biology and Chemistry of Proteins, Lyon, France
  - November 7, 2006



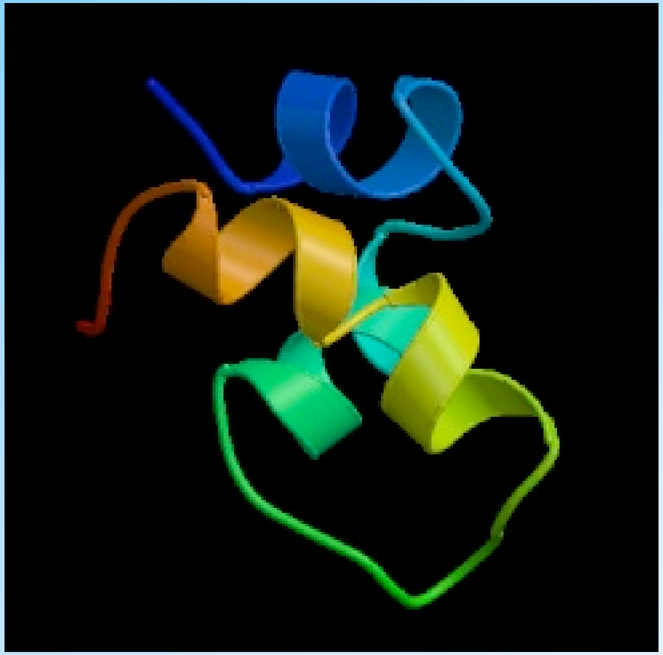
Outline

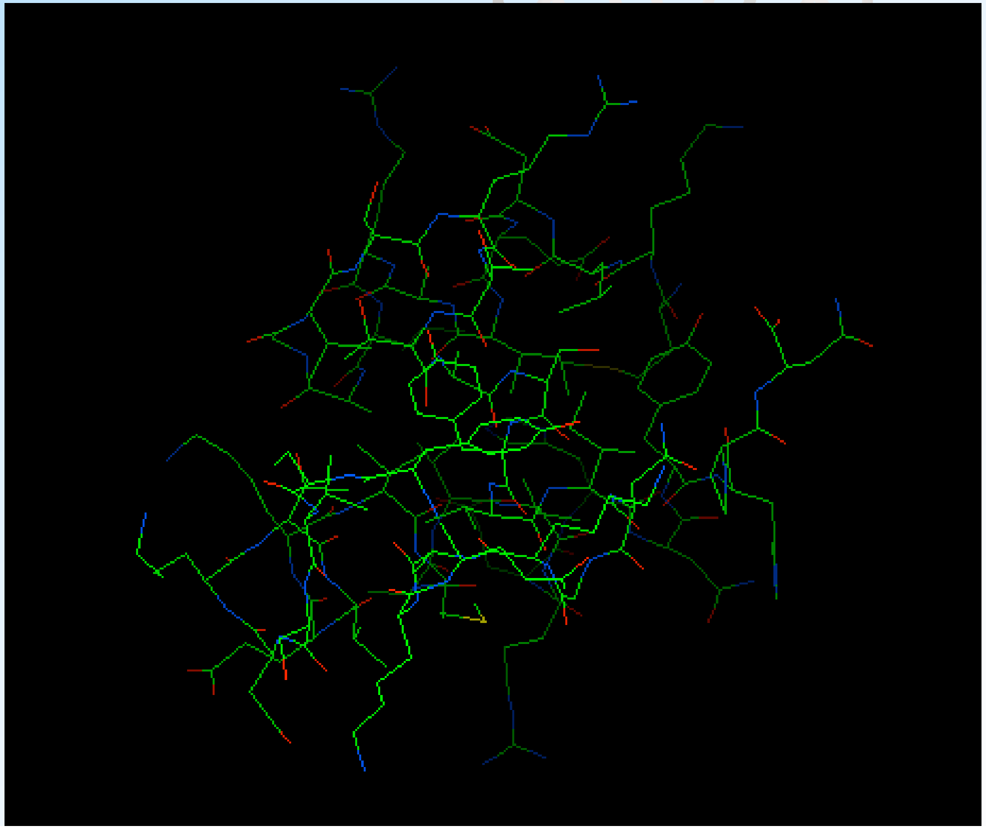
- EU-EGEE and Biomedical Applications

- GPS@: ex. of bioinformatics application
  - Biological database
  - Legacy bioinformatics tools
  - Protein sequence analysis
- Virtualization of biological data on EGEE
  - EGEE-enabling Parrot
  - Local copy vs I/O access forward
  - Application to grid Web portal

Protein ?

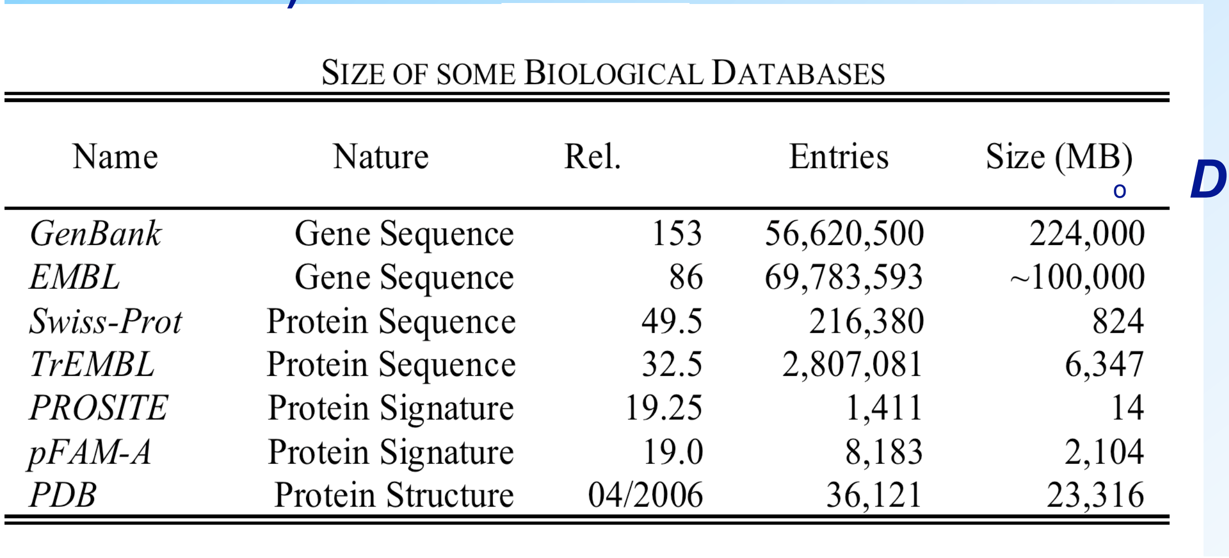
- **FRUCTOSE REPRESSOR DNA-BINDING DOMAIN, NMR, MINIMIZED STRUCTURE**
  - Penin, F., Geourjon, C., Montserret, R., Bockmann, A., Lesage, A., Yang, Y., Bonod-Bidaud, C., Cortay, J.C., Negre, D., Cozzone, A.J., Deleage, G
  - >1UXC:_|PDBID|CHAIN|SEQUENCE MKLDEIARLAGVSRTTASYVINGKAKQYRVSDKTVEKVMAVVREHNYHPNAVAAGLRLQHHHHHH





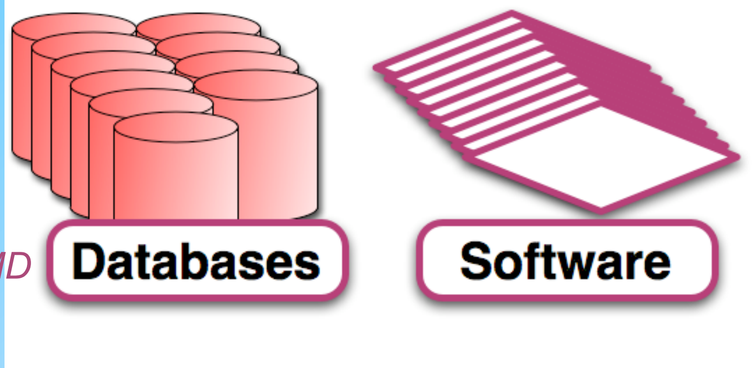
Biological Data and Tools

- **Numerous**
  - + 800 (Galperin *et al.*, 2006)
- **Heterogeneous**
  - Data & metadata (MD)
    - *Swiss-Prot: 12 % of data, 88% MD*
    - *TrEMBL: 19% data, 81% MD*
  - Size: kB to 100s GB
  - Authors & initial location
  - Storage: file, object, image
  - Format: EMBL, GenBank, Pearson-Fasta, PDB, pubmed, …
- **Updatable !!**
- **In some case sensitive (Patient, Industrial, Scientific)**



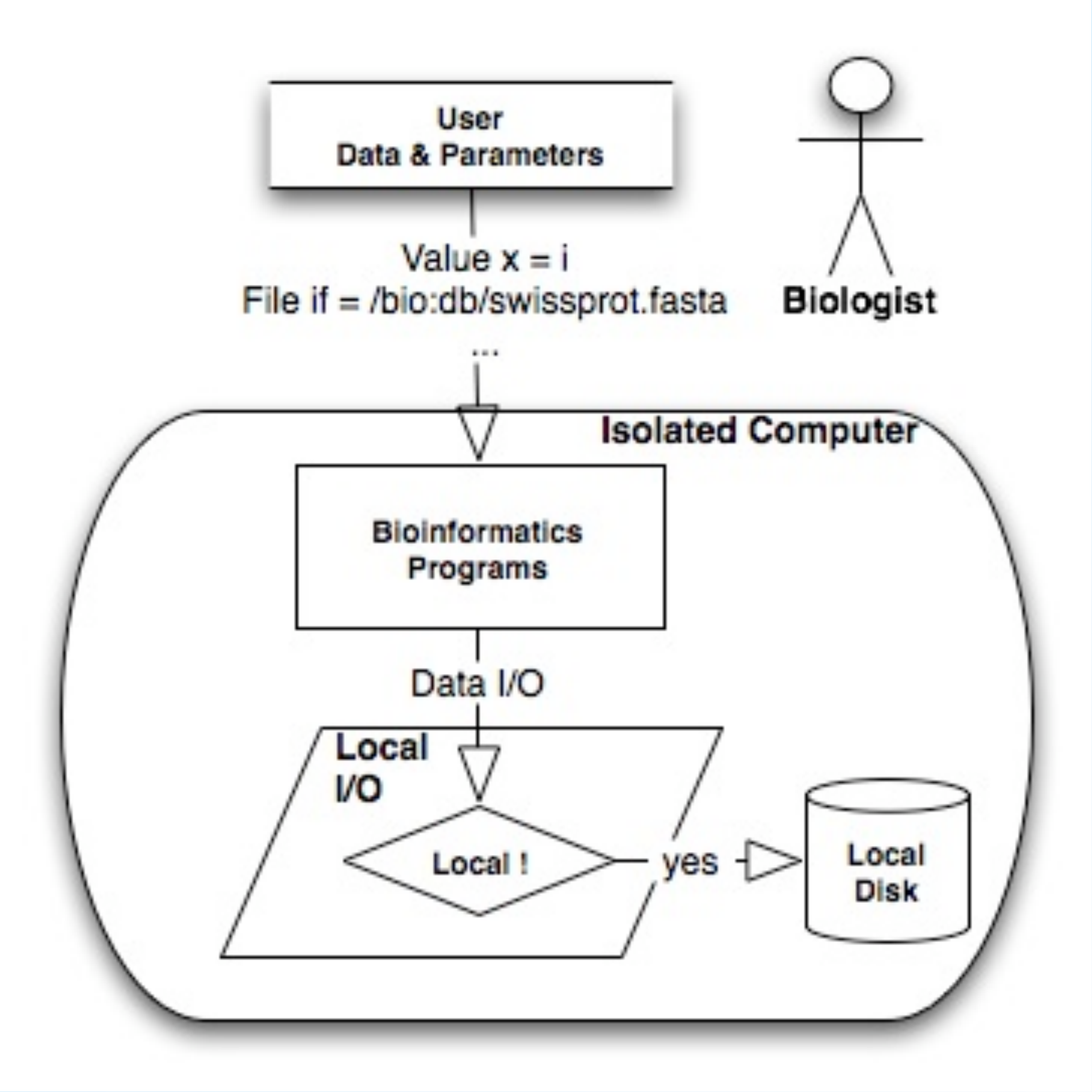
SIZE OF SOME BIOLOGICAL DATABASES

Name	Nature	Rel.	Entries	Size (MB)
GenBank	Gene Sequence	153	56,620,500	224,000
EMBL	Gene Sequence	86	69,783,593	~100,000
Swiss-Prot	Protein Sequence	49.5	216,380	824
TrEMBL	Protein Sequence	32.5	2,807,081	6,347
PROSITE	Protein Signature	19.25	1,411	14
pFAM-A	Protein Signature	19.0	8,183	2,104
PDB	Protein Structure	04/2006	36,121	23,316



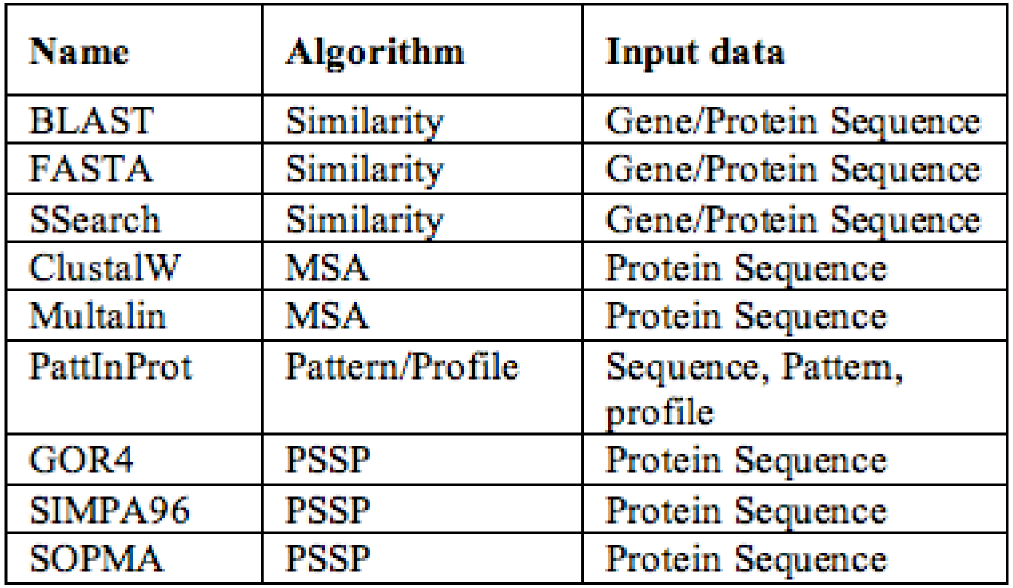
- **Numerous**
  - BioCatalog: > 600 (end of 90s)
  - EMBOSS toolkit: > 200
- **Heterogeneous**
  - Bioinformatics algorithm: Sequence similarity, Multiple alignment, Structural prediction, …
  - Execution: sequential, parallel, workflow
- ***Data I/O***
  - *Text files*
  - *Specific format*
  - *Local I/O only*

Legacy bioinformatics tools



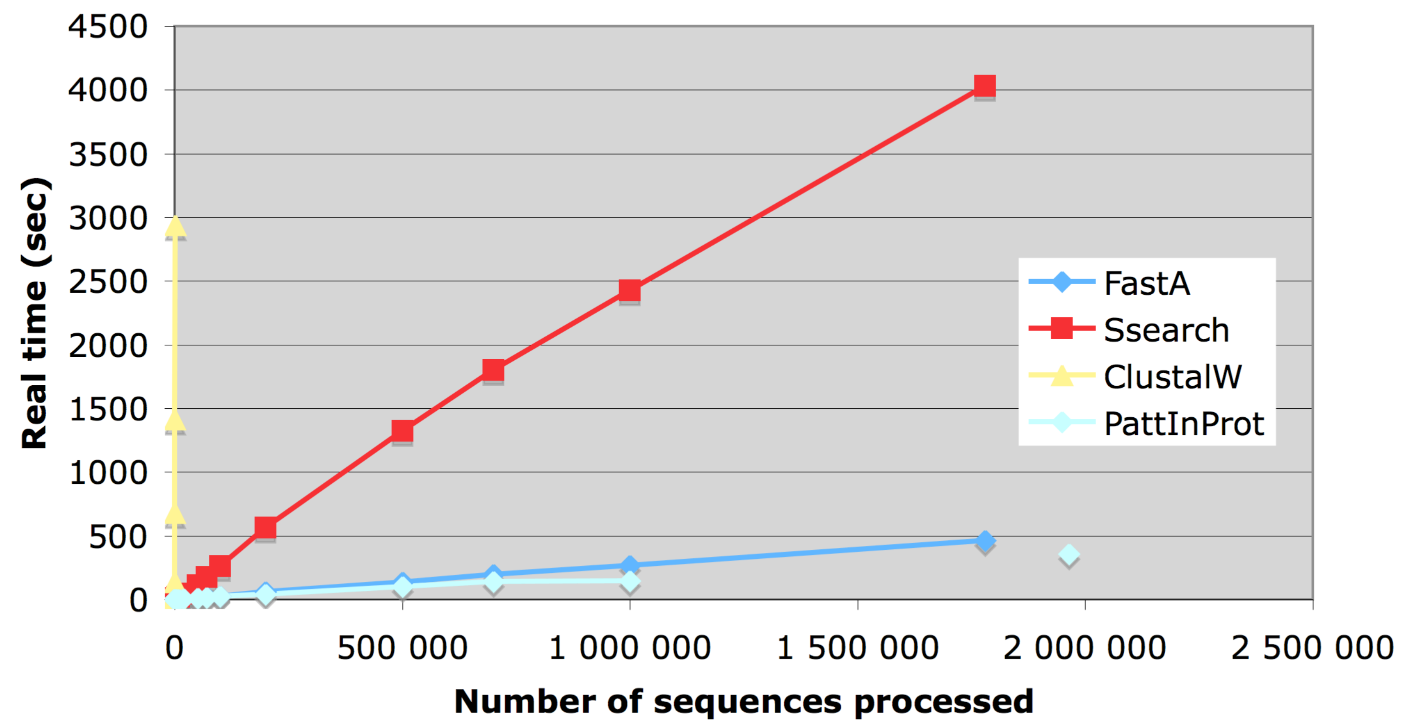
CPU- or data-intensive tools

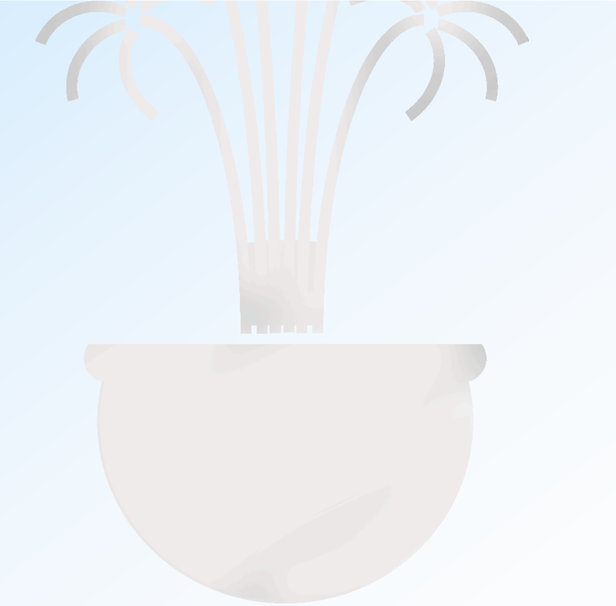
Blanchet, C., Mollon, R., and Deleage, G.. Integrating Bioinformatics Resources on the EGEE platform. ccgrid, p. 48, Sixth IEEE International Symposium on Cluster Computing and the Grid Workshops (CCGRIDW'06), 2006



Name	Algorithm	Input data
BLAST	Similarity	Gene/Protein Sequence
FASTA	Similarity	Gene/Protein Sequence
SSearch	Similarity	Gene/Protein Sequence
ClustalW	MSA	Protein Sequence
Multalin	MSA	Protein Sequence
PattInProt	Pattern/Profile	Sequence, Pattern, profile
GOR4	PSSP	Protein Sequence
SIMPA96	PSSP	Protein Sequence
SOPMA	PSSP	Protein Sequence

Local execution of bioinformatics al



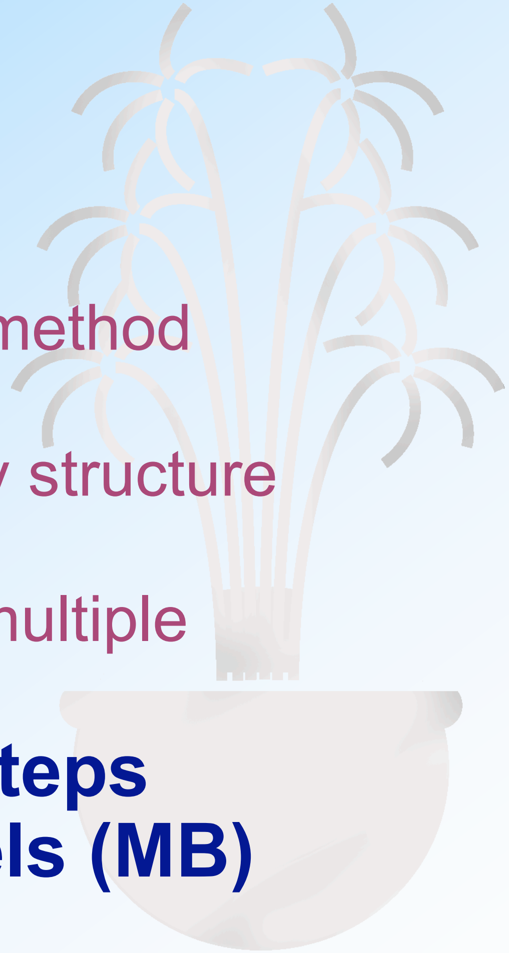


Typical usecase

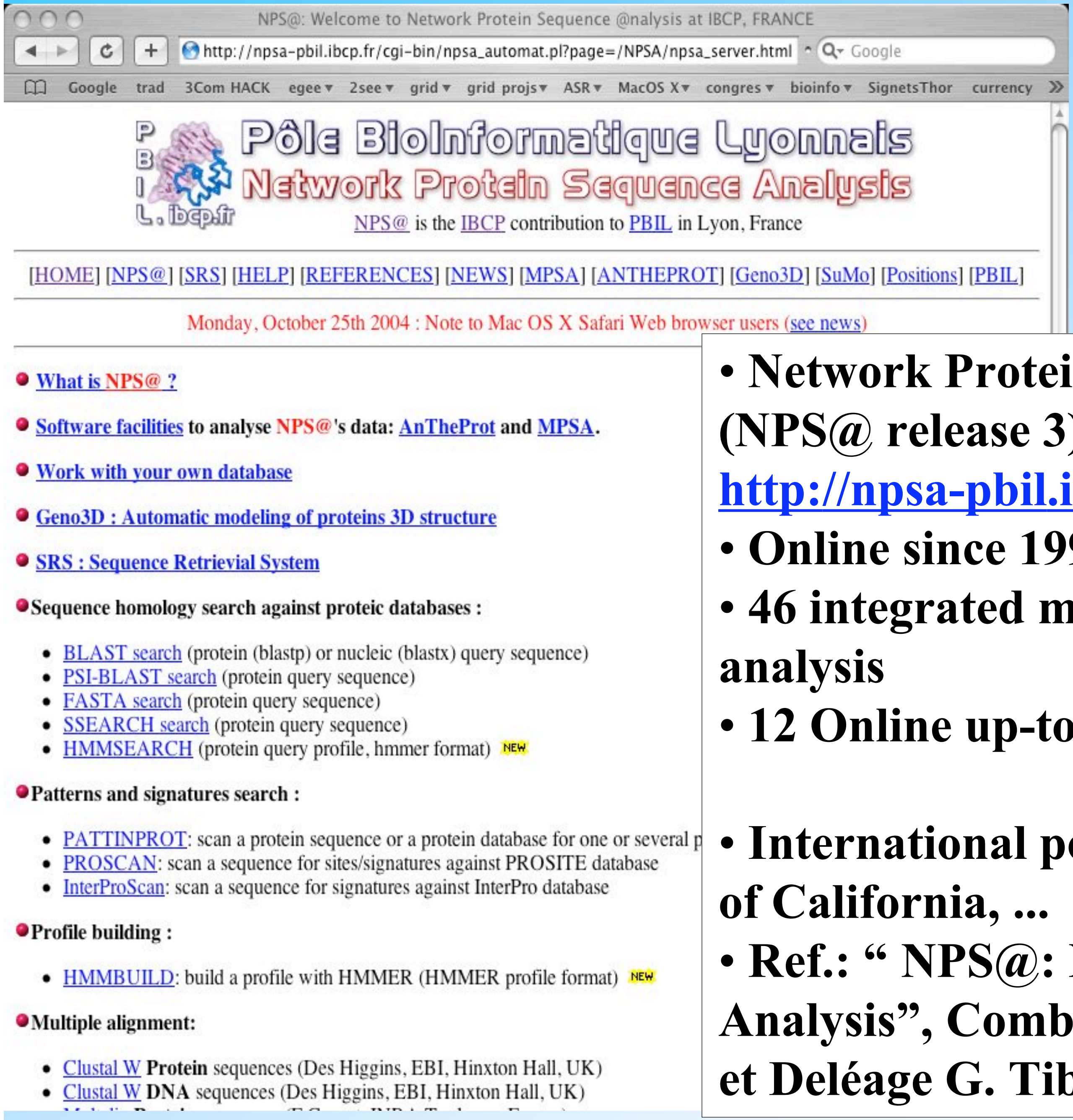
- **Example of usual bioinformatics workflow:**

1. Protein database
2. Semantic selection of sequences
   e.g.: species = human
3. Similarity selection of sequences
   e.g.: with BLAST algorithm
4. Aligning these subset of sequences
   e.g.: with ClustalW multiple alignment method
5. Validating this alignment
   e.g.: with insertion of protein secondary structure predictions : SOPMA, GOR4, PHD, …
6. Building a 3D structural model with this multiple alignement …

- **Transfering data among all these steps**
  ***e.g.* from TrEMBL (6.4 GB) to models (MB)**



NPS@: Bioinformatics Web Portal


NPS@: Welcome to Network Protein Sequence @nalysis at IBCP, FRANCE
http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_server.html
Google
Google trad 3Com HACK egee 2see grid grid projs ASR MacOS X congres bioinfo SignetsThor currency
PBIL ibcp.fr

Pôle BioInformatique Lyonnais

Network Protein Sequence Analysis

NPS@ is the IBCP contribution to PBIL in Lyon, France

[HOME] [NPS@] [SRS] [HELP] [REFERENCES] [NEWS] [MPSA] [ANTHEPROT] [Geno3D] [SuMo] [Positions] [PBIL]

Monday, October 25th 2004 : Note to Mac OS X Safari Web browser users (see news)

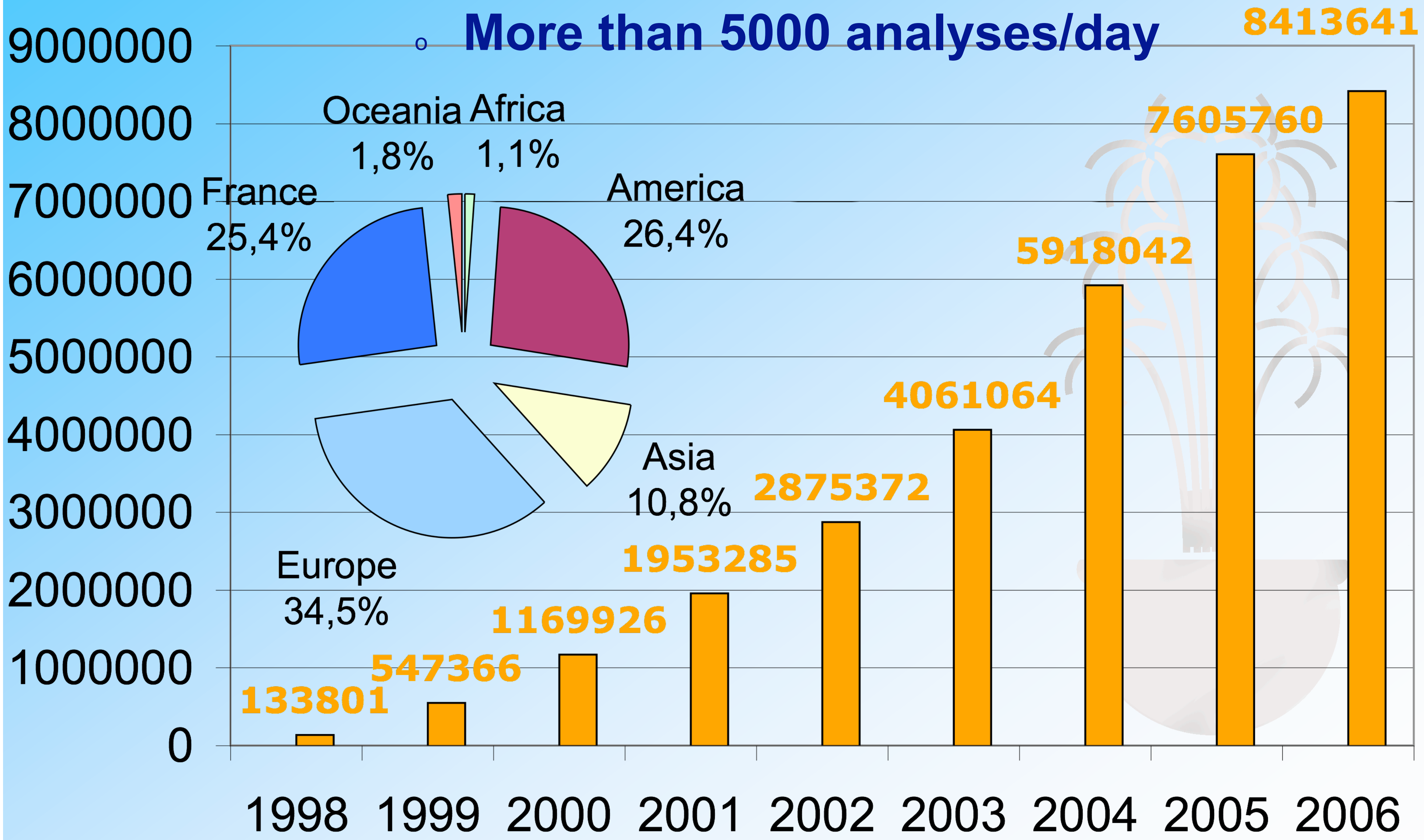
- What is NPS@ ?
- Software facilities to analyse NPS@'s data: AnTheProt and MPSA.
- Work with your own database
- Geno3D : Automatic modeling of proteins 3D structure
- SRS : Sequence Retrievial System
- Sequence homology search against proteic databases :
  - BLAST search (protein (blastp) or nucleic (blastx) query sequence)
  - PSI-BLAST search (protein query sequence)
  - FASTA search (protein query sequence)
  - SSEARCH search (protein query sequence)
  - HMMSEARCH (protein query profile, hmmer format) NEW
- Patterns and signatures search :
  - PATTINPROT: scan a protein sequence or a protein database for one or several p
  - PROSCAN: scan a sequence for sites/signatures against PROSITE database
  - InterProScan: scan a sequence for signatures against InterPro database
- Profile building :
  - HMMBUILD: build a profile with HMMER (HMMER profile format) NEW
- Multiple alignment:
  - Clustal W **Protein** sequences (Des Higgins, EBI, Hinxton Hall, UK)
  - Clustal W **DNA** sequences (Des Higgins, EBI, Hinxton Hall, UK)

• Network Protein Sequence Analysis (NPS@ release 3)
http://npsa-pbil.ibcp.fr
• Online since 1998
• 46 integrated methods for protein sequence analysis
• 12 Online up-to-date biological databanks

• International pointers: Expasy (Ch) , University of California, ...
• Ref.: “ NPS@: Network Protein Sequence Analysis”, Combet C., Blanchet C., Geourjon C. et Deléage G. Tibs, 2000, 25, 147-150.

NPS@ hits

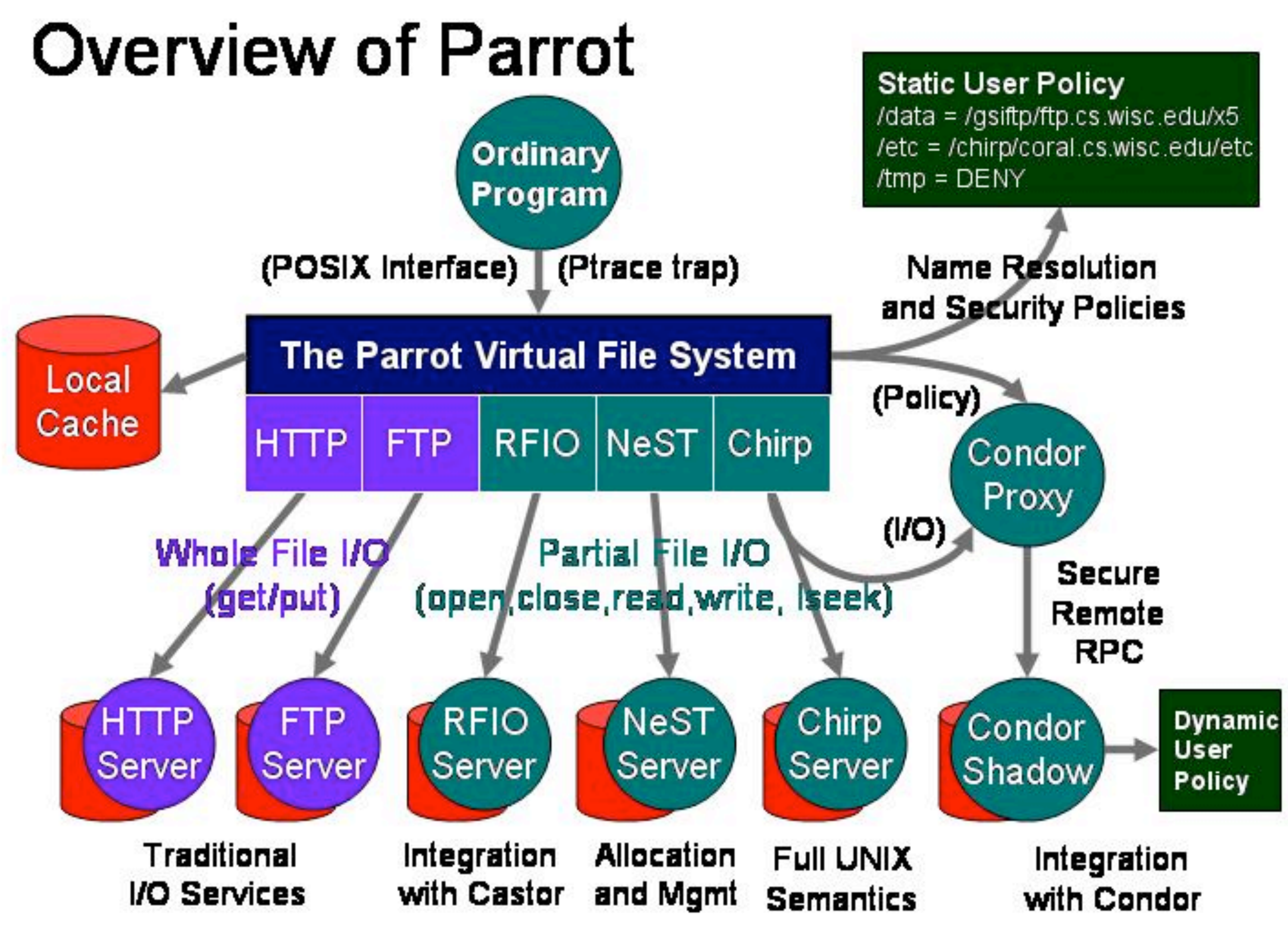
- More than 8 millions analyses since 1998
    - More than 5000 analyses/day



Outline

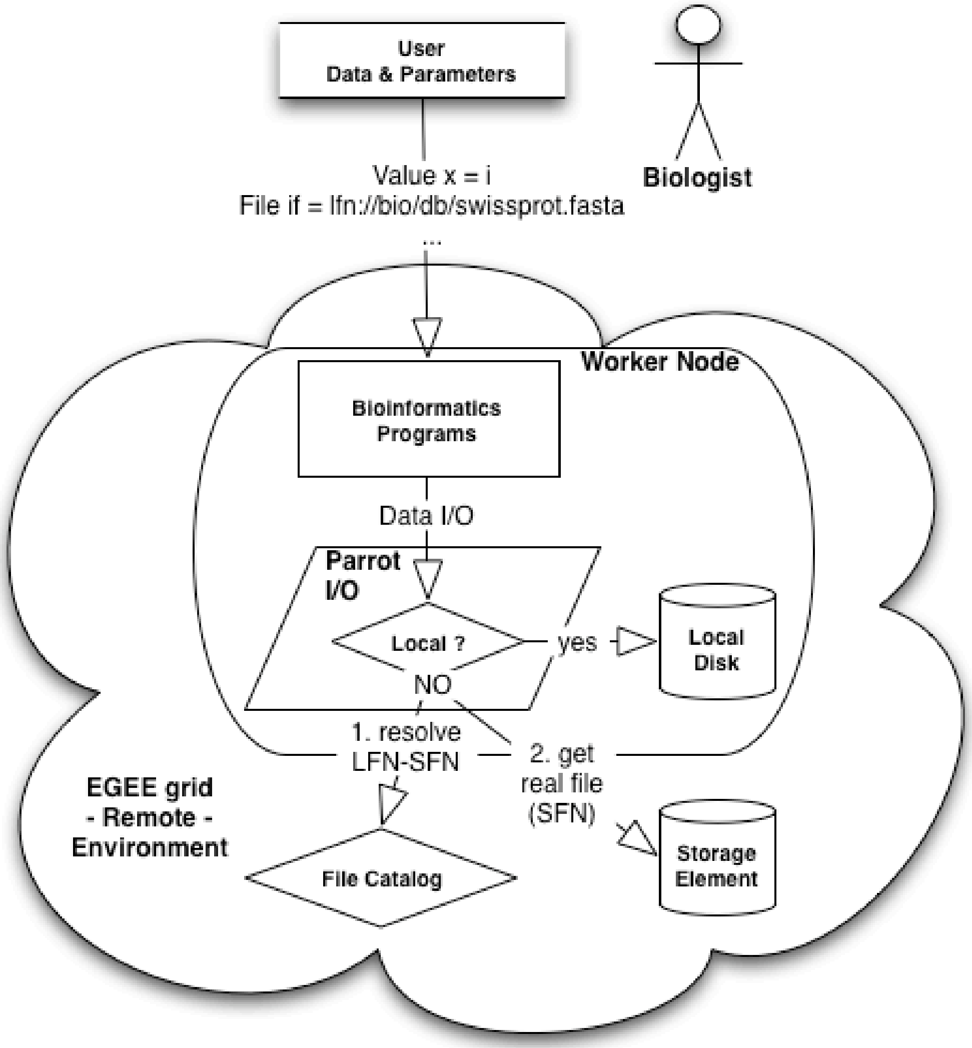
- EU-EGEE and Biomedical Applications
- GPS@: ex. of bioinformatics application
  - Biological database
  - Legacy bioinformatics tools
  - Protein sequence analysis
- Virtualization of biological data on EGEE
  - EGEE-enabling Parrot
  - Local copy vs I/O access forward
  - Application to grid Web portal

Parrot Tool



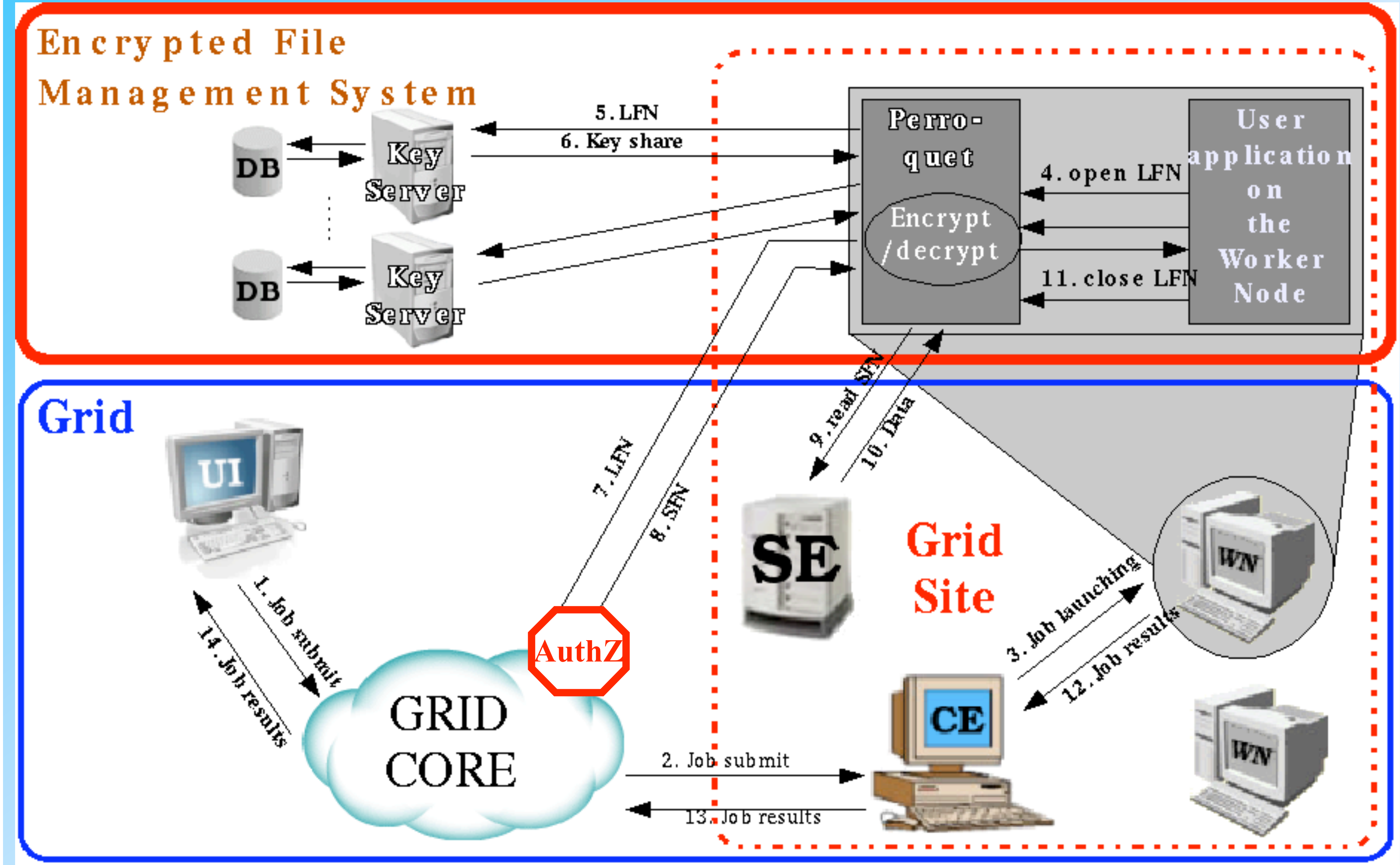
- D. L. Thain (Univ. ND, USA). http://www.cctools.org/parrot

Data Virtualization on EGEE DMS



- ***Adding LFN namespace***
  - Adding EGEE file namespace
    - *LFN recognition*
      lfn://logical/path/to/file
  - Adding EGEE name resolution
    - *Querying File Catalog (RLS, LFC)*
      SFN <-> LFN

EGEE Name Resolution



Transparent use of logical filename

- **Put a file on the EGEE Grid**
  ```
  parrot cp /local/path/to/my/file lfn:/grid/path/to/my/file
  ```

- **Get a file from the EGEE Grid**
  ```
  parrot cp lfn:/grid/path/to/my/file /local/path/to/my/file
  ```

- **EGEE-run of a BLAST on Swiss-Prot**
  ```
  parrot blastall -i my_sequence.fas -d
  lfn:/grid/biomed/db/swissprot/last/sprot.fas -o
  lfn:/grid/biomed/myspace/blast.out -p blastp
  ```

Performance tests

Biological databases

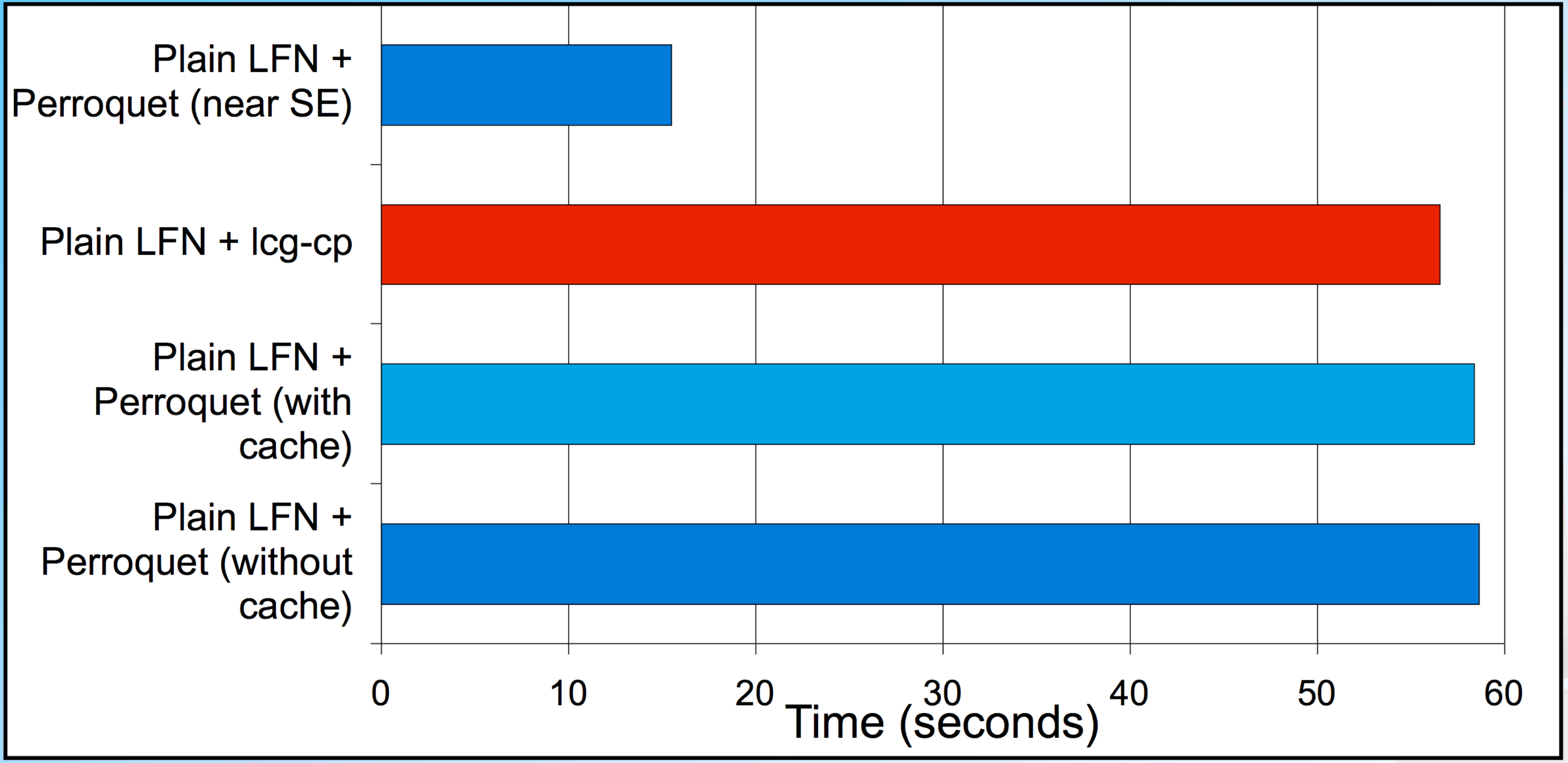
- Splitting Swiss-Prot and TrEMBL into several subsets of different size, from one-sequence file to the file containing the whole bank:
  - 184,034 entries for Swiss-Prot release 47.2
  - 1,779,481 entries for TrEMBL release 30.2.
- Deploying all the subsets onto the EGEE grid platform.
- Naming them with adequate logical filenames (LFNs) into the replica location system (RLS), and randomly replicated these LFNs on the storage elements of several grid nodes without applying any particular model of replication

Grid execution of bioinformatics tools

- Sequence similarity search
- 3 tools as models
  - *SSEarch : Exact algorithm, CPU-intensive, sequential access to file*
  - *FastA: heuristic algorithm, sequential access to file*
  - *BLAST: heuristic algorith, use 3 indexes, hidden file (no argv)*

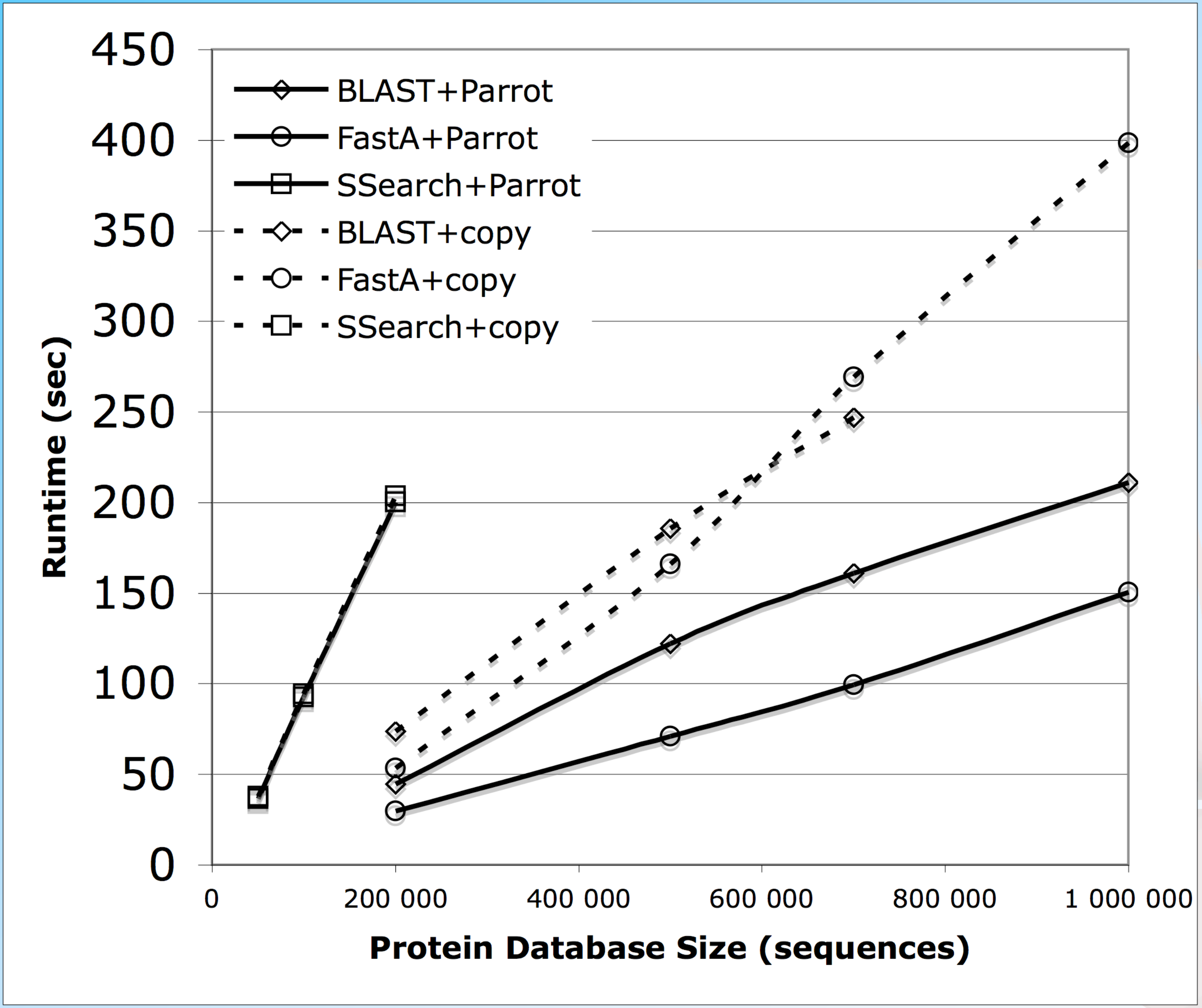
Resource	Grid Descriptor
Swiss-Prot And Blast indexes	lfn://genomics_gpsa/db/swissprot/swissprot.fasta
	lfn://genomics_gpsa/db/swissprot/swissprot.fasta.phr
	lfn://genomics_gpsa/db/swissprot/swissprot.fasta.pin
	lfn://genomics_gpsa/db/swissprot/swissprot.fasta.psq
TrEMBL	lfn://genomics_gpsa/db/trembl/trembl.fasta
PROSITE	lfn://genomics_gpsa/db/prosite/prosite.dat

Local copy vs I/O access forward (1)



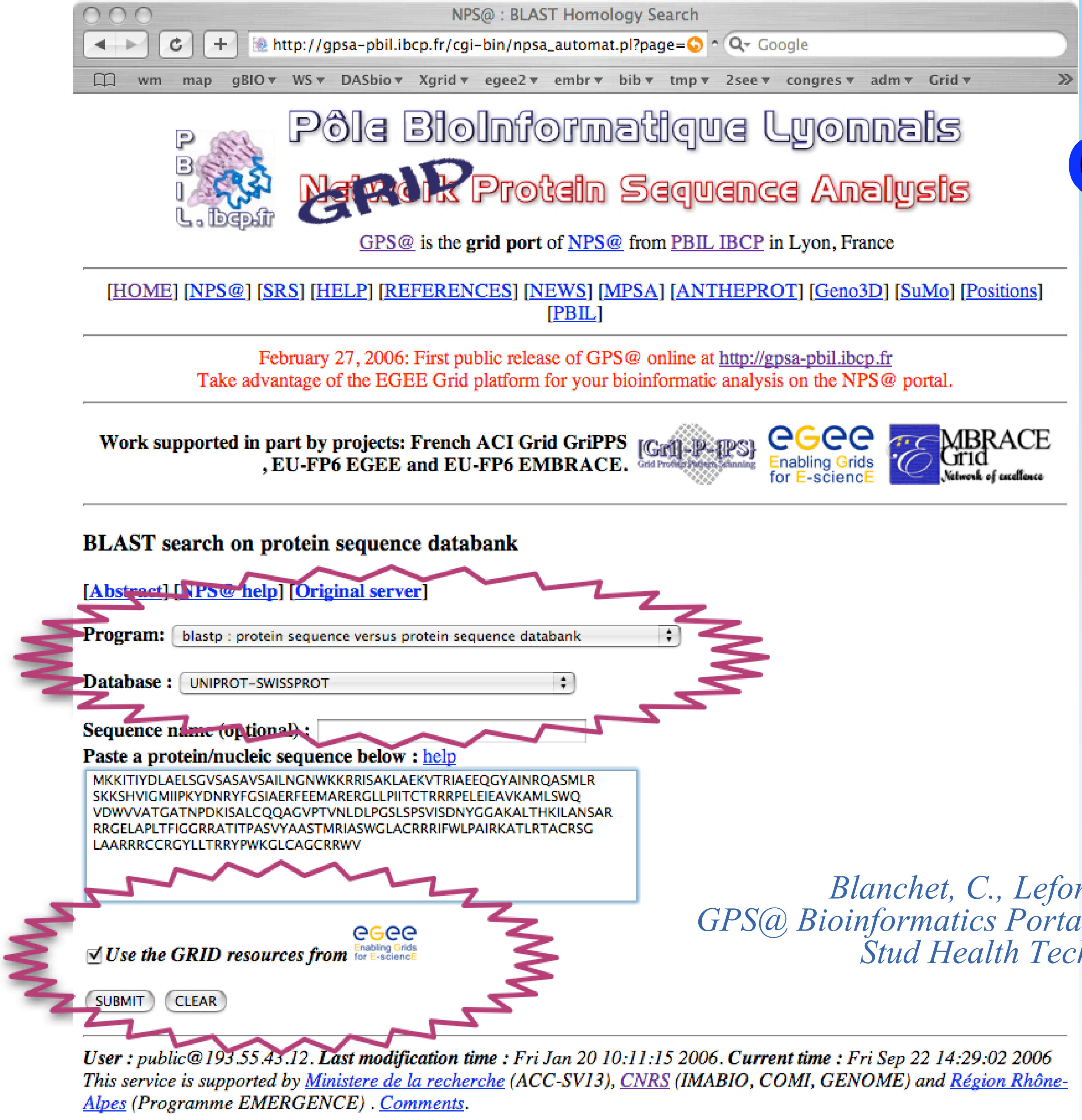
Christophe Blanchet, Rémi Mollon, Douglas L. Thain and Gilbert Deléage
Grid Deployment of Legacy Bioinformatics Applications with Transparent Data Access.
IEEE Proceedings of the International Conference GRID 2006 (GRID06), Barcelona, Sept. 28-29, 2006

Local copy vs I/O access forward (2)



Christophe Blanchet, Rémi Mollon, Douglas L. Thain and Gilbert Deléage
Grid Deployment of Legacy Bioinformatics Applications with Transparent Data Access.
IEEE Proceedings of the International Conference GRID 2006 (GRID06), Barcelona, Sept. 28-29, 2006

Grid Protein Sequence Analysis (GPS@)


NPS@ : BLAST Homology Search
http://gpsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=
Google
wm map gBIO WS DASbio Xgrid egee2 embr bib tmp 2see congres adm Grid
PBIL. ibcp.fr
Pôle BioInformatique Lyonnais
Network Protein Sequence Analysis
GRID

GPS@ is the **grid port** of NPS@ from PBIL IBCP in Lyon, France

[HOME] [NPS@] [SRS] [HELP] [REFERENCES] [NEWS] [MPSA] [ANTHEPROT] [Geno3D] [SuMo] [Positions] [PBIL]

February 27, 2006: First public release of GPS@ online at http://gpsa-pbil.ibcp.fr
Take advantage of the EGEE Grid platform for your bioinformatic analysis on the NPS@ portal.

Work supported in part by projects: French ACI Grid GriPPS , EU-FP6 EGEE and EU-FP6 EMBRACE.

[Gri]-P-[PS]
Grid Protein Patterns Scanning
egee Enabling Grids for E-sciencE
EMBRACE Grid Network of excellence

BLAST search on protein sequence databank

[Abstract] [NPS@ help] [Original server]

Program: blastp : protein sequence versus protein sequence databank

Database : UNIPROT-SWISSPROT

Sequence name (optional) :

Paste a protein/nucleic sequence below : help

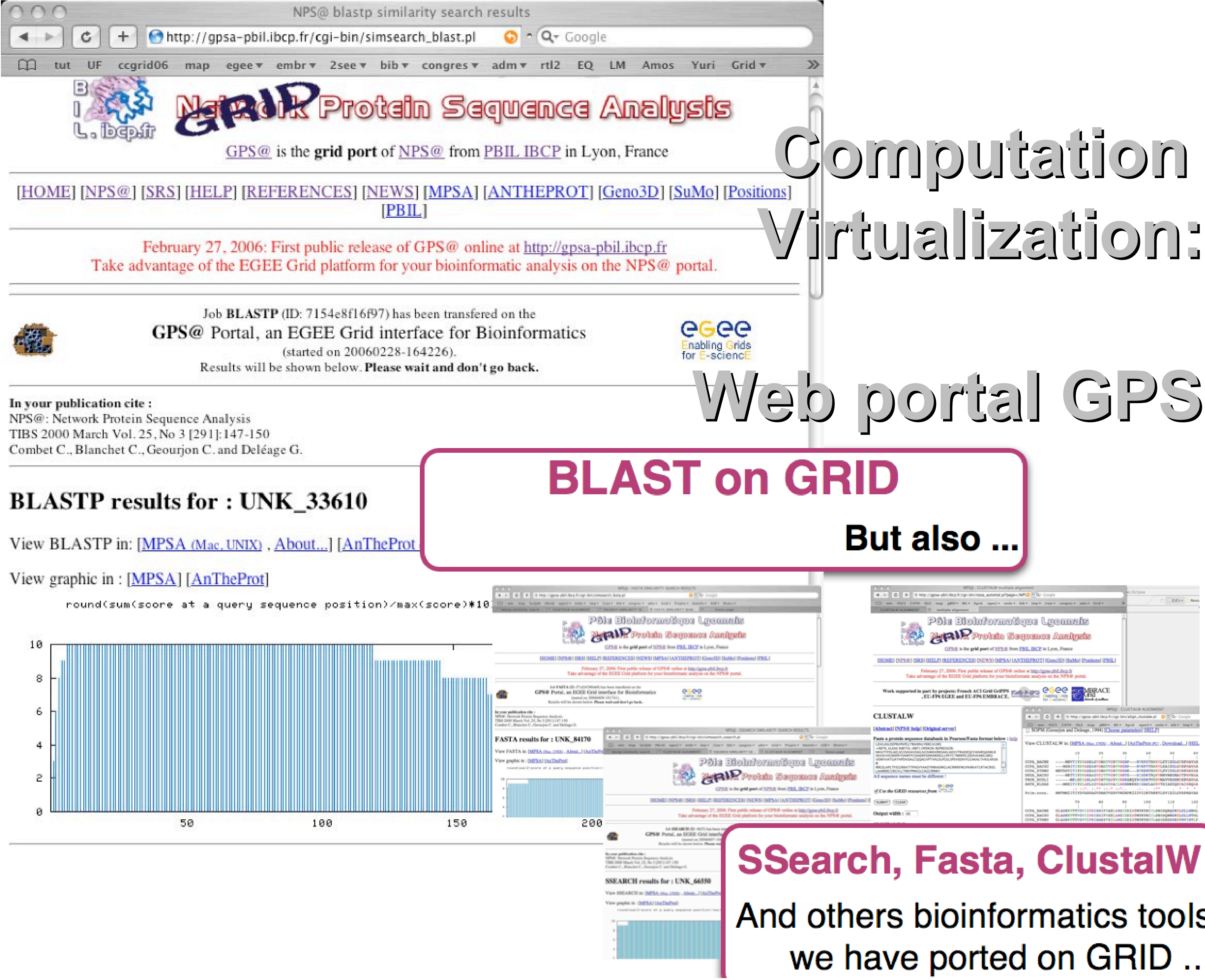
```
MKKITIYDLAELSGVSASAVSAILNGNWKKRRISAKLAEKVTRIAEEQGYAINRQASMLR
SKKSHVIGMIIPKYDNRYFGSIAERFEEMARERGLLPIITCTRRRPELEIEAVKAMLSWQ
VDWVVATGATNPDKISALCQQAGVPTVNLDLPGSLSPSVISDNYGGAKALTHKILANSAR
RRGELAPLTFIGGRRATITPASVYAASTMRIASWGLACRRRIFWLPAIRKATLRTACRSG
LAARRRCCRGYLLTRRYPWKGLCAGCRRWV
```

☑ ***Use the GRID resources from*** egee Enabling Grids for E-sciencE

SUBMIT CLEAR

User : *public@193.55.43.12.* ***Last modification time :*** *Fri Jan 20 10:11:15 2006.* ***Current time :*** *Fri Sep 22 14:29:02 2006*
This service is supported by Ministere de la recherche (ACC-SV13), CNRS (IMABIO, COMI, GENOME) and Région Rhône-Alpes (Programme EMERGENCE) . Comments.

Blanchet, C., Lefort, V., Combet, C., Deleage, G..
GPS@ Bioinformatics Portal: from Network to EGEE Grid
Stud Health Technol Inform. 2006;120:187-93.



NPS@ blastp similarity search results
http://gpsa-pbil.ibcp.fr/cgi-bin/simsearch_blast.pl
Google
tut UF ccgrid06 map egee embr 2see bib congres adm rtl2 EQ LM Amos Yuri Grid
PBIL.ibcp.fr
GRID
Network Protein Sequence Analysis
eGee Enabling Grids for E-sciencE

GPS@ is the **grid port** of NPS@ from PBIL IBCP in Lyon, France

[HOME] [NPS@] [SRS] [HELP] [REFERENCES] [NEWS] [MPSA] [ANTHEPROT] [Geno3D] [SuMo] [Positions] [PBIL]

February 27, 2006: First public release of GPS@ online at http://gpsa-pbil.ibcp.fr
Take advantage of the EGEE Grid platform for your bioinformatic analysis on the NPS@ portal.

Job **BLASTP** (ID: 7154e8f16f97) has been transfered on the
GPS@ Portal, an EGEE Grid interface for Bioinformatics
(started on 20060228-164226).
Results will be shown below. **Please wait and don't go back.**

In your publication cite :
NPS@: Network Protein Sequence Analysis
TIBS 2000 March Vol. 25, No 3 [291]:147-150
Combet C., Blanchet C., Geourjon C. and Deléage G.

BLASTP results for : UNK_33610

View BLASTP in: [MPSA (Mac, UNIX) , About...] [AnTheProt

View graphic in : [MPSA] [AnTheProt]

```
round(sum(score at a query sequence position)/max(score)*10
```

Computation Virtualization:

Web portal GPS@

BLAST on GRID

But also ...

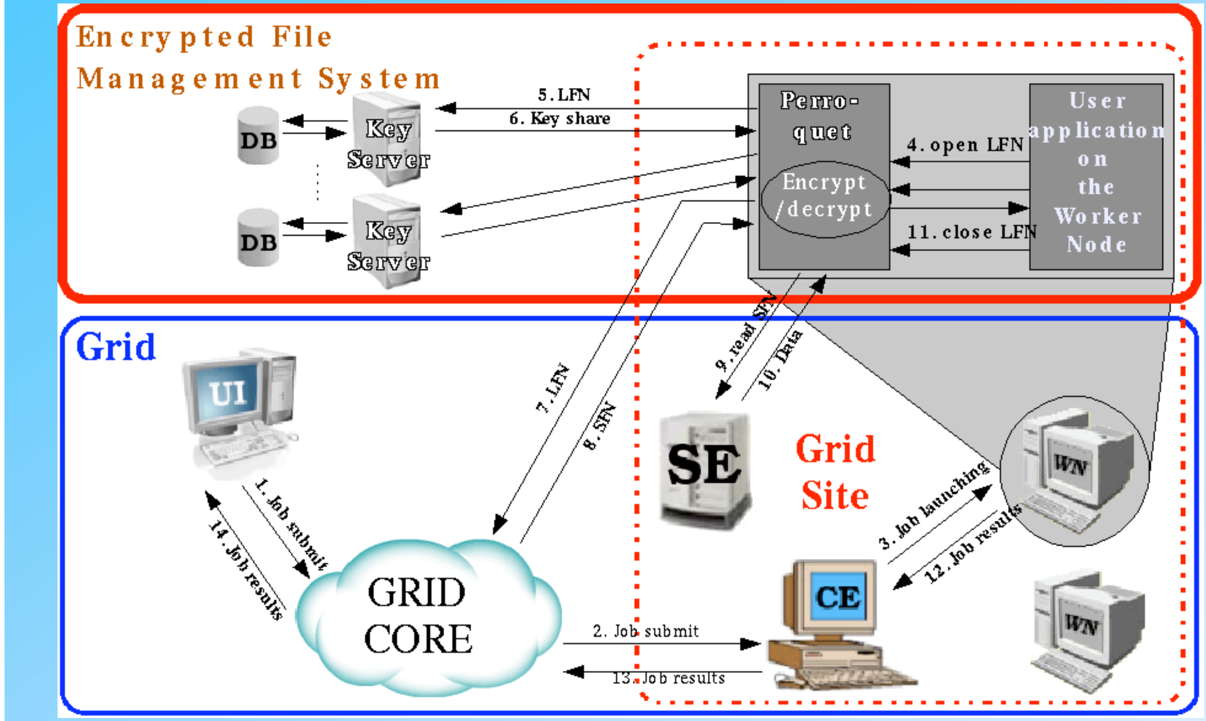
Pôle Bioinformatique Lyonnais
GRID Protein Sequence Analysis
FASTA results for : UNK_84170
CLUSTALW
SSEARCH results for : UNK_66550

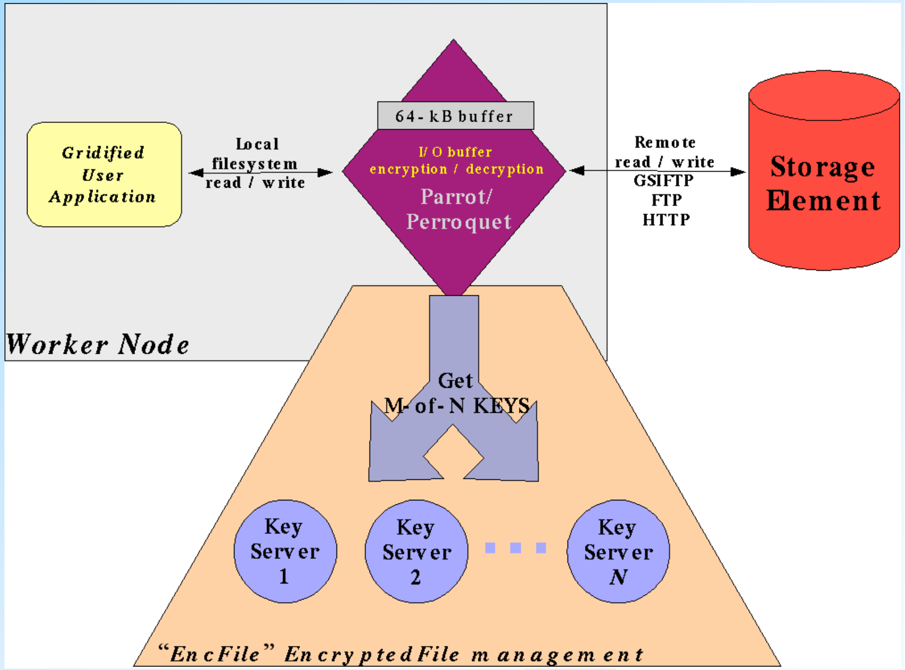
SSearch, Fasta, ClustalW

And others bioinformatics tools
we have ported on GRID ...

Secure Data Virtualization

C. Blanchet, R. Mollon and G. Deleage.
Building an Encrypted File System on the EGEE grid: Application to Protein Sequence Analysis.
IEEE Proceedings of the First International Conference on Availability, Reliability and Security (ARES'06)





Gridified User Application
Local filesystem read / write
64-kB buffer
I/O buffer encryption / decryption
Parrot/ Perroquet
Remote read / write GSIFTP FTP HTTP
Storage Element
Worker Node
Get M-of-N KEYS
Key Server 1
Key Server 2
Key Server N
"EncFile" Encrypted File management

- Put and encrypt a file on the EGEE Grid
  - parrot -e cp /local/path/to/my/clear/file lfn:/grid/path/to/my/encrypted/file
- Get and decrypt a file from the EGEE Grid
  - parrot cp lfn:/grid/path/to/my/file /local/path/to/my/clear/file
- EGEE-run of a BLAST on an encrypted database
  - parrot blastall -i my_sequence.fas -d lfn:/grid/biomed/db/encripted-db.fas -o blast.out -p blastp

Acknowledgement

Science collaborators

- D.G. Thain (Univ. ND, US)
- Y. Denneulin (IMAG, Fr)
- Members of the grid projects we collaborate with

Team collaborators

- *C. Blanchet*
- R. Mollon (EGEE fellow)
- V. Daric (EMBRACE fellow)
- C. Combet
- G. Deléage (Team Leader)

[Gri]-P-{PS}
Grid Protein Pattern Scanning

Work supported in part by projects:
French ACI GriPPS (FR-GRID-PPL02-05),
EU-FP6 EGEE-II (INFSO-RI-031688)
EU-FP6 EMBRACE (LHSG-CT-2004-512092)

EMBRACE Grid
Network of excellence

egee
Enabling Grids for E-sciencE

Questions ?

• • •