Computational Challenges at EMBL-Grenoble

Sandor Brockhauser

ESRFUP WP11 Workshop on

Exploiting the GRID

ESRF, Grenoble, France

Dec 8-9, 2008

EMBL ?

Austria 1974
Belgium 1990
Croatia 2006
Denmark 1974
Finland 1984
France 1974
Germany 1974
Greece 1984
Iceland 2004
Ireland 2003
Israel 1974
Italy 1974
Luxembourg 2007
Netherlands 1974
Norway 1985
Portugal 1998
Spain 1986
Sweden 1974
Switzerland 1974
United Kingdom 1974

EMBL Facts

- 21 Member Countries
- 5 Facilities (Heidelberg, Hamburg, Cambridge, Rome, **Grenoble**)
- ~100 Research Groups
- >1440 Employee from 60 nations
- Ranking as the highest non-US institute in research performance by ISI Science Indicator for 1992-2002

EMBL Units

- **Cell Biology**- Heidelberg
- **Biophysics** - HD
- Developmental Biology - HD
- Gene Expression - HD
- **Computational Biology** - HD
- **Structural Biology** – HD, HH, GR
- **Bioinformatics** - Cambridge
- Mouse Biology - Rome

- **Center for Disease Mechanisms**
- **Center for High Throughput Functional Genomics**
- **Center for Molecular and Cellular Imaging**

EMBL Facilities

- Advanced Light Microscopy, Genomics, Proteomics, Protein Expression and Purification, Electron Microscopy, Flow Cytometry, Monoclonal Antibody and Chemical Biology

- Microcomputing and data acquisition (detector electronics for high speed time-resolved synchrotron radiation scattering experiments; use of microstructures to carry out work on single cells; and image processing methods for biological applications)

EMBL Software Solutions

- EMBL-EBI Dbs: EMBL-Bank – *DNA/RNA sequence*;
  UniProt – *protein sequence*;
  Ensemble – *genomes*;
  EMSD – *wwPDB*;
  ArrayExpress – *gene expression*

- EMBL-EBI Tools: >200, eg: SMART; STRING; ELM; Harvester;...

- Hamburg Tools: ABRA – *EXAFS*;
  ARP/wARP; Autow-Rickshow; BEST; XREC – *MX*;
  ATSAS – *SAX*;

- Heidelberg Tools: BLAST2GENE;
  4DXpress;
  MEPD;
  SIRW;
  Genes2Diseases
  ...

Job Types

- Small data – long run: *ideal for job submission*
  SAX iterations
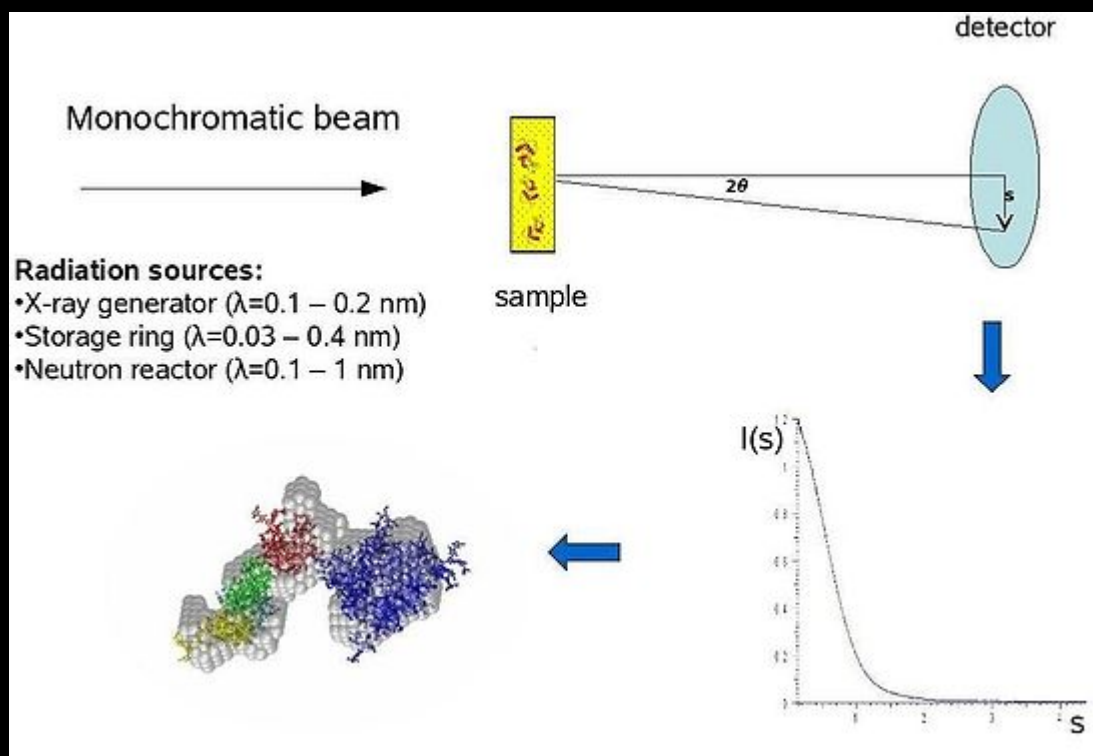
Job Types

- Huge DB + small data – long(?) run: *dedicated servers with fast DB access (DB replica/indexing for job submission?)* sequence search; Molecule Replacement
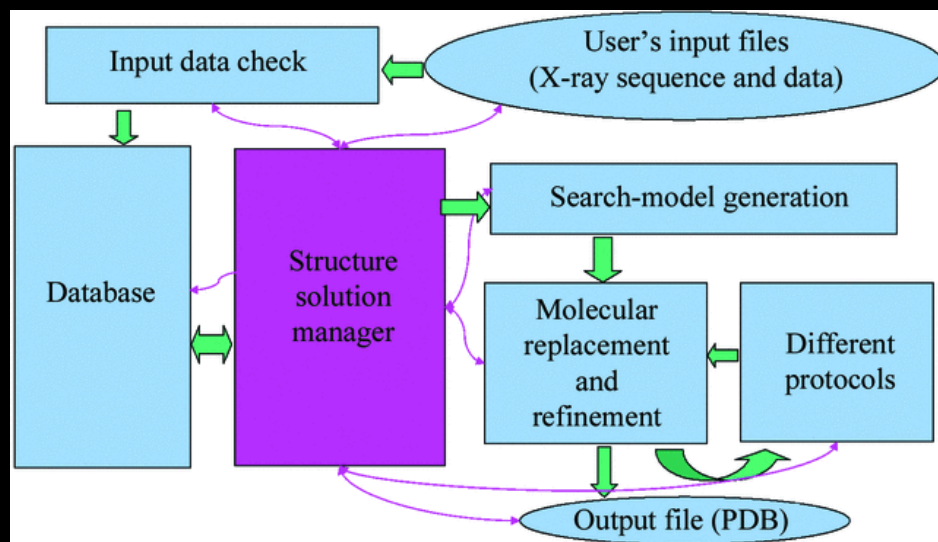
Job Types

- Huge DB + small data – long(?) run: *dedicated servers with fast DB access (DB replica/indexing for job submission?)* sequence search; Molecule Replacement
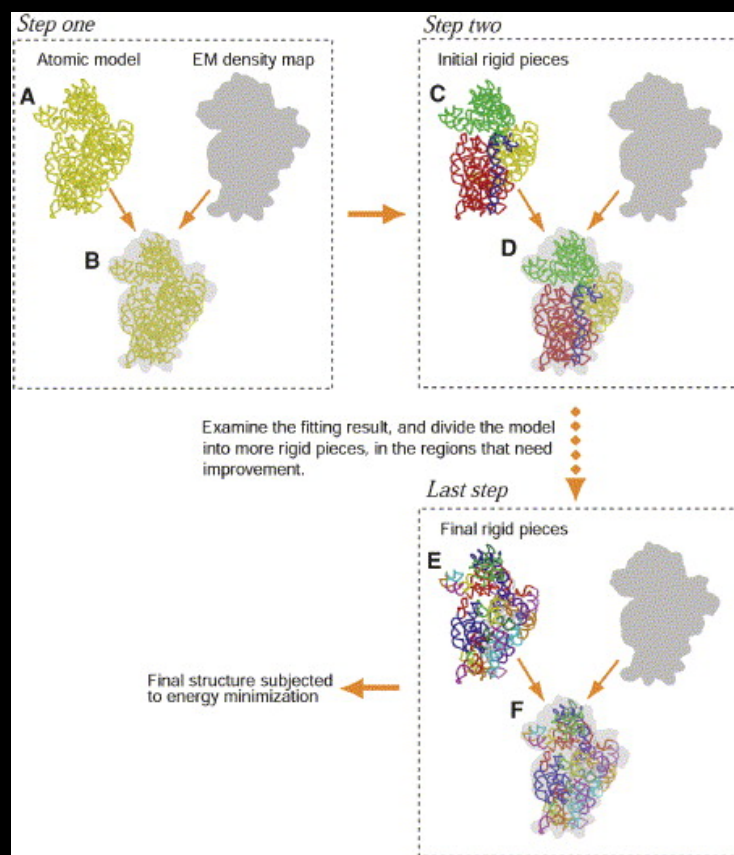


```
AAB24882    TYHMCQFHCRYVNNHSGEKLYECNERSKAFSCPSHLQCHKRRQIGEKTHEHNQCGKAFPT 60
AAB24881    -------------------YECNQCGKAFAQHSSLKCHYRTHIGEKPYECNQCGKAFSK 40
            ****: .***:   * *:** * :****.:* *******..

AAB24882    PSHLQYHERTHTGEKPYECHQCGQAFKKCSLLQRHKRTHTGEKPYE-CNQCGKAFAQ- 116
AAB24881    HSHLQCHKRTHTGEKPYECNQCGKAFSQHGLLQRHKRTHTGEKPYMNVINMVKPLHNS 98
            **** *:***********:***:**.: .***************    :   *.: :
```


BALBES
http://journals.iucr.org/d/issues/2008/01/00/ba5114/index.html

Job Types

- Modest data – long run: *good for job submission*
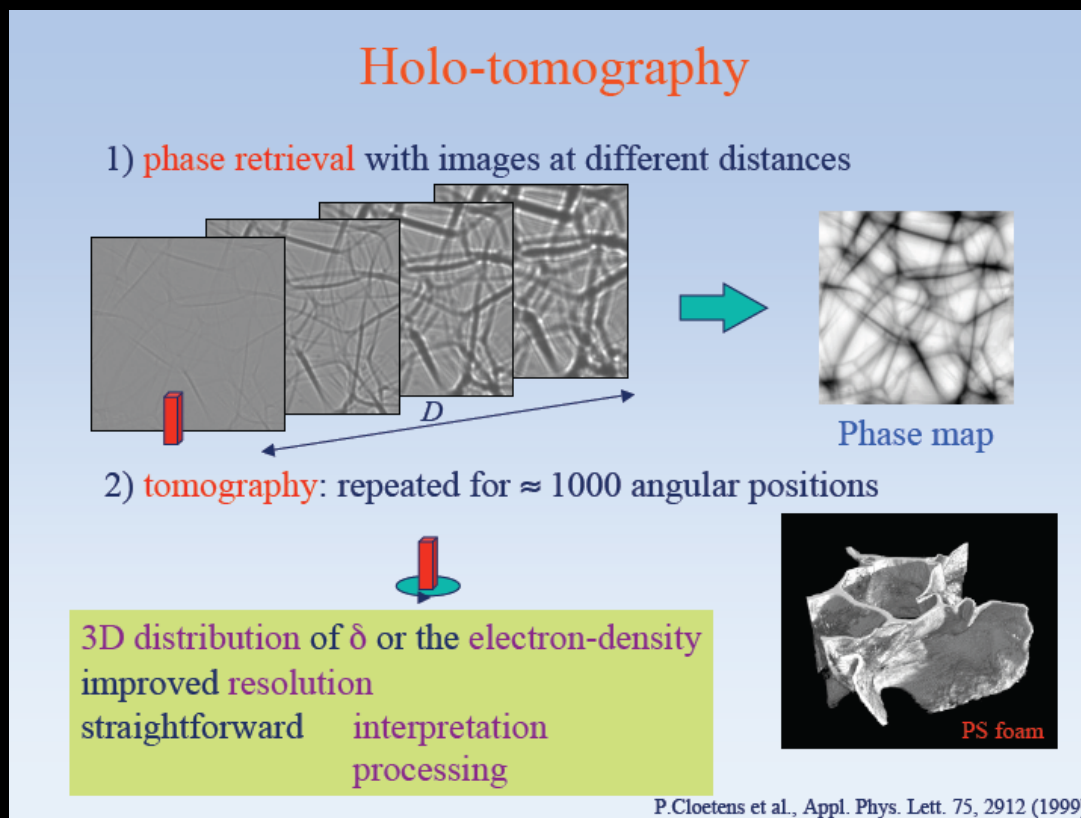 Rigid Body Fitting; EM classification / reconstruction

Job Types

- Big data – long run: *good for job submission*
  Holo-; NanoScan-Tomography

Holo-tomography

1) phase retrieval with images at different distances

Phase map

2) tomography: repeated for ≈ 1000 angular positions

3D distribution of δ or the electron-density
improved resolution
straightforward interpretation
processing

PS foam

P.Cloetens et al., Appl. Phys. Lett. 75, 2912 (1999)

Job Types

- (Modest) data – SHORT RUN: *must be submitted!*
  MX data processing
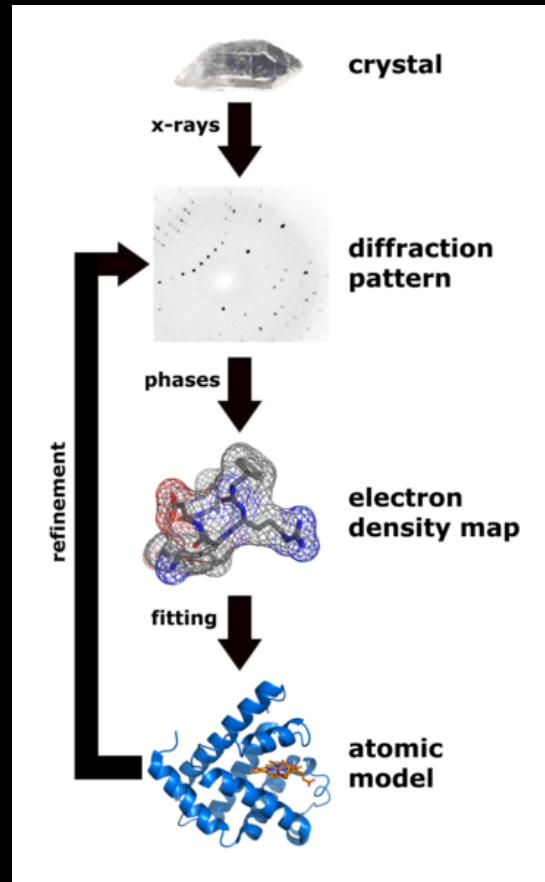
Job Types

- Small data – long run: *ideal for job submission*
  SAX iterations

- Huge DB + small data – long(?) run: *dedicated servers with fast DB access*
                          (DB replica/indexing for job submission?)
  sequence search; Molecule Replacement

- Modest data – long run: *good for job submission*
  Rigid Body Fitting; EM classification / reconstruction

- Big data – long run: *good for job submission*
  Holo-; NanoScan-Tomography

- (Modest) data – SHORT RUN: *must be submitted!*
  MX data processing

Thank you for your attention!