



E-MSD & SMART 4.0


Integrating data resources

*Maarten Altorf
Dorieke Schipper*



PDB Protein Data Bank

- Macromolecular structure data
- Protein and nucleic acid structures
- Ligands
- Maintained
 - MSD-group
 - Research Collaboratory for Structural Bioinformatics (RCSB)
 - PDBj (Japan)



MSD Macromolecular Structure Database

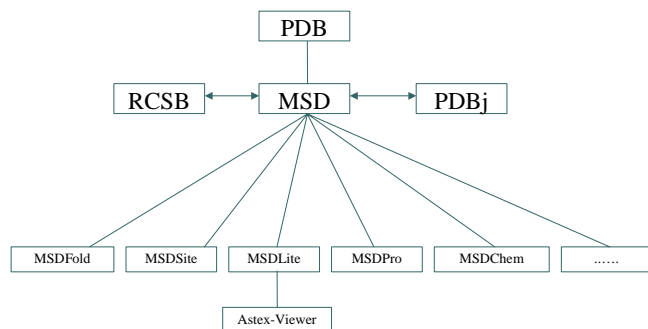
- Single access point
- Better storing, maintaining and cleaning of PDB
- Enhancing functionality
- Series of Java servlets running in the Tomcat servlet container
- Meta-data-driven query engine



Partners

- Allows MSD to enhance usefulness and scope
- Acquire more data
- Partners benefit from access to MSD
- partners among others:
 - SCOP
 - CATH
 - Swiss Prot
 - InterPro
 - Pfam
 - ProSite
 - PubMed

Overview



Search tools

- MSD Lite
- MSD Site
- MSD Pro
- Astex Viewer @ MSD-EBI

MSD Lite

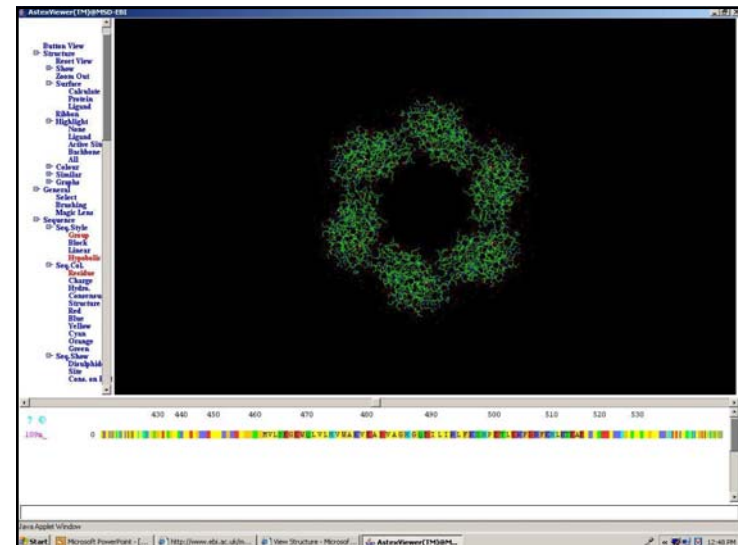
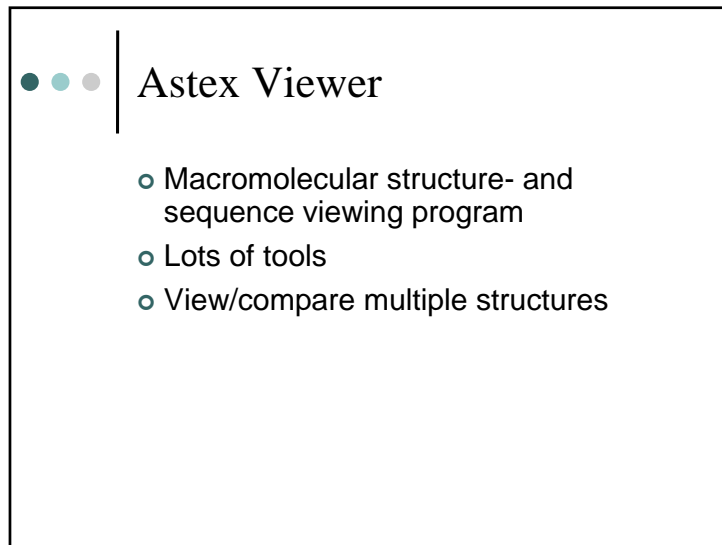
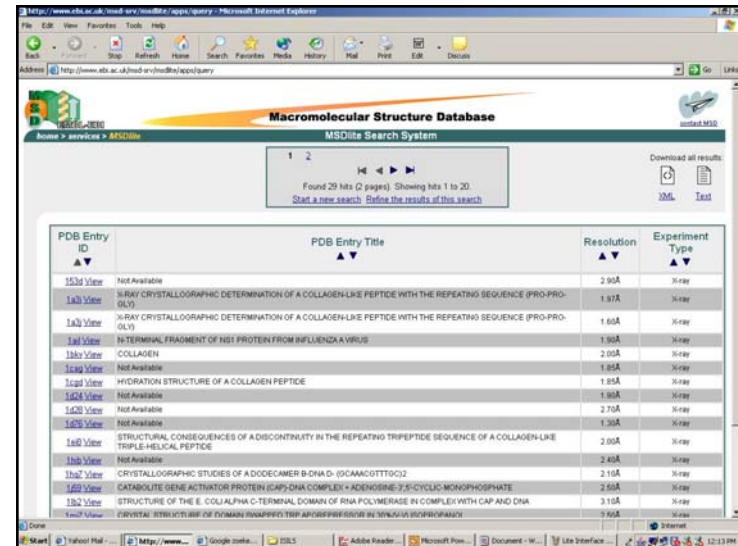
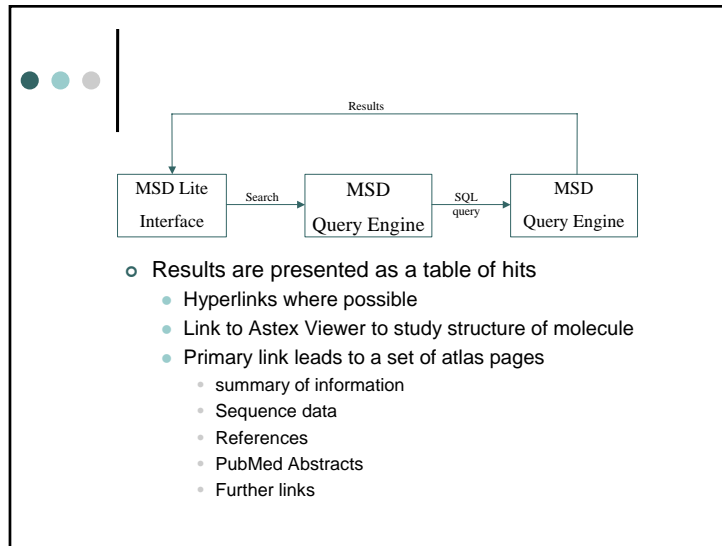
- Easy to use search tool
- Cross reference to other data resources
- Sequence similarity, FASTA
- Select result fields

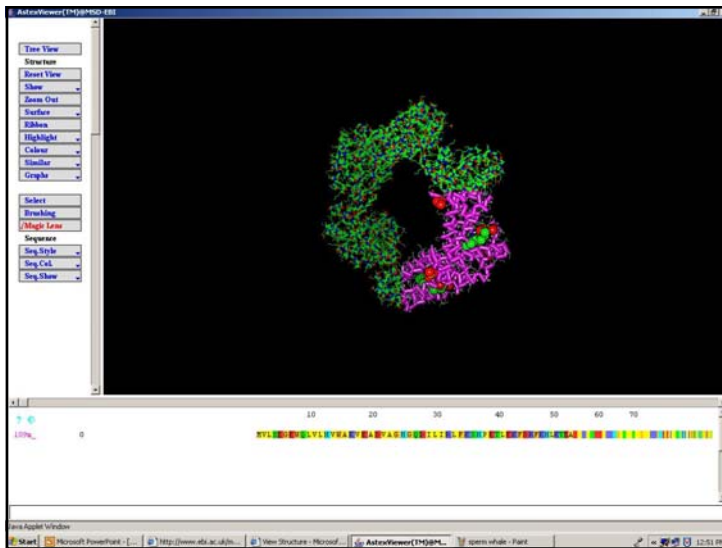
Macromolecular Structure Database
MSD Lite Search System

Search form fields:
ID code: PDB
Author Last Name: Berman
Associated small molecule: [Molecule Name]
Experiment type: X-ray, Electron microscopy, Theoretical model, Electron diffraction, NMR, Electron tomography, Fibre diffraction, Fluorescence transfer, Infrared spectroscopy, Neutron diffraction, Powder diffraction (X-ray), Solid state NMR, Other method

Results section:
Results per page: 20
Start search, Reset form, New search

Last database update: 29-SEP-2004



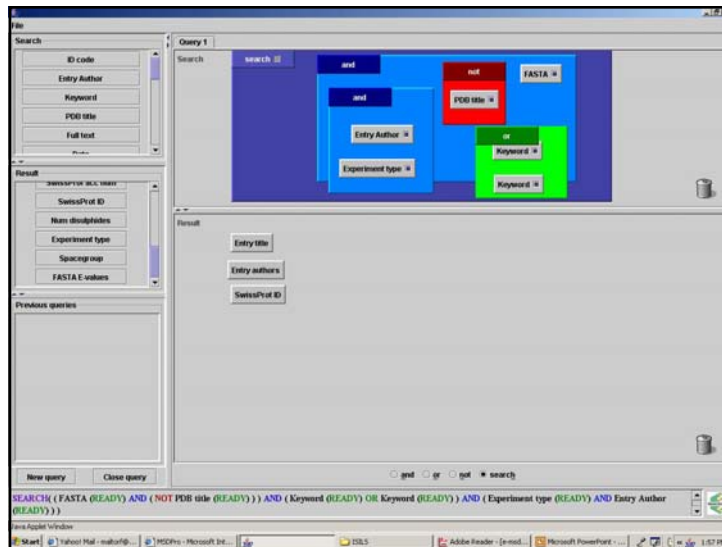


MSD Site

- Form based interface
- Interaction between macro molecules and bound ligands
- Can be used as identification tool for molecules of unknown function
- Interactions categorized by type
- Extract statistics describing similar interactions throughout PDB

MSDpro

- Specific biological question
- Interactive Java applet
- Query not mapped directly to tables or columns of the database
- Data known to expert

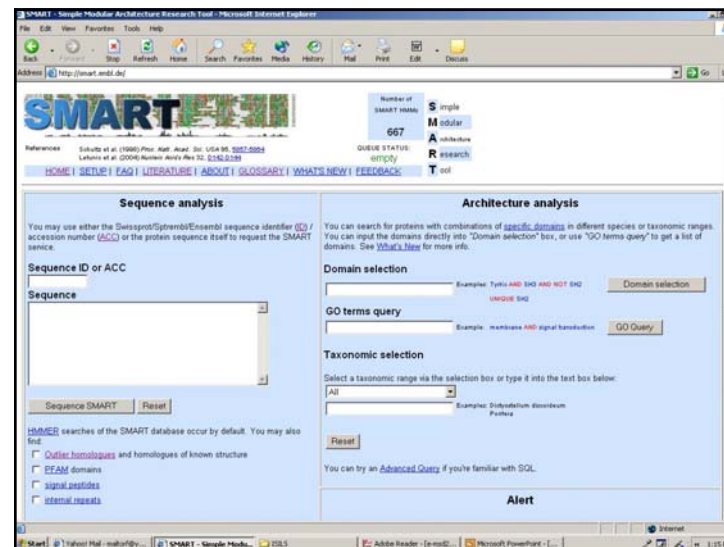


Future plans

- Download database and installation on third-party systems
- Application Programming Interface to extract data programmatically

SMART 4.0

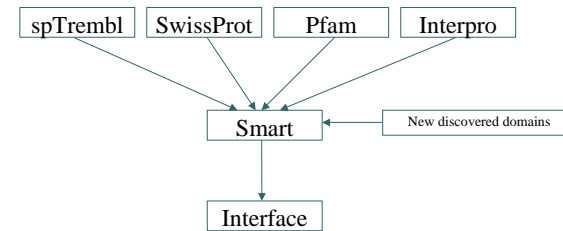
- Simple Modular Architecture Research Tool
- Identification and annotation of protein domains



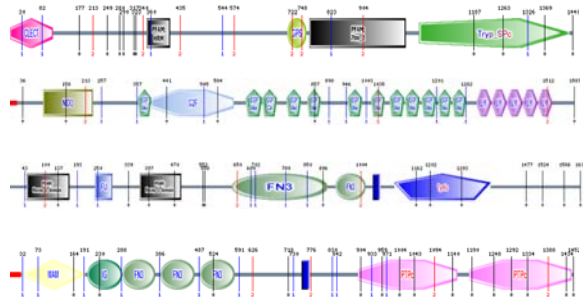
What is new?

- Centered on integration of data from completed metazoan genomes
- New visualization tools to allow analysis of gene intron-exon structure within the context of protein domain structure.
- Improved structure database searching
- Batch retrieval of multiple entries
- Query by Gene Ontology terms

Data integration



Visualization Tools



Conclusions

- Differences between SMART and MSD
 - SMART DB contains lots of data, while MSD contains only meta-data
 - MSD uses multiple interfaces, while SMART tries to integrate all functions into one Interface



Questions

- Visualizations in SMART
- Is data copied into the SMART-database or does SMART also store links