

Protein Prospector as a Large Scale MS Proteomics Data Management Tool

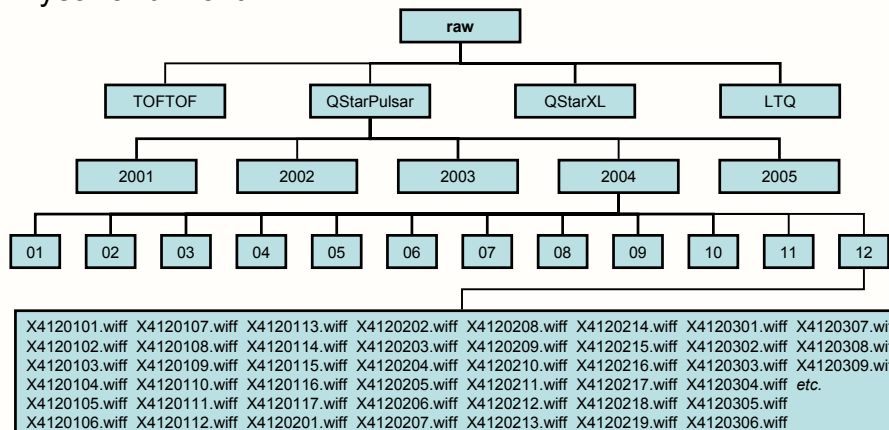
Aenoch J. Lynn¹, Peter R. Baker¹,
Yinhe Chang², Robert J. Chalkley¹,
and Alma L. Burlingame¹

¹ Mass Spectrometry Facility, Dept. of Pharmaceutical Chemistry,
University of California, San Francisco, USA

² IBM Systems Group, Austin, TX, USA

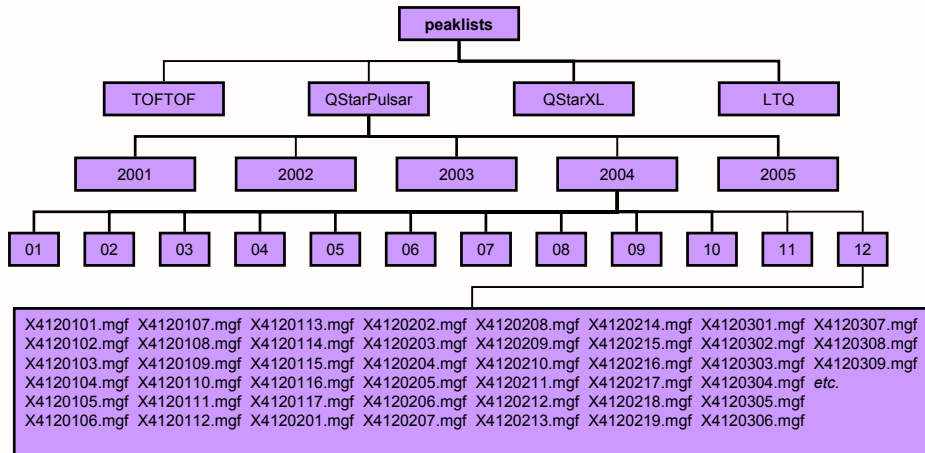
Central Data Store

The raw data files from an experiment are automatically transferred to a centrally accessible directory hierarchy after acquisition. Here they are arranged by instrument, year and month.

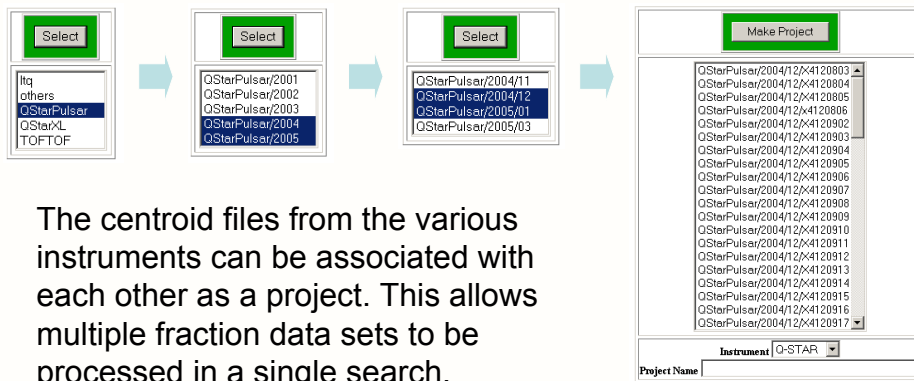


Automatic Centroiding Program

Once the raw data arrive a centroiding program running as a daemon automatically creates centroided peak list files in an equivalent directory hierarchy.



Central Data Store Data Selection



Parallel Searches

To increase speed the database search software has been rewritten to make use of the MPI parallel processing library. We have tested this on Windows, Macintosh and UNIX platforms.

Both cluster and SMP configurations are possible.

The spectra are split in both a parallel and serial manner so that each process is typically working on 1000 spectra at a time. This prevents the process memory size from getting too large.

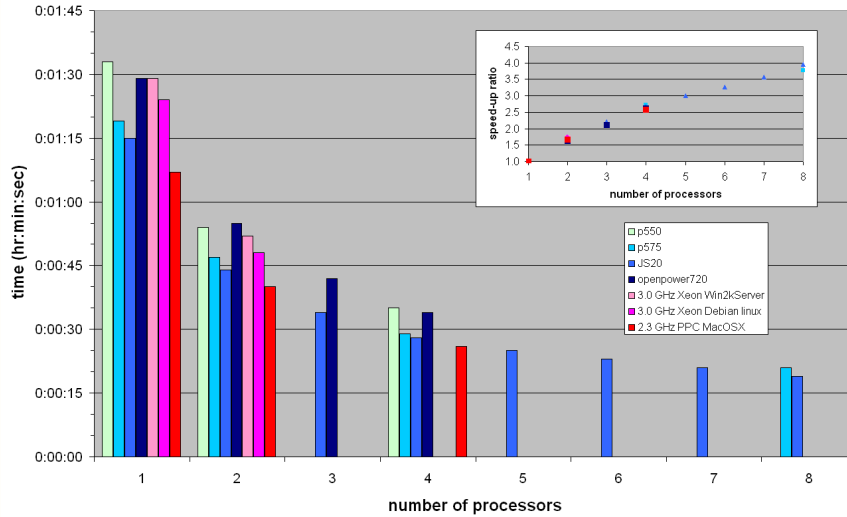
A typical results file containing the top 5 hits from 116000 spectra is around 50 Mbytes.

Computers in Parallel Search

Machine	Processor	Speed	CPU/node	OS	Compiler
IBM p550	POWER5	1.65 GHz	4	SUSE SLES9	xlC: IBM XL C/C++ Advanced Edition v7.0 for Linux
IBM p575	POWER5	1.90 GHz	8	SUSE SLES9	
IBM JS20	PowerPC	2.20 GHz	2	SUSE SLES9	
IBM openpower720	POWER5	1.65 GHz	4	SUSE SLES9	
Xeon Win2kServer	Intel Xeon	3.00 GHz	2	Win2KServer	Visual C++ v6.0
Xeon Debian Linux	Intel Xeon	3.00 GHz	2	Debian Linux (3.1)	g++ (GCC) 3.3.5 (Debian 1:3.3.5-8)
Apple XServe	PowerPC	2.30 GHz	2	MacOSX (10.3)	g++ (GCC) 3.3 20030304 (Apple Computer, Inc. build 1666)

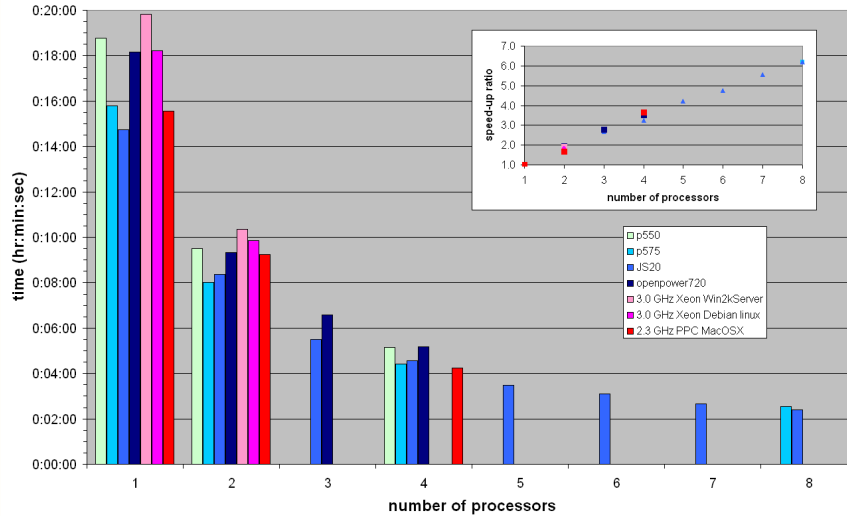
Search 1 Benchmarks

271 spectra, Swiss Prot, 2 missed cleavages, Parent tol 200 ppm, Fragment tol 300 ppm



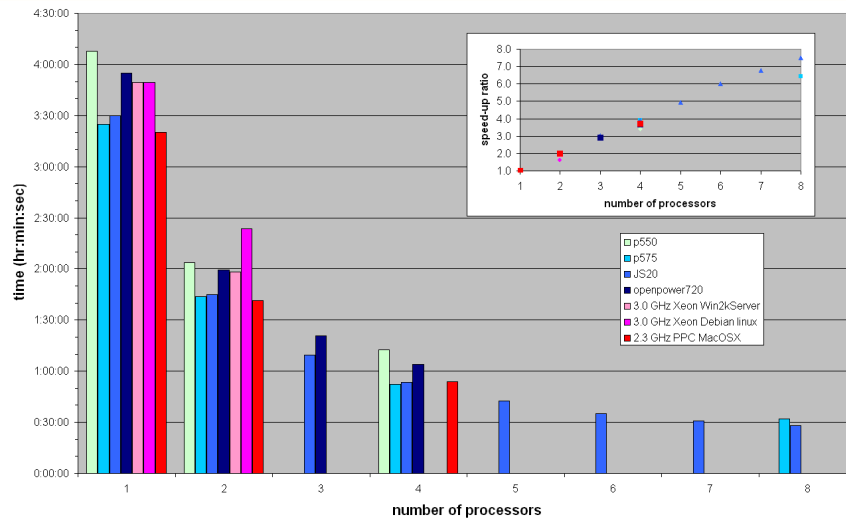
Search 2 Benchmarks

3269 spectra, Swiss Prot, 2 missed cleavages, Parent tol 200 ppm, Fragment tol 300 ppm



Search 3 Benchmarks

Search 3: 86604 spectra, Swiss Prot RODENT, STY Phosphorylation, 3 missed cleavages, Parent tol 150 ppm, Fragment tol 300 ppm



Acknowledgements

- Funding: NIH NCRR grant 01614 and the Vincent Coates Foundation.