

HUPO BPP DCC Data Reprocessing Guideline

Table of Contents:

1. Single Protein or low protein mixtures
(e.g. separated proteins by gel electrophoresis)
 - 1.1 Processing parameters for MS (PMF) spectra
 - 1.1.1 MS Search, parameters and validation criteria
 - 1.1.2 Apply calibrant list and new search?
 - 1.1.2.1 Calibrant list generation
 - 1.2 Processing parameters for MS/MS (PFF) spectra
 - 1.2.1 MS/MS Search, MALDI parameters
 - 1.2.1.1 Peptide Scoring
 - 1.2.1.2 Protein validation
 - 1.2.2 MS/MS Search, LC ESI parameters
 - 1.2.2.1 Peptide Scoring
 - 1.2.2.2 Protein validation

Appendix A: SequestMetaScore

Appendix B: Search parameter for different mass spectrometers

Appendix C: Flow chart of data reprocessing

Appendix D: Determining false positive rate and protein assignment by a composite decoy protein database

(No combination of PFF and PMF data planned for protein identification)

HUPO BPP DCC Data Reprocessing Guideline

1. Single Protein or low protein mixtures (*eg. separated proteins by gel electrophoresis*)

1.1 Processing parameters for MS (PMF) spectra

1.1.1 MS Search, parameters and validation criteria

Search parameter

- Search engines: Mascot, ProFound,
- Variable modifications: Oxidation (Methionine)
- Fixed modifications: Carbamidomethyl (Cysteine) for IPG-Gels; Propionamide (Cysteine) for Klose- and 1D-Gels
- Cleavage enzyme: Trypsin(*KR)
- Max missed cleavages: 1
- Mass type: mono
- Mass tolerance: 0.5 Da
- Database: IPI.Decoy[mouse, human].shuffle, depending on sample organism
- Taxonomy: all taxa
- MW range:
 - 0.0 to 250.0 kDa (ProFound search algorithm)
 - 0.0, i.e. infinite (Mascot search algorithm)
- pI range: 0.0 – 14.0

- Spectra calibration with Scorebooster
 - calibrant list: default list of ProteinScape (see 1.1.2)
 - Calibration: yes
 - incl. masses explained by identified protein: no
 - removed masses:
 - automatic: yes
 - known calibrants from list: yes
 - suspected calibrants from list: yes
 - polymers: yes
 - explained by identified protein: no

Validation criteria

To evaluate the search results for an MS spectrum the established protein identification scoring system called "Metascore" is used. The best result with a Metascore higher than 90 will be set as identified ¹.

¹ Chamrad DC, Koerting G, Gobom J, Thiele H, Klose J, Meyer HE, Blueggel M., Interpretation of mass spectrometry data for high-throughput proteomics. Anal Bioanal Chem. 2003 Aug;376(7):1014-22. Epub 2003 Jul 05.

HUPO BPP DCC Data Reprocessing Guideline

1.1.2 Apply new calibrant list and new search?

To increase the identification rate of a sample or gel it is useful to repeat searches with iterative generated calibrant lists¹. The first protein search will be performed with the default calibration list of ProteinScape, two additional searches will be performed with iterative calibration lists if the criteria in 1.1.2.1 are fulfilled. The search method parameter will be the same as described in 1.1.1.

1.1.2.1 Calibrant list generation

The parameter for the generation of the calibrant lists for the additional searches are:

- participant related calibrant list
- calibration status: Good, Excellent
- mass has to be found in at least:
 - 15 % of selected spectra
 - 10 spectra

1.2 Processing parameters for MS/MS (PFF) spectra

For peptide identification the MS/MS spectra are analysed in parallel by four different search engines to benefit from each search engines particular identification strengths.

Mascot, ProteinSolver, Sequest and Phenyx search engines are used. For ESI MS/MS the identified peptide lists from each search engine are then combined by the ProteinExtractor algorithm of ProteinScape.

1.2.1 MS/MS Search, MALDI Parameters

Search parameter for MALDI TOF/TOF are mostly depending on specifications of the different mass spectrometers and sample processing steps. In (*appendix B*) a PFF search parameter list for different mass spectrometers is shown.

In contrast to LC ESI MS/MS the ProteinExtractor is not used.

Search Method:

HUPO BPP DCC Data Reprocessing Guideline

All searches with these engines are performed at once.

Selected search engines:

- ProteinSolver
- Phenyx
- Sequest
- Mascot

Search parameters:

- do not combine search results: no
combine results by search engine: no
- prepare results for ProteinExtractor: yes
- Fixed modifications: Carbamidomethyl (Cysteine) for IPG-Gels;
Propionamide (Cysteine) for Klose- and 1D-Gels or others
depending on sample preparation
- Variable modifications: Oxidation (Methionine)
- Cleavage enzyme: Trypsin (*KR) (or others depending on sample
preparation)
- Max missed cleavages: 1
- Mass spectrometer: depending on used instrument

Masses:

- Peptide/Fragment mass tol.: Depending on instrument, see
Appendix B
- Search for SILE tags: no

ProteinExtractor Method:

- None
Processing will be started afterwards manually

Database parameters:

- Database: IPI.Decoy[Human, Mouse].shuffle
- Reading frame: none
- Taxonomy: all taxa
- MW range:
 - 0.0 – 250.0 kDa
 - 0.0 kDa (Mascot)
- pI range: 0.0 – 14.0
- Protein name filter: -
- Sequence tag: -

Evaluation of results:

HUPO BPP DCC Data Reprocessing Guideline

- Result is assigned correct: no selection will be processed afterwards manually

1.2.1.1 Peptide Scoring

For every peptide search result an individual score is computed. In case of Sequest a ProteinScape internal score named SequestMetaScore is used (*see appendix A*).

1.2.1.2 Protein validation

Proteins are assigned correct if at least 2 different peptides are identified significantly with a false positive rate of max. 5%, determined by the search in the shuffled decoy composite database:

1.2.2 MS/MS Search, LC ESI parameters

Search parameter for LC ESI MS/MS are mostly depending on specifications of the different mass spectrometers and sample processing steps. In (*appendix B*) a PFF search parameter list for different mass spectrometers is shown.

Search parameters:

- prepare results for ProteinExtractor: yes
- Fixed modification: Carbamidomethyl (Cysteine) for IPG-Gels; Propionamide (Cysteine) for Klose- and 1D-Gels or others depending on sample preparation
- Variable modifications: Oxidation (Methionine)
- Cleavage enzyme: Trypsin (*KR) (or others depending on sample preparation)
- Max missed cleavages: 1
- Mass spectrometer: depending on used instrument

Masses:

- Peptide/Fragment mass tol.: Depending on instrument, see Appendix B
- Search for SILE tags: no

ProteinExtractor Method:

- Depending on search engine, see 1.2.2.1

Database parameters:

Version 2.0beta4
11/23/2005

HUPO BPP DCC Data Reprocessing Guideline

- Database: IPI.Decoy[Human, Mouse].shuffle
- Reading frame: none
- Taxonomy: all taxa
- MW range:
 - 0.0 – 250.0 kDa
 - 0.0 kDa (Mascot)
- pI range: 0.0 – 14.0
- Protein name filter: -
- Sequence tag: -

Evaluation of results:

- No automatic correct assignment, ProteinExtractor is used manual assignment of correct protein identification by the use of the decoy composite database and different parallel search engines.

1.2.2.1 Peptide scoring

(see 1.2.1.1)

Peptides identified by a search engine are further processed with ProteinExtractor if following criteria are met:

- Mascot peptide score: >20
- Sequest peptide score: >5.5
- Phenyx peptide score: to be evaluated
- ProteinSolver: -, manual evaluation afterwards

1.2.2.2 Protein validation criteria

The peptide results for each search engine are merged by ProteinExtractor to get reliable protein results:

ProteinExtractor Method:

Peptide Filter:

- Protein score calculation: peptide score depending on search engine, (see 1.2.1.1)
- Protein display: same value as for “protein score calculation”

Extracted Proteins:

- Min. number of peptides per protein: 2
- Min. protein score: 0
- Max. number of displayed proteins: 3000

HUPO BPP DCC Data Reprocessing Guideline

- Remove redundancies and proteins which can't be distinguished:
yes

HUPO BPP DCC Data Reprocessing Guideline

Appendix A: SequestMetaScore

For each peptide assignment of Sequest a SequestMetaScore is computed. This value is a combination of the following parameter:

- Rank
- Xcorrelation
- DeltaCn
- Rank SP
- percentage of matched vs. theoretical fragment ions of the candidate peptide
- percentage of matched signals vs. number of signals in the spectrum
- spectrum quality
- total spectrum intensity
- additional score points, if a signal matched to a proline is intensive

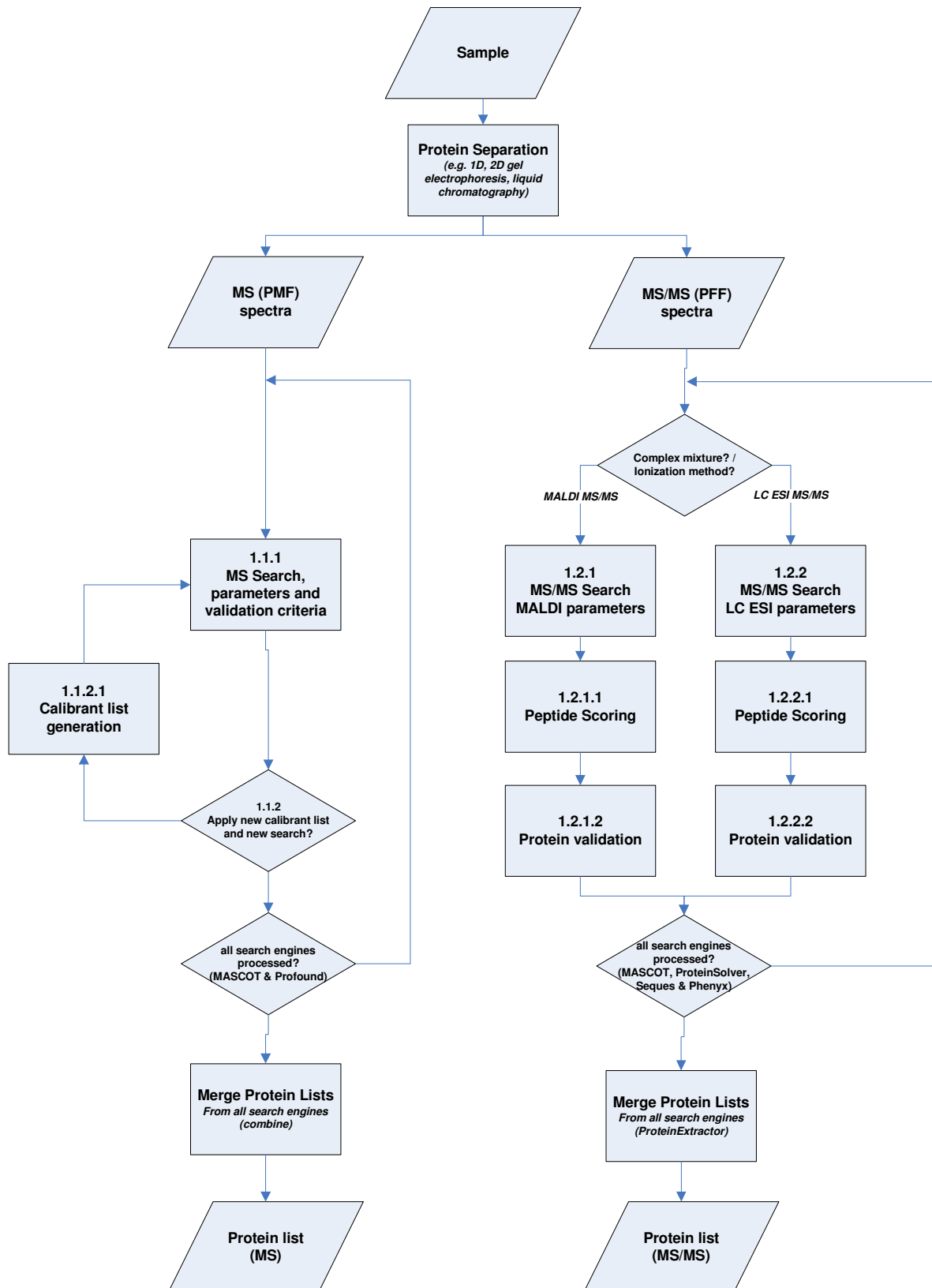
HUPO BPP DCC Data Reprocessing Guideline

Appendix B: MS/MS search parameter for different mass spectrometers

	Bruker *flex	Bruker Esquire	ABI 4700	LCQ Thermo	LTQ Thermo	Q-Star ABI	Micromass Qtof	Waters MALDI Tof
Search engines	Sequest, Mascot, Protein Solver, Phenyx							
Combine results	yes							
Fixed modifications	Carbamidomethyl or Propionamide (Cystein)							
Variable modifications	Oxidation (Methionine)							
Cleavage enzyme	Trypsin KR (or others depending on sample preparation)							
Max missed cleavages	1							
Neutral losses	b,y ion							
Mass type	mono	avg	mono	avg	mono	mono	mono	mono
Parent mass tol.	1.0 Da	1.0 Da	1.0 Da	1.5 Da	1.5 Da	1.0 Da	0.5 Da	0.5 Da
Fragment mass tol	0.5 Da	0.5 Da	1.0 Da	1.5 Da	1.0 Da	0.2 Da	0.5 Da	0.5 Da
Database	IPI.Decoy[mouse, human].shuffle							
Taxonomy	no restriction							
MW Range	infinite							
pI range	infinite							

HUPO BPP DCC Data Reprocessing Guideline

Appendix C: Data reprocessing flow chart



HUPO BPP DCC Data Reprocessing Guideline

Appendix D: Determining false positive rate and protein assignment by a composite decoy protein database

For each organism (human and mouse) a composite decoy database has been generated.

This database contains the original protein database entries as well as decoy proteins. The purpose is to determine the false positive rate of protein identification. If a decoy protein is identified by a database search it can be regarded as a false positive hit, because the decoy protein was not in the original database.

The rank of the first decoy hit can be used as a measurement for the false positive rate in regard to the true positive hits from the original database.

...