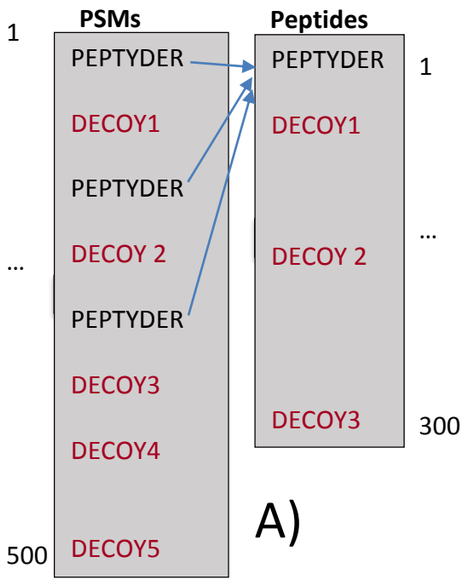



```
<SpectrumIdentificationProtocol analysisSoftware_ref="ID_software" id="SearchProtocol_1">
      <SearchType>
        <cvParam accession="MS:1001083" cvRef="PSI-MS" name="ms-ms search"/>
      </SearchType>
      <AdditionalSearchParams>
        <cvParam accession="MS:1001211" cvRef="PSI-MS" name="parent mass type mono"/>
        <cvParam accession="MS:1001256" cvRef="PSI-MS" name="fragment mass type mono"/>
        <cvParam accession="MS:1002490" cvRef="PSI-MS" name="peptide-level scoring"/>
        <cvParam accession = "MS:1002496" cvRef="PSI-MS" name="group PSMs by
sequence"/>
      </AdditionalSearchParams>
...
      <Threshold>
        <cvParam accession="MS:1001868" cvRef="PSI-MS" name="distinct peptide-level q-
value" value="0.01"/>
        <cvParam accession="MS:1002354" cvRef="PSI-MS" name="PSM-level q-value"
value="0.01"/>
      </Threshold>
  </SpectrumIdentificationProtocol>
```

B)

Protocol

Special processing = peptide-level scoring; group PSMs by sequence

Threshold = distinct peptide-level q-value

```
<Peptide id="SSHAPVPHGVRLWK">
  <PeptideSequence>SSHAPVPHGVRLWK</PeptideSequence>
</Peptide>
```

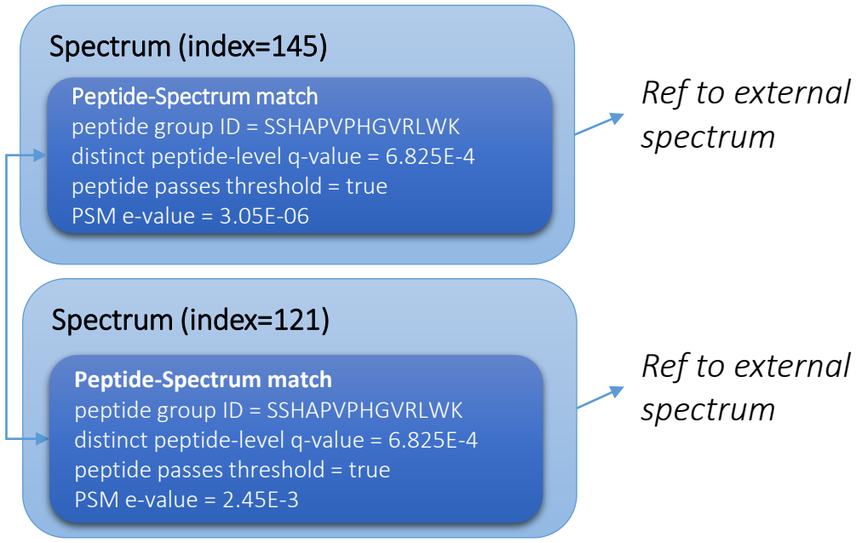
C)

Peptide

NH_2- **SSHAPVPHGVRLWK** -COOH

```
<SpectrumIdentificationResult spectraData_ref="SID_1" spectrumID="index=145" id="SIR_5">
  <SpectrumIdentificationItem passThreshold="false" rank="1" peptide_ref="SSHAPVPHGVRLWK"
calculatedMassToCharge="523.284" experimentalMassToCharge="523.194" chargeState="3" id="SII_5_1">
    <PeptideEvidenceRef peptideEvidence_ref="PE5_2_9"/>
    <cvParam accession="MS:1001328" cvRef="PSI-MS" value="3.05E-06" name="OMSSA:evalue"/>
    <cvParam accession="MS:1002520" cvRef="PSI-MS" name="peptide group ID" value="SSHAPVPHGVRLWK" />
    <cvParam accession="MS:1001868" cvRef="PSI-MS" name="distinct peptide-level q-value" value="6.825E-4"/>
    <cvParam accession="MS:1002500" cvRef="PSI-MS" name="peptide passes threshold" value="true"/>
  </SpectrumIdentificationItem>
  <cvParam accession="MS:1000796" cvRef="PSI-MS" value="55.6021.6024.3.dta" name="spectrum title"/>
</SpectrumIdentificationResult>
<SpectrumIdentificationResult spectraData_ref="SID_1" spectrumID="index=121" id="SIR_6">
  <SpectrumIdentificationItem passThreshold="false" rank="1" peptide_ref="SSHAPVPHGVRLWK"
calculatedMassToCharge="523.284" experimentalMassToCharge="523.194" chargeState="3" id="SII_6_1">
    <PeptideEvidenceRef peptideEvidence_ref="PE5_2_9"/>
    <cvParam accession="MS:1001328" cvRef="PSI-MS" value="2.45E-03" name="OMSSA:evalue"/>
    <cvParam accession="MS:1002520" cvRef="PSI-MS" name="peptide group ID" value="SSHAPVPHGVRLWK" />
    <cvParam accession="MS:1001868" cvRef="PSI-MS" name="distinct peptide-level q-value" value="6.825E-4"/>
    <cvParam accession="MS:1002500" cvRef="PSI-MS" name="peptide passes threshold" value="true"/>
  </SpectrumIdentificationItem>
  <cvParam accession="MS:1000796" cvRef="PSI-MS" value="55.6021.6025.3.dta" name="spectrum title"/>
</SpectrumIdentificationResult>
```

D)



Suppl. Figure 1. A) A graphical representation of the process by which redundant PSMs identifying the same peptide can be collapsed down into a peptide list and re-scored. Decoys are often identified by a single PSM only, and thus peptide-level FDR estimation tends to be more conservative than PSM-level FDR estimation. B) The search protocol must include a CV term to indicate that peptide-level re-scoring has been encoded, and the threshold implemented in the file for accepting or rejecting peptide identifications. C) Different PSMS reference to the same re-usable *<Peptide>* element stored in the file D) Peptide-level scores are represented in *<SpectrumIdentificationItem>* (SII) by grouping different SIIs with a shared value for "peptide group ID", shared scores and shared values for "peptide passes threshold".

A)

```
<SpectrumIdentificationProtocol analysisSoftware_ref="ID_software" id="SearchProtocol_1">
  <SearchType>
   <cvParam cvRef="PSI-MS" accession="MS:1001083" name="ms-ms search"/>
  </SearchType>
  <AdditionalSearchParams>
   <cvParam cvRef="PSI-MS" accession="MS:1001211" name="parent mass type mono"/>
   <cvParam cvRef="PSI-MS" accession="MS:1001256" name="fragment mass type mono"/>
   <cvParam cvRef="PSI-MS" accession="MS:1002635" name="proteogenomics search"/>
...
  </AdditionalSearchParams>
```

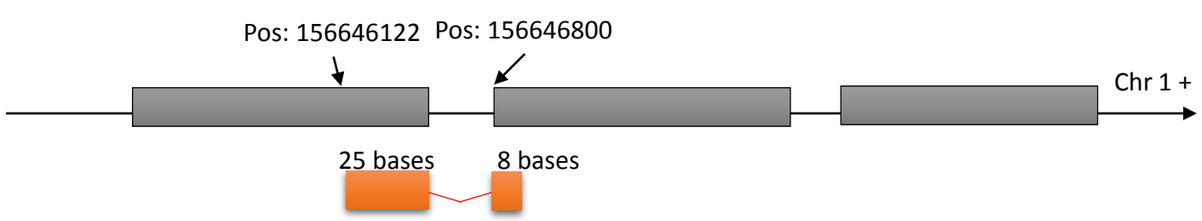
B)

```
<PeptideEvidence dBSequence_ref="dbseq_generic|A_ENSP00000354925|" peptide_ref="DVLEGDSSEDR_"
start="23" end="33" pre="A" post="A" isDecoy="false" id="DVLEGDSSEDR_generic|A_ENSP00000354925|_23_33">
  <cvParam cvRef="PSI-MS" accession="MS:1002640" name="peptide end on chromosome" value="156646808"/>
  <cvParam cvRef="PSI-MS" accession="MS:1002641" name="peptide exon count" value="2"/>
  <cvParam cvRef="PSI-MS" accession="MS:1002642" name="peptide exon nucleotide sizes" value="25,8"/>
  <cvParam cvRef="PSI-MS" accession="MS:1002643" name="peptide start positions on chromosome"
value="156646122,156646800"/>
 </PeptideEvidence>
```

C)

```
<DBSequence searchDatabase_ref="SearchDB_1" accession="generic|A_ENSP00000389898|"
id="dbseq_generic|A_ENSP00000389898|">
  <cvParam cvRef="PSI-MS" accession="MS:1002637" name="chromosome name" value="1"/>
  <cvParam cvRef="PSI-MS" accession="MS:1002638" name="chromosome strand" value="+"/>
  <cvParam cvRef="PSI-MS" accession="MS:1002644" name="genome reference version" value="Ensembl release
84"/>
</DBSequence>
```

D)



Protocol

Special processing = proteogenomics search

Referenced from PSMs (not shown)

PeptideEvidence

Peptide start positions on chrom. = 156646122,156646800
Peptide exon count = 2
Peptide end on chrom. = 156646808
Peptide exon nt sizes = 25,8

Database protein

Accession = ENSP00000389898
Chromosome name = 1
Chromosome strand = +
Genome reference version = Ensembl r 84

Suppl. Figure 2. A) If a proteogenomics approach has been followed, there is a mandatory CV term inserted into the protocol. B) There are four mandatory attributes encoding the mapping of the peptide onto the chromosome, according to a particular protein/gene product in which it can be identified. C) The chromosome strand, number/name and genome reference version are stored on the database protein record. D) A visualization of an example mapped intron-spanning peptide (in orange) onto exons is indicated in grey.