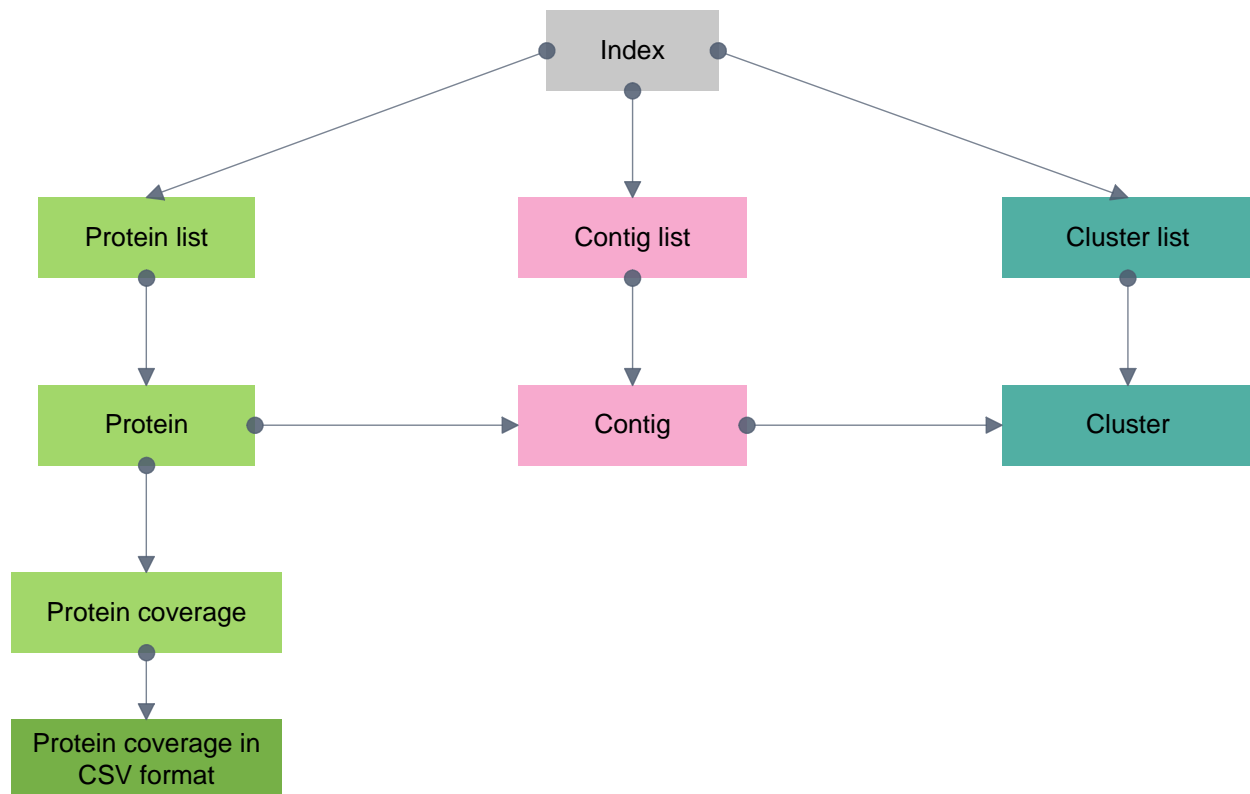


SPS results

Introduction

SPS reports are organized to reflect the hierarchical nature of the results, in the following areas:

- Input spectra: A set of spectra read from MS/MS device produced files.
- Clustered spectra: similar spectra are clustered and a consensus spectrum is defined. The clustering process determines which spectra are grouped together.
- *Contigs*: Consensus spectra that are related to the same protein are grouped together. Then an overlapping algorithm builds a contiguous large spectrum (*contig*) based on the grouped consensus spectra.
- Proteins: one or several *contigs* may map to one protein. The identified protein, along with the identified protein sequences, are shown.



Report structure

Project results

The report’s initial page, *index.html*, contains a summary and links to access the various report areas. General information, which include job name (Job), user who submitted the job (User), current report generation status (100% means the report is fully generated and can be viewed), and the time needed for report generation.

Job Status	
Job	sps
User	jcanhita
Status	100 %
Elapsed	25:47
Log	Warnings
Data	Group by Contig
	Group by Protein
Cluster Data	All Clusters (txt)
	Group by aBTLA LC AspN 042707.mgf
	Group by aBTLA LC chymotrypsin 042707.mgf
	Group by aBTLA LC pepsin 30min 042707.mgf
	Group by aBTLA LC pepsin 3h 042707.mgf
	Group by aBTLA LC trypsin 042707.mgf
	Group by aBTLA hybrid LC DTT IAA AspN ON 100407.mgf
	Group by aBTLA hybrid LC DTT IAA chymotryp 30min 100407.mgf
	Group by aBTLA hybrid LC DTT IAA chymotryp 3h 100407.mgf
	Group by aBTLA hybrid LC DTT IAA tryp 30m 100407.mgf
	Group by aBTLA hybrid LC DTT IAA tryp ON 100407.mgf

Report’s initial page

The **Log** section provides access to the report generation engine log file, where warning messages related to SPS internal works are displayed.

Through the **Data** section, the user may view directly the identified proteins or the generated *contigs*. Going further into the report hierarchy, clusters and spectra data may be accessed in context.

The **Cluster Data** section allows the user to access directly to generated clusters of spectra, and subsequently to the spectra associated with them.

Proteins

The **Group by Proteins** section contains the list of identified proteins. The table, as shown below, contains, per each protein, its name, a brief description, the number of *contigs* and spectra that lead to its identification, the number of matched amino-acids.

Protein	Description	Contigs	Spectra	Amino Acids	Coverage (%)
gi 39995102 ref NP_076376.3 	glucosaminyl (N-acetyl) transferase 2 isoform B [Mus musculus]	9	30	70	17.4
gi 29244579 ref NP_080249.2 	progastricsin (pepsinogen C) [Mus musculus]	20	83	139	35.4
gi 16716569 ref NP_444473.1 	protease, serine, 1 [Mus musculus]	18	87	133	54
gi 71274146 ref NP_001025048.1 	neuro-oncological ventral antigen 2 [Mus musculus]	19	57	185	31.7
gi 42543442 pdb 1Q9V A	Chain A, S25-2- Kdo Monosaccharide Complex	122	754	209	95.4
gi 6978717 ref NP_036668.1 	Chymotrypsinogen B [Rattus norvegicus]	24	90	191	72.6

Protein list

Selecting a protein, specific protein information is accessed. At the top of the page the protein's name can be found, as well as the number of *contigs* and spectra that lead to its identification, the number of matched amino-acids.

GI 42543442 PDB 1Q9V A

122 contigs, 754 spectra, 209 amino acids, 95.4% coverage

1

DIVMSQSPSS

LAVSAGEKVT

MSCKSSQSLL

NSRTRKNYLA

WYQQKPGQSP

51

KLLIYWASTR

ESGVDPDRFTG

SGSGTDFTLT

ITSVQAEDLA

VYYCKQSYNL

101

RTFGGGTKLE

IKRADAAPTV

SIFPPSSEQL

TSGGASVVCV

LNNFYPKDIN

151

VKWKIDGSR

QNGVLNSWID

QDSKDYSTM

SSTLTLTKDE

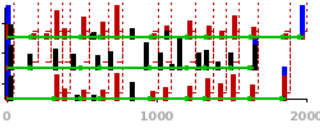
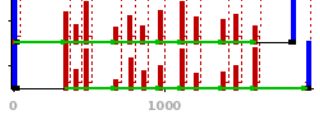
YERHNSYTCE

201

ATHKTSTSPI

VKSFNRNEC

[Protein coverage](#)

Index	Spectra	Contig	Contig Sequence	Protein
1	3	<div> <div> <div>AVSAGEKV(TM)SCKSSQ(SL)L</div> <div> <div>SAGEKV</div> <div>SCKSSK</div> <div>L</div> </div> </div>  </div>	<div>Homolog</div> <div>AVSAGEKV(TM,16.0)SCKSSQ(SL,17.9)L</div> <div>De Novo</div> <div>[170.2]SAGEKV[248.1]SCKSSK[218.0]L</div> <div> <input type="text"/> <input type="button" value="Update"/> </div>	<div>gi 42543442 pdb 1Q9V A</div> <div>Chain A, S25-2- Kdo Monosaccharide Complex</div>
3	2	<div> <div> <div>RADAAPTVSIFFPPSSEQLT</div> <div> <div>RADAAPTVSI</div> <div>FFPPSSEQLT</div> </div> </div>  </div>	<div>Reference</div> <div>RADAAPTVSIFFPPSSEQLT</div> <div>Homolog</div> <div>RADAAPTVSIFFPPSSEQLT</div> <div>De Novo</div> <div>[342.3]AA[198.2]VSLFP[184.0]S[0.3]E[342.1]</div> <div> <input type="text"/> <input type="button" value="Update"/> </div>	<div>gi 42543442 pdb 1Q9V A</div> <div>Chain A, S25-2- Kdo Monosaccharide Complex</div>

Protein page

The protein sequence is shown, with the identified amino-acids displayed in black and the not identified in gray.

Further down the page, the *contig* list associated with the protein, where each row corresponds to one *contig*. Each column contains (from left to right) the *contig* index number, the number of spectra used to generate the *contig*, the *contig* image, the *contig* sequence and the protein associated with the *contig*.

In the *contig* sequence column there is the homolog sequence and the **De Novo** sequence. The edit box allows the user to specify its own sequence and modify the report according. This functionality allows the user to specify a sequence and perform the mapping to the spectrum in context.

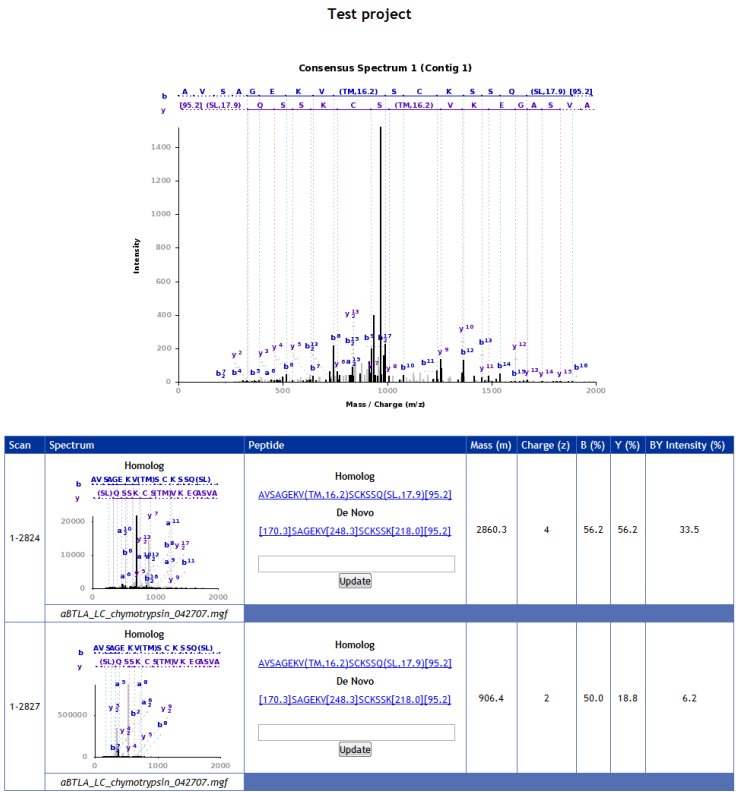
Clicking on one of the *contig* images, information about the *contig* itself is accessed.

Contigs

The *contig* information page shows, at the top, the *contig* image. The table below contains the homolog spectra used to produce the *contig*.

In each row, from left to right, are shown the spectrum index number, the homolog peptide spectrum image, the peptide sequence, the peptide mass, charge, the percentage of Y and B ions and the signal intensity.

The clusters section displays information about the cluster of spectra groups together to identify the consensus spectrum, further used to generated *contigs*.



61	E	S	G	V	P	D	R	F	T	G	S	G	S	G	T	D	F	T	L	T
CSPS 11	E	S	G	V	PD			RF		TG	S	G	S	G	T	D	F	T	L	(T, -44.03)
sps:338	E																			
sps:94	E	S	G	(VPD, -77.04)																
sps:263	E	(SG, 28.31)	(V, -28.40)	PD		R														
sps:116	E	S	G	V	PD		RF													
sps:262	E	S	GV		PD		RF													
sps:313	(S, 62.14)		GV		P	(D, -27.96)	(RFTG, -35.22)		S											
sps:14	S	G	V	PD			(RFTGSG, 74.94)				(S, 60.05)									
sps:130	SGV				P	D	RF		TG	S	G	S	G	T	D	F	T	L		
sps:8					(PD, 20.10)		(RF, -20.03)		T	G	S	GS		G	T	D	F	TL		
sps:180					DR			F	T	G	S	(GSGT, -17.87)			(DFT, -84.08)					
sps:110																		T	(L, 57.01)	(T, -44.04)

81	I	T	S	V	Q	A	E	D	L	A	V	Y	Y	C	K	Q	S	Y	N	L		
CSPS 11	I	T	SV																			
CSPS 14													YC	K	Q	SY		N		L		
sps:110	I	T	SV																			
sps:287													YC	K	Q	SY		N		L		
sps:19															K	QS	(YN, -149.07)			(LR, 87.92)		
sps:300																QS	YN			L		
sps:4																QSY		N		L		
sps:295																QS	Y		N		L	
sps:296																			NL			
sps:298																					(LR, 54.96)	

Protein coverage

The protein coverage page contains the protein sequence (shown in dark blue) and the *contigs* that were mapped to it shown in light blue. Each box around an amino-acid or around a group of amino acids means that it corresponds to a single spectrum value (mass value). If the amino acid (or group of amino acids) mass differs from the mass value, it is shown inside brackets with the mass difference.

Below the protein are shown the **csps contigs**, which are larger *contigs* generated from sets of *contigs* that overlap, and are identified on the leftmost column by “csp” followed by its index number.

In this examples are shown two groups of *contigs* (csp contigs) generated from two sets of *contigs*, which then are mapped to the protein.

Interactivity

As stated earlier, SPS allows specifying an amino-acid sequence to map to a *contig* or spectrum in context. This functionality works by sending back the specified sequence by the user to the server and rebuilding the report files by the server, which means that some time (from a few seconds to a few minutes, depending on report size and server load) is needed. After, a page refresh is needed to clear the browser cache and reload the modified page.