nature portfolio

https://doi.org/10.1038/s41586-021-03368-8

Supplementary information

Identification of bacteria-derived HLA-bound peptides in melanoma

In the format provided by the authors and unedited

Supplementary Information Guide

Identification of microbial-derived HLA-bound peptides in melanoma

Shelly Kalaora*, Adi Nagler*, Deborah Nejman, Michal Alon, Chaya Barbolin, Eilon Barnea, Steven L.C. Ketelaars, Kuoyuan Cheng, Kevin Vervier, Noam Shental, Yuval Bussi, Ron Rotkopf, Ronen Levy, Gil Benedek, Sophie Trabish, Tali Dadosh, Smadar Levin-Zaidman, Leore Geller, Kun Wang, Polina Greenberg, Gal Yagel, Aviyah Peri, Garold Fuks, Neerupma Bhardwaj, Alexandre Reuben, Leandro Hermida, Sarah B Johnson, Jessica R Galloway-Pena, William C Shropshire, Chantale Bernatchez, Cara Haymaker, Reetakshi Arora, Lior Roitman, Raya Eilam, Adina Weinberger, Maya Lotan-Pompan, Michal Lotem, Arie Admon, Yishai Levin, Trevor D. Lawley, David Adams, Mitchell P. Levesque, Michal J Besser, Jacob Schachter, Ofra Golani, Eran Segal, Naama Geva-Zatorsky, Eytan Ruppin, Pia Kvistborg, Scott N. Peterson, Jennifer A. Wargo, Ravid Straussman, Yardena Samuels

Supplementary Information Guide

Table of content

- **Supplementary Figure legends: pages 3-6**
- **Supplementary Video legends: page 7**
- **Supplementary Table legends: pages 8-14**

Supplementary File contains:

- **Methods**
- **Supplementary Figures**
- **Supplementary Note**
- **Supplementary Videos**

Supplementary Figure legends

Supplementary Figure 1: Gibbs clustering of identified bacteria-derived HLA-I peptides. Bacteria- and human-derived HLA-I peptides were Gibbs clustered to assign the peptides to the different HLA alleles of the patient. Logos were created to identify the HLA allele's motif using all the peptides that match this allele in the IEDB database. N - number of all peptides, B - number of bacterial peptides, H - number of human peptides in each cluster. The trash represents the peptides that were not assigned to any cluster. Clusters with special characteristics are marked with an asterisk.

Supplementary Figure 2: The length distribution of bacteria-derived peptides is similar to the expected length of HLA-I and HLA-II peptides. Presented is the length of bacterial (red) and human (gray) HLA-I and HLA-II peptides identified in the melanoma tumors.

Supplementary Figure 3: Tandem mass spectra of the bacteria peptides identified in melanoma tumors and their corresponding synthetic peptides. Fregmantaion images of endogenous and synthetic peptides were extracted from MaxQuant. A head to tail figure was plotted to compare between the fragmentation of the endogenous and synthetic peptides. The Pearson correlation and the dot product score are indicated for each comparison.

Supplementary Figure 4: Frequency of the HLA alleles among bacterial and human peptides. The percentages of the bacterial (y-axis) and human (x-axis) peptides that match each of the HLA alleles of the patients. Data is shown for each metastasis separately or for all the metastases of each patient together.

Supplementary Figure 5: Correlation between the percentage of identified peptides per HLA-I allele to the respective HLA-I's RNA expression level. a, The percentage of human and bacterial peptides matching HLA-A/B/C from the sum of peptides was calculated for each metastasis. Each peptide was assigned an allele with the best binding prediction score (%rank). This was compared to the relative RNA expression of HLA-A/B/C (i.e., percentage of total HLA-I expression).

HLA-A/B/C were marked red, blue and green, respectively; bacteria are marked by a darker shade, human by a lighter shade. b, The TPM values of HLA-A/B/C expression by RNA sequencing.

Supplementary Figure 6: Hydrophobicity comparison between bacterial and human peptides presented on HLA-I melanoma tumors. Bacterial peptides from the same patient (combining all the patients' metastases) or from each metastasis were more hydrophobic than human peptides (the statistical analysis used is found in the Material and Methods under "Peptide hydrophobicity"). In box plots, the centre lines represent the medians, the boxes the range between the 25th and 75th percentile, and the whiskers the range between the smallest and largest data point.

Supplementary Figure 7: CD45+/- population isolation and HLA-II expression in melanoma tumors. a, Flow cytometry analysis of tumor HG38, from which CD45+/- populations were isolated using MACS CD45 MicroBeads and stained with anti-HLA-II and anti-CD45 antibodies, indicating the efficiency of separation and HLA-II expression by the CD45- population. The top panel indicates the gating strategy used for the selection of live single cells. The last figure in this panel describes the unstained cells before separation, which were used as a negative control for the CD45 and HLA-II staining. The panel below shows the stained cells before isolation and the CD45+/- populations after the isolation based on the CD45 marker. b, Immunoblot analysis of CD45 and HLA-II expression in CD45+/- populations isolated from tumor 422.

Supplementary Figure 8: THP1 differentiation into macrophages and their HLA-I and HLA-II expression in B cells and melanoma cell lines. a, THP1 cells that differentiated into macrophages, using PMA, were stained with anti-HLA-I and anti-HLA-II and analyzed using flow cytometry. Macrophages show decreased anti-CD14 staining compared to THP1 cells. b, Images of THP1 cells and THP1 differentiated into macrophages. Images are presented at 40x magnification, scale 10 μm. Figures are representative of at least three independent experiments. c,d,

Flow cytometry analysis of IHW1070 B cells and 721.221 cells with A*01:01 OE stained with anti-HLA-I and anti-HLA-II. e, f, Flow-cytometry analysis of 51AL and 55A3 KO of B2M or CIITA and 58A melanoma cell lines stained with anti-HLA-I and anti-HLA-II. Unstained cells were used as control.

Supplementary Figure 9: Schematic overview of bacteria sequencing prior to the HLA-I and HLA-II peptidomics workflows. a, 17 tumor samples were used for 16S rRNA gene sequencing to identify their intratumor bacteria and in parallel were used for HLA peptidomics. The tumors' bacterial peptides were identified by analyzing the HLA peptidomics data with the Uniprot proteome of the identified bacterial species. In addition, commercial bacterial species were used for in vitro experiments to show entry of the bacteria and presentation of bacterial peptides by melanoma lines derived from the same tumors and APCs. We also performed WGS of the bacteria, and used the proteome assembled from it or the bacteria proteome from Uniprot to analyze the HLA peptidomics data. b, To show that both melanoma cells and APCs from the same tumor can present HLA-I and HLA-II peptides, we used a fresh tumor that was separated based on the CD45 marker. The bacteria isolated from the tumor was identified using WGS. The bacteria proteome to be used to analyze the HLA peptidomics data of the tumor was assembled from the bacteria WGS as well as its Uniprot. c-d, To show that bacterial strains that were isolated from the tumor can enter melanoma cells, we isolated bacteria from tumors 58A and 261MS, and used the bacteria for co-culture experiments with melanoma cells and HLA peptidomics. We also performed WGS of the isolated bacteria, and used the proteome assembled from it or proteome from Uniprot to analyze the HLA peptidomics data.

Supplementary Figure 10: Immunogenic bacterial peptides in IFNγ secretion assay. a, Flow cytometry analysis of IFNγ–secreting TILs after the co-culture of 51AL and 55A TILs with B cells loaded with the indicated bacterial peptide or DMSO control, as described in Extended Data Fig. 12. The figure contains peptides in which the ratio of IFNγ secretion to DMSO control was two-fold or above. b, TILs were gated as single, live and CD19-negative cells.

Supplementary Figure 11: CD69 expression by TILs reactive towards immunogenic bacteria-derived antigens. a, Flow-cytometry analysis of CD69-positive TILs after co-culture of 51AL and 55A TILs with B cells loaded with the bacterial peptides shown to increase IFNγ secretion in Supplementary Fig. 10 or DMSO control. b, TILs were gated as single, live and CD19-negative cells.

Supplementary Figure 12: Recurrent and immunogenic bacterial peptides were derived from prevalent HLA alleles in melanoma TCGA patients. The distribution of the 10 most abundant HLA-A/B/C alleles were plotted in pie chart. All other alleles were grouped together. Of the 10 most abundant HLA alleles for each of the HLA-A/B/C molecules, we identified recurrent and immunogenic bacterial peptides from the A*02:01, A*01:01, A*11:01, A*03:01, A*24:02, A*68:01, B*44:02, B*15:01, B*07:02, B*27:05, C*03:03, C*05:01, C*03:04, C*06:02, C*07:02 and C*02:02 alleles.

Supplementary Figure 13: Correlation between peptides' experimental retention time to the predicted peptides' hydrophobicity. The peptides' RTs were plotted against the calculated hydrophobicity index for each sample and identified bacteria. The measured RTs were regressed against the calculated hydrophobicity index and the residual absolute errors of the lm-regression were plotted using a boxplot. In box plots, the center lines represent the medians, the boxes the range between the 25th and 75th percentile, and the whiskers the range between the smallest and largest data point. In the scatter plots, the red dots represent the bacterial peptides, with the human peptides marked in gray. In the box plots, the red dots represent the outlier human and bacterial peptides of the lm-regression.

Supplementary Figure 14: Uncropped Immunoblot t described in Supplementary Figure 7b.

Supplementary Video legends

Supplementary Video 1-9: Z-stacks of melanoma cells co-cultured with bacteria. Movies were compiled from Z stack images, that their middle stack image is represented in Extended Data Fig. 9-10. Scale 10 μm.

Supplementary Video 10-18: 3D segmentation of confocal Z stacks of melanoma cells co-cultured with bacteria. GFP expressing melanoma cells (green) / Alexa flour 546 (red) stained melanoma cells were co-cultured with Alexa flour 488 (green) / Alexa flour 546 (red) stained bacteria. Segmentation was achieved using the Imaris software. Melanoma segmentation is shown in magenta, bacteria segmentation is shown in turquoise.

Supplementary Table legends

Supplementary Table1: Patients and cells information. Melanoma tumors and cell lines: The table includes clinical information of the patients used for the study; age, gender, stage, tumor site, HLA-I and HLA-II typing of the patient, treatment, from which tumor the TILs and patient derived cell lines were derived, by which method the bacterial species were identified, and in which experiments the tumor and/or cell line were used. Other cell lines: Information for other cell lines used in the study, including their HLA typing and in which experiment they were used.

Supplementary Table 2: Bacterial composition in tumor samples. A table summary of the identified bacterial species in each tumor sample, their taxonomy information and their relative abundance. Under each bacterial species name and taxonomy classification, the group of sequences can be found (under the “+” sign on the left side). For each sequence in the group we indicated its GreenGenes ID and NCBI accession number. Sequences whose species name in the GreenGenes database is missing were searched by NCBI BLAST for the closest species of these sequences. The name of the closest species is indicated in the table ("Closest species (BLAST)") as well as the percentage of identity ("% identity") and its NCBI accession number ("NCBI accession number of similar species"). In case closest species had >96% sequence identity, it was included for further analyses.

Supplementary Table 3: Bacterial derived HLA-I peptides from tumor samples. HLA-I bacterial peptides that were identified in melanoma bulk tumor samples. The length and HLA alleles to which the peptide is predicted to bind (using NetMHCpan) are indicated for each peptide.

Supplementary Table 4: Bacterial derived HLA-II peptides from tumor samples. HLA-II bacterial peptides that were identified in melanoma bulk tumor samples. The length and HLA alleles to which the peptide is predicted to bind (using NetMHCpan) are indicated for each peptide.

Supplementary Table 5: Number of bacterial and human peptides in bulk tumors identified by HLA peptidomics. Summary of the number of identified

bacteria, human derived peptides and the number of peptides that remain after the different filtration steps. Total peptides - number of peptides before filtration. Non-coding/pseudogenes - number of peptides remaining after eliminating peptides with the same sequence as possible non-coding regions or pseudogenes. I/L change - number of peptides remaining after eliminating peptides with the same sequence as coding human sequences, possible non-coding regions or pseudogenes when changing the I/L in the sequence to all possible combinations of I/L. lm(RT~Hydrophobicity) - number of peptides remaining after eliminating outliers of a linear model regression between experimental retention time and predicted hydrophobicity. MS/MS coverage - number of peptides remaining after eliminating HLA-I and HLA-II peptides with MS/MS coverage below 60% or 40% respectively. Predicted to bind by NetMHCpan - number of peptides remaining after eliminating peptides that were not predicted to bind the patients' HLA alleles by NetMHCpan. Note - as not all HLA-II alleles are known (due to typing), much more HLA-II peptides are not predicted by NetMHCpan compared to HLA-I.

Supplementary Table 6: References - intracellular bacteria. Literature references of known intracellular bacteria.

Supplementary Table 7: Recurrent bacterial HLA peptides. The table summarizes the recurrent bacterial HLA peptides that were repeated in different patients, different metastases of the same patient, different samples of the same metastasis or different cell lines and tumor samples. Or if a peptide that was identified in a tumor sample might be derived from two possible bacterial species. Count in tumor samples - sum of tumor samples in which the peptide was identified. Derived from two possible bacteria in tumor - marked with ""+"" if a peptide was identified in one tumor sample might be derived from two possible bacterial species. Count in all samples - sum of all samples (tumor or cell line samples) in which the peptide was identified.

Supplementary Table 8: Statistics for hydrophobicity index. The statistical significance of the difference between bacterial derived and human derived peptides was calculated by unpaired two-samples Wilcoxon test. The p-value and

the number of human and bacterial peptides are indicated for each test. The test was done in different levels: all identified peptides in all tumors and bacteria, for each patient and for each metastasis. As the number of bacteria derived peptides is relatively low, compared to the number of identified human derived peptides, we performed a permutation analysis in which we randomly selected two groups of peptides from a combined list of all human and bacterial peptides, and iterated it 10,000 times. We compared the actual difference to a histogram of differences from the randomized runs and looked at the proportion of runs that gave a difference higher than the observed difference. If a sample passed the permutation test we indicate it as “passed” (corresponding to $P<0.0001$).

Supplementary Table 9: Tumor 422 CD45+/- bacterial HLA peptides. Tumor 422 was digested and used for cell separation based on CD45 marker to two populations; immune cells (CD45+ cells) and non-immune cells (CD45- cells, which are mainly the melanoma cells), followed by HLA peptidomics. Bacteria was isolated from the tumor and underwent WGS to identify the bacterial species as well as to create a bacterial specific proteome database (WGS DB). The samples were analyzed with the proteome database from Uniprot or from WGS. Identified peptides are presented in the table as well as their length and predicted binding (using NetMHCpan) to the patient's HLA alleles.

Supplementary Table 10: APCs derived bacterial HLA peptides. B-cell line IHW01070 and THP1 differentiated macrophages were infected with Fusobacterium nucleatum and bacterial peptides were identified using HLA peptidomics. HLA peptidomics data was analyzed with the bacterial specific proteome database (WGS DB) as well as the database from Uniprot (Uniprot DB). Identified peptides are presented in the table as well as their length and predicted binding to the patient's HLA alleles. All experiments were done in triplicates, and the pepetides presented in the table were identified in two or more replicates of the same experiment.

Supplementary Table 11: APCs derived bacterial HLA peptides - 721.221 cells. B-cell line 721.221 with or without HLA-A*01:01 overexpression ("A*01:01

OE" or "HLA-I null") were infected with Fusobacterium nucleatum and bacterial peptides were identified using HLA peptidomics. HLA peptidomics data was analyzed with the bacterial specific proteome database (WGS DB) as well as the database from Uniprot (Uniprot DB). Identified peptides are presented in the table as well as their length and predicted binding to the patient's HLA alleles. As expected, no HLA-I peptides were identified in the HLA-I null sample. All experiments were done in triplicates, and the pepetides presented in the table were identified in two or more replicates of the same experiment.

Supplementary Table 12: Gentamicin assay statistics. Statistic analysis for gentamicin Colony-forming units (CFU) data: Table 1 - P-values and FDR corresponding to Student's t-test between supernatant sample to No gentamicin/With gentamicin samples. Table 2 - P-adj values comparing *Actinomyces odontolyticus*/*Staphylococcus caprae* With gentamicin samples to *Lactobacillus animalis* / *Lactobacillus plantarum* with gentamicin control samples, correspond to one-way ANOVA followed by Tukey's test. Table 3 - Raw CFU counts.

Supplementary Table 13: Bacterial HLA peptides in infected melanoma cells. 51AL and 55A3 melanoma cell lines were genome edited to reduce HLA-I and HLA-II expression after B2M or CIITA KO, respectively. Cells with scrembeled sgRNA control were marked as "SC". 51AL cells were infected with Fusobacterium nucleatum and 55A3 were infected with Staphylococcus caprae and bacterial peptides were identified using HLA peptidomics. HLA peptidomics data was analyzed with the bacterial specific proteome database (WGS DB) as well as the database from Uniprot (Uniprot DB). Identified peptides are presented in the table as well as their length and predicted binding to the patient's HLA alleles. As expected, less or no peptides were identified in the cells with the KO. Less invasive bacteria Lactobacillus plantarum and Lactobacillus animalis were used as controls in 51AL and 55A3, respectively. As expected a lower number of peptides were identified in these samples. All experiments were done in triplicates, and the

pepetides presented in the table were identified in two or more replicates of the same experiment.

Supplementary Table 14: Contaminating peptides in the HLA peptidomics. Samples containing only the bacteria (with no cells) in the same amount as in the infection experiments were used for HLA peptidomics. HLA peptidomics data was analyzed with the bacterial specific proteome database (WGS DB) as well as the database from Uniprot (Uniprot DB) and with all HLA alleles of the cell lines used in this study (in order to repeat the same analysis as done on the infected cell lines). Identified peptides are presented in the table as well as their length and predicted binding to the patient's HLA alleles. Non of the contaminating peptides were identified in the infected cells experiments. All experiments were done in triplicates, and the pepetides presented in the table were identified in two or more replicates of the same experiment.

Supplementary Table 15: Bacterial HLA peptides identified in 58A cell line. Mlelanoma cell line 58A was infected with Staphylococcus capitis isolated from tumor 58A and bacterial peptides were identified using HLA peptidomics. HLA peptidomics data was analyzed with the bacterial specific proteome database (WGS DB) as well as the database from Uniprot (Uniprot DB). Identified peptides are presented in the table as well as their length and predicted binding to the patient's HLA alleles. All experiments were done in triplicates, and the pepetides presented in the table were identified in two or more replicates of the same experiment.

Supplementary Table 16: Comparison between the use of uniprot and WGS database. Summarizing tables containing the number of identified bacterial HLA peptides in all cell line experiments which were analyzed with the Uniprot database as well as the WGS database of the specific bacteria used for the infection experiments. The table indicates the number of overlapping bacterial peptides identified using both databases.

Supplementary Table 17: Bacterial proteins with multiple peptides. All bacterial peptides that were identified in both tumors and cell line experiments, that

were derived from the same protein were grouped under the same protein name. All protein entries from uniprot and WGS databases sharing the same set of peptide sequences were grouped together, and all the names are indicated in the column "Database entries". All proteins that share similar peptides were grouped under the same "Protein group". Peptide count column indicates the number of unique peptide sequences in the group. Sample count column sum the number of different samples in which a specific peptide was identified.

Supplementary Table 18: Absolute abundance values of sequence reads mapped to the Kraken database. Shown are the counts of sequence reads mapped to each bacterial family (F), genus (G), species (S) or strain (-). For each patient the read count in the tumor and in the blood sample (negative control, normal) of each patient is shown.

Supplementary Table 19: Sanger sequencing of 16S rRNA gene in isolated bacteria. Results of sanger sequencing of bacteria isolated from tumors used in this study. The bacterial species were determined by performing BLAST in order to identify the closest bacterial species, the % of identity is indicated in the table.

Supplementary Table 20: WGS of isolated bacteria and commercial bacteria. Tumor - name of the tumor from which the bacteria was isolated, or if the bacteria is a commercial bacteria.Species Classification- bacteria species as determined by GTDB-tk and verified with 16S and % reads mapping. Assembly Length (Mbp). Number of contigs greater than 500bp. Longest contig (kbp). N50 (kbp)- minimum contig length needed to cover 50% of the genome, i.e. half of the total bp of the genome assembly is in contigs larger than or equal to N50. CheckM Completeness (%) – CheckM estimates genome completeness and contamination by presence of collocated sets of single-copy genes (SCG), which are ubiquitous in the determined phylogenetic lineage. CheckM Contamination (%) – see CheckM Completeness (%). Cluster # - Genomes were clustered according to pairwise ANI with the dRep tool. Each species clustered (≥95% ANI) was numbered. Reference NCBI ID – RefSeq (GCF) or GenBank (GCA) ID for the representative genome of the species identified. ANI (Average Nucleotide Identity) – calculated between the

assembly and the reference genome which is the representative of the identified species in GTDB (ANI computed with fastANI in GTDB-tk software). Alignment Fraction (AF) – ratio of the number of contigs that mapped to the reference and the total number of contigs (AF computed with fastANI in GTDB-tk software). % Overall Read Alignment – percentage of reads that aligned to the reference genome (calculated with the fast aligner SNAP). Annotated Protein CDS – number of protein coding sequences predicted by Prokka.

Supplementary Table 21: Bacteria used in the study. Details of the bacteria used in the study, including growth conditions and media, strain source and if the bacteria was received from another lab.

Methods

Patients and cells

All patients included in the analysis were diagnosed with metastatic melanoma. Tumor samples from patients 19, 42, 51, 55, 58, 70, 86, 92, 112, 152 and 422 were received from eleven different patients treated at the University of Texas MD Anderson Cancer Center, all of who had signed an informed consent for the collection and analysis of their tumor samples. The protocol for tumor samples was approved by the MD Anderson Institutional Review Board (protocol numbers 2012-0846, LAB00-063 and 2004-0069, NCT00338377). Synchronous metastatic tumors were resected via surgery at the same time. Sample and patient information are provided in Extended Data Table 1. From each patient, we collected between 1-3 different metastases. From two metastases, 51BR and 55A7, we collected two different samples, which were later analyzed separately by HLA peptidomics, these samples are marked as 51BR-1/2 and 55A7-1/2. For sample 152A we did not purify HLA-II peptides. From three metastases, 51AL, 55A3 and 58A, we produced patient-derived cell lines. From samples 51AL and 55A3 TILs were produced. Tumor sample HG38 was received from patient biopsies removed at the University of Zurich Hospital. The patient had signed a patient release form, which has been approved by an ethics committee and assigned the numbers EK647 and EK800. Tumor 261MS was received from patient biopsies removed at the Ella Lemelbaum Institute for Immuno-Oncology (Sheba medical center, Israel), under Institutional Review Board number 6786-20-SMC.

EBV-transformed B-cells were purchased from the IHWG Cell and DNA Bank, their HLA typing if indicated in Extended Data Table 1. Hybridoma cells HB95 and HB145 were purchased from the ATCC and were used to purify pan-HLA-I and pan-HLA-II antibodies for the preparation of the HLA affinity columns. LCL 721.221 HLA-I null cells were purchased from the ATCC. THP-1 cells were a kind gift from Prof. Steffen Jung (Weizmann Institute of Science). All cell lines were tested regularly and were found negative for mycoplasma contamination (EZ-PCR

Mycoplasma Kit, Biological Industries). All cells were maintained in RPMI 1640 containing 10% fetal bovine serum (FBS) and 1% penicillin-streptomycin, glutamine and sodium pyruvate.

HLA typing

HLA typing for patients 19, 42, 51, 55, 86, 92 was determined from the WES data of tumors/ blood samples by POLYSOLVER [51]. WES data is available in the NCBI Sequence Read Archive public repository (BioProject ID PRJNA316754). For patients 70, 112 and 152 genomic DNA extracted from the tumor using a QIAgen DNAeasy Blood & tissue kit (cat.69506). HLA-A, -B, -C, -DQB1 and DRB1 were identified using Sequence Specific Oligonucleotide (SSO) and Sequence Based Typing (SBT). Genomic DNA for cell lines 58A, THP1 and 721.221 was extracted from $2*10^6$ cells. DNA samples were typed for six loci: HLA-A, -B, -C, -DPB1, -DQB1 and DRB1, using the MX6-1 NGS typing kit (GenDx). HLA typing information is found in Supplementary Table 1.

Bacterial DNA extraction, 16S sequencing and data analysis from tumor samples

Samples were collected, processed and sequenced as described in [5]. Briefly, frozen tissue samples (40-70 mg) were extracted with the UltraClean Tissue & Cells DNA Isolation kit (MoBio, #12334). All negative controls were processed according to the exact same protocols. Different controls were included, such as DNA extraction controls and no-template PCR amplification controls, to account for the various sources of contamination from the hospital and laboratory environments as well as the different stages of handling and processing of the samples, as described in [5].

16S rRNA amplification and sequencing was done by amplifying 5 regions on the 16S rRNA gene in multiplex. The libraries were sequenced on Illumina Hi-seq system. Reads were demultiplexed per sample, filtered and aligned to each of the five amplified regions based on the primers' sequences. The Short MUlitiple Regions Framework (SMURF) method was applied to combine read counts from

the five regions into a coherent profiling results solving a maximum likelihood problem [52]. The GreenGenes database (May 2013 version, with some improvements) was used as a reference. The SMURF's output comprises a list of 'groups' and their relative abundances, where a group is a set of full length GreenGenes 16S rRNA sequences that share the same sequence over the amplified regions. In some cases the group contains the same bacterial species and in some cases there is more than one species (in these cases all of the bacterial species were reported). Supplementary Table 2 summarize the identified bacterial groups. Under each group, the species name and taxonomy classification is designated, together with its corresponding full length 16S rRNA sequences (under the "+" sign on the left side). For each sequence in the group we indicated its GreenGenes ID and NCBI accession number. Whenever the GreenGenes ID is in the form "XX.Y" rather than "XX", it corresponds to a GreenGenes sequences "XX", where ambiguous nucleotides were replaced (for example "N"), by real nucleotides. Sequences whose species name in the GreenGenes database is missing were searched by NCBI BLAST for the closest species of these sequences. The name of the closest species is indicated in the table (Closest species (BLAST))as well as the percent identity (% identity) and its NCBI accession number (NCBI accession number of similar species). In cases where the closest species had >96% sequence identity, it was included for further analyses.

Phylogenetic tree assembly

A schematic phylogenetic tree was constructed based on the taxonomy classification, and was imported to R using ape [53] package. The phylogenetic tree figure was generated in R by ggplot2 [54] and ggtree [55,56] packages with a phyloseq object from phyloseq [57] package, that combines the phylogenetic tree data and the bacteria species in samples.

Similarity between samples' bacterial composition

The degree of similarity of bacterial composition between samples was calculated as Jaccard distances and plotted with the ggplot2 package in R.

Taxonomic profiling of a melanoma WGS data

High-throughput DNA sequencing of 108 pairs of tumor and blood samples was used for this analysis [15]. There were 222,149 (95% confidence interval 128,226–316,071) reads (1.1% of the raw reads) per sample that were not mapped to the human genome (hs37d5 using BWA-MEM software) and were kept for downstream analysis as potential microbial DNA. The raw read counts of the bacteria in the samples are indicated in Supplementary Table 18. Taxonomy analysis was performed with the Kraken v.1.0 [58] metagenomics read classification tool, using a human microbiome database, as previously described [59]. Statistical analysis was performed using pairs of tumor and blood samples from the same patient. This process helps account for technical noise (including handling and kit reagent contaminants) observed in both tumor and blood from the same pair. Additionally, we tested for correlation between bacterial abundance and clinical metadata by applying generalized linear mixed models (GLMMs) using a tool called MaaSlin2 (/https://huttenhower.sph.harvard.edu/maaslin/). TSS normalization was applied to measured read count. Random effects were modeled by patient. We investigated associations against fixed effects such as sex, age, specimen type, subtype, telomere length, melanoma subgroup, ploidy, MSI sensor, mutation load, and UV signature as described in [15]. The R packages phyloseq [57] and vegan version 2.5-4 [60], were used for metagenomics data analysis. Results were visualized using ggplot2 [54] package.

Production and purification of membrane HLA molecules

Tumor sample weights ranged from 0.1-0.8 grams. Cell-line pellets were collected from $2x10^8$ cells. Samples were processed as previously described [61-63]. Tumor samples and cell pellets were homogenized through a cell strainer on ice with a lysis buffer containing 0.25% sodium deoxycholate, 0.2 mM iodoacetamide, 1 mM EDTA, 1:200 Protease Inhibitor Cocktail (Sigma-Aldrich, P8340), 1 mM PMSF and

1% octyl-b-D glucopyranoside in PBS. Samples were then incubated in rotation at 4˚C for 1 hour. The lysates were cleared by centrifugation at 48,000 g for 60 minutes at 4˚C and then passed through a pre-clearing column containing Protein-A Sepharose beads.

HLA-I molecules were immunoaffinity purified from cleared lysate with the pan-HLA-I antibody (W6/32 antibody purified from HB95 hybridoma cells) covalently bound to Protein-A Sepharose beads [61,64]. HLA-II molecules were then purified by transferring the flow-through to similar affinity columns containing a pan-HLA-II antibody (purified from HB-145 hybridoma cells). Affinity columns were washed first with 400 mM NaCl, 20 mM Tris–HCl and then with 20 mM Tris–HCl pH 8.0. The HLA-peptide complexes were then eluted with 1% trifluoracetic acid followed by separation of the peptides from the proteins by binding the eluted fraction to Sep-Pak (Waters). Elution of the peptides was done with 28% acetonitrile in 0.1% trifluoracetic acid for HLA-I and 32% acetonitrile in 0.1% trifluoracetic acid for HLA-II.

Mass-spectrometry analysis of eluted HLA peptides

Tumor samples 42RF, 42RS, 27, 51AL, 51BR-1, 55A3, 55B3, 55A7-1, 86A, 86B2, 92B3, 92C3 were analyzed as follows:

The HLA peptides were dried by vacuum centrifugation, solubilized with 0.1% formic acid, and resolved on capillary reversed phase chromatography on 0.075x300 mm laser-pulled capillaries self-packed with C18 reversed-phase 3.5 µm beads (Reprosil-C18-Aqua, Dr. Maisch GmbH, Ammerbuch-Entringen, Germany) [65]. Chromatography was performed on the UltiMate 3000 RSLCnano-capillary UHPLC system (Thermo Fisher Scientific), which was coupled by electrospray to tandem mass spectrometry on Q-Exactive-Plus (Thermo Fisher Scientific). The HLA peptides were eluted with a linear gradient over 2 hours from 5 to 28% acetonitrile with 0.1% formic acid at a flow rate of 0.15 µl/minute. Data was acquired using a data-dependent “top 10” method, fragmenting the peptides by higher-energy collisional dissociation. Full scan mass-spectrometry (MS)

spectra were acquired at a resolution of 70,000 at 200 m/z with a target value of $3x10^6$ ions. Ions were accumulated to an AGC target value of 10^5 with a general maximal injection time of 100 milliseconds. The peptide match option was set to Preferred. Normalized collision energy was set to 25% and the MS/MS resolution was 17,500 at 200 m/z. Fragmented m/z values were dynamically excluded from further selection for 20 seconds.

Tumor samples 51BR-2, 55A7-2, 70.1, 86B, 112, 152A, 422 and all cell lines derived samples were analyzed as follows:

ULC/MS grade solvents were used for all chromatographic steps. Each sample was solubilized in 12 µL 97:3 water: acetonitrile with 0.1% formic acid. Samples were loaded using split-less nano-Ultra Performance Liquid Chromatography (10 kpsi nanoAcquity; Waters, Milford, MA, USA). The mobile phase was: A) H_2O + 0.1% formic acid and B) acetonitrile + 0.1% formic acid. Desalting of the samples was performed online using a reversed-phase Symmetry C18 trapping column (180 µm internal diameter, 20 mm length, 5 µm particle size, Waters). The peptides were then separated using a T3 HSS nano-column (75 µm internal diameter, 250 mm length, 1.8 µm particle size; Waters) at 0.35 µL/minute. The column was heated to 45˚C. Peptides were eluted from the column into the mass spectrometer using the following gradient: 5% to 28%B in 120 minutes, 28% to 35%B in 15 minutes, 35% to 95% in 15 minutes, maintained at 95% for 10 minutes and then back to initial conditions.

The nanoUPLC was coupled online through a nanoESI emitter (10 µm tip; New Objective, Woburn, MA, USA) to a quadrupole orbitrap mass spectrometer (Q Exactive Plus, Thermo Scientific) using a FlexIon nanospray apparatus (Proxeon). For the discovery experiments, data was acquired in data dependent acquisition (DDA) mode, using a Top20 method. MS1 resolution was set to 70,000 (at 400m/z), mass range of 300-1800m/z, automatic gain control (AGC) of 3e6 and maximum injection time was set to 100 msec. MS2 resolution was set to 17,500, quadrupole isolation 1.8m/z, AGC of 1e5, dynamic exclusion of 20 seconds and

maximum injection time of 150 msec. Charge exclusion was set to 'unassigned', 4-8, greater than 8 and an exclusion list of singly charge, background ions.

Identification of bacterial-derived HLA peptides

FASTA files of the bacterial species proteome, as well as the human proteome, were downloaded from UniProt (https://www.uniprot.org/). Proteome datasets were available for all bacteria except to *Porphyromonas bennonis*, *Sphingomonas kyeonggiensis* and *Sphingomonas roseiflava*. For bacterial species with multiple entries, we downloaded all entries, to avoid missing information, and then removed redundant sequences using CD-HIT software [66,67] (http://weizhongli-lab.org/cd-hit/). The longest sequences with 100% identity were kept.

MS data was analyzed using MaxQuant [68] version 1.5.0.25. Peptides were searched against the UniProt human proteome and the specific bacterial species that were identified in the sample (both datasets were used together in each analysis). In patients where the same bacteria was identified in a few metastases, we ran all metastases samples together. N-terminal acetylation (42.010565 Da) and methionine oxidation (15.994915 Da) were set as variable modifications. Enzyme specificity was set as "unspecific" and peptides' FDR was set to 0.05. The "match between runs" option was disabled to avoid the matching of identifications across the samples. Mass tolerance was set to 20 ppm.

For the identification of bacteria derived peptides, only peptides that were identified from the bacterial proteome were selected. As Leucine and Isoleucine residues are generally considered indistinguishable by MS, because their molecular masses are the same, we removed from our list all peptides in which replacing their I/L position (by either I/L) resulted in peptides that can be derived from a human origin. We also removed peptides that have the same sequence (or with a change of I/L) as possible non-coding regions (https://www.gencodegenes.org/human/release_19.html) as well as possible pseudogenes (http://www.pseudogene.org/Human/Human90.txt). We then filtered

out outliers of a linear model regression between experimental retention time and predicted hydrophobicity (see methods: “Hydrophobicity index calculation”). Peptides were further filtered according to the quality of their identification; we defined the peptide’s “MS/MS coverage” as the number of amino acids that start b or y ion, divided by the peptide length. In order to determine a threshold for sufficient MS/MS coverage, we checked the expected MS/MS coverage of peptides with a high certainty of being correct such as peptides found in the IEDB [69] or peptides that were found to be immunogenic in our samples. We found that the threshold for HLA-I peptides is a MS/MS coverage of >60% and for HLA-II peptides >40% (see Supplementary information).

NetMHCpan [70-72] version 4.0 (http://www.cbs.dtu.dk/services/NetMHCpan/) and NetMHCIIpan [73] version 3.2 (http://www.cbs.dtu.dk/services/NetMHCIIpan/) were used to check if the peptides can bind the patient’s HLA alleles. Peptides with %rank< 2 were kept.

Finally, we validated the identification of selected peptides by comparing their MS/MS spectra fragmentation to that of synthetic peptides (see methods: “Validation using synthetic peptides”). We also clustered human and bacteria derived peptides using Gibbs clustering, to see if they clustered according to the expected HLA binding motifs of the patient (see methods: “Gibbs clustering”).

In samples with different bacterial species that had different proteome size, we observed some difference in the number of human peptides, although it was derived from the same sample and raw file. This change was more pronounced when the difference in the bacterial proteome sizes was higher.

All raw MS files as well as human and bacteria proteomes and the MaxQuant version used for analyzing the data were deposited to the ProteomeXchange Consortium [74] via the PRIDE [75] partner repository with the dataset identifier PXD022150.

Hydrophobicity index calculation

Sequence specific hydrophobicity index was calculated using the ssrc function from the R package specL v1.6.2 [76] with default parameters. Observed retention times (RT) were obtained from MaxQuant output file "msms.txt". If a peptide was detected multiple times in the same sample, the best RT was used (with the highest number of fragments). Peptides' RTs were plotted against the calculated hydrophobicity index as in Supplementary Fig. 13.

We regressed the measured RTs against the calculated hydrophobicity index using the lm function from the R package stats v3.6.2, in order to calculate the standard errors of the hydrophobicity index. The residual absolute errors of the lm-regression were plotted using a boxplot (Supplementary Fig. 13) with a median and hinges (25% and 75%). The outliers were determined as residual absolute errors greater than 75th percentile + 1.5 * interquartile range (IQR). These outliers were excluded from further analyses.

Validation using synthetic peptides

Synthetic peptides (GenScript, 50 fmol/μl) were used to validate the peptides' fragmentation. Peptides were analyzed in the same MS and conditions as the eluted peptide samples from cell lines. The MS2 spectra as they appear in the MaxQuant software are found in Supplementary Fig. 3. The MSnbase R package [77] was used to calculate the correlation between the matched y and b ions of the synthetic peptides and the endogenous peptides. Pearson correlation and dot product score are indicated for each comparison in the figure. SpectrumSimilarity function from the R package OrgMassSpecR was used to plot a head to tail figure of the endogenous and synthetic peptides fragmentation (Supplementary Fig. 3).

Gibbs clustering

Each set of peptides was clustered using the GibbsCluster 2.0 server [78,79] (www.cbs.dtu.dk/services/GibbsCluster), with the "MHC class I ligands of length 8-

13" parameters; the number of clusters was set to six and the trash cluster option was enabled. Since the number of peptides per allele is different, for alleles with a higher number of peptides (as the HLA-A alleles), the unbiased clustering sometimes resulted in more than one cluster for these alleles. In these cases, we selected the number of clusters in which we receive only one cluster per HLA allele. The variation in the number of peptides per allele also resulted in clusters with mixed motifs that were similar. In these cases, we assigned the cluster to the allele that had the highest representation in the cluster and added a note as to which other alleles are mixed within. For each cluster, we indicated the number of human- and bacterial-derived peptides clustered to this allele. All motifs were generated by Seq2Logo 2.0 [80] (http://www.cbs.dtu.dk/biotools/Seq2Logo) with the default settings. In order to classify the clustered peptides into HLA alleles, we first identified for each allele its motif. This entailed retrieving all HLA-I epitopes registered under this allele from the Immune epitope database [69] (IEDB, www.iedb.org). All peptides annotated as positive in "MHC ligand assays" to the specific HLA-I allele and were 8-13 amino acids in length. The GibbsCluster 2.0 server was used to align the peptides using the "MHC class I ligands of length 8-13" parameters; the number of clusters was set to one and the trash cluster option was disabled. HLA alleles to which there were no peptides in the IEDB to create a representative logo for their motif were searched in the SYFPEITHI database [81] (www.syfpeithi.de) or in Bassani-Sternberg et al. [82].

Peptide hydrophobicity

Kyte-Doolittle hydrophobicity scale [83] was calculated for the identified HLA-I peptides using the Peptides package in R (https://github.com/dosorio/Peptides). The statistical significance of the difference between bacteria derived and human derived peptides was calculated by unpaired two-samples Wilcoxon test. As the HLA peptidome of the different metastases was analyzed separately with each bacterial proteome, we collected all the bacteria and human peptides identified in the cohort (Fig. 2), for each patient and for each metastasis (Supplementary Fig. 6) and removed redundant sequence for the statistical analysis. We also did the

same test for each sample that was analyzed with every relevant bacteria (Supplementary Table 8 and Supplementary Fig. 6). As the number of bacterial derived peptides is relatively low, compared to the number of identified human derived peptides, we performed a permutation analysis in which we randomly selected two groups of peptides from a combined list of all human and bacterial peptides, and iterated it 10,000 times. We compared the actual difference to a histogram of differences from the randomized runs and looked at the proportion of runs that gave a difference higher than the observed difference. If a sample passed the permutation test we indicate it as “passed” (corresponding to $P<0.0001$) in Supplementary Table 8.

Separation of CD45 +/- cell populations from tumor samples

Tumors 422 and HG38 were dissociated by manual mincing and suspended with RPMI-1640 supplemented with DNase (2 mg/ml, Sigma-Adrich) and collagenase IV (1 mg/ml, Roche). Tissues were homogenized by GentleMACS tissue dissociator (Miltenyi Biotec), and incubated at 37˚C for 10 minutes. Following three times of mechanic and enzymatic dissociation. Cells were washed with cold PBS, passed through 100 mm mesh filter, centrifuged at 400 g for 5 minutes at 4°C and then resuspended in cold FACS buffer. CD45+/- populations were isolated using MACS CD45 MicroBeads (Miltenyi Biotec, cat# 130-045-801, Lot# 5191204025). Sample HG38 was small and used for flow cytometry (see “Flow cytometry”) to show the efficiency of separation and HLA-II expression by CD45- population. Tumor 422 was approximately 1.5 cm when it was resected, a small fraction of it was used for bacteria isolation and the rest was used for CD45 separation (see “Bacterial isolation from melanoma tumors”) followed by HLA peptidomics. To access the quality of separation the flow through of the CD45+/- populations after HLA peptidomics was analyzed by western blot with anti-CD45, anti-HLA-DP1 and anti-GAPDH (see “Immunoblotting”).

Immunoblotting

Immunoblotting was performed using the HLA peptidomics flow-through of CD45+/- populations of tumor 422. Protein amount of CD45+/- populations' lysates was quantified using Pierce™ BCA Protein Assay Kit (Thermo Fisher Scientific), CD45- had 13-fold more protein compared to CD45+ populations, indicating on the size of separated populations. Sample buffer 2X was added to the flow-through and boiled for 5 minutes. The samples were then resolved by 10% SDS-PAGE and transferred to nitrocellulose membranes (Bio-Rad). Immunoblots were probed with the following antibodies: CD45 (D9M8I, Cell Signaling), HLA-DP1 (EPR11226,Abcam), GAPDH (6C5, Millipore). All antibodies used were diluted 1:1000. Blots were developed with HRP-conjugated anti-mouse or anti-rabbit Abs, using SuperSignal West Femto Chemiluminescent Substrate from Thermo Scientific. Images of the blots were taken using Bio-Rad ChemiDoc MP System. The uncropped images are provided in Supplementary Fig. 14.

Bacterial isolation from melanoma tumors

Bacterial isolation from tumor 422

Tumor was homogenized in 2 ml of PBS. The tumor homogenate was then plated in varying volumes (25µl, 50µl, 75µl, 100µl, 200µl and 250µl) on BD™ Columbia CNA Agar with 5% Sheep Blood, BD BBL™ Trypticase Soy Agar (TSA II) with 5% Sheep Blood, or Cooked meat agar. Blood agar plates were grown at 37°C in aerobic conditions, while the cooked meat agar plates were grown at 37°C in an anaerobic chamber. 25µl of homogenate were also added directly to Sodium thioglycolate broth and grown at 37°C to allow for the differentiation of obligate aerobes, obligate anaerobes, facultative anaerobes, microaerophiles and aerotolerant organisms. Homogenates were grown for 24 hours in aerobic conditions, and 48 hours in anaerobic conditions. A single colony was isolated from TSA Blood agar, and one from cooked meat agar. These isolates were then identified on a Bruker MALDI Biotyper System as *Bacillus megaterium* and *Clostridium intestinalis*, respectively. *Bacillus megaterium* colony was grown in 2 ml of Luria-Bertani Broth at 37°C with shaking, and *Clostridium intestinalis* colony was grown at 37°C in anaerobic conditions in reduced cooked meat media, for 6

hours, DNA was extracted from their pellets using the Lucigen MasterPure Gram Positives DNA purification kit (Lucigen, cat#04100) with the addition of mutanolysin (~80U). . Whole genome sequencing was performed using 150bp paired end reads using Ilumina NextSeq500. KAPA Hyper Kit was used for the library preparation and ligated to NEXTflex dual index adapters. The assemblies were performed in SPAdes, with sequence and assembly quality checked with fastqc and quast, respectively. Data analysis was performed as described in “Bacteria isolates WGS analysis”.

Bacterial isolation from tumor 58A and Mel261

Frozen tumor sample 58 and fresh tumor Mel261 were pulverized extensively in PBS (1 ml) to apparent homogeneity. 400 ml of the homogenate was plated on Brucella agar plates supplemented with 5% sheep blood (Fisher cat# R04011). The remaining material was used to inoculate 600 ml of SHI medium. All cultures were incubated at 37°C overnight under aerobic and anaerobic conditions. The liquid cultures were diluted serially and then plated onto solid media as described. Plates were incubated for 5-6 days at 37°C to capture slow growing isolates. Isolated colonies were picked as they appeared into 3 ml liquid BHIY (Beef heart infusion with yeast extract) and incubated overnight at 37°C. Genomic DNA was extracted using a QIAgen DNAeasy Blood & tissue kit (cat# 69506). Genomic DNA was used as template to amplify the 16S rRNA gene using primers F: 5’ AGRGTTTGATCMTGGCTCAG 3’ and R: 5’ CCGTCAATTCMTTTRAGT 3’. Products were subjected to Sanger sequencing using the R primer. Resulting sequence data was used to perform BLAST to identify the best bacterial species match (Supplementary Table 19). Bacteria isolates were also subjected to WGS as described in “Bacterial isolates WGS analysis”. Sanger sequencing and WGS of the isolates gave similar results, and identified *Staphylococcus capitis* in tumor 58 and *Staphylococcus succinus* and *Enterobacter cloacae* in tumor Mel261.

Bacterial isolates WGS analysis

Library preparation and sequencing of bacterial isolates from tumors 58A and Mel261 and commercial bacteria

Illumina libraries were prepared using Nextera DNA library kit (Ref# 15028211), by Tecan Freedom Evo 200 robot platform. IDT® for Illumina Nextera DNA Unique Dual Index Set E were used for library preparation. Library concentration was measured using iQuantTM dsDNA HS Assay Kit, ABP biosciences (Cat# AP-N011) and library size quantified by automated electrophoresis nucleic acid QC - Tape-Station system. Libraries were sequenced to a minimum depth of 1.3 million reads by NextSeq 500 machine with IlluminaNS 500/550 High Output V2.5 150 cycle kit (Cat# 20024907).

Analysis of WGS data for all bacterial isolates and commercial bacteria

Illumina reads were trimmed for adapters and low-quality nucleotides with Trimmomatic v0.38 [84]. Quality filtered paired-end reads were *de novo* assembled using SPAdes v3.14 [85].Genome assembly quality was assessed with QUAST v5.0 [86] and CheckM v1.0.18 [87]. For the few samples where multiple genomes were detected, reads were mapped by BWA-MEM v0.7 [88] and contigs were grouped according to coverage and tetranucleotide frequencies by metabat2 v2.14 [89]. Taxonomic classifications were assigned based on the Genome Taxonomy Database (GTDB release 04-RS89) [90] using the toolkit GTDB-tk v0.3.2 [91] .Included in GTDB-tk, fastANI v1.1 [92] was used to calculate the alignment fraction and average nucleotide identity (ANI) between the genome assembly and the identified reference genome (95% ANI threshold for species identification). Clustering was also performed according to ANI with dRep [93] .Protein coding sequences were annotated with Prokka v1.14 [94]. Finally, for confirmation of taxonomic classifications, 16S rRNA sequences detected by barrnap (https://github.com/tseemann/barrnap) were verified with BLAST [95]and the percentage of reads aligning to the reference genome (retrieved from the NCBI RefSeq/GenBank Databases [96,97]) was calculated with the SNAP aligner (v1.0) [98]. Supplementary Table 20 summarizes the WGS data. Raw sequence data were

deposited in the NCBI Sequence Read Archive (SRA), under BioProject accession number PRJNA669827.

Differentiation of THP1 cells to macrophages

THP-1 cells were differentiated into macrophages by growing the cells (10^6 cells/ml) in the presence of phorbol 12-myristate 13-acetate (PMA, 100 ng/ml, sigma, cat: P1585) for 60 hours [29,30]. As expected, after differentiation of the cells the CD14 marker was lower compared to untreated THP1 cells [29,99] (Supplementary Fig. 8b).

Pooled stable expression of HLA-A*01:01 in 721.221 cells

DNA sequences coding for HLA allele A*01:01 were taken from the IPD-IMGT/HLA database (https://www.ebi.ac.uk/ipd/imgt/hla/allele.html) and purchased as synthetic dsDNA from Twist bioscience. Coding sequence was cloned into pCDH-CMV-MCS-EF1α-Neo vector (SBI, #CD514B-1) and lentiviral particles were produced by co-transfection with envelope and packaging plasmids (PMD2.G and psPAX2) into HEK293T cells using Lipofectamine 2000 (Invitrogen). At 48 hours post transfection, the virus containing media was harvested, filtered, aliquoted and stored at -80˚C. Human B-LCL 721.221 (HLA-I null) were infected, and after 72 hours were selected with neomycin (G418). Overexpression of HLA-A*01:01 was validated by flow cytometry (Supplementary Fig. 8d).

Co-culture of melanoma cells and APCs with bacteria

Bacteria were cultured overnight in their respective growth conditions (Supplementary Table 21). The next morning, bacteria were washed three times in PBS and re-suspended in RPMI1640 without penicillin/streptomycin. Cells were co-cultured with bacteria at a multiplicity of infection (MOI) of 500 and co-cultured for 4 hours and/or 8 hours at 37˚C. Cells were then washed three times with PBS and fresh RPMI1640 with 200 µg/ml gentamicin was added and the cells were incubated at 37˚C for 1 hour.

Co-cultured APCs and melanoma cell lines HLA peptidomics analysis

All *in vitro* HLA peptidomics experiments were done in triplicates. Samples were processed similarly to the tumor samples and analyzed using the proteome derived from UniProt and the proteome derived from WGS of the bacteria (analyzed separately for each database). After the HLA peptidomics data filtration (as described for the tumors), only peptides that were identified in at least two replicated were kept for further analyses.

Gentamicin protection assay

Cells were co-cultured with bacteria for 4 and 8 hours. Cells were then treated or not treated with gentamicin for one hour and then lysed by liquid homogenization. Media from cells that were treated with gentamicin was collected as control for bacteria cells left after gentamicin treatment. All samples were plated on agar plates for 12 hours at 37˚C. The number of colonies was counted, and colony forming units (CFU) CFU/ml were calculated in order to estimate the number of intracellular bacteria. CFU of samples that were not treated gives an indication of the number of bacteria that invade the cells or only adhere, while the CFU of samples with gentamicin treatment gives indication of the number of bacteria that invaded the cells. Bars represent standard error between replicates (n=3). *Lactobacillus animalis* and *Lactobacillus plantarum* were used as controls for bacteria that were less expected to invade the cells.

Pooled stable expression of GFP-expressing melanoma cells

To produce the lentivirus, GFP-pLEX-TRC206 plasmids were co-transfected with pCMV-VSV-G and psPAX2 helper plasmids using Lipofectamine 2000 (Life Technologies) into HEK293T-cells. The cells were seeded at 2.5×10^6 per T75 flask. Virus-containing medium was collected 72 hours after transfection, filtered, aliquotted and stored at -80°C. 51AL, 55A3 and 58A cells were infected with the virus for 48 hours and then selected with Blasticidin.

Immunofluorescence imaging of aerobically grown bacteria

51AL and 55A3 GFP-expressing cells ($1.5x10^5$) were seeded onto 18 mm coverslips in 24-well plates and grown for 24 hours under humid conditions at 37°C. Cells were washed once with PBS, and each well was filled with RPMI medium containing no penicillin/streptomycin. A 10 µl volume of *Staphylococcus caprae, Staphylococcus capitis* and *Staphylococcus succinus* bacterial cells was added to each well at an MOI of 500. Eight hours later, coverslips were removed from the plates and the plates were washed three times with PBS. The cells on the coverslips were fixed for 10 minutes in 2% paraformaldehyde and then washed three times in PBS. The cells were then permeabilized with 0.1% Triton X-100 for 5 minutes, washed twice with PBS, and incubated with anti-Lipoteichoic acid antibody (HM2018, HycultBiotech,1:100 in PBS with 0.5% goat serum) for 1 hour. Coverslips were washed three times and incubated with the anti-mouse secondary antibody labeled with Alexa flour 546 (A-11030, Invitrogen,1:500 in PBS with 0.5% goat serum), in light-protected conditions for 1 hour at room temperature. Coverslips were extensively washed and then stained for 10 minutes with DAPI nuclear staining (1 µg/ml). Confocal images were acquired with a Zeiss LSM 800 confocal microscope using a 63x objective and processed with ImageJ (NIH).

Immunofluorescence imaging of anaerobically grown bacteria using copper-free click-chemistry labelling

Actinomyces odontolyticus and *Fusobacterium nucleatum* were grown overnight at 37°C under anaerobic conditions in a medium containing a final concentration of 100 µM D-tetraacetylated-N-azidoacetylgalactosamine (D-GalNAz) to an OD_{600} of 0.7. Next, the cultures were labeled with DIBO-AF488, as previously described [32]. The labeled bacteria were added to 51AL and 55A3 ($1.5x10^5$) cells seeded onto 18 mm coverslips in 12-well plates and grown for 24 hours at 37°C. Cells were washed once with PBS, and each well was filled with RPMI medium containing no penicillin/streptomycin. A 10 µl volume of bacterial cells was added to each well at an MOI of 500 for 8 hours. Coverslips were then removed from the plates and washed three times in PBS. The cells on the coverslips were fixed for 10 minutes in 2% paraformaldehyde and washed three times in PBS. Next, the cells were

permeabilized with 0.1% Triton X-100 for 5 minutes and then washed twice with PBS. Anti-pan-HLA–I antibody (H1650, Sigma, 1:100 in PBS with 0.5% goat serum) was added for 1 hour at room temperature. Coverslips were washed three times and incubated with the anti-mouse secondary antibody Alexa flour 546 (A-11030, Invitrogen,1:500 in PBS with 0.5% goat serum), in light-protected conditions for 1 hour at room temperature. Coverslips were extensively washed and stained for 10 minutes with DAPI nuclear staining (1 µg/ml). Confocal images were acquired with a Zeiss LSM 800 confocal microscope using a 63x objective and processed with ImageJ (NIH).

Immunofluorescence imaging and image processing

Z-stack images were reconstructed from 20 serial confocal images 8.28 µm apart, acquired with a Zeiss LSM 800 confocal microscope using a 63x objective. Movie animation of the of the image stack was created to capture the 3D structure from different points of view using Imaris software (9.6.0, by Oxford Instruments). To identify the location of bacteria in relation to the melanoma cells we created surface objects for the melanoma cells and bacteria, and visualize them. The melanoma cells are transparent enabling the identification of bacteria within the cells. The threshold were manually tuned to best fit the borders of the objects. We separated between touching bacteria. The Supplementary Movies start with showing all three channels (bacteria, melanoma cells and nuclei) in 3D volume rendering (bland mode), then the objects surfaces are added to the bacteria and melanoma cells and finally only the surface is shown.

Correlative Light and Electron Microscopy (CLEM)

51AL and 55A3 ($5x10^6$) cells were fixed in 4% paraformaldehyde with 0.2% glutaraldehyde in 0.1 M cacodylate buffer (pH=7.4) for 1 hour at room temperature. Samples were then soaked overnight in 2.3 M sucrose and rapidly frozen in liquid nitrogen. Frozen ultrathin (70–90 nm) sections were cut with a diamond knife at −120°C on a Leica EM UC7 ultramicrotome. Sections were collected on 200-mesh Formvar-coated nickel grids.

Actinomyces odontolyticus and *Staphylococcus caprae* were incubated with anti-Lipoteichoic acid antibody (HM5018, Hycult Biotech) followed by anti-mouse secondary antibody labeled with AF-488 (115-545-146, Jackson Immunoresearch*). Fusobacterium nucleatum* was labeled by click-chemistry with AF-488 as previously described. In addition, sections were stained with DAPI (0.2 µg/ml for 30 minutes at room temperature).

Wide-field fluorescence images were taken in order to identify cells containing bacteria-AF488 using a VUTARA SR352 system (Bruker) in a buffer composed of 5 mM cysteamine, oxygen scavengers (7 µM glucose oxidase and 56 nM catalase) in 50 mM Tris with 10 mM NaCl and 10% glucose at pH 8.0. Images were recorded using 1.3 NA 60x silicon oil immersion objective (Olympus) with 405 and 488 nm lasers for excitation. AlexaFluor488 was used to identify the bacteria and the DAPI labeling was used to identify the nucleus of melanoma cells.
Contrast staining and embedding of the sections were performed as previously described [100]. The same grids were viewed with a FEI Tecnai SPIRIT (FEI, Eidhoven, Netherlands) transmission electron microscope operated at 120 kV and equipped with an OneView camera (Gatan). The AF-488 fluorescence was used to identify bacteria in the section and to provide a desired region for TEM imaging. Overlaying the fluorescence with the TEM images verified that the bacteria are located inside a cell. The TEM grid mesh was used to detect the general region of a selected cell, while the nucleus, marked with DAPI and easily identified in the TEM, was used as fiducial markers in order to overlay the fluorescence image and the TEM image with high accuracy [101].

Pooled stable melanoma cells with knockout of B2M and CIITA

Cloning of sgRNAs into the LentiCRISPRv2 vector was performed as described on (http://www.genome-engineering.org/crispr/). Briefly, the LentiCRISPRv2 plasmid (#52961, Addgene) was digested with BsmBI and gel-purified. DNA oligonucleotides (Invitrogen) were annealed and ligated into the digested vector. Specific sgRNA oligonucleotide sequences used for B2M F: 5’

GGCCGAGATGTCTCGTCCG 3', R: 5' CGGACGAGACATCTCGGCC 3'. And for CIITA F: 5' TCAACTGCGACCAGTTCAGC 3', R: 5' GCTGAACTGGTCGCAGTTGA 3' or 5' GATATTGGCATAAGCCTCCC3', R: 5' GGGAGGCTTATGCCAATATC 3'. Lentivirus was produced as previously described [102]. Lentivirus was produced by the co–transfection of 4.4 µg of the LentiCRISPRv2 vector together with 2.2 µg of PMD2.G (#12259, Addgene) and 3.4 µg of psPAX2 (#12260, Addgene) packaging plasmids into HEK293T cells in a 10 cm^2 dish using TurboFect (Thermo Fisher) as described by the manufacturer. After 60 hours, media was collected, filtered, aliquoted and stored at -80°C. 51AL and 55A3 cells were lentivirally transduced and selected with puromycin (1 µg/ml) for 14 days. Single clones were produced and expanded. To validate the B2M and CIITA KO the HLA-I and HLA-II expression was examine using flow cytometry (Supplementary Fig. 8e).

16S RNA FISH

FFPE tissue sections were deparaffinized, rehydrated and incubated in 70% ethanol at 4˚C for 2 hours. Slides were washed in 2XSSC (Ambion #AM9765) and incubated with proteinase K (10µg/ml in 2XSSC, Ambion #AM2546) for 10 minutes. Slides then underwent two washes with 2XSSC followed by two washes in 2XSSC with 15% formamide (Ambion #AM9342). Cy5 labelled Probes (EUB338 [103]- GCTGCCTCCCGTAGGAGT and non-specific complement probe - CGACGGAGGGCATCCTCA, at 1680nM) were hybridized overnight at 30˚C in 2XSSC, 10% Dextran sulfate (Sigma #D8906), 1mg/ml E.coli tRNA (Sigma #R4251), 0.02% BSA (Ambion #AM2616), 2mM Vanadyl-ribonucleoside (New England Biolabs #S1402S) and 15% formamide. Sections were washed for 30 minutes at 30˚C in 2XSSC with 15% formamide followed by a 30 minutes incubation step with 2XSSC, 15% formamide and DAPI at 30˚C. Slides were washed in 2XSSC, 10mM TRIS pH8 and 0.4% glucose and mounted with ProLong Gold Antifade Mountant (Life technologies #P36930). Confocal images were acquired with a Zeiss LSM 800 confocal microscope using a 20x and 10x objective and processed with ImageJ (NIH).

Bacterial peptides reactivity assay

Autologous TILs (10^6) were co-cultured for 5 hours with EBV-transformed human B-cells loaded with synthetic peptides (GenScript, 10 µg/ml for 2 hours at 37°C) at a 1:1 ratio. The peptides were tested with different B-cells to cover all HLA alleles to which they were predicted to bind. Cells were labeled with an IFNγ capture antibody (130-054-202, Miltenyi Biotec) and then incubated for 45 minutes at 37°C. Then cells were stained with the detection antibody, APC/Cy7-conjugated CD19 (clone HIB19, BioLegend) and SYTOX Blue Dead Cell Stain (ThermoFisher). The peptides found to be immunogenic in the IFNγ assay were also tested for CD69 expression on the TILs. For CD69 staining, after 6 hours of co-culture, the cells were stained with BV421-conjugated CD69 (clone FN50, BioLegend), APC-conjugated CD19 (clone HIB19, BioLegend) and SYTOX Green Dead Cell Stain (ThermoFisher). A BD LSR II flow cytometer (BD Biosciences) was used and the data was analyzed using FlowJo software (FlowJo, LLC). Cells were first gated for single cells (FSC-A / SSC-A, FSC-A / FSC-H, SSC-A / SSC-H), then live cells were selected, and TILs were gated as CD19-cells.

Flow cytometry

Supplementary Fig. 7a – digested tumor single cells were stained with AF647-conjugated CD45 (clone 2D1), APC-conjugated HLA-DR,DP,DQ (clone Tu39) and SYTOX Blue Dead Cell Stain (ThermoFisher). Cells were first gated for single cells (FSC-A / SSC-A, FSC-A / FSC-H, SSC-A / SSC-H), then live cells were selected. Unstained sample was used to determine the CD45 and HLA-II negative cells.

Supplementary Fig. 8a – THP1 cells that were differentiated to macrophages were stained with APC-conjugated HLA-A,B,C (clone W6/32) or APC-conjugated HLA-DR,DP,DQ (clone Tu39). THP1 cells and macrophages were stained with APC/Cy7-conjugated CD14 (clone HCD14). Cells were first gated for single cells and unstained cells were used as control.

Supplementary Fig. 8c-f – Cell lines were stained with APC-conjugated HLA-A,B,C (clone W6/32) or APC-conjugated HLA-DR,DP,DQ (clone Tu39). Cells were first gated for single cells and unstained cells were used as control.
All antibodies were obtained from BioLegend and diluted 1:50. A BD LSR II flow cytometer (BD Biosciences) was used and the data was analyzed using FlowJo software (FlowJo, LLC).

Methods references

51 Shukla, S. A. *et al.* Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. *Nat Biotechnol* **33**, 1152-1158, doi:10.1038/nbt.3344 (2015).

52 Fuks, G. *et al.* Combining 16S rRNA gene variable regions enables high-resolution microbial community profiling. *Microbiome* **6**, 17, doi:10.1186/s40168-017-0396-x (2018).

53 Paradis, E. & Schliep, K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526-528, doi:10.1093/bioinformatics/bty633 (2019).

54 Wickham, H. ggplot2: Elegant Graphics for Data Analysis. *Use R*, 1-212, doi:10.1007/978-0-387-98141-3 (2009).

55 Yu, G. C., Smith, D. K., Zhu, H. C., Guan, Y. & Lam, T. T. Y. GGTREE: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol* **8**, 28-36, doi:10.1111/2041-210x.12628 (2017).

56 Yu, G. C., Lam, T. T. Y., Zhu, H. C. & Guan, Y. Two Methods for Mapping and Visualizing Associated Data on Phylogeny Using Ggtree. *Mol Biol Evol* **35**, 3041-3043, doi:10.1093/molbev/msy194 (2018).

57 McMurdie, P. J. & Holmes, S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* **8**, e61217, doi:10.1371/journal.pone.0061217 (2013).

58 Wood, D. E. & Salzberg, S. L. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol* **15**, R46, doi:10.1186/gb-2014-15-3-r46 (2014).

59 Forster, S. C. *et al.* A human gut bacterial genome and culture collection for improved metagenomic analyses. *Nat Biotechnol* **37**, 186-192, doi:10.1038/s41587-018-0009-7 (2019).

60 Dixon, P. VEGAN, a package of R functions for community ecology. *J Veg Sci* **14**, 927-930, doi:DOI 10.1111/j.1654-1103.2003.tb02228.x (2003).

61 Kalaora, S. *et al.* Use of HLA peptidomics and whole exome sequencing to identify human immunogenic neo-antigens. *Oncotarget* **7**, 5110-5117, doi:10.18632/oncotarget.6960 (2016).

62 Kalaora, S. *et al.* Combined Analysis of Antigen Presentation and T-cell Recognition Reveals Restricted Immune Responses in Melanoma. *Cancer Discov* **8**, 1366-1375, doi:10.1158/2159-8290.CD-17-1418 (2018).

63 Kalaora, S. & Samuels, Y. Cancer Exome-Based Identification of Tumor Neo-Antigens Using Mass Spectrometry. *Methods Mol Biol* **1884**, 203-214, doi:10.1007/978-1-4939-8885-3_14 (2019).

64 Bassani-Sternberg, M. *et al.* Soluble plasma HLA peptidome as a potential source for cancer biomarkers. *Proc Natl Acad Sci U S A* **107**, 18769-18776, doi:10.1073/pnas.1008501107 (2010).
65 Ishihama, Y., Rappsilber, J., Andersen, J. S. & Mann, M. Microcolumns with self-assembled particle frits for proteomics. *J Chromatogr A* **979**, 233-239, doi:Pii S0021-9673(02)01402-4 Doi 10.1016/S0021-9673(02)01402-4 (2002).
66 Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658-1659, doi:10.1093/bioinformatics/btl158 (2006).
67 Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150-3152, doi:10.1093/bioinformatics/bts565 (2012).
68 Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology* **26**, 1367-1372, doi:10.1038/nbt.1511 (2008).
69 Vita, R. *et al.* The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Res* **47**, D339-D343, doi:10.1093/nar/gky1006 (2019).
70 Jurtz, V. *et al.* NetMHCpan-4.0: Improved Peptide-MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. *J Immunol* **199**, 3360-3368, doi:10.4049/jimmunol.1700893 (2017).
71 Nielsen, M. & Andreatta, M. NetMHCpan-3.0; improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets. *Genome Med* **8**, doi:ARTN 33 10.1186/s13073-016-0288-x (2016).
72 Hoof, I. *et al.* NetMHCpan, a method for MHC class I binding prediction beyond humans. *Immunogenetics* **61**, 1-13, doi:10.1007/s00251-008-0341-z (2009).
73 Jensen, K. K. *et al.* Improved methods for predicting peptide binding affinity to MHC class II molecules. *Immunology* **154**, 394-406, doi:10.1111/imm.12889 (2018).
74 Deutsch, E. W. *et al.* The ProteomeXchange consortium in 2020: enabling 'big data' approaches in proteomics. *Nucleic Acids Research* **48**, D1145-D1152, doi:10.1093/nar/gkz984 (2020).
75 Perez-Riverol, Y. *et al.* The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Research* **47**, D442-D450, doi:10.1093/nar/gky1106 (2019).
76 Krokhin, O. V. *et al.* An improved model for prediction of retention times of tryptic peptides in ion pair reversed-phase HPLC: its application to protein peptide mapping by off-line HPLC-MALDI MS. *Mol Cell Proteomics* **3**, 908-919, doi:10.1074/mcp.M400031-MCP200 (2004).
77 Gatto, L. & Lilley, K. S. MSnbase-an R/Bioconductor package for isobaric tagged mass spectrometry data visualization, processing and quantitation. *Bioinformatics* **28**, 288-289, doi:10.1093/bioinformatics/btr645 (2012).
78 Andreatta, M., Alvarez, B. & Nielsen, M. GibbsCluster: unsupervised clustering and alignment of peptide sequences. *Nucleic Acids Res* **45**, W458-W463, doi:10.1093/nar/gkx248 (2017).
79 Andreatta, M., Lund, O. & Nielsen, M. Simultaneous alignment and clustering of peptide data using a Gibbs sampling approach. *Bioinformatics* **29**, 8-14, doi:10.1093/bioinformatics/bts621 (2013).
80 Thomsen, M. C. F. & Nielsen, M. Seq2Logo: a method for construction and visualization of amino acid binding motifs and sequence profiles including sequence weighting,

pseudo counts and two-sided representation of amino acid enrichment and depletion. *Nucleic Acids Research* **40**, W281-W287, doi:10.1093/nar/gks469 (2012).
81 Rammensee, H. G., Bachmann, J., Emmerich, N. P. N., Bachor, O. A. & Stevanovic, S. SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics* **50**, 213-219, doi:DOI 10.1007/s002510050595 (1999).
82 Bassani-Sternberg, M. *et al.* Deciphering HLA-I motifs across HLA peptidomes improves neo-antigen predictions and identifies allostery regulating HLA specificity. *Plos Comput Biol* **13**, doi:ARTN e100572510.1371/journal.pcbi.1005725 (2017).
83 Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J Mol Biol* **157**, 105-132, doi:10.1016/0022-2836(82)90515-0 (1982).
84 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120, doi:10.1093/bioinformatics/btu170 (2014).
85 Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19**, 455-477, doi:10.1089/cmb.2012.0021 (2012).
86 Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072-1075, doi:10.1093/bioinformatics/btt086 (2013).
87 Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25**, 1043-1055, doi:10.1101/gr.186072.114 (2015).
88 Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997* (2013).
89 Kang, D. D., Froula, J., Egan, R. & Wang, Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* **3**, e1165, doi:10.7717/peerj.1165 (2015).
90 Parks, D. H. *et al.* A complete domain-to-species taxonomy for Bacteria and Archaea. *Nat Biotechnol*, doi:10.1038/s41587-020-0501-8 (2020).
91 Chaumeil, P. A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics*, doi:10.1093/bioinformatics/btz848 (2019).
92 Jain, C., Rodriguez, R. L., Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* **9**, 5114, doi:10.1038/s41467-018-07641-9 (2018).
93 Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J* **11**, 2864-2868, doi:10.1038/ismej.2017.126 (2017).
94 Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068-2069, doi:10.1093/bioinformatics/btu153 (2014).
95 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J Mol Biol* **215**, 403-410, doi:10.1016/S0022-2836(05)80360-2 (1990).
96 O'Leary, N. A. *et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* **44**, D733-745, doi:10.1093/nar/gkv1189 (2016).
97 Benson, D. A. *et al.* GenBank. *Nucleic Acids Res* **41**, D36-42, doi:10.1093/nar/gks1195 (2013).
98 Zaharia, M. *et al.* Faster and More Accurate Sequence Alignment with SNAP. *arXiv preprint arXiv* **1111** (2011).

99 Liu, H., Dong, H., Jiang, L., Li, Z. & Ma, X. Bleomycin inhibits proliferation and induces apoptosis in TPC-1 cells through reversing M2-macrophages polarization. *Oncol Lett* **16**, 3858-3866, doi:10.3892/ol.2018.9103 (2018).

100 Tokuyasu, K. T. Application of cryoultramicrotomy to immunocytochemistry. *J Microsc* **143**, 139-149 (1986).

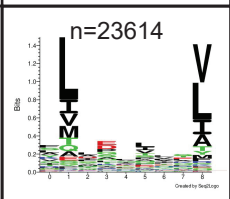
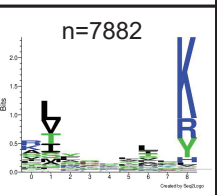
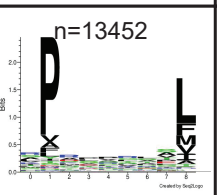
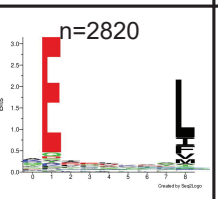
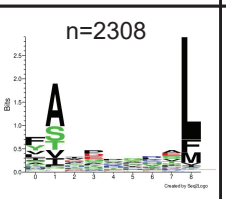
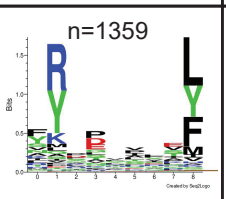
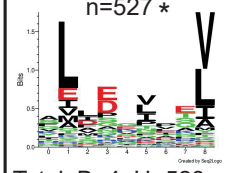
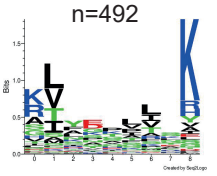
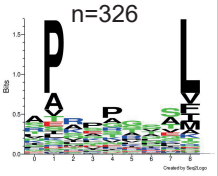
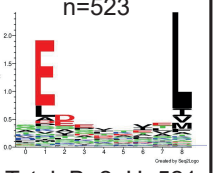
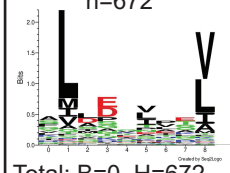
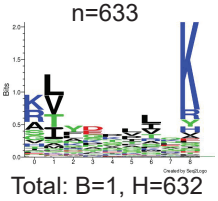
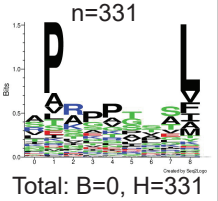
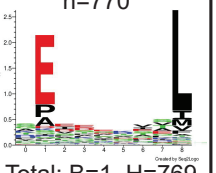
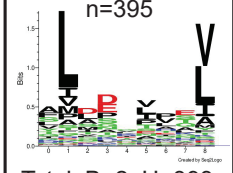
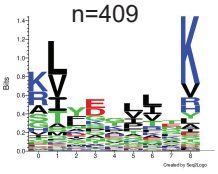
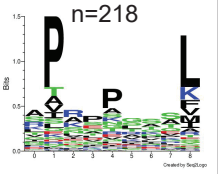
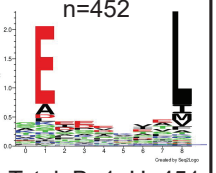
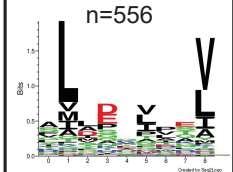
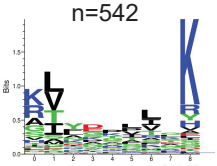
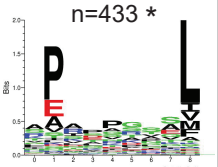
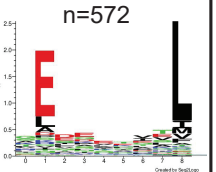
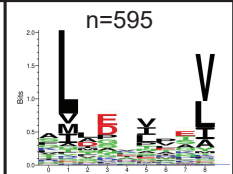
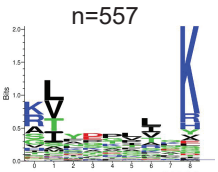
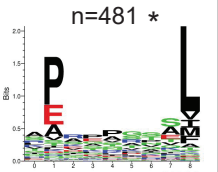
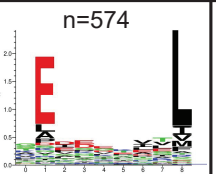
101 Abada, A., Levin-Zaidman, S., Porat, Z., Dadosh, T. & Elazar, Z. SNARE priming is essential for maturation of autophagosomes but not for their formation. *Proc Natl Acad Sci U S A* **114**, 12749-12754, doi:10.1073/pnas.1705572114 (2017).

102 Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84-87, doi:10.1126/science.1247005 (2014).

103 Amann, R. I. *et al.* Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl Environ Microbiol* **56**, 1919-1925, doi:10.1128/AEM.56.6.1919-1925.1990 (1990).

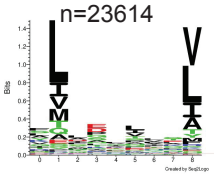
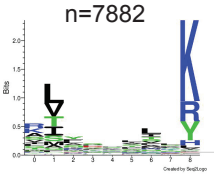
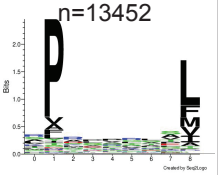
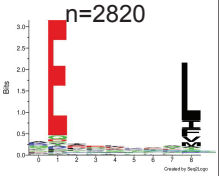
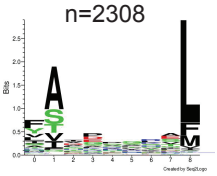
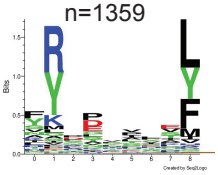
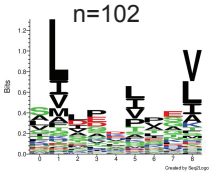
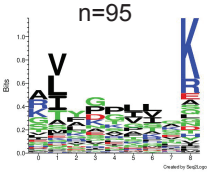
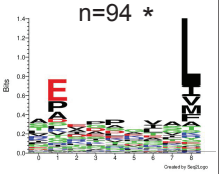
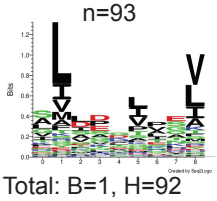
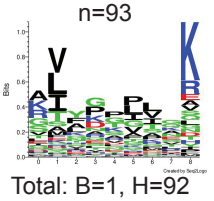
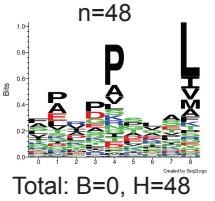
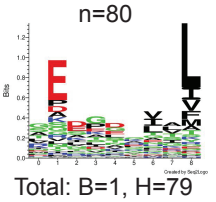
Supplementary Figure 1

Gibbs clustering of identified HLA peptides 42RF

	A*02:01	A*03:01	B*07:02	B*40:01	C*03:04	C*07:02	Trash
IEDB	 n=23614	 n=7882	 n=13452	 n=2820	 n=2308	 n=1359	
Campylobacter concisus	 n=527 * Total: B=4, H=523	 n=492 Total: B=2, H=490	 n=326 Total: B=0, H=326	 n=523 Total: B=2, H=521			n=10 B=0 H=10
Campylobacter lanienae	 n=672 Total: B=0, H=672	 n=633 Total: B=1, H=632	 n=331 Total: B=0, H=331	 n=770 Total: B=1, H=769			n=12 B=0 H=12
Klebsiella oxytoca	 n=395 Total: B=2, H=393	 n=409 Total: B=2, H=407	 n=218 Total: B=0, H=218	 n=452 Total: B=1, H=451			n=7 B=1 H=6
Paracoccus marcusii	 n=556 Total: B=1, H=555	 n=542 Total: B=0, H=542	 n=433 * Total: B=0, H=433	 n=572 Total: B=0, H=572			n=10 B=0 H=10
Shewanella decolorationis	 n=595 Total: B=3, H=592	 n=557 Total: B=1, H=556	 n=481 * Total: B=2, H=479	 n=574 Total: B=1, H=573			n=10 B=1 H=9

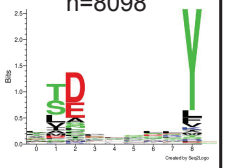
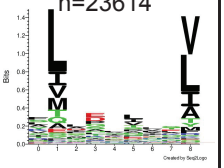
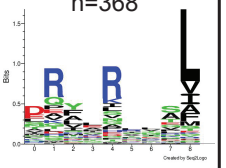
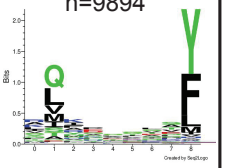
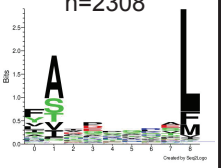
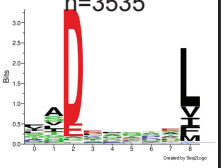
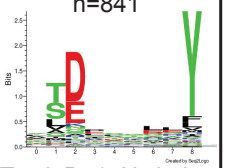
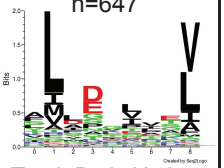
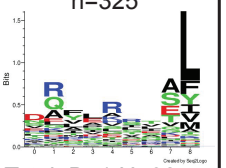
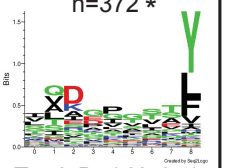
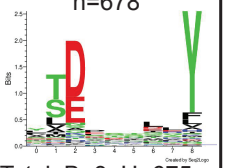
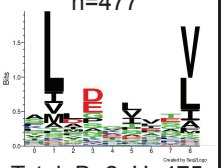
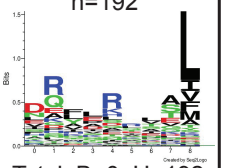
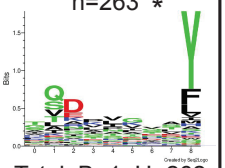
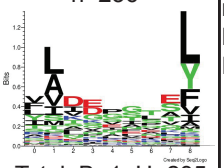
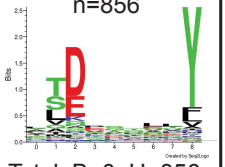
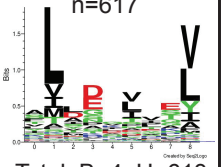
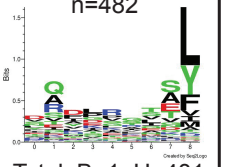
* This cluster contains also peptides with the B*40:01 motif

Gibbs clustering of identified HLA peptides 42RS

	A*02:01	A*03:01	B*07:02	B*40:01	C*03:04	C*07:02	Trash
IEDB	n=23614 	n=7882 	n=13452 	n=2820 	n=2308 	n=1359 	
Enterobacter cancerogenus	n=102  Total: B=0, H=102	n=95  Total: B=0, H=95		n=94 *  Total: B=1, H=93			n=0 B=0 H=0
Sphingomonas melonis	n=93  Total: B=1, H=92	n=93  Total: B=1, H=92	n=48  Total: B=0, H=48	n=80  Total: B=1, H=79			n=0 B=0 H=0

* This cluster contains also peptides with the C*03:04 motif

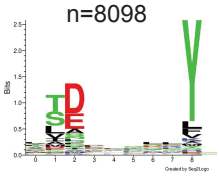
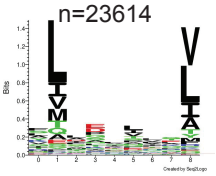
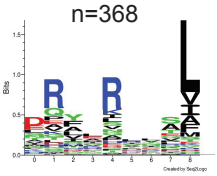
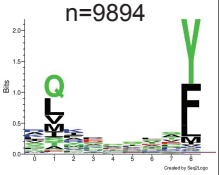
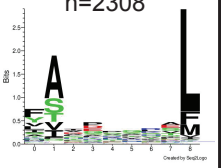
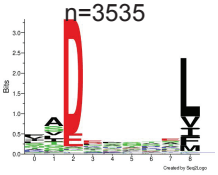
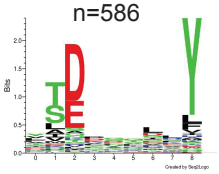
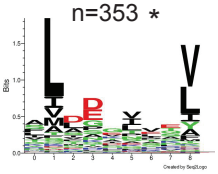
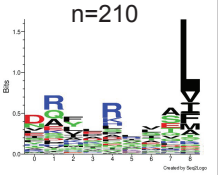
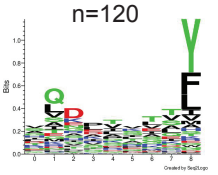
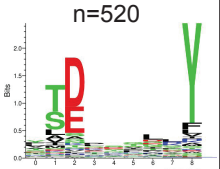
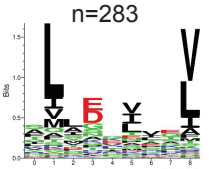
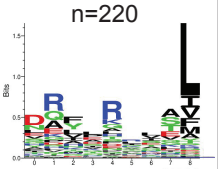
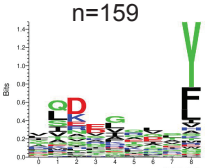
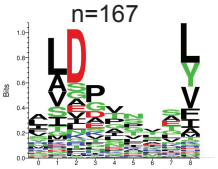
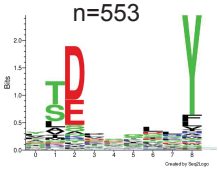
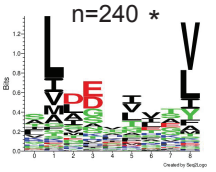
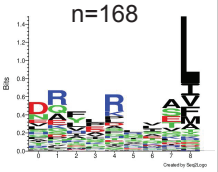
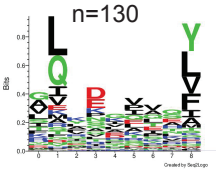
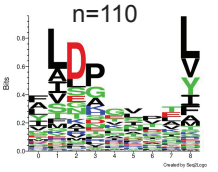
Gibbs clustering of identified HLA peptides 27

	A*01:01	A*02:01	B*14:02	B*15:01	C*03:04	C*08:02	Trash
IEDB	n=8098 	n=23614 	n=368 	n=9894 	n=2308 	n=3535 	
Actinomyces odontolyticus	n=841  Total: B=1, H=840	n=647  Total: B=2, H=645	n=325  Total: B=1 H=324	n=372 *  Total: B=1 H=371			n=8 B=0 H=8
Fusobacterium nucleatum	n=678  Total: B=3, H=675	n=477  Total: B=2, H=475	n=192  Total: B=0, H=192	n=263 *  Total: B=1, H=262		n=236 **  Total: B=1, H=235	n=6 B=0 H=6
Schaalia odontolytica	n=856  Total: B=0, H=856	n=617  Total: B=4, H=613	n=482  Total: B=1, H=481				n=22 B=0 H=22

* This cluster contains also peptides with the A*02:01 motif

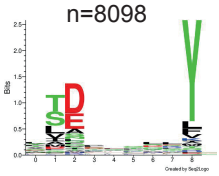
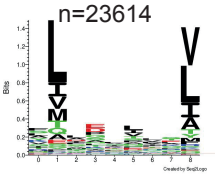
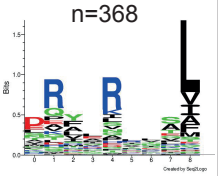
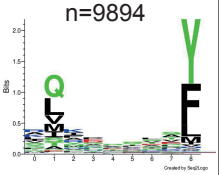
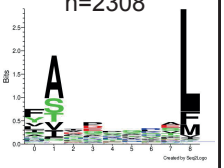
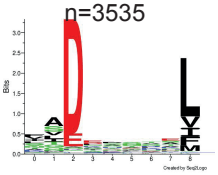
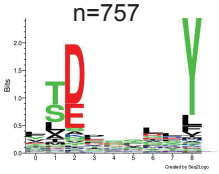
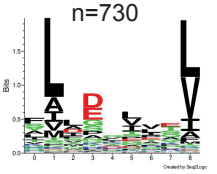
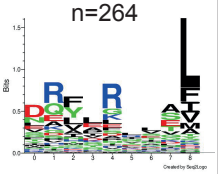
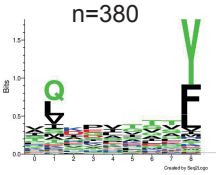
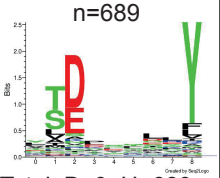
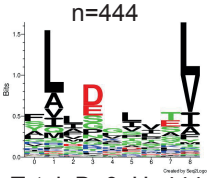
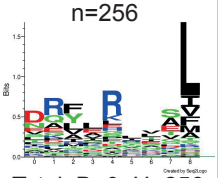
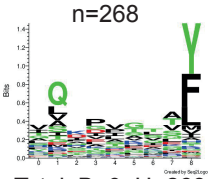
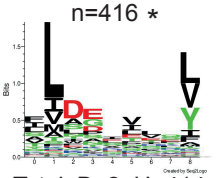
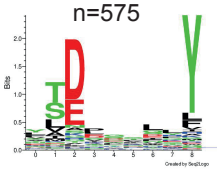
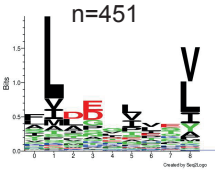
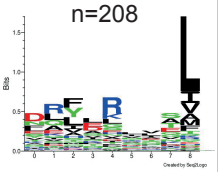
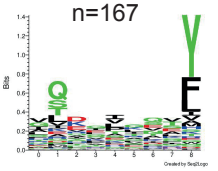
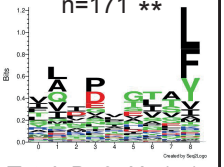
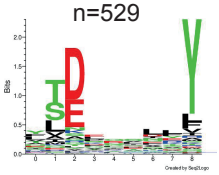
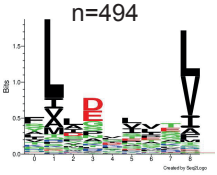
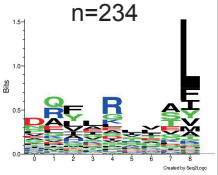
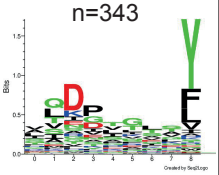
** This cluster contains also peptides with the A*01:01/B*15:01 motif

Gibbs clustering of identified HLA peptides 51AL

	A*01:01	A*02:01	B*14:02	B*15:01	C*03:04	C*08:02	Trash
IEDB	n=8098 	n=23614 	n=368 	n=9894 	n=2308 	n=3535 	
Corynebacterium afermentans	n=586  Total: B=1, H=585	n=353 *  Total: B=1, H=352	n=210  Total: B=1, H=209	n=120  Total: B=1, H=119			n=6 B=0 H=6
Eikenella corrodens	n=520  Total: B=0, H=520	n=283  Total: B=0, H=283	n=220  Total: B=0, H=220	n=159  Total: B=0, H=159		n=167  Total: B=1, H=166	n=1 B=1 H=0
Staphylococcus aureus	n=553  Total: B=1, H=552	n=240 *  Total: B=0, H=240	n=168  Total: B=2, H=166	n=130  Total: B=0, H=130		n=110  Total: B=0, H=110	n=2 B=0 H=2

* This cluster contains also peptides with the C*08:02 motif

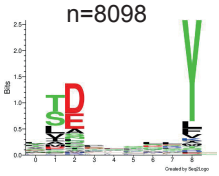
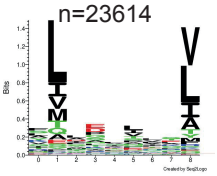
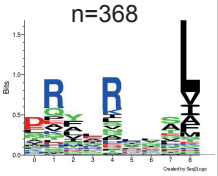
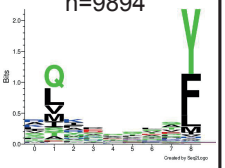
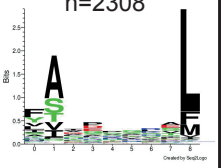
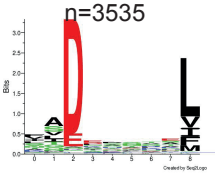
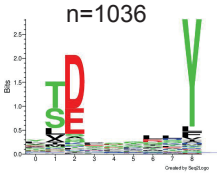
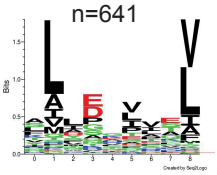
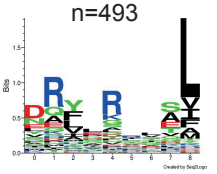
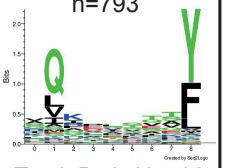
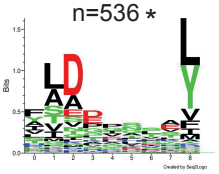
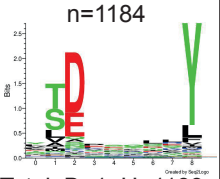
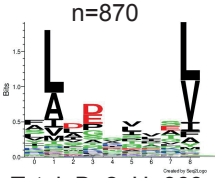
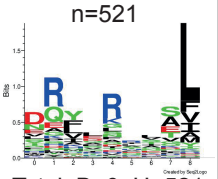
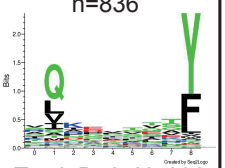
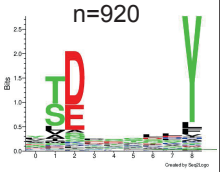
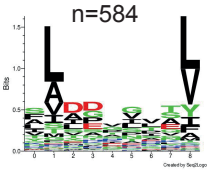
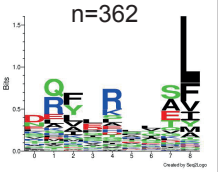
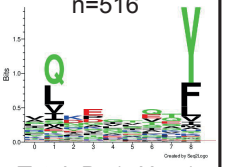
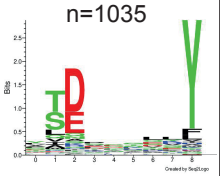
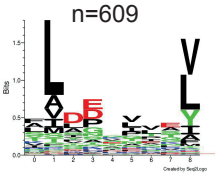
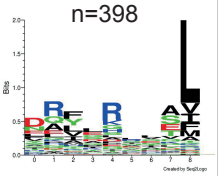
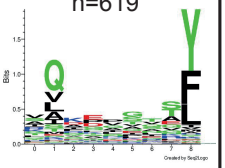
Gibbs clustering of identified HLA peptides 51BR-1

	A*01:01	A*02:01	B*14:02	B*15:01	C*03:04	C*08:02	Trash
IEDB	n=8098 	n=23614 	n=368 	n=9894 	n=2308 	n=3535 	
Brachybacterium alimentarium	n=757  Total: B=0, H=757	n=730  Total: B=3, H=727	n=264  Total: B=1, H=263	n=380  Total: B=0, H=380			n=11 B=0 H=11
Eikenella corrodens	n=689  Total: B=0, H=689	n=444  Total: B=0, H=444	n=256  Total: B=0, H=256	n=268  Total: B=0, H=268		n=416 *  Total: B=2, H=414	n=8 B=1 H=7
Fusobacterium nucleatum	n=575  Total: B=0, H=575	n=451  Total: B=2, H=449	n=208  Total: B=0, H=208	n=167  Total: B=0, H=167	n=171 **  Total: B=0, H=171		n=3 B=0 H=3
Staphylococcus aureus	n=529  Total: B=1, H=528	n=494  Total: B=4, H=490	n=234  Total: B=1, H=233	n=343  Total: B=1, H=342			n=8 B=1 H=7

* This cluster contains also peptides with the A*02:01 motif

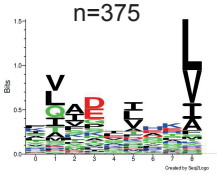
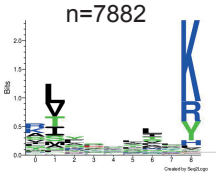
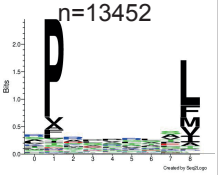
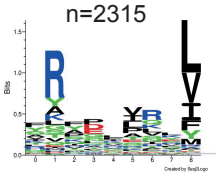
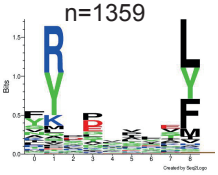
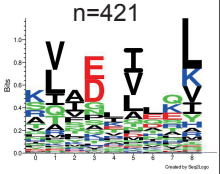
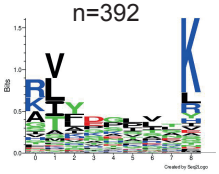
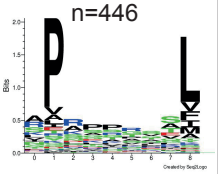
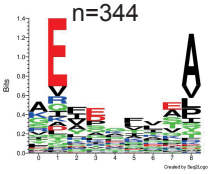
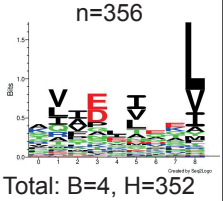
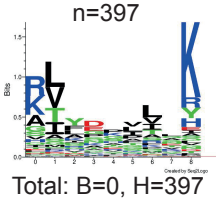
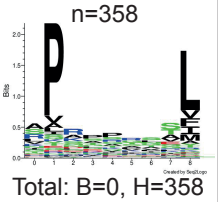
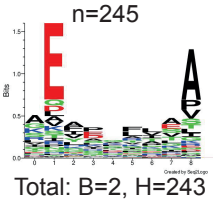
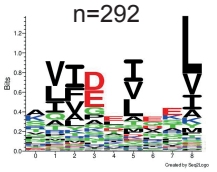
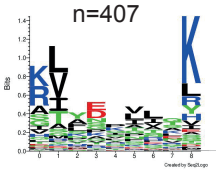
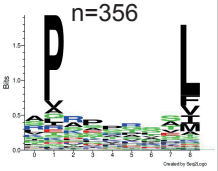
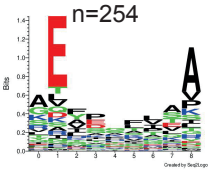
** This cluster contains also peptides with the B*15:01 motif

Gibbs clustering of identified HLA peptides 51BR-2

	A*01:01	A*02:01	B*14:02	B*15:01	C*03:04	C*08:02	Trash
IEDB	n=8098 	n=23614 	n=368 	n=9894 	n=2308 	n=3535 	
Brachybacterium alimentarium	n=1036  Total: B=1, H=1035	n=641  Total: B=5, H=636	n=493  Total: B=0, H=493	n=793  Total: B=0, H=793		n=536 *  Total: B=5, H=531	n=22 B=1 H=21
Eikenella corrodens	n=1184  Total: B=1, H=1183	n=870  Total: B=2, H=868	n=521  Total: B=0, H=521	n=836  Total: B=0, H=836			n=21 B=4 H=17
Fusobacterium nucleatum	n=920  Total: B=1, H=919	n=584  Total: B=6, H=578	n=362  Total: B=0, H=362	n=516  Total: B=1, H=515			n=10 B=0 H=10
Staphylococcus aureus	n=1035  Total: B=2, H=1033	n=609  Total: B=2, H=607	n=398  Total: B=0, H=398	n=619  Total: B=1, H=618			n=19 B=2 H=17

* This cluster contains also peptides with the A*01:01 motif

Gibbs clustering of identified HLA peptides 55A3

	A*02:05	A*03:01	B*07:02	B*40:32	C*06:02	C*07:02	Trash
IEDB	n=375 	n=7882 	n=13452 	N/A *	n=2315 	n=1359 	
Kingella denitrificans	n=421  Total: B=1, H=420	n=392  Total: B=0, H=392	n=446  Total: B=0, H=446	n=344  Total: B=0, H=344			n=29 B=0 H=29
Staphylococcus capitis	n=356  Total: B=4, H=352	n=397  Total: B=0, H=397	n=358  Total: B=0, H=358	n=245  Total: B=2, H=243			n=7 B=1 H=6
Streptococcus gordonii	n=292  Total: B=0, H=292	n=407  Total: B=1, H=406	n=356  Total: B=0, H=356	n=254  Total: B=0, H=254			n=18 B=0 H=18

* There are no peptides in the IEDB to create a motif. These clusters were assigned to this allele as they are similar to the B*40 motif with E in second position and L/A in ninth position (SYFPEITHI)

Gibbs clustering of identified HLA peptides 55A7-1

	A*02:05	A*03:01	B*07:02	B*40:32	C*06:02	C*07:02	Trash
IEDB	n=375	n=7882	n=13452	N/A **	n=2315	n=1359	
Clostridium clostridioforme	n=253 Total: B=1, H=252	n=333 Total: B=1, H=332	n=308 Total: B=0, H=308	n=127 Total: B=1 H=126			n=9 B=0 H=9
Dialister invisus	n=283 Total: B=0, H=283	n=337 Total: B=0, H=337	n=349 Total: B=0, H=349	n=134 Total: B=0, H=134			n=12 B=1 H=11
Dialister pneumosintes	n=252 Total: B=0, H=252	n=304 Total: B=0, H=304	n=330 Total: B=1, H=329	n=192 *** Total: B=1, H=191			n=15 B=0 H=15
Enterobacter cancerogenus	n=233 Total: B=0, H=233	n=327 Total: B=1, H=326	n=275 Total: B=2, H=273		n=111 * Total: B=0, H=111		n=10 B=0 H=10
Enterobacter hormaechei	n=284 Total: B=0, H=284	n=406 Total: B=1, H=405	n=298 Total: B=0, H=298				n=16 B=0 H=16
Veillonella dispar	n=279 Total: B=0, H=279	n=361 Total: B=0, H=361	n=343 Total: B=0, H=343	n=83 Total: B=0, H=83			n=14 B=1 H=13
Veillonella parvula	n=267 Total: B=1, H=266	n=343 Total: B=0, H=343	n=310 Total: B=0, H=310	n=101 Total: B=0, H=101			n=8 B=0 H=8

* This cluster contains also peptides with the A*02:05 motif

** There are no peptides in the IEDB to create a motif. These clusters were assigned to this allele as they are similar to the B*40 motif (E in second position and L/A in ninth position)

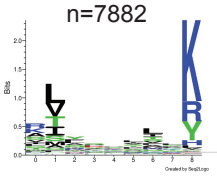
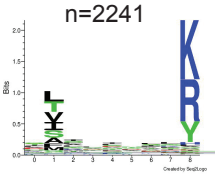
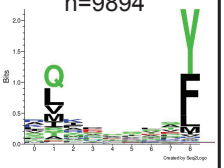
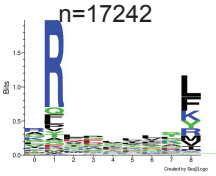
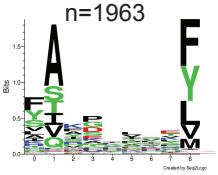
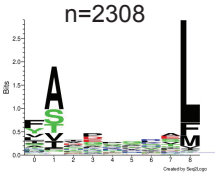
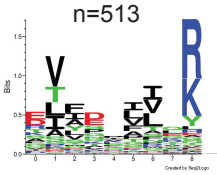
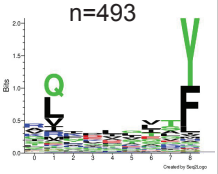
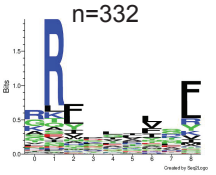
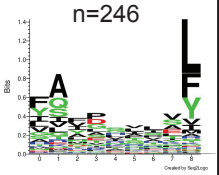
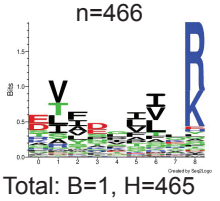
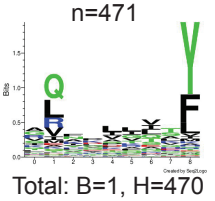
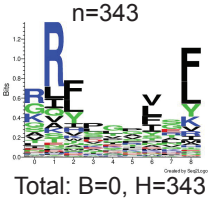
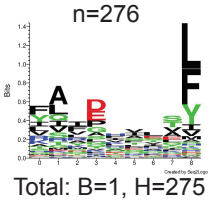
*** This cluster contains also peptides with the A*03:01 motif

Gibbs clustering of identified HLA peptides 55B3

	A*02:05	A*03:01	B*07:02	B*40:32	C*06:02	C*07:02	Trash
IEDB	n=375	n=7882	n=13452	N/A *	n=2315	n=1359	
Gardnerella vaginalis	n=524 Total: B=2, H=522	n=596 Total: B=2, H=594	n=563 Total: B=0, H=563	n=489 Total: B=2, H=487			n=24 B=1 H=23
Sphingomonas dokdonensis	n=848 Total: B=2, H=846	n=478 Total: B=2, H=476	n=682 Total: B=1, H=681	n=620 Total: B=1, H=619			n=41 B=0 H=41
Sphingomonas melonis	n=428 Total: B=2, H=426	n=855 Total: B=0, H=855	n=644 Total: B=0, H=644	n=582 Total: B=3, H=579			n=23 B=2 H=21
Staphylococcus capitis	n=606 Total: B=2, H=604	n=624 Total: B=0, H=624	n=632 Total: B=4, H=628	n=548 Total: B=1, H=547			n=24 B=1 H=23
Staphylococcus caprae	n=567 Total: B=1, H=566	n=693 Total: B=0, H=693	n=618 Total: B=0, H=618	n=663 Total: B=0, H=663			n=29 B=1 H=28

* There are no peptides in the IEDB to create a motif. These clusters were assigned to this allele as they are similar to the B*40 motif (E in second position and L/A in ninth position)

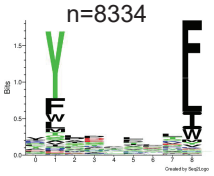
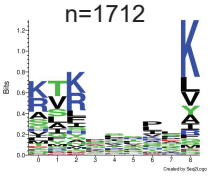
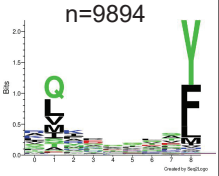
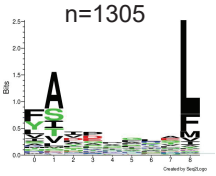
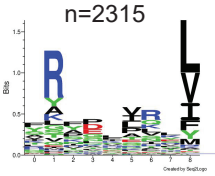
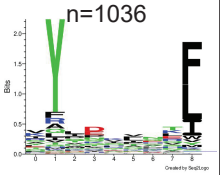
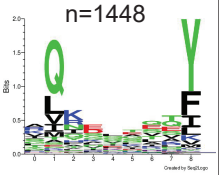
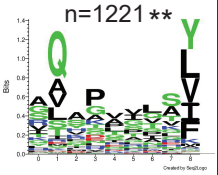
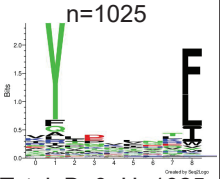
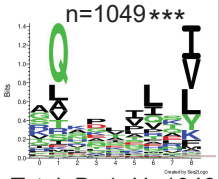
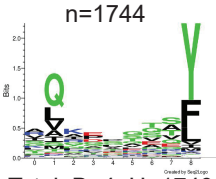
Gibbs clustering of identified HLA peptides 70.1

	A*03:01	A*68:01 *	B*15:01	B*27:05	C*02:02 **	C*03:04 **	Trash
IEDB	n=7882 	n=2241 	n=9894 	n=17242 	n=1963 	n=2308 	
Staphylococcus capitis	n=513  Total: B=2, H=511		n=493  Total: B=2, H=491	n=332  Total: B=0, H=332	n=246  Total: B=4, H=242		n=12 B=1 H=11
Streptococcus gordonii	n=466  Total: B=1, H=465		n=471  Total: B=1, H=470	n=343  Total: B=0, H=343	n=276  Total: B=1, H=275		n=9 B=0 H=9

* As the motif of A*03:01 and A*68:01 are similar, all peptides that match to these alleles were assigned to the same cluster found under A*03:01

** As the motif of C*02:02 and C*03:04 are similar, all peptides that match to these alleles were assigned to the same cluster found under C*02:02

Gibbs clustering of identified HLA peptides 86B

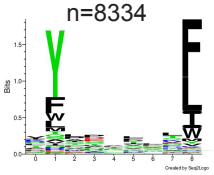
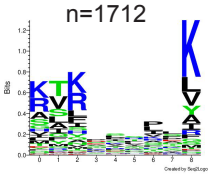
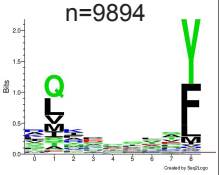
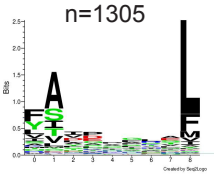
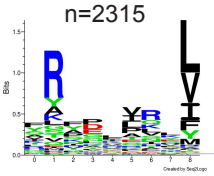
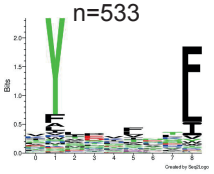
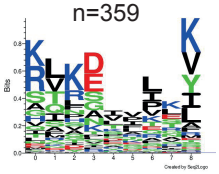
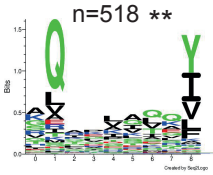
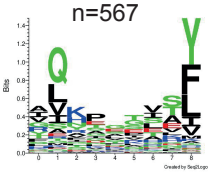
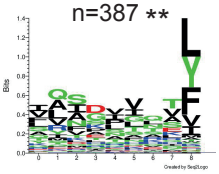
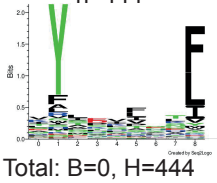
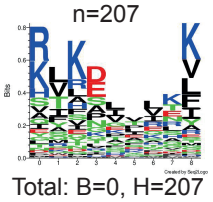
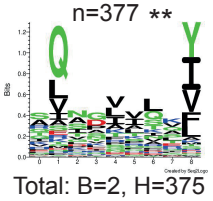
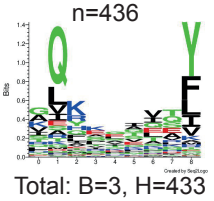
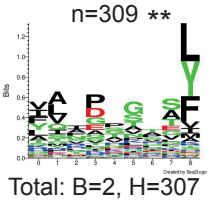
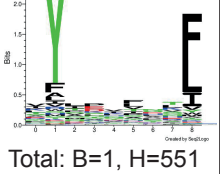
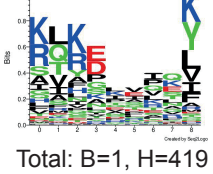
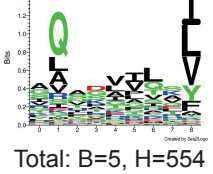
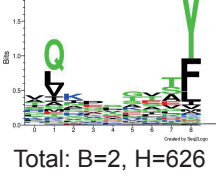
	A*24:02	A*30:01	B*13:02	B*15:01	C*03:03	C*06:02	Trash
IEDB	n=8334 	n=1712 	N/A *	n=9894 	n=1305 	n=2315 	
Corynebacterium humireducens	n=1036  Total: B=0, H=1036			n=1448  Total: B=0, H=1448	n=1221 **  Total: B=8, H=1213		n=132 B=2 H=130
Staphylococcus lugdunensis	n=1025  Total: B=0, H=1025		n=1049 ***  Total: B=1, H=1048	n=1744  Total: B=4, H=1740			n=114 B=1 H=113

* There are no peptides in the IEDB to create a motif. The motif for this allele was found in Bassani-Sternberg et al. 2017, with Q in second position and I/V/L in ninth position

** This cluster contains also peptides with the B*13:02/B*15:01 motif

*** This cluster contains also peptides with the B*15:01/C*03:03 motif

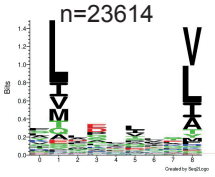
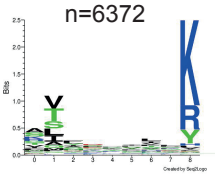
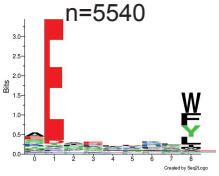
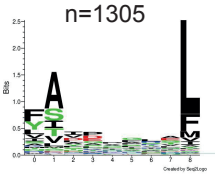
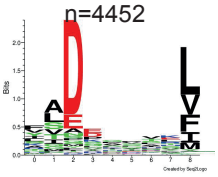
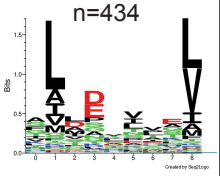
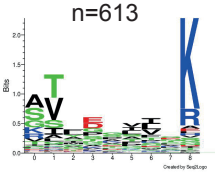
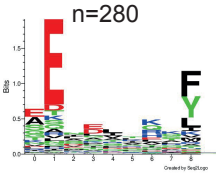
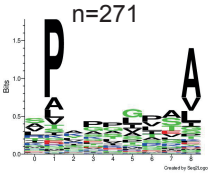
Gibbs clustering of identified HLA peptides 86B2

	A*24:02	A*30:01	B*13:02	B*15:01	C*03:03	C*06:02	Trash
IEDB	n=8334 	n=1712 	N/A *	n=9894 	n=1305 	n=2315 	
Acinetobacter cumulans	n=533  Total: B=0, H=533	n=359  Total: B=0, H=359	n=518 **  Total: B=1, H=517	n=567  Total: B=1, H=566	n=387 **  Total: B=2, H=385		n=40 B=0 H=40
Bacteroides vulgatus	n=444  Total: B=0, H=444	n=207  Total: B=0, H=207	n=377 **  Total: B=2, H=375	n=436  Total: B=3, H=433	n=309 **  Total: B=2, H=307		n=16 B=0 H=16
Staphylococcus capitis	n=552  Total: B=1, H=551	n=420  Total: B=1, H=419	n=559  Total: B=5, H=554	n=628  Total: B=2, H=626			n=45 B=4 H=41

* There are no peptides in the IEDB to create a motif. The motif for this allele was found in Bassani-Sternberg et al. 2017, with Q in second position and I/V/L in ninth position

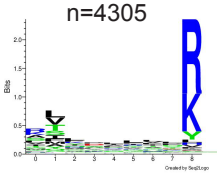
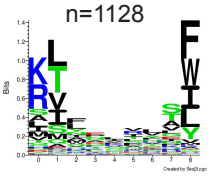
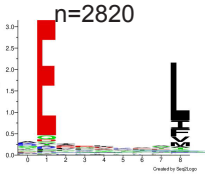
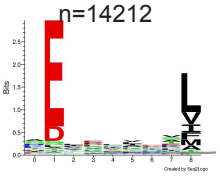
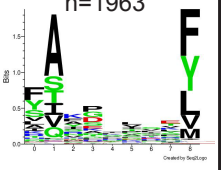
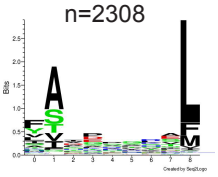
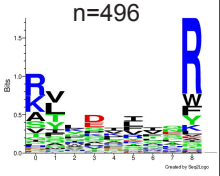
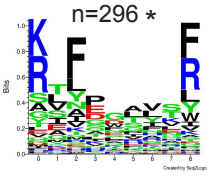
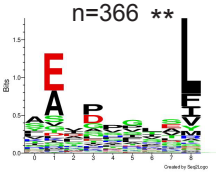
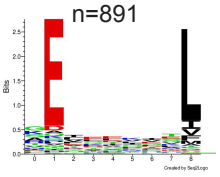
** This cluster contains also peptides with the B*15:01 motif

Gibbs clustering of identified HLA peptides 92B3

	A*02:01	A*11:01	B*44:02	B*55:01	C*03:03	C*05:01	Trash
IEDB	n=23614 	n=6372 	n=5540 	N/A *	n=1305 	n=4452 	
Fusobacterium nucleatum	n=434  Total: B=3, H=431	n=613  Total: B=1, H=612	n=280  Total: B=2, H=278	n=271  Total: B=0, H=271			n=7 B=1 H=6

* There are no peptides in the IEDB to create a motif. These clusters were assigned to this allele as they are similar to the B*55:01 motif as described in SYFPEITHI with P in second position

Gibbs clustering of identified HLA peptides 112

	A*31:01	A*32:01	B*40:01	B*40:02	C*02:02	C*03:04	Trash
IEDB	n=4305 	n=1128 	n=2820 	n=14212 	n=1963 	n=2308 	
Staphylococcus aureus	n=496  Total: B=1, H=495	n=296 *  Total: B=2, H=294	n=366 **  Total: B=2, H=364	n=891  Total: B=7, H=884			n=9 B=0 H=9

* This cluster contains also peptides with the A*31:01 motif

** This cluster contains also peptides with the C*02:02/C*03:04 motif

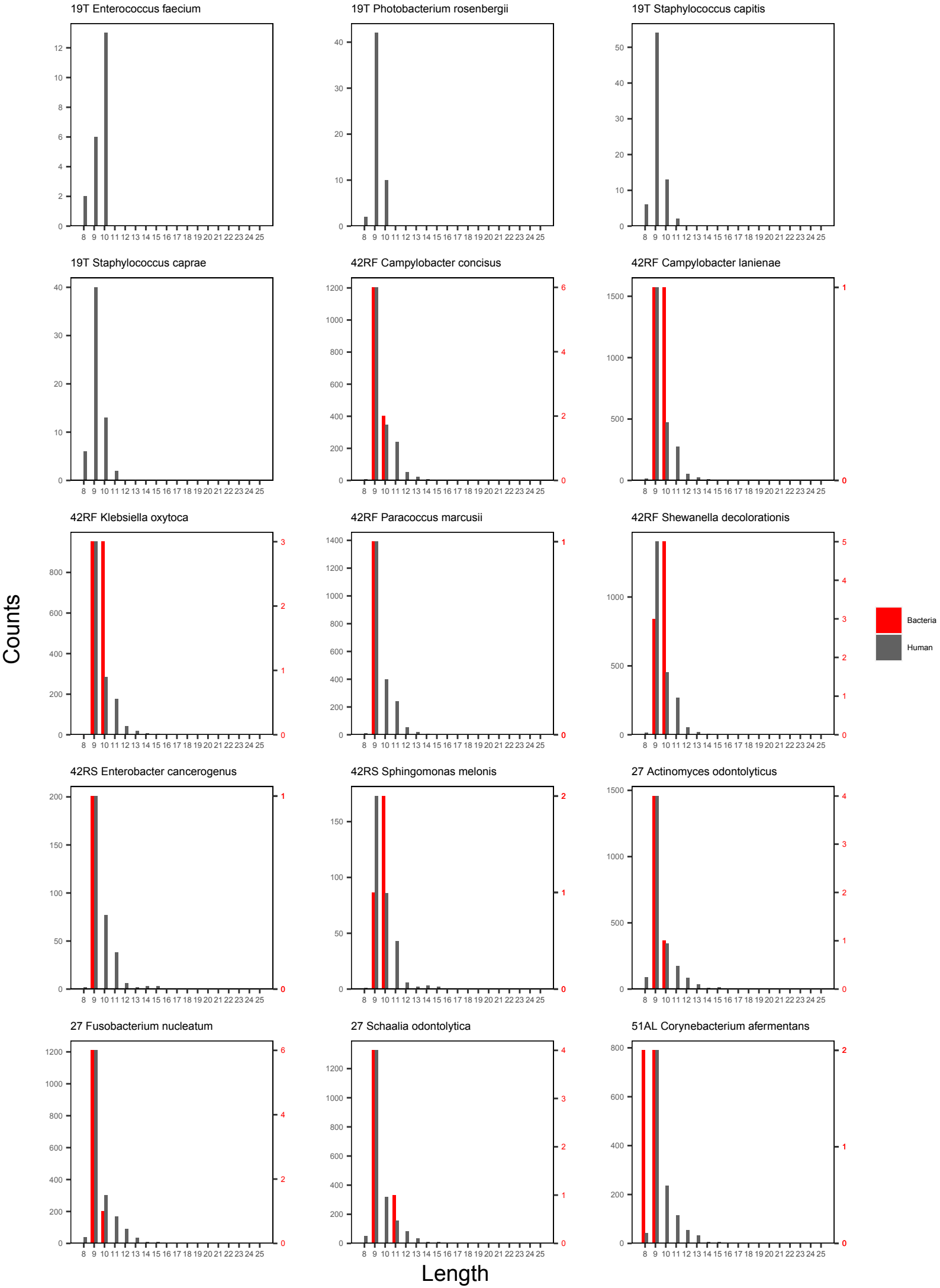
Gibbs clustering of identified HLA peptides 152A2

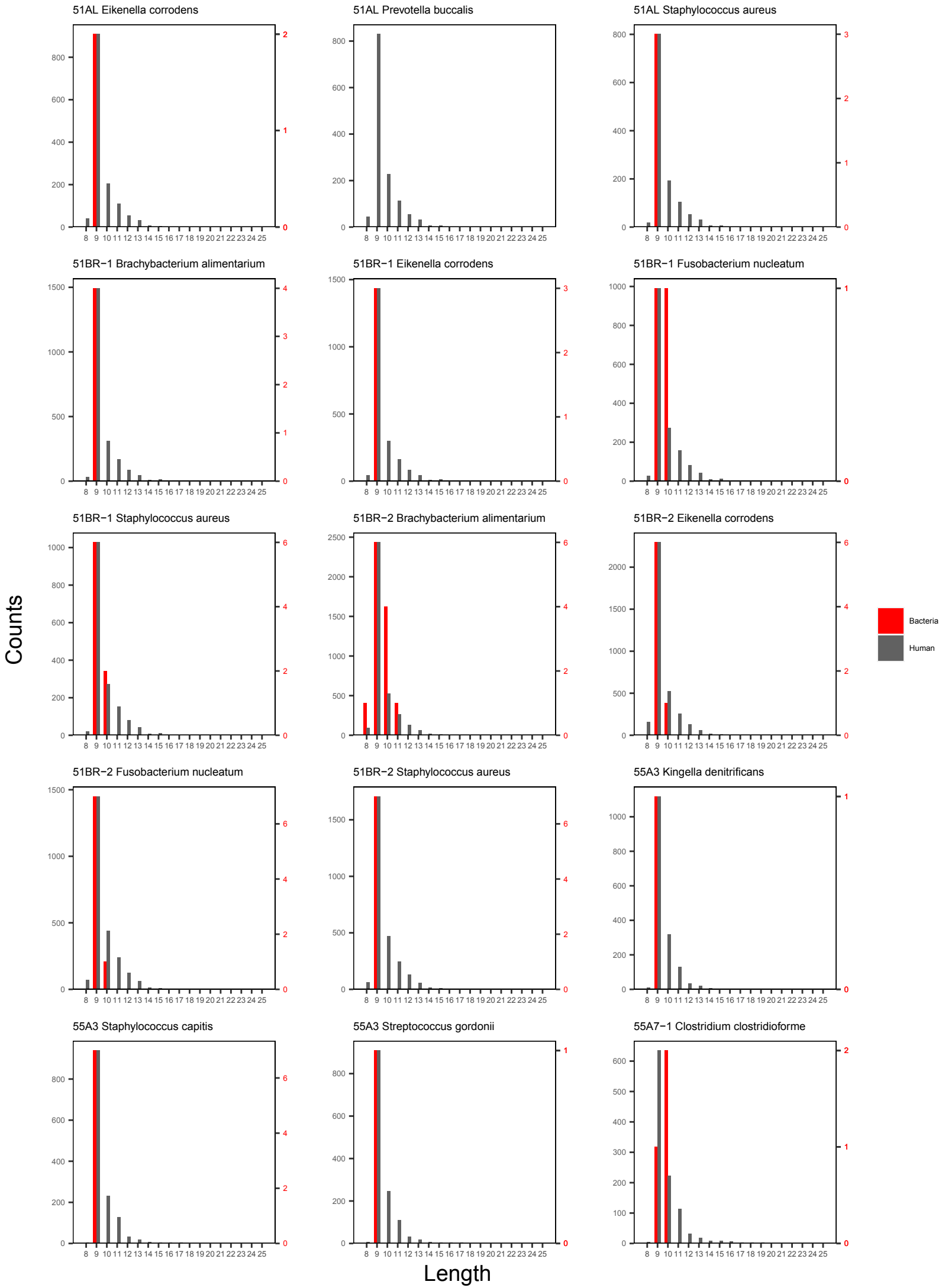
	A*03:01	B*08:01	B*35:01	C*04:01	C*07:01	Trash
IEDB	n=7882	n=5989	n=3642	n=2424	n=455	
Bacteroides dorei	n=1117 Total: B=5, H=1112	n=313 Total: B=2, H=311	n=643 Total: B=0 H=643			n=33 B=1 H=32
Bacteroides ovatus	n=1016 Total: B=2, H=1014	n=340 Total: B=2, H=338	n=519 Total: B=2, H=517			n=17 B=1 H=16
Bacteroides vulgatus	n=1067 Total: B=2, H=1065		n=547 Total: B=2, H=545	n=333 * Total: B=3, H=330		n=21 B=1 H=20
Campylobacter concisus	n=1028 Total: B=2, H=1026	n=400 Total: B=1, H=399	n=568 Total: B=0, H=568			n=44 B=1 H=43
Eikenella corrodens	n=1108 Total: B=1, H=1107	n=294 Total: B=2, H=292	n=586 Total: B=2, H=584	n=215 Total: B=0, H=215		n=15 B=0 H=15
Fusobacterium nucleatum	n=868 Total: B=1, H=867	n=263 Total: B=3, H=260	n=464 Total: B=3, H=461	n=162 Total: B=0, H=162		n=11 B=1 H=10
Veillonella dispar	n=1025 Total: B=0, H=1025	n=324 Total: B=1, H=323	n=587 Total: B=1, H=586	n=303 Total: B=1, H=302		n=18 B=0 H=18

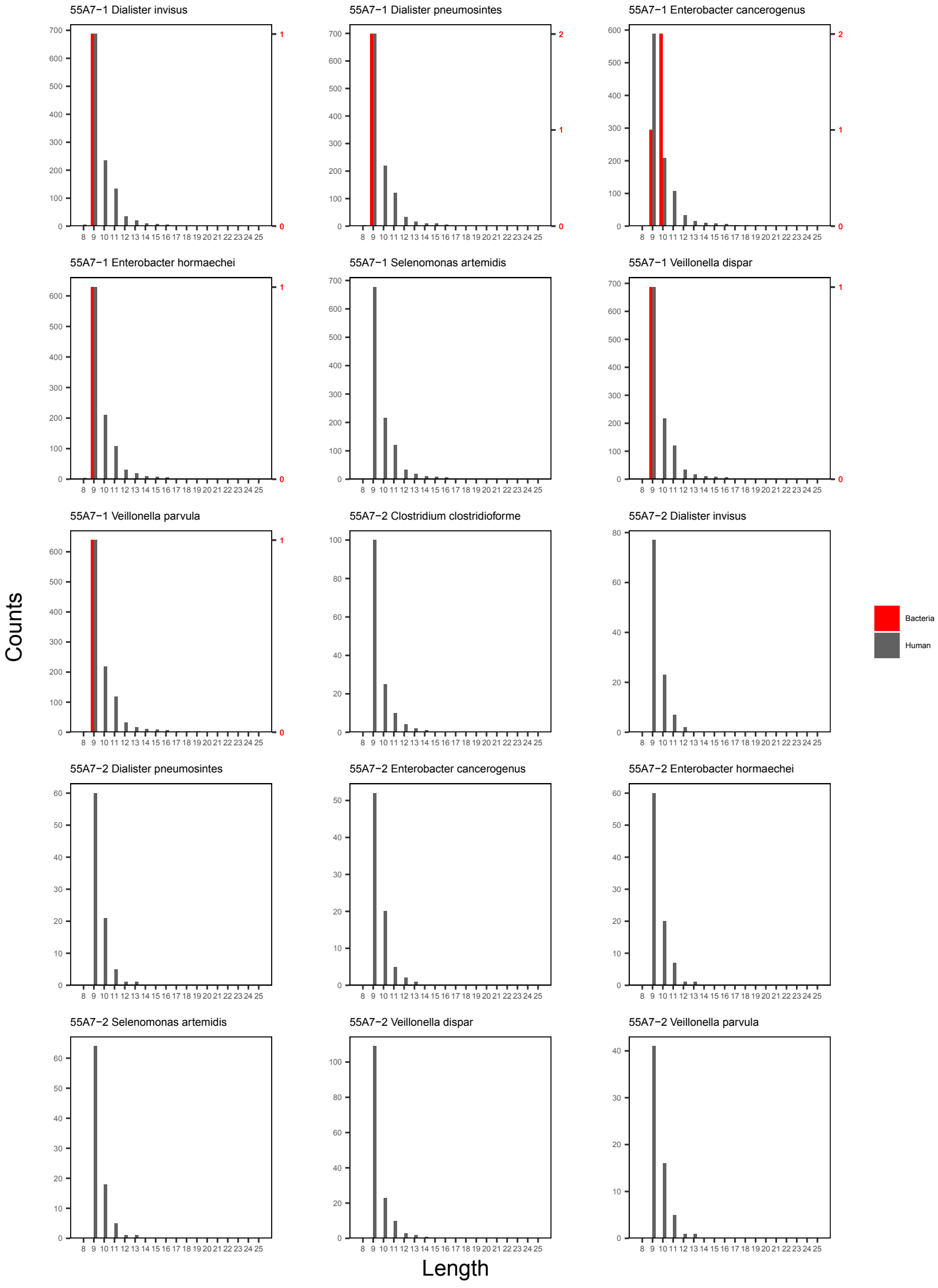
* This cluster contains also peptides with the B*08:01 motif

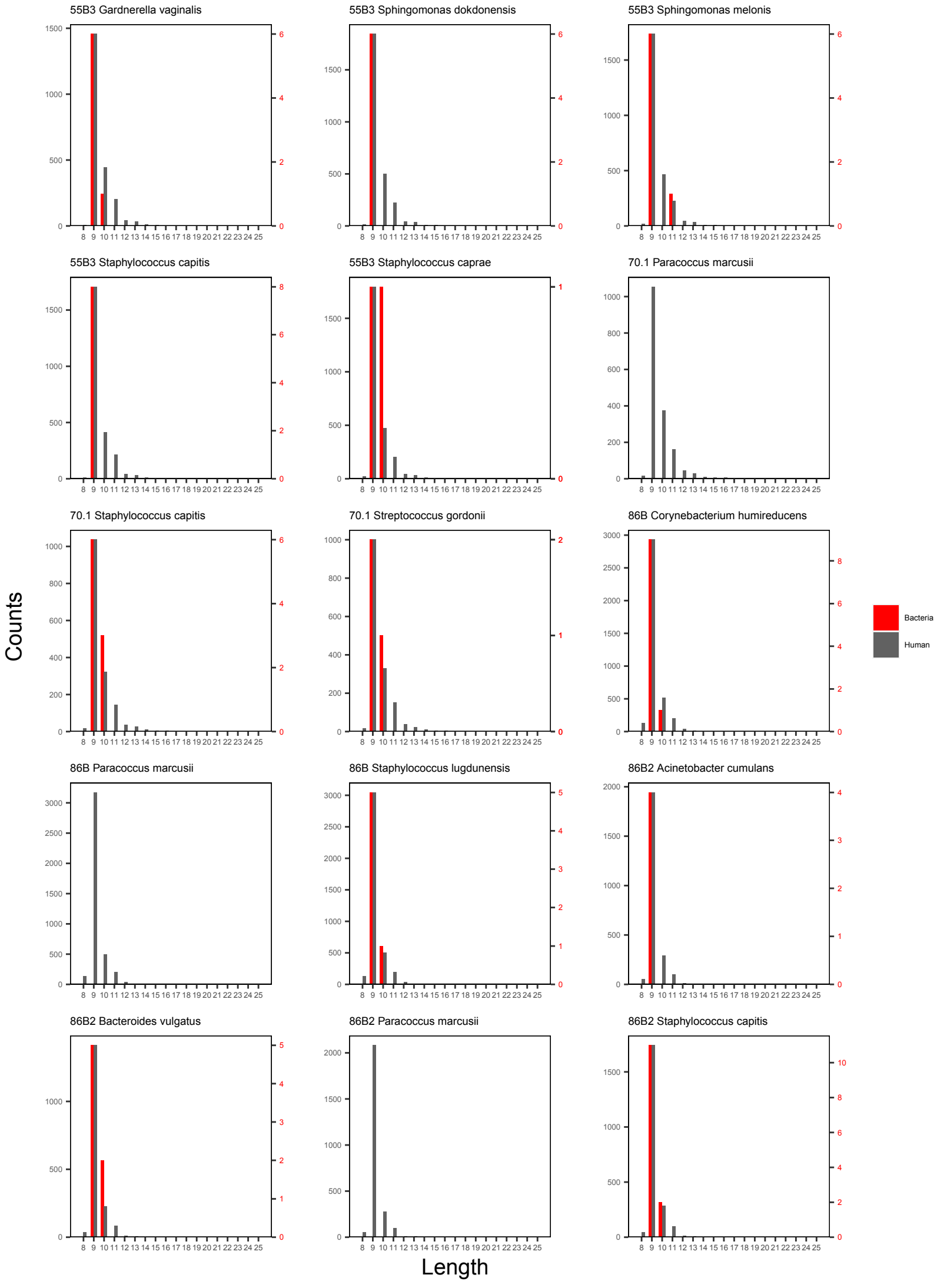
Supplementary Figure 1: Gibbs clustering of identified bacteria-derived HLA-I peptides. Bacteria- and human-derived HLA-I peptides were Gibbs clustered to assign the peptides to the different HLA alleles of the patient. Logos were created to identify the HLA allele's motif using all the peptides that match this allele in the IEDB database. N - number of all peptides, B - number of bacterial peptides, H - number of human peptides in each cluster. The trash represents the peptides that were not assigned to any cluster. Clusters with special characteristics are marked with an asterisk.

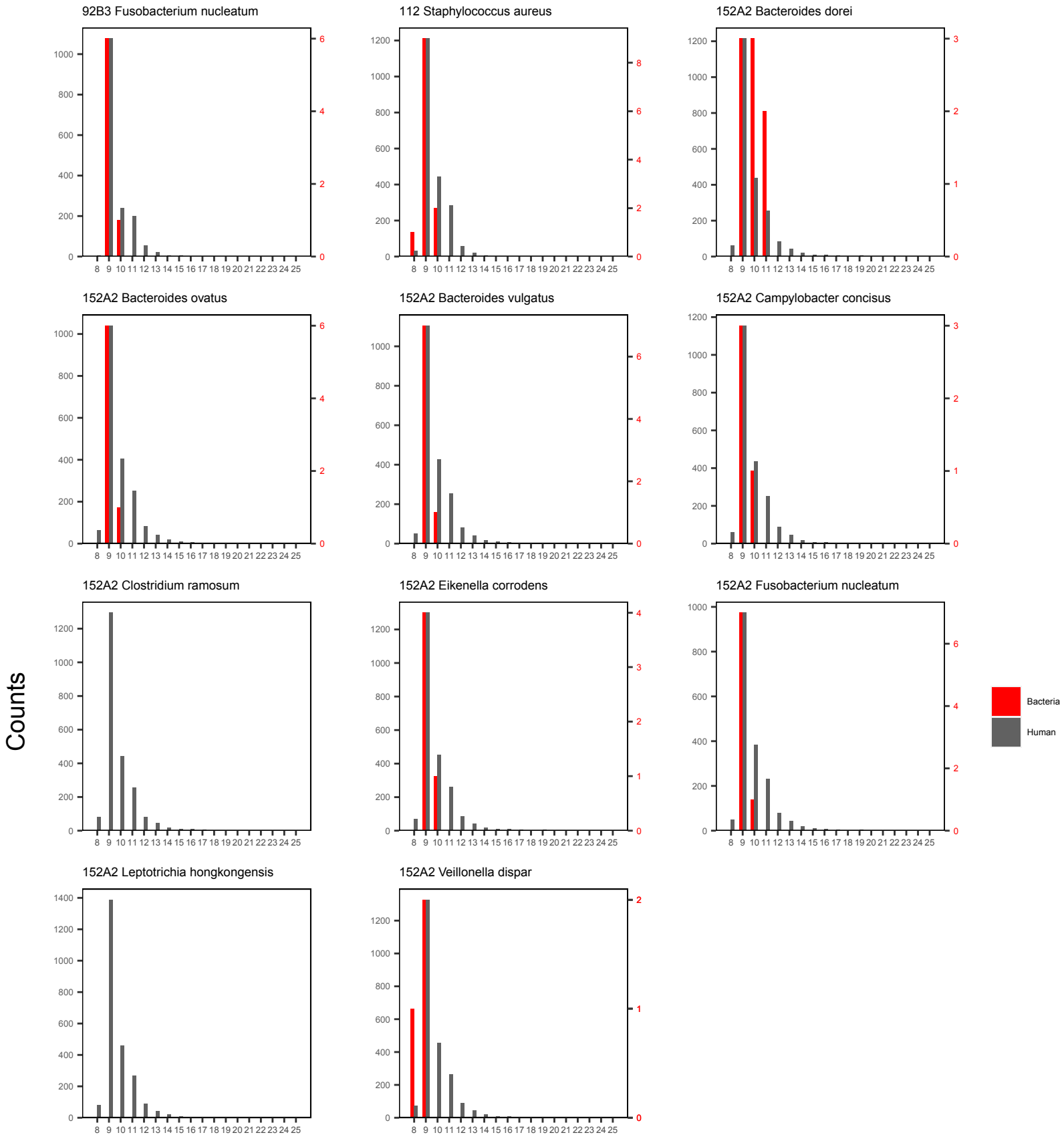
Supplementary Figure 2
Length distribution HLA-I peptides

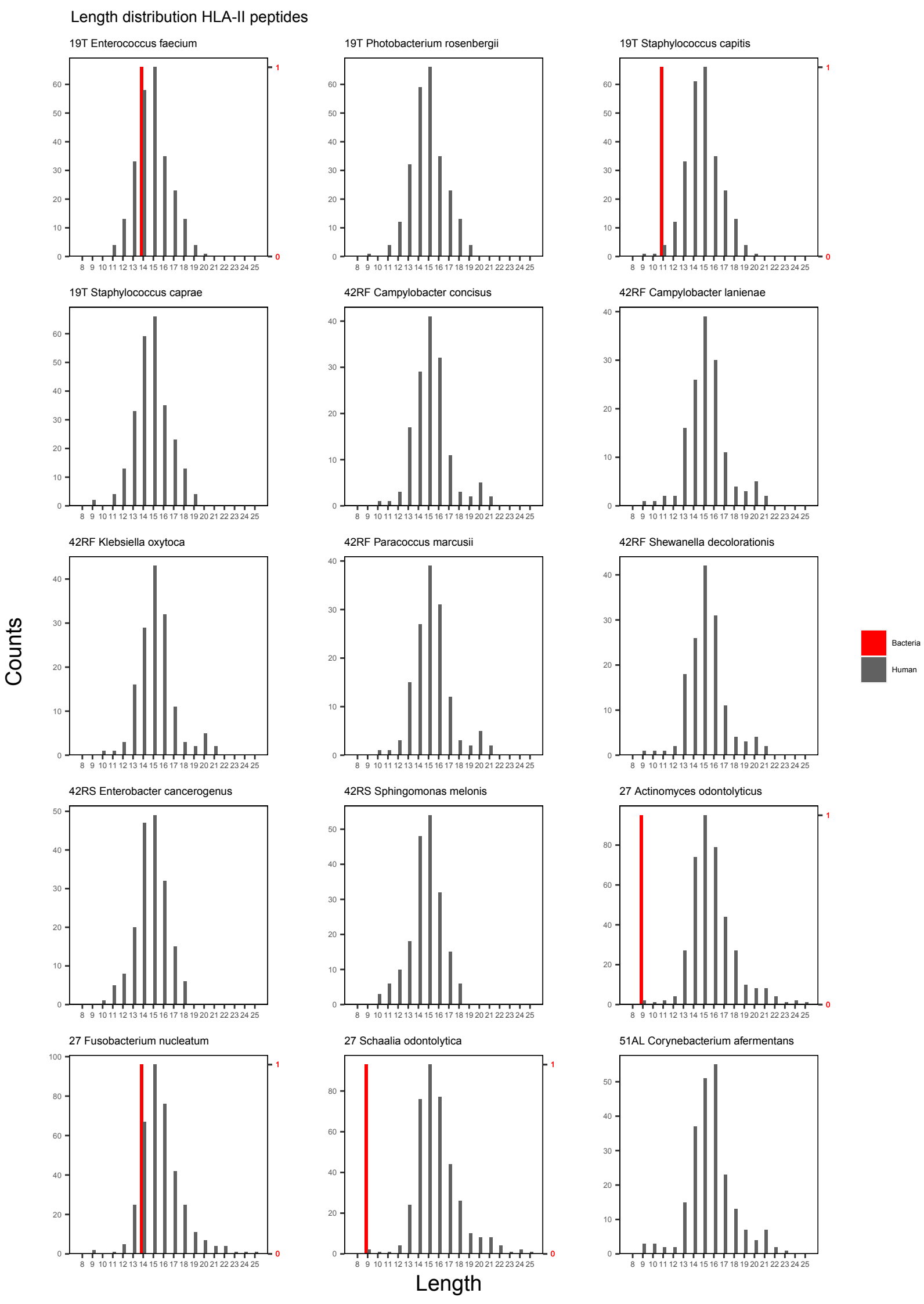


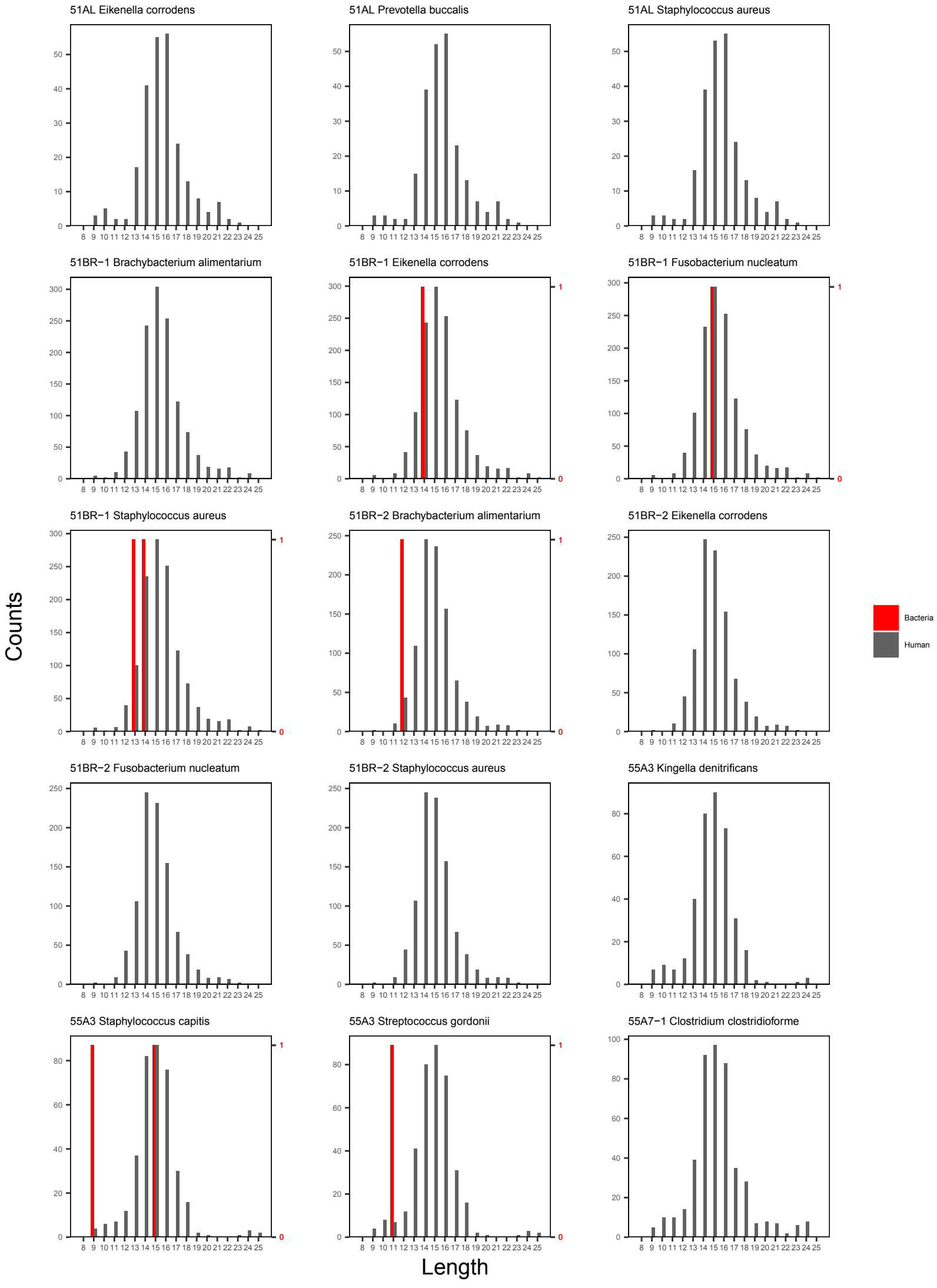

51AL Eikenella corrodens
51AL Prevotella buccalis
51AL Staphylococcus aureus
51BR−1 Brachybacterium alimentarium
51BR−1 Eikenella corrodens
51BR−1 Fusobacterium nucleatum
51BR−1 Staphylococcus aureus
51BR−2 Brachybacterium alimentarium
51BR−2 Eikenella corrodens
51BR−2 Fusobacterium nucleatum
51BR−2 Staphylococcus aureus
55A3 Kingella denitrificans
55A3 Staphylococcus capitis
55A3 Streptococcus gordonii
55A7−1 Clostridium clostridioforme
Counts
Length
Bacteria
Human

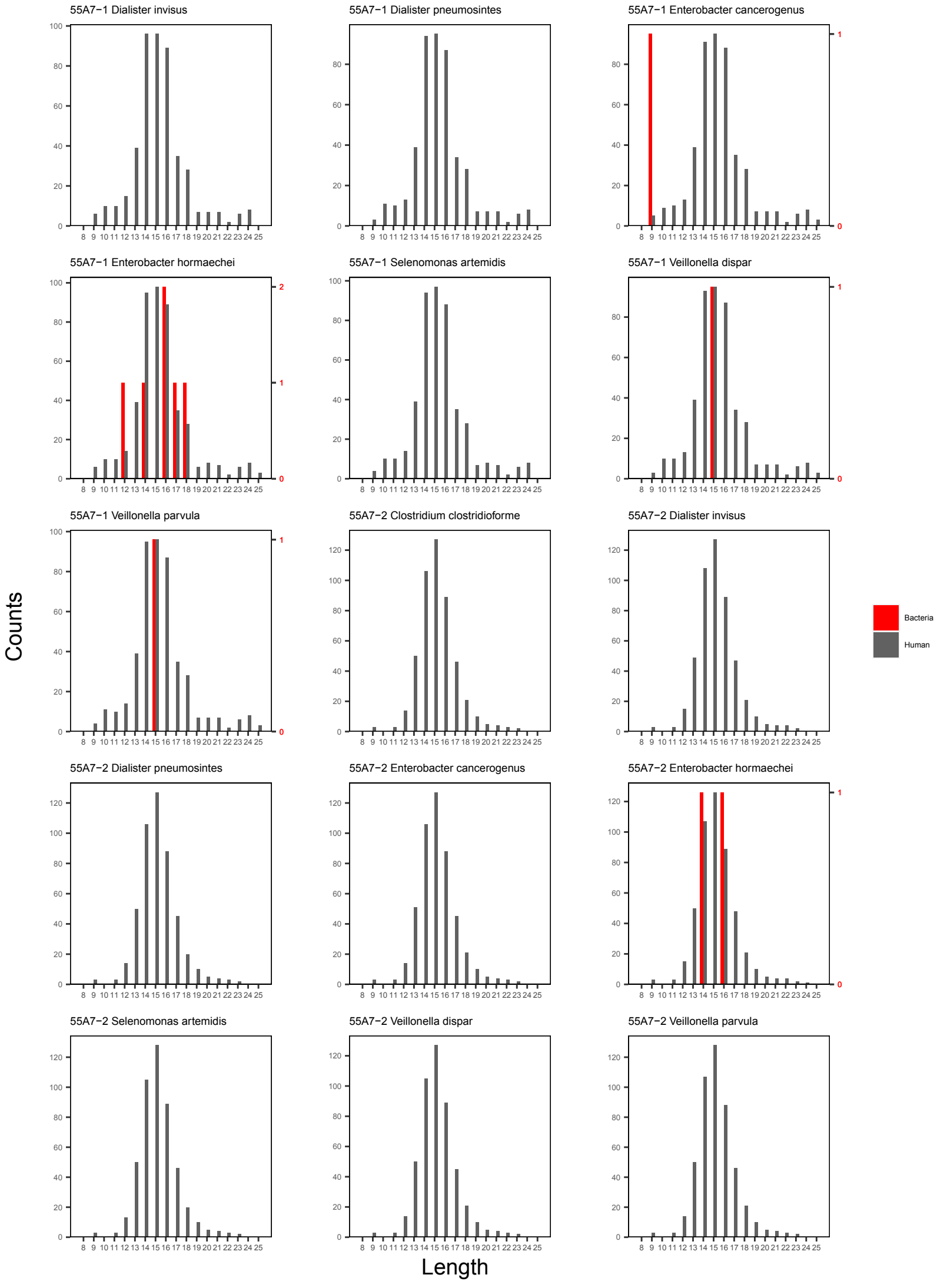

55A7−1 Dialister invisus
55A7−1 Dialister pneumosintes
55A7−1 Enterobacter cancerogenus
55A7−1 Enterobacter hormaechei
55A7−1 Selenomonas artemidis
55A7−1 Veillonella dispar
55A7−1 Veillonella parvula
55A7−2 Clostridium clostridioforme
55A7−2 Dialister invisus
55A7−2 Dialister pneumosintes
55A7−2 Enterobacter cancerogenus
55A7−2 Enterobacter hormaechei
55A7−2 Selenomonas artemidis
55A7−2 Veillonella dispar
55A7−2 Veillonella parvula
Counts
Length
Bacteria
Human

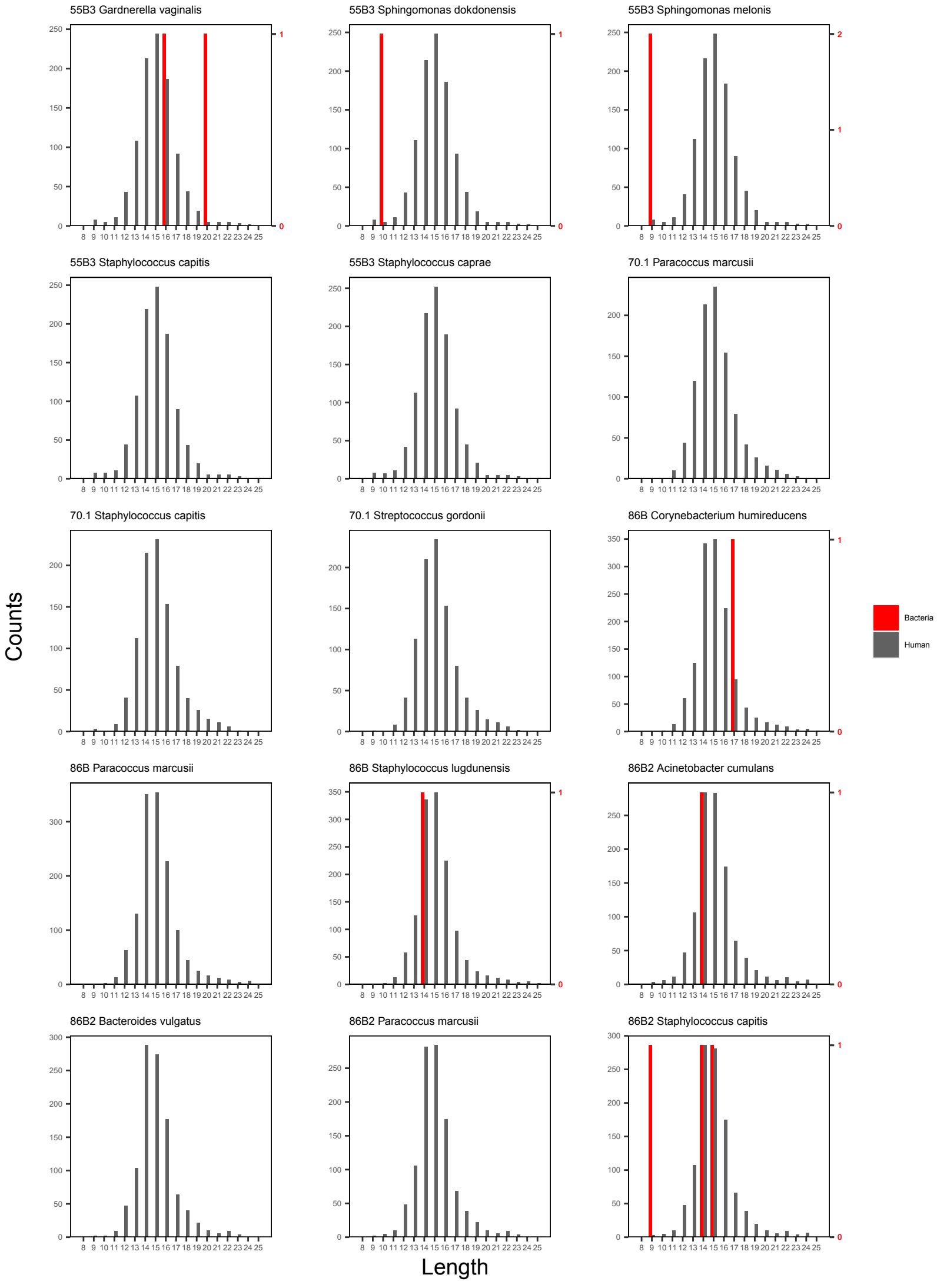

55B3 Gardnerella vaginalis
55B3 Sphingomonas dokdonensis
55B3 Sphingomonas melonis
55B3 Staphylococcus capitis
55B3 Staphylococcus caprae
70.1 Paracoccus marcusii
70.1 Staphylococcus capitis
70.1 Streptococcus gordonii
86B Corynebacterium humireducens
86B Paracoccus marcusii
86B Staphylococcus lugdunensis
86B2 Acinetobacter cumulans
86B2 Bacteroides vulgatus
86B2 Paracoccus marcusii
86B2 Staphylococcus capitis
Counts
Length
Bacteria
Human

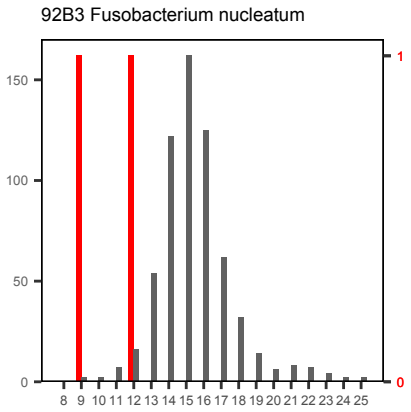

92B3 Fusobacterium nucleatum
112 Staphylococcus aureus
152A2 Bacteroides dorei
152A2 Bacteroides ovatus
152A2 Bacteroides vulgatus
152A2 Campylobacter concisus
152A2 Clostridium ramosum
152A2 Eikenella corrodens
152A2 Fusobacterium nucleatum
152A2 Leptotrichia hongkongensis
152A2 Veillonella dispar
Counts
Length
Bacteria
Human


Length distribution HLA-II peptides
19T Enterococcus faecium
19T Photobacterium rosenbergii
19T Staphylococcus capitis
19T Staphylococcus caprae
42RF Campylobacter concisus
42RF Campylobacter lanienae
42RF Klebsiella oxytoca
42RF Paracoccus marcusii
42RF Shewanella decolorationis
42RS Enterobacter cancerogenus
42RS Sphingomonas melonis
27 Actinomyces odontolyticus
27 Fusobacterium nucleatum
27 Schaalia odontolytica
51AL Corynebacterium afermentans
Counts
Length
Bacteria
Human


51AL Eikenella corrodens
51AL Prevotella buccalis
51AL Staphylococcus aureus
51BR−1 Brachybacterium alimentarium
51BR−1 Eikenella corrodens
51BR−1 Fusobacterium nucleatum
51BR−1 Staphylococcus aureus
51BR−2 Brachybacterium alimentarium
51BR−2 Eikenella corrodens
51BR−2 Fusobacterium nucleatum
51BR−2 Staphylococcus aureus
55A3 Kingella denitrificans
55A3 Staphylococcus capitis
55A3 Streptococcus gordonii
55A7−1 Clostridium clostridioforme
Counts
Length
Bacteria
Human


55A7−1 Dialister invisus
55A7−1 Dialister pneumosintes
55A7−1 Enterobacter cancerogenus
55A7−1 Enterobacter hormaechei
55A7−1 Selenomonas artemidis
55A7−1 Veillonella dispar
55A7−1 Veillonella parvula
55A7−2 Clostridium clostridioforme
55A7−2 Dialister invisus
55A7−2 Dialister pneumosintes
55A7−2 Enterobacter cancerogenus
55A7−2 Enterobacter hormaechei
55A7−2 Selenomonas artemidis
55A7−2 Veillonella dispar
55A7−2 Veillonella parvula
Counts
Length
Bacteria
Human


55B3 Gardnerella vaginalis
55B3 Sphingomonas dokdonensis
55B3 Sphingomonas melonis
55B3 Staphylococcus capitis
55B3 Staphylococcus caprae
70.1 Paracoccus marcusii
70.1 Staphylococcus capitis
70.1 Streptococcus gordonii
86B Corynebacterium humireducens
86B Paracoccus marcusii
86B Staphylococcus lugdunensis
86B2 Acinetobacter cumulans
86B2 Bacteroides vulgatus
86B2 Paracoccus marcusii
86B2 Staphylococcus capitis
Counts
Length
Bacteria
Human



Counts

Bacteria

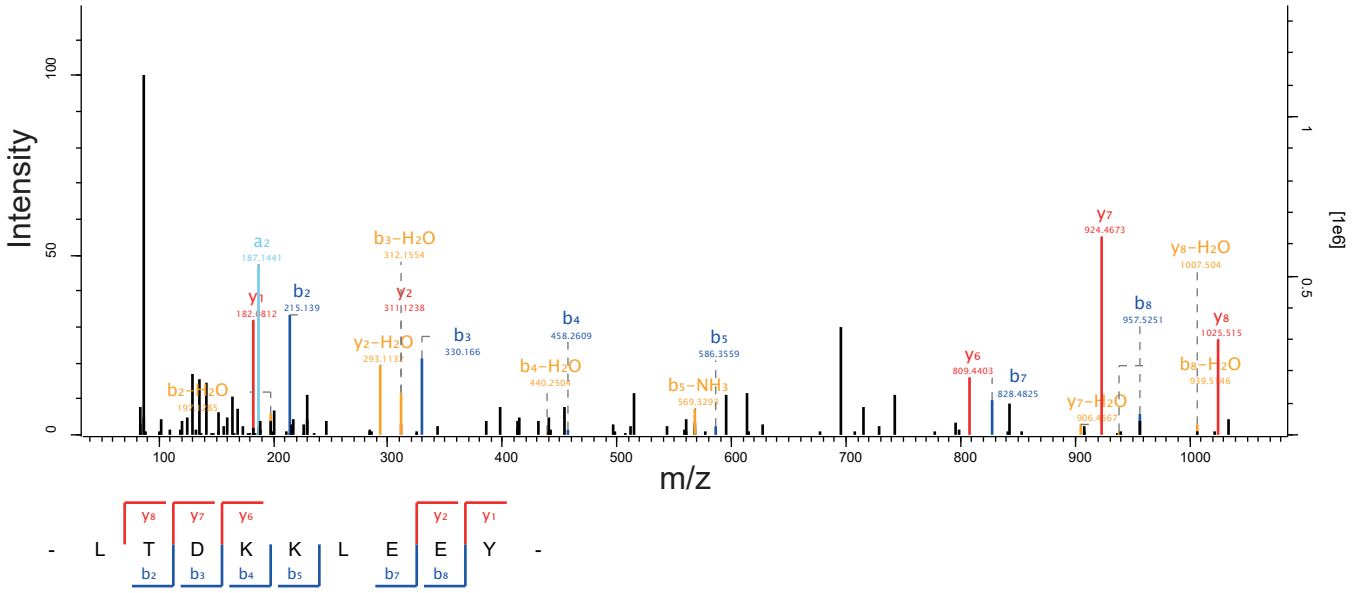
Human

Length

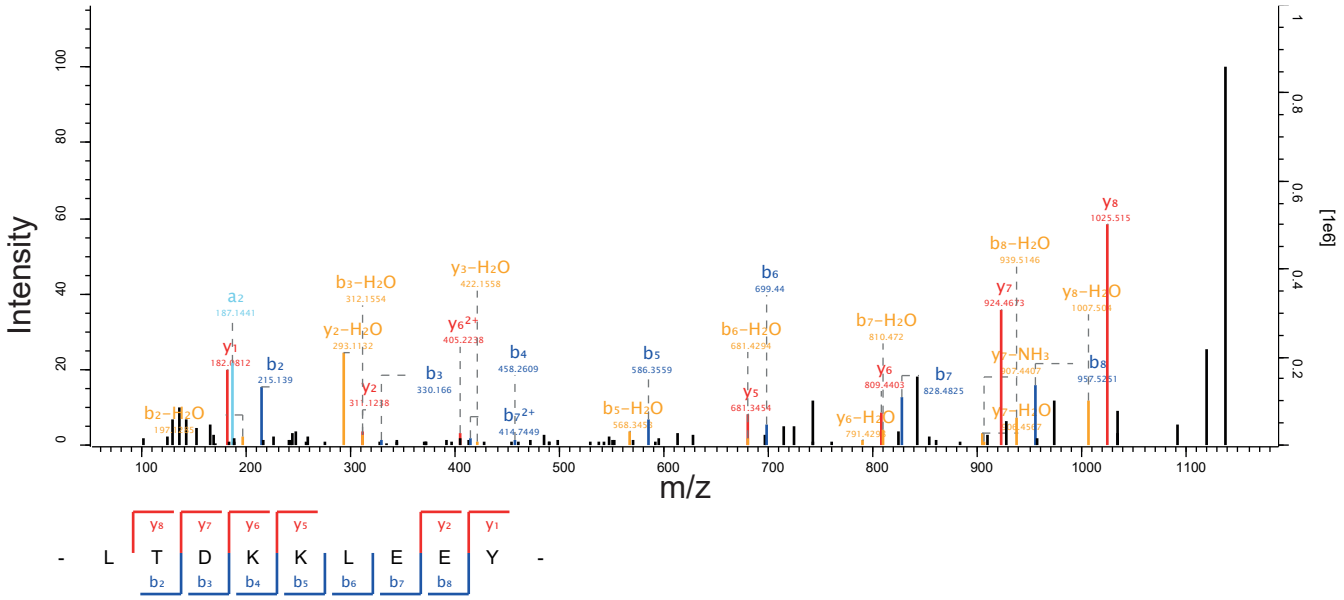
Supplementary Figure 2: The length distribution of bacteria-derived peptides is similar to the expected length of HLA-I and HLA-II peptides. Presented is the length of bacterial (red) and human (gray) HLA-I and HLA-II peptides identified in the melanoma tumors.

Supplementary Figure 3
27 Fusobacterium nucleatum HLA-I LTDKKLEEY

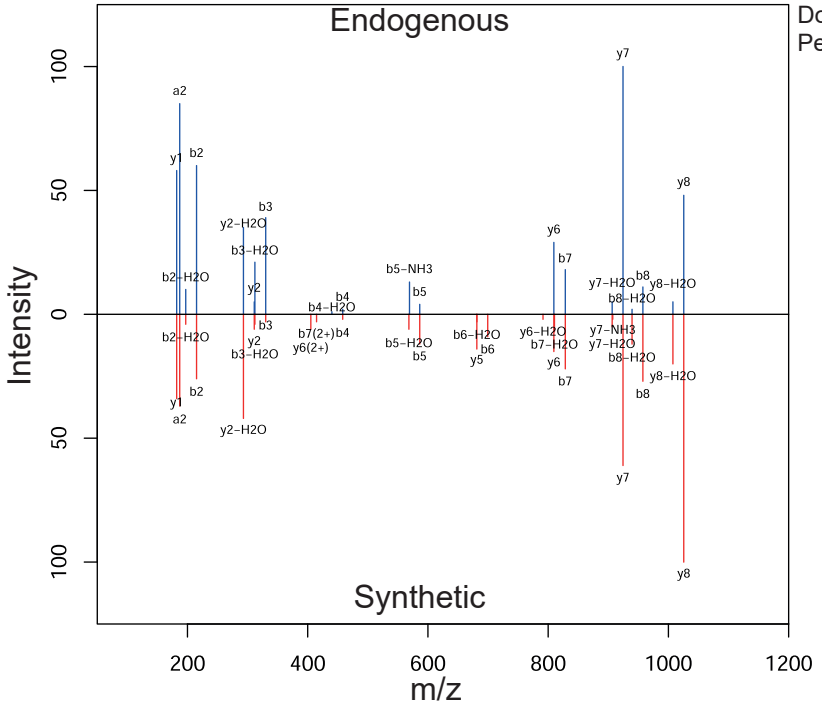
Endogenous peptide



Synthetic peptide



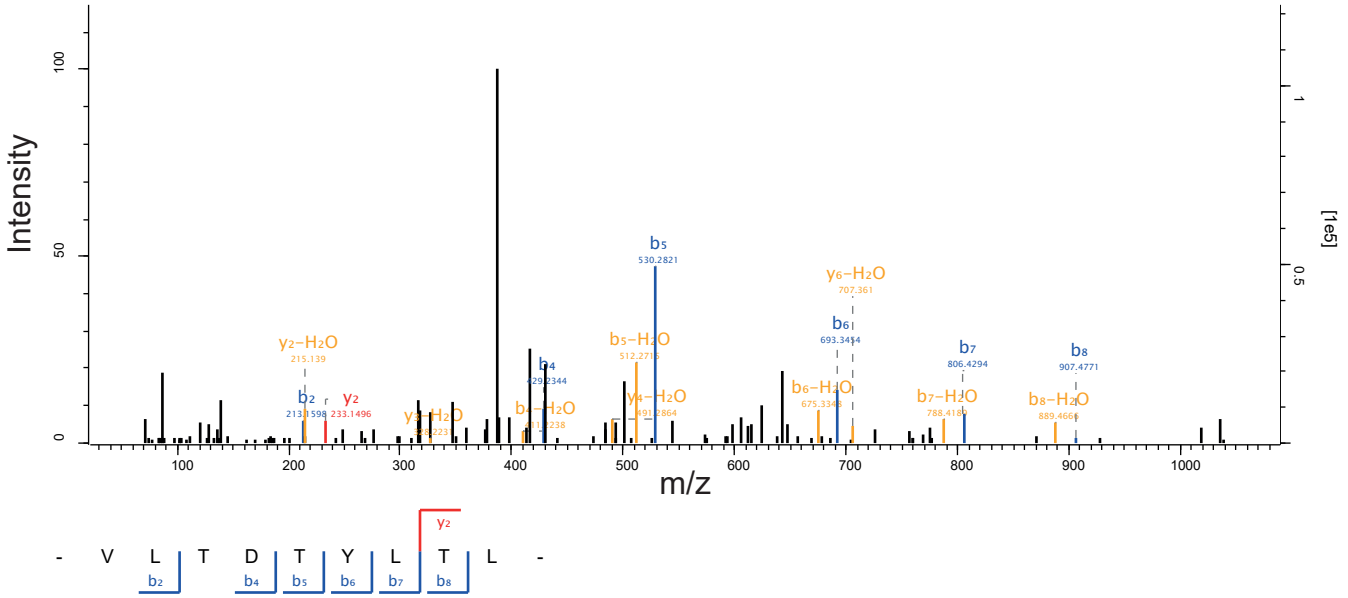
Endogenous/Synthetic peptide correlation



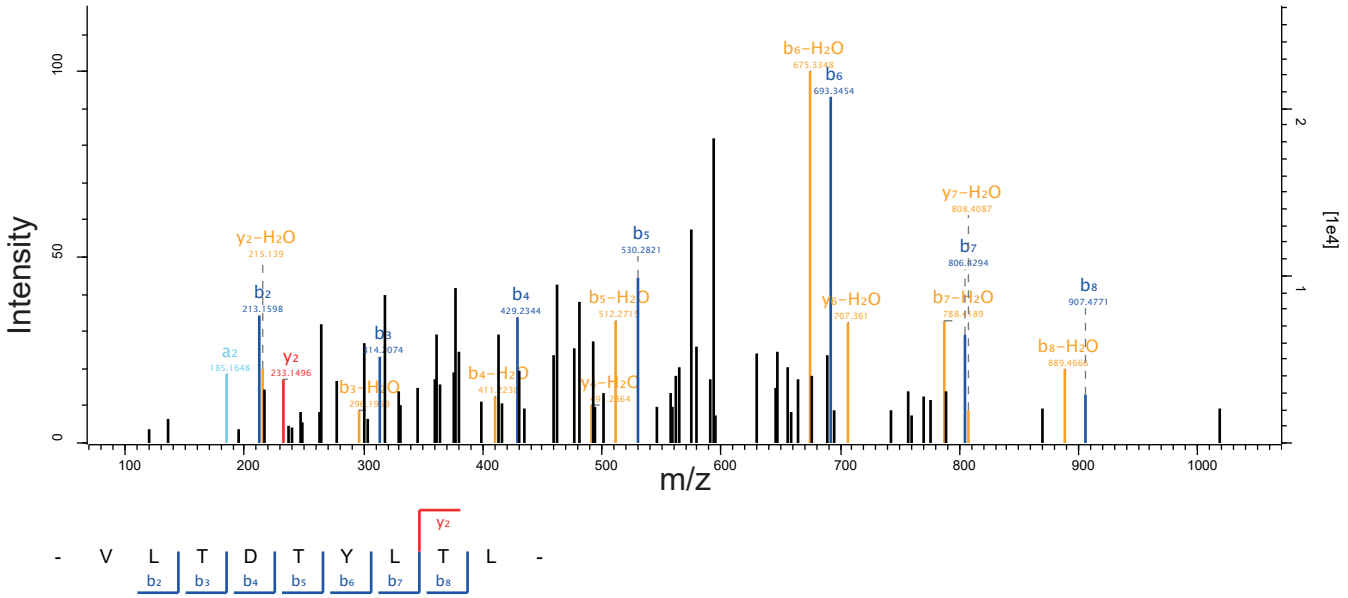
Dot product score=0.808
Pearson correlation=0.807

27 Fusobacterium nucleatum HLA-I VLTDTYLTL

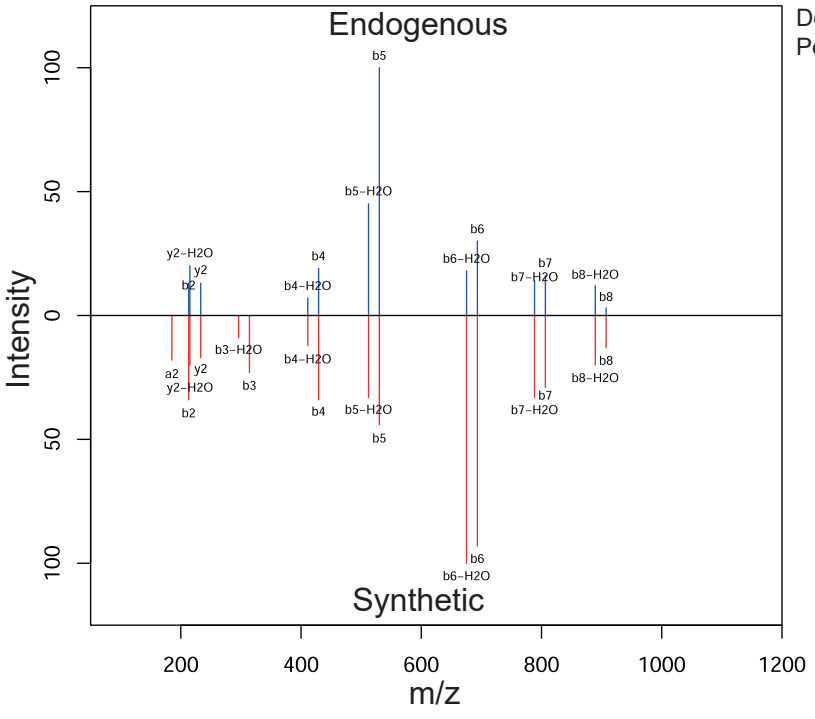
Endogenous peptide



Synthetic peptide



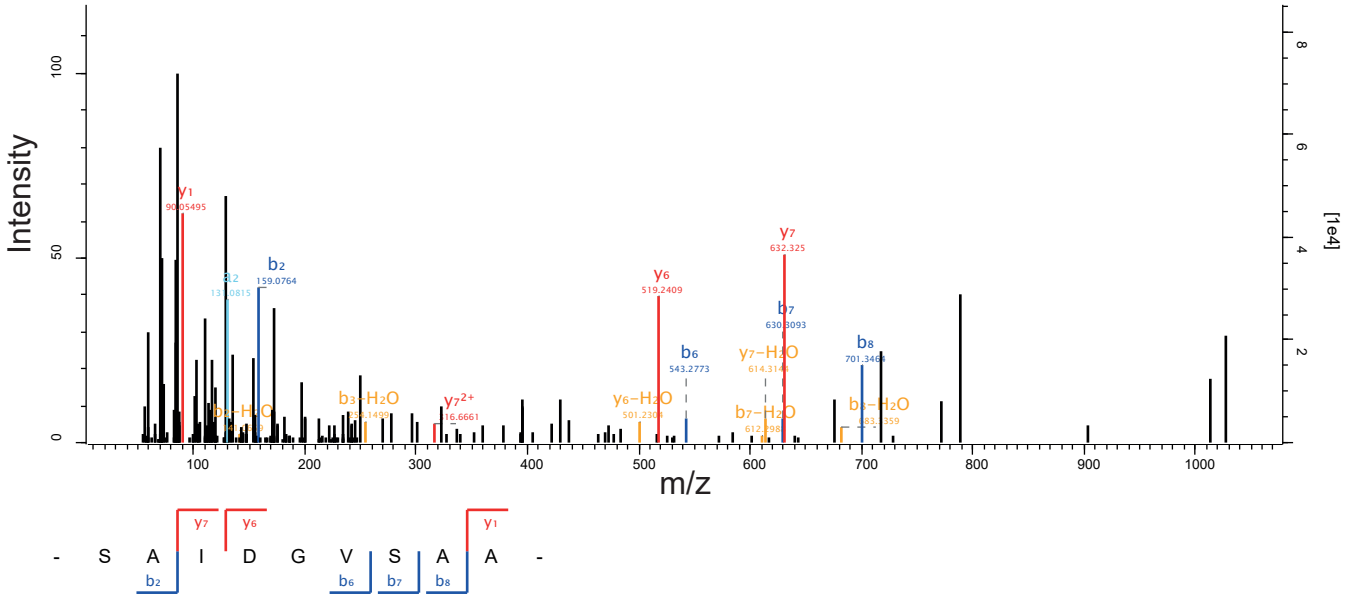
Endogenous/Synthetic peptide correlation



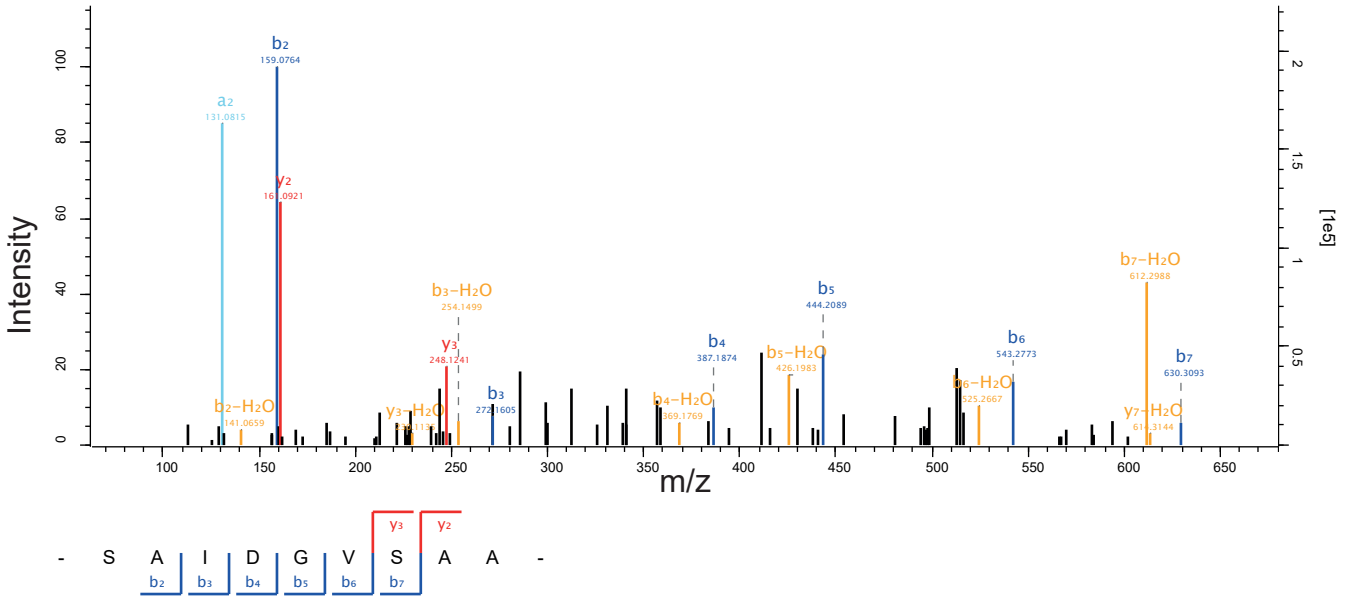
Dot product score=0.695
Pearson correlation=0.693

27 Schaalia odontolytica HLA-I SAIDGVSAA

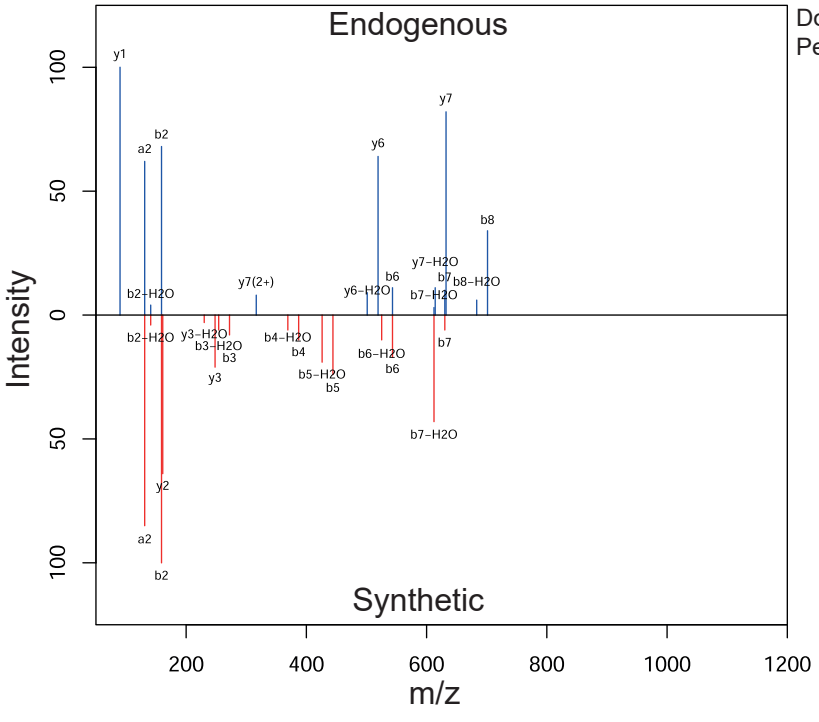
Endogenous peptide



Synthetic peptide



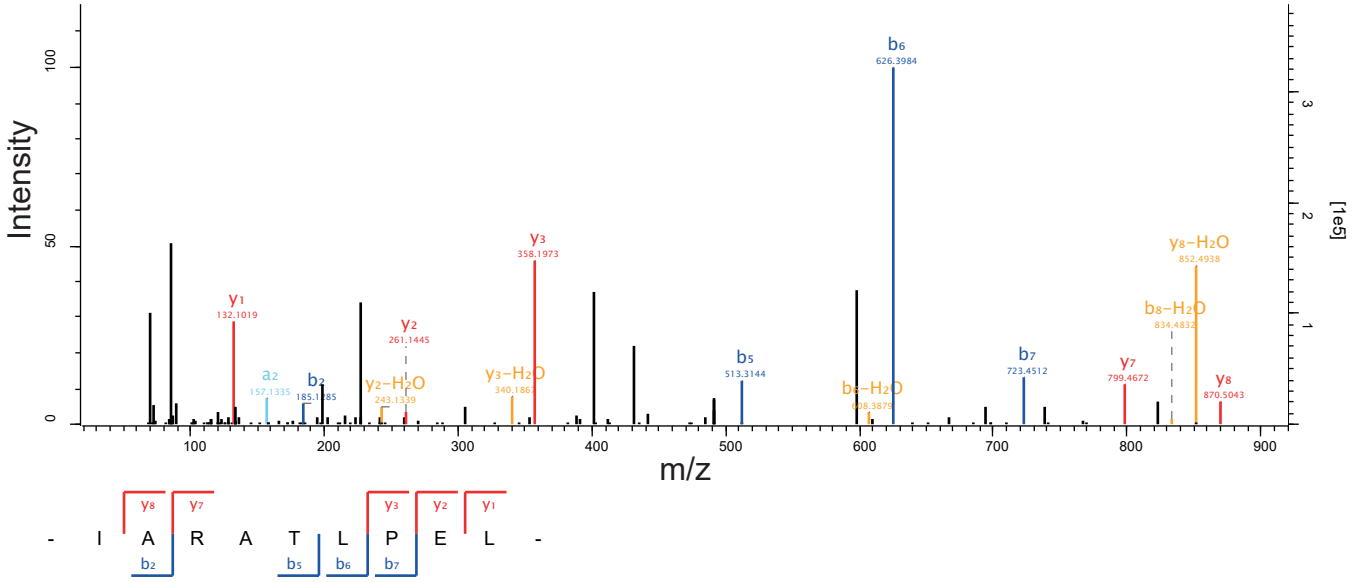
Endogenous/Synthetic peptide correlation



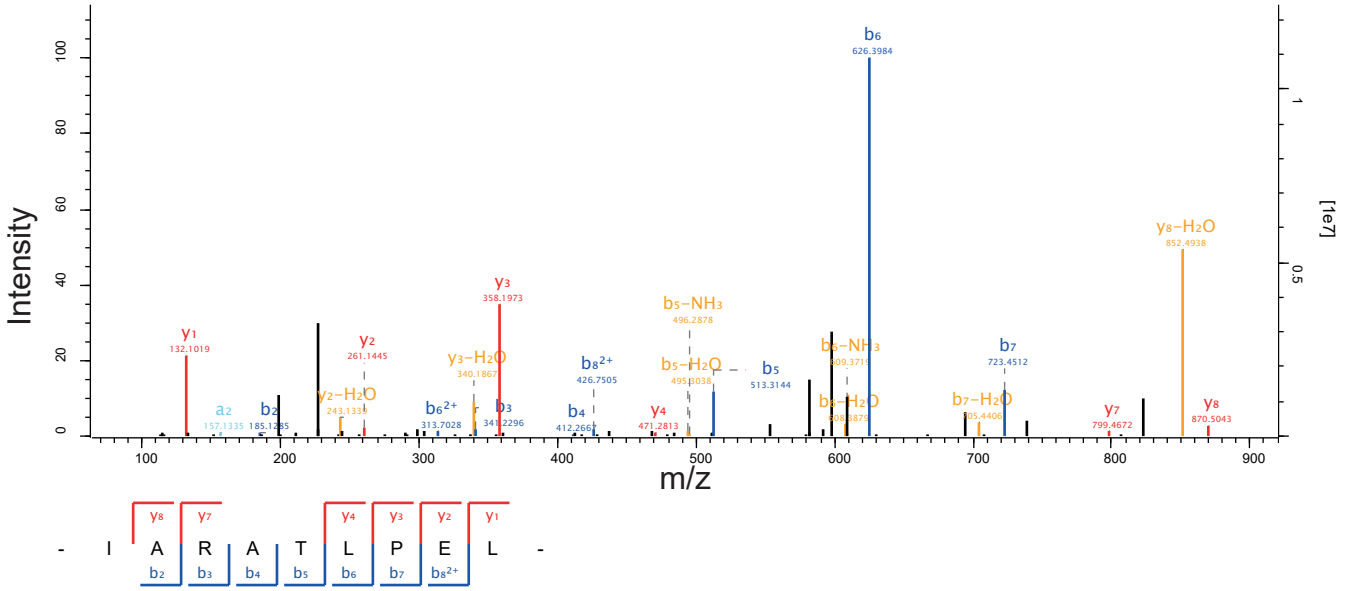
Dot product score=0.995
Pearson correlation=0.995

51AL Corynebacterium afermentans HLA-I IARATLPEL

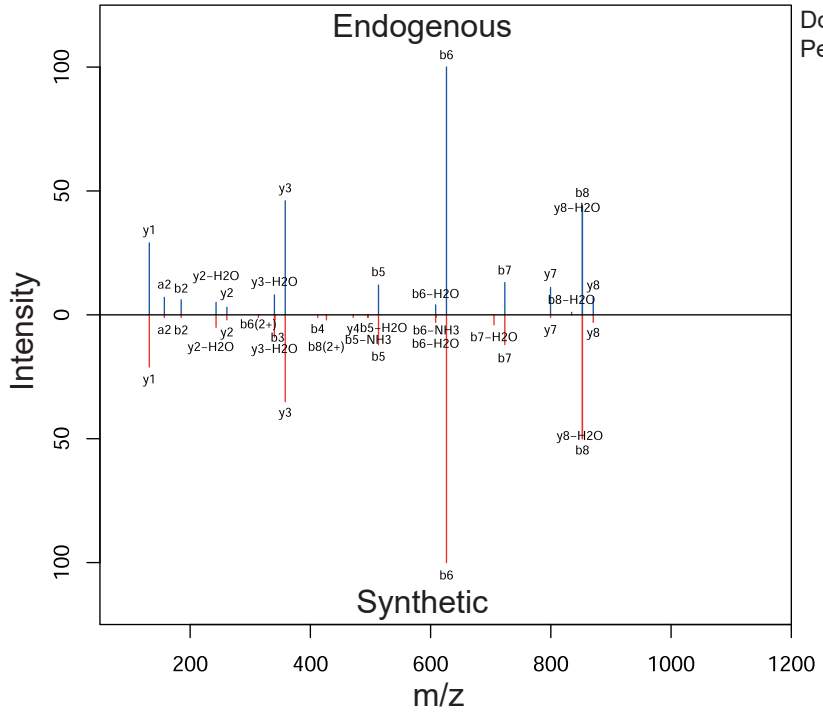
Endogenous peptide



Synthetic peptide



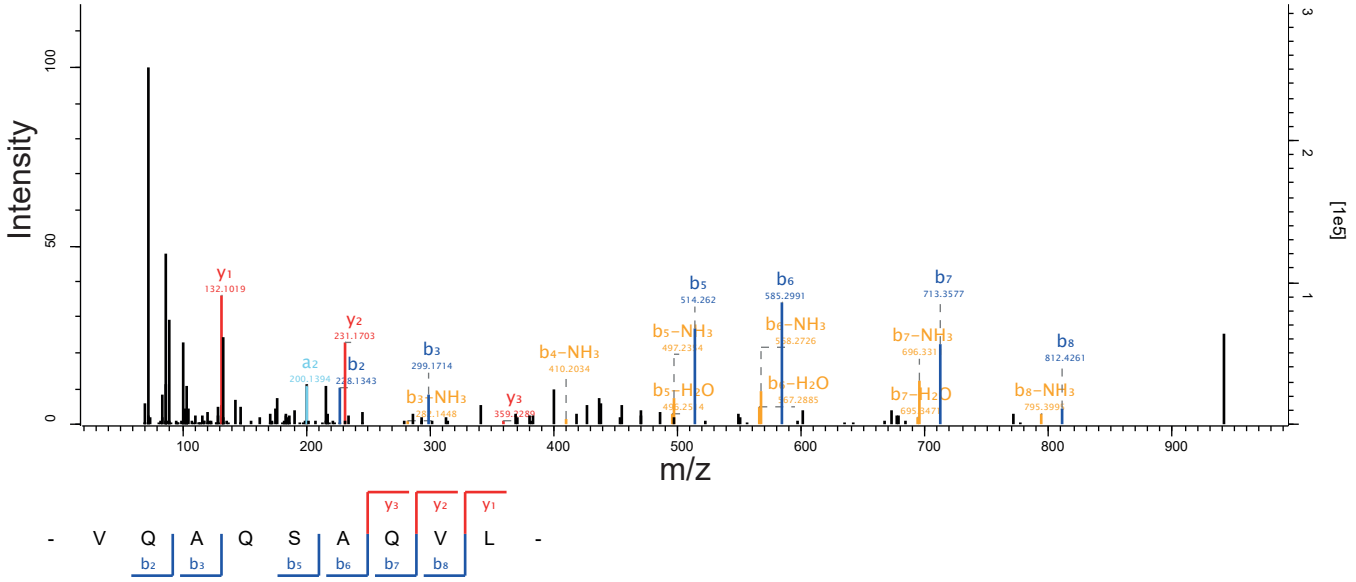
Endogenous/Synthetic peptide correlation



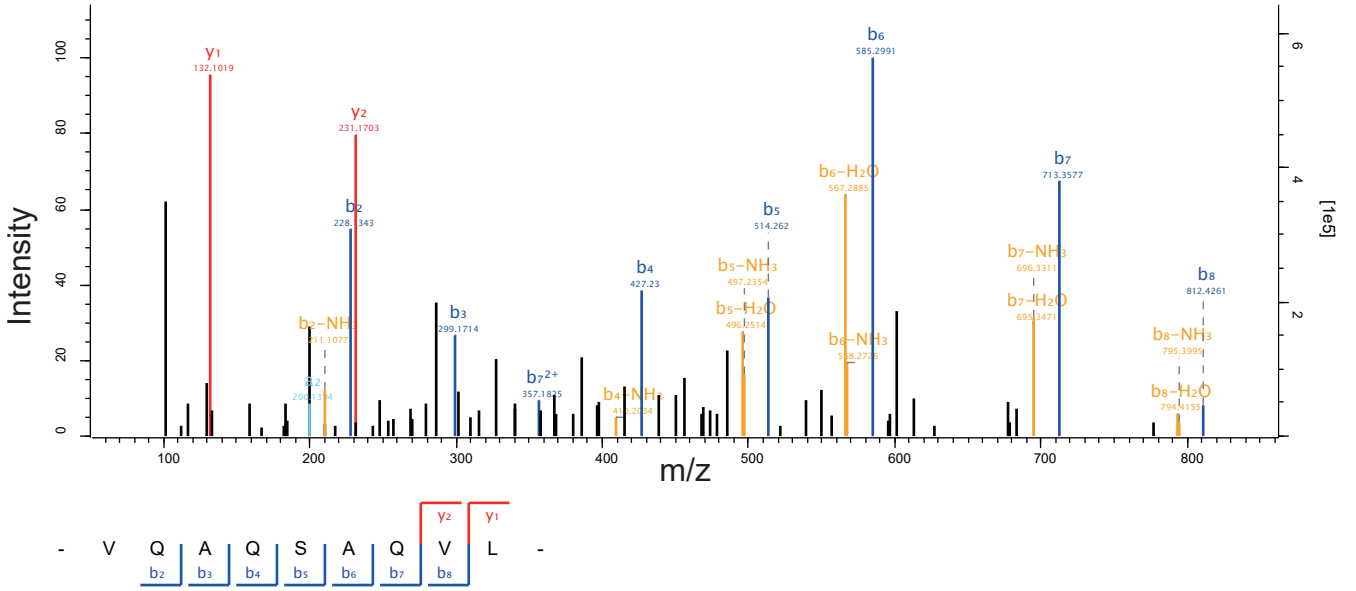
Dot product score=0.989
Pearson correlation=0.989

51AL Corynebacterium afermentans HLA-I VQAQSAQVL

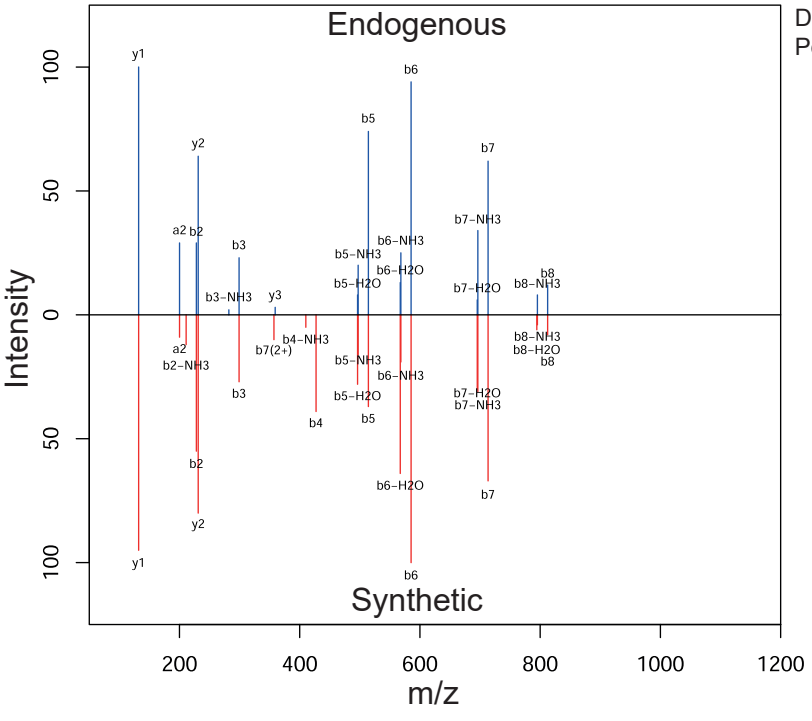
Endogenous peptide



Synthetic peptide



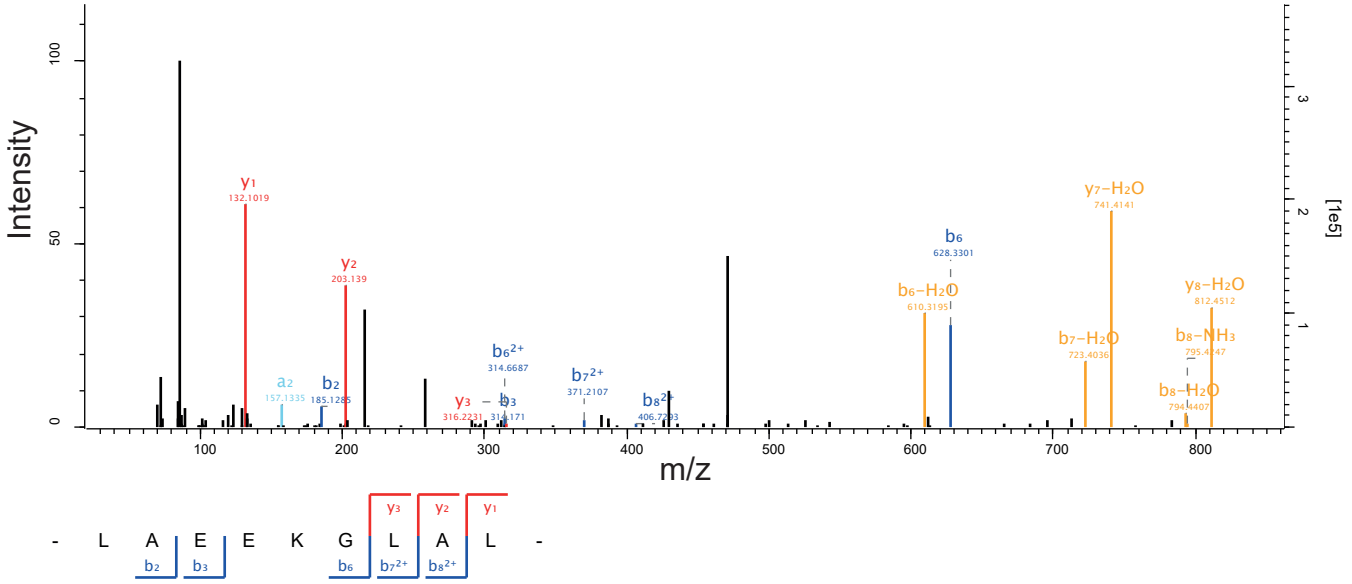
Endogenous/Synthetic peptide correlation



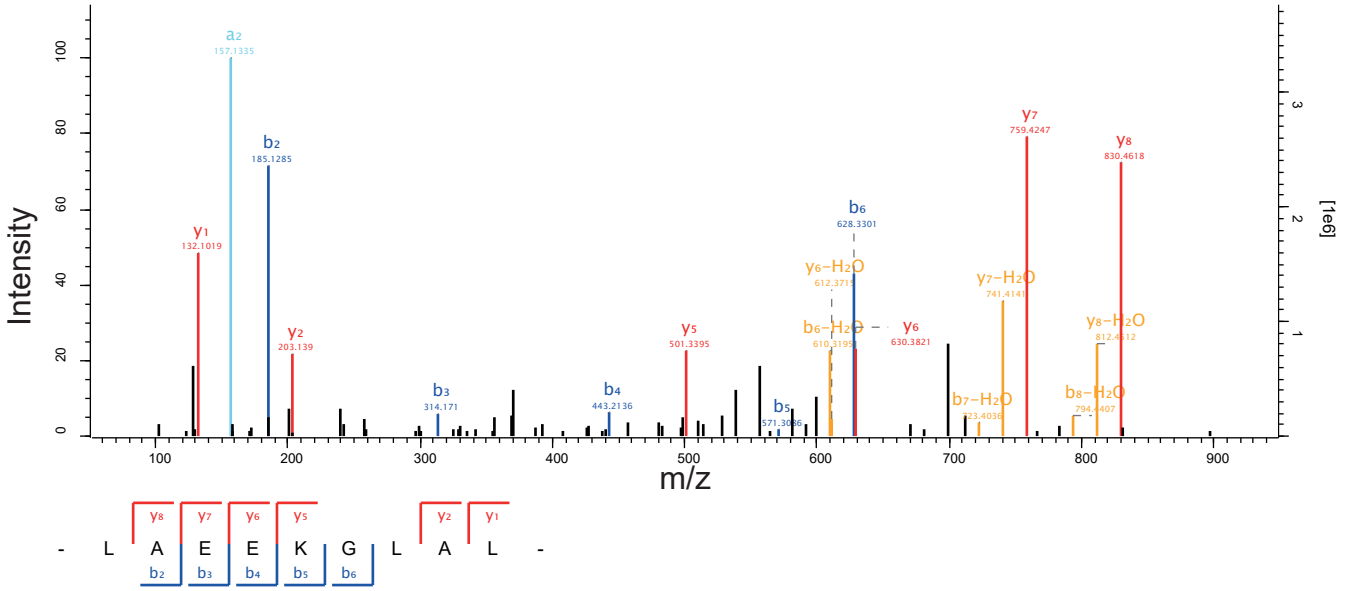
Dot product score=0.964
Pearson correlation=0.964

51AL Staphylococcus aureus HLA-I LAEEKGLAL

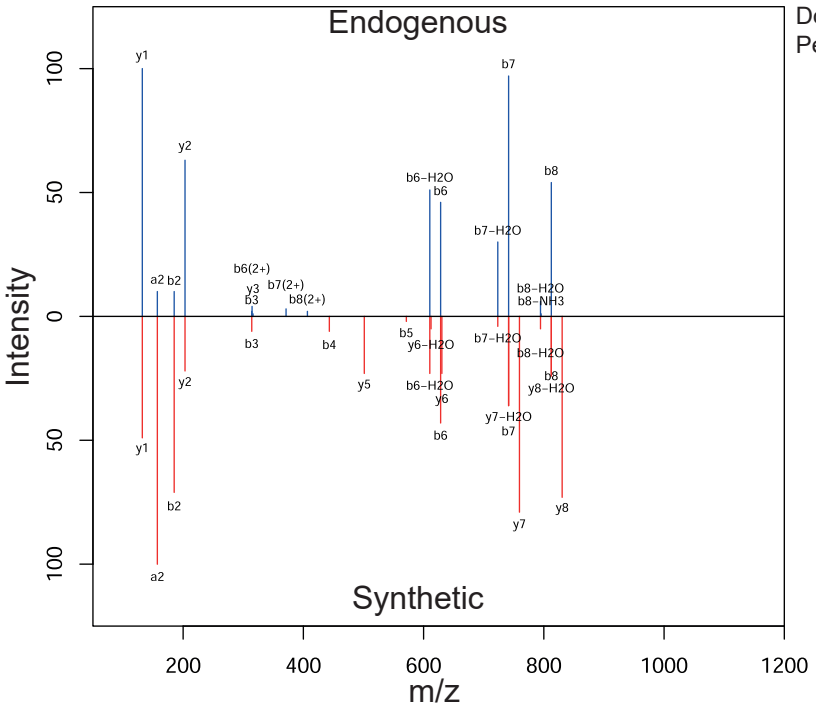
Endogenous peptide



Synthetic peptide



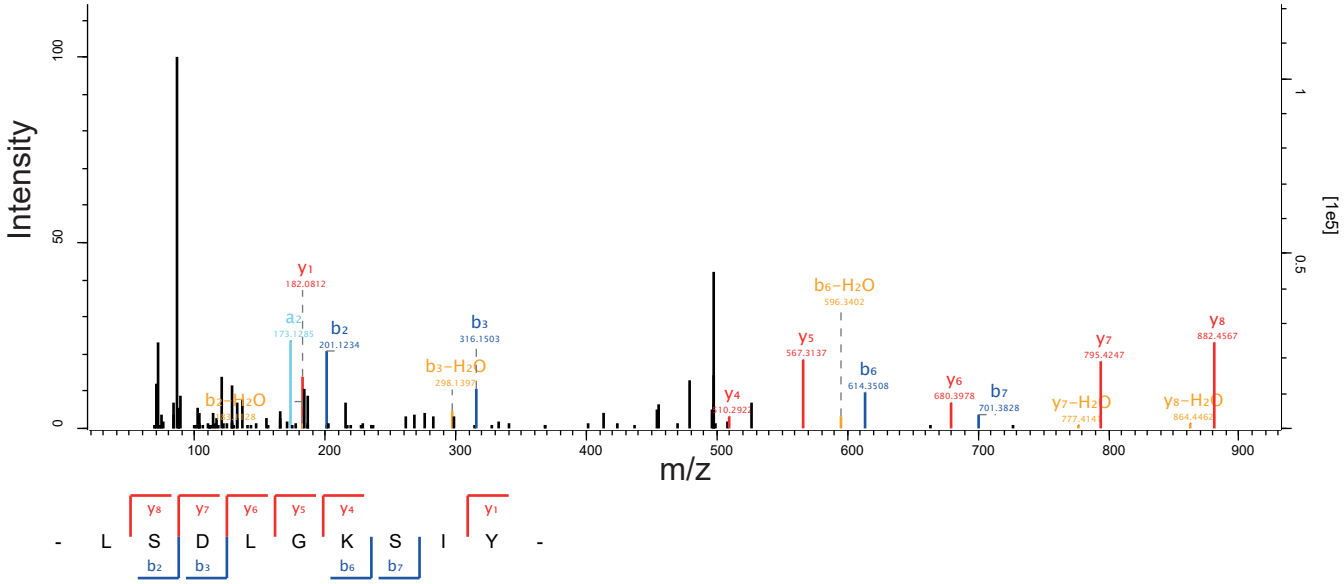
Endogenous/Synthetic peptide correlation



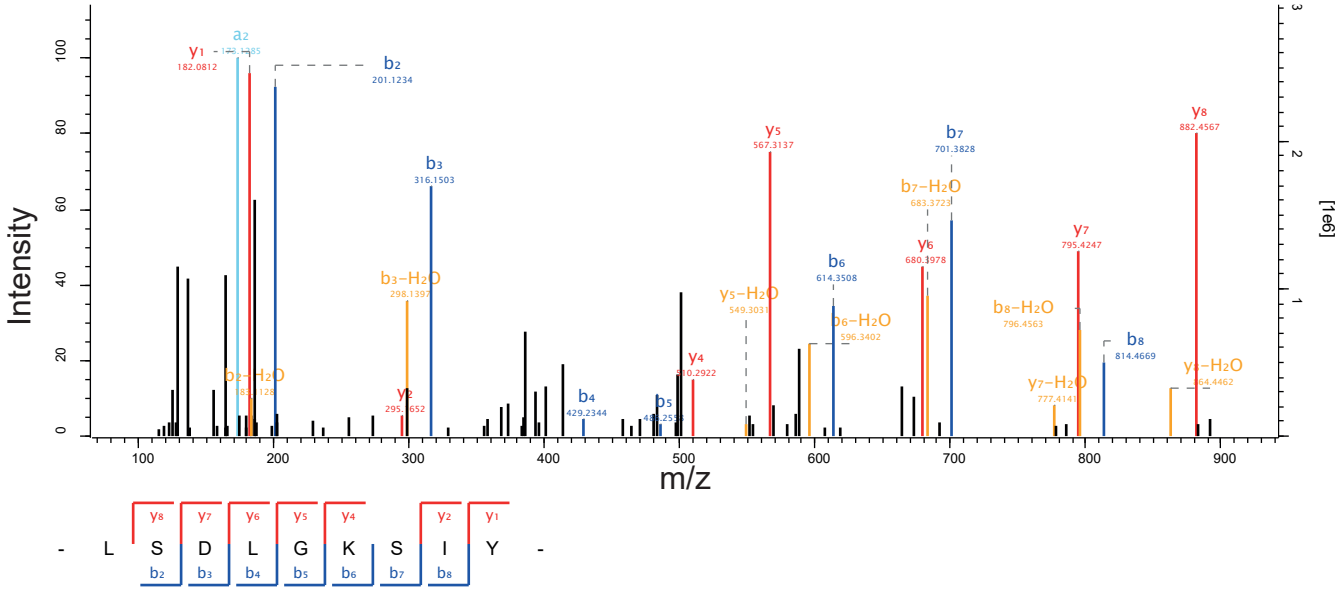
Dot product score=0.749
Pearson correlation=0.747

51AL Staphylococcus aureus HLA-I LSDLGKSIY

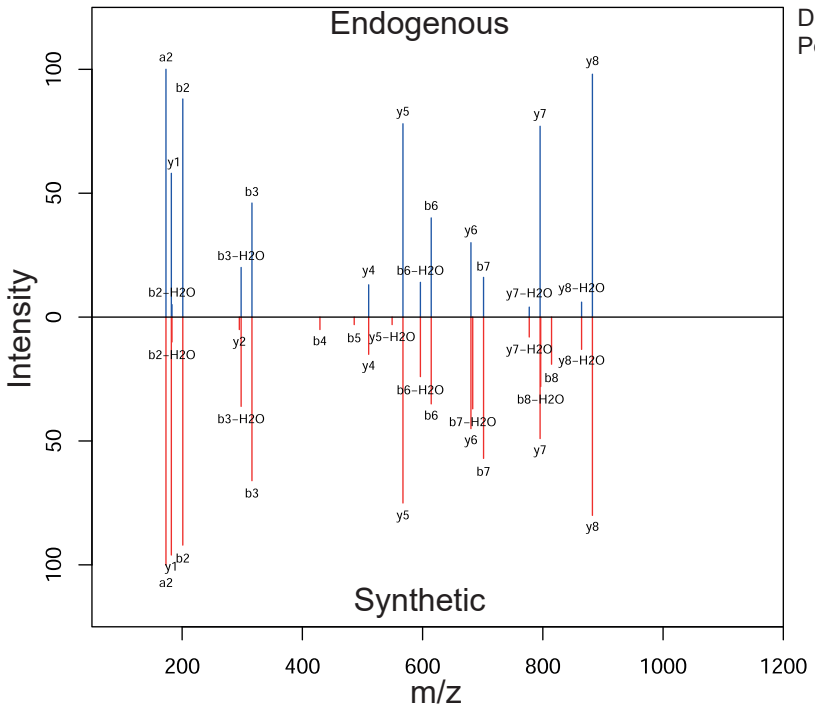
Endogenous peptide



Synthetic peptide



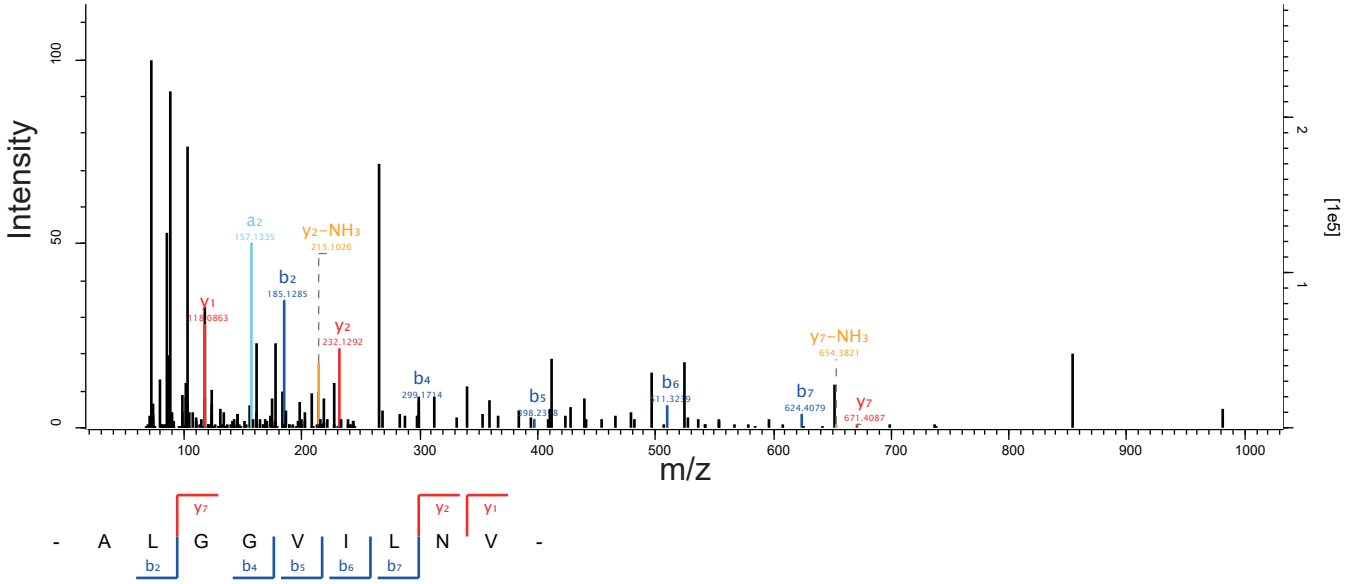
Endogenous/Synthetic peptide correlation



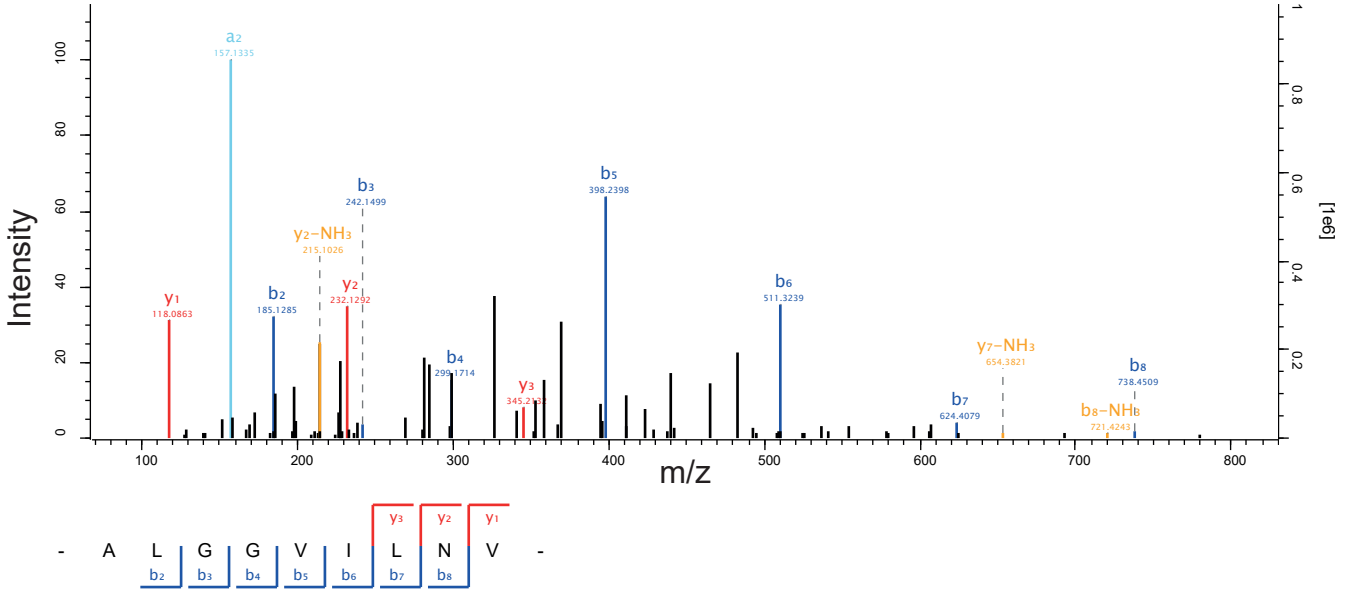
Dot product score=0.941
Pearson correlation=0.940

51BR-1 Brachybacterium alimentarium HLA-I ALGGVILNV

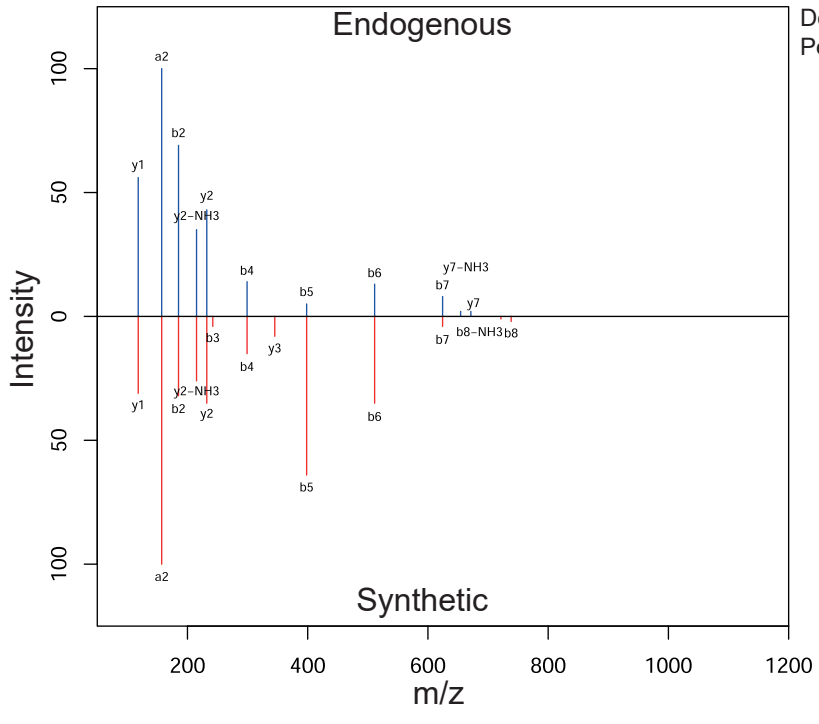
Endogenous peptide



Synthetic peptide



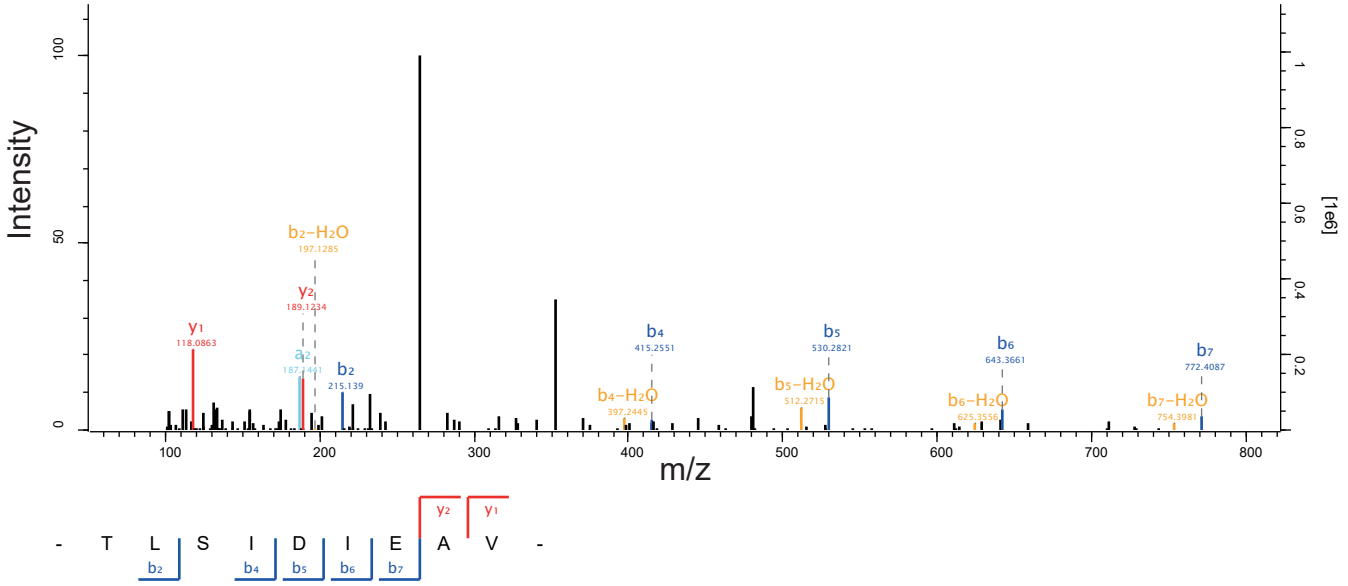
Endogenous/Synthetic peptide correlation



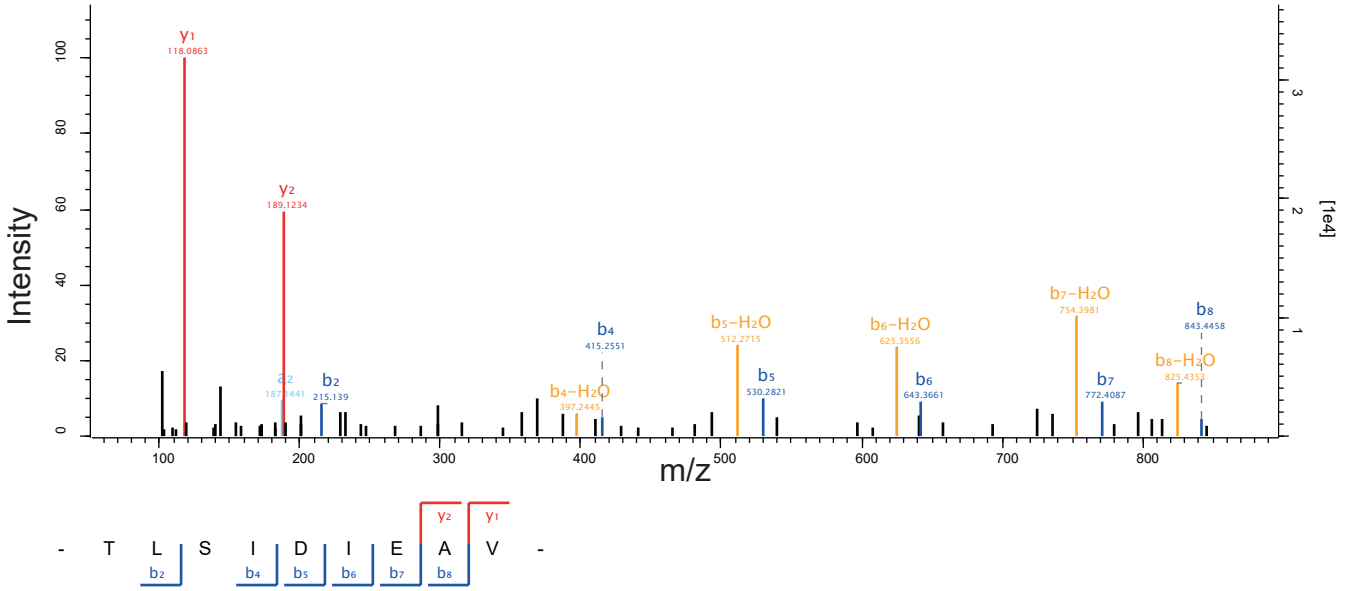
Dot product score=0.683
Pearson correlation=0.680

51BR-1 Brachybacterium alimentarium HLA-I TLSIDIEAV

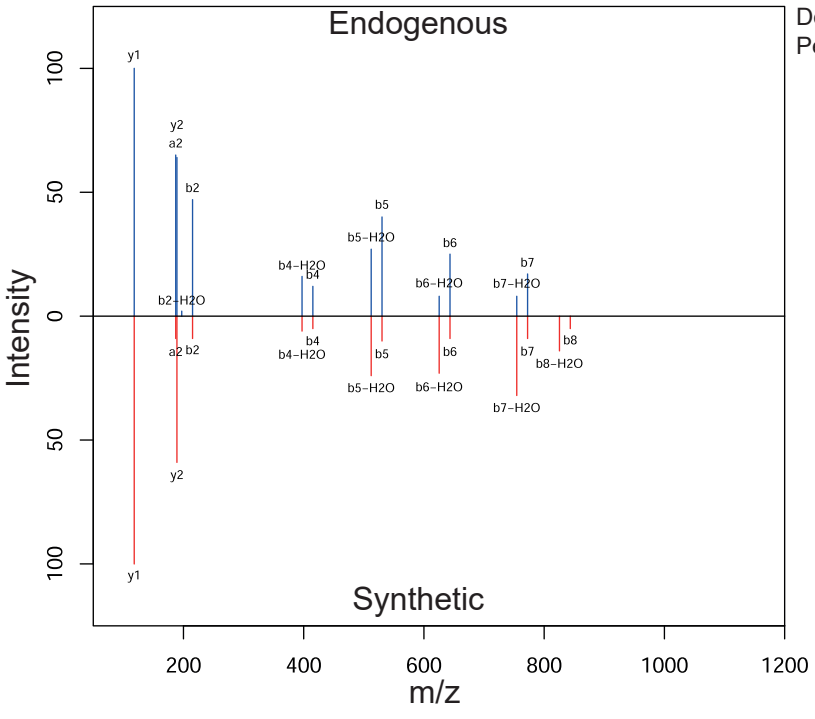
Endogenous peptide



Synthetic peptide



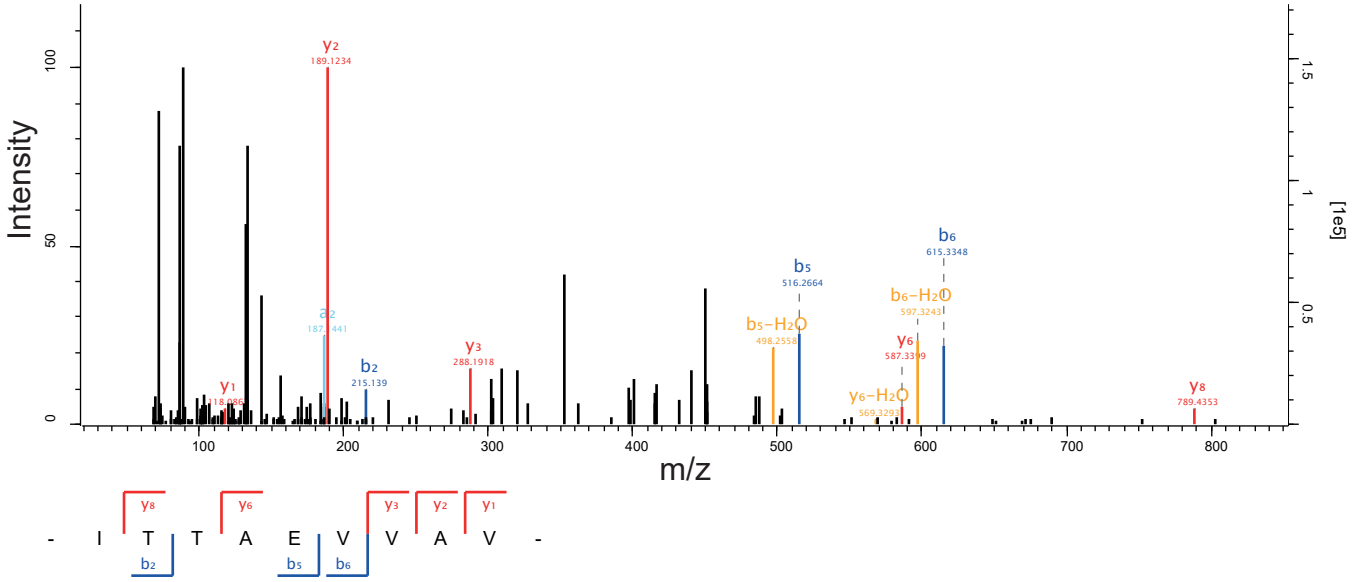
Endogenous/Synthetic peptide correlation



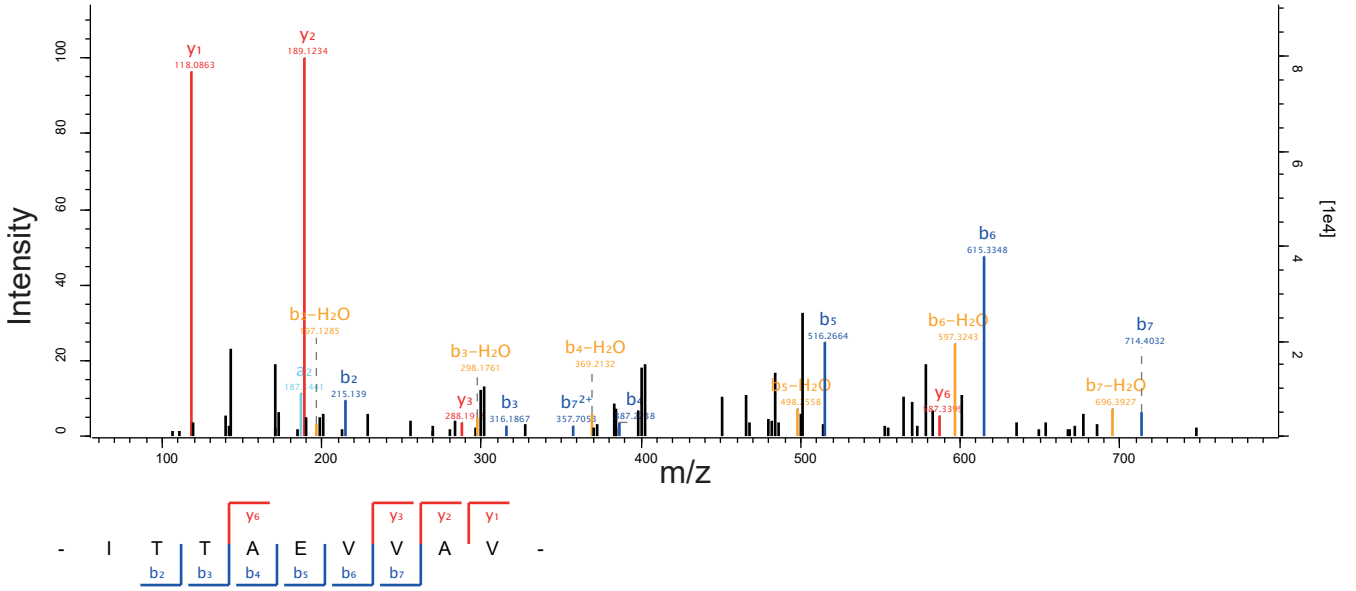
Dot product score=0.927
Pearson correlation=0.927

51BR-1 Eikenella corrodens HLA-I ITTAEVVAV

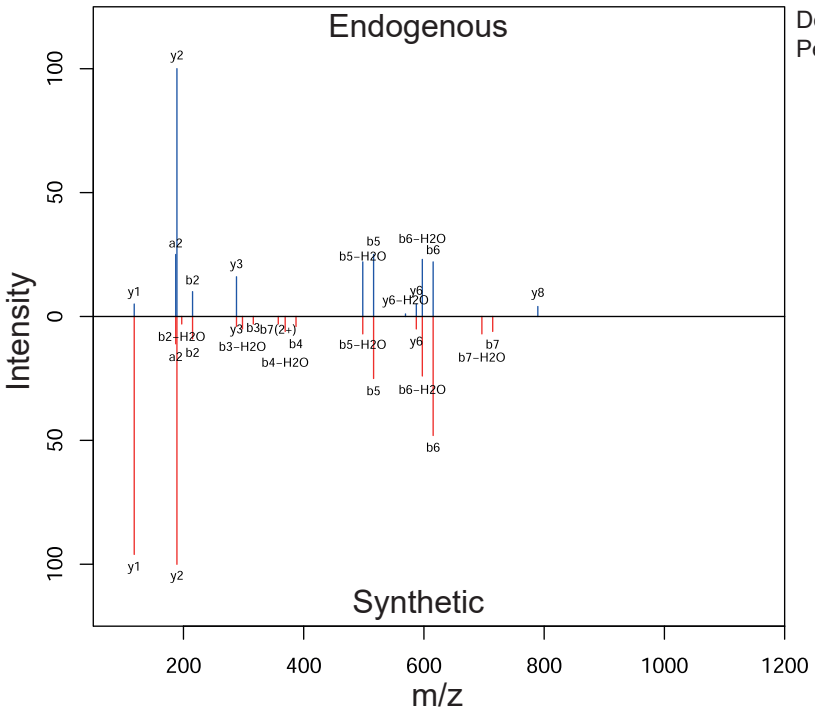
Endogenous peptide



Synthetic peptide



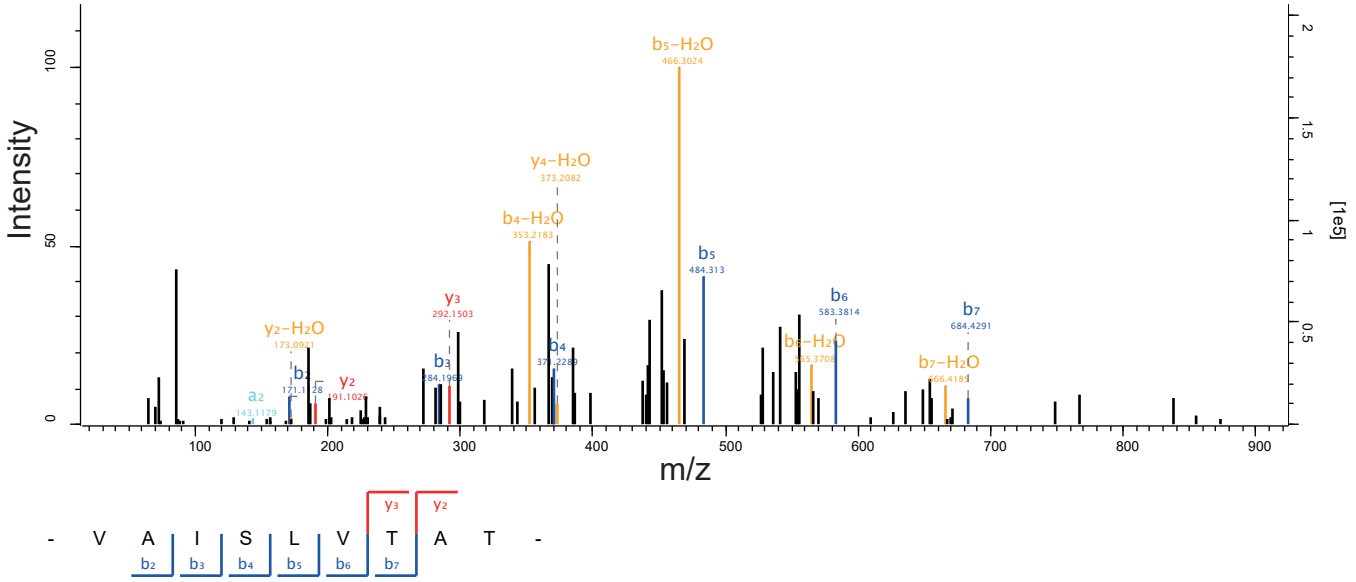
Endogenous/Synthetic peptide correlation



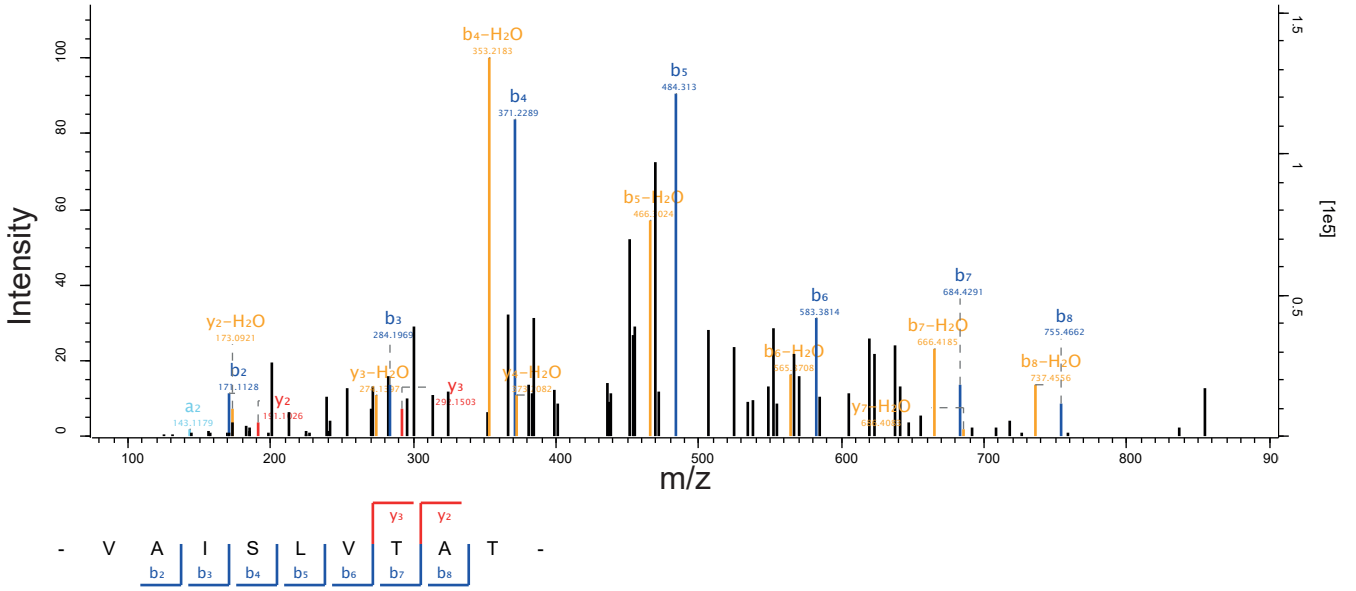
Dot product score=0.767
Pearson correlation=0.765

51BR-1 Eikenella corrodens HLA-I VAISLVTAT

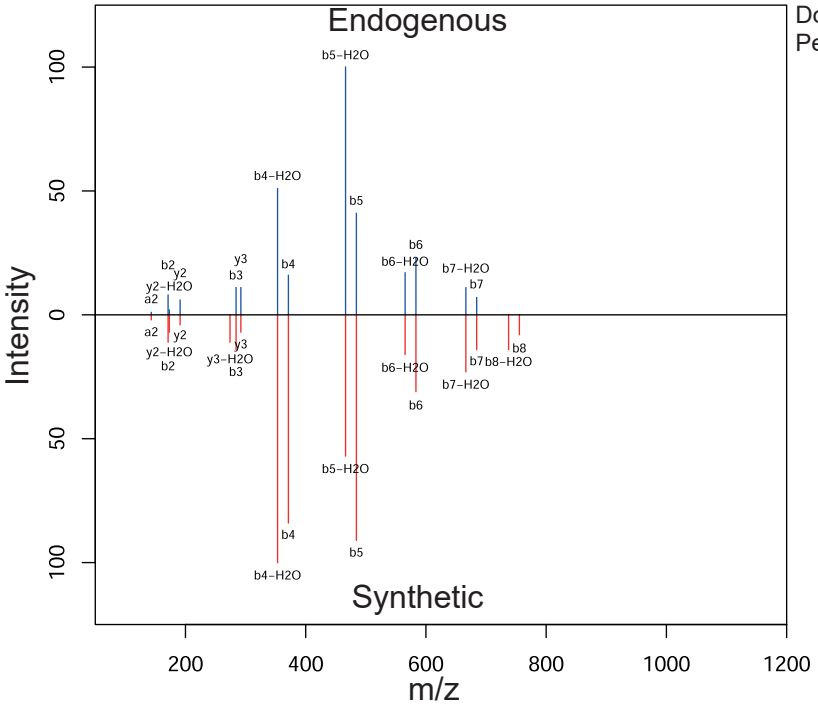
Endogenous peptide



Synthetic peptide



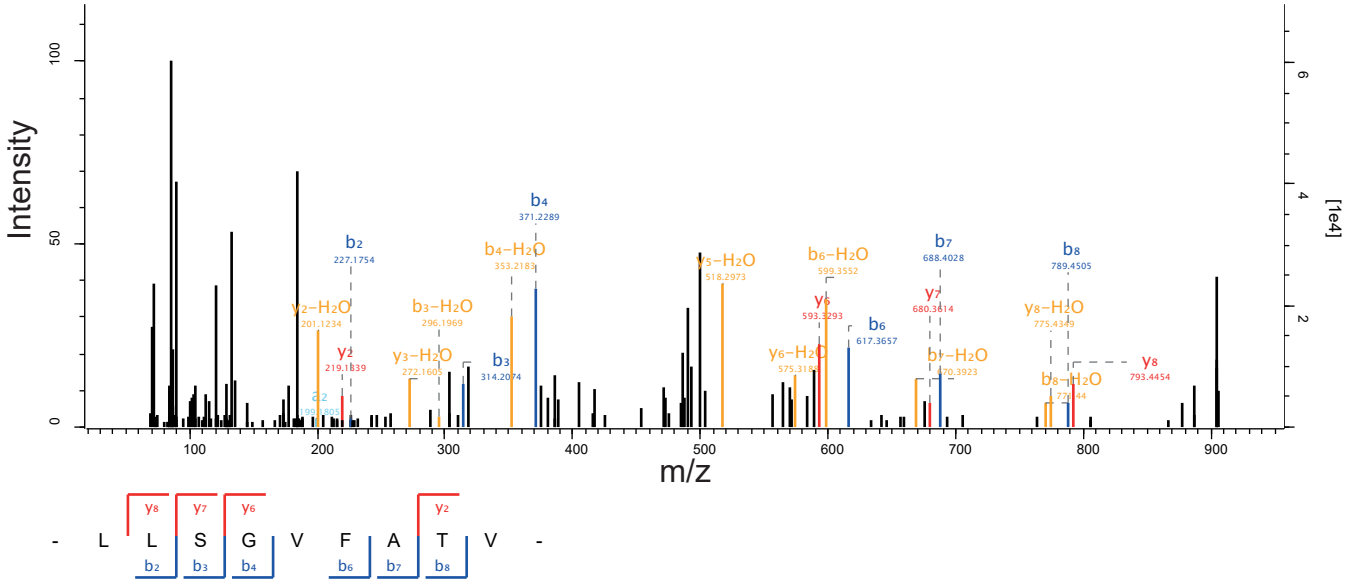
Endogenous/Synthetic peptide correlation



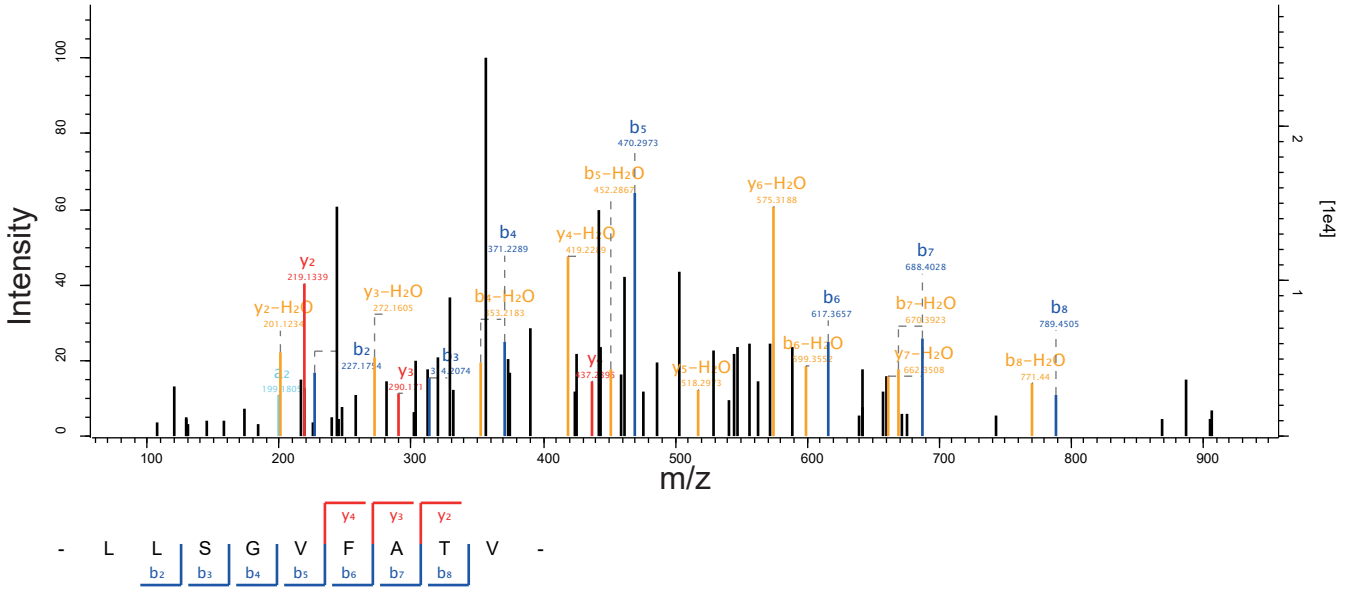
Dot product score=0.894
Pearson correlation=0.893

51BR-1 Staphylococcus aureus HLA-I LLSGVFATV

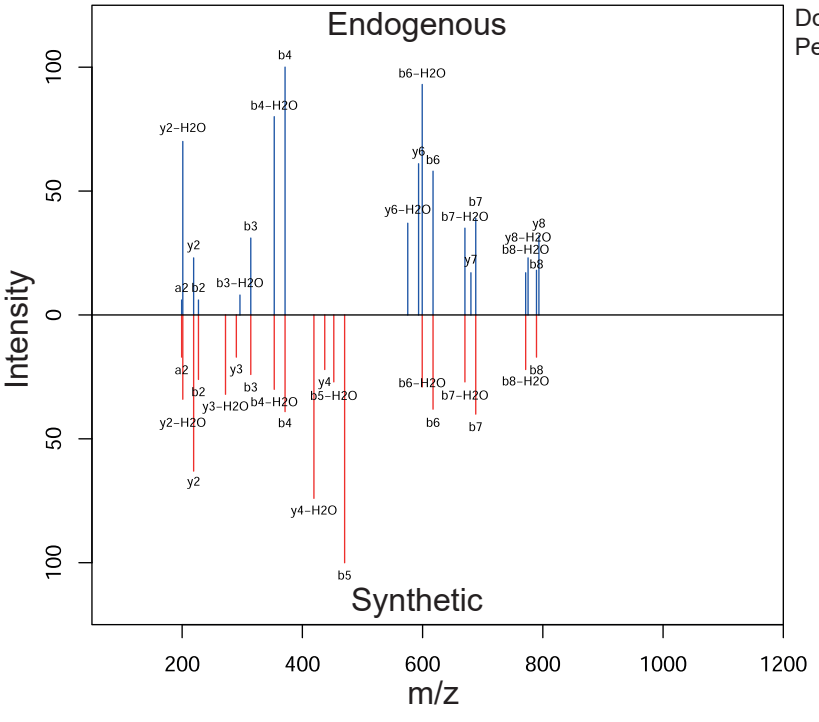
Endogenous peptide



Synthetic peptide



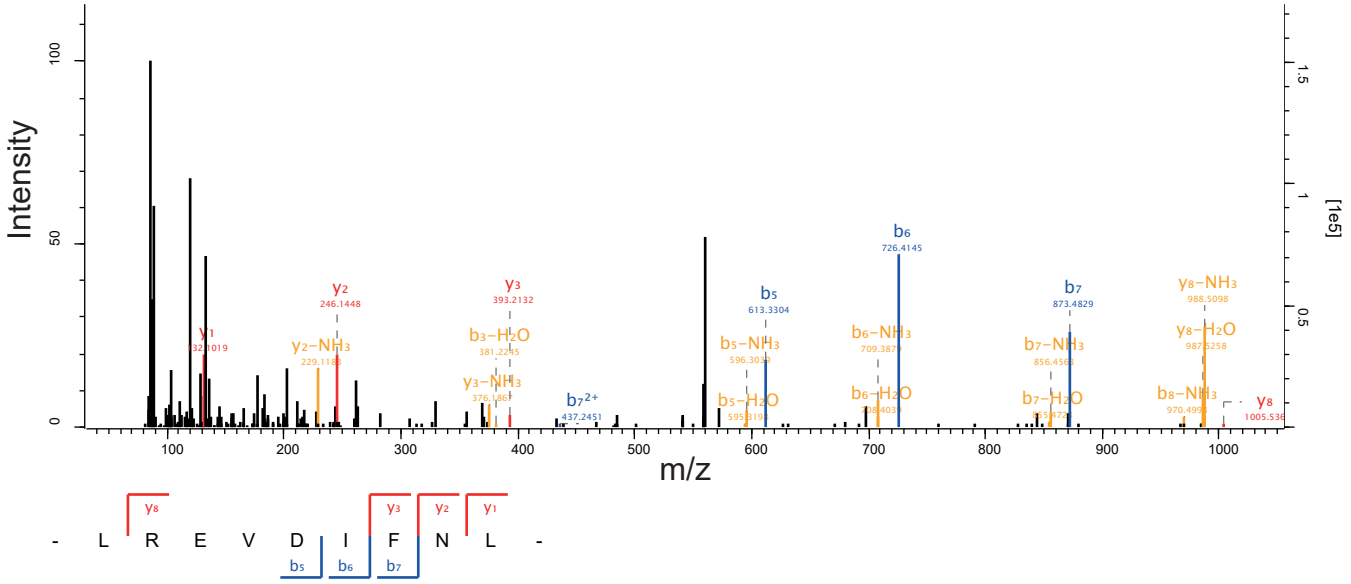
Endogenous/Synthetic peptide correlation



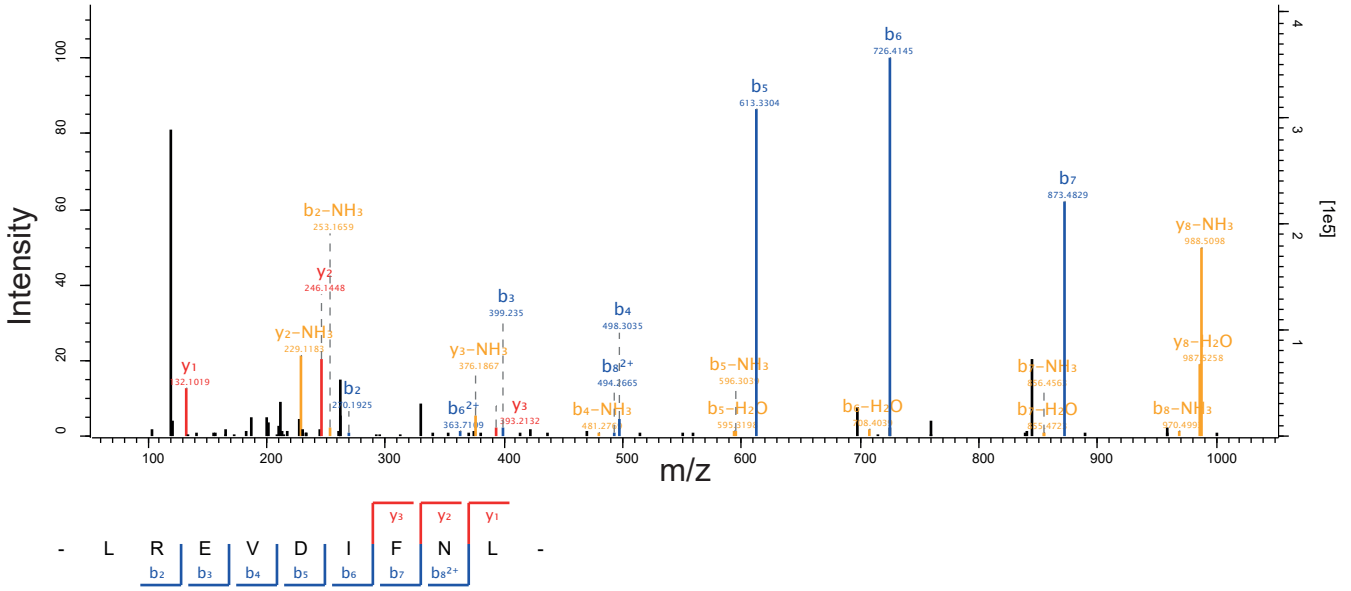
Dot product score=0.792
Pearson correlation=0.790

51BR-1 Staphylococcus aureus HLA-I LREVDIFNL

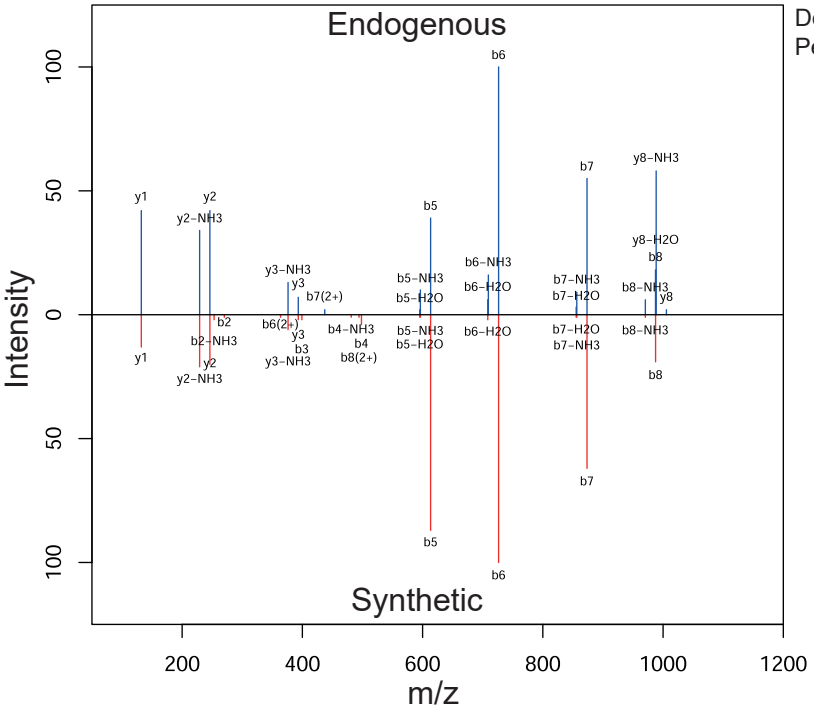
Endogenous peptide



Synthetic peptide



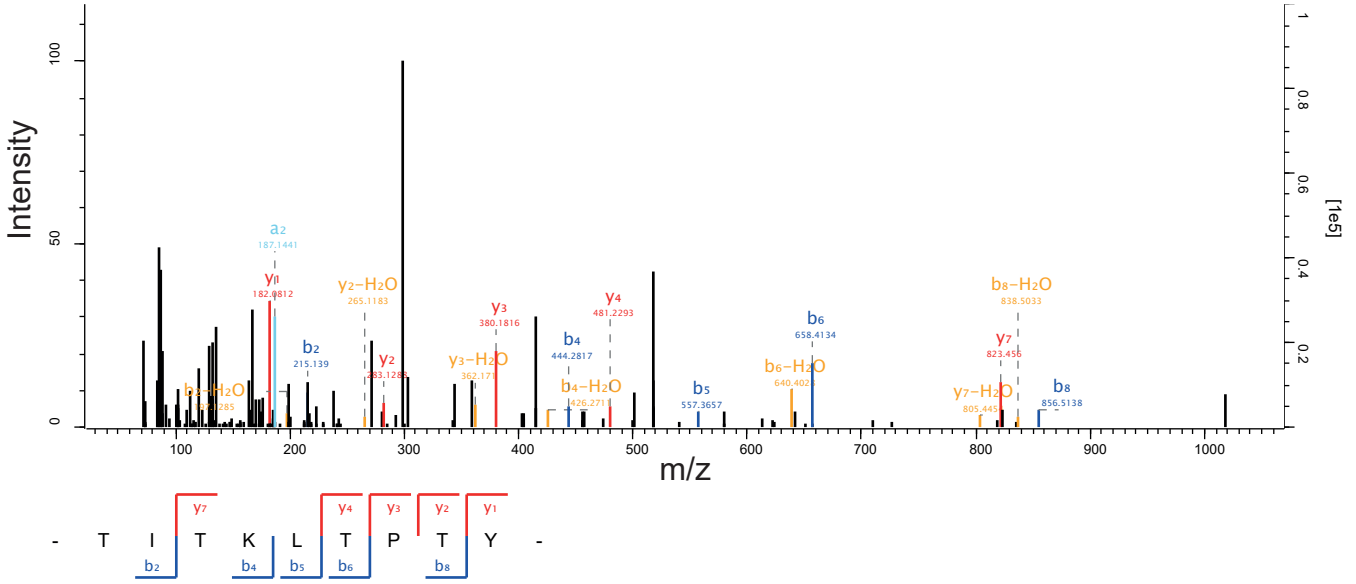
Endogenous/Synthetic peptide correlation



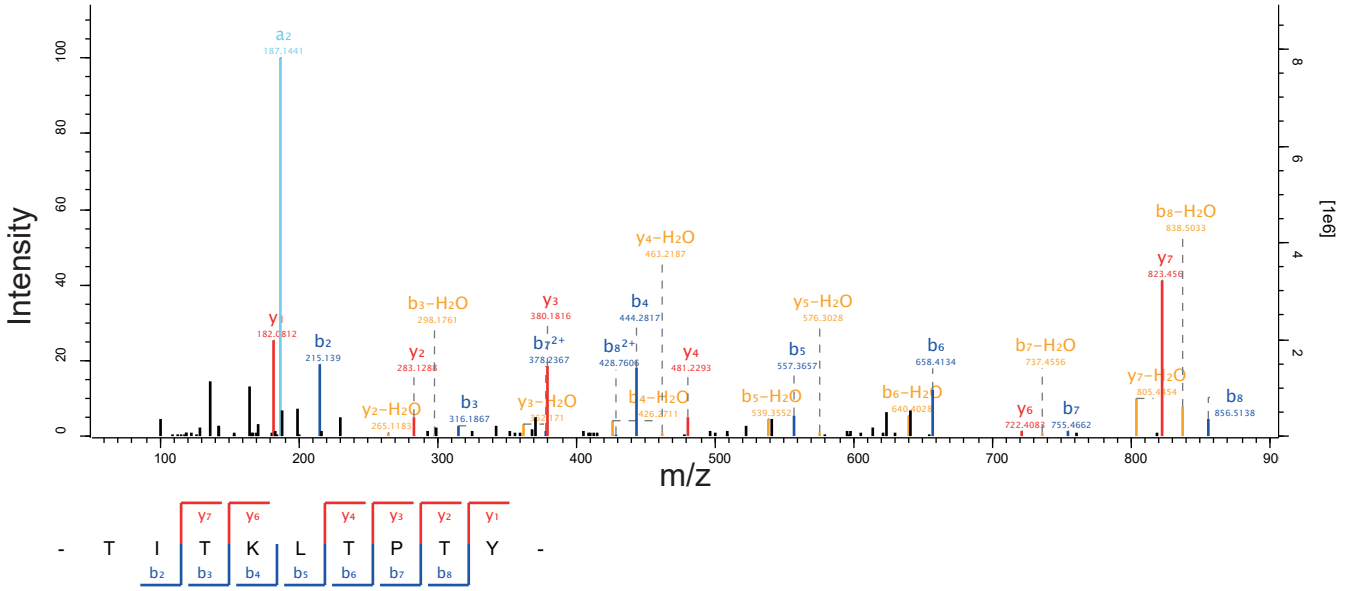
Dot product score=0.913
Pearson correlation=0.912

51BR-1 Staphylococcus aureus HLA-I TITKLTPTY

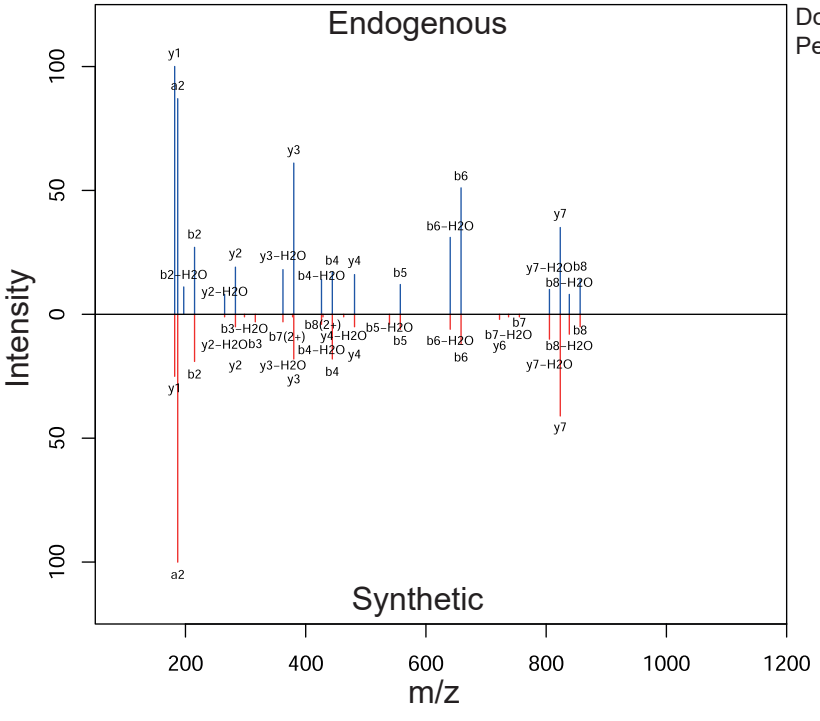
Endogenous peptide



Synthetic peptide



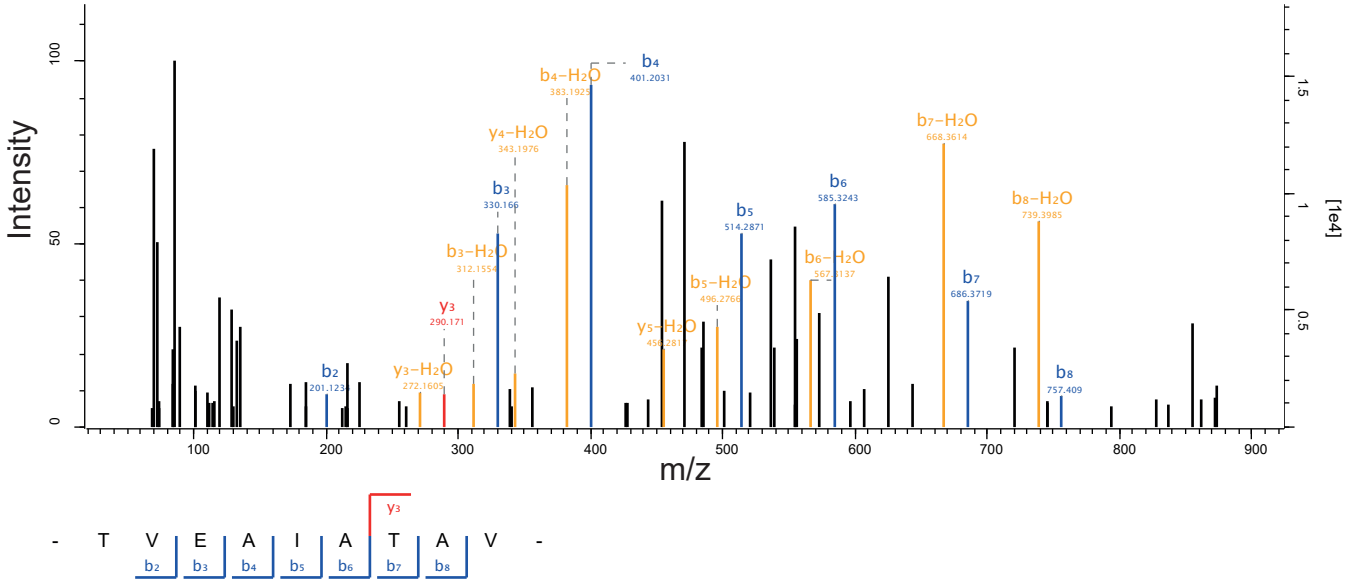
Endogenous/Synthetic peptide correlation



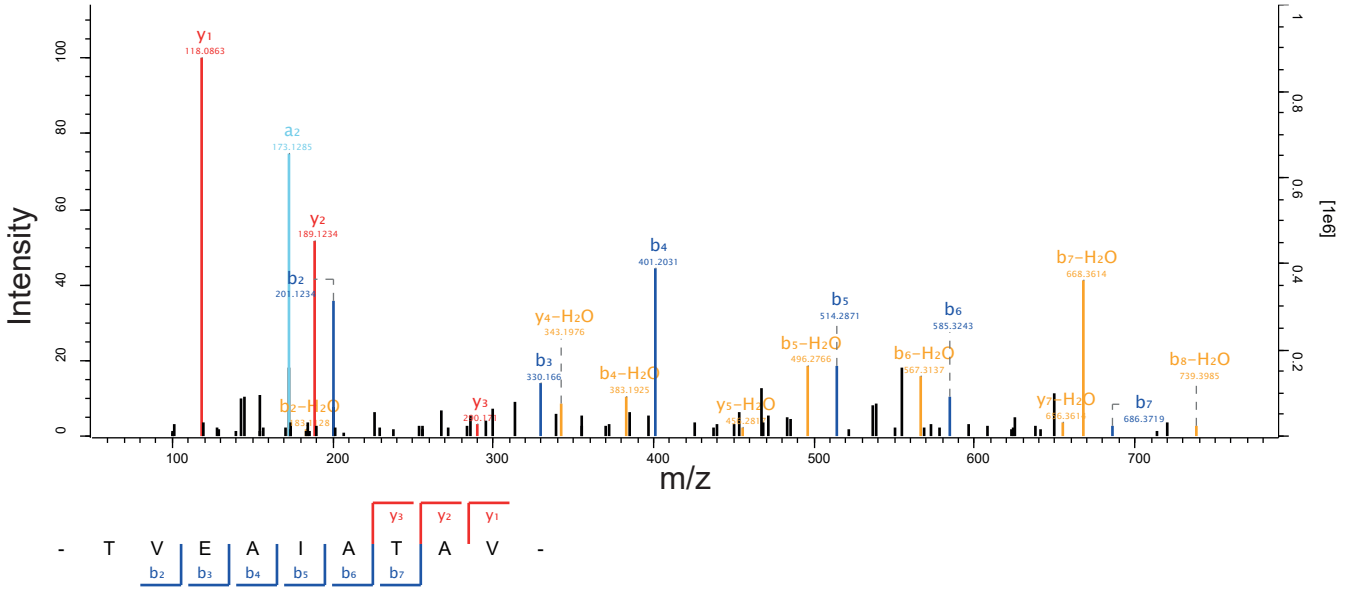
Dot product score=0.816
Pearson correlation=0.814

51BR-1 Staphylococcus aureus HLA-I TVEAIATAV

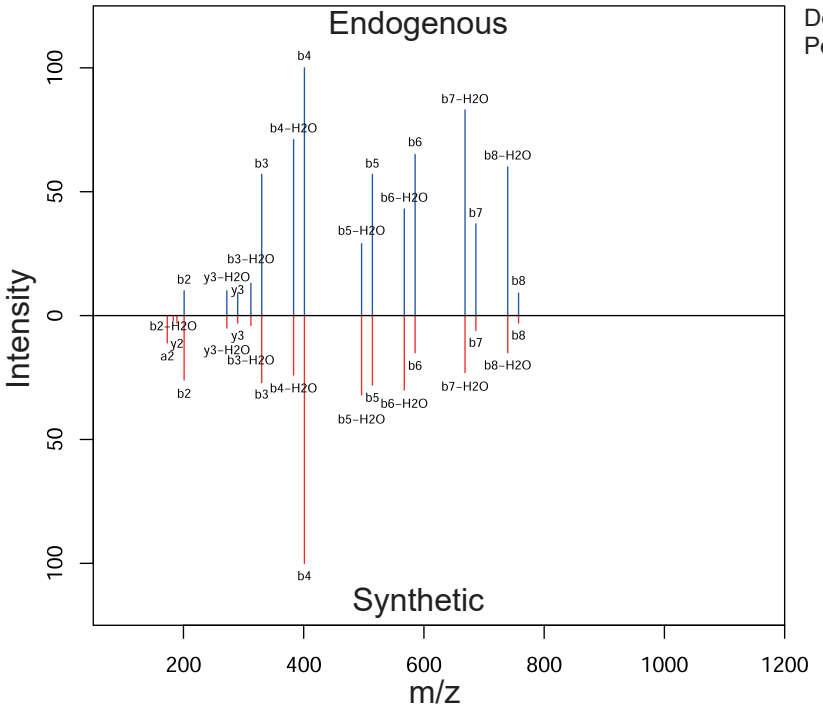
Endogenous peptide



Synthetic peptide



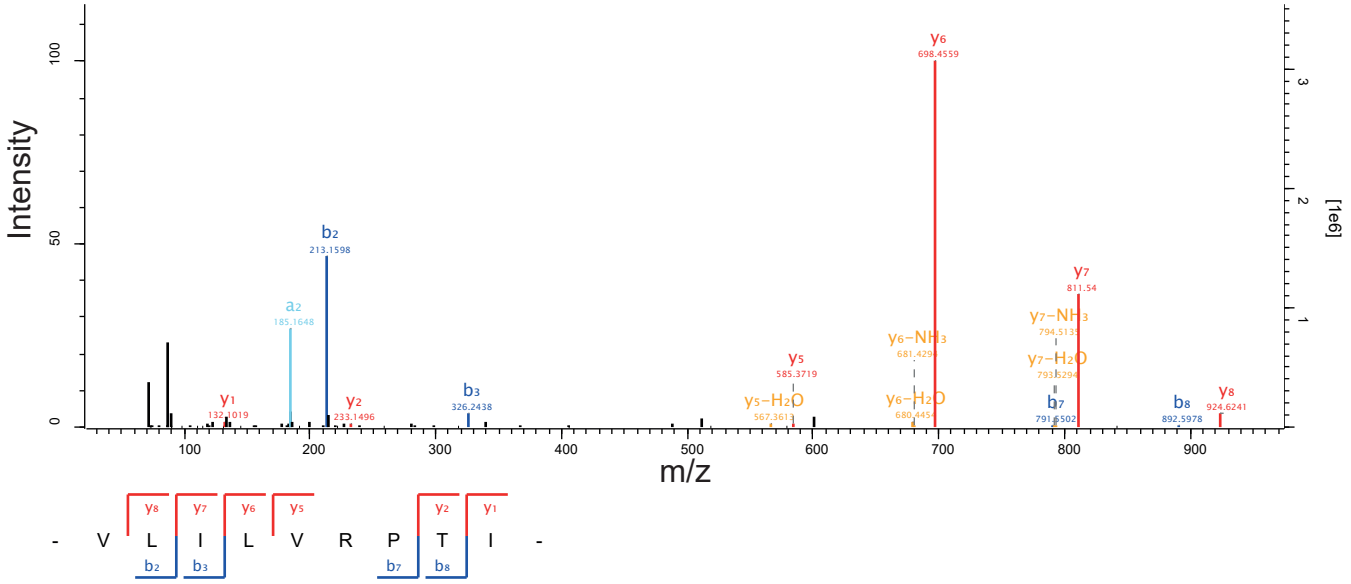
Endogenous/Synthetic peptide correlation



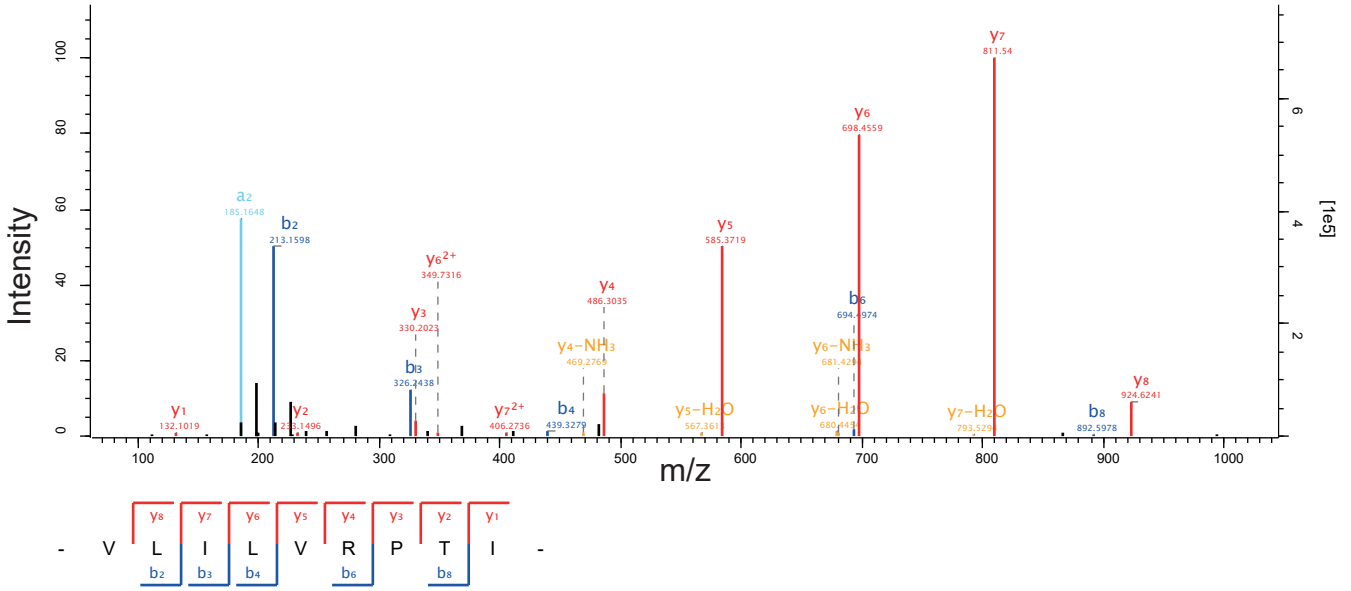
Dot product score=0.876
Pearson correlation=0.875

51BR-1 Staphylococcus aureus HLA-I VLILVRPTI

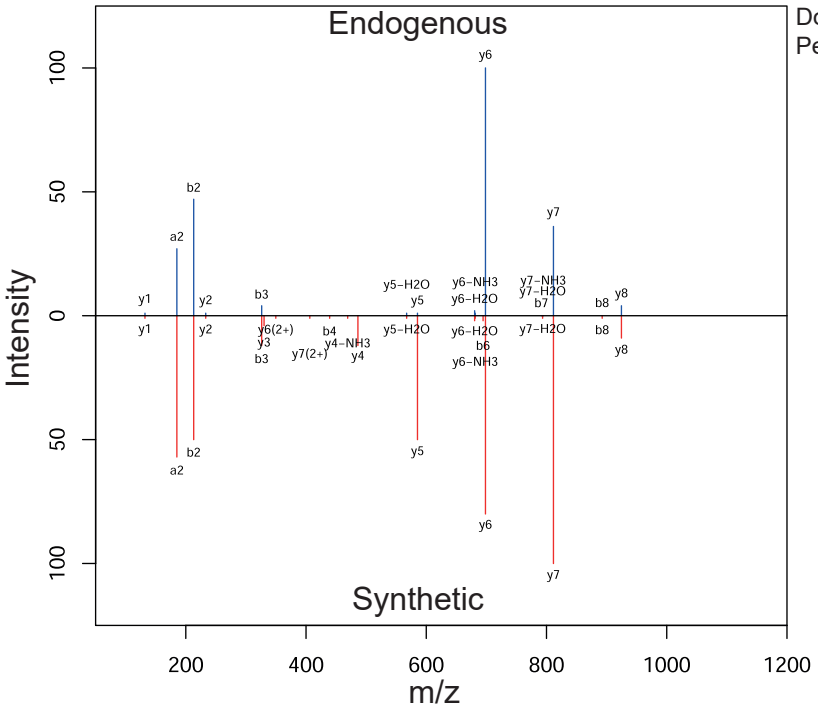
Endogenous peptide



Synthetic peptide



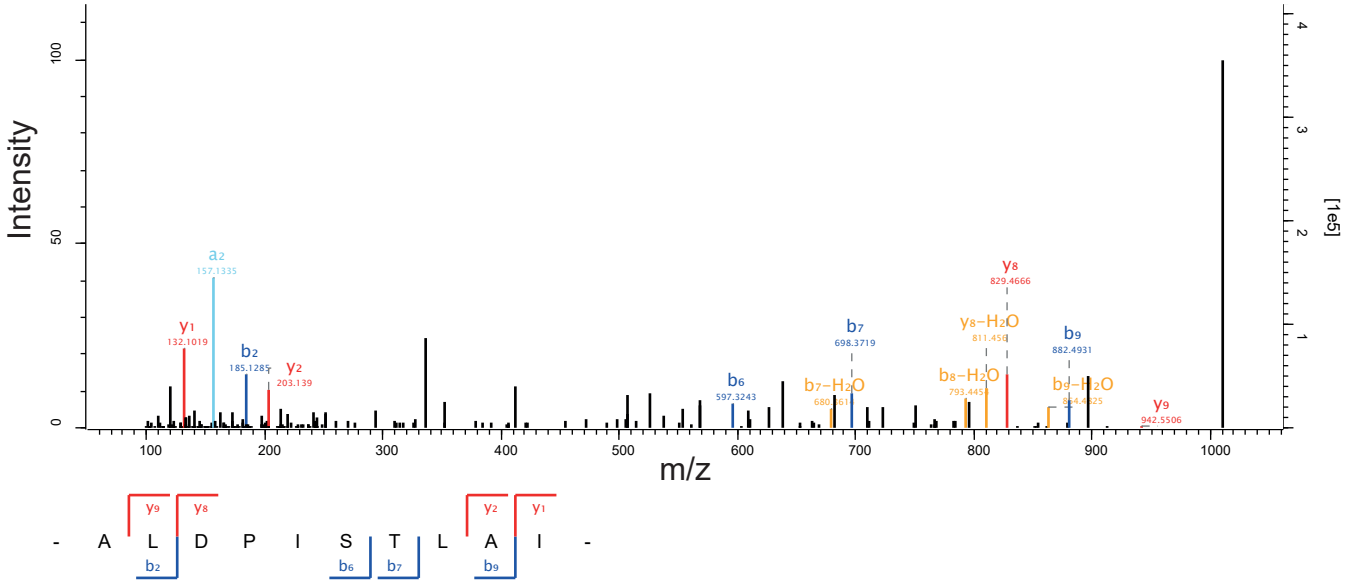
Endogenous/Synthetic peptide correlation



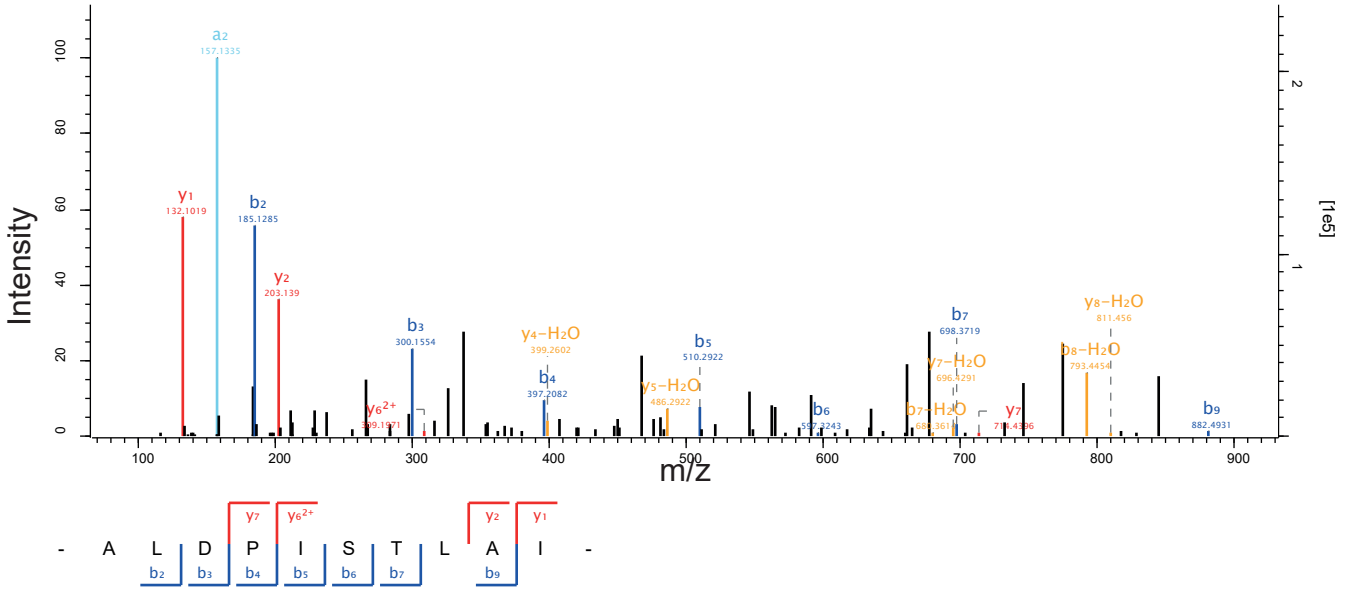
Dot product score=0.822
Pearson correlation=0.822

51BR-2 Brachybacterium alimentarium HLA-I ALDPISTLAI

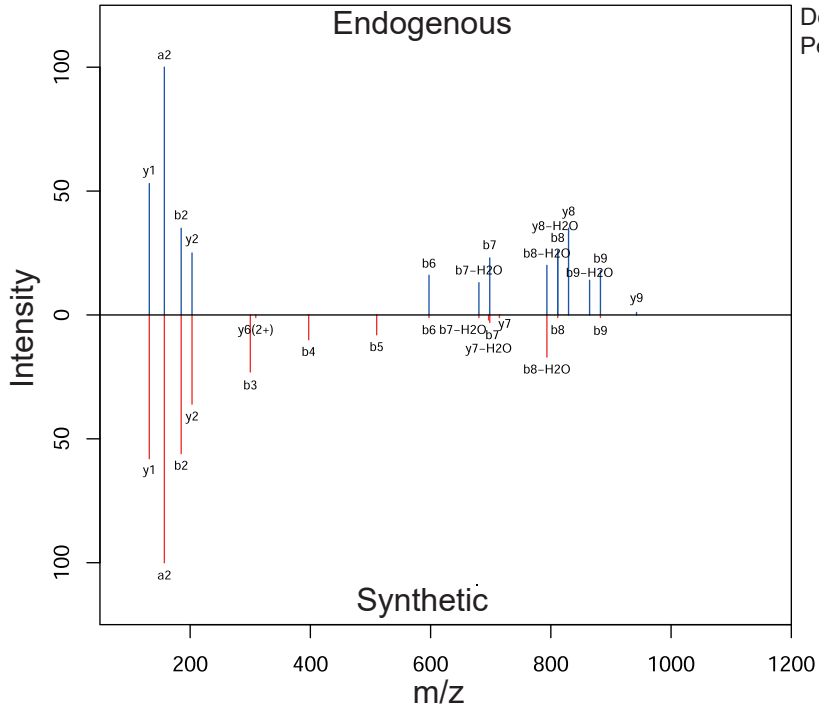
Endogenous peptide



Synthetic peptide



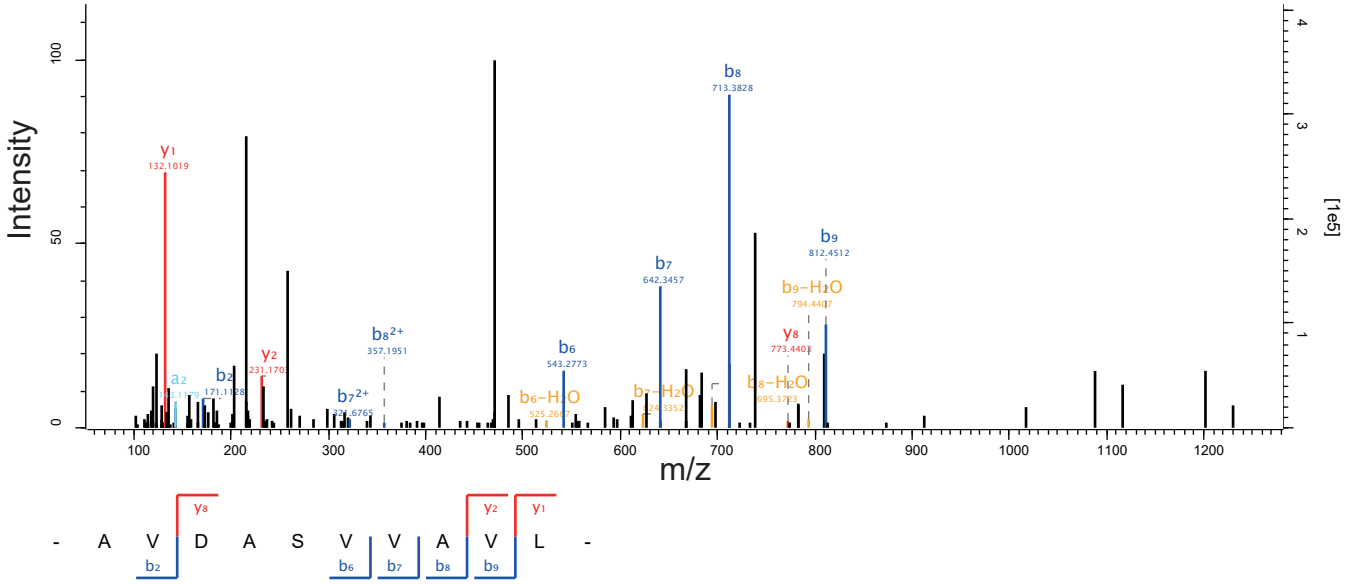
Endogenous/Synthetic peptide correlation



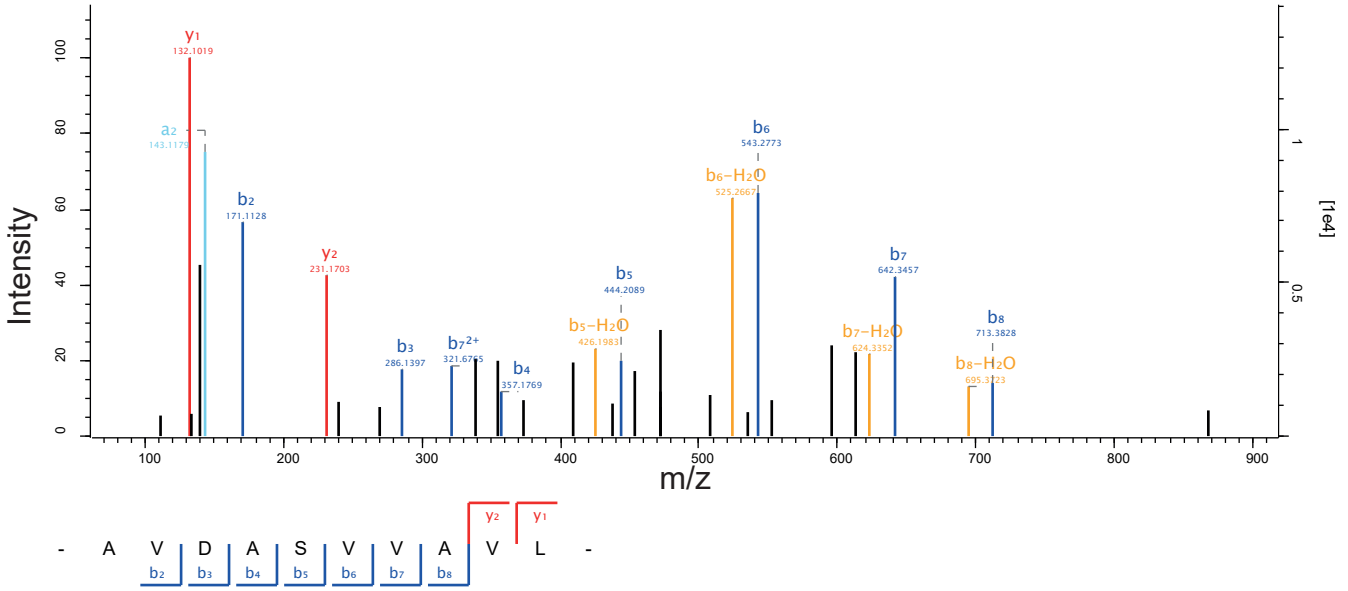
Dot product score=0.855
Pearson correlation=0.854

51BR-2 Brachybacterium alimentarium HLA-I AVDASVVAVL

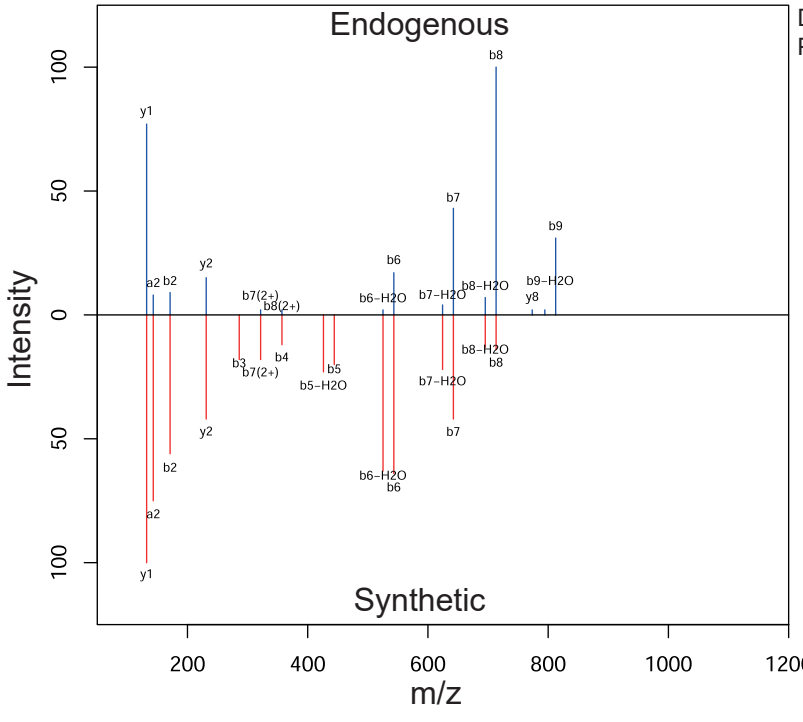
Endogenous peptide



Synthetic peptide



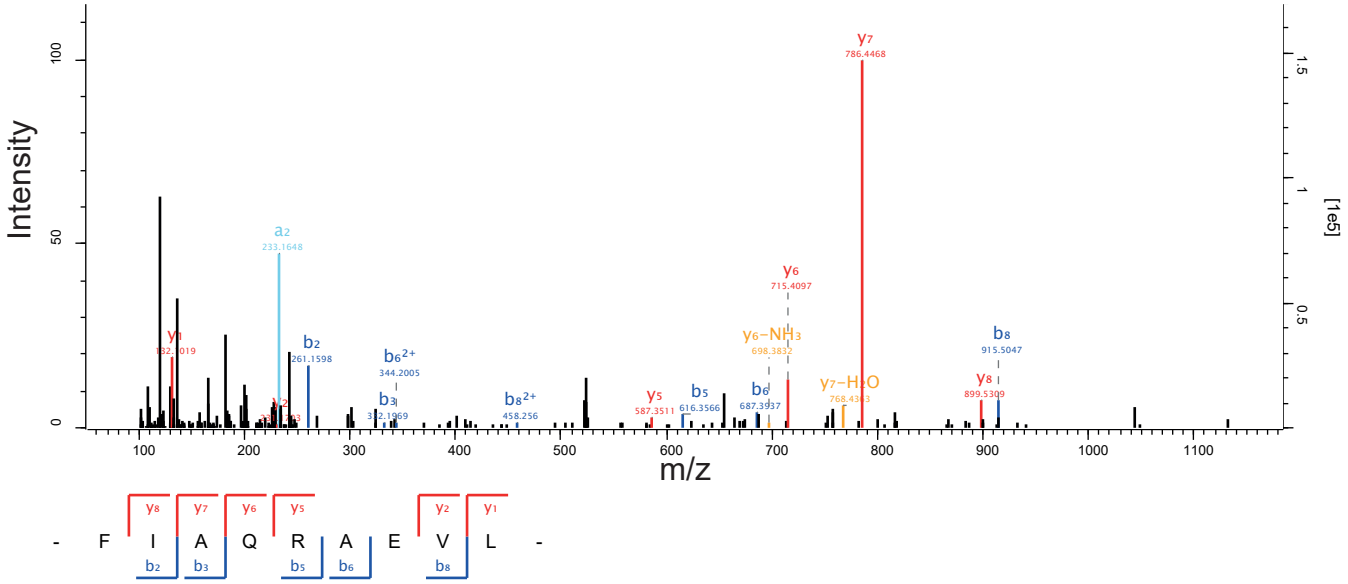
Endogenous/Synthetic peptide correlation



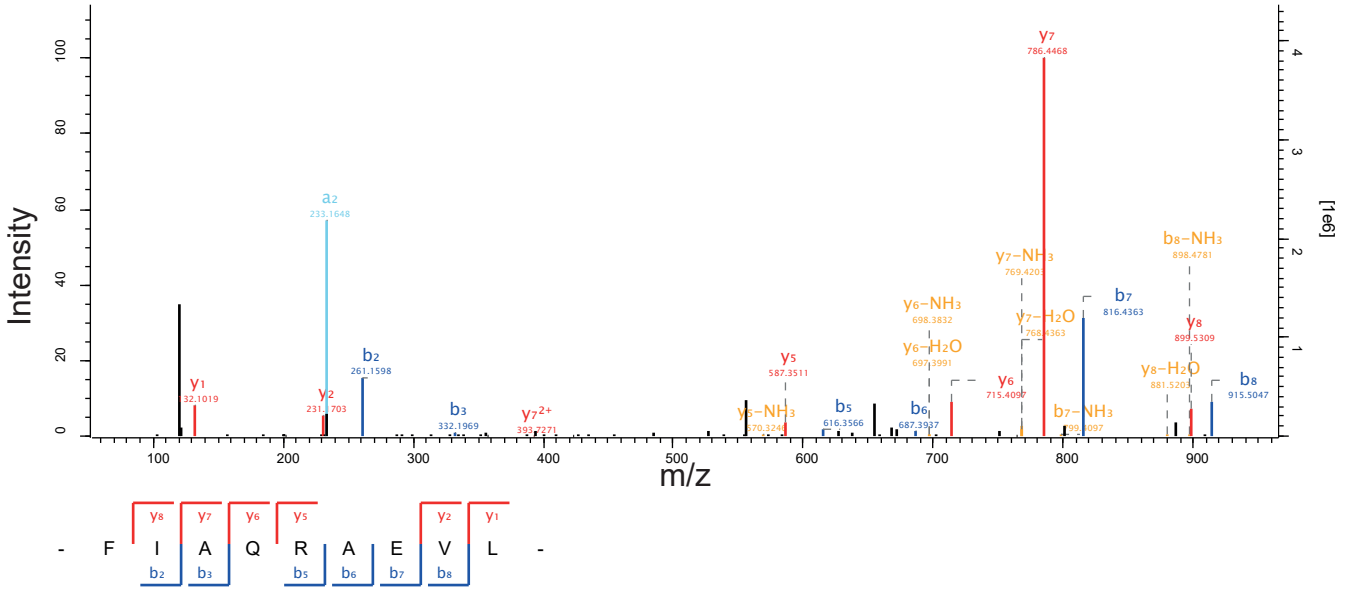
Dot product score=0.665
Pearson correlation=0.663

51BR-2 Brachybacterium alimentarium HLA-I FIAQRAEVL

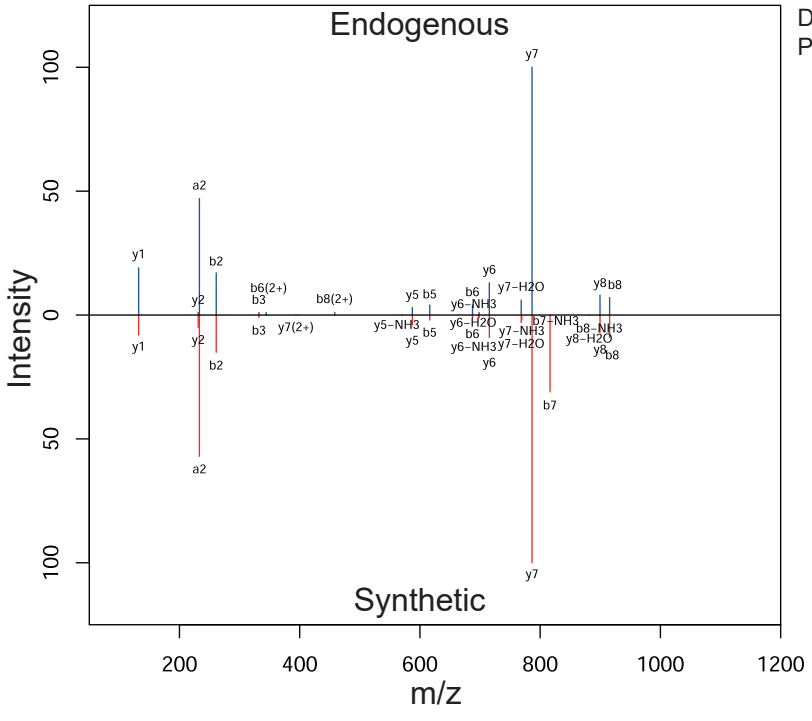
Endogenous peptide



Synthetic peptide



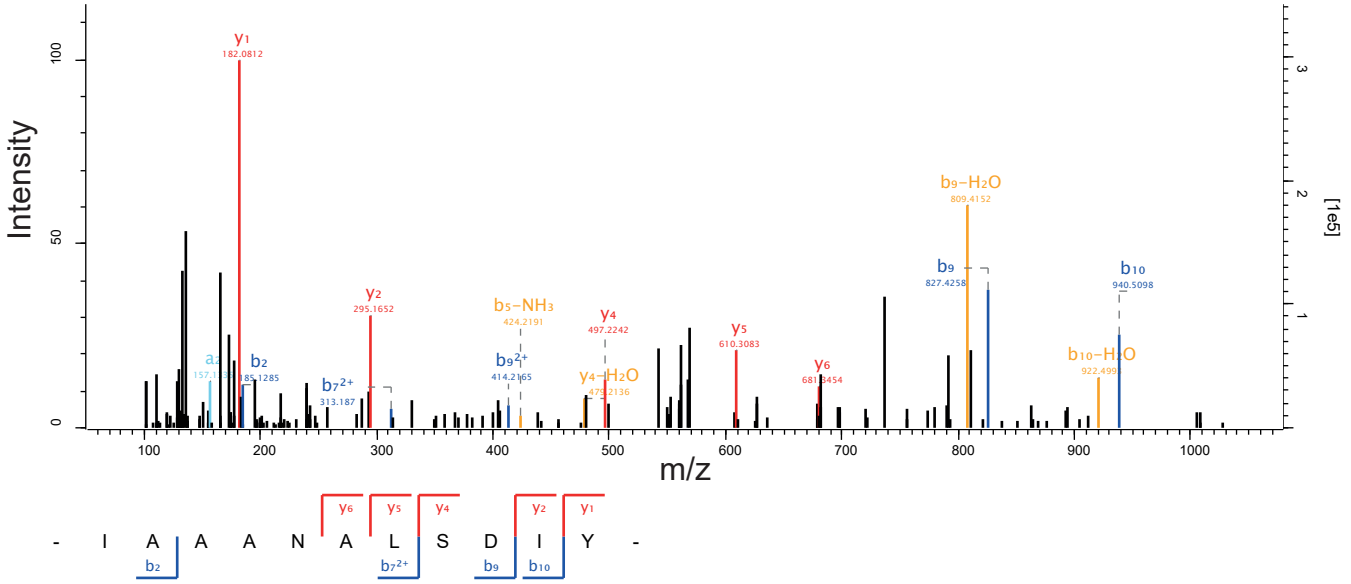
Endogenous/Synthetic peptide correlation



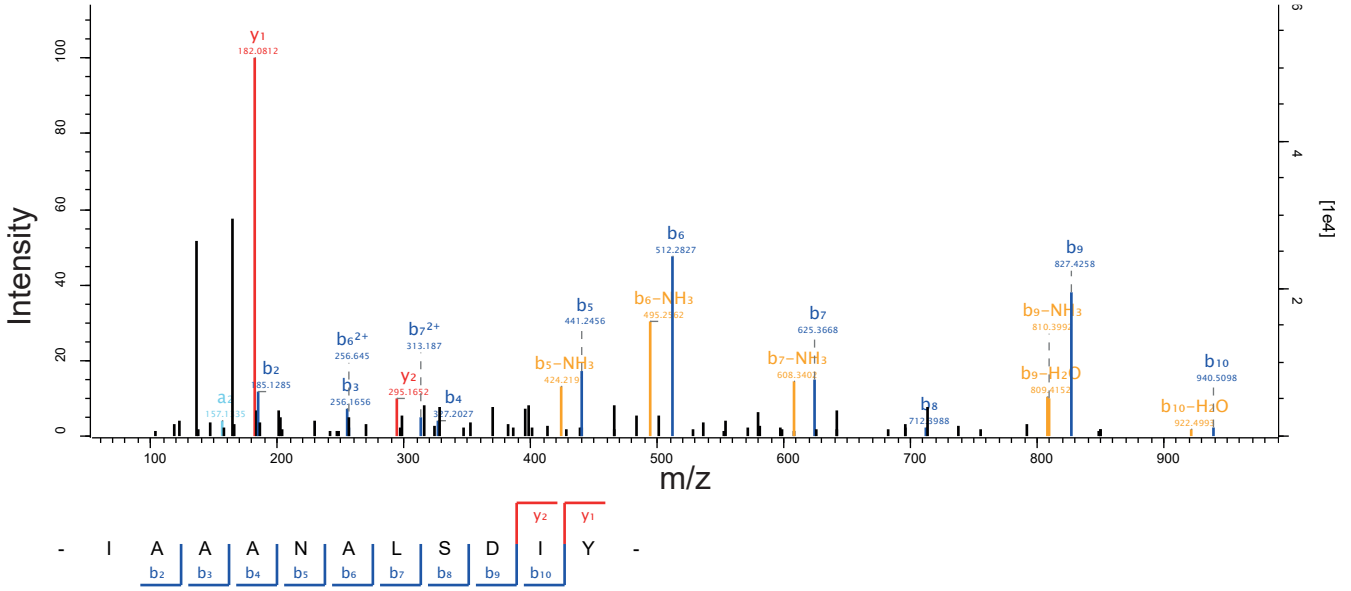
Dot product score=0.991
Pearson correlation=0.991

51BR-2 Brachybacterium alimentarium HLA-I IAAANALSDIY

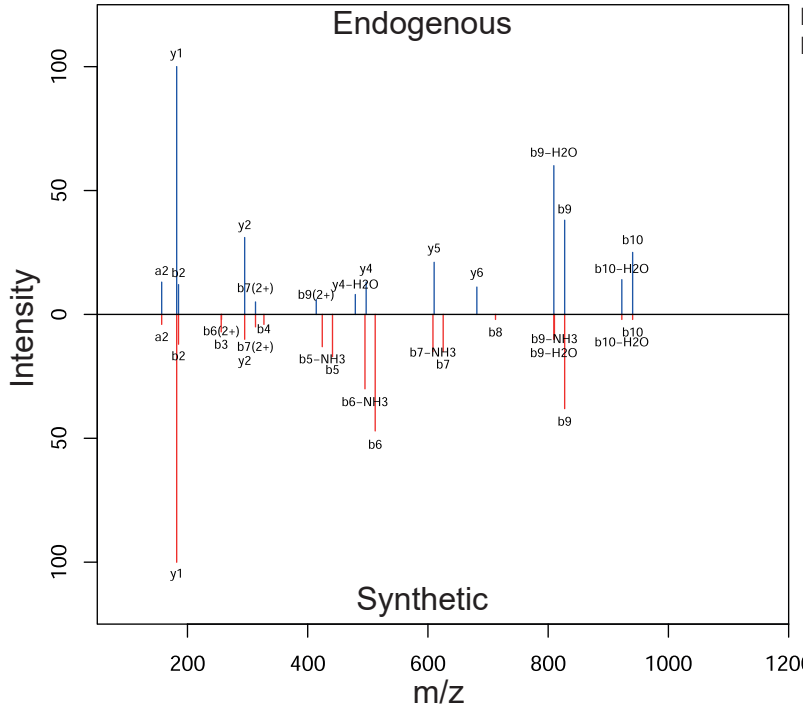
Endogenous peptide



Synthetic peptide



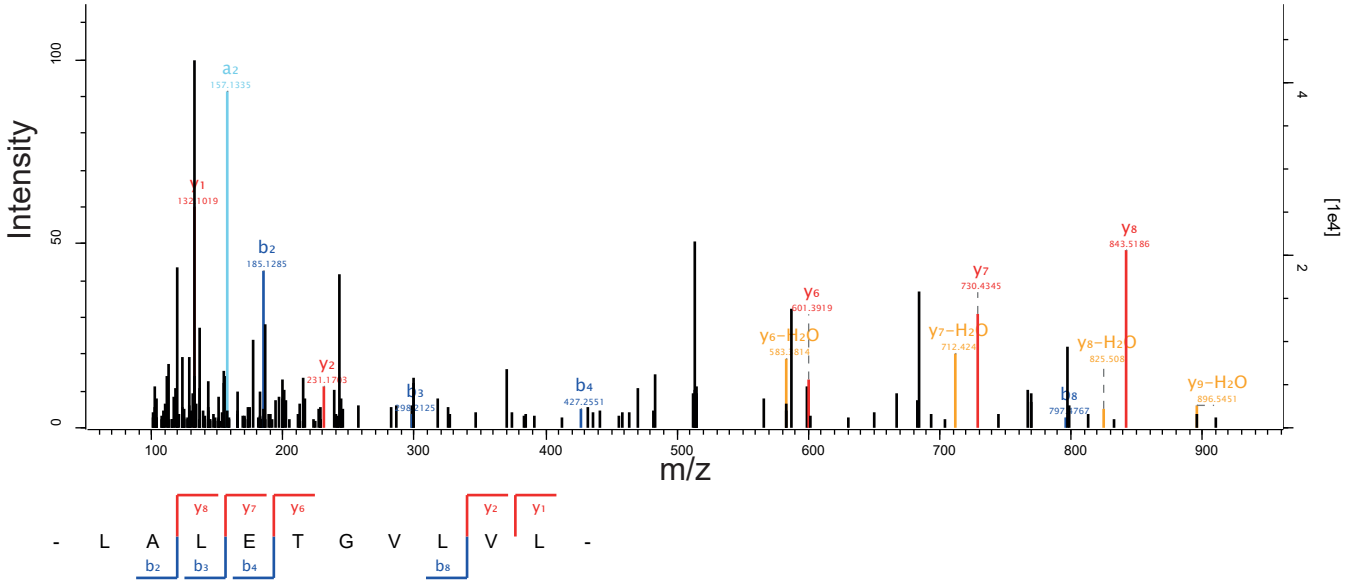
Endogenous/Synthetic peptide correlation



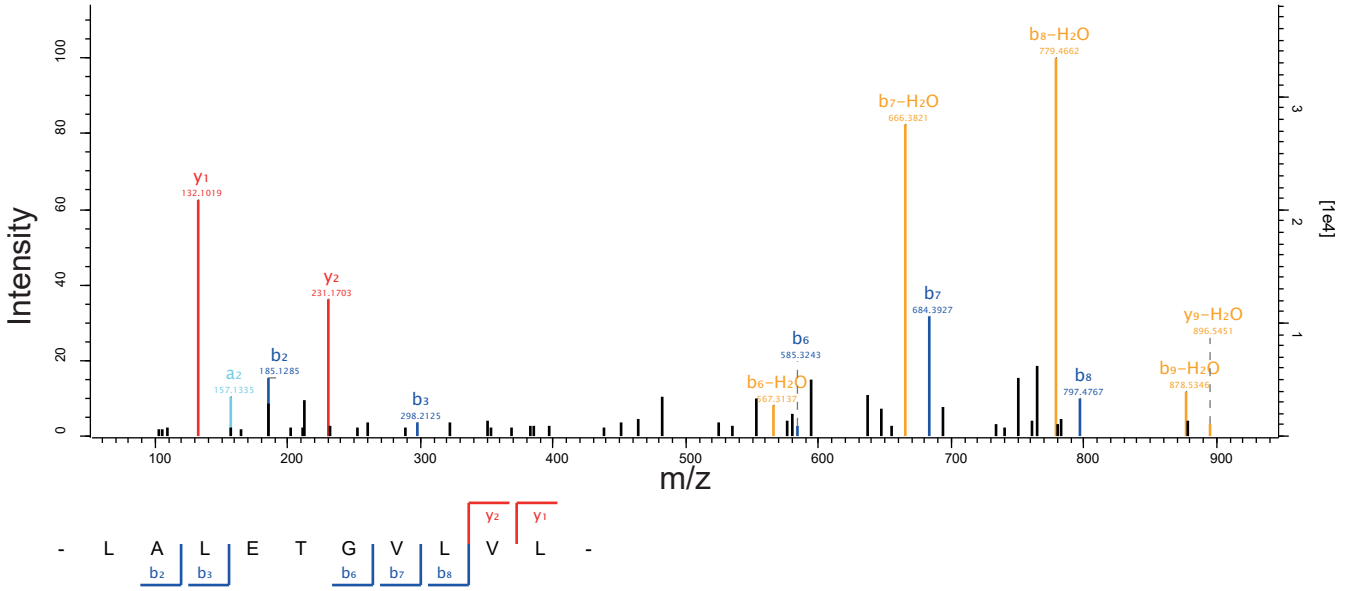
Dot product score=0.962
Pearson correlation=0.962

51BR-2 Brachybacterium alimentarium HLA-I LALETGVLVL

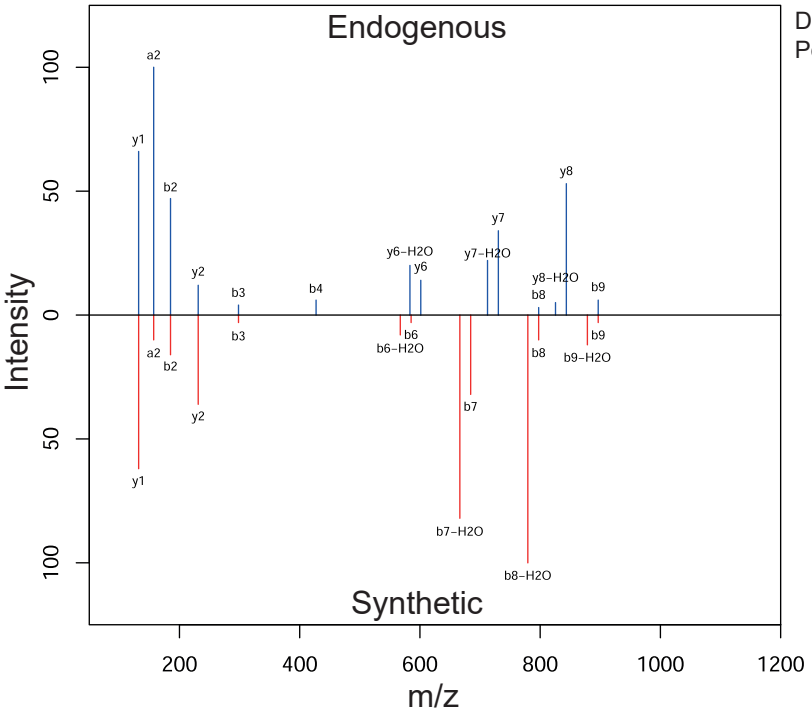
Endogenous peptide



Synthetic peptide



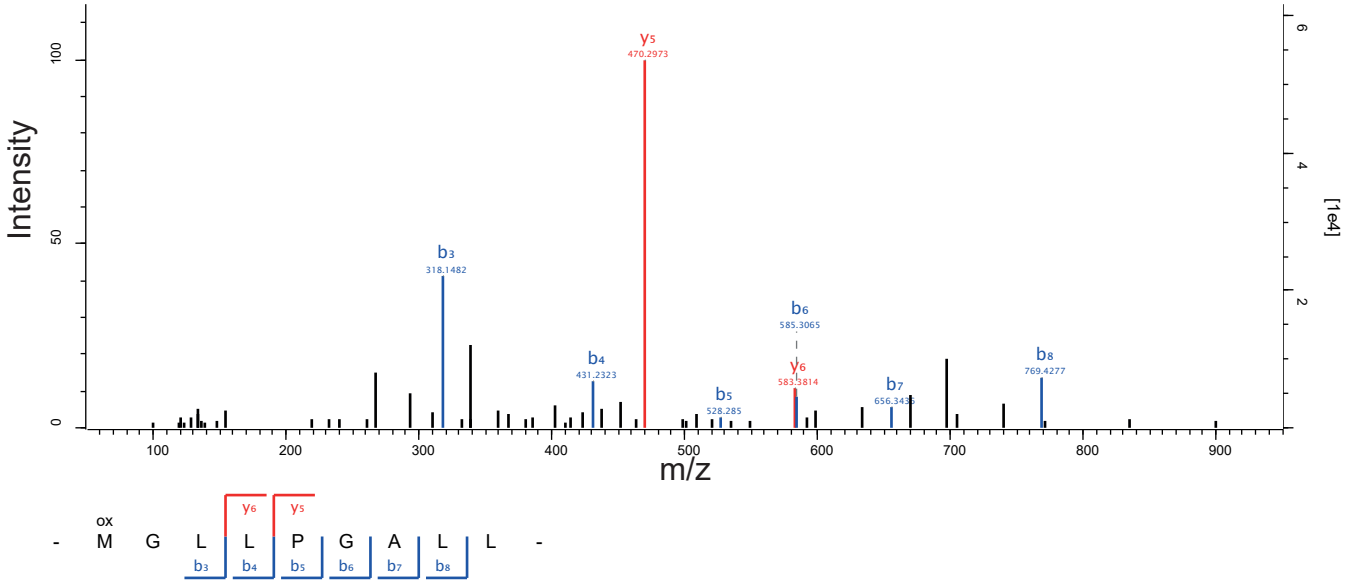
Endogenous/Synthetic peptide correlation



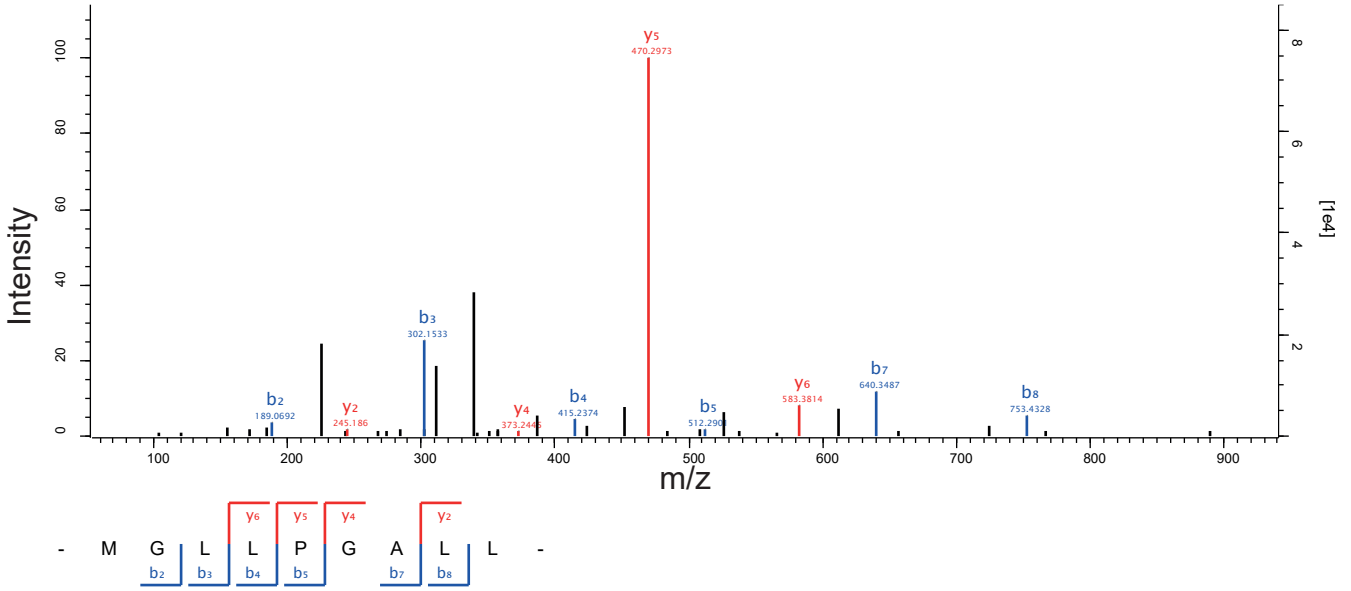
Dot product score=0.871
Pearson correlation=0.871

51BR-2 Brachybacterium alimentarium HLA-I MGLLPGALL

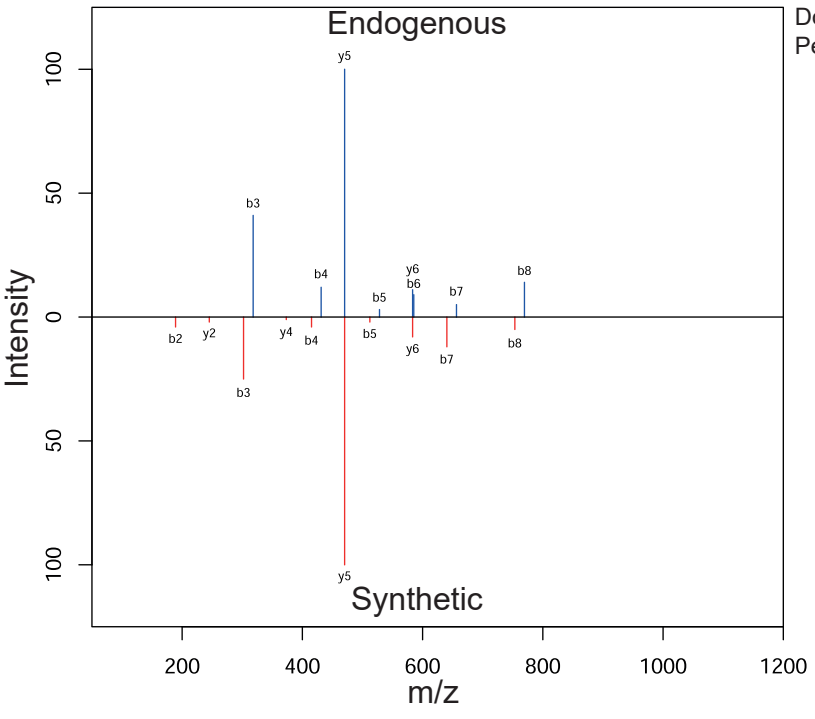
Endogenous peptide



Synthetic peptide



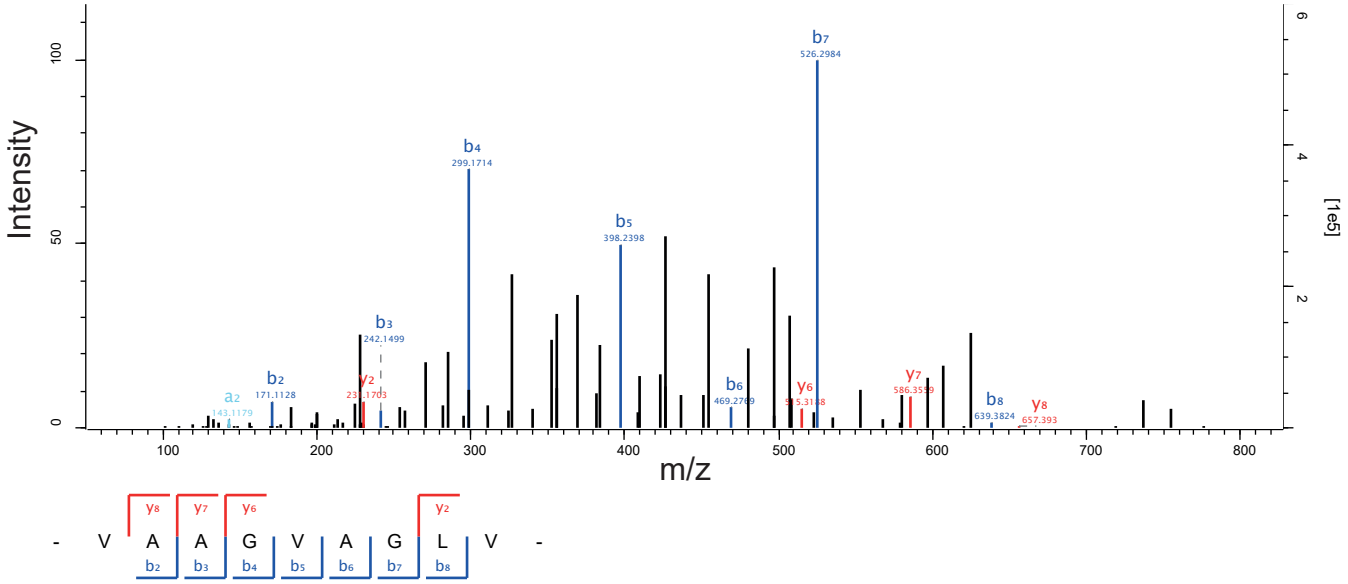
Endogenous/Synthetic peptide correlation



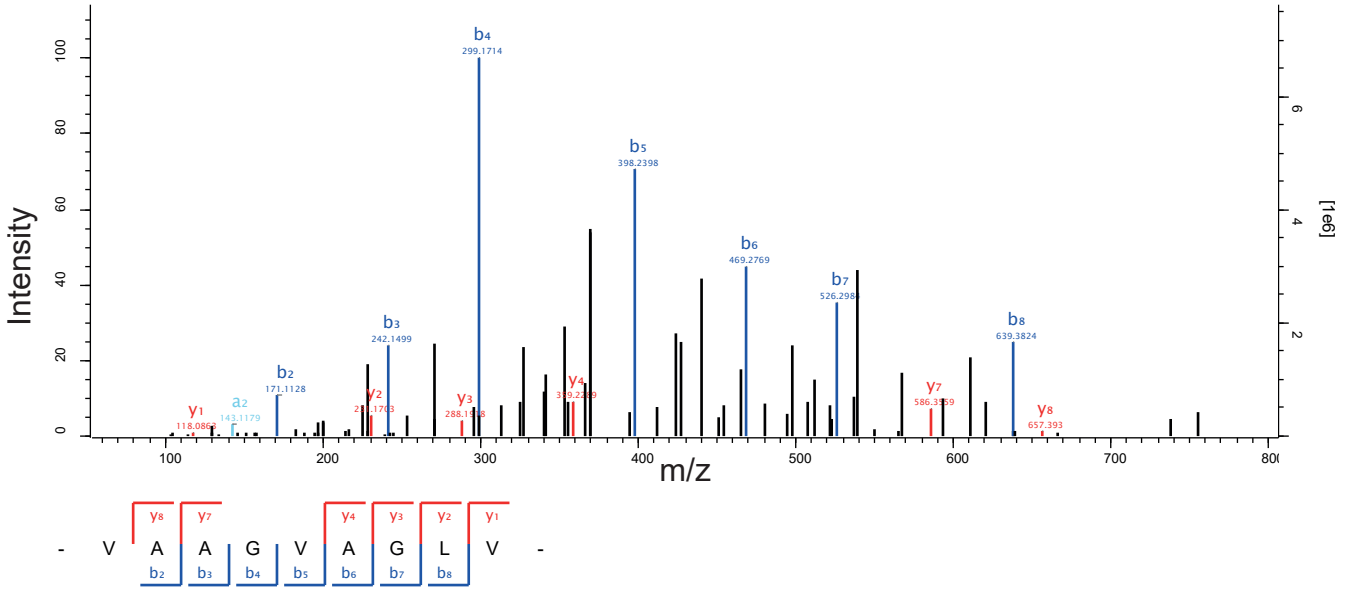
Dot product score=0.875
Pearson correlation=0.874

51BR-2 Brachybacterium alimentarium HLA-I VAAGVAGLV

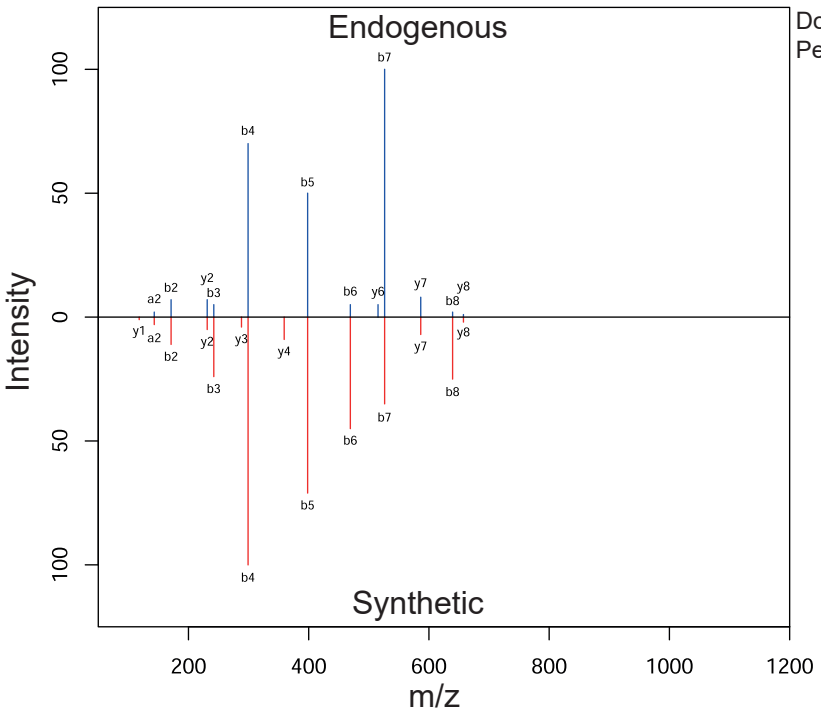
Endogenous peptide



Synthetic peptide



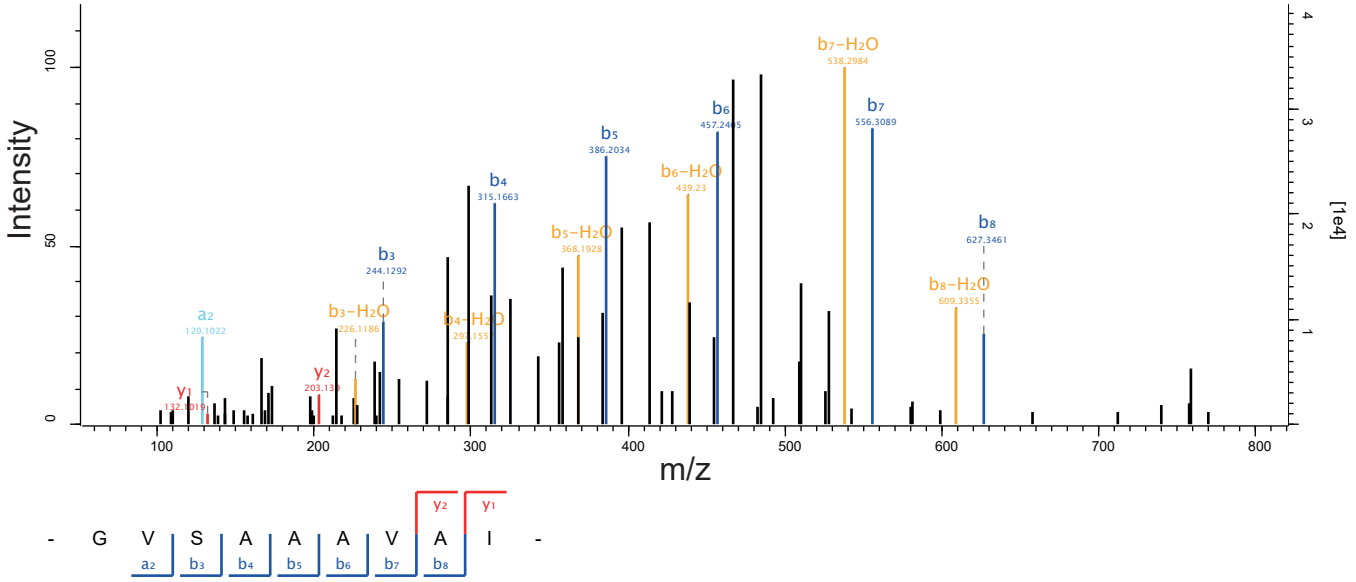
Endogenous/Synthetic peptide correlation



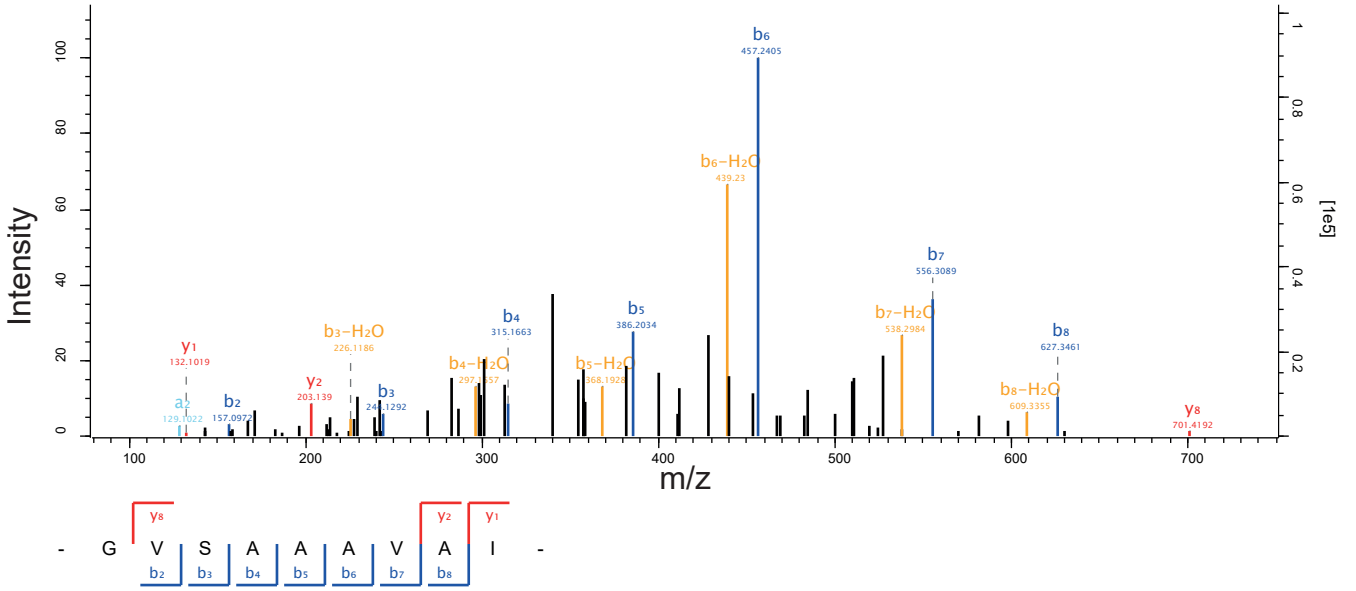
Dot product score=0.787
Pearson correlation=0.785

51BR-2 Eikenella corrodens HLA-I GVSAAAVAI

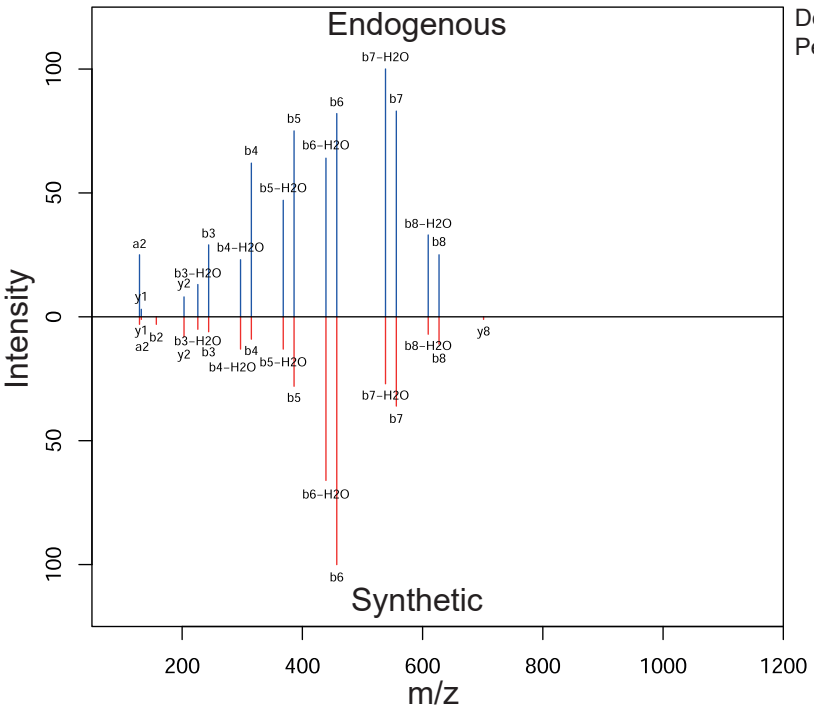
Endogenous peptide



Synthetic peptide



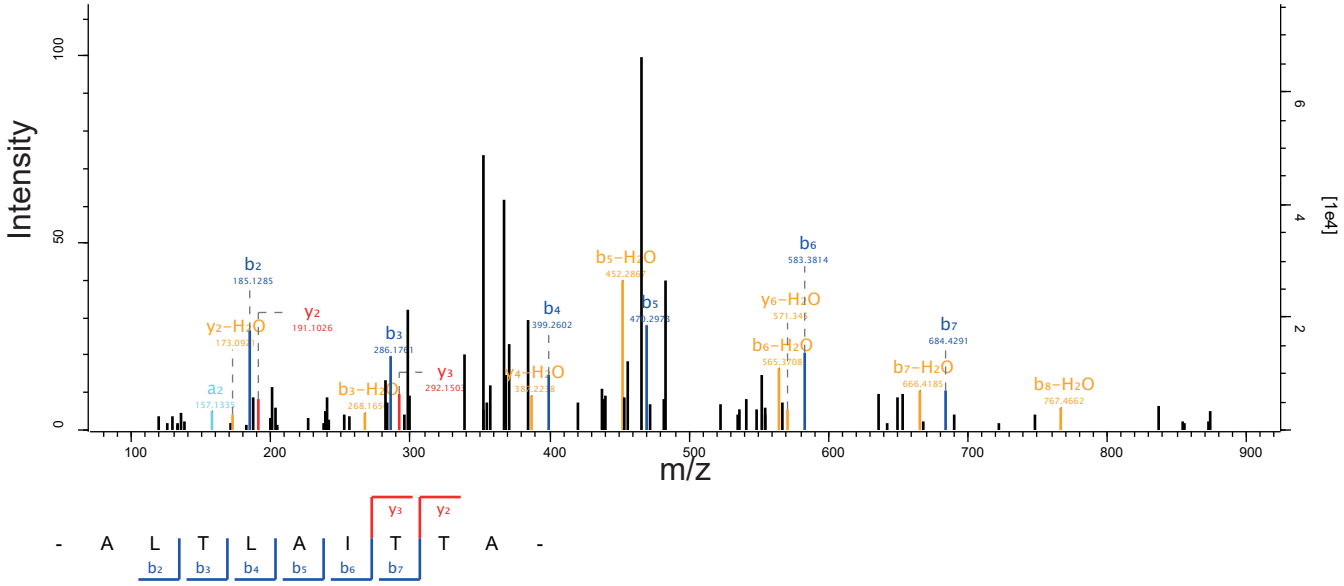
Endogenous/Synthetic peptide correlation



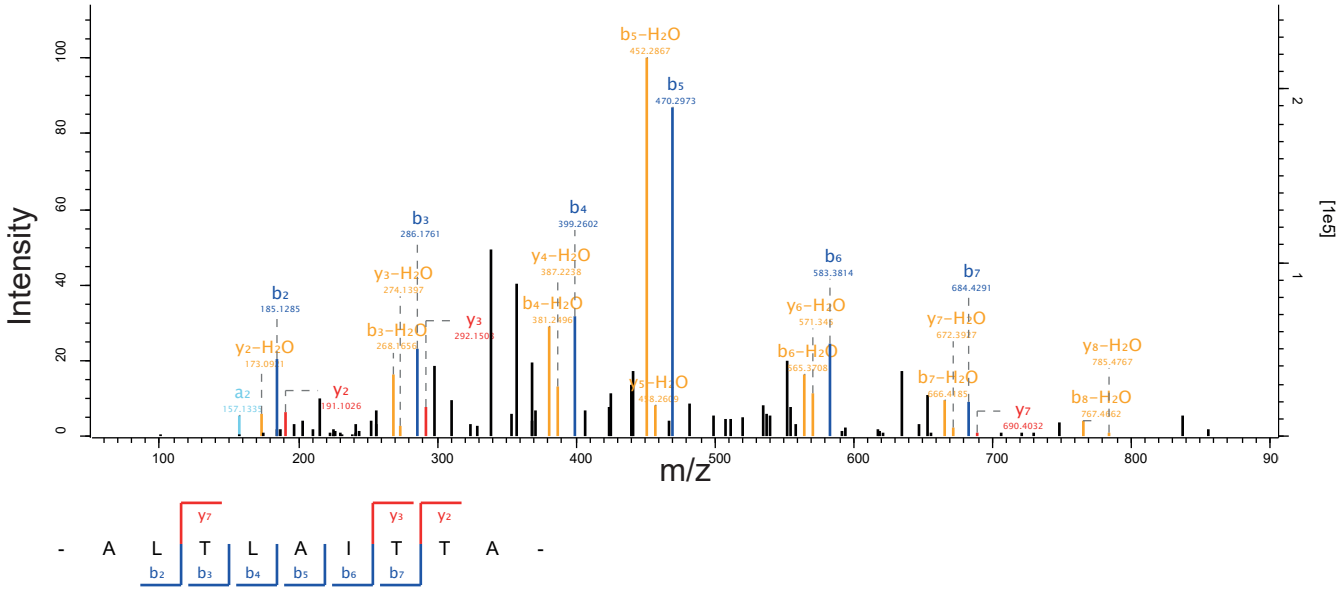
Dot product score=0.818
Pearson correlation=0.817

51BR-2 Fusobacterium nucleatim HLA-I ALTLAITTA

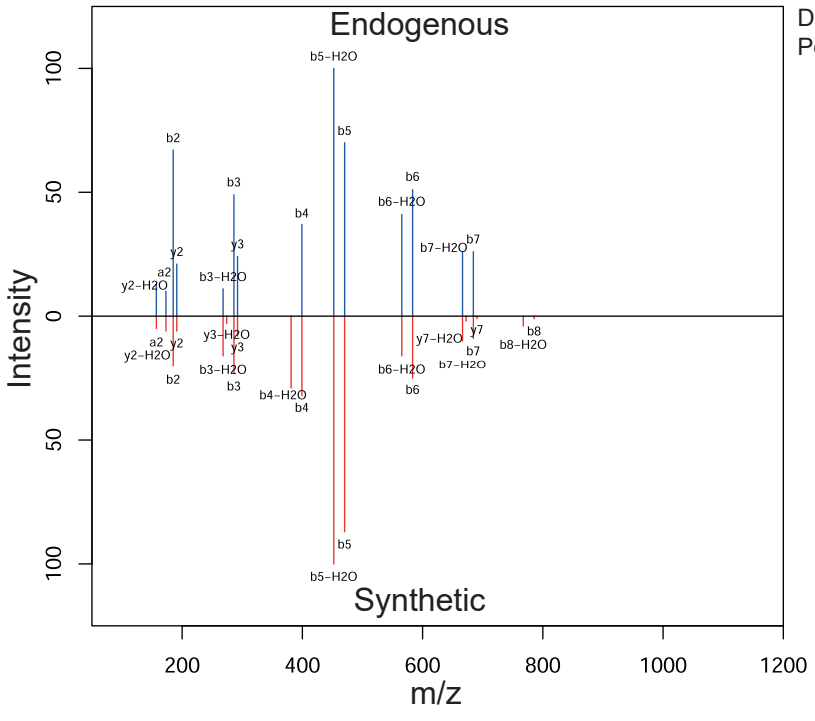
Endogenous peptide



Synthetic peptide



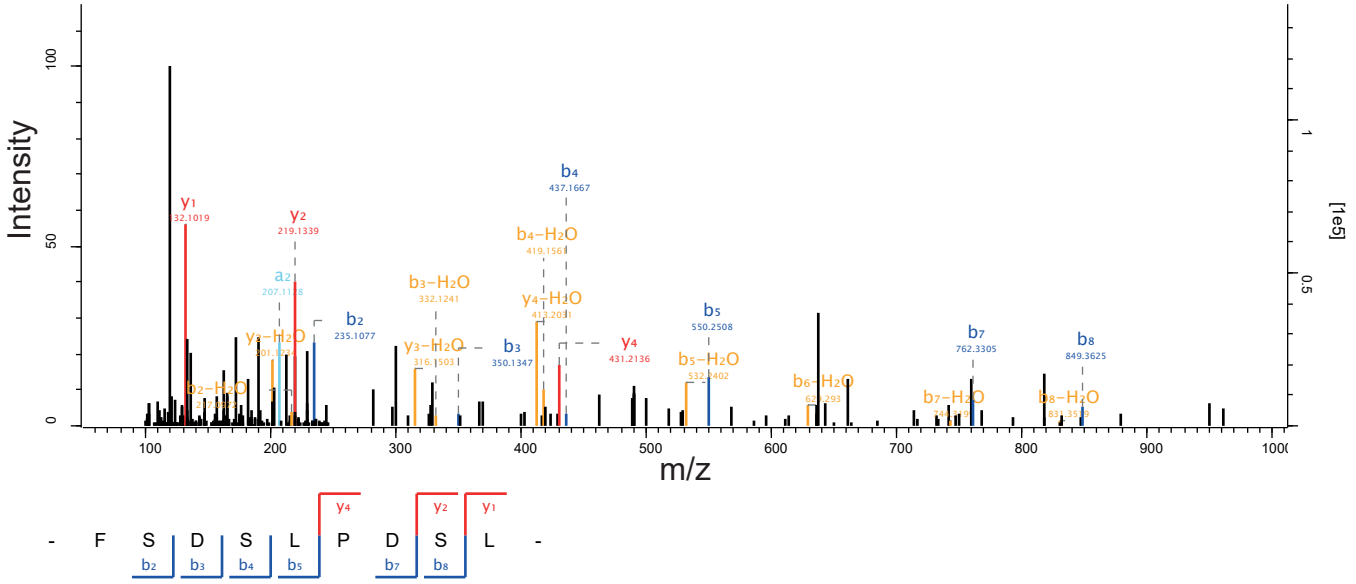
Endogenous/Synthetic peptide correlation



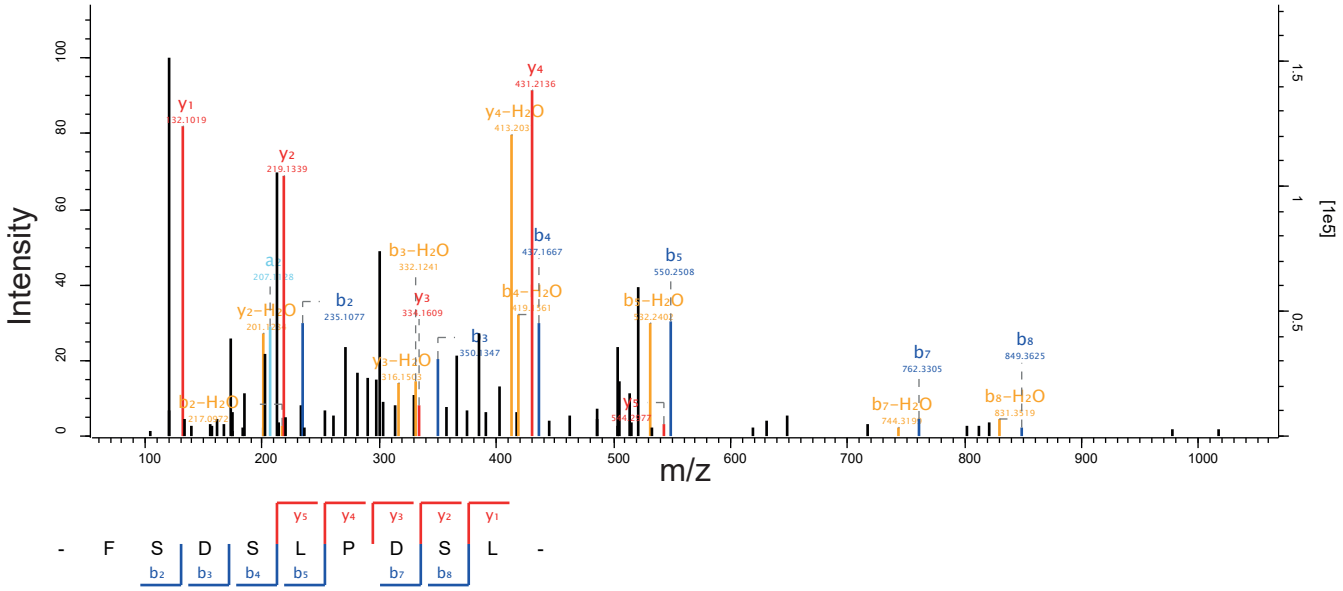
Dot product score=0.862
Pearson correlation=0.861

51BR-2 Fusobacterium nucleatim HLA-I FSDSLPDSL

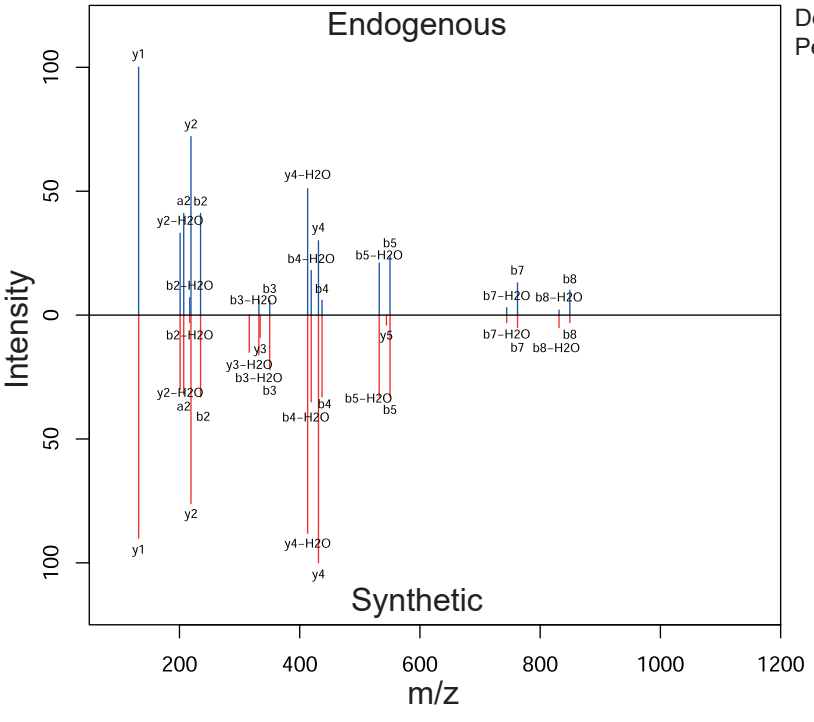
Endogenous peptide



Synthetic peptide



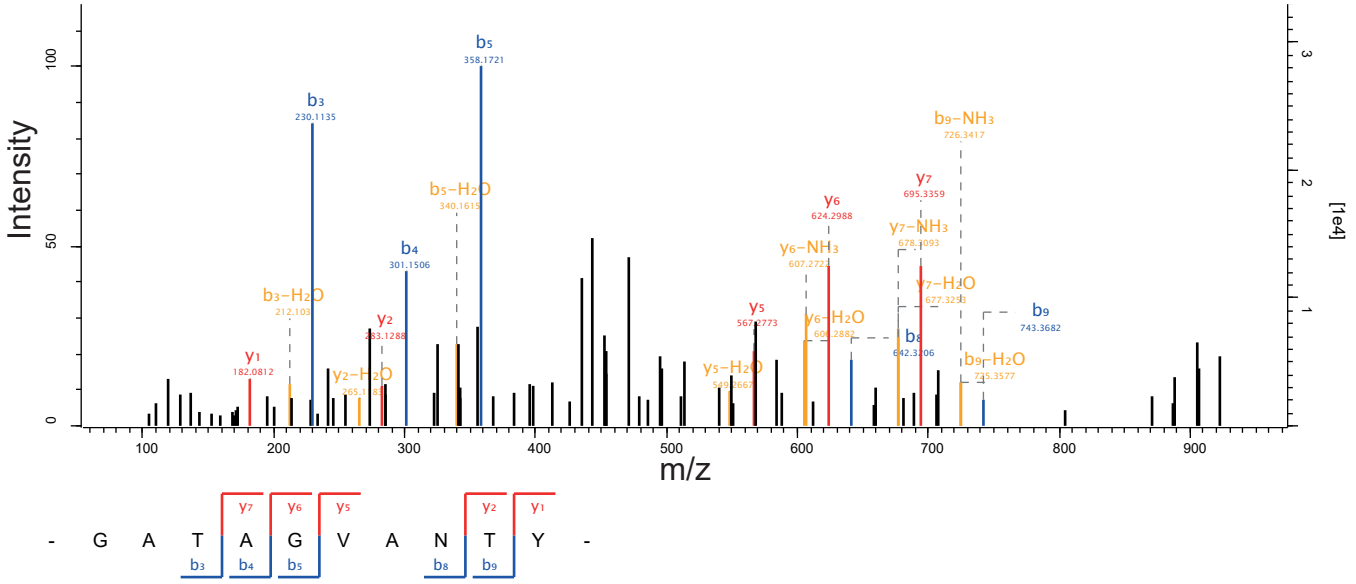
Endogenous/Synthetic peptide correlation



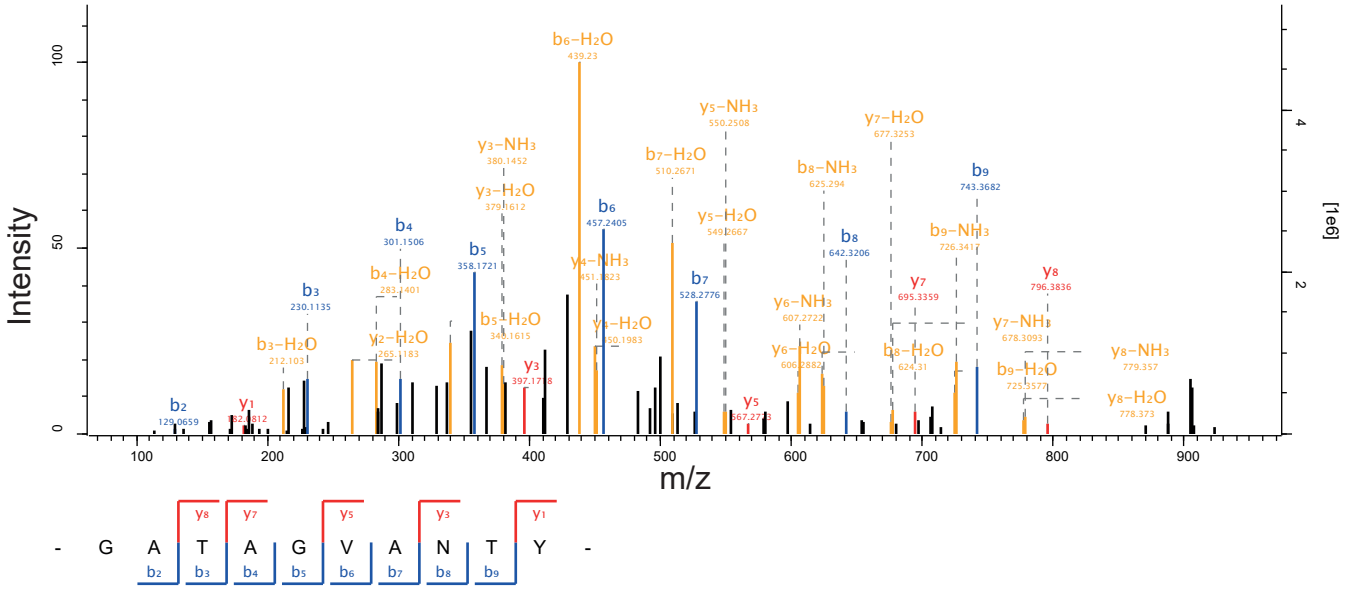
Dot product score=0.880
Pearson correlation=0.879

51BR-2 Fusobacterium nucleatim HLA-I GATAGVANTY

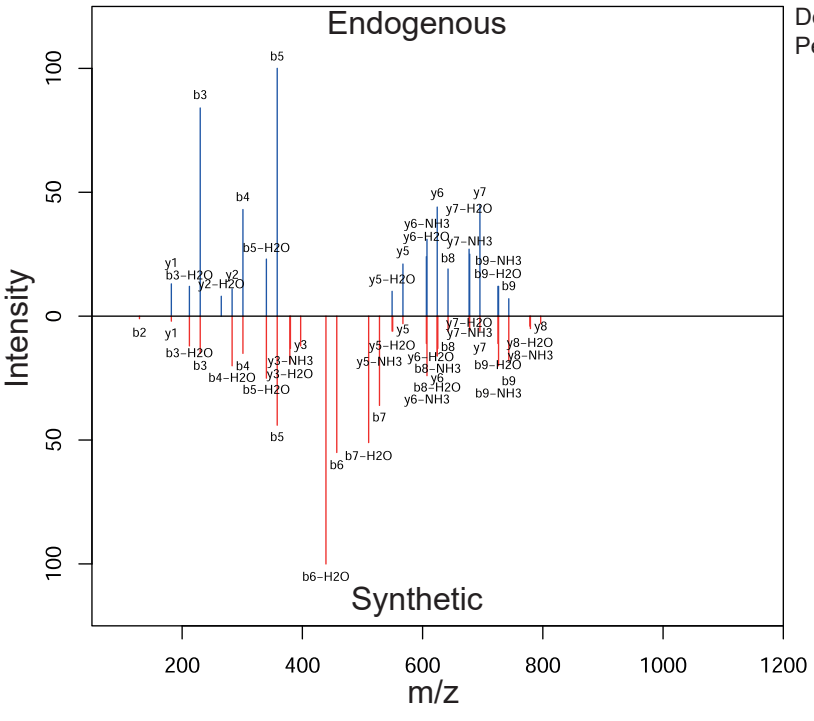
Endogenous peptide



Synthetic peptide



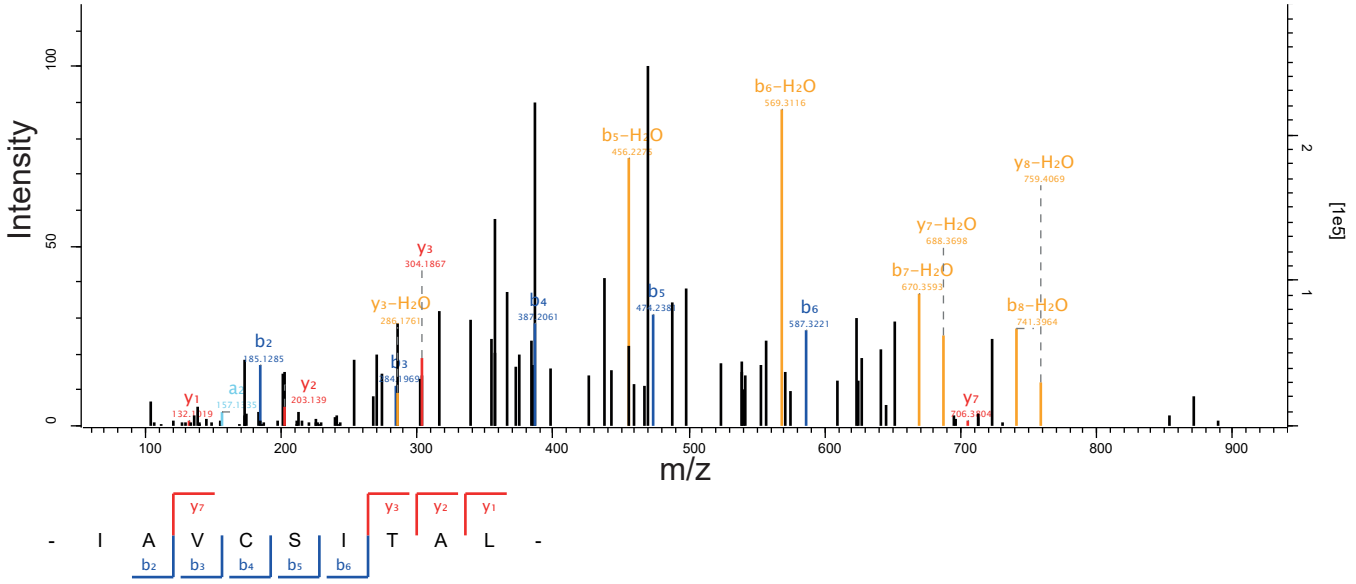
Endogenous/Synthetic peptide correlation



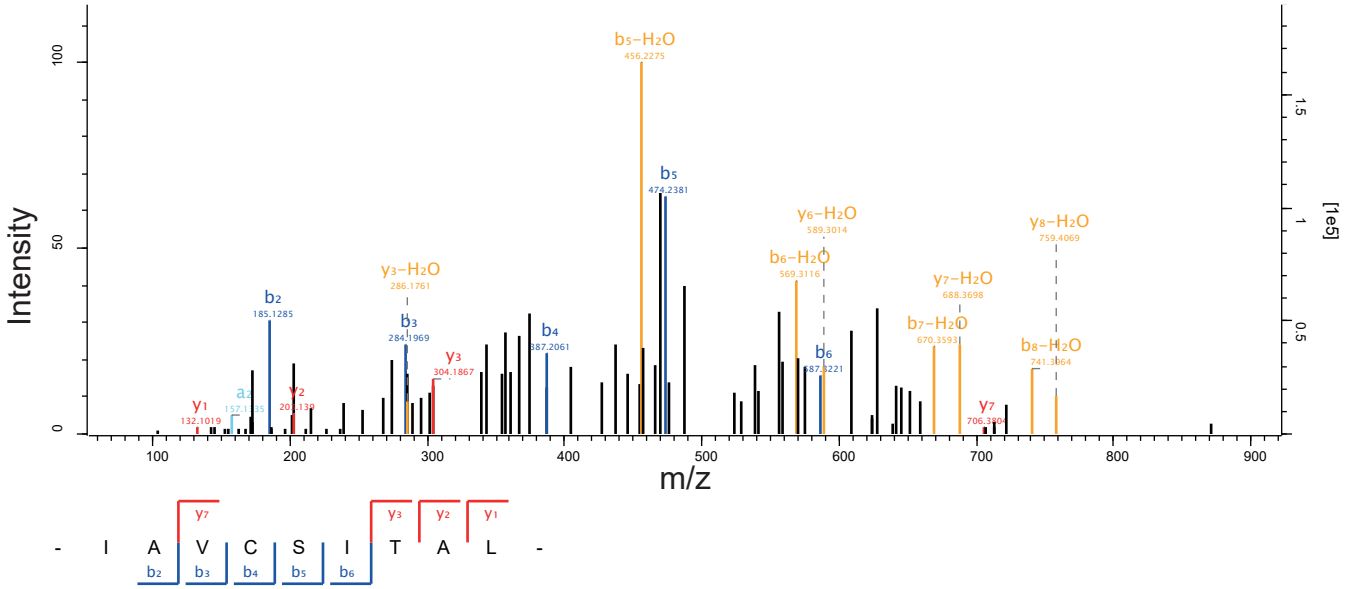
Dot product score=0.891
Pearson correlation=0.890

51BR-2 Fusobacterium nucleatim HLA-I IAVCSITAL

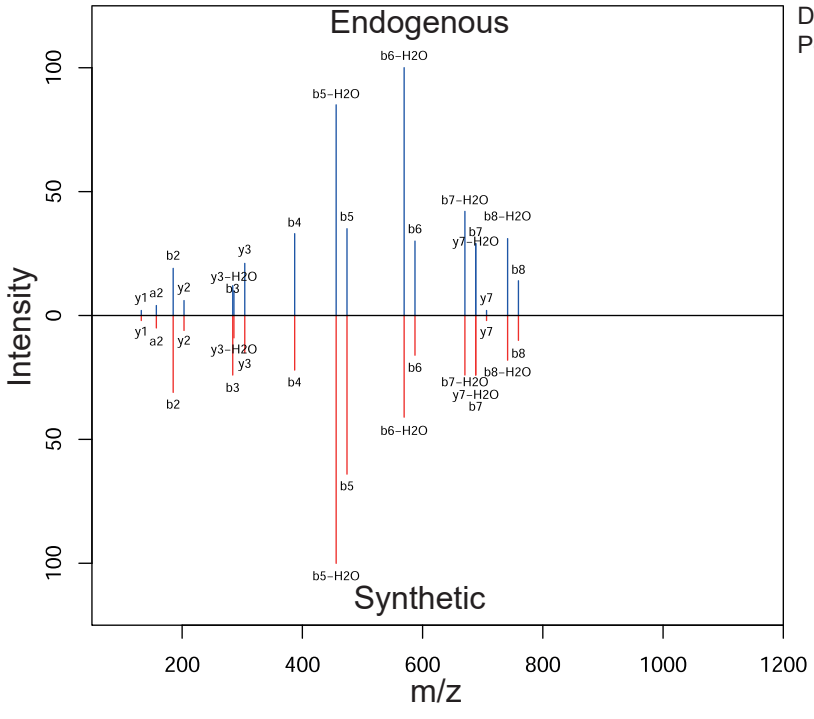
Endogenous peptide



Synthetic peptide



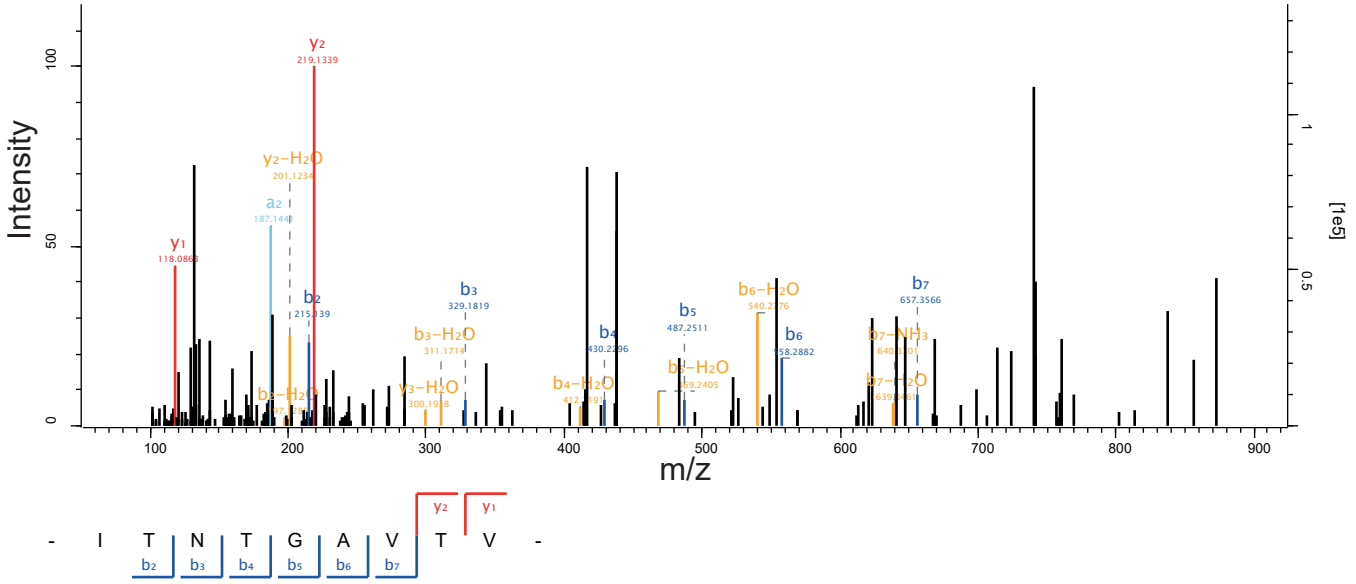
Endogenous/Synthetic peptide correlation



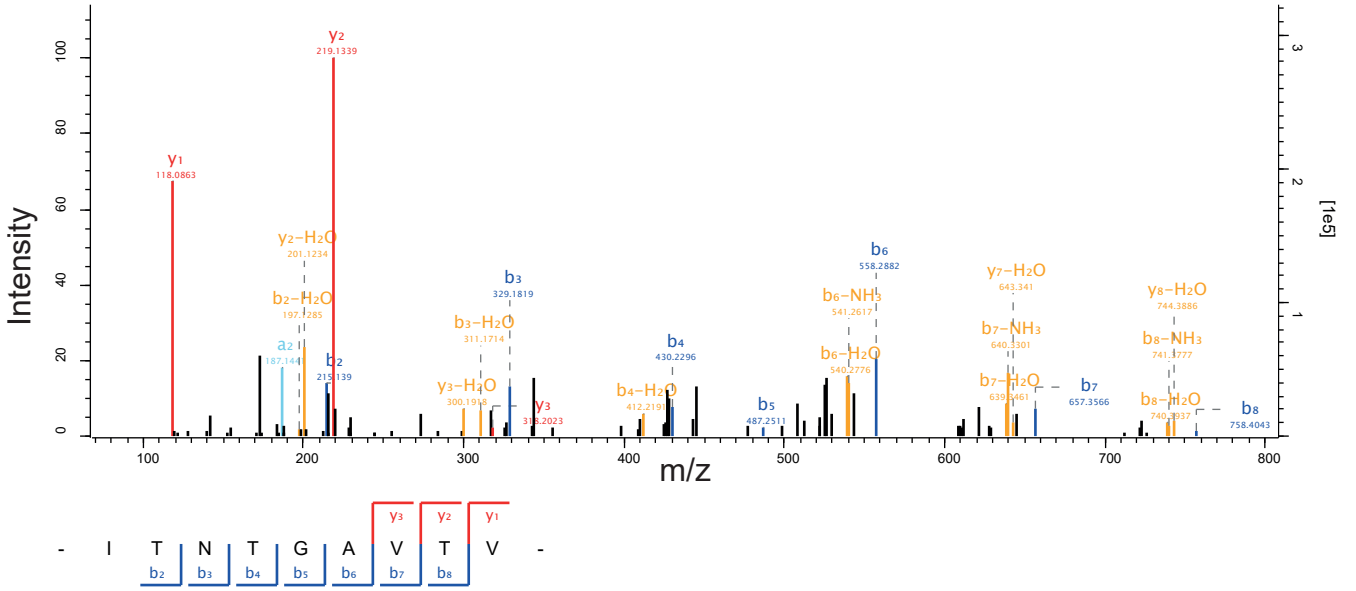
Dot product score=0.893
Pearson correlation=0.892

51BR-2 Fusobacterium nucleatim HLA-I ITNTGAVTV

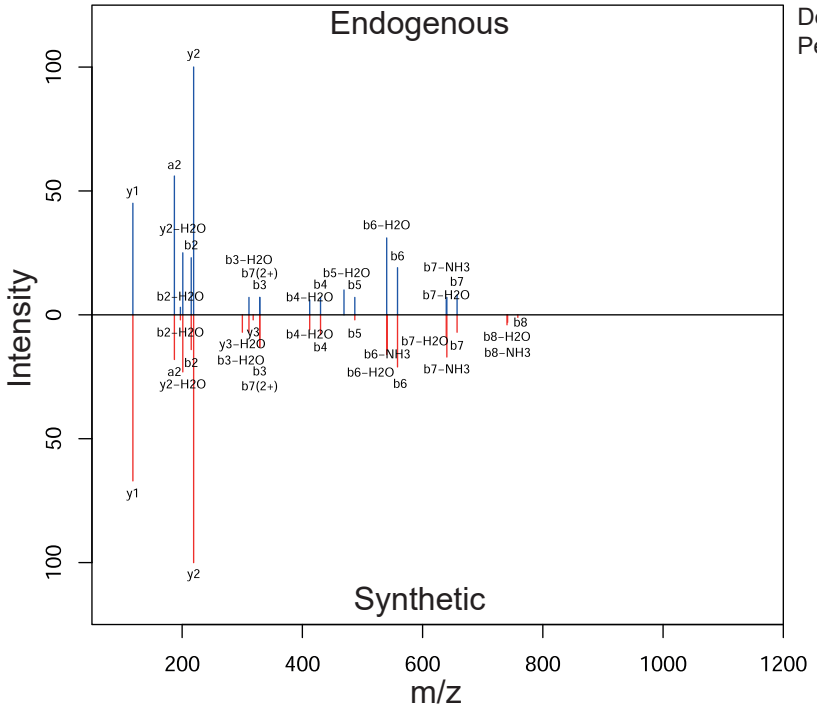
Endogenous peptide



Synthetic peptide



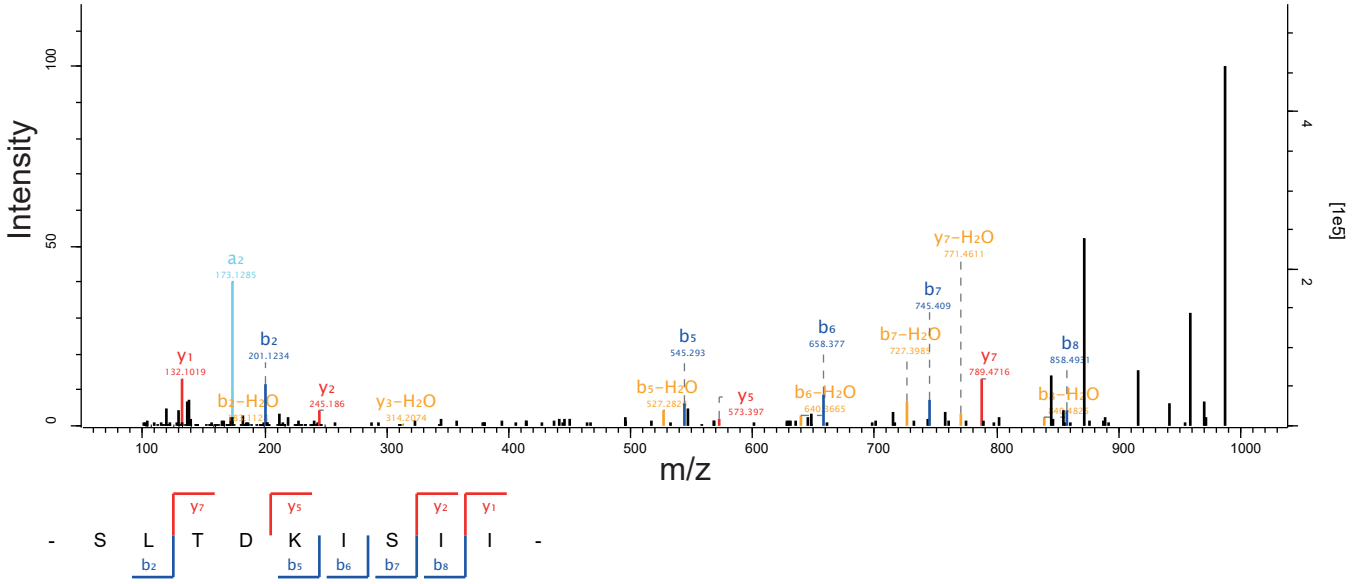
Endogenous/Synthetic peptide correlation



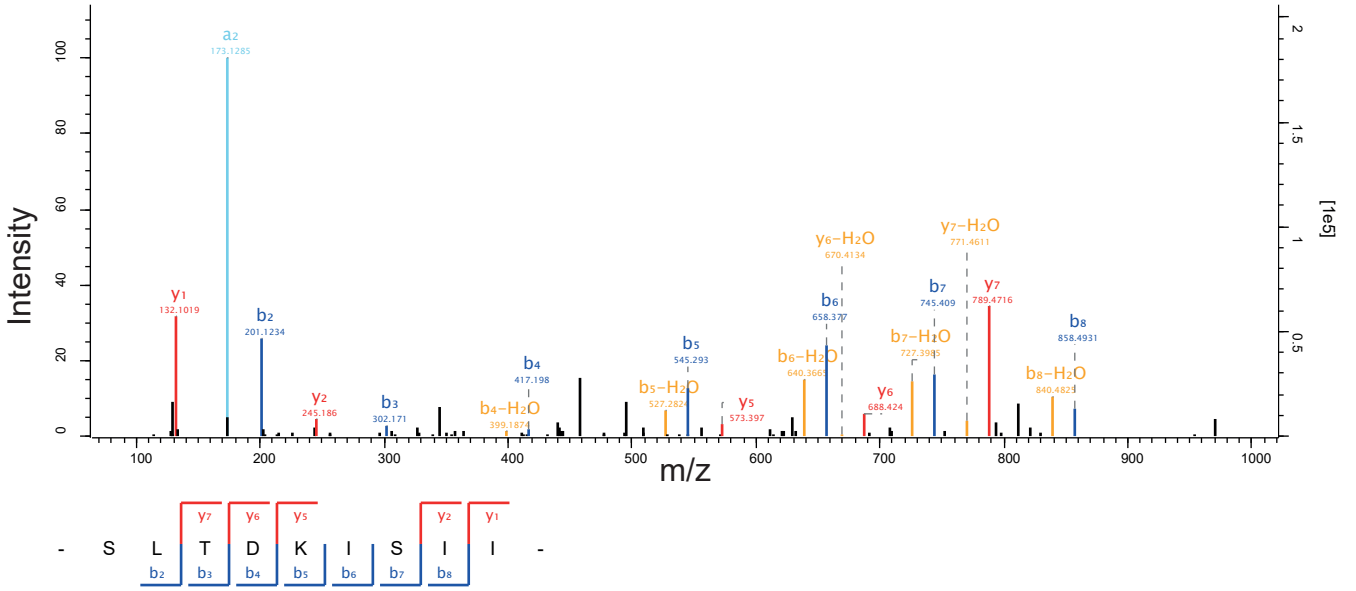
Dot product score=0.977
Pearson correlation=0.977

51BR-2 Fusobacterium nucleatim HLA-I SLTDKISII

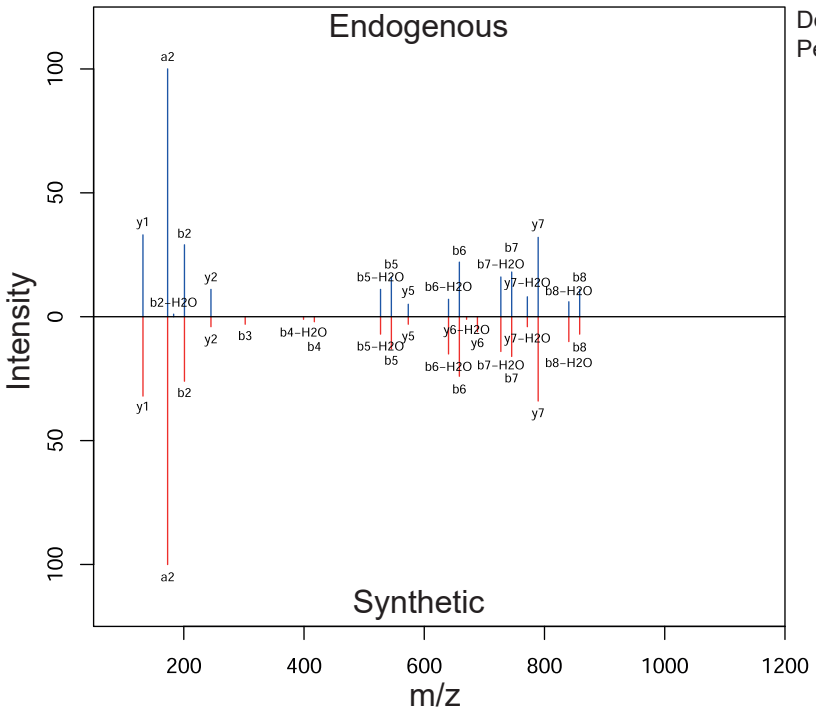
Endogenous peptide



Synthetic peptide



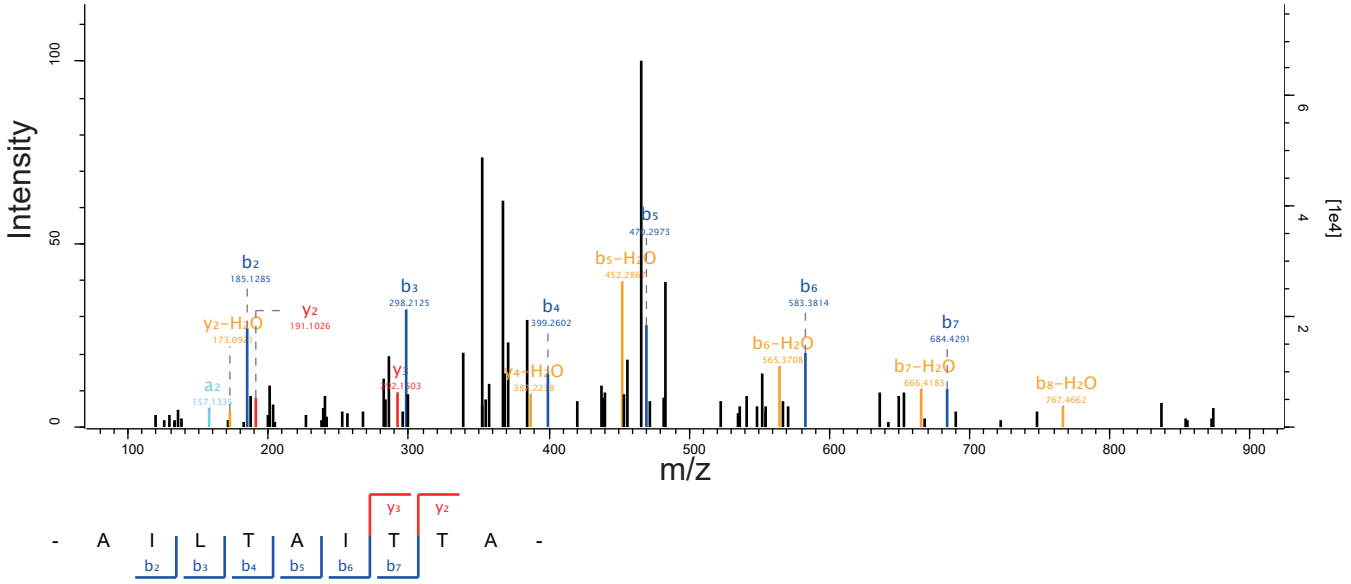
Endogenous/Synthetic peptide correlation



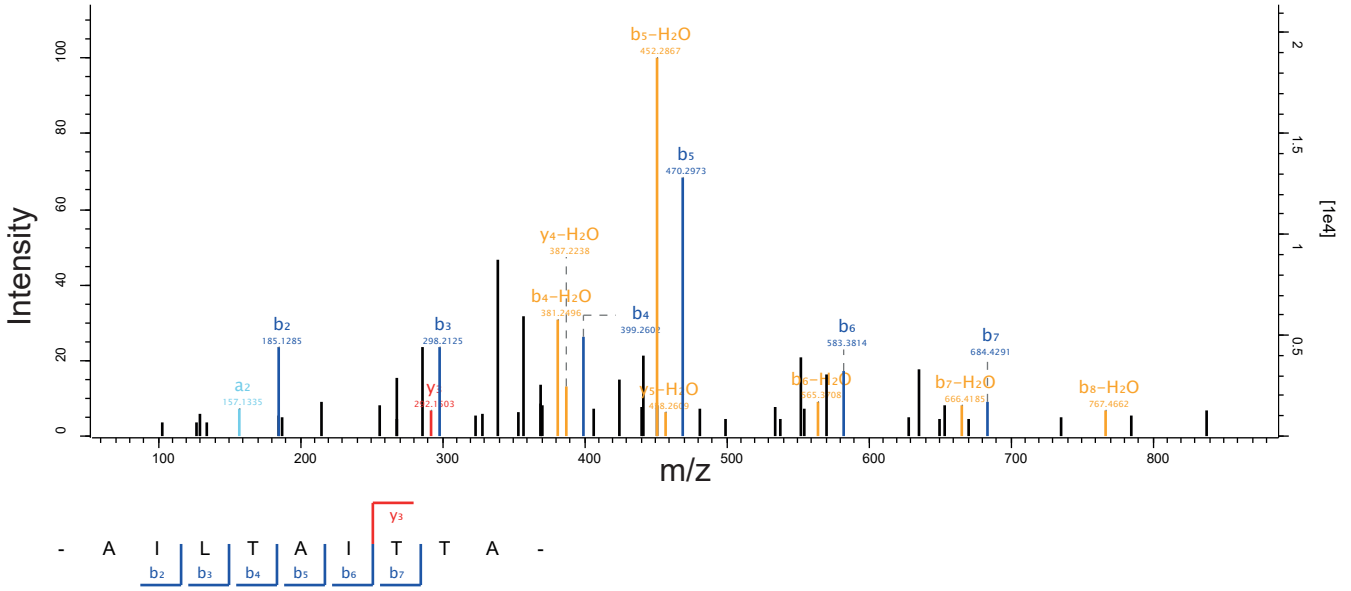
Dot product score=0.989
Pearson correlation=0.989

51BR-2 Staphylococcus aureus HLA-I AILTAITTA

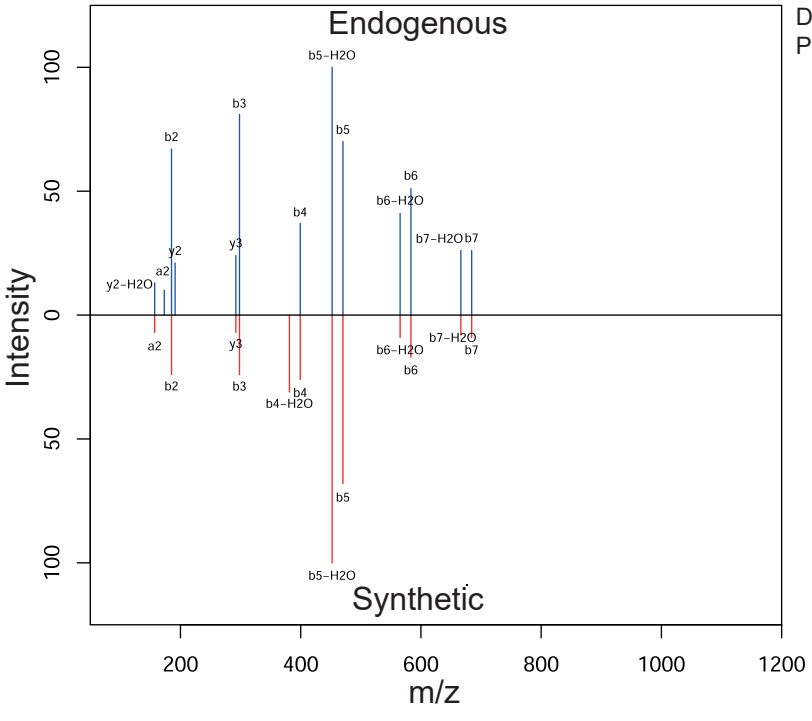
Endogenous peptide



Synthetic peptide



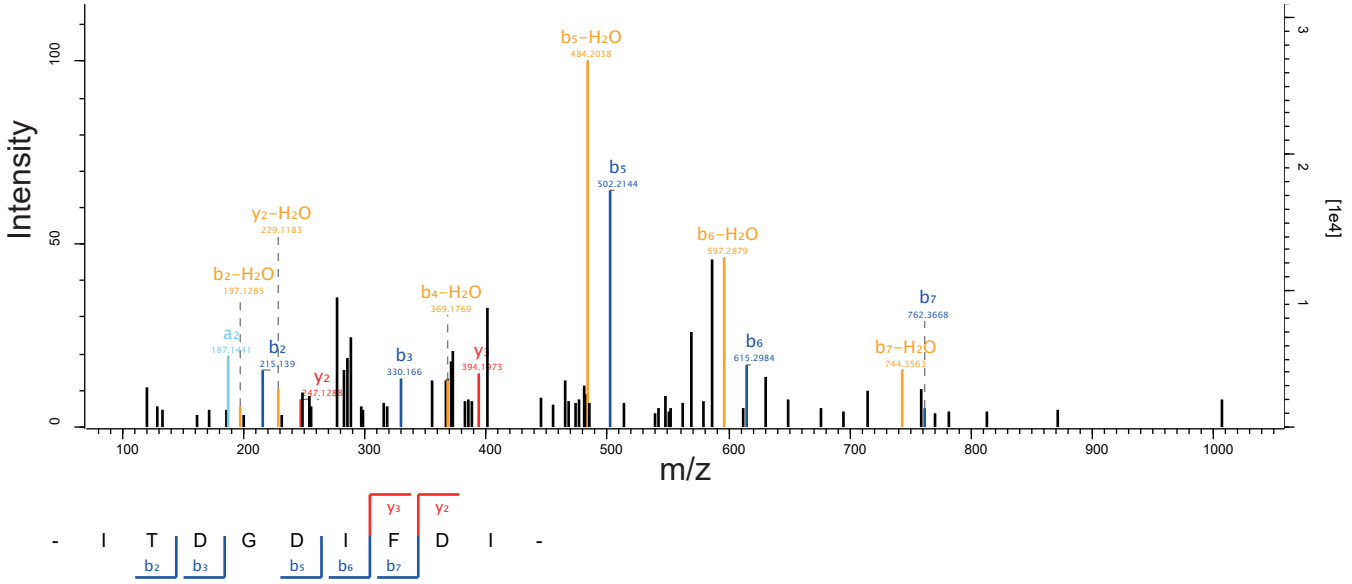
Endogenous/Synthetic peptide correlation



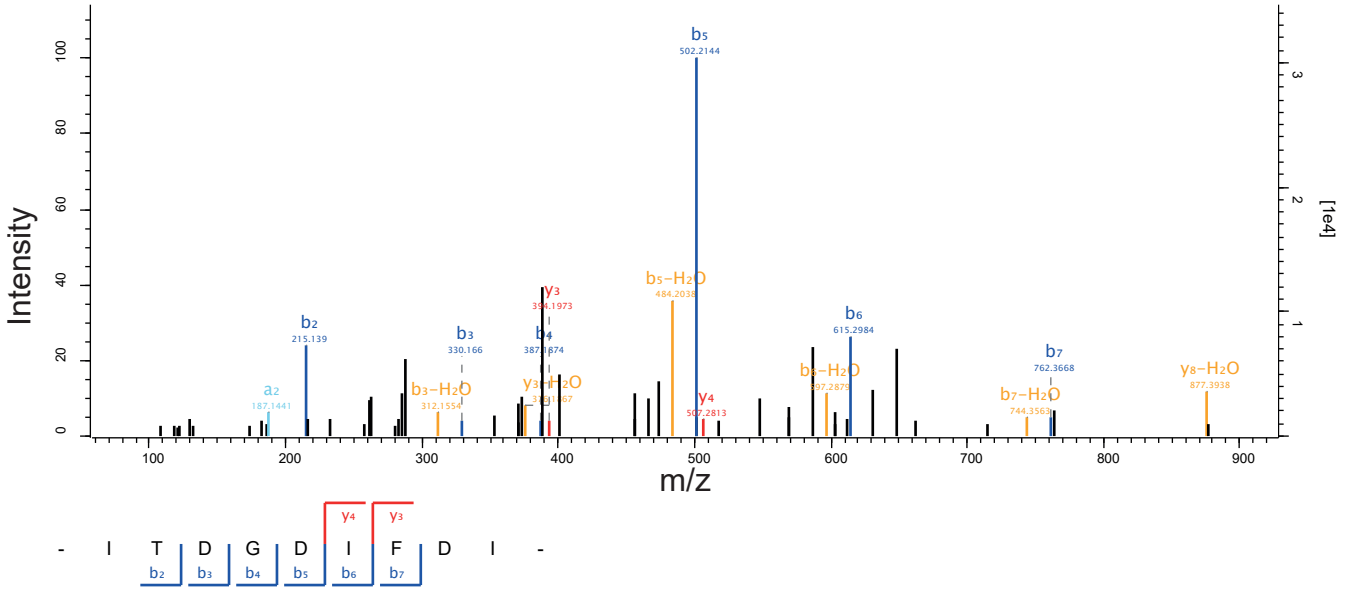
Dot product score=0.870
Pearson correlation=0.869

51BR-2 Staphylococcus aureus HLA-I ITDGDIFDI

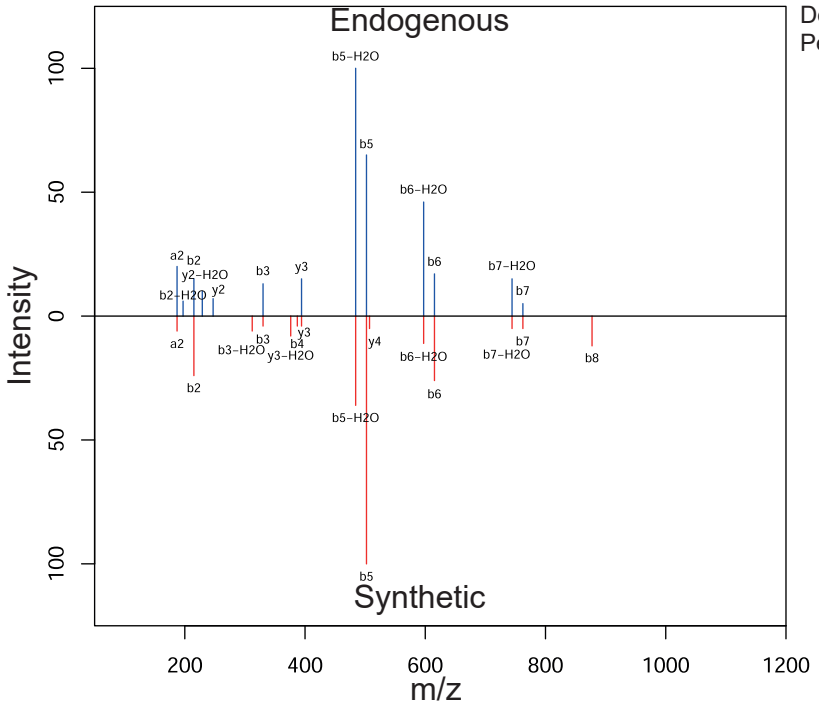
Endogenous peptide



Synthetic peptide



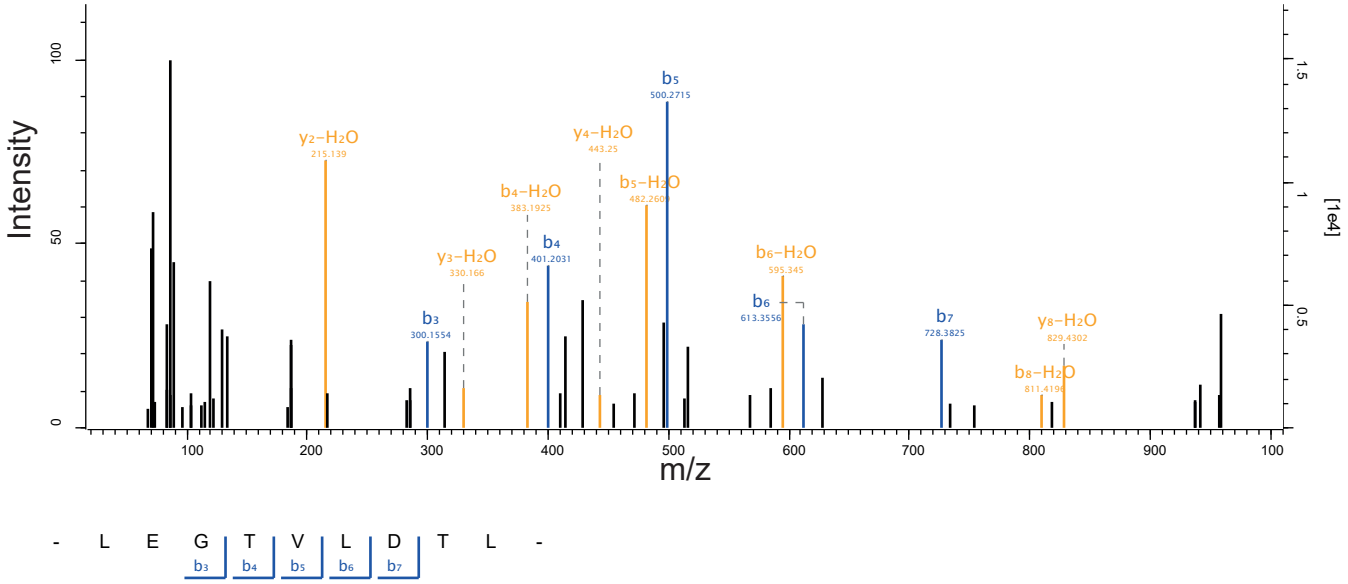
Endogenous/Synthetic peptide correlation



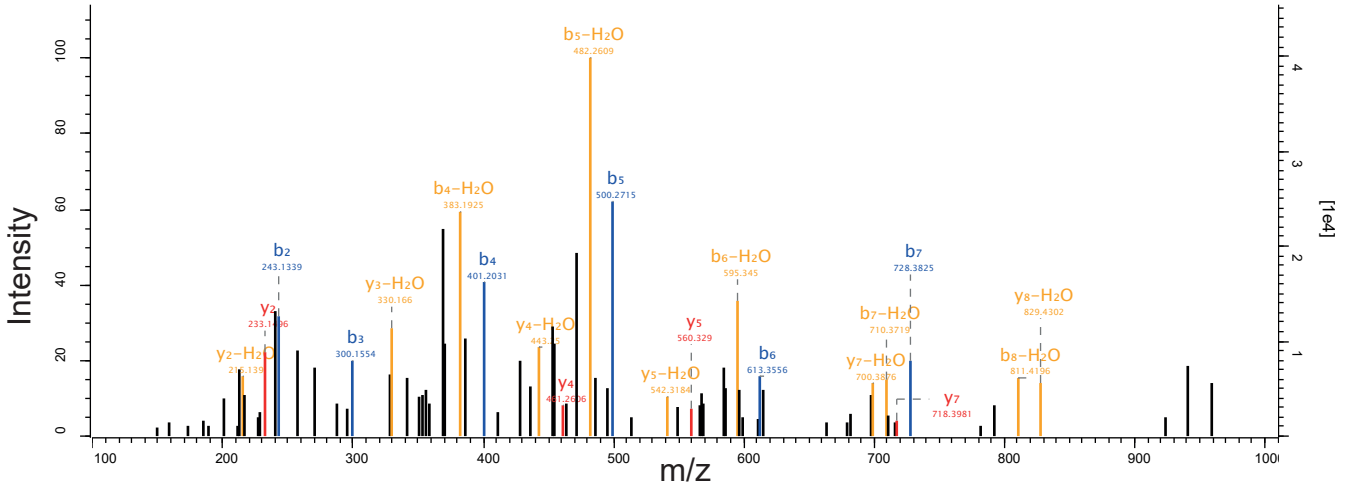
Dot product score=0.974
Pearson correlation=0.974

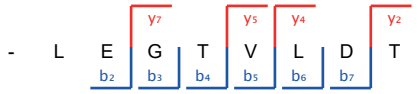
55A3 Staphylococcus capitis HLA-I LEGTVLDTL

Endogenous peptide

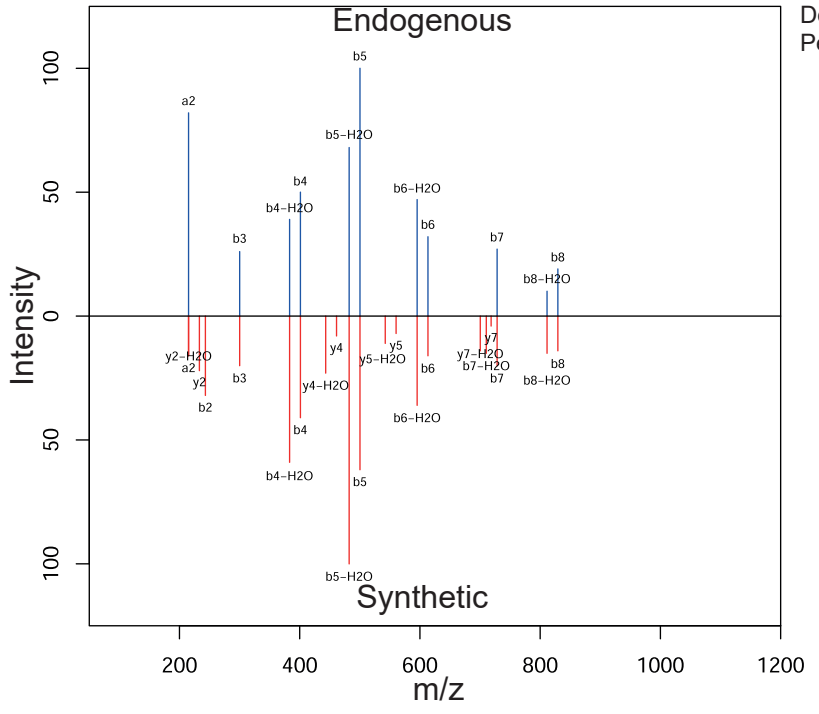


Synthetic peptide





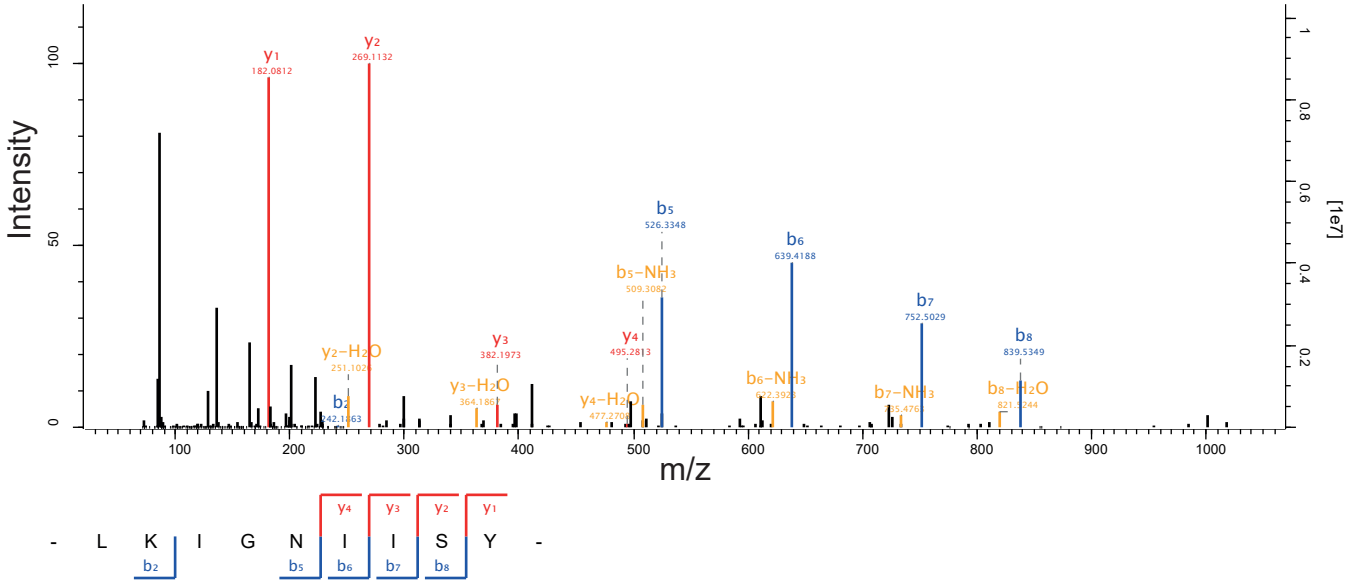
Endogenous/Synthetic peptide correlation



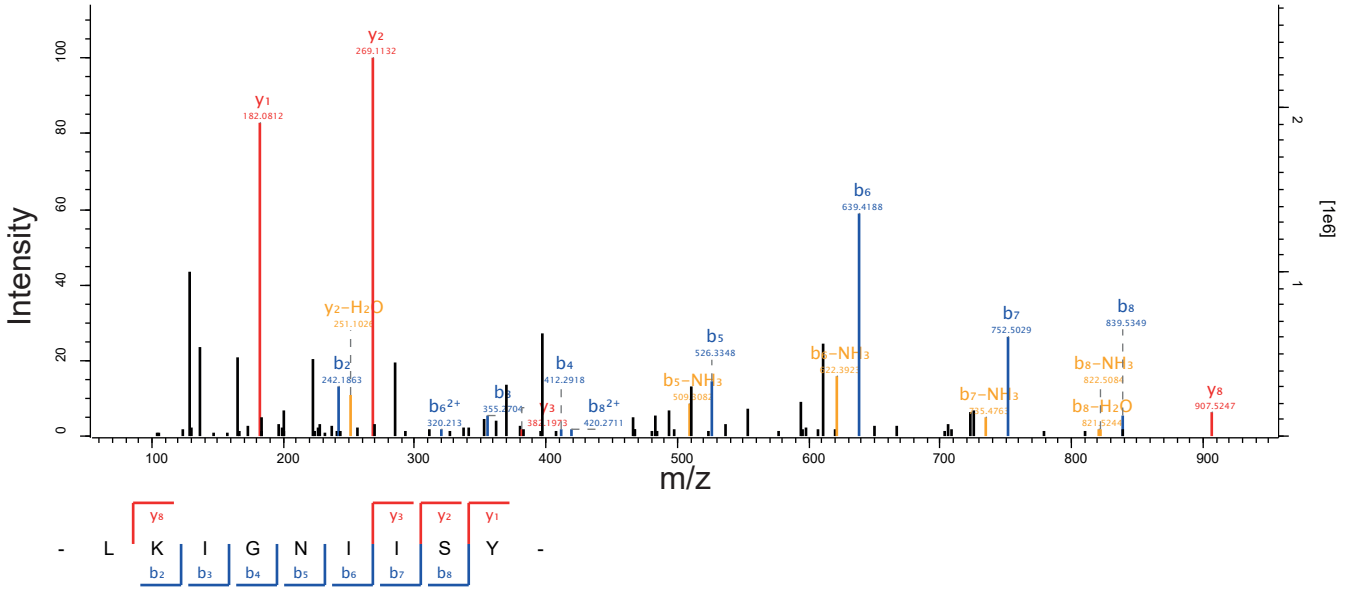
Dot product score=0.990
Pearson correlation=0.990

55A7-1 Clostridium clostridioforme HLA-I LKIGNIISY

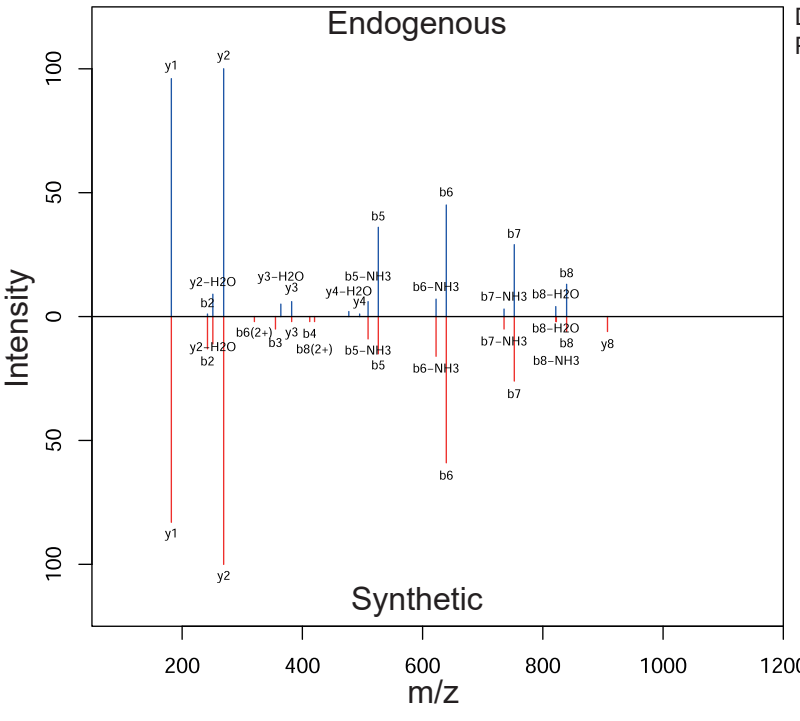
Endogenous peptide



Synthetic peptide



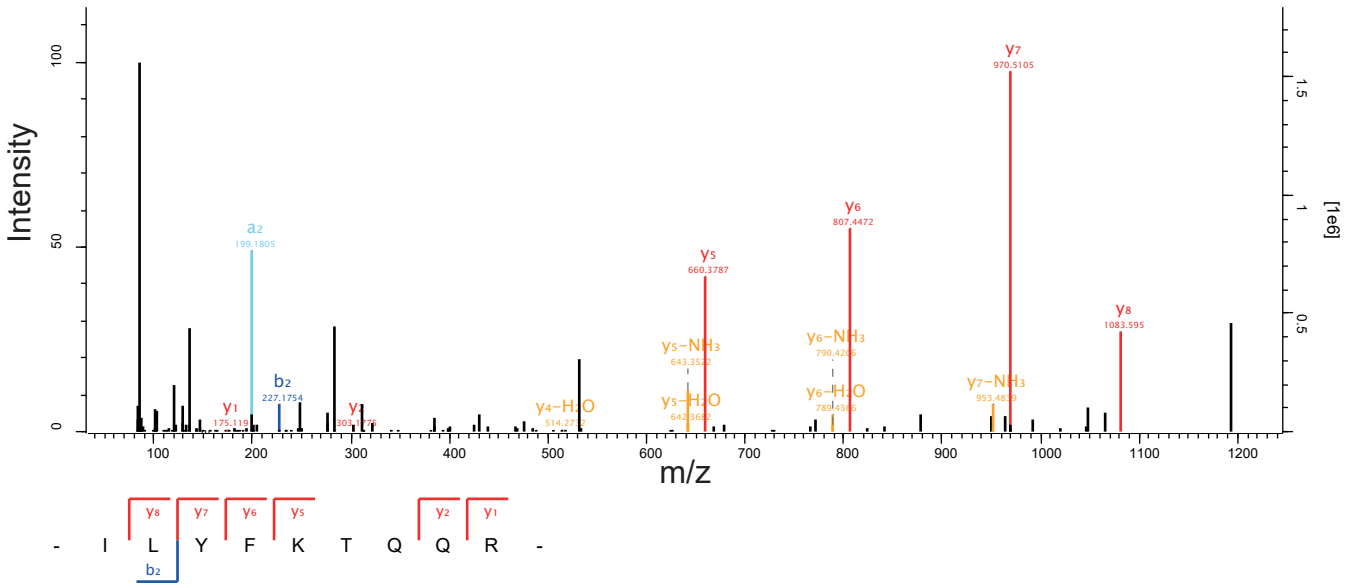
Endogenous/Synthetic peptide correlation



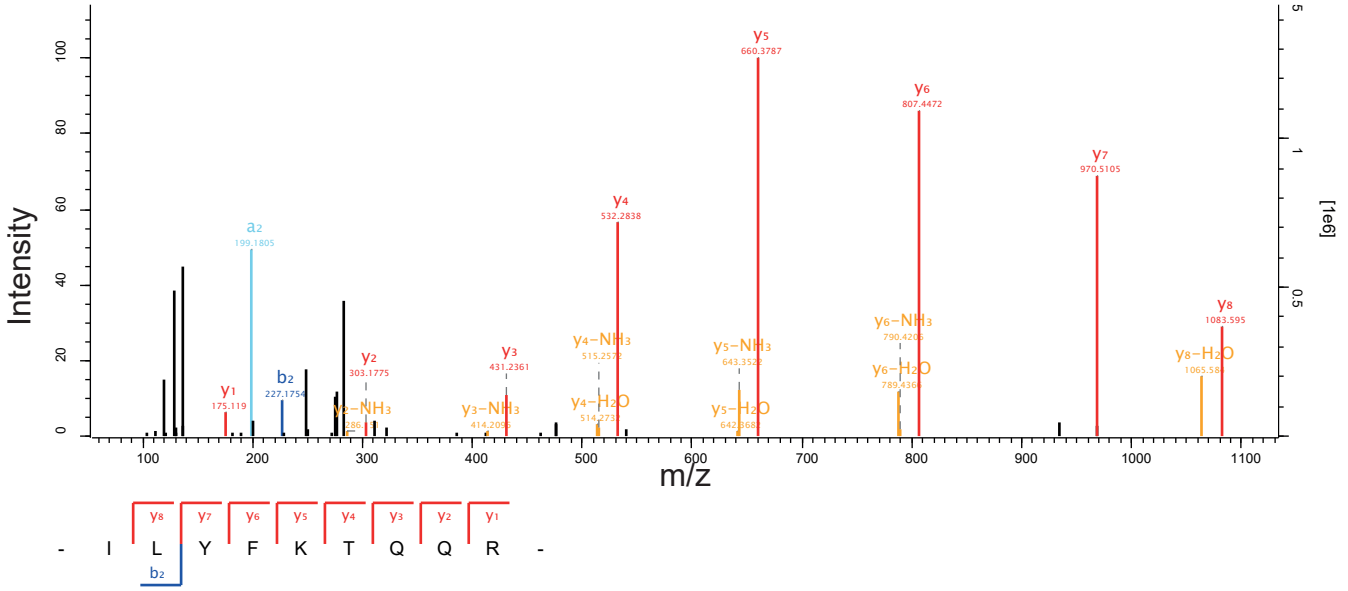
Dot product score=0.978
Pearson correlation=0.977

55A7-1 Dialister pneumosintes HLA-I ILYFKTQQR

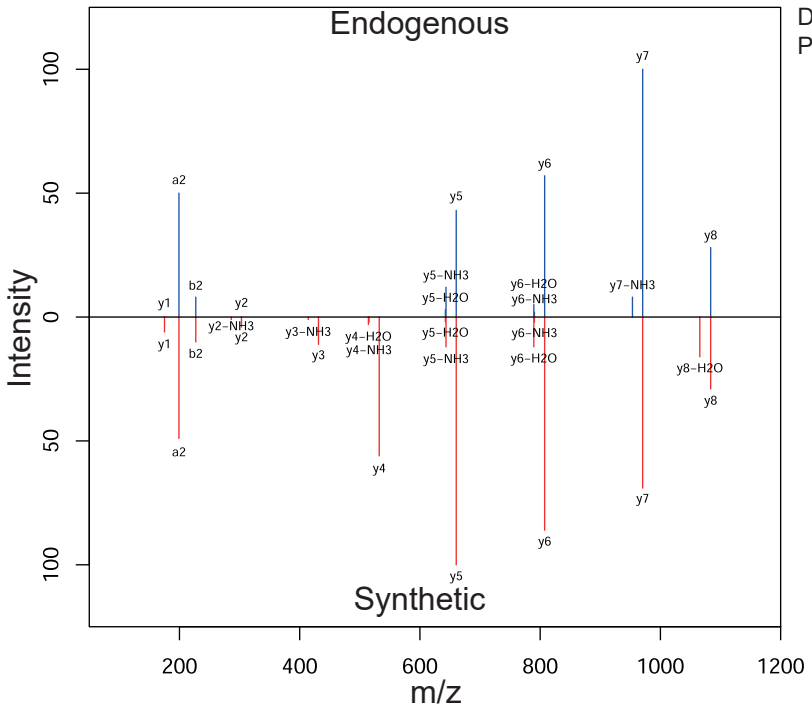
Endogenous peptide



Synthetic peptide



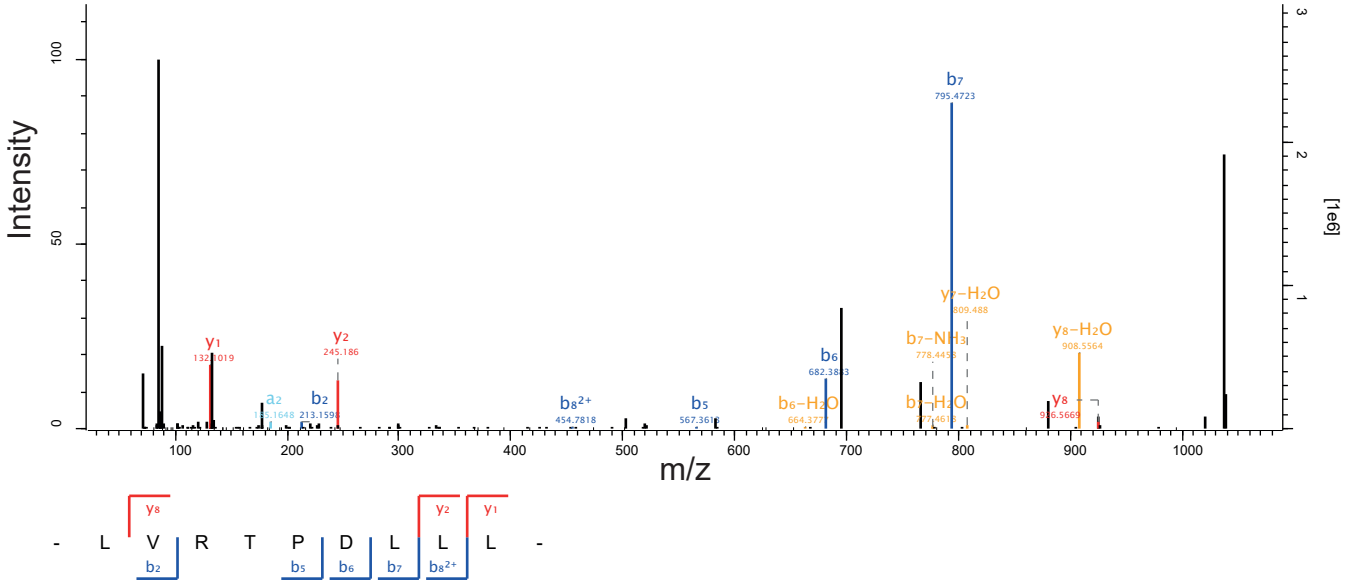
Endogenous/Synthetic peptide correlation



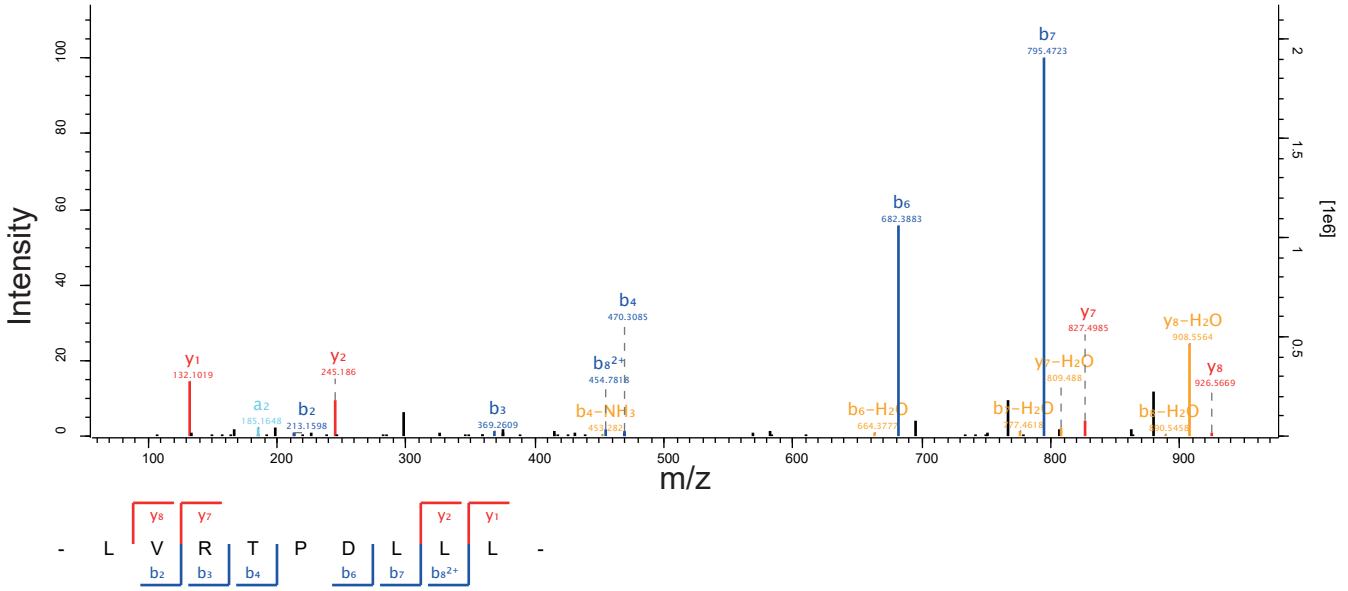
Dot product score=0.884
Pearson correlation=0.883

55A7-1 Dialister pneumosintes HLA-I LVRTPDLLL

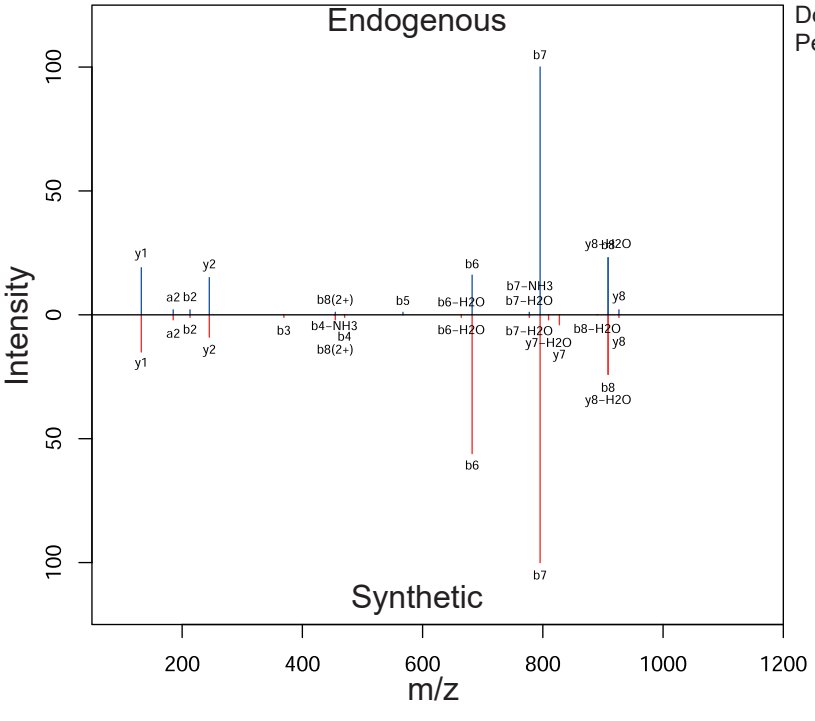
Endogenous peptide



Synthetic peptide



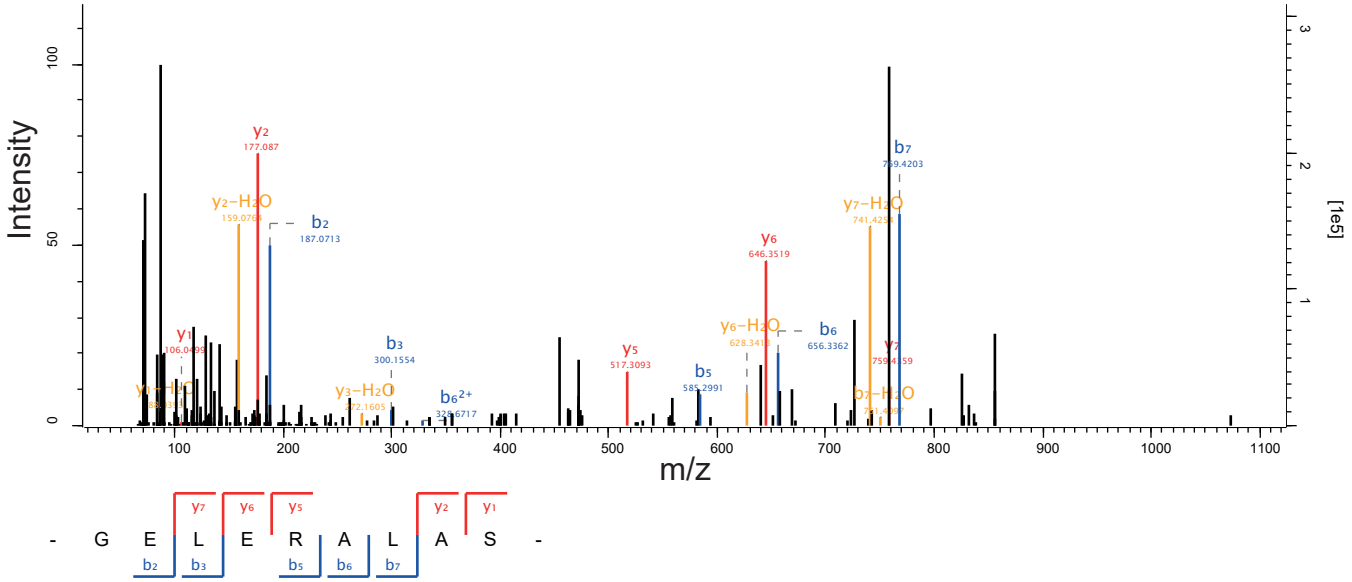
Endogenous/Synthetic peptide correlation



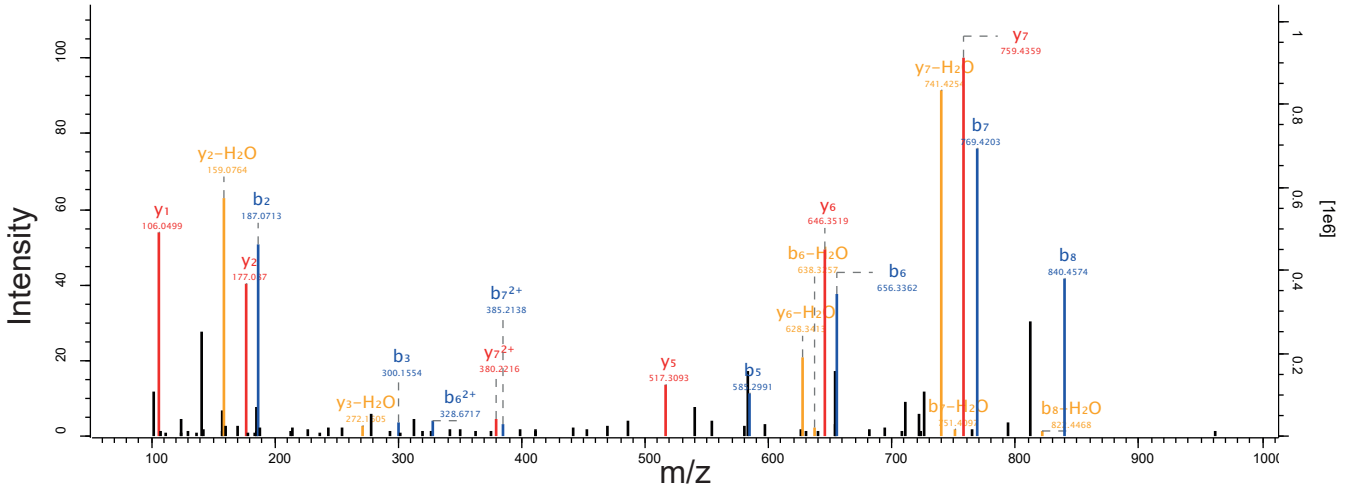
Dot product score=0.939
Pearson correlation=0.939

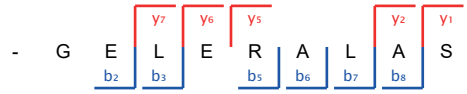
55A7-1 Enterobacter hormaechei HLA-I GELERALAS

Endogenous peptide

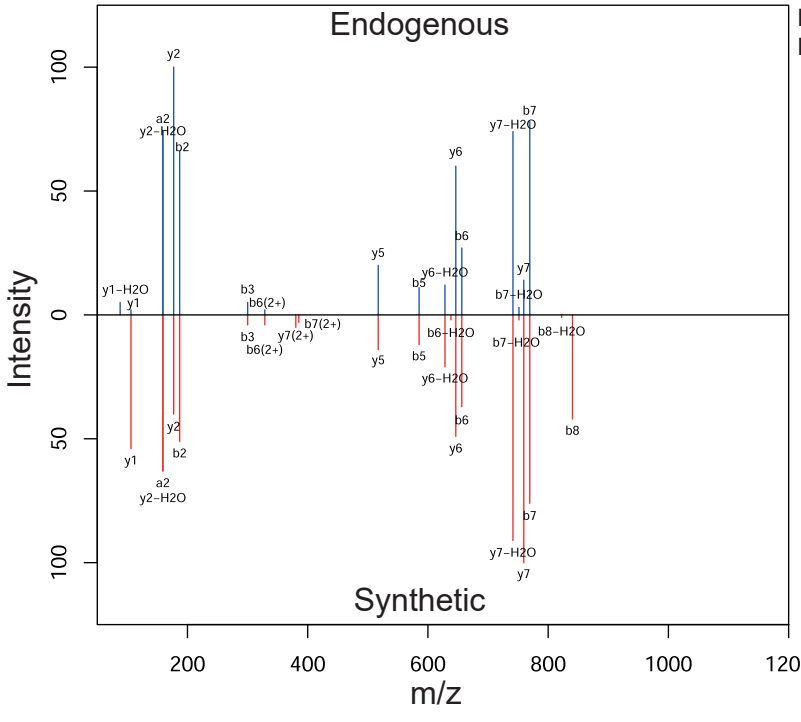


Synthetic peptide





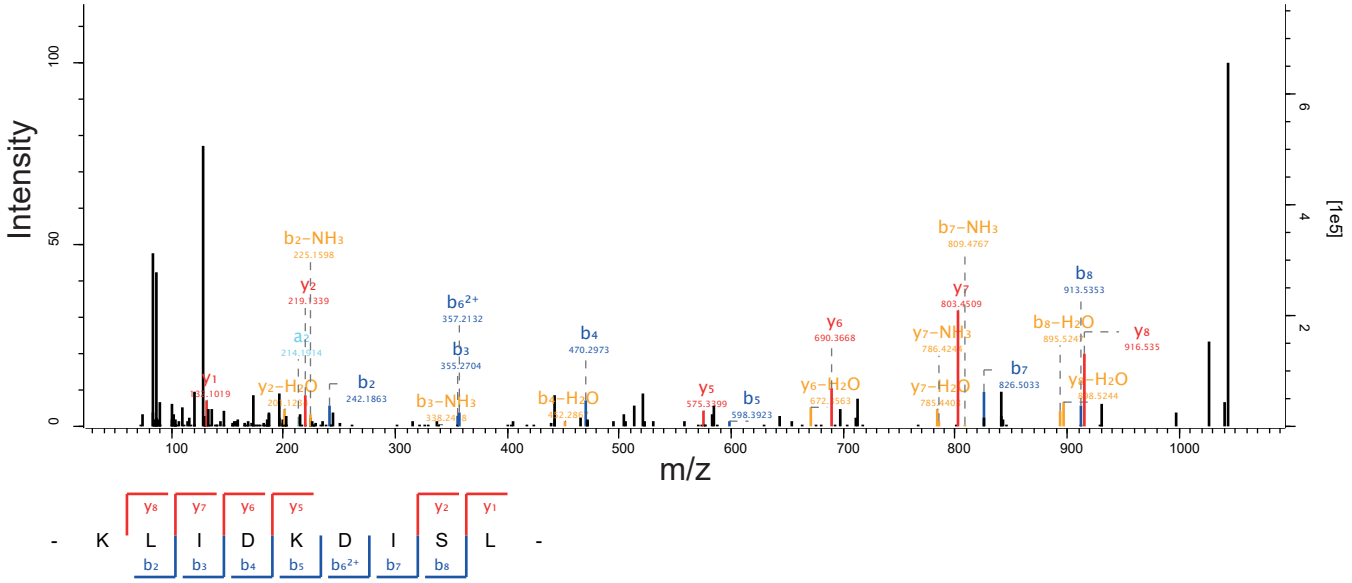
Endogenous/Synthetic peptide correlation



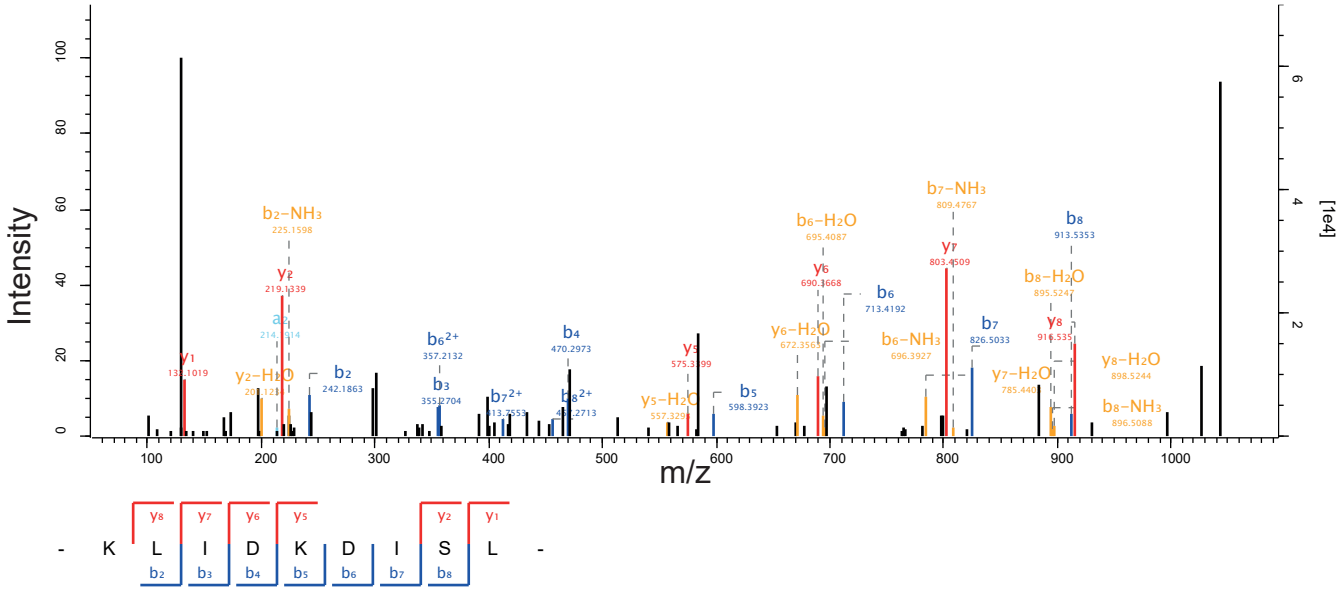
Dot product score=0.728
Pearson correlation=0.726

55A7-1 Veillonella parvula HLA-I KLIDKDISL

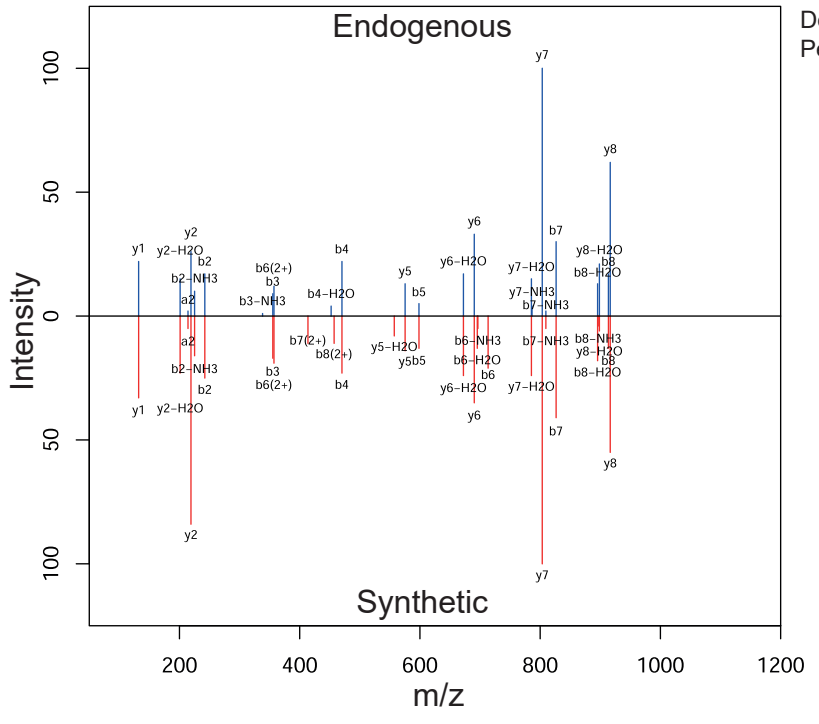
Endogenous peptide



Synthetic peptide



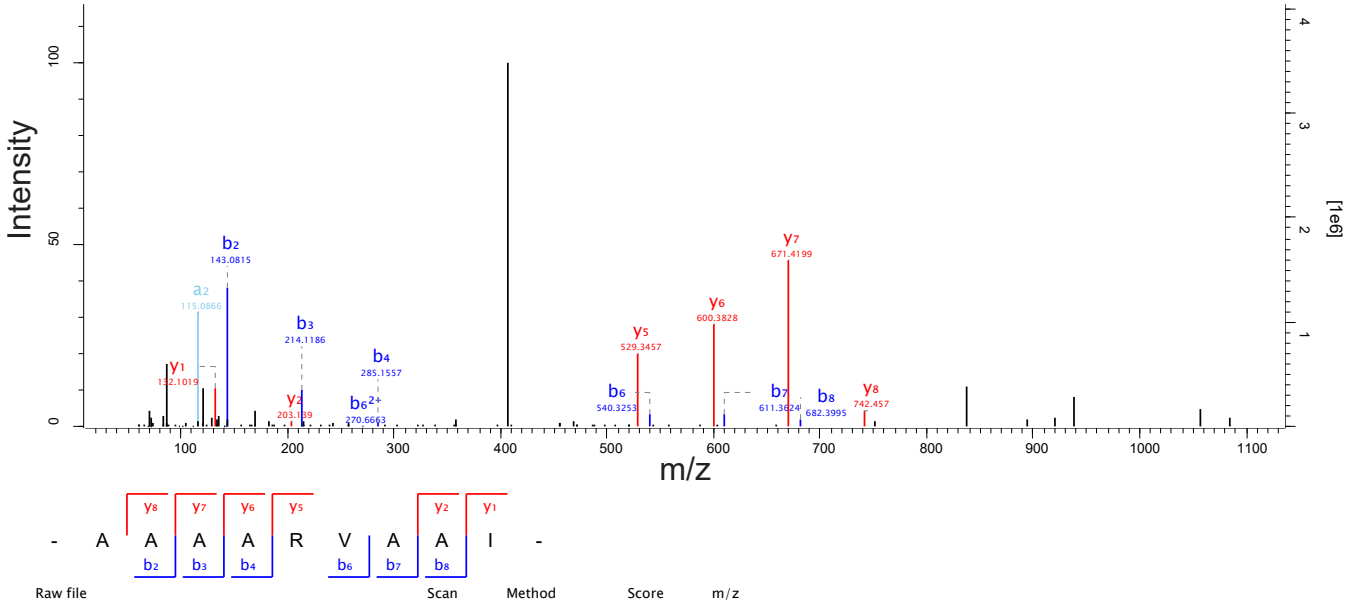
Endogenous/Synthetic peptide correlation



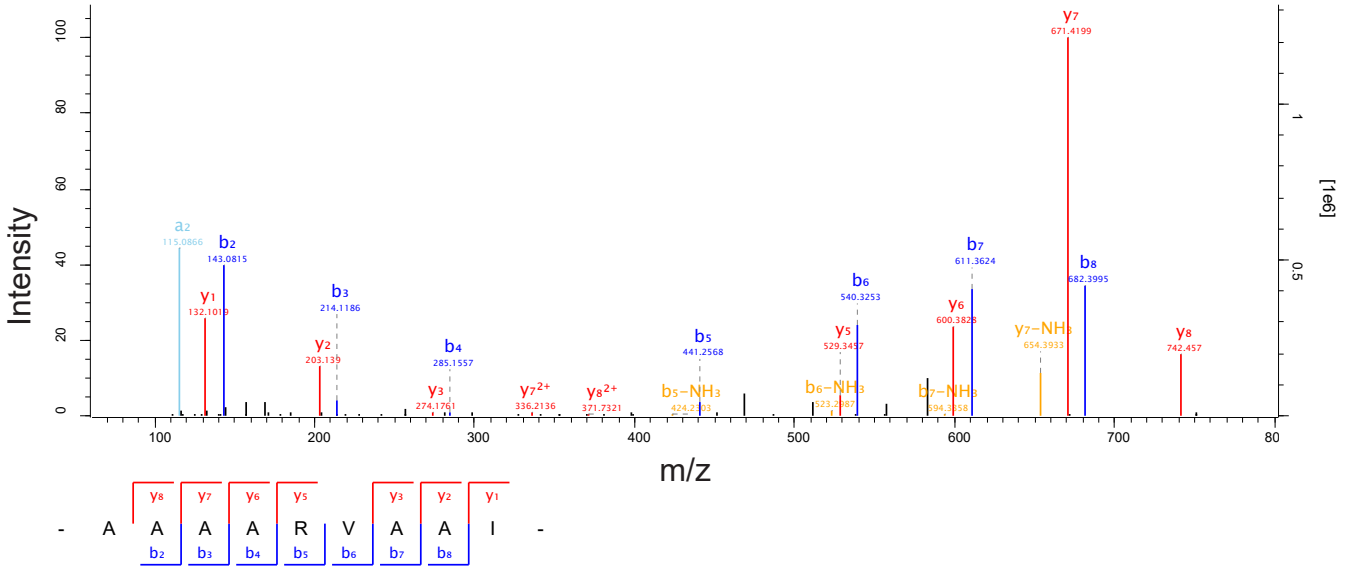
Dot product score=0.928
Pearson correlation=0.927

55B3 Sphingomonas dokdonensis HLA-I AAAARVAAI

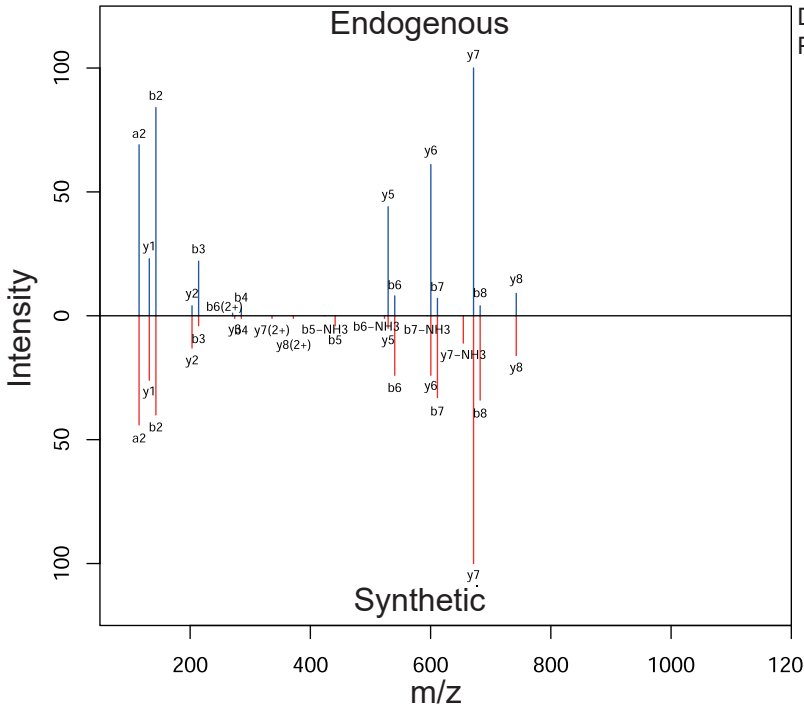
Endogenous peptide



Synthetic peptide

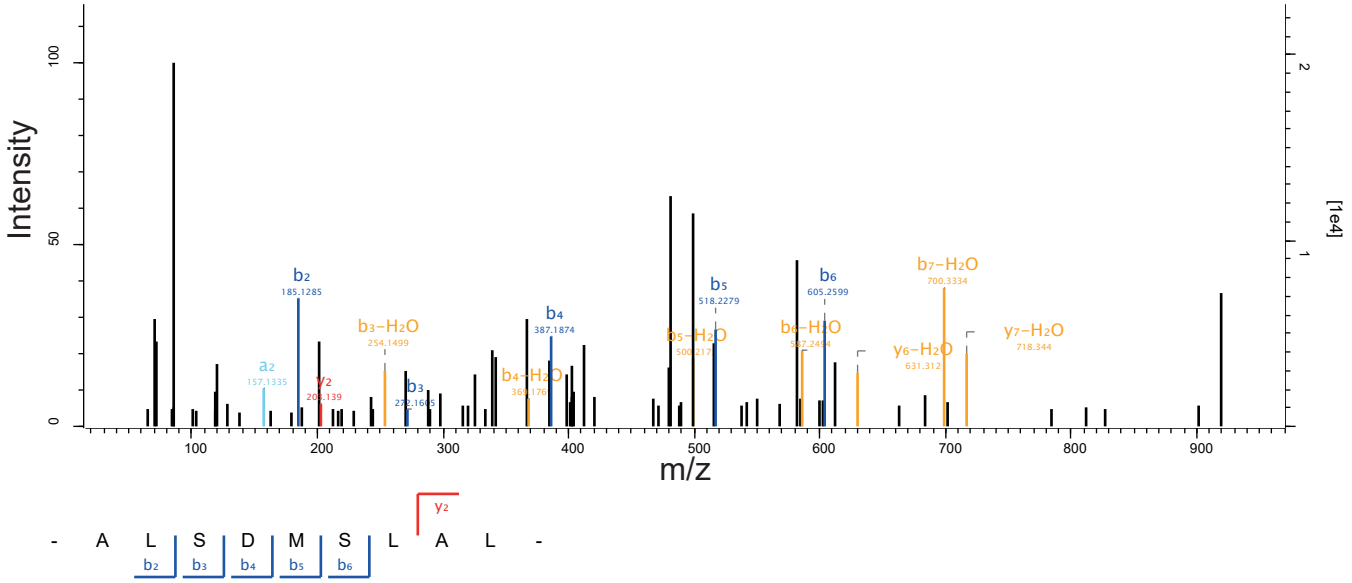


Endogenous/Synthetic peptide correlation

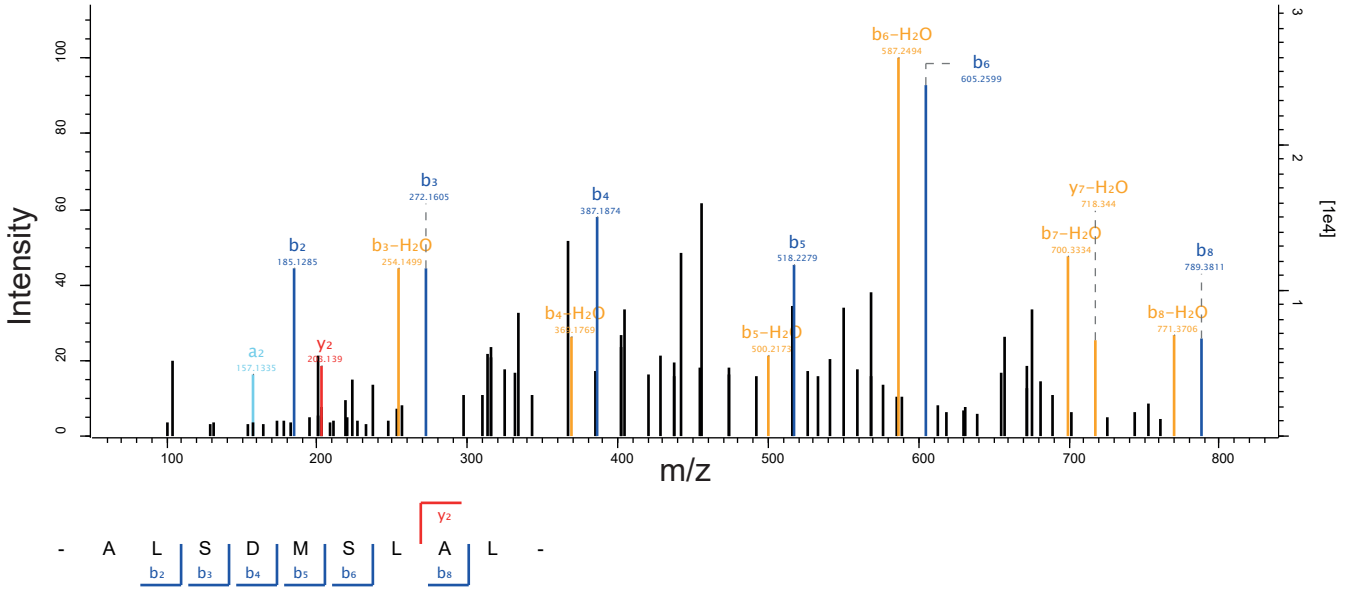


55B3 Sphingomonas dokdonensis HLA-I ALSDMSLAL

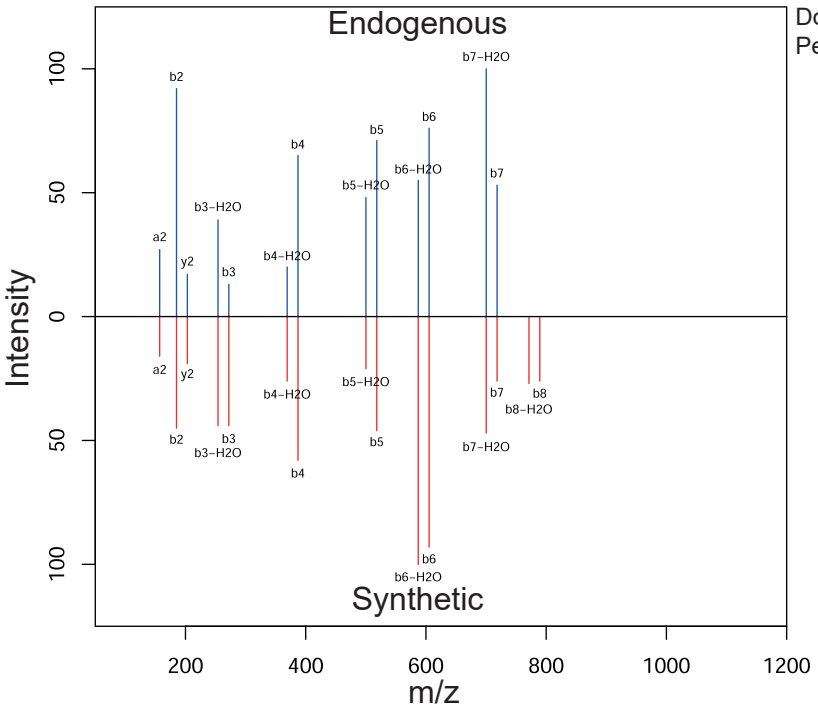
Endogenous peptide



Synthetic peptide



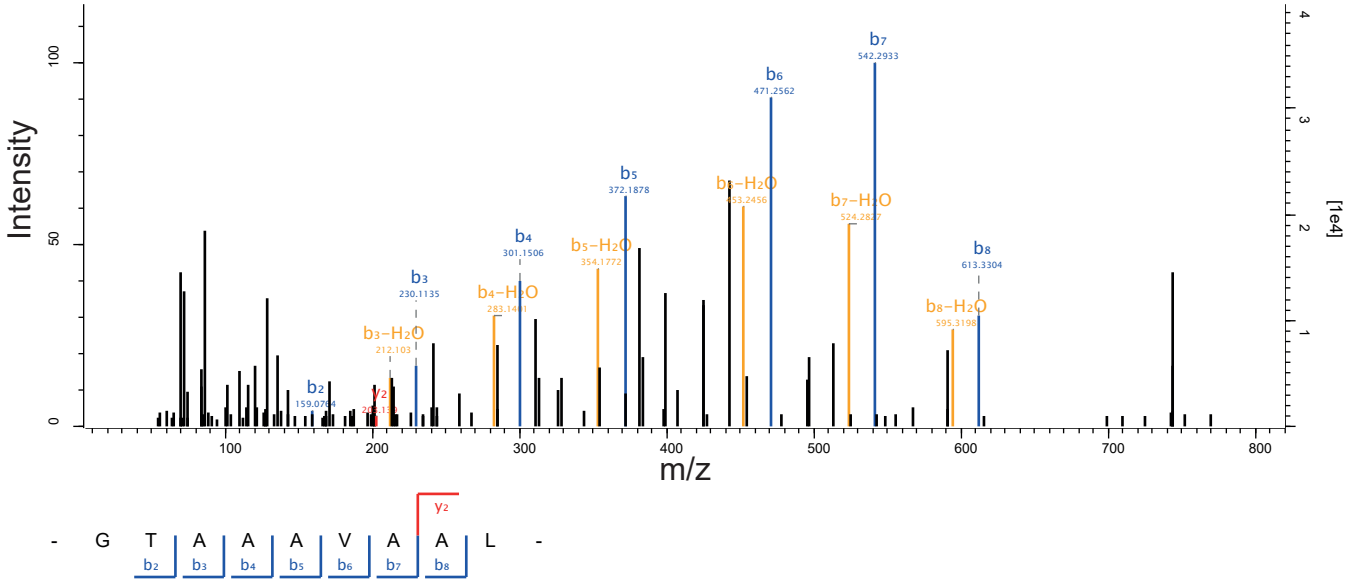
Endogenous/Synthetic peptide correlation



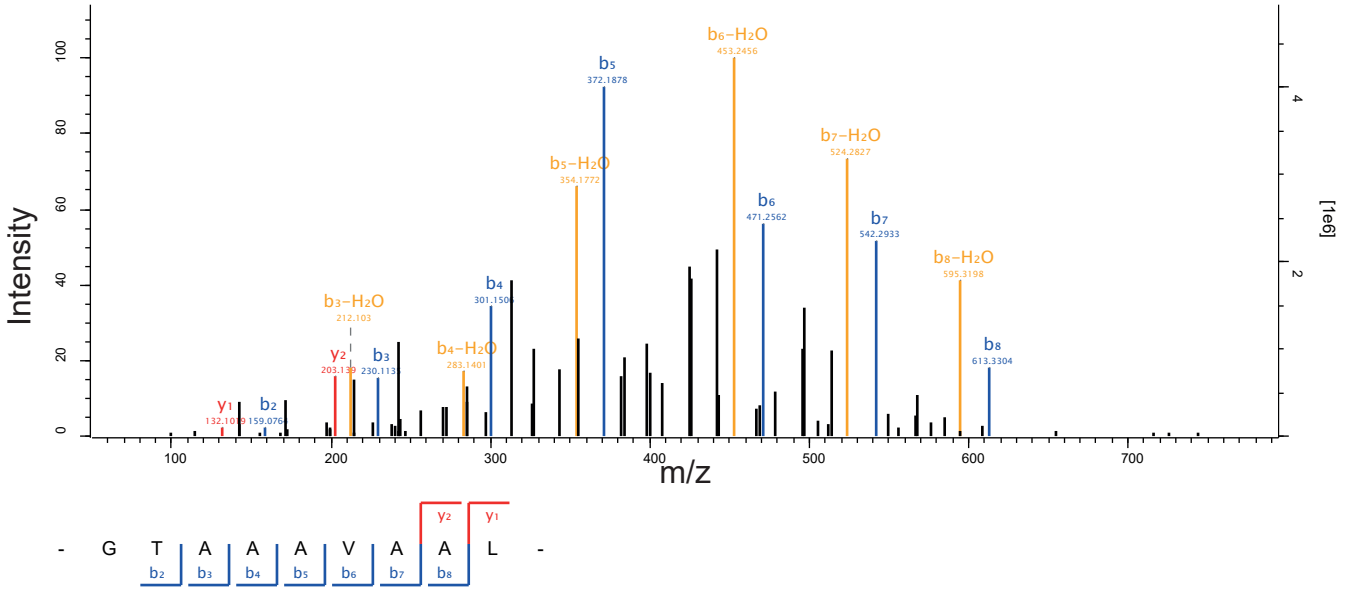
Dot product score=0.904
Pearson correlation=0.903

55B3 Sphingomonas dokdonensis HLA-I GTAAAVAAL

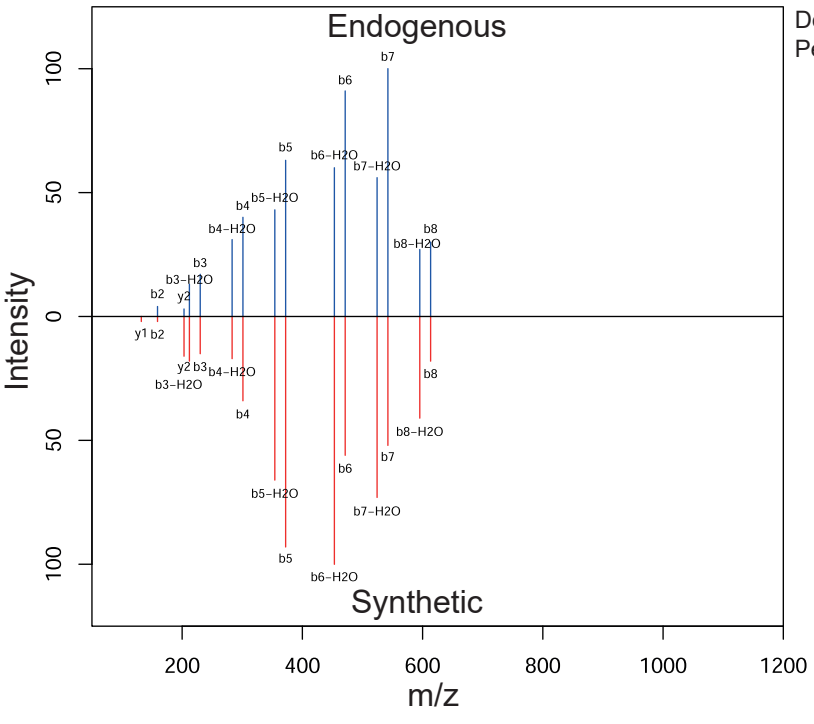
Endogenous peptide



Synthetic peptide



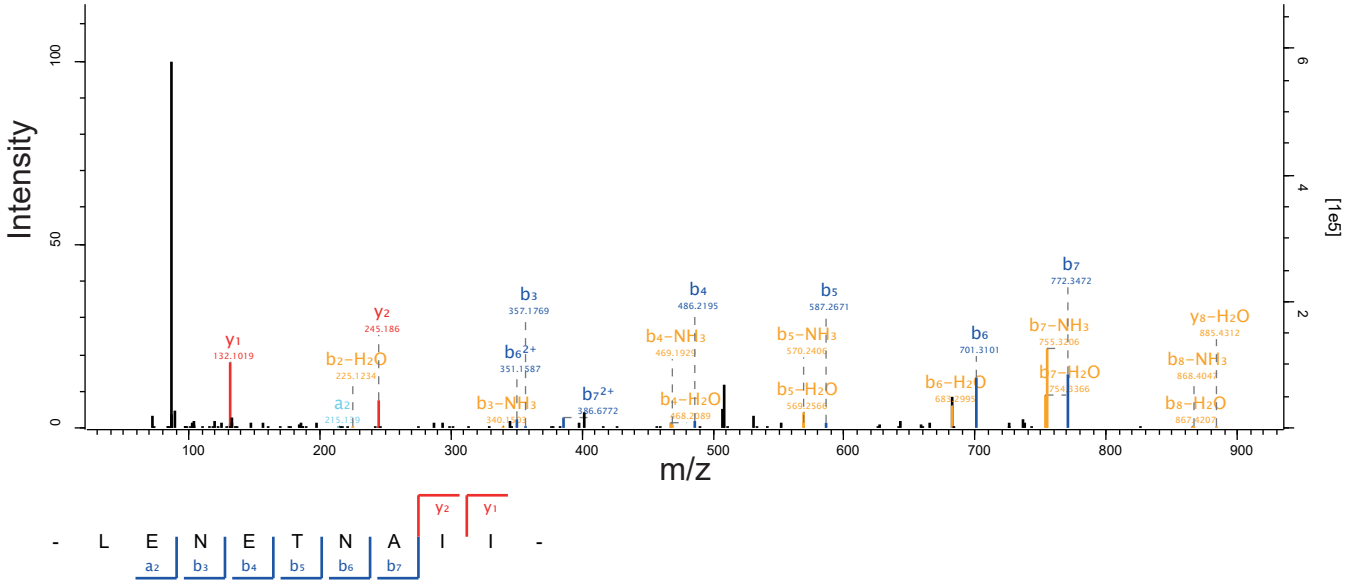
Endogenous/Synthetic peptide correlation



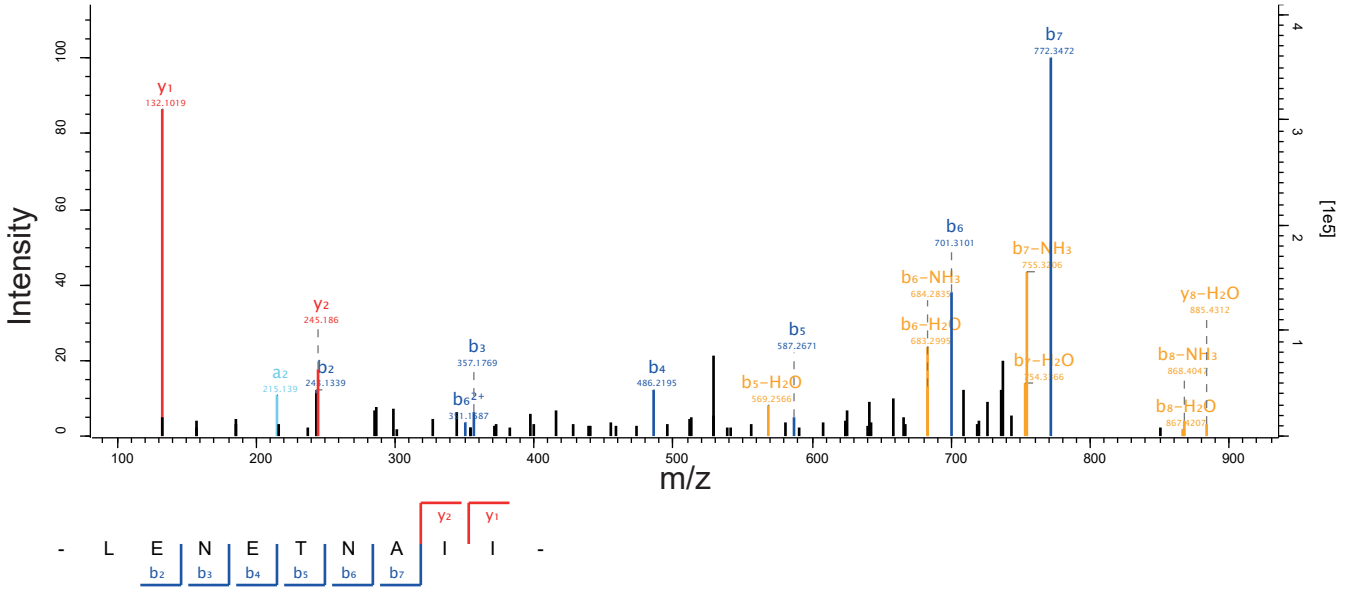
Dot product score=0.905
Pearson correlation=0.904

92B3 Fusobacterium nucleatum HLA-I LENETNAII

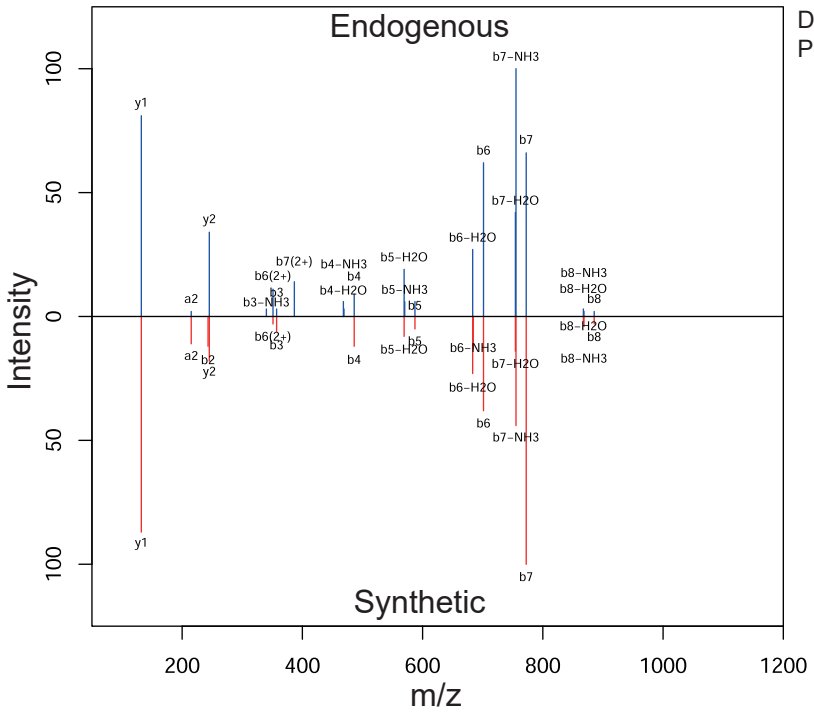
Endogenous peptide



Synthetic peptide



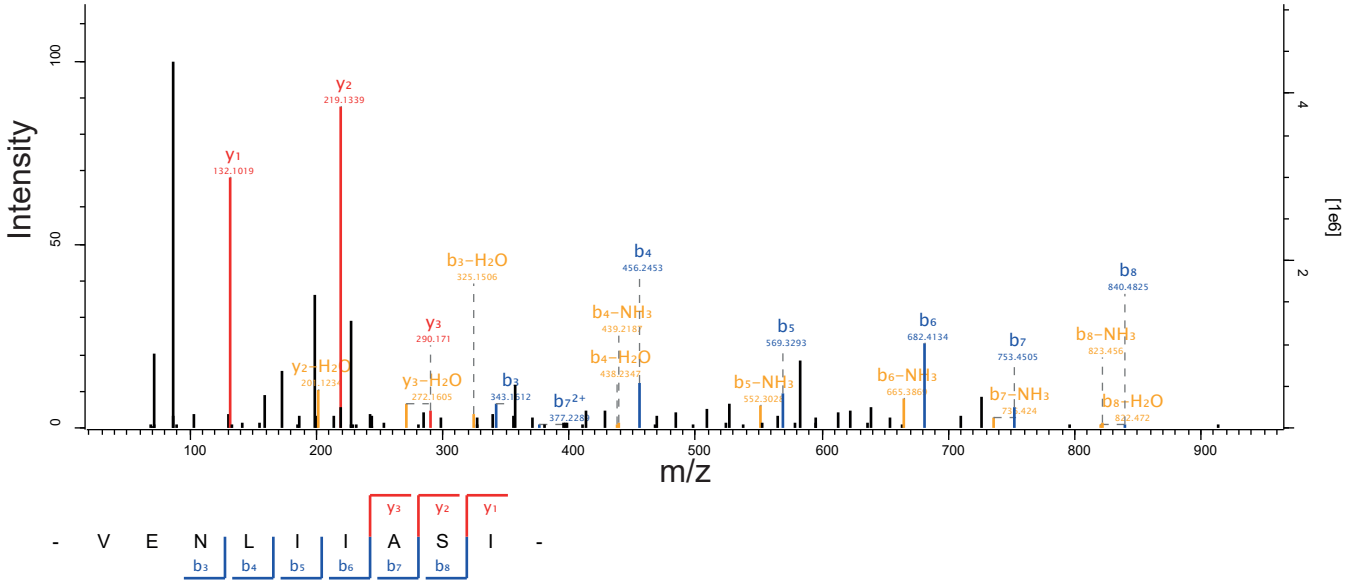
Endogenous/Synthetic peptide correlation



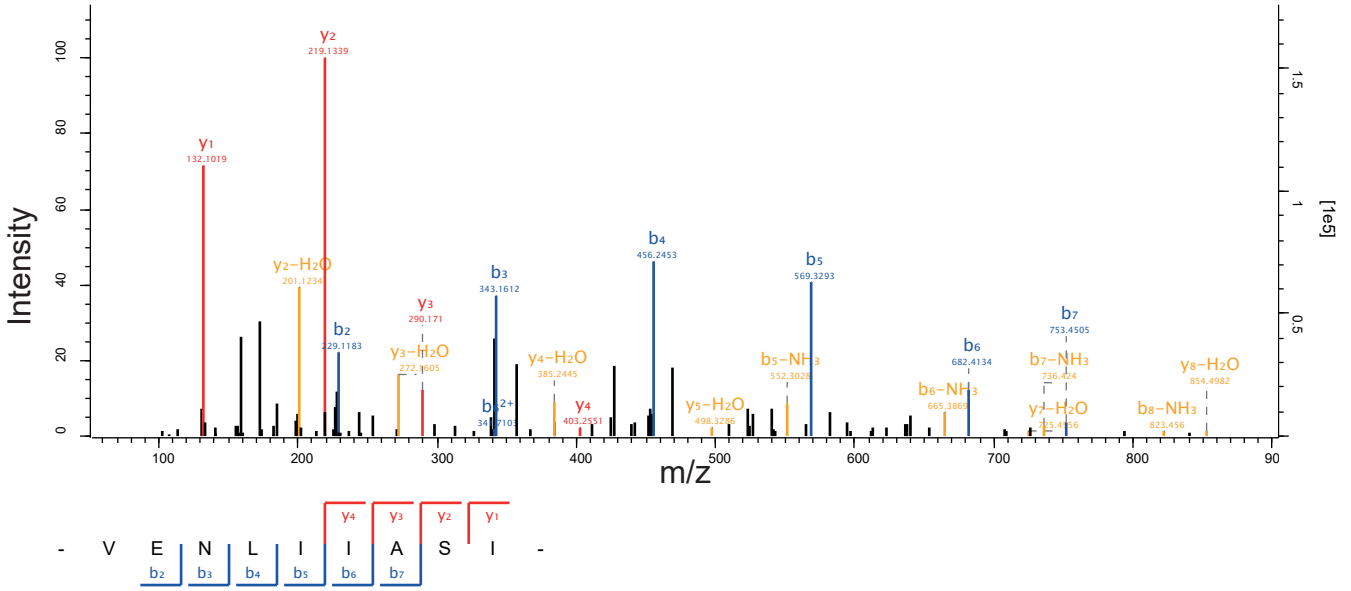
Dot product score=0.945
Pearson correlation=0.945

92B3 Fusobacterium nucleatum HLA-I VENLIIASI

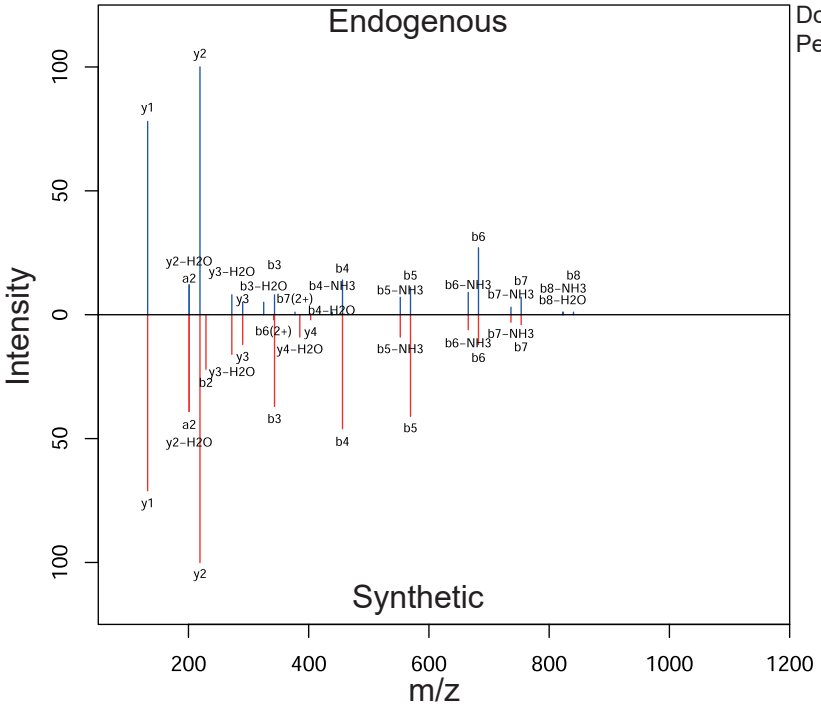
Endogenous peptide



Synthetic peptide



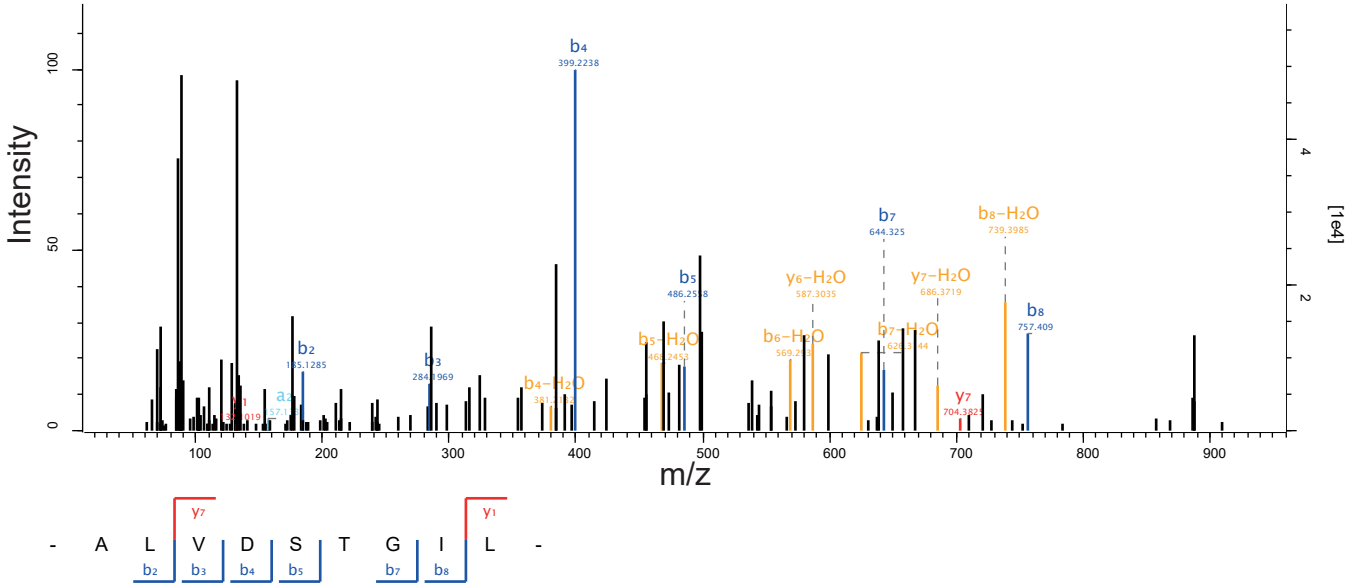
Endogenous/Synthetic peptide correlation



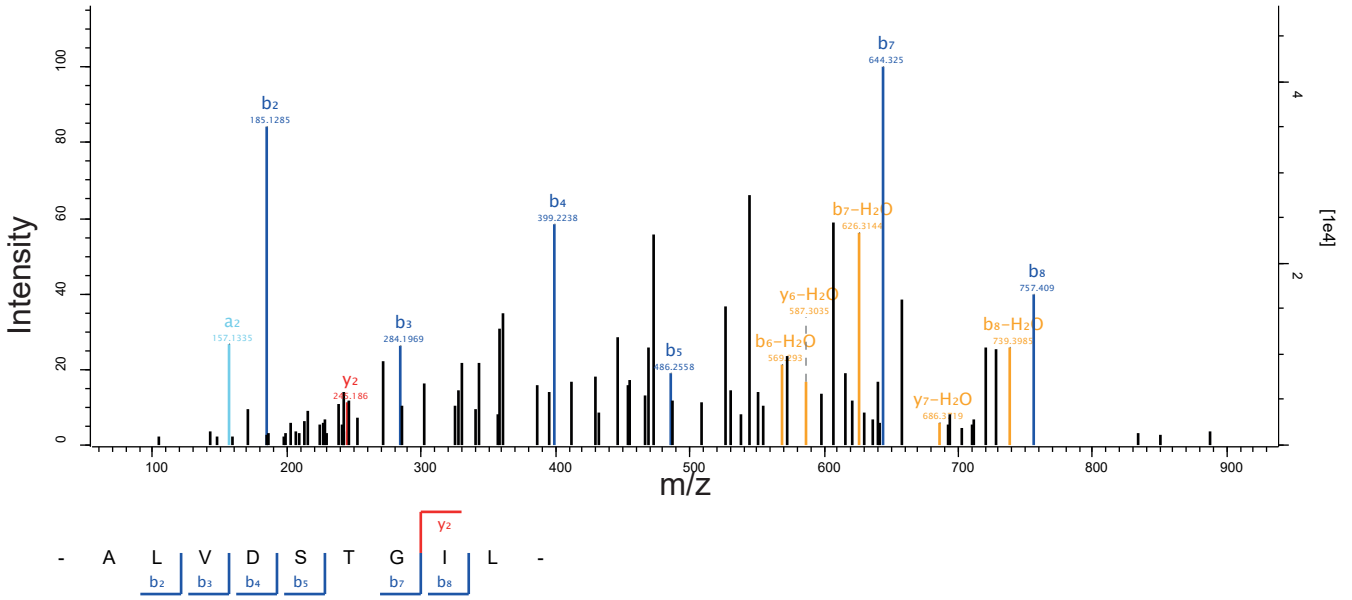
Dot product score=0.921
Pearson correlation=0.921

27 Actinomyces odontolyticus HLA-I ALVDSTGIL

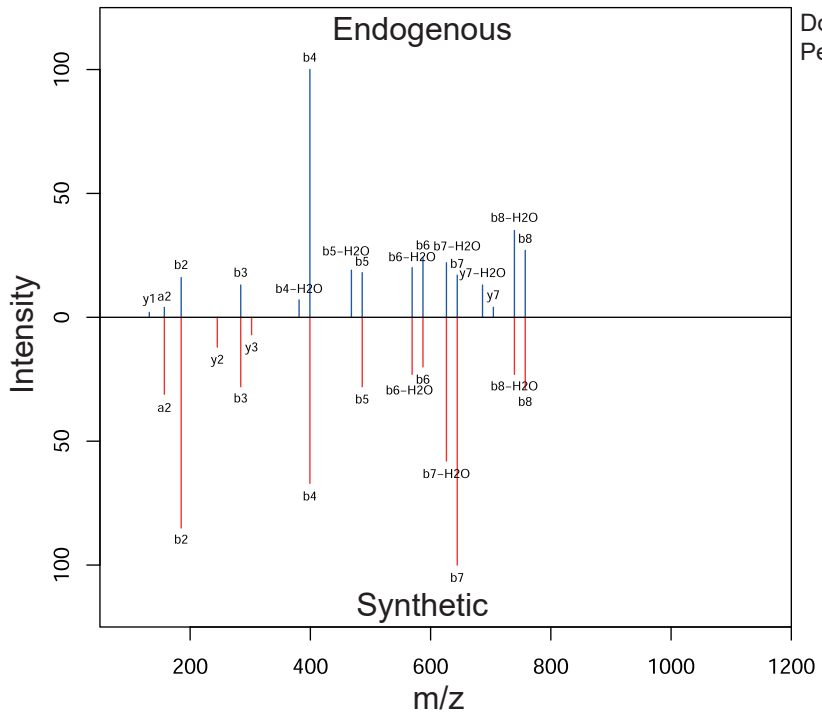
Endogenous peptide



Synthetic peptide



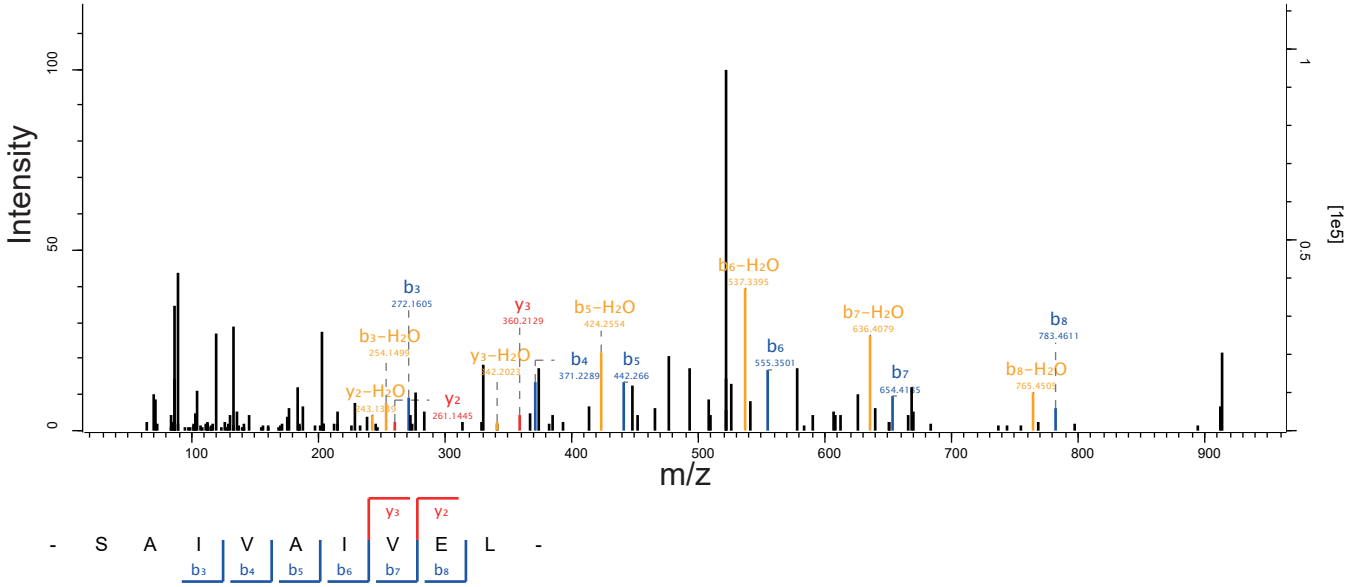
Endogenous/Synthetic peptide correlation



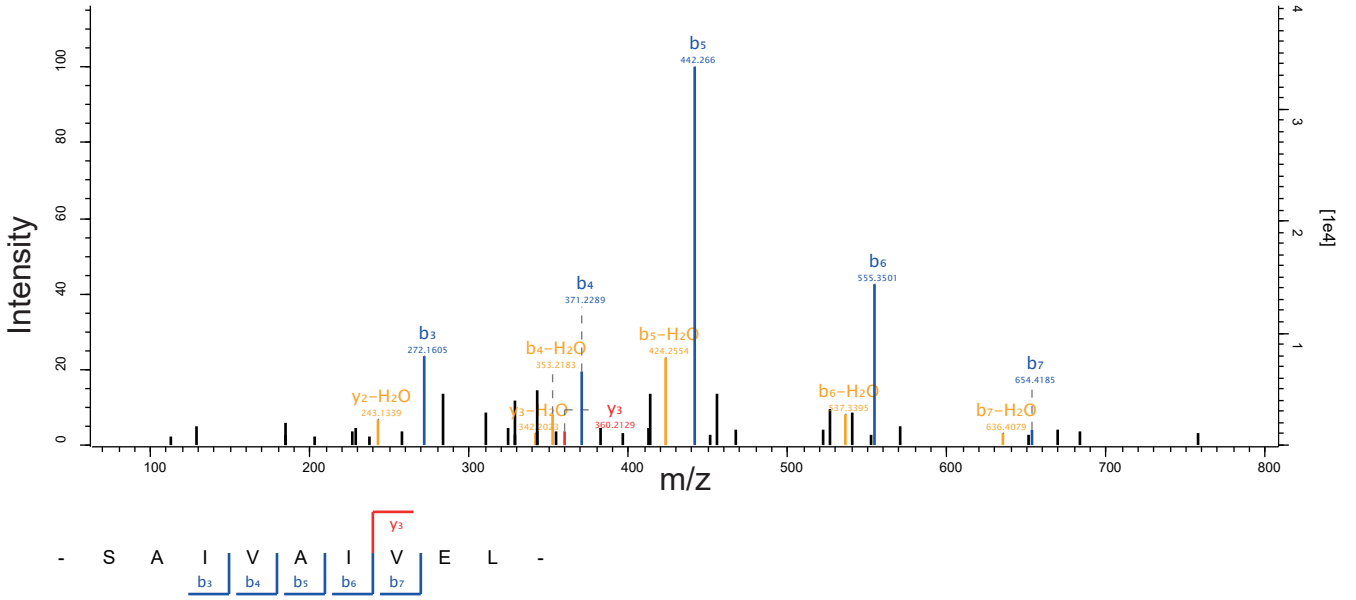
Dot product score=0.684
Pearson correlation=0.681

27 Actinomyces odontolyticus HLA-I SAIVAIVEL

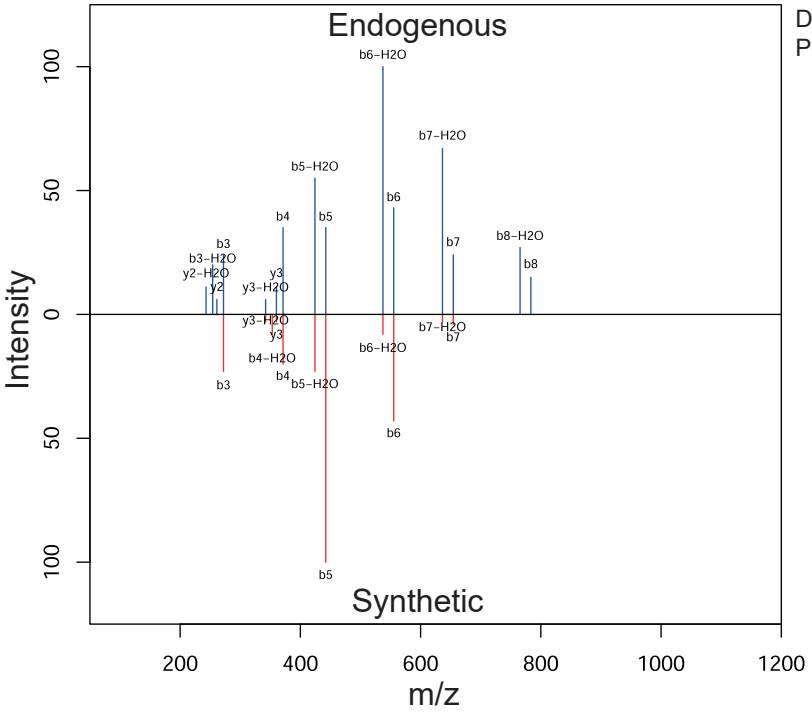
Endogenous peptide



Synthetic peptide



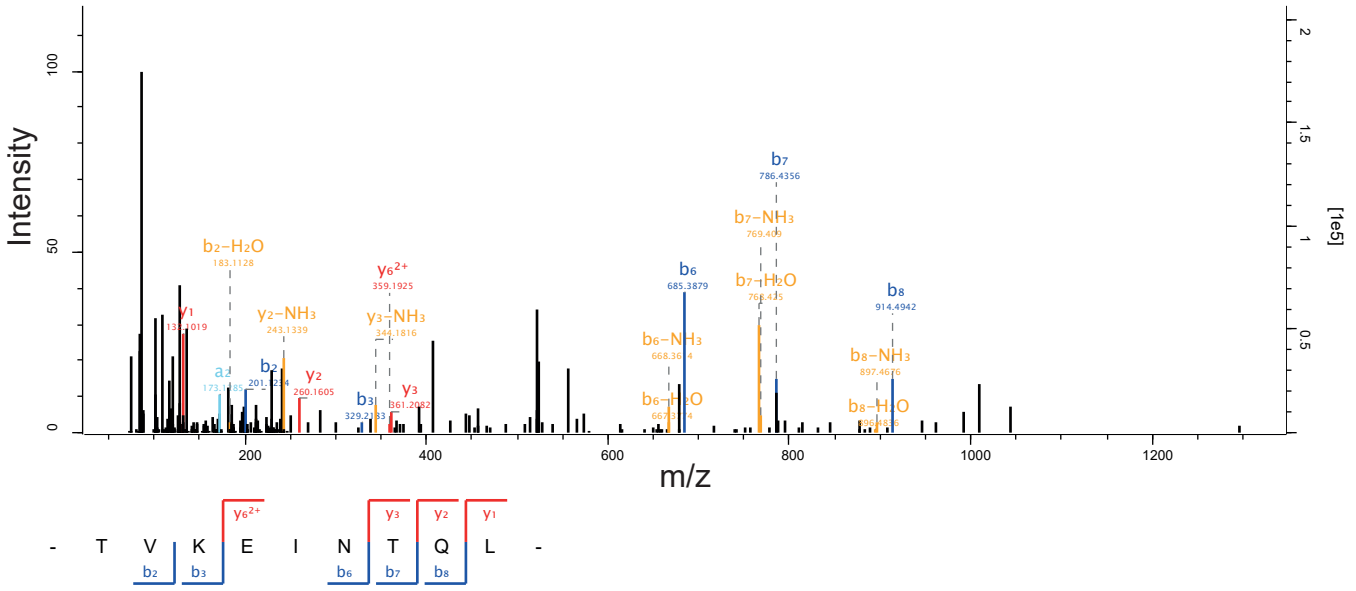
Endogenous/Synthetic peptide correlation



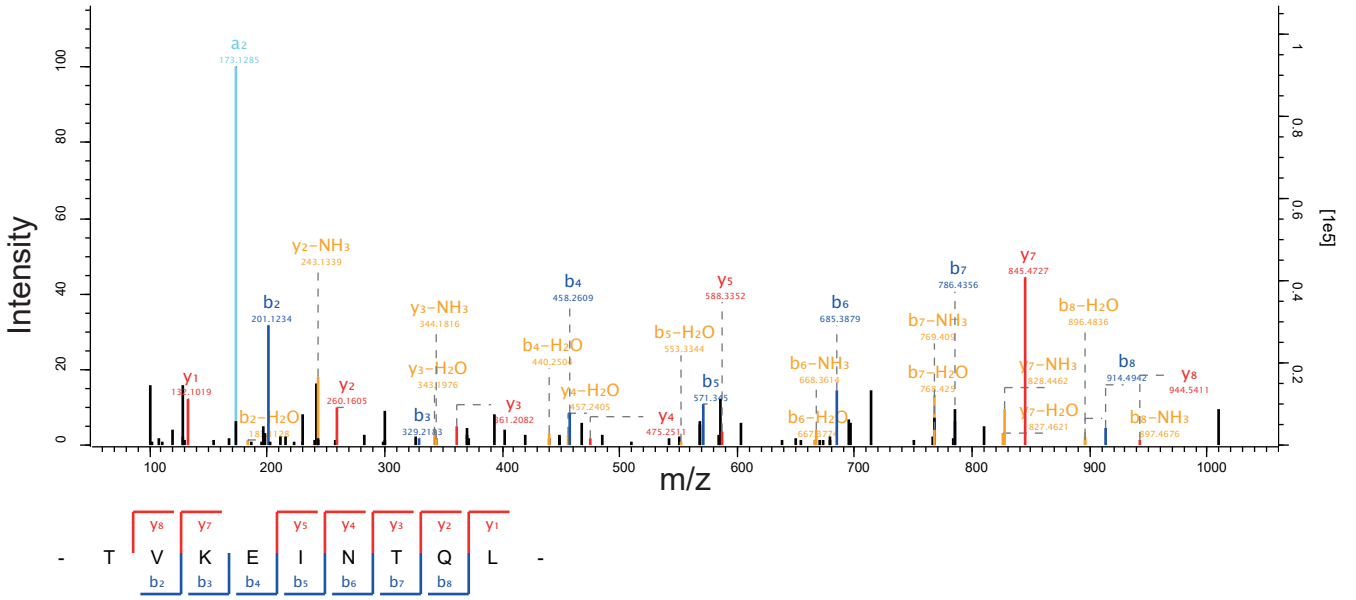
Dot product score=0.794
Pearson correlation=0.792

27 Actinomyces odontolyticus HLA-I TVKEINTQL

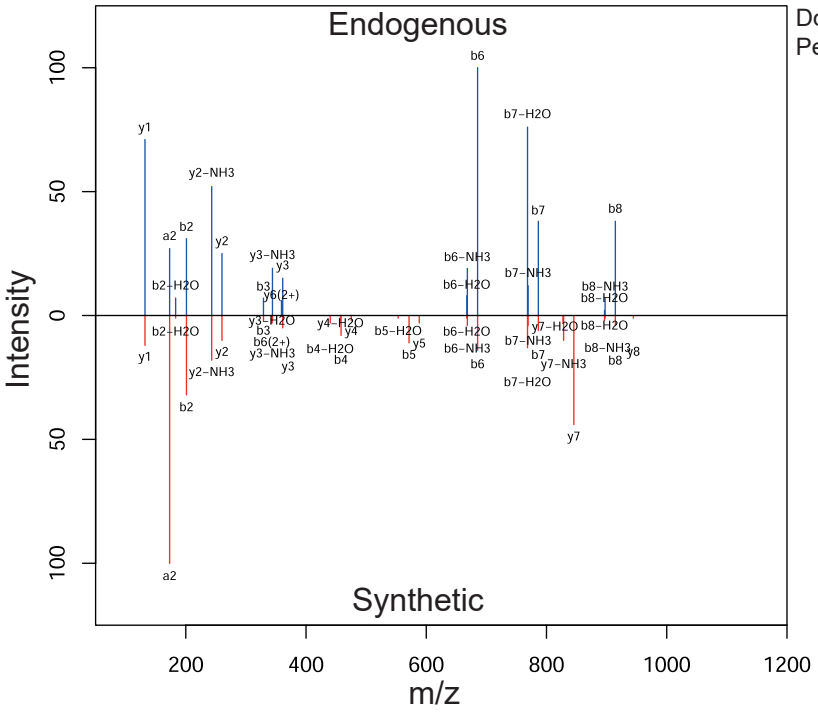
Endogenous peptide



Synthetic peptide



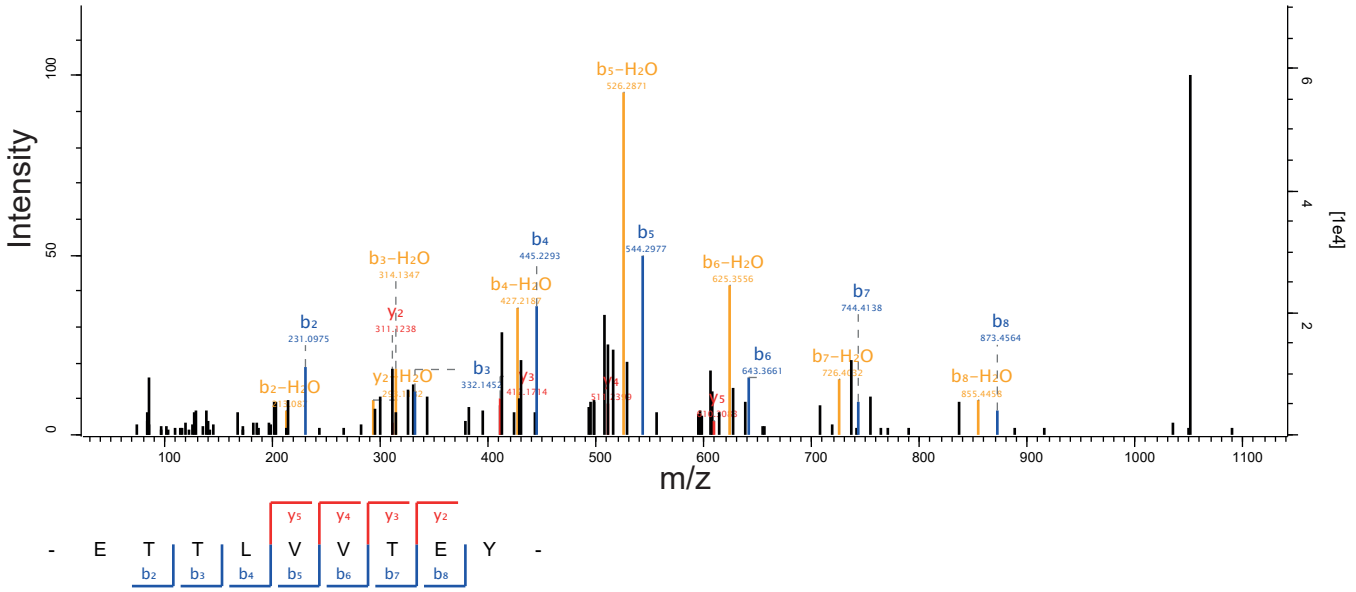
Endogenous/Synthetic peptide correlation



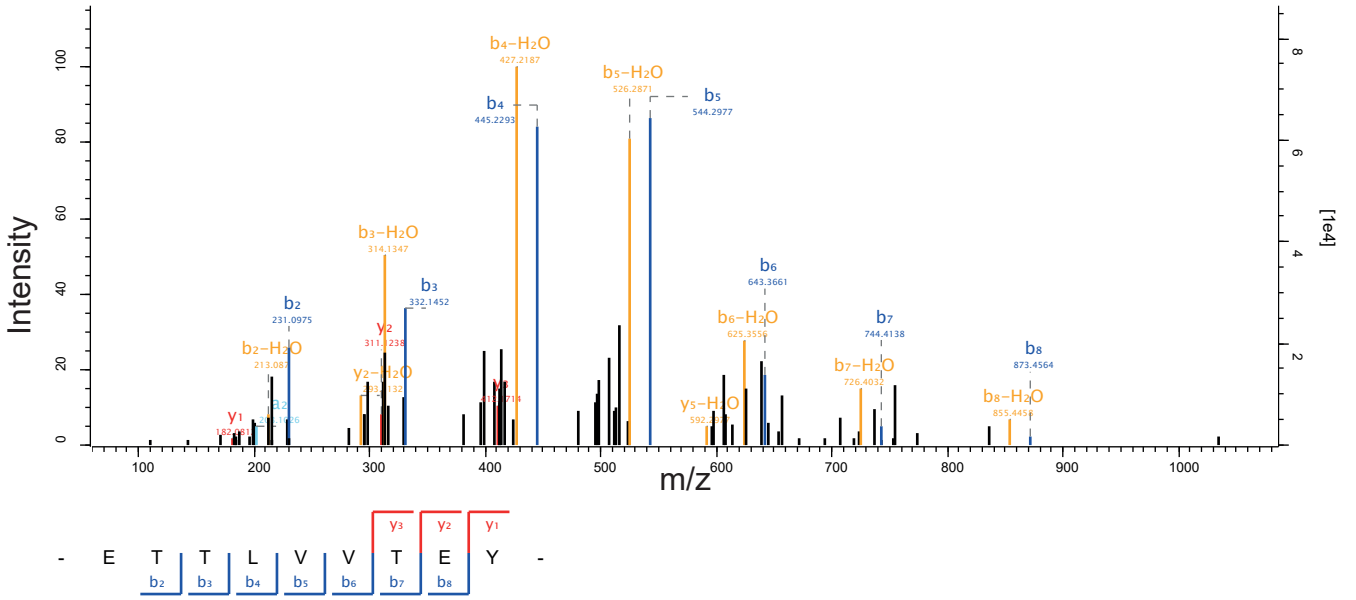
Dot product score=0.727
Pearson correlation=0.725

27 Fusobacterium nucleatum HLA-I

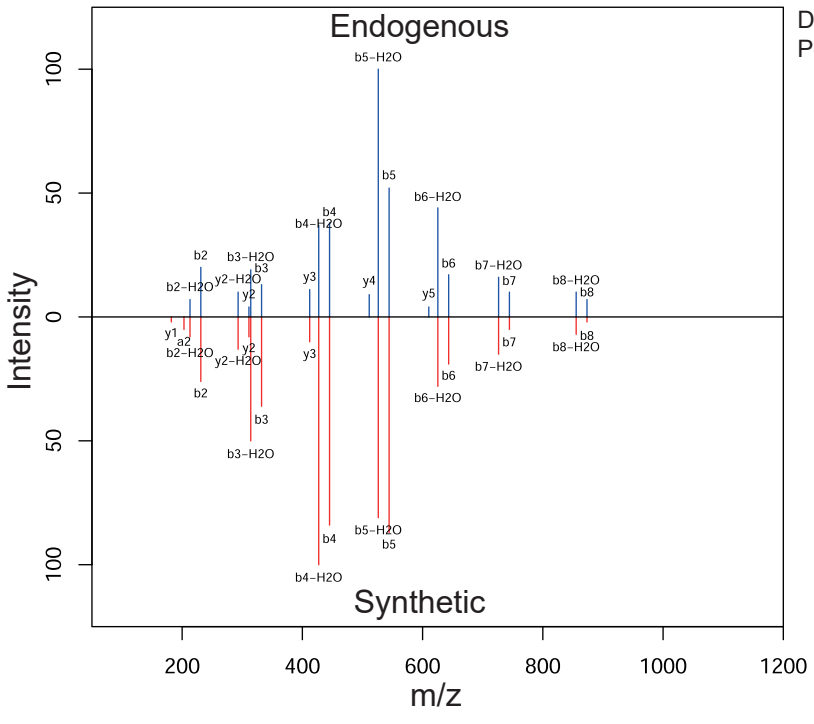
Endogenous peptide



Synthetic peptide



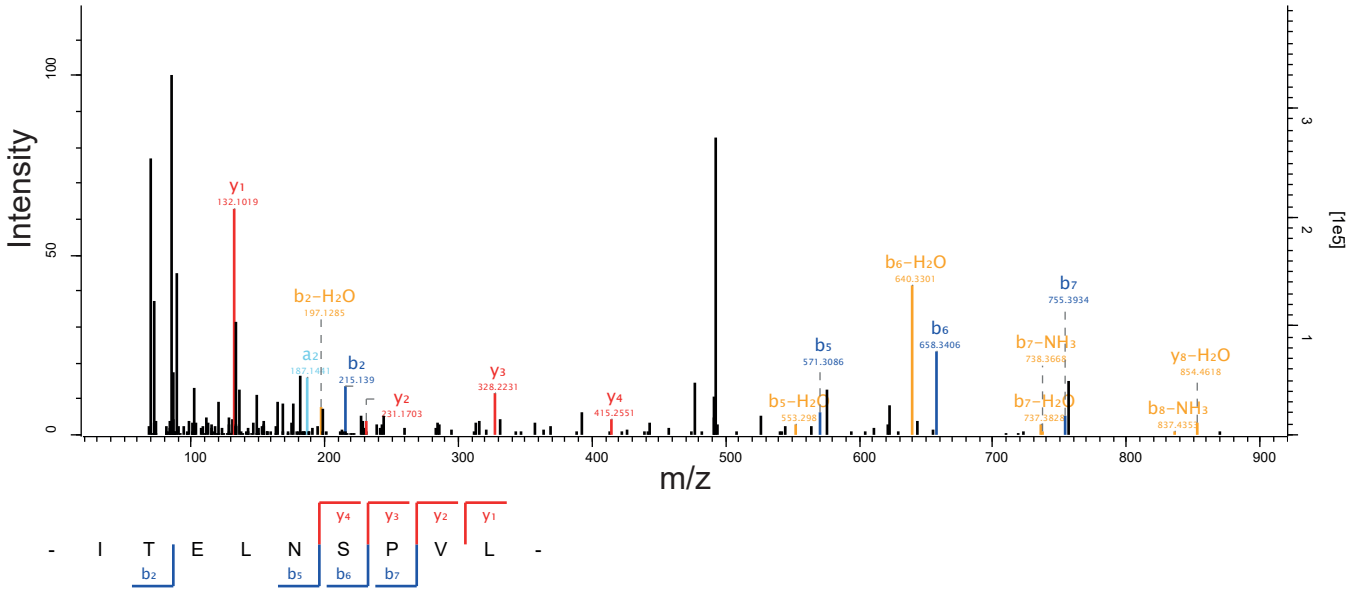
Endogenous/Synthetic peptide correlation



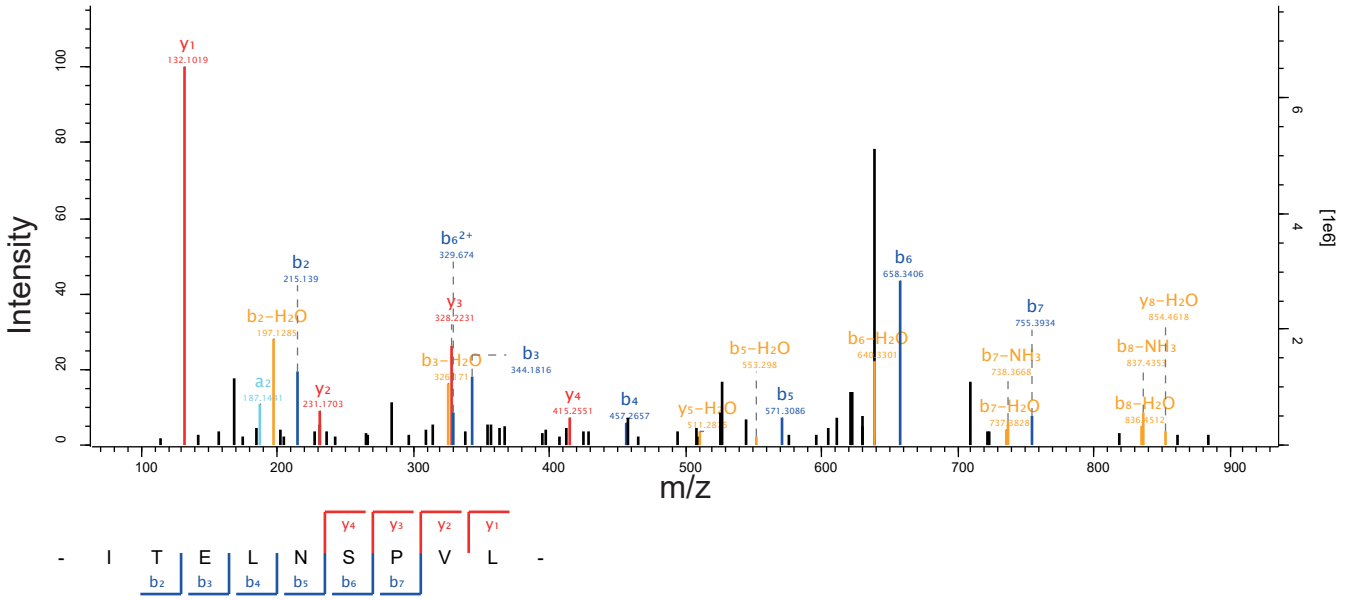
Dot product score=0.967
Pearson correlation=0.967

27 Fusobacterium nucleatum HLA-I ITELNSPVL

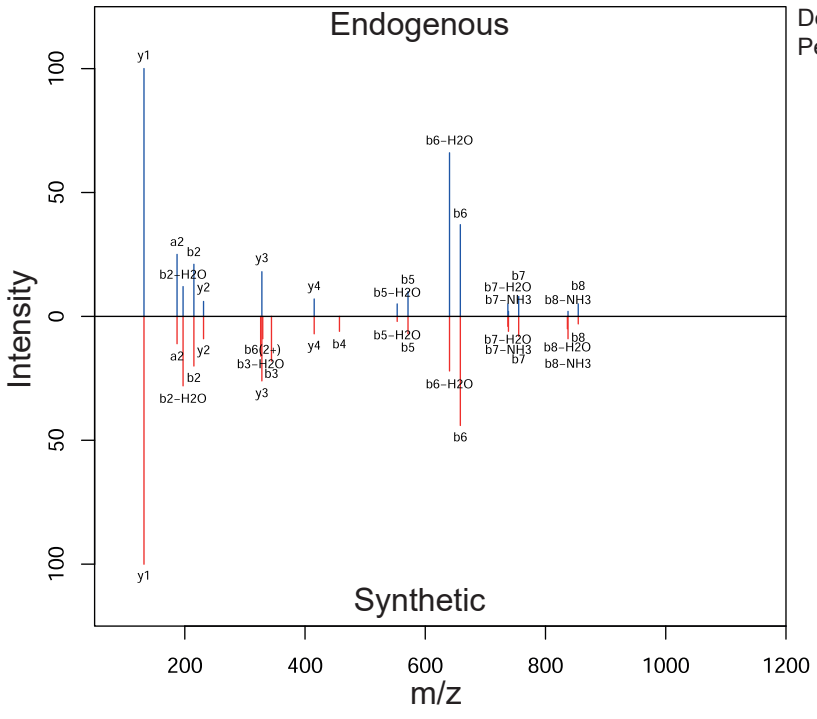
Endogenous peptide



Synthetic peptide



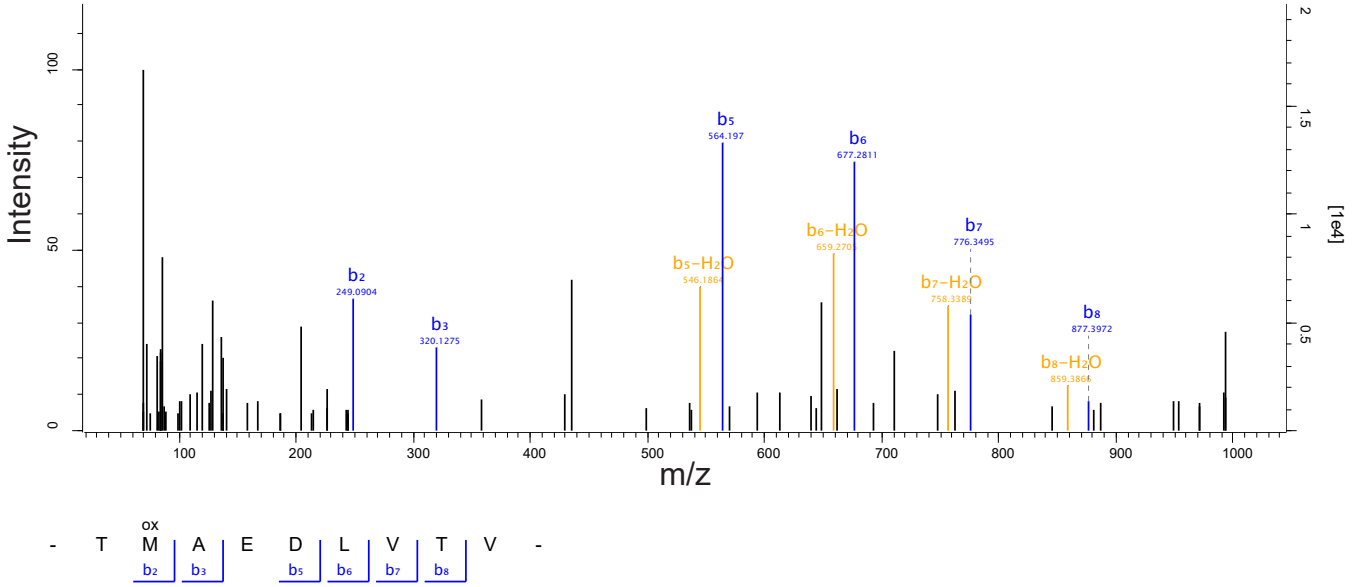
Endogenous/Synthetic peptide correlation



Dot product score=0.995
Pearson correlation=0.995

27 Actinomyces odontolyticus HLA-I

Endogenous peptide



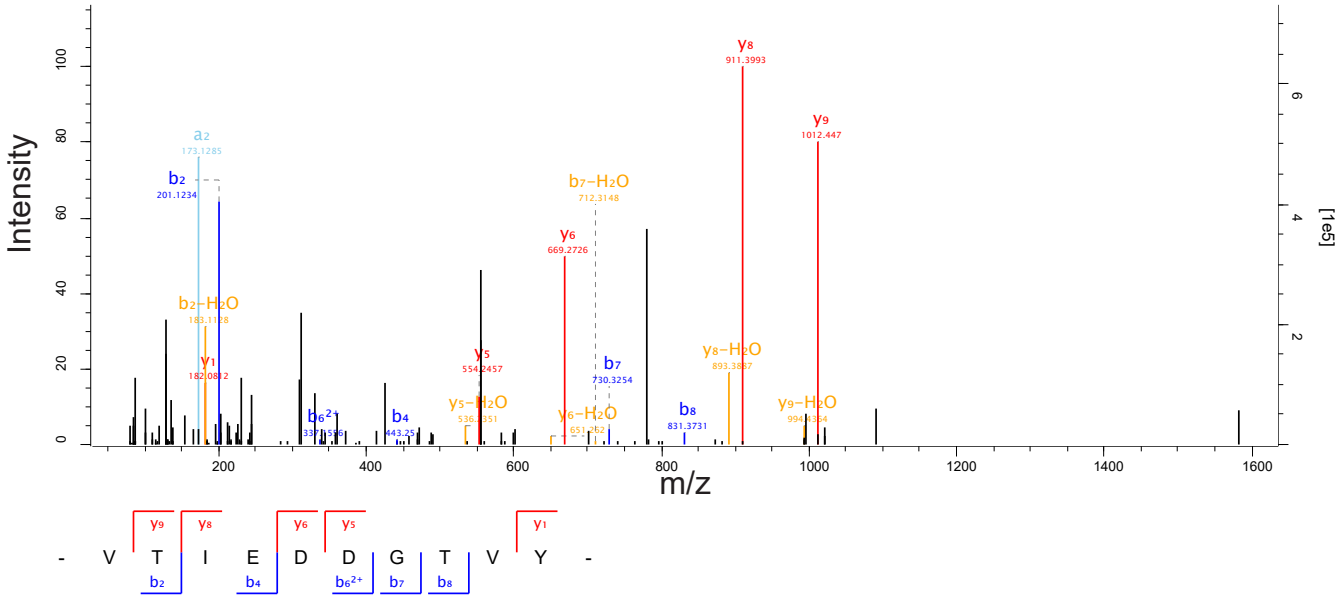
ox

- T M A E D L V T V -

b2 b3 b5 b6 b7 b8

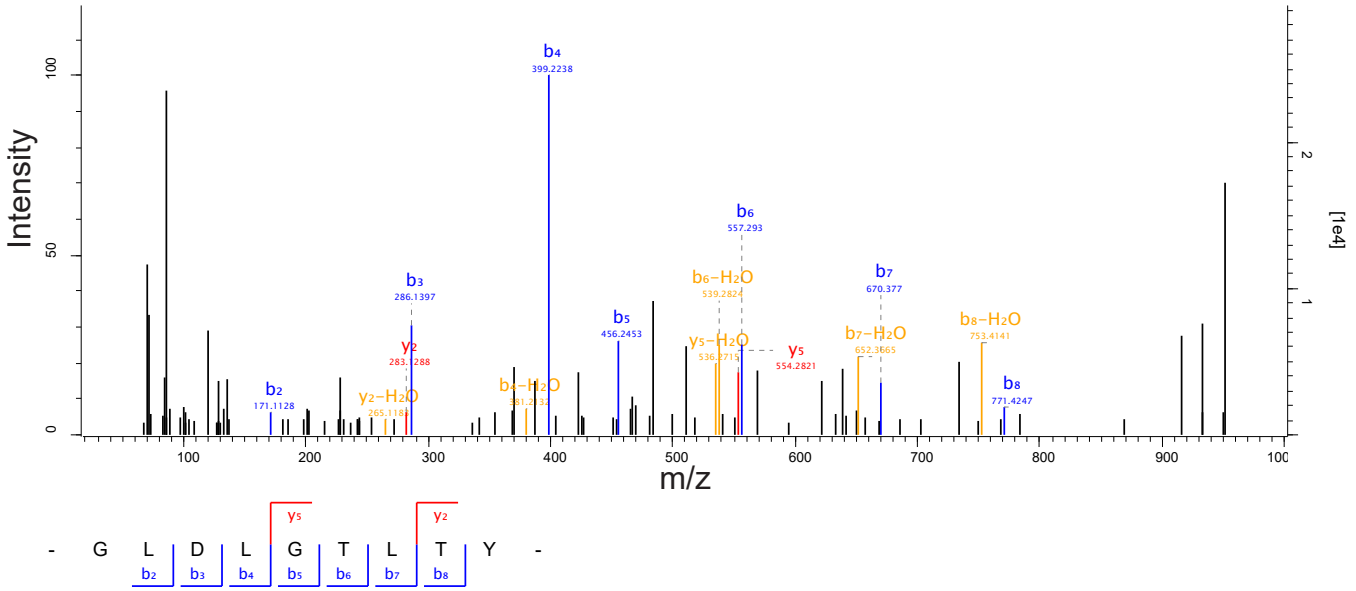
27 Actinomyces odontolyticus HLA-I VTIEDDGTVY

Endogenous peptide



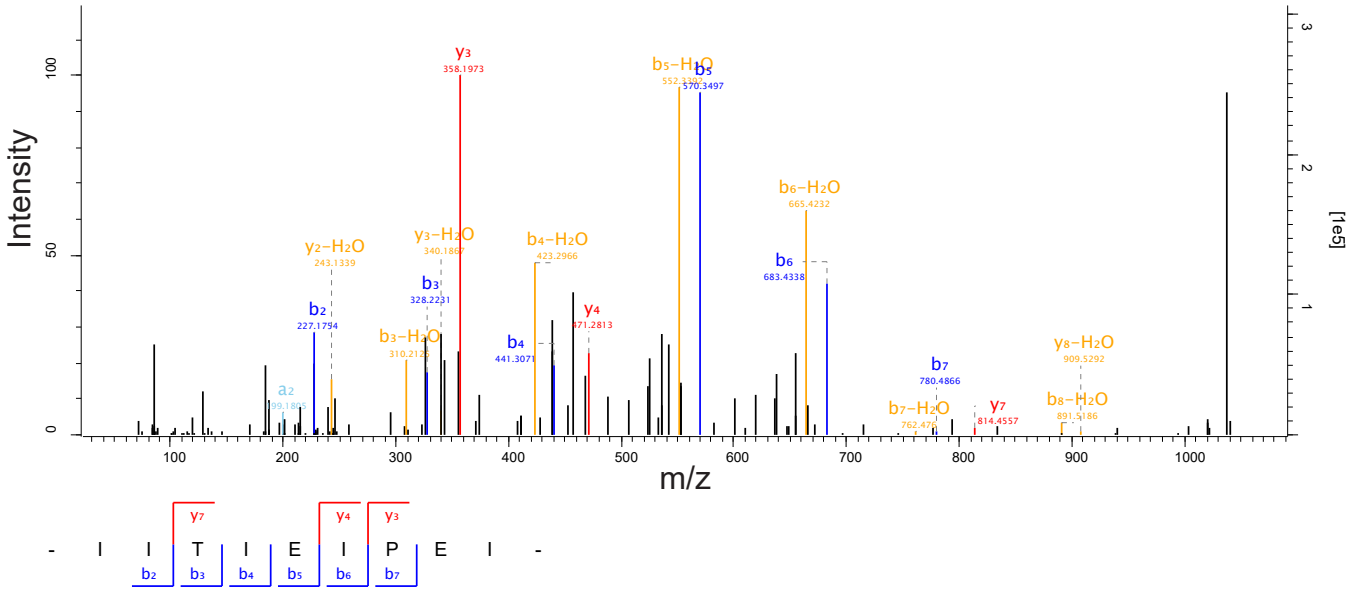
27 Fusobacterium nucleatum HLA-I GLDLGTLTY

Endogenous peptide



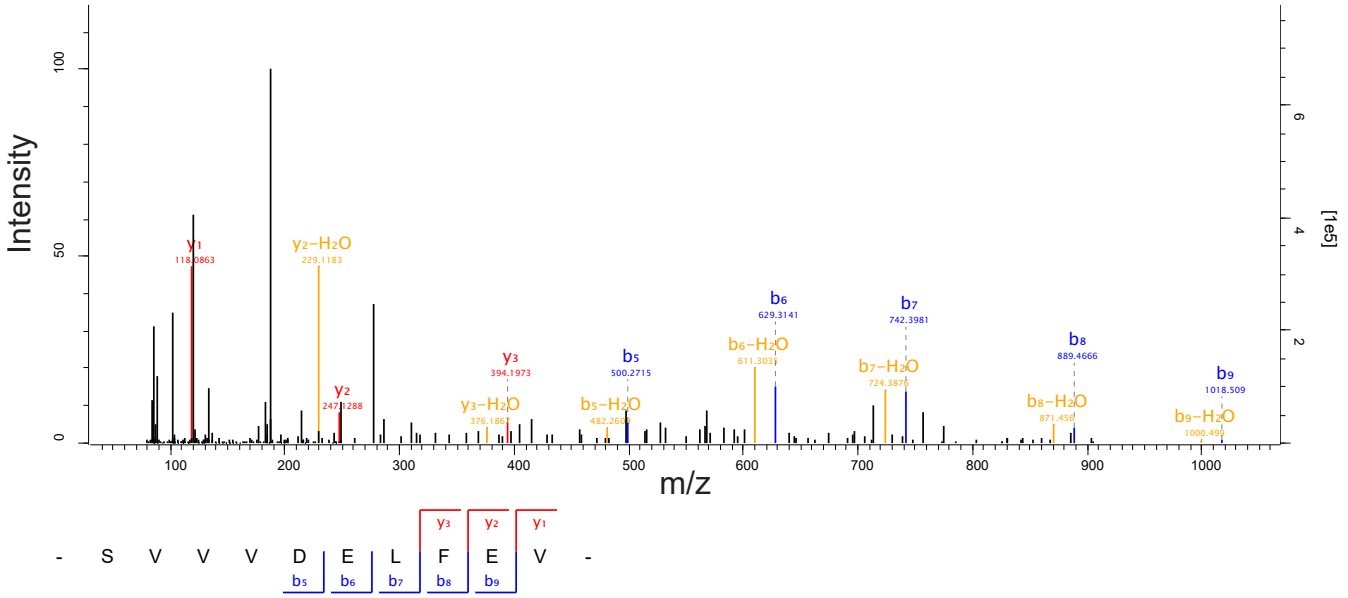
27 Fusobacterium nucleatum HLA-I IITIEIPEI

Endogenous peptide



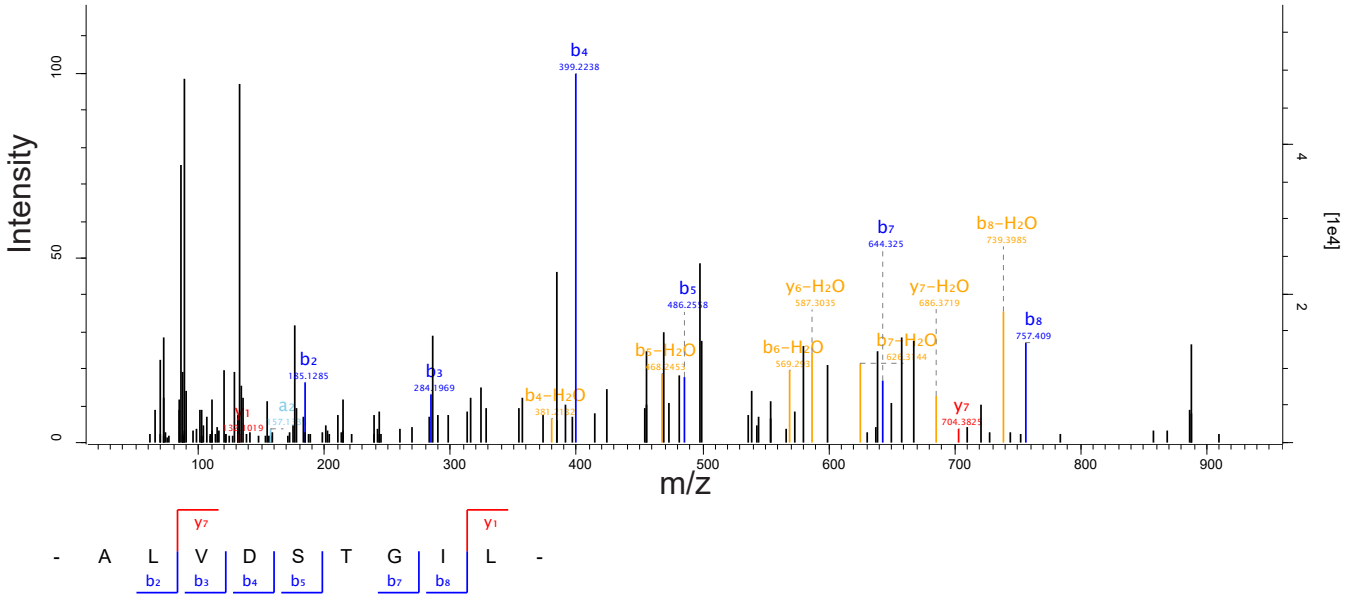
27 Fusobacterium nucleatum HLA-I SVVVDELFEV

Endogenous peptide



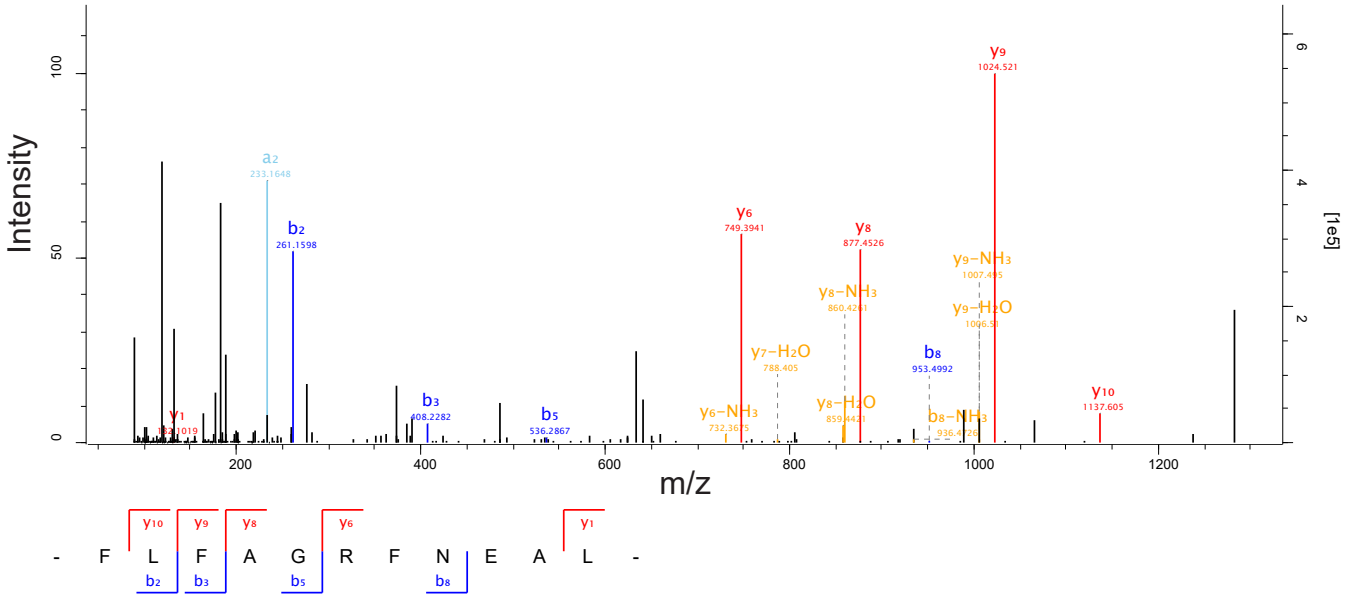
27 Schaalia odontolytica HLA-I ALVDSTGIL

Endogenous peptide



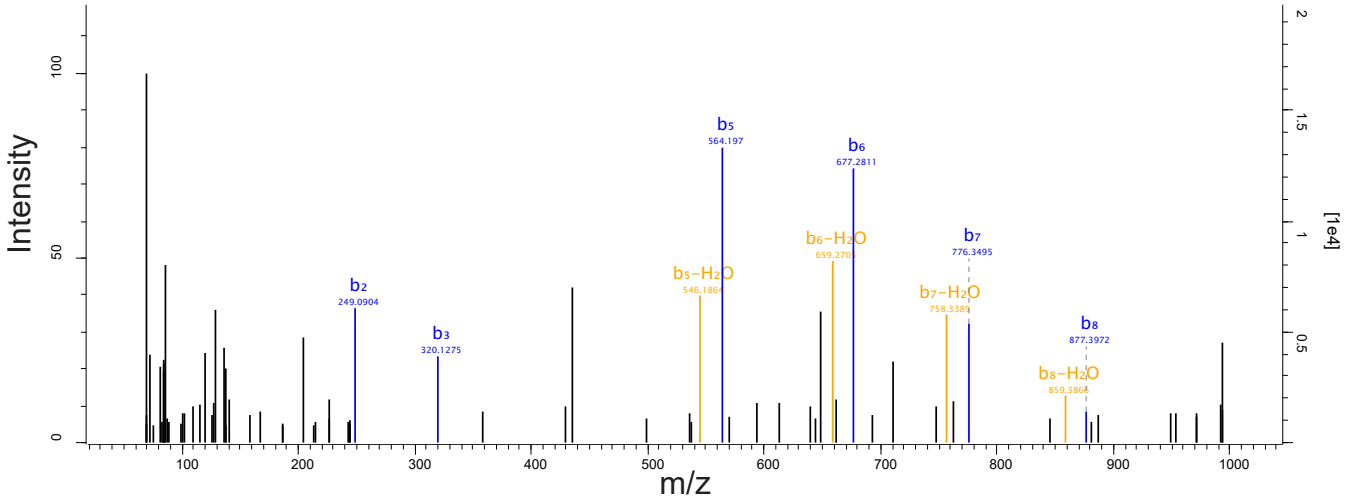
27 Schaalia odontolytica HLA-I FLFAGRFNEAL

Endogenous peptide



27 Schaalia odontolytica HLA-I TMAEDLVTV

Endogenous peptide



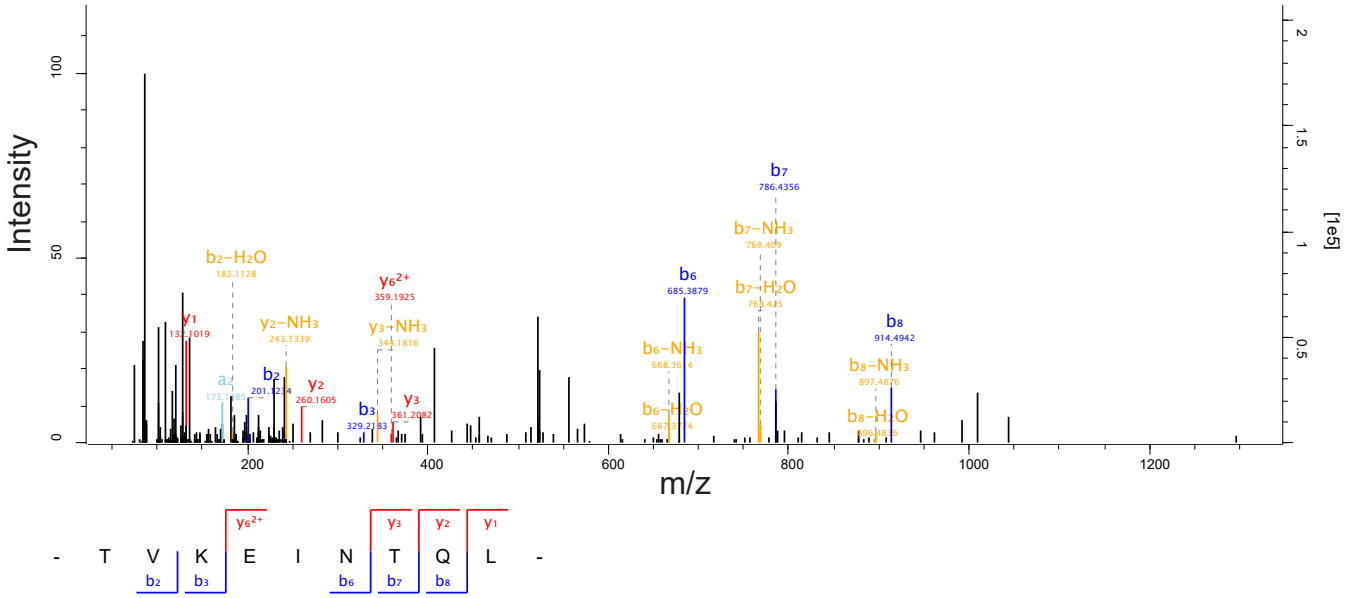
ox

- T M A E D L V T V -

b2 b3 b5 b6 b7 b8

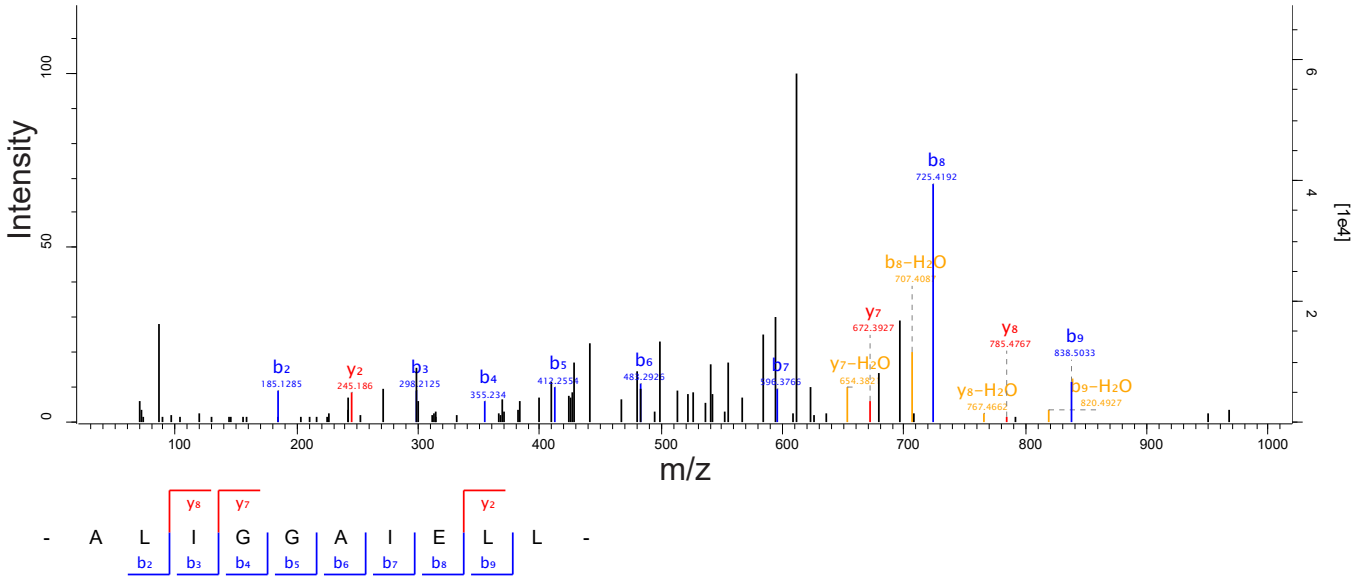
27 Schaalia odontolytica HLA-I TVKEINTQL

Endogenous peptide



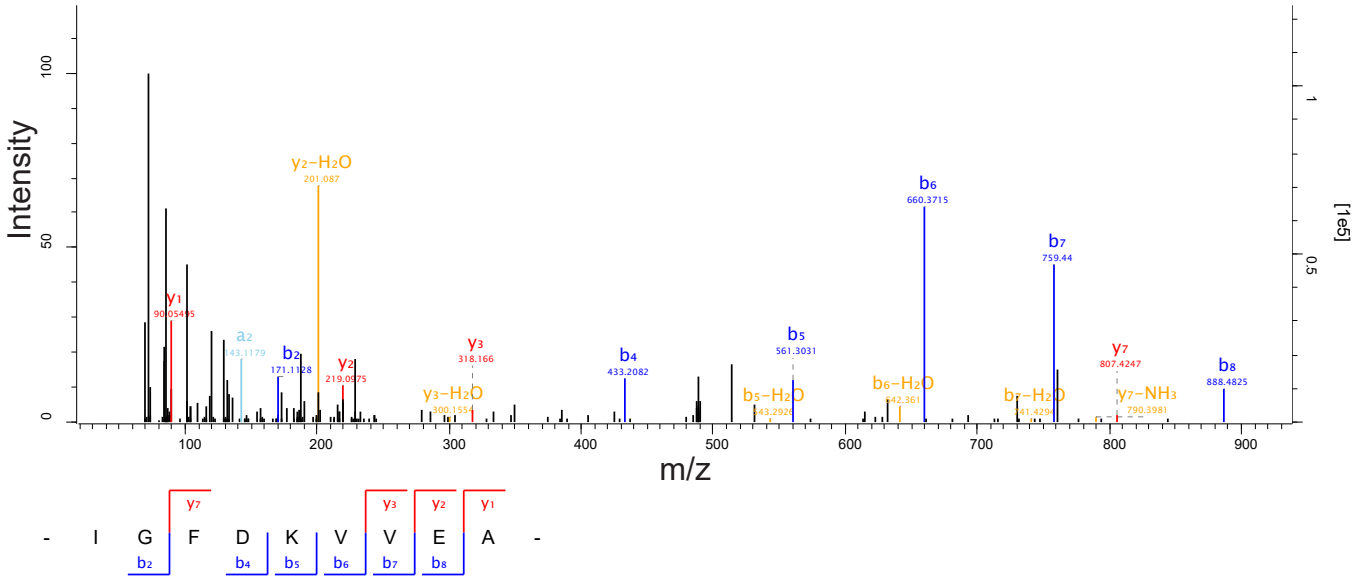
42RF Campylobacter concisus HLA-I ALIGGAIELL

Endogenous peptide



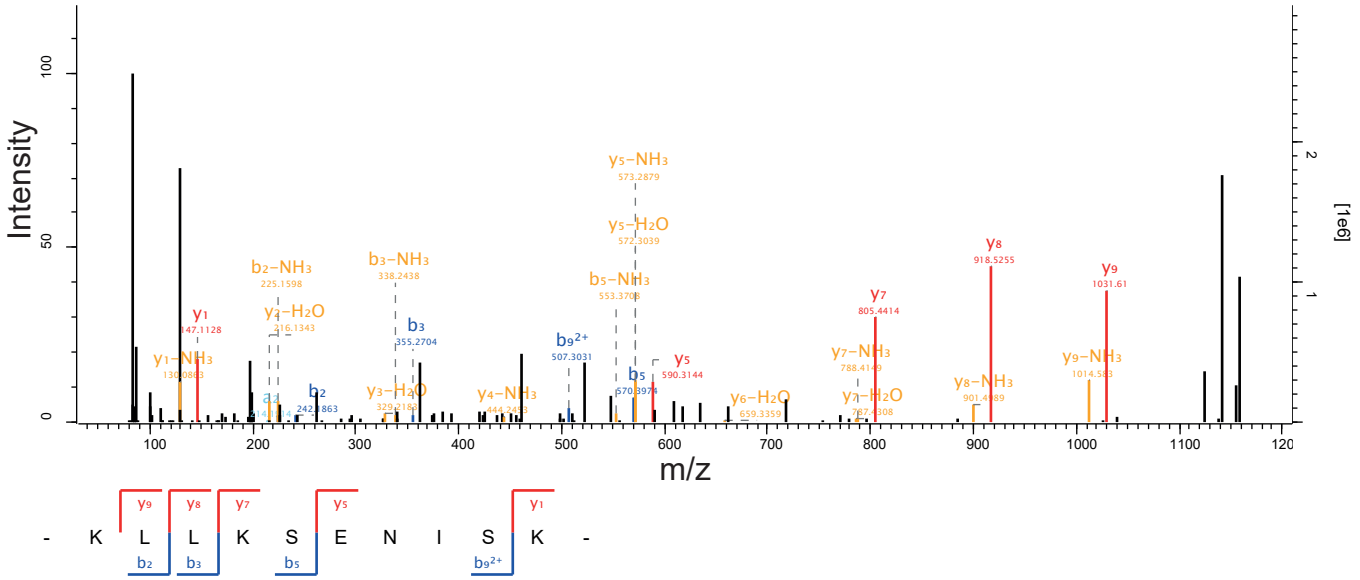
42RF Campylobacter concisus HLA-I IGFDKVVEA

Endogenous peptide



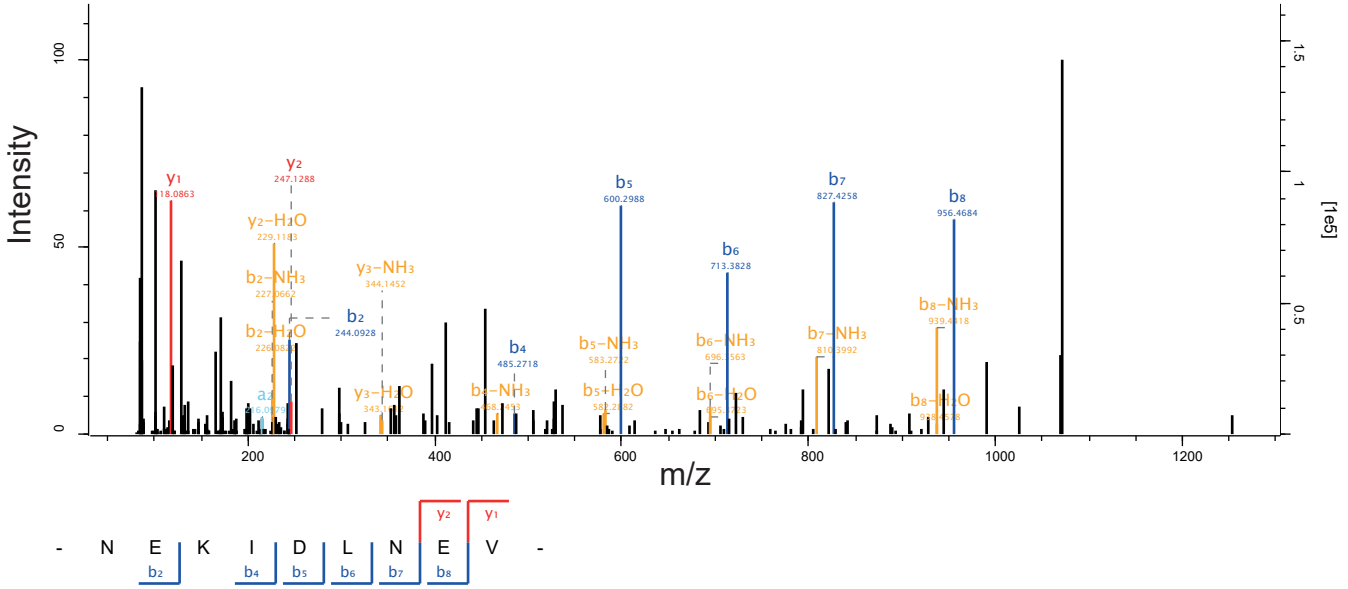
42RF Campylobacter concisus HLA-I KLLKSENISK

Endogenous peptide



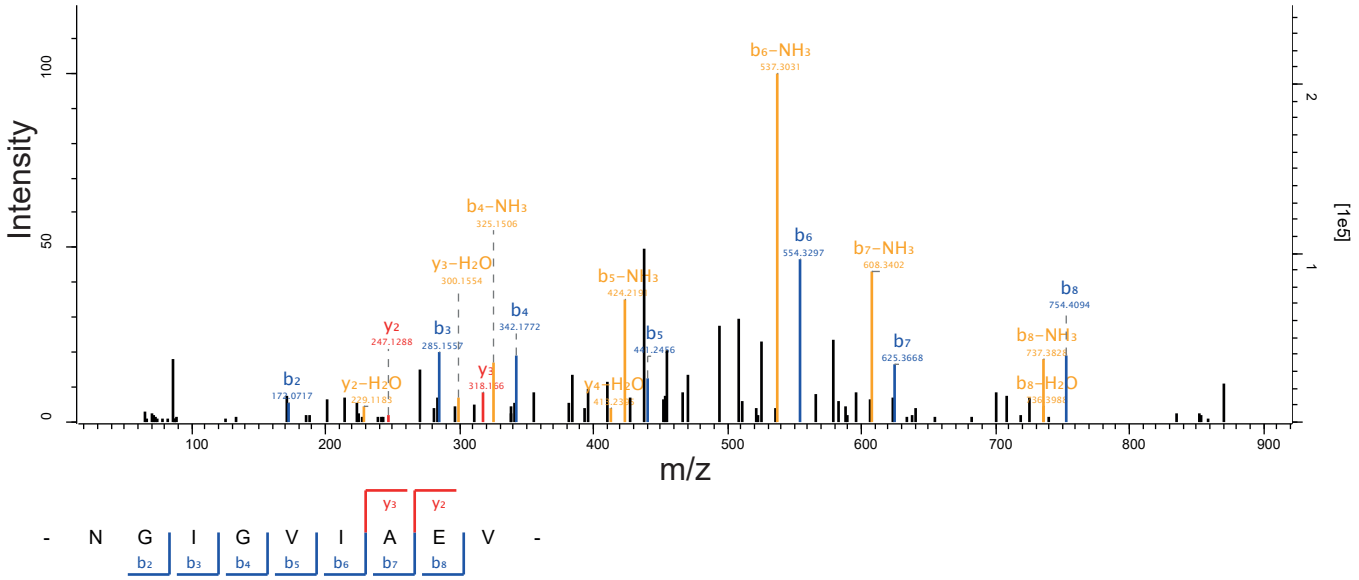
42RF Campylobacter concisus HLA-I NEKIDLNEV

Endogenous peptide



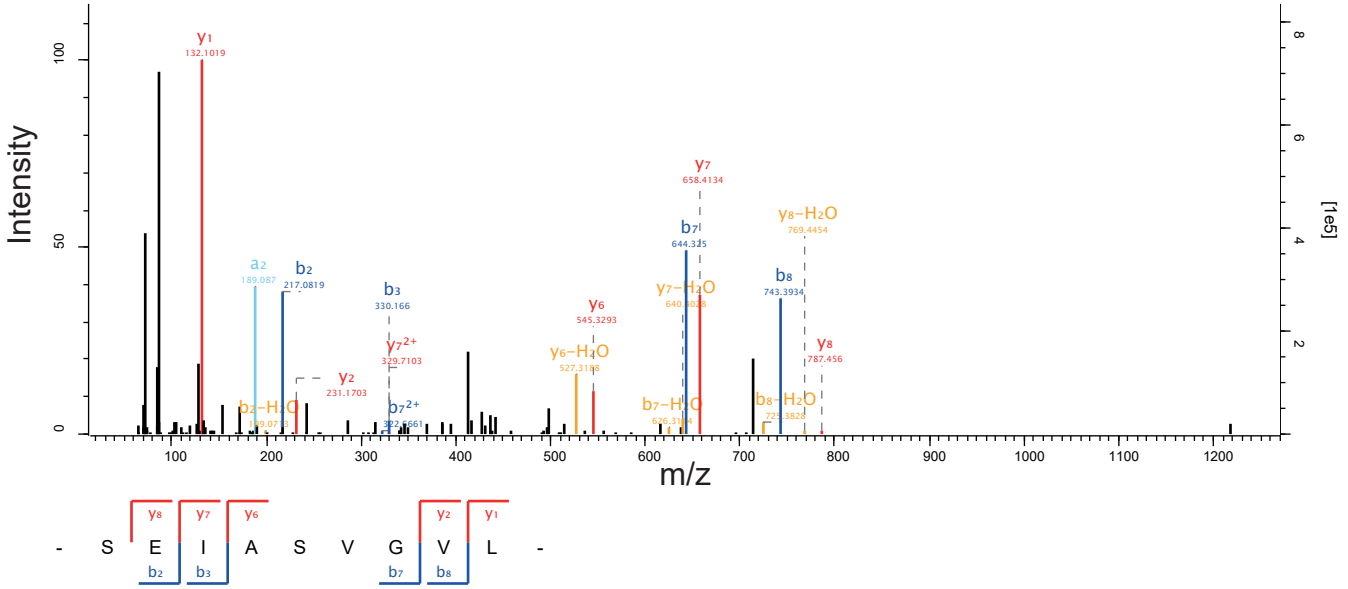
42RF Campylobacter concisus HLA-I NGIGVIAEV

Endogenous peptide



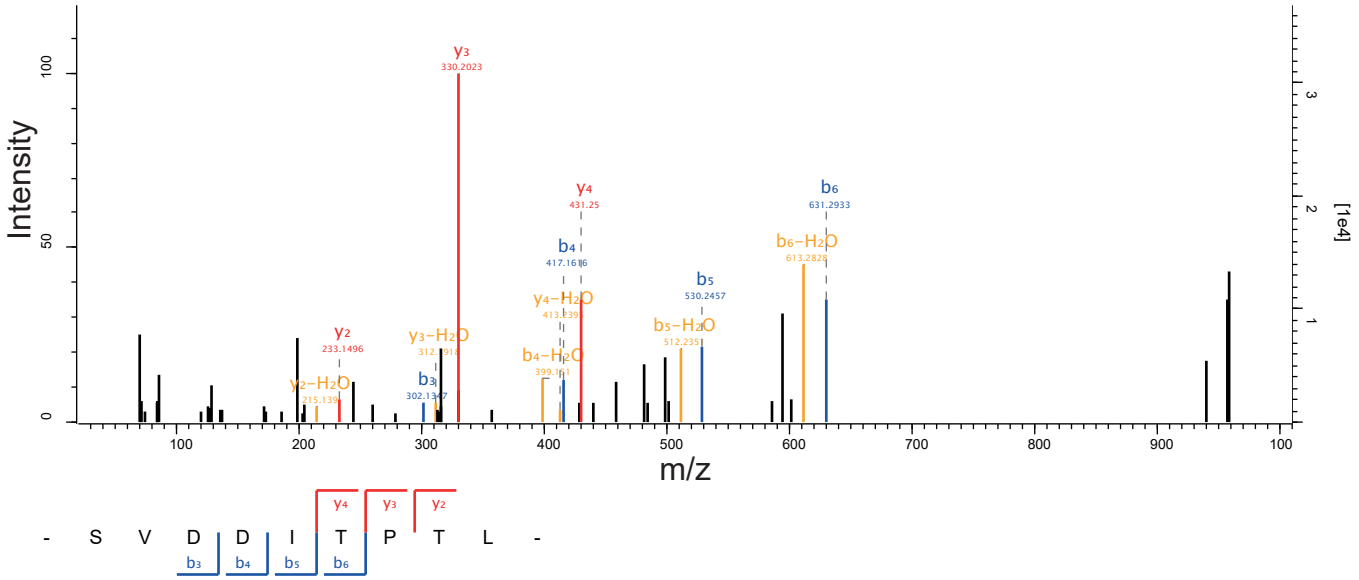
42RF Campylobacter concisus HLA-I SEIASVGVL

Endogenous peptide



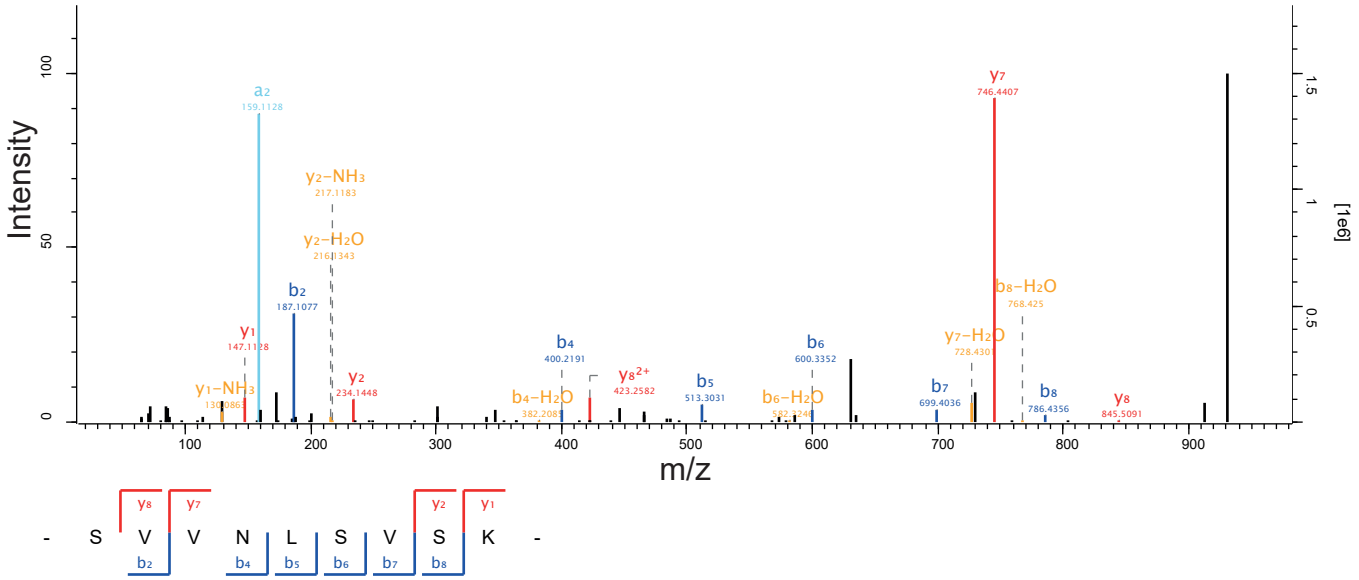
42RF Campylobacter concisus HLA-I SVDDITPTL

Endogenous peptide



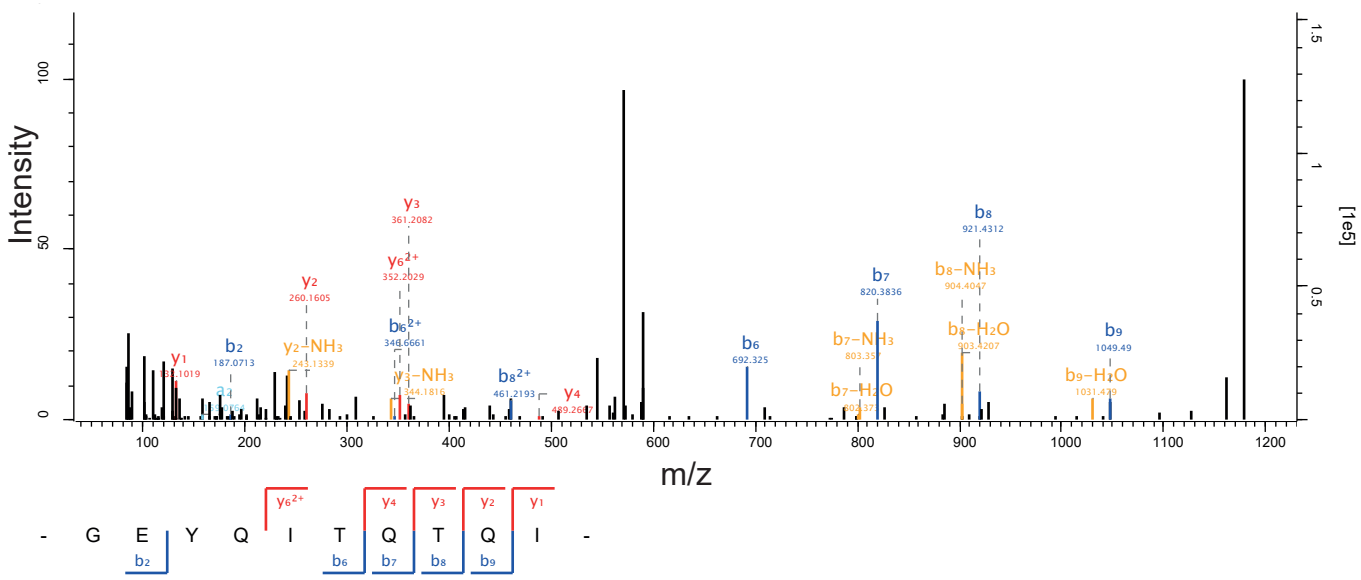
42RF Campylobacter concisus HLA-I SVVNLSVSK

Endogenous peptide



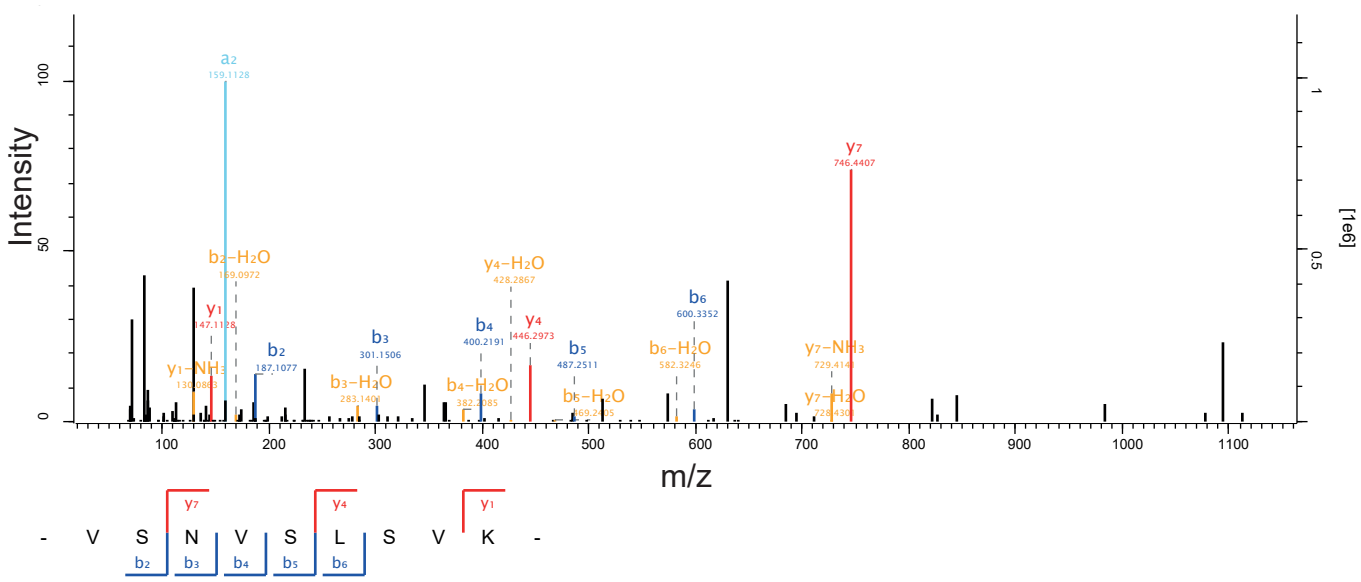
42RF Campylobacter lanienae HLA-I GEYQITQTQI

Endogenous peptide



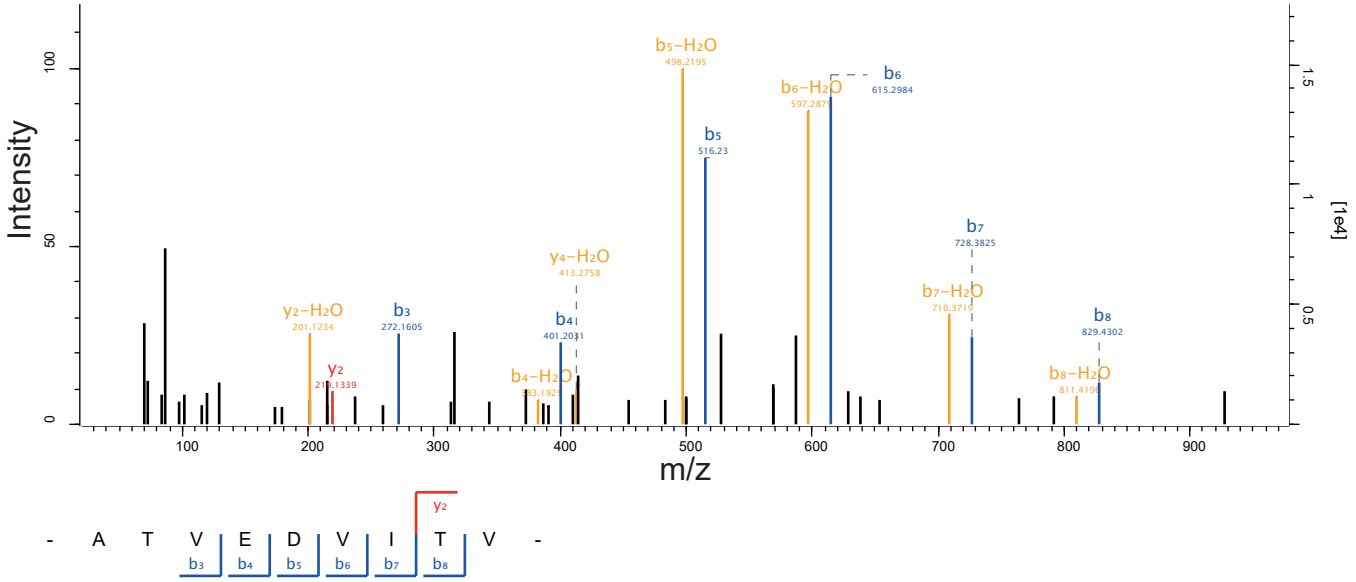
42RF Campylobacter lanienae HLA-I VSNVSLSVK

Endogenous peptide



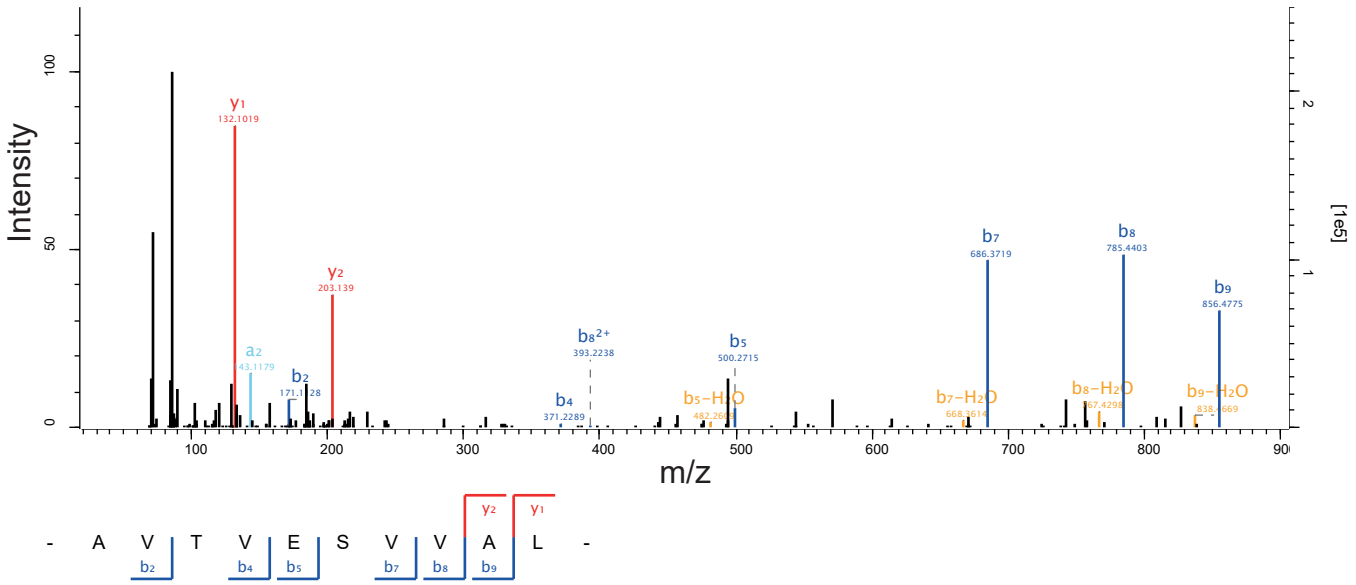
42RF Klebsiella oxytoca HLA-I ATVEDVITV

Endogenous peptide



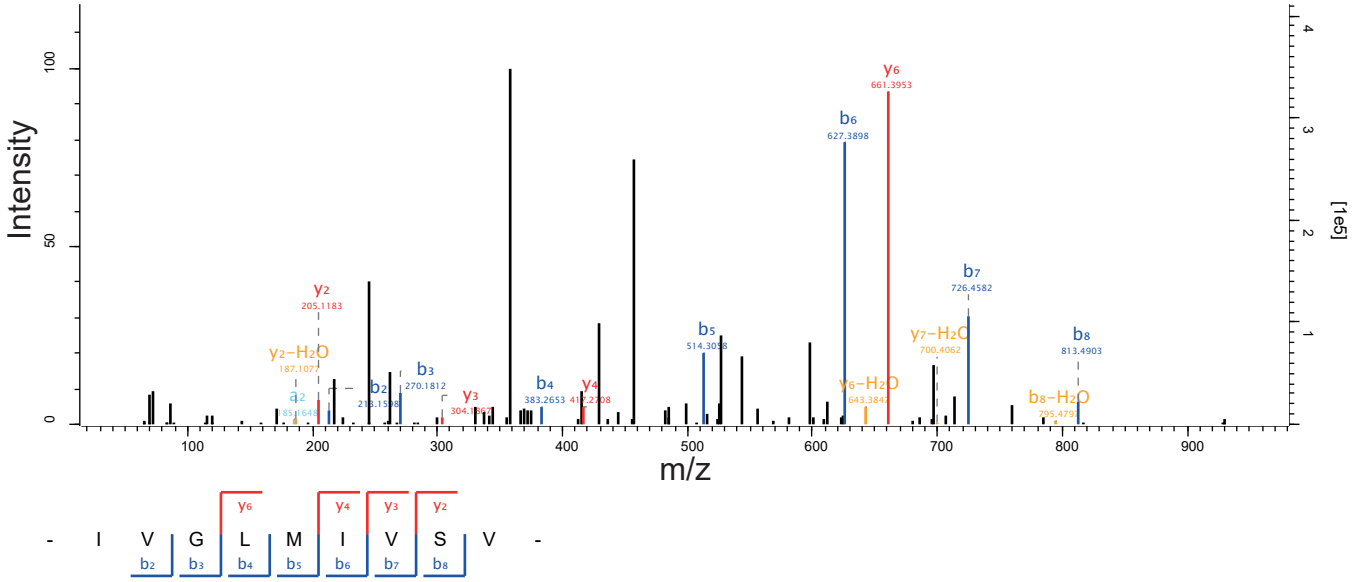
42RF Klebsiella oxytoca HLA-I AVTVESVVAL

Endogenous peptide



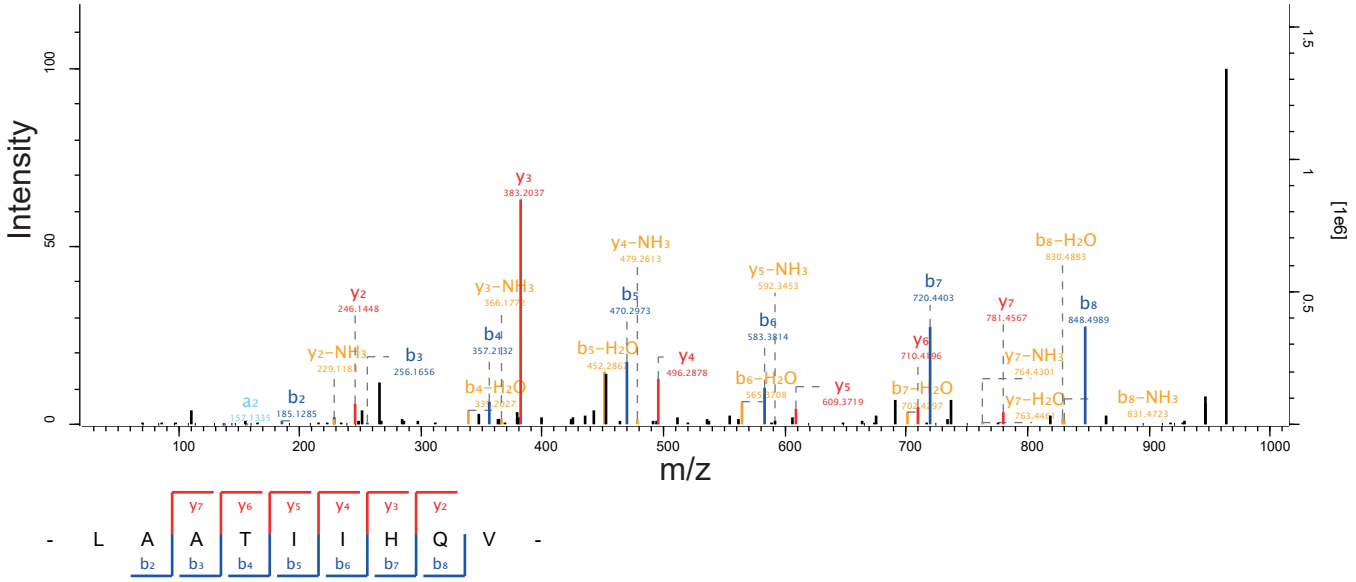
42RF Klebsiella oxytoca HLA-I IVGLMIVSV

Endogenous peptide



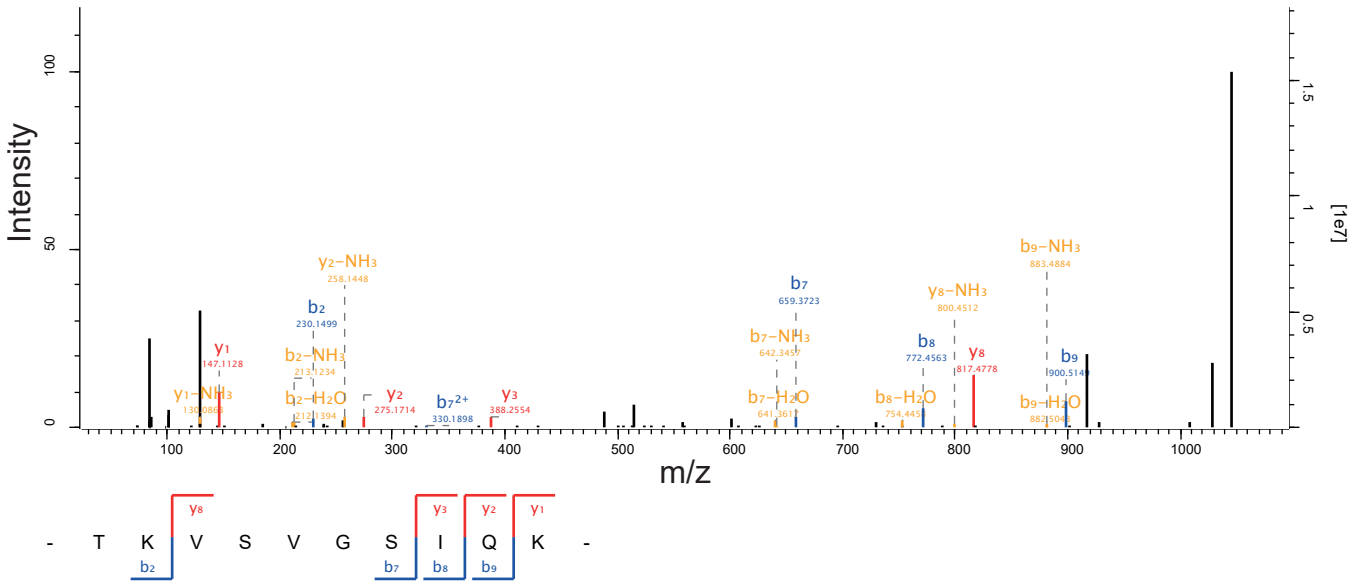
42RF Klebsiella oxytoca HLA-I LAATIIHQV

Endogenous peptide



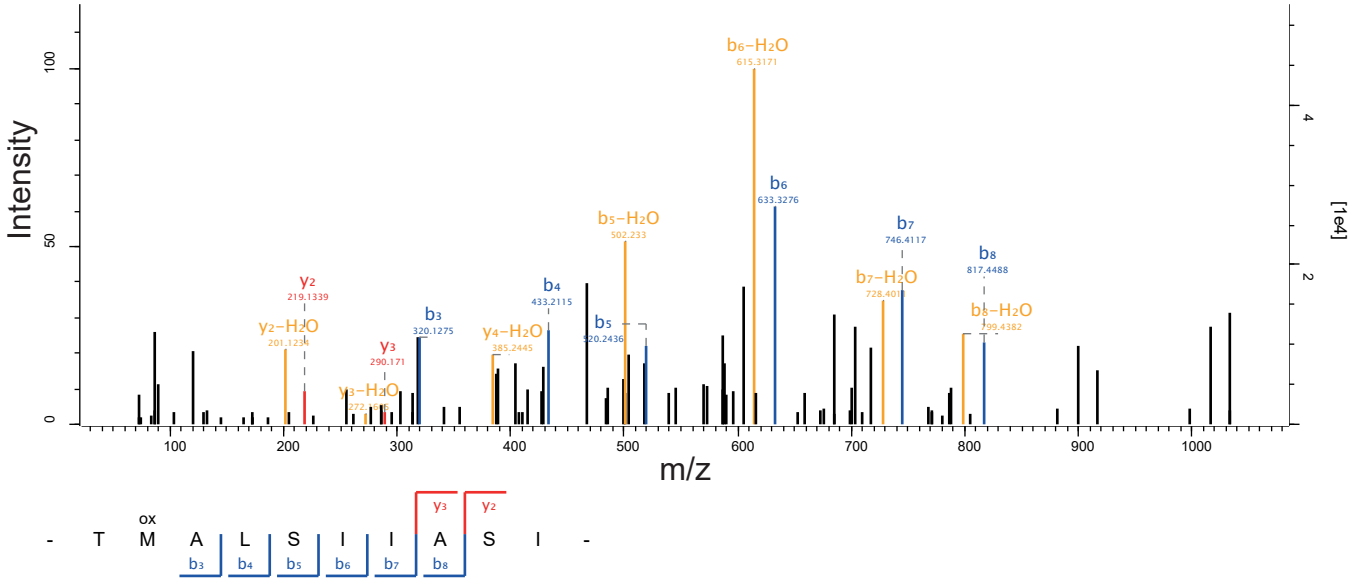
42RF Klebsiella oxytoca HLA-I TKVSVGSIQK

Endogenous peptide



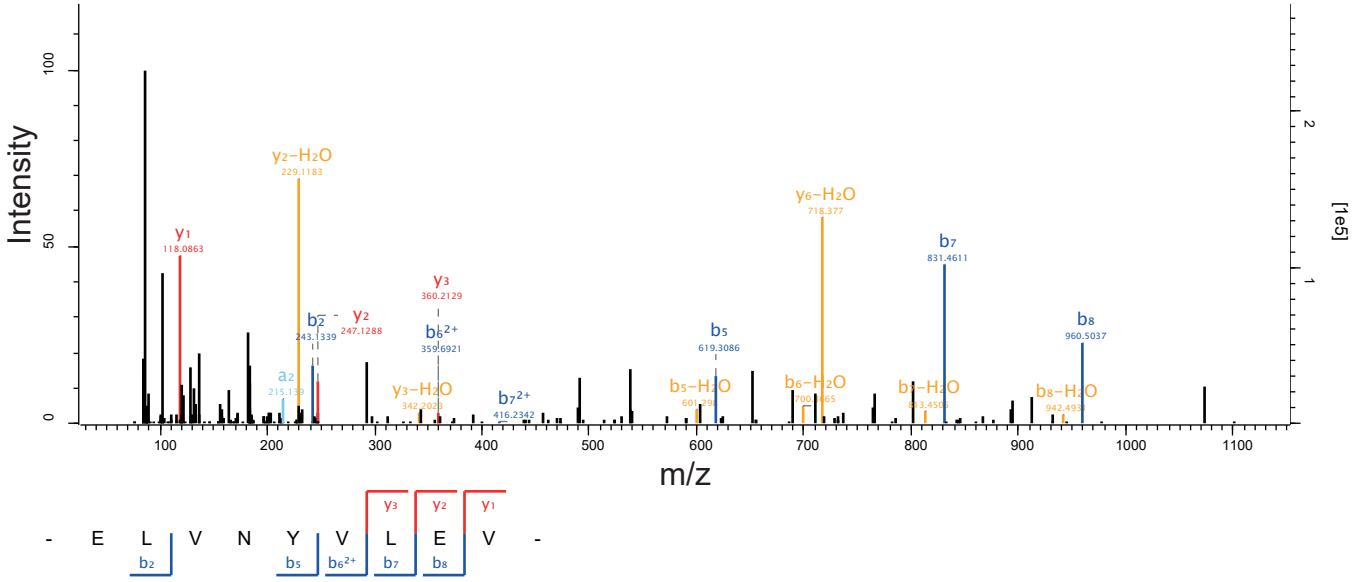
42RF Klebsiella oxytoca HLA-I TMALSIIASI

Endogenous peptide



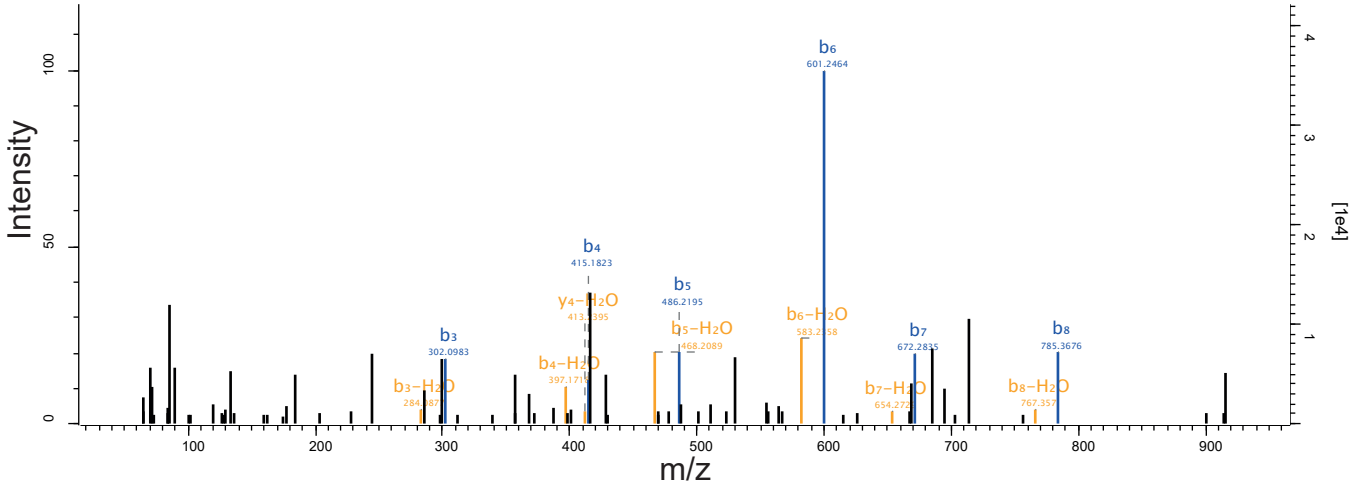
42RF Paracoccus marcusii HLA-I ELVNYVLEV

Endogenous peptide



42RF Shewanella decolorationis HLA-I ADDLADAIL

Endogenous peptide

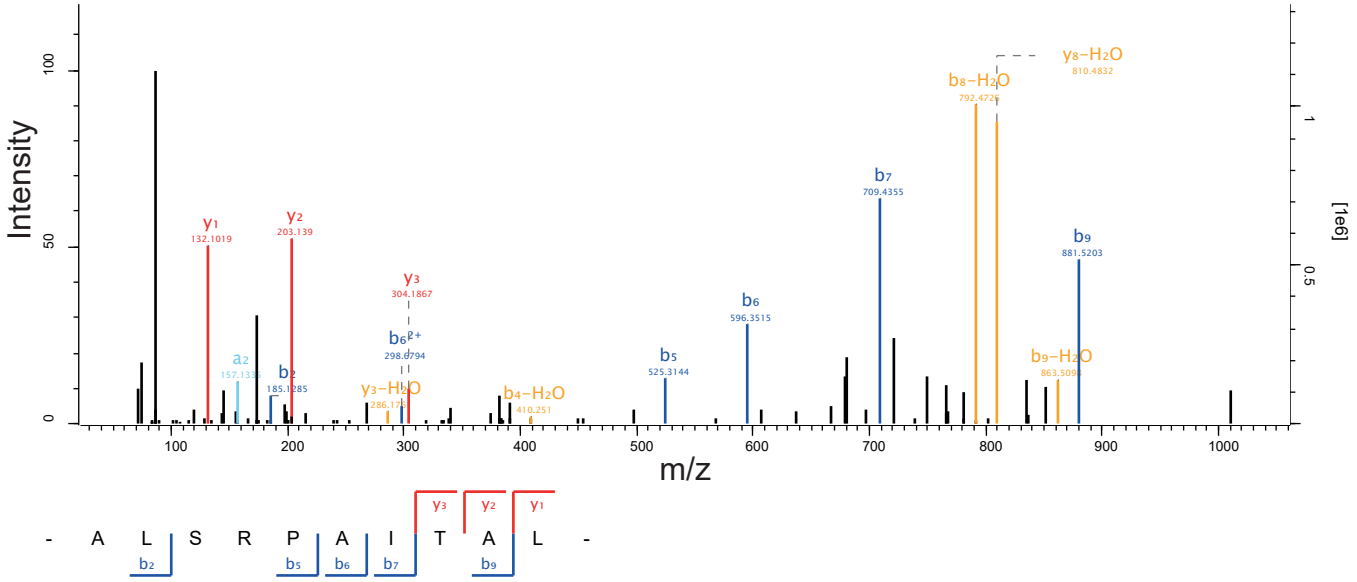


- A D D L A D A I L -

b3 b4 b5 b6 b7 b8

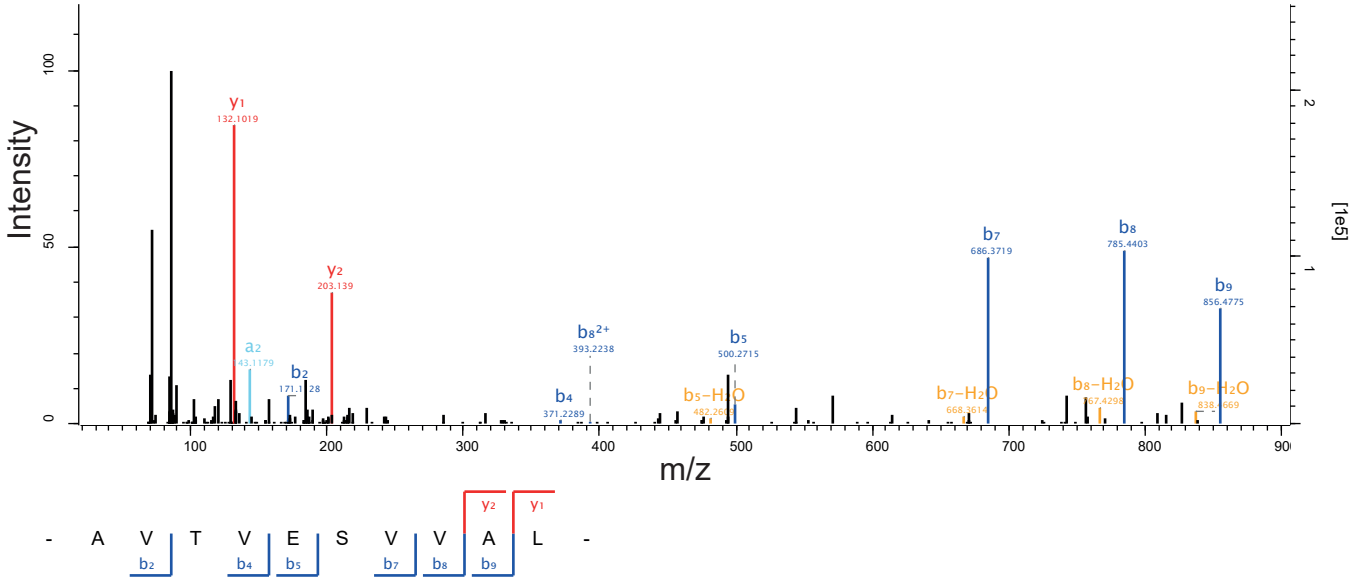
42RF Shewanella decolorationis HLA-I ALSRPAITAL

Endogenous peptide



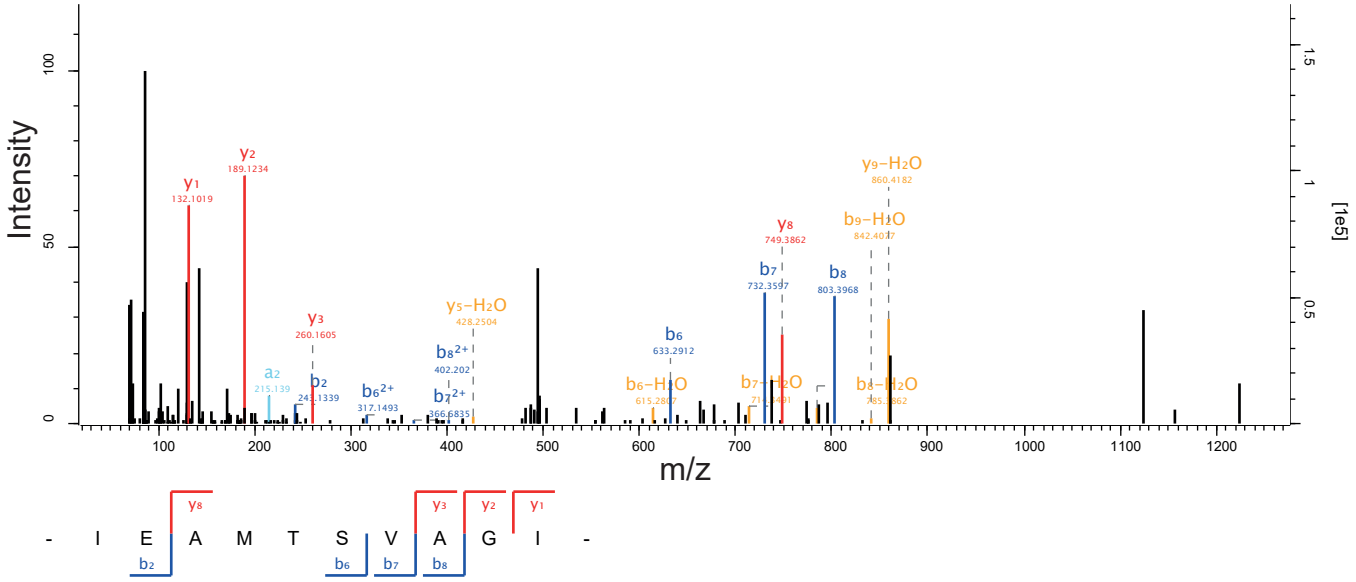
42RF Shewanella decolorationis HLA-I AVTVESVVAL

Endogenous peptide



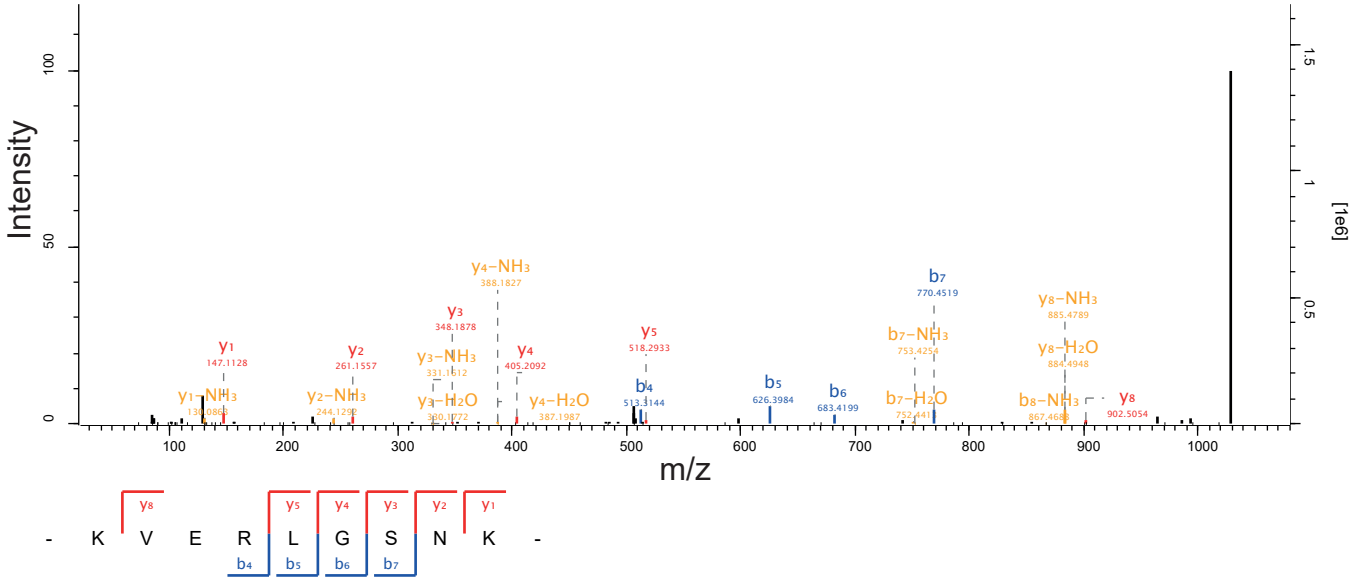
42RF Shewanella decolorationis HLA-I IEAMTSVAGI

Endogenous peptide



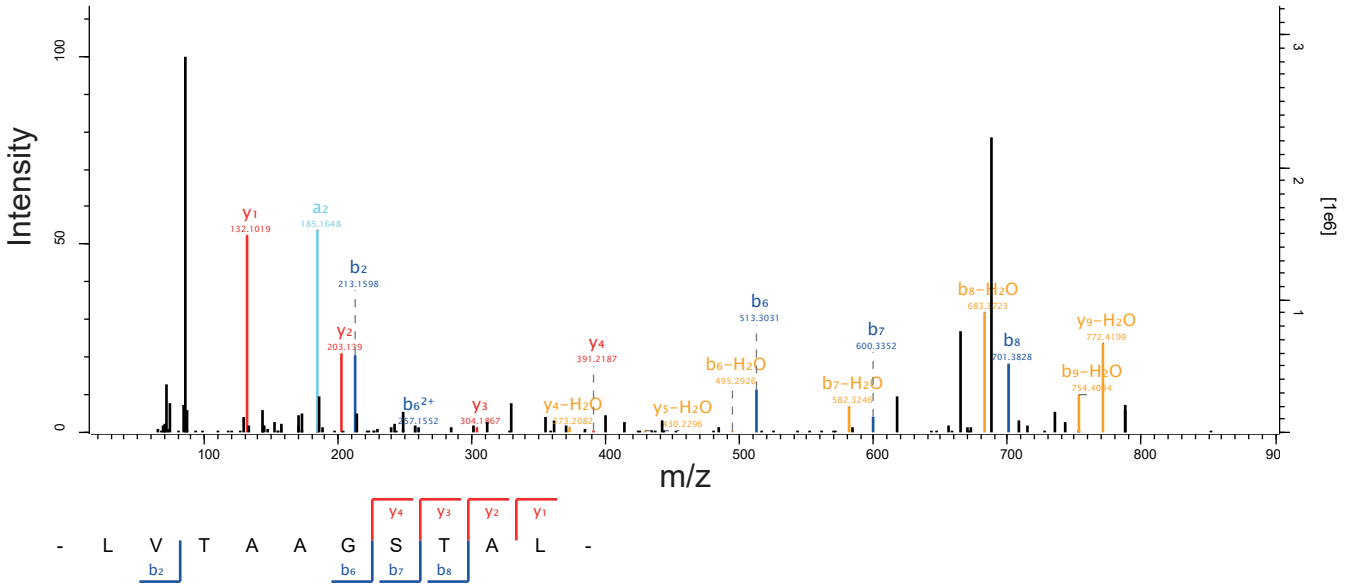
42RF Shewanella decolorationis HLA-I KVERLGSNK

Endogenous peptide



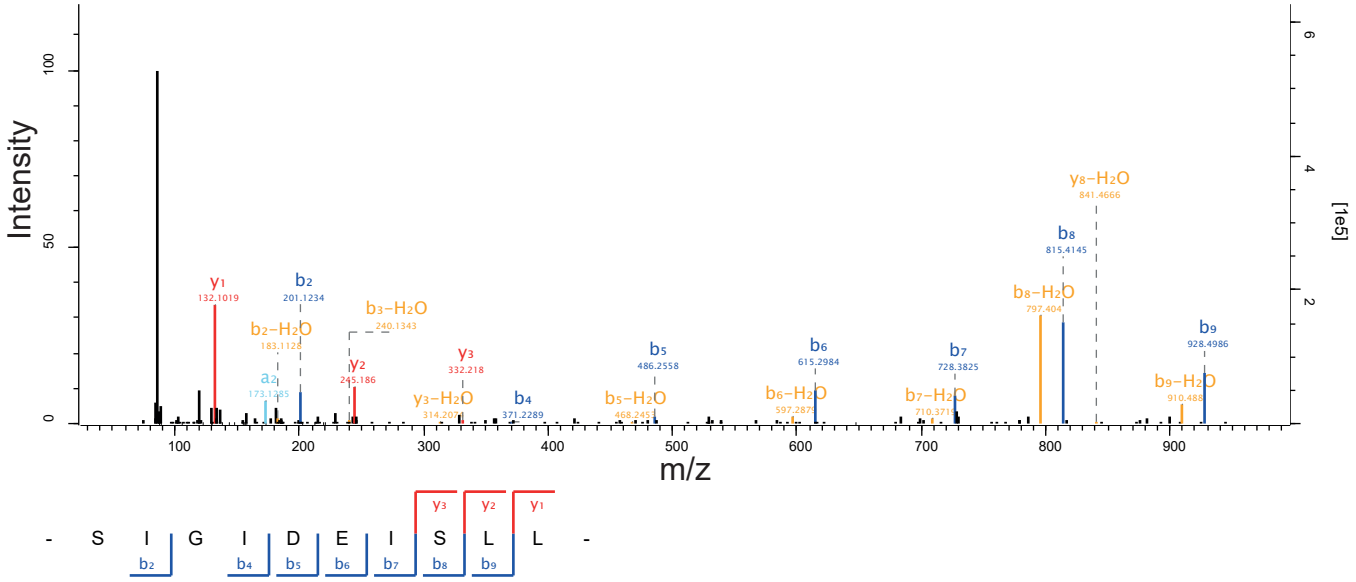
42RF Shewanella decolorationis HLA-I LVTAAGSTAL

Endogenous peptide



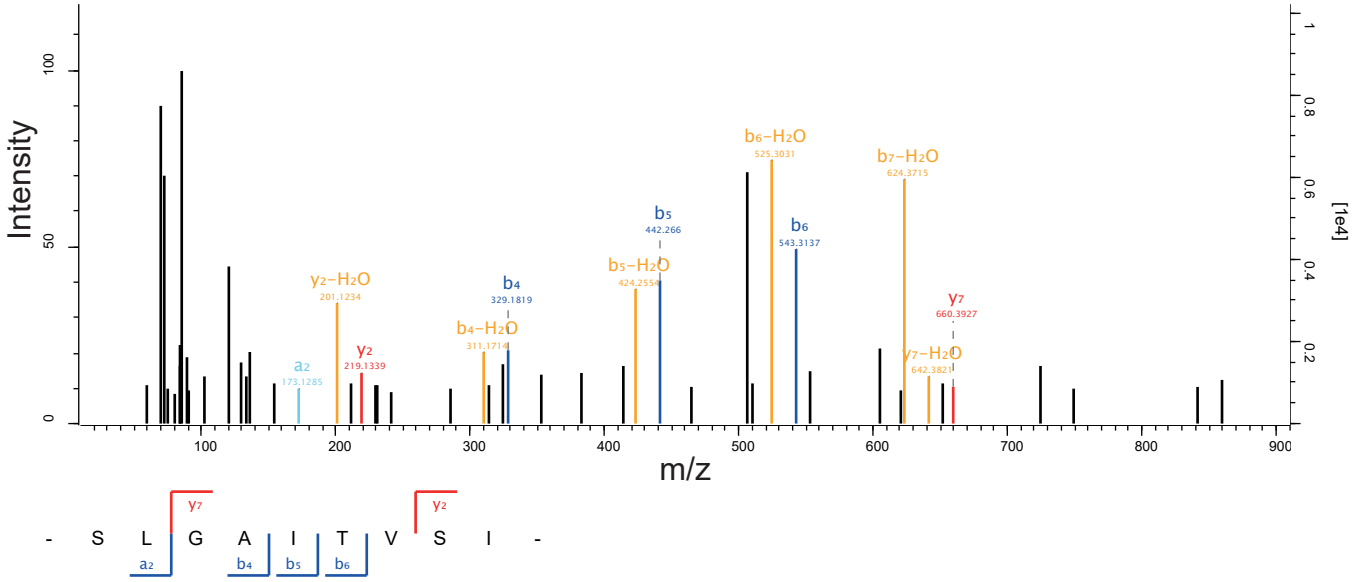
42RF Shewanella decolorationis HLA-I SIGIDEISLL

Endogenous peptide



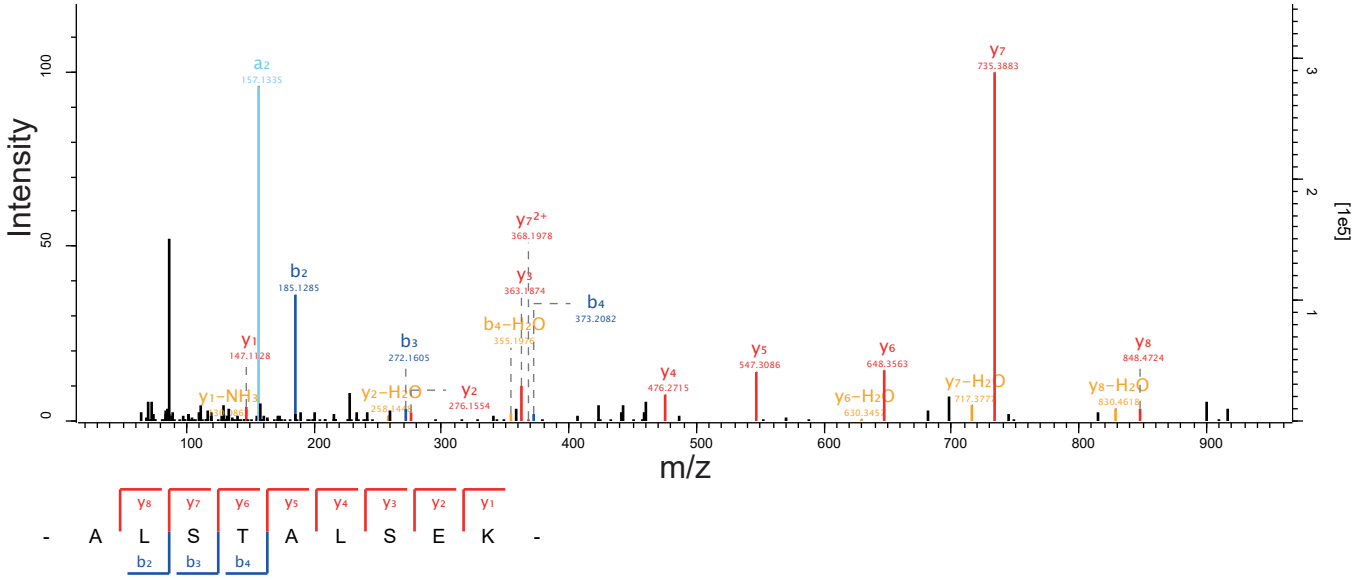
42RF Shewanella decolorationis HLA-I SLGAITVSI

Endogenous peptide



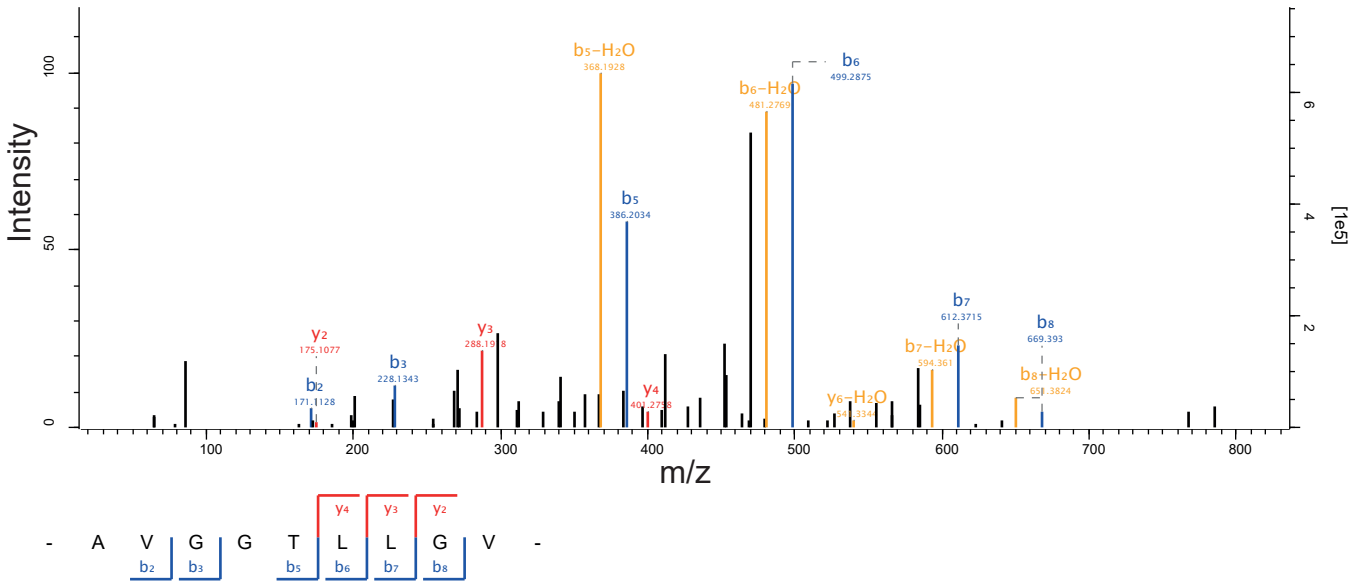
42RS Enterobacter cancerogenus HLA-I ALSASDIAL

Endogenous peptide



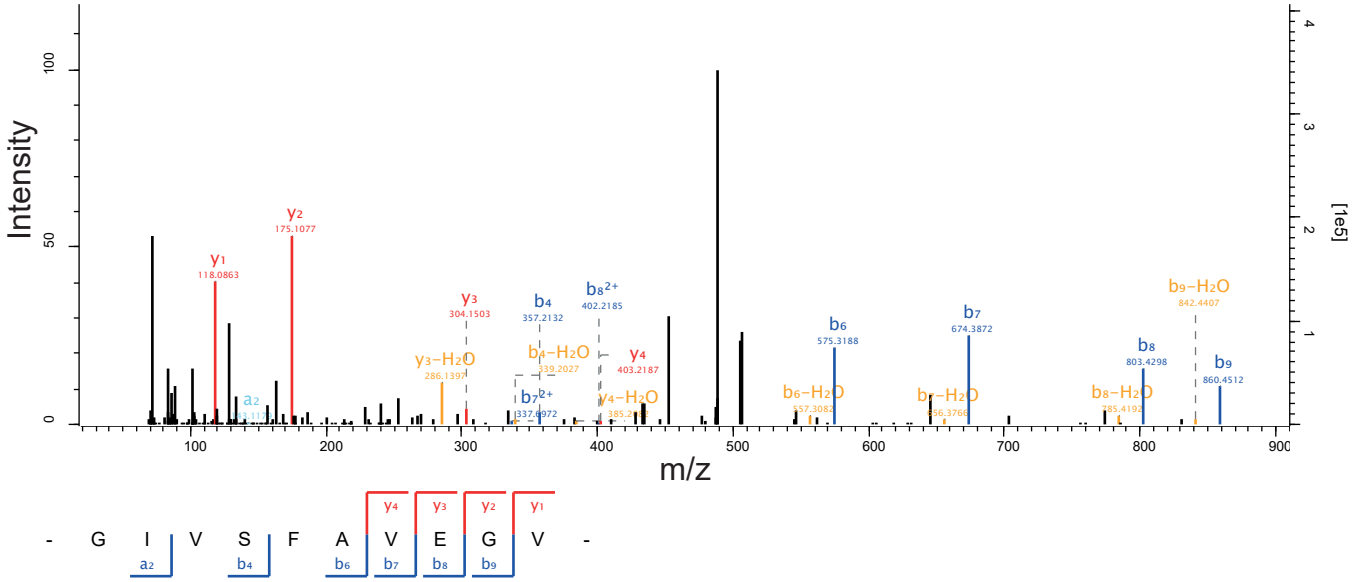
42RS Sphingomonas melonis HLA-I AVGGTLLGV

Endogenous peptide



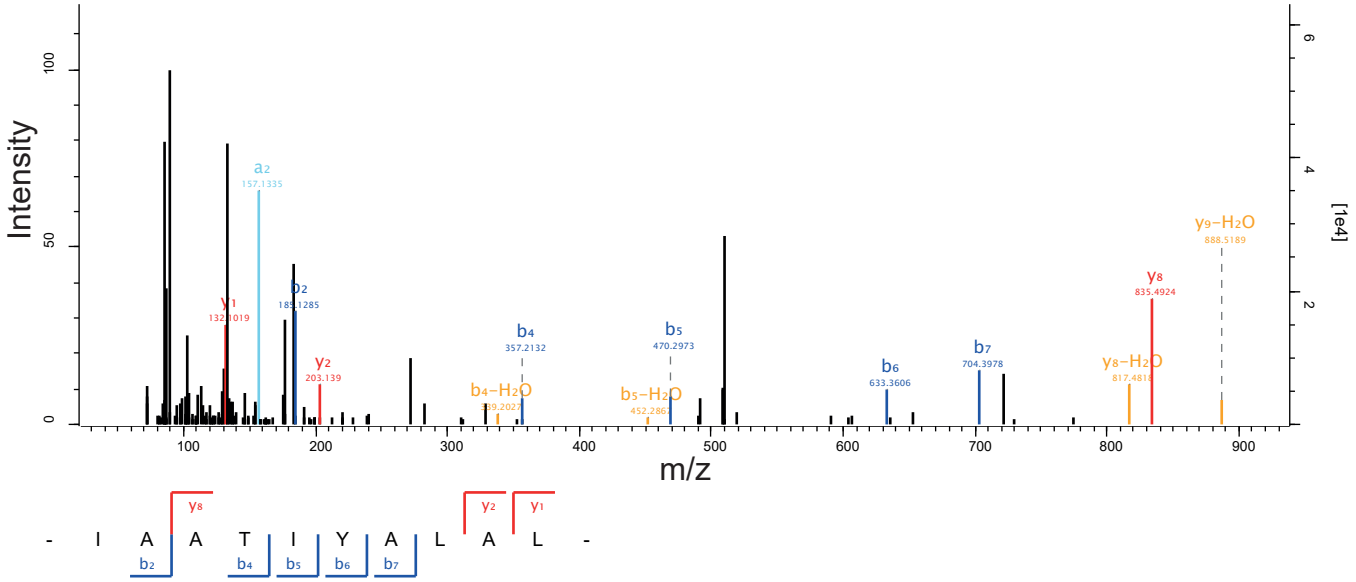
42RS Sphingomonas melonis HLA-I GIVSFAVEGV

Endogenous peptide



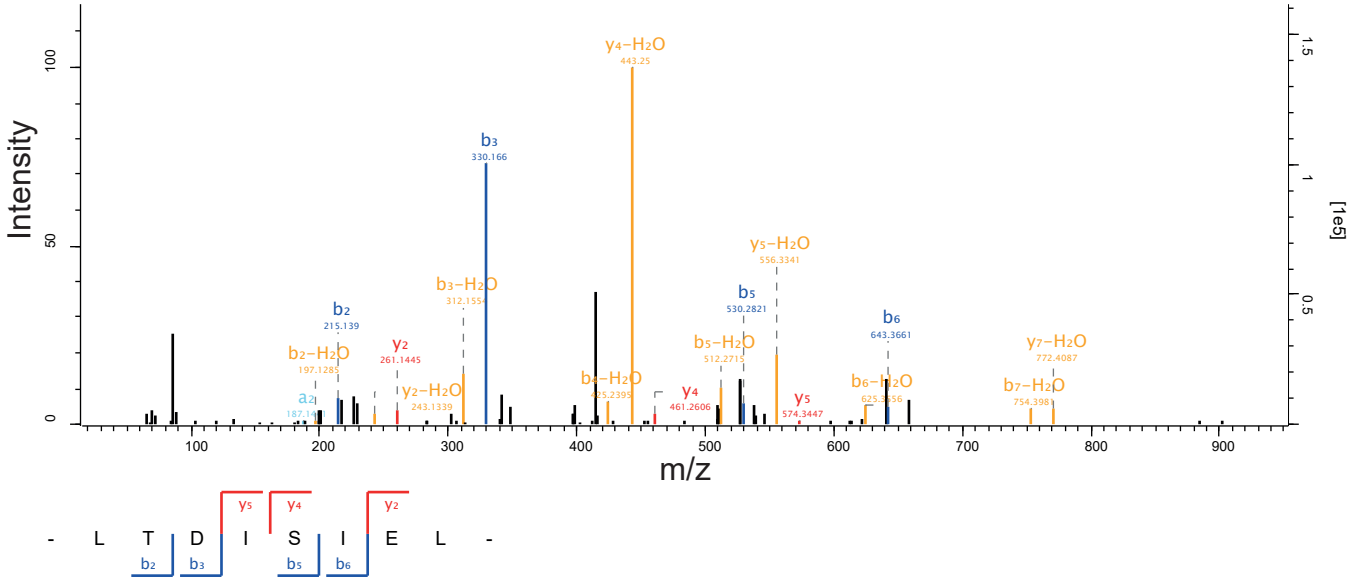
42RS Sphingomonas melonis HLA-I IAATIYALAL

Endogenous peptide



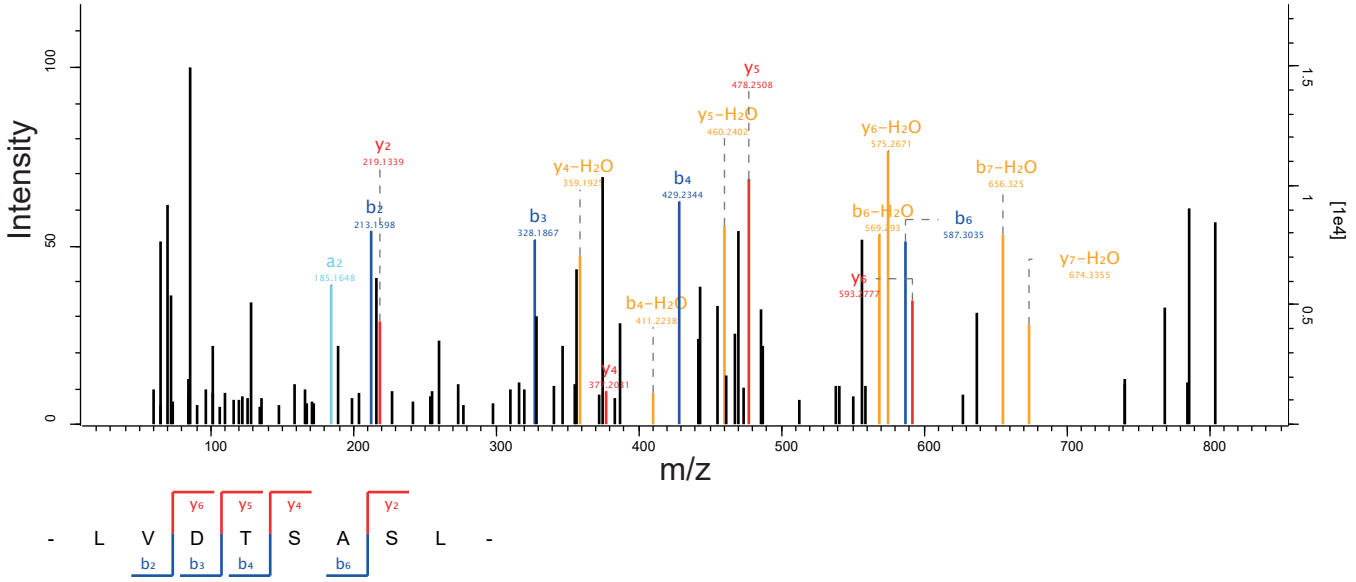
51AL Corynebacterium afermentans HLA-I LTDISIEL

Endogenous peptide



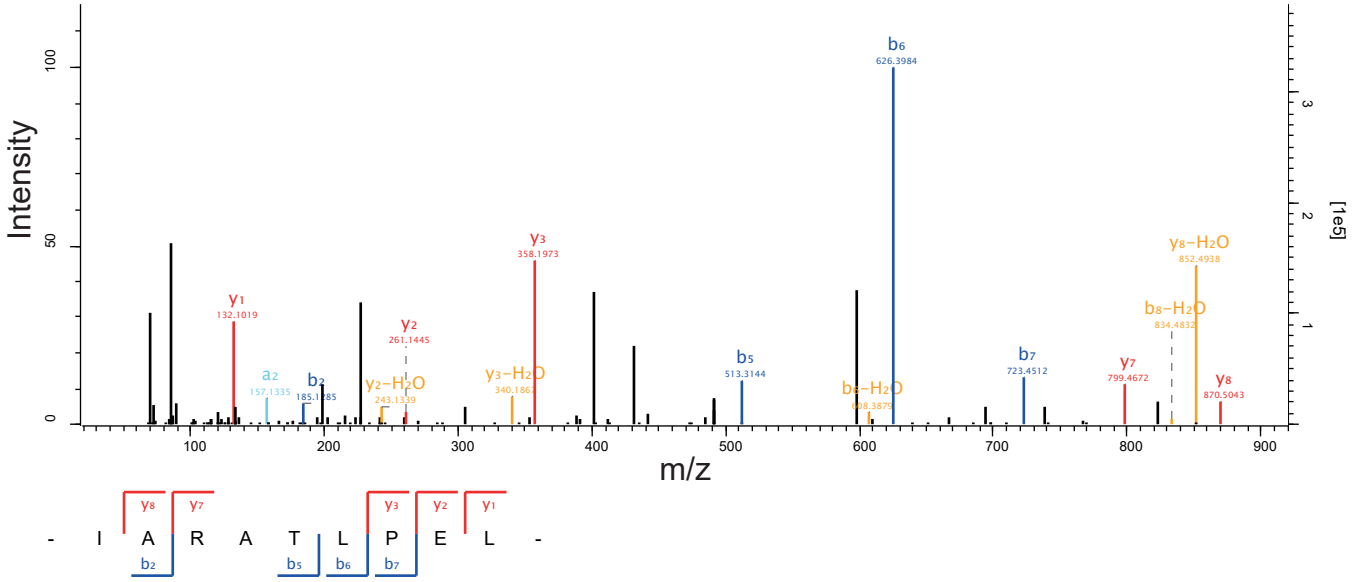
51AL Corynebacterium afermentans HLA-I LVDTSASL

Endogenous peptide



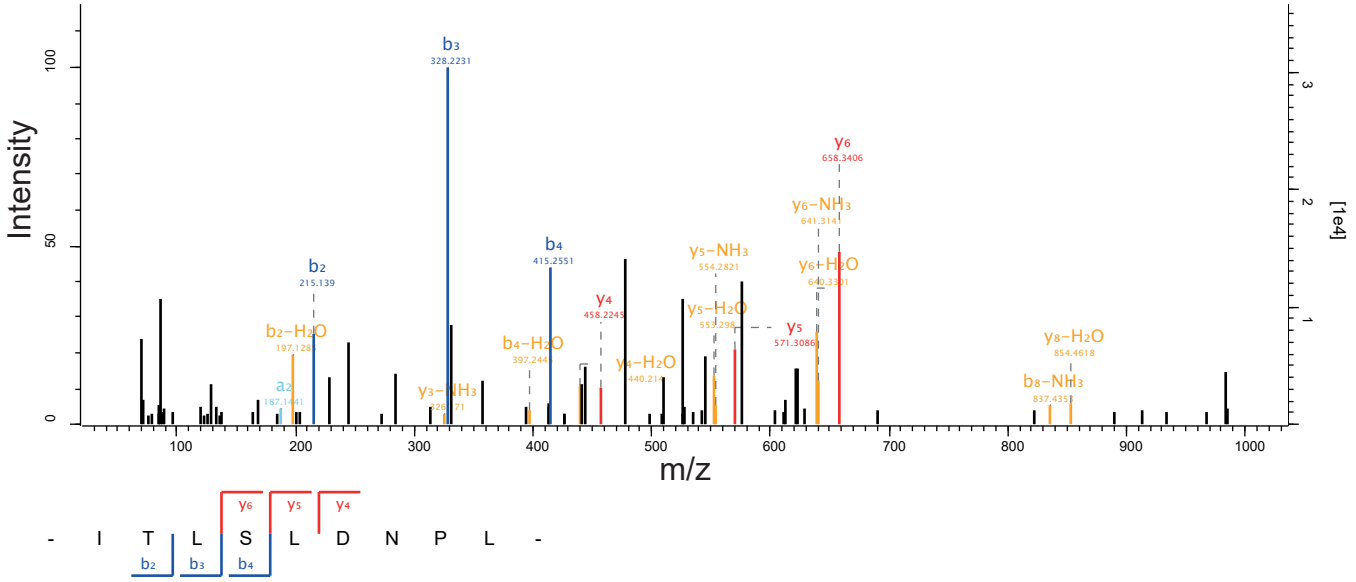
51AL Eikenella corrodens HLA-I IARATLPEL

Endogenous peptide



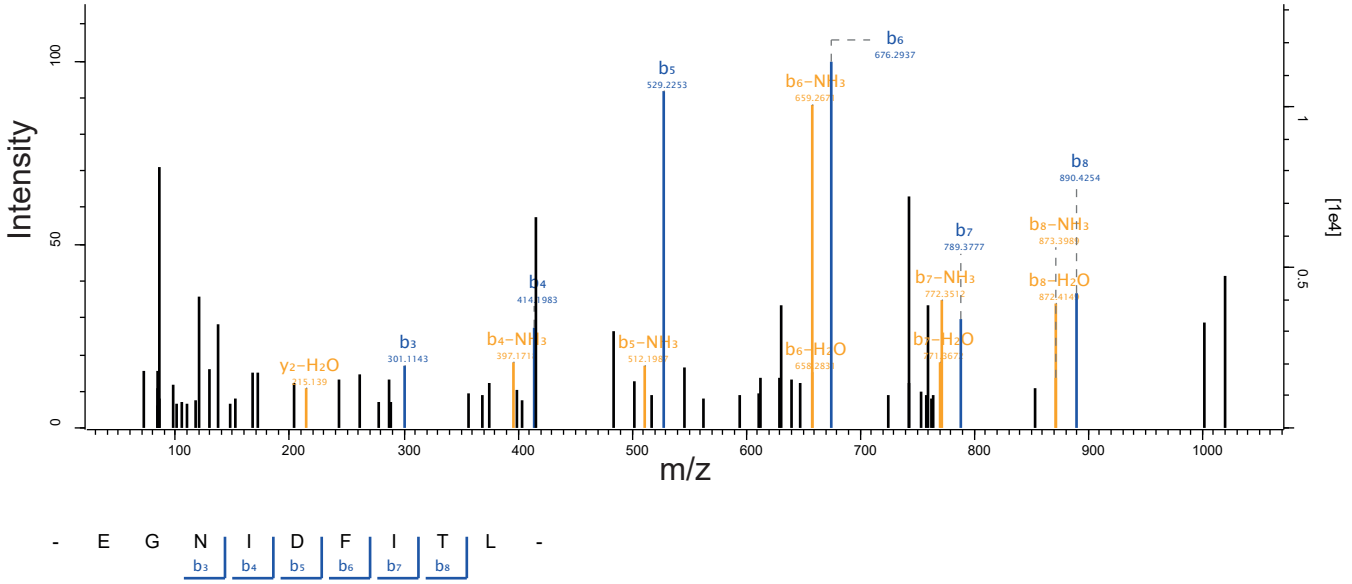
51AL Eikenella corrodens HLA-I ITLSLDNPL

Endogenous peptide



51AL Staphylococcus aureus HLA-I EGNIDFITL

Endogenous peptide

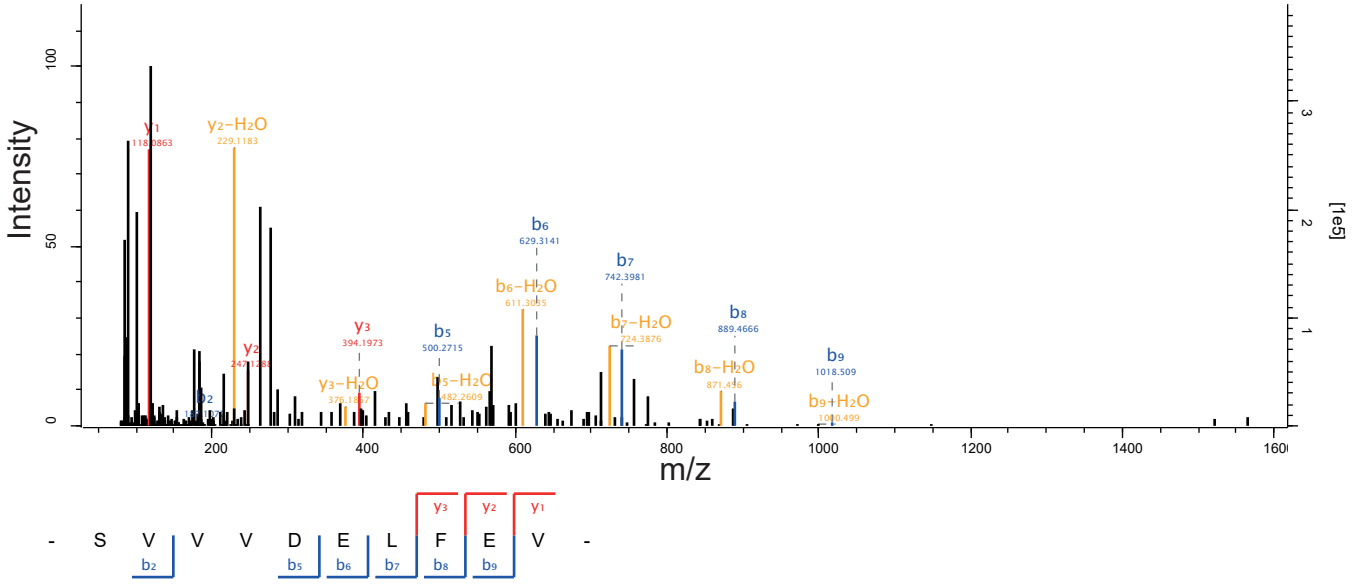


- E G N I D F I T L -

b3 b4 b5 b6 b7 b8

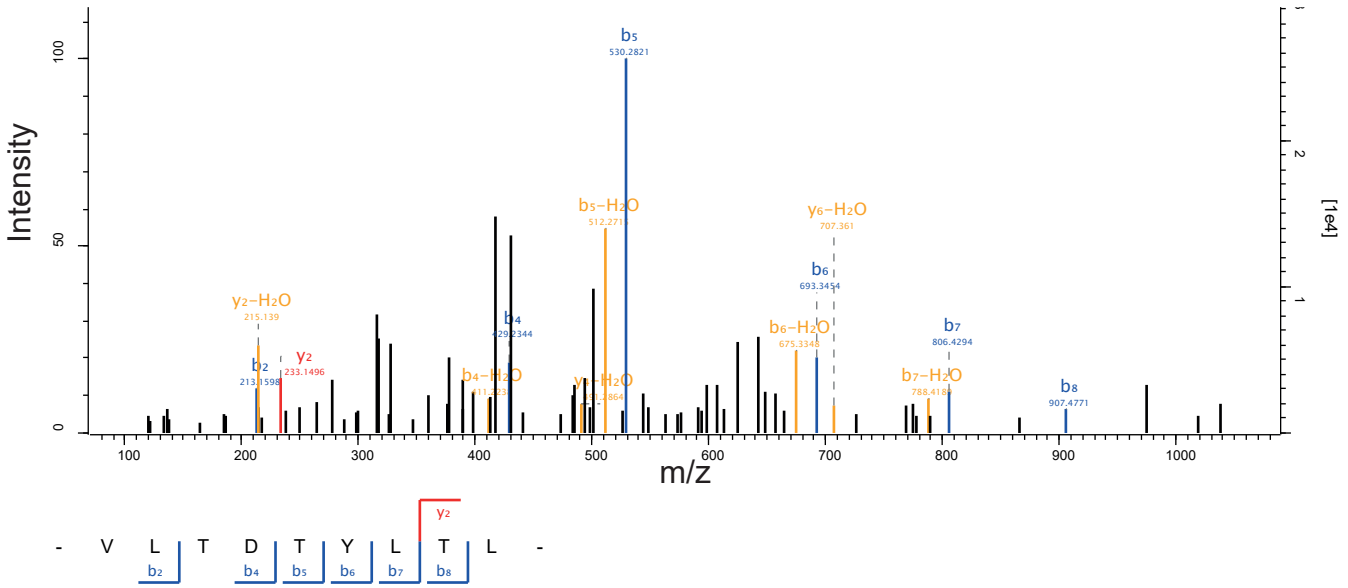
51BR-1 Fusobacterium nucleatum HLA-I SVVVDELFEV

Endogenous peptide



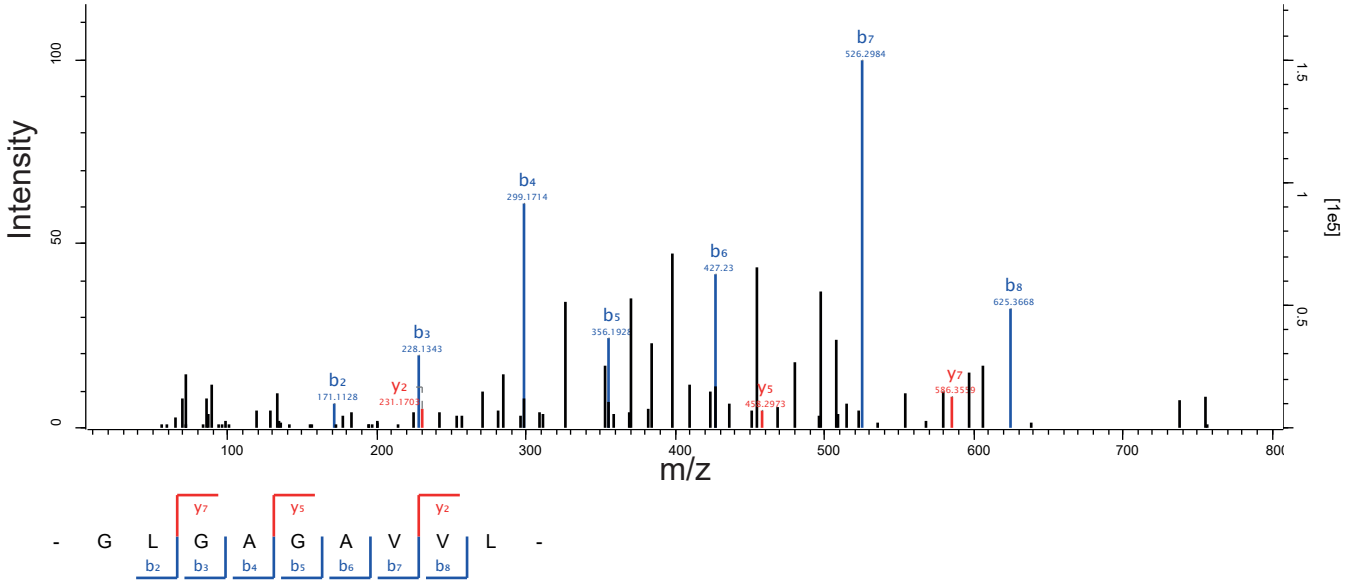
51BR-1 Fusobacterium nucleatum HLA-I VLTDTYLTL

Endogenous peptide



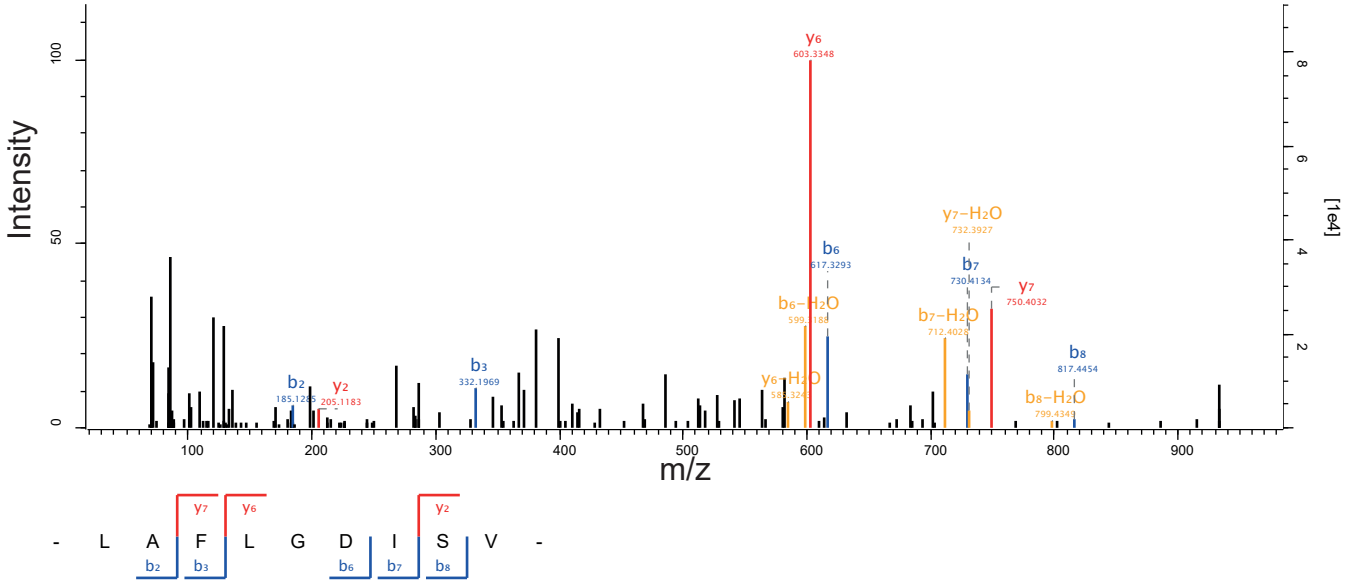
51BR-1 Brachybacterium alimentarium HLA-I GLGAGAVVL

Endogenous peptide



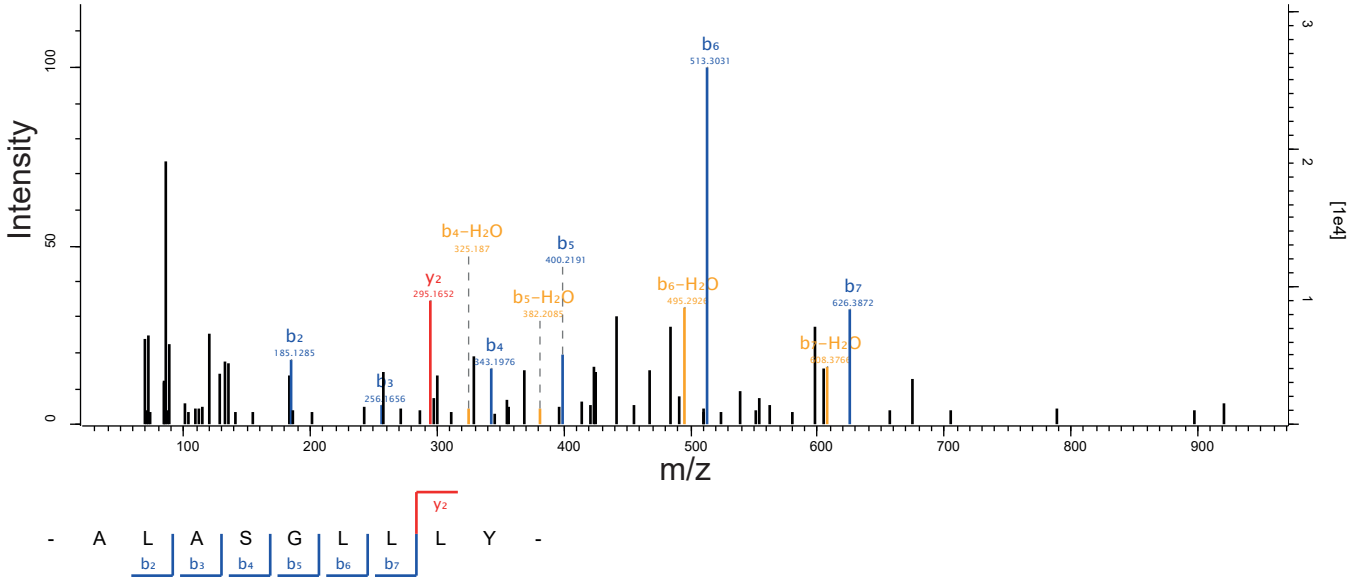
51BR-1 Brachybacterium alimentarium HLA-I LAFLGDISV

Endogenous peptide



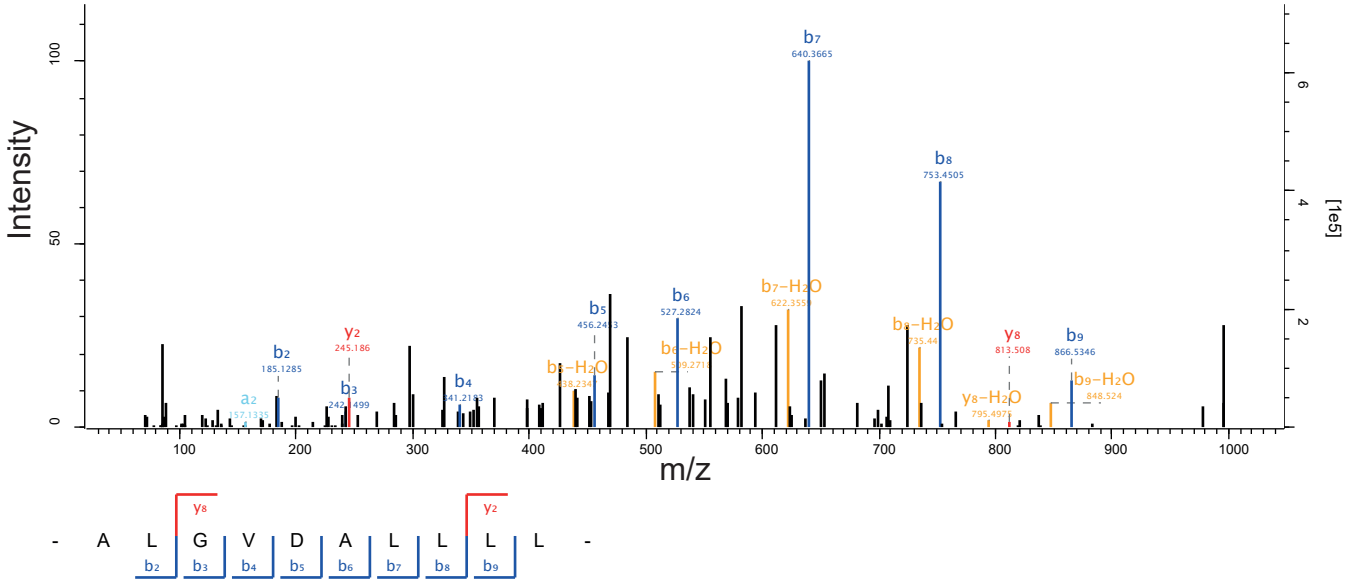
51BR-1 Eikenella corrodens HLA-I ALASGLLLY

Endogenous peptide



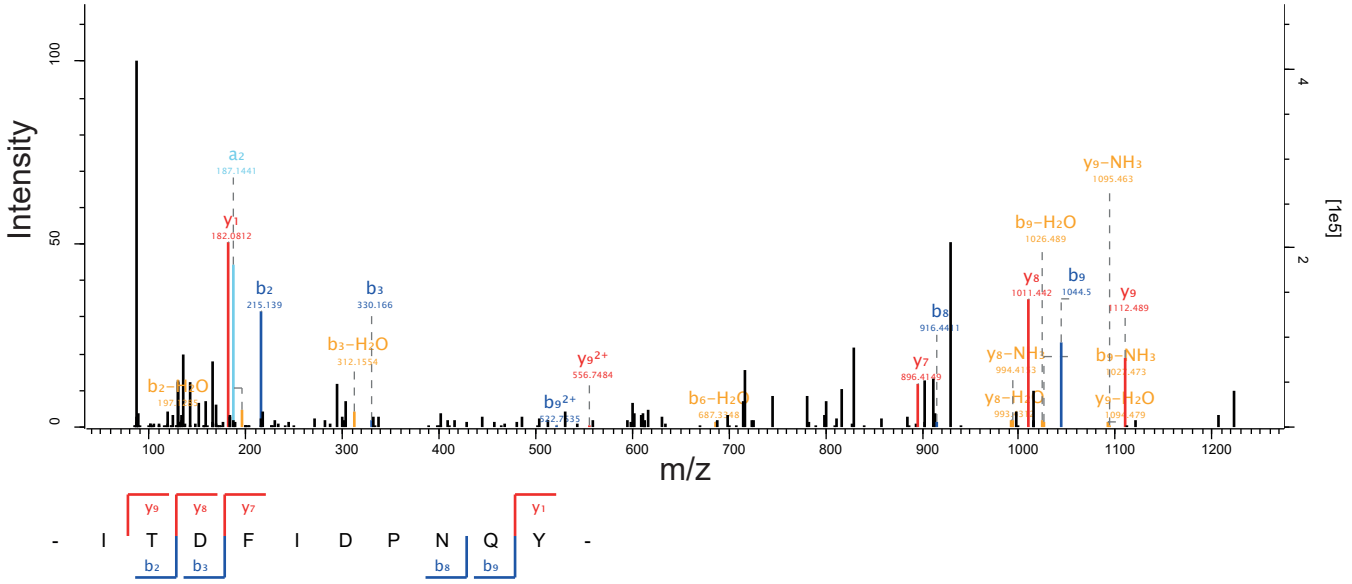
51BR-1 Staphylococcus aureus HLA-I ALGVDALLLL

Endogenous peptide



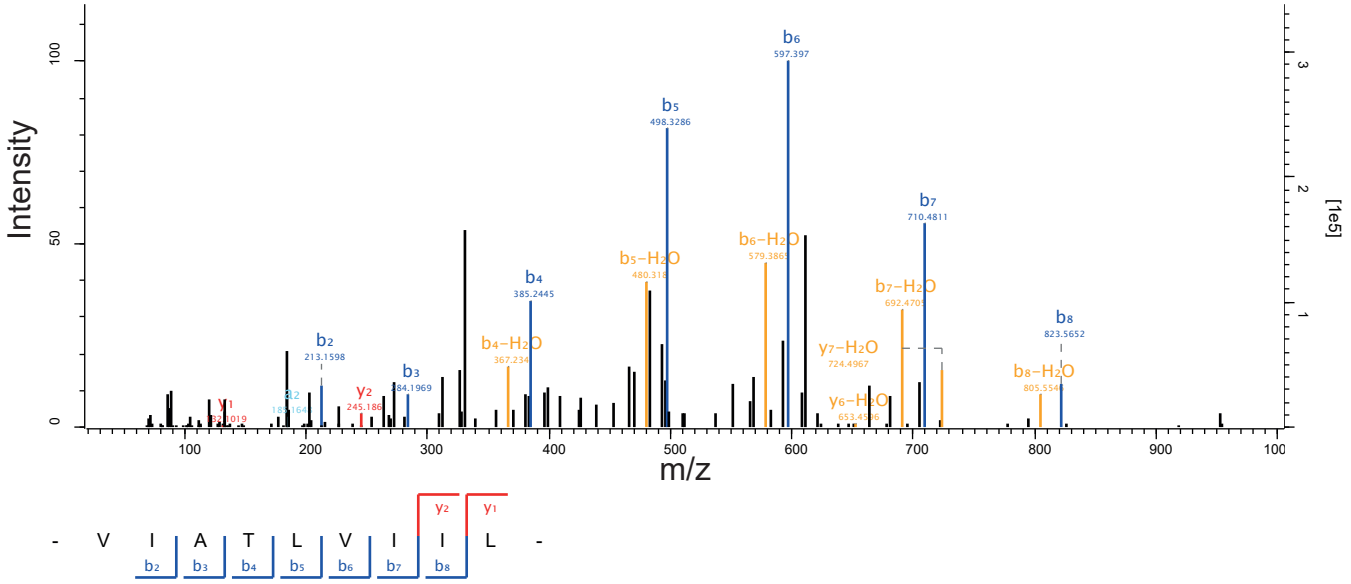
51BR-1 Staphylococcus aureus HLA-I ITDFIDPNQY

Endogenous peptide



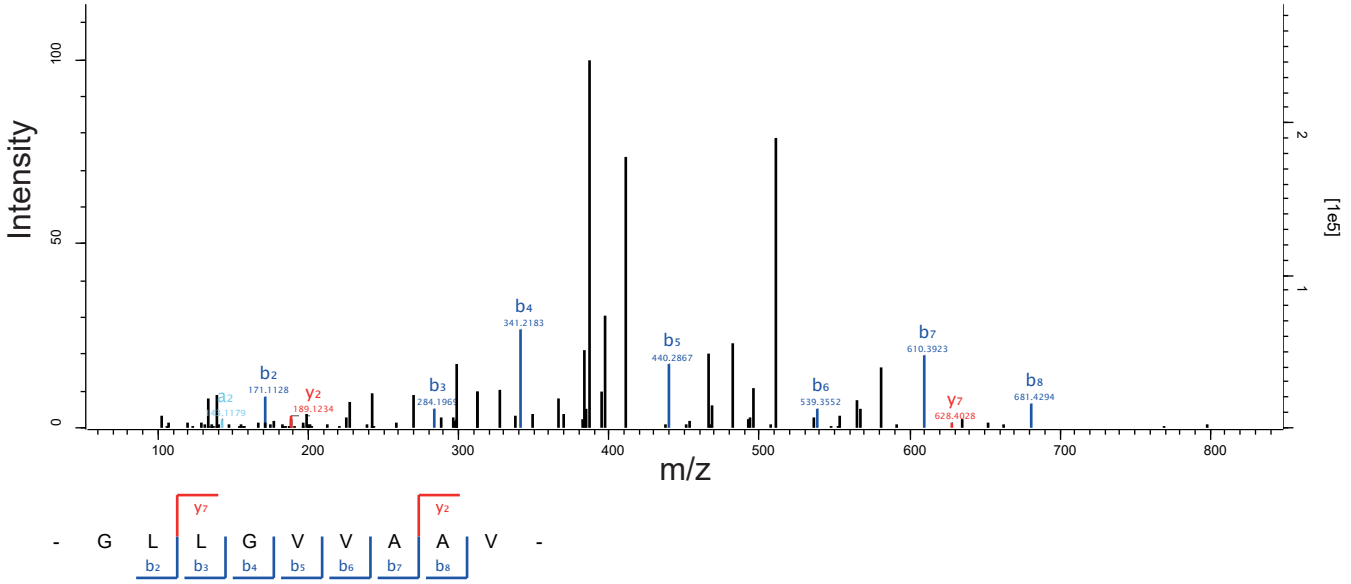
51BR-1 Staphylococcus aureus HLA-I VIATLVIIL

Endogenous peptide



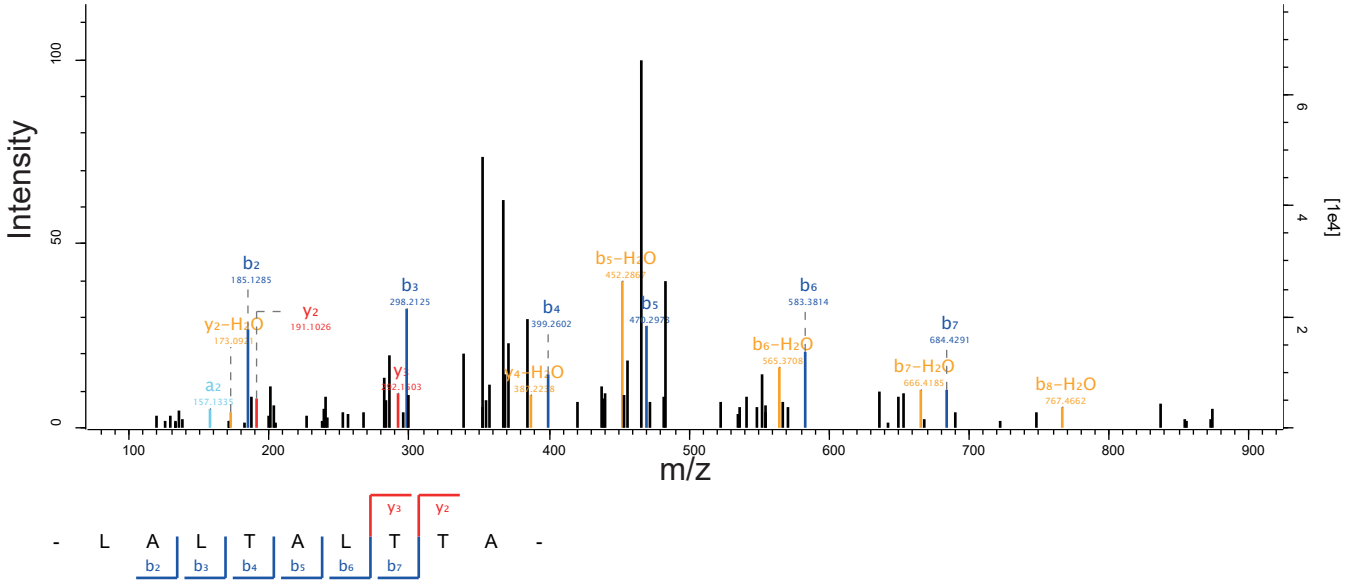
51BR-2 Brachybacterium alimentarium HLA-I GLLGVVAAV

Endogenous peptide



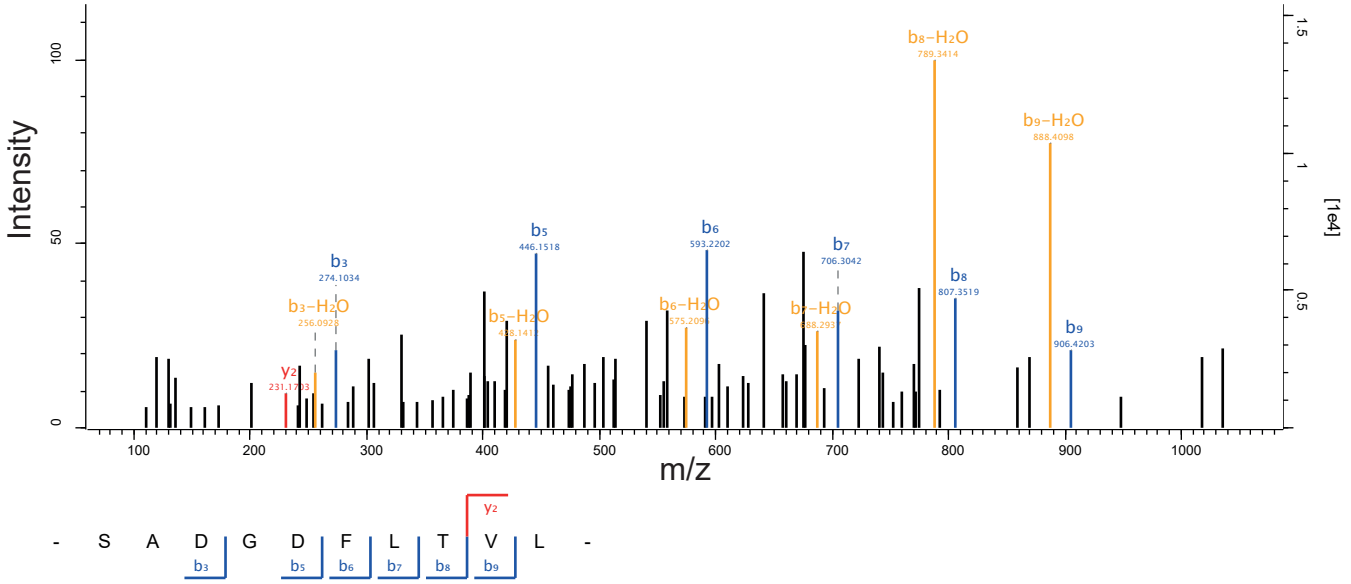
51BR-2 Brachybacterium alimentarium HLA-I LALTALTTA

Endogenous peptide



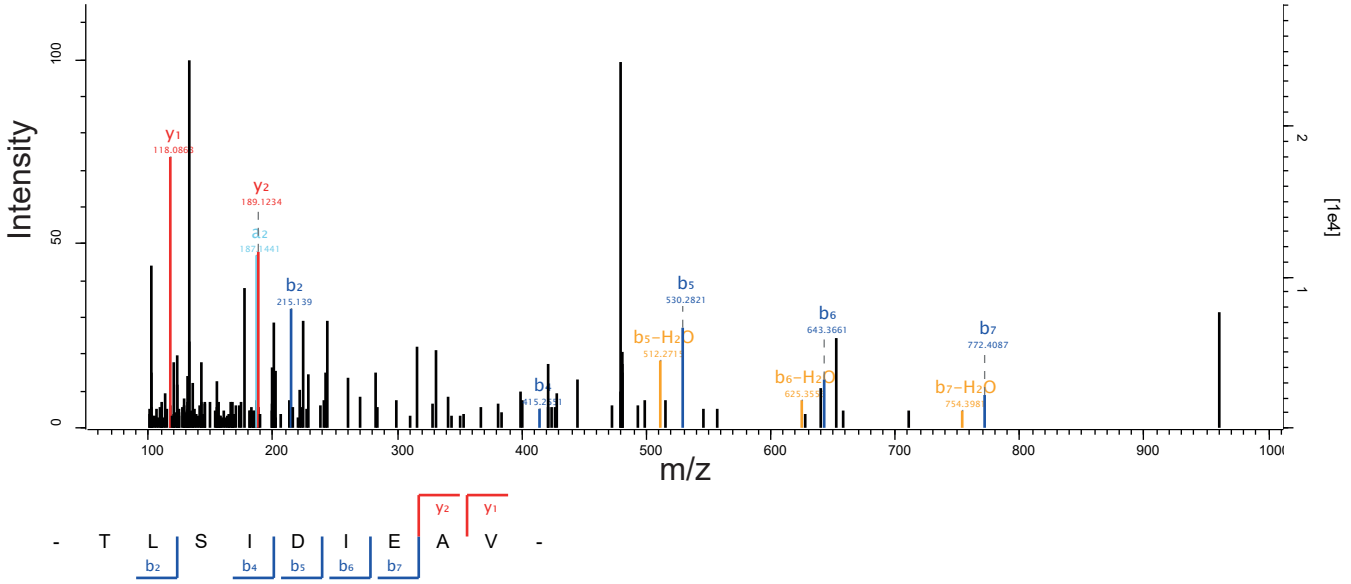
51BR-2 Brachybacterium alimentarium HLA-I SADGDFLTVL

Endogenous peptide



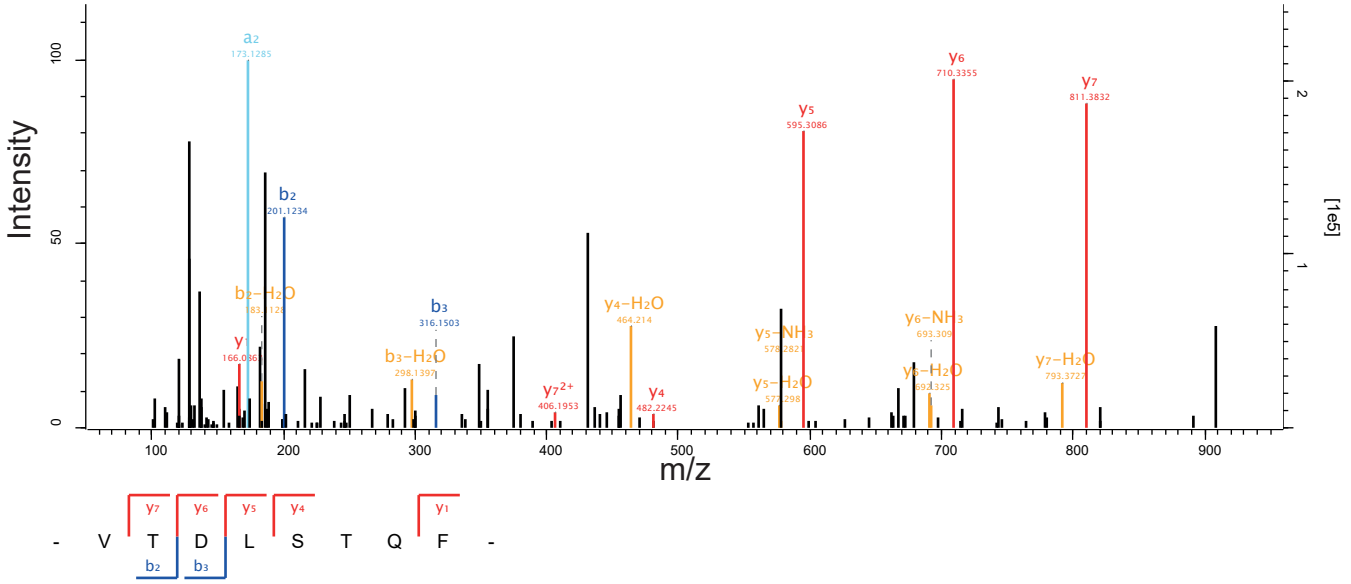
51BR-2 Brachybacterium alimentarium HLA-I TLSIDIEAV

Endogenous peptide



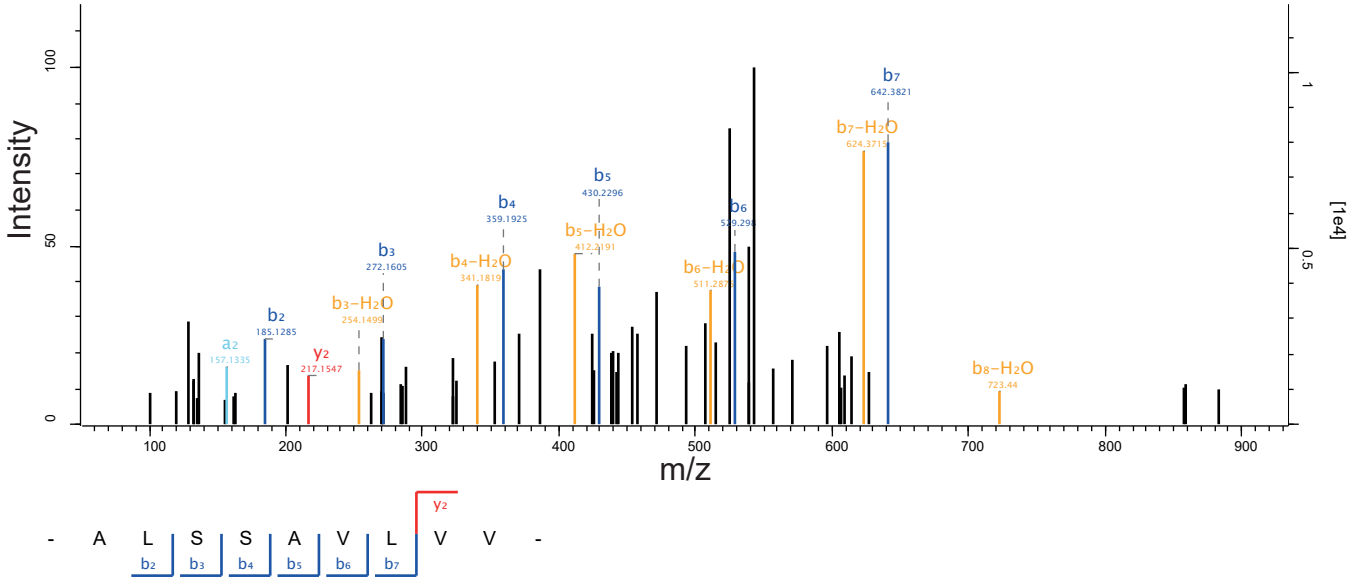
51BR-2 Brachybacterium alimentarium HLA-I VTDLSTQF

Endogenous peptide



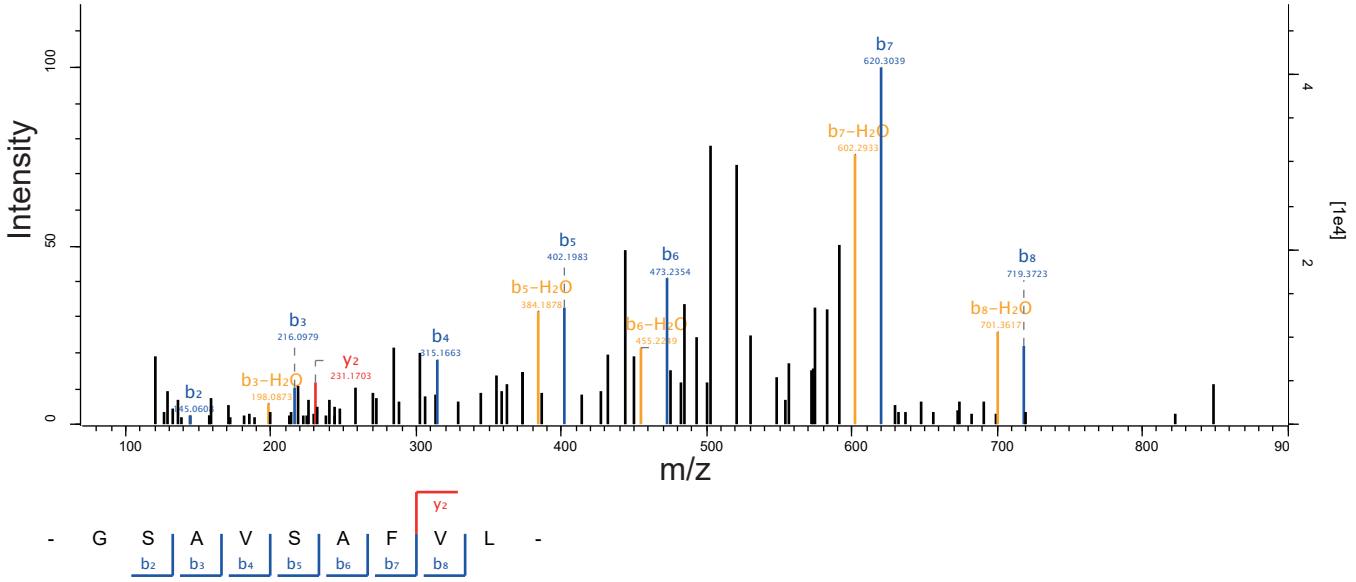
51BR-2 Eikenella corrodens HLA-I ALSSAVLVV

Endogenous peptide



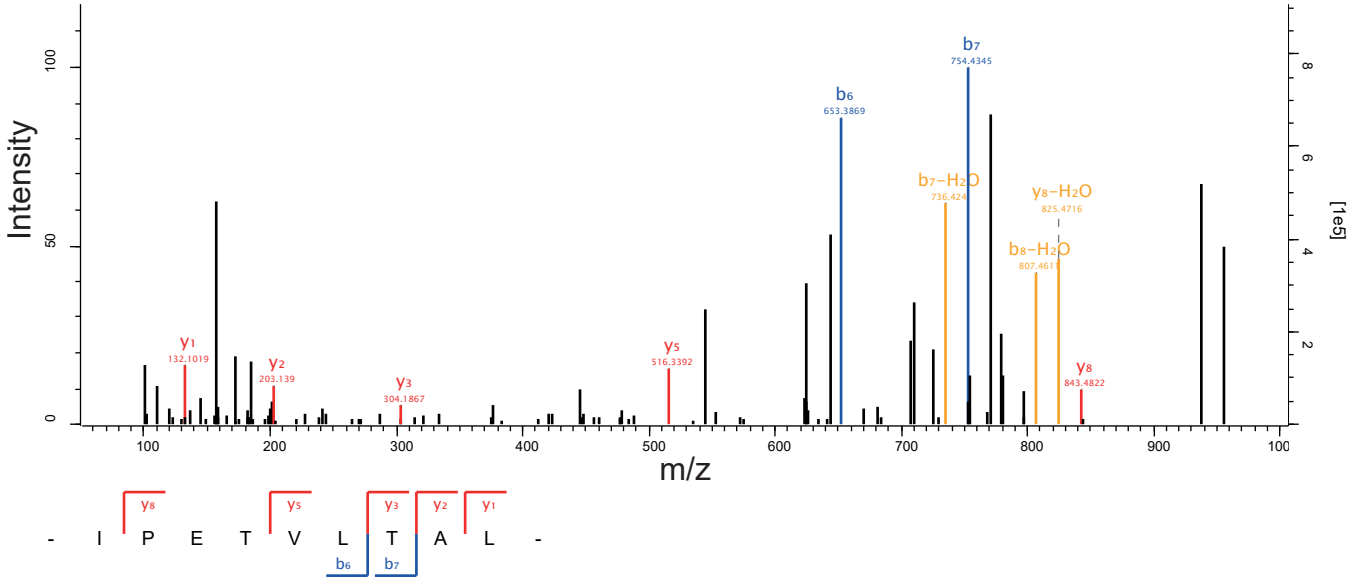
51BR-2 Eikenella corrodens HLA-I GSAVSAFVL

Endogenous peptide



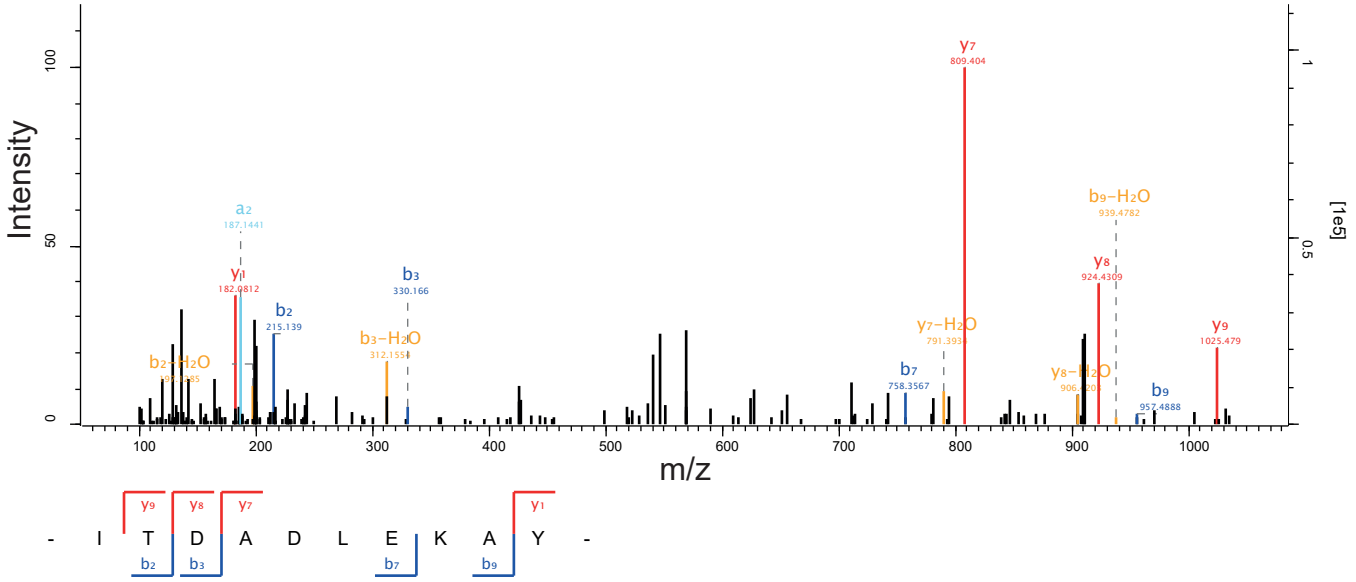
51BR-2 Eikenella corrodens HLA-I IPETVLTAL

Endogenous peptide



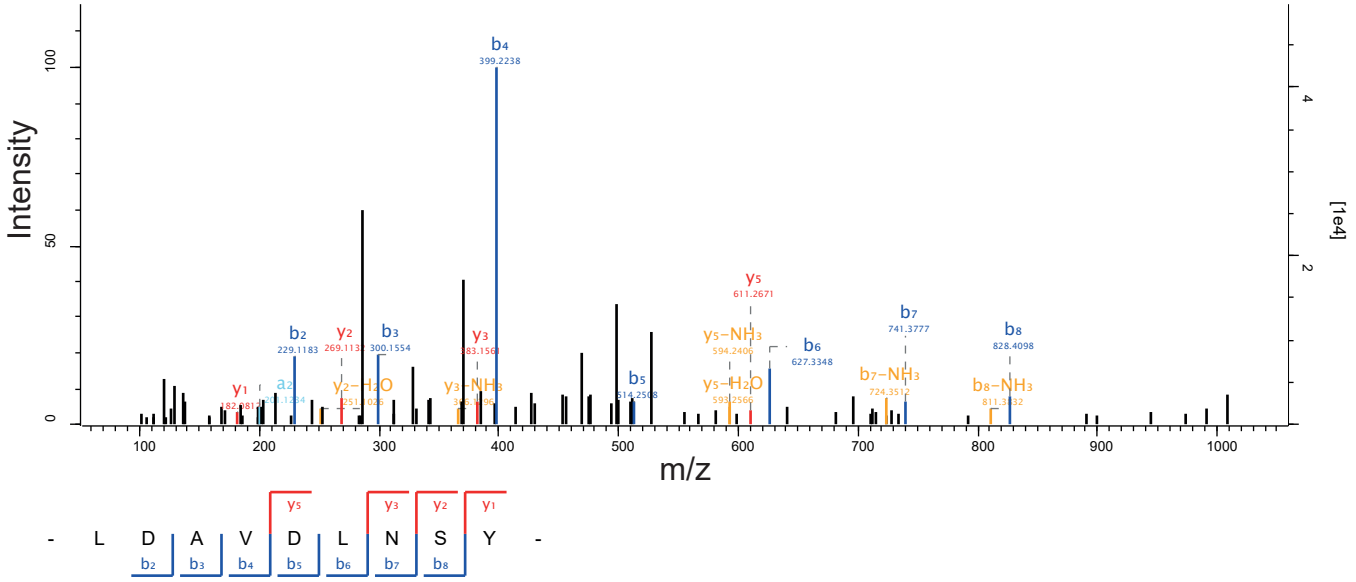
51BR-2 Eikenella corrodens HLA-I ITDADLEKAY

Endogenous peptide



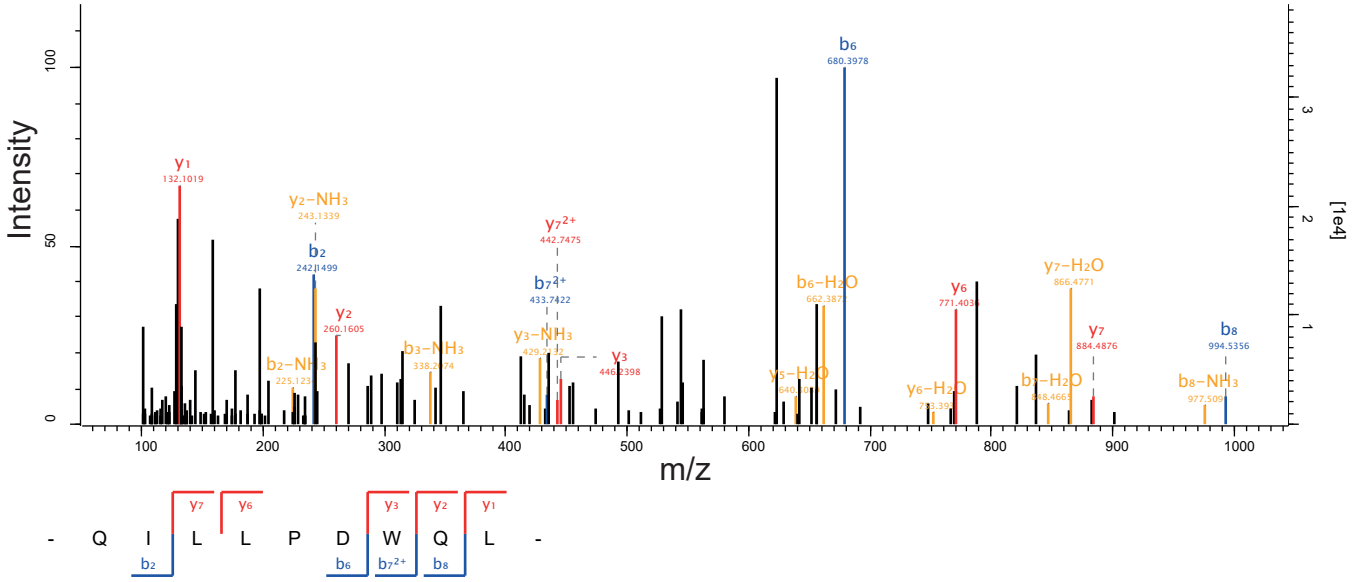
51BR-2 Eikenella corrodens HLA-I LDAVDLNSY

Endogenous peptide



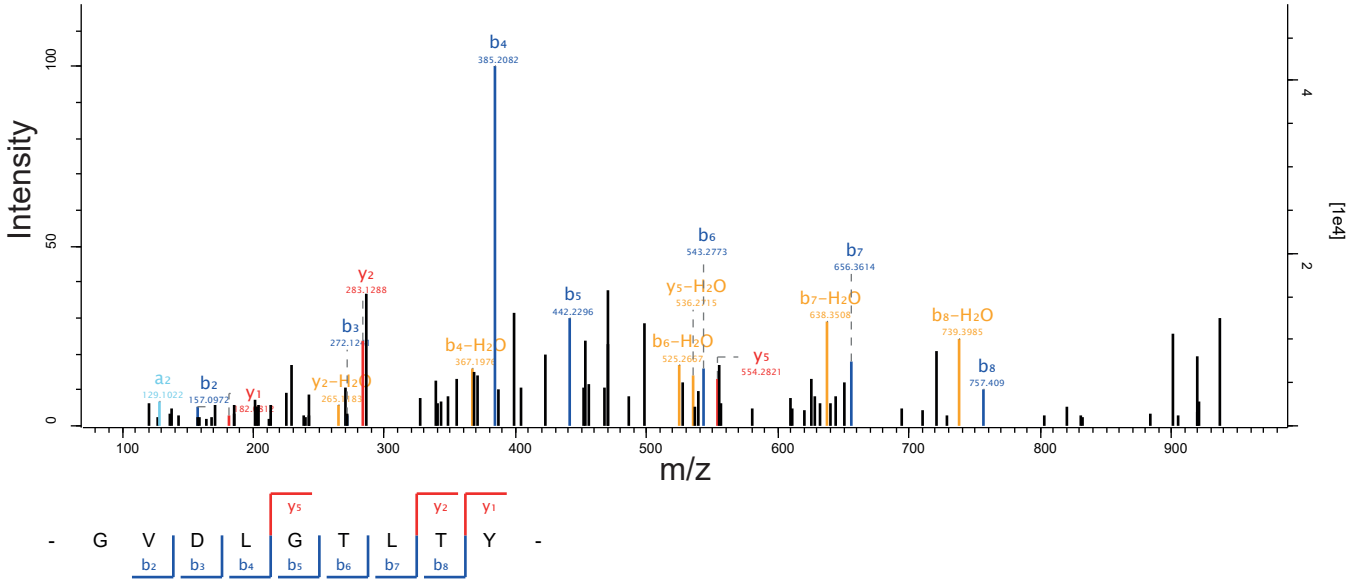
51BR-2 Eikenella corrodens HLA-I QILLPDWQL

Endogenous peptide



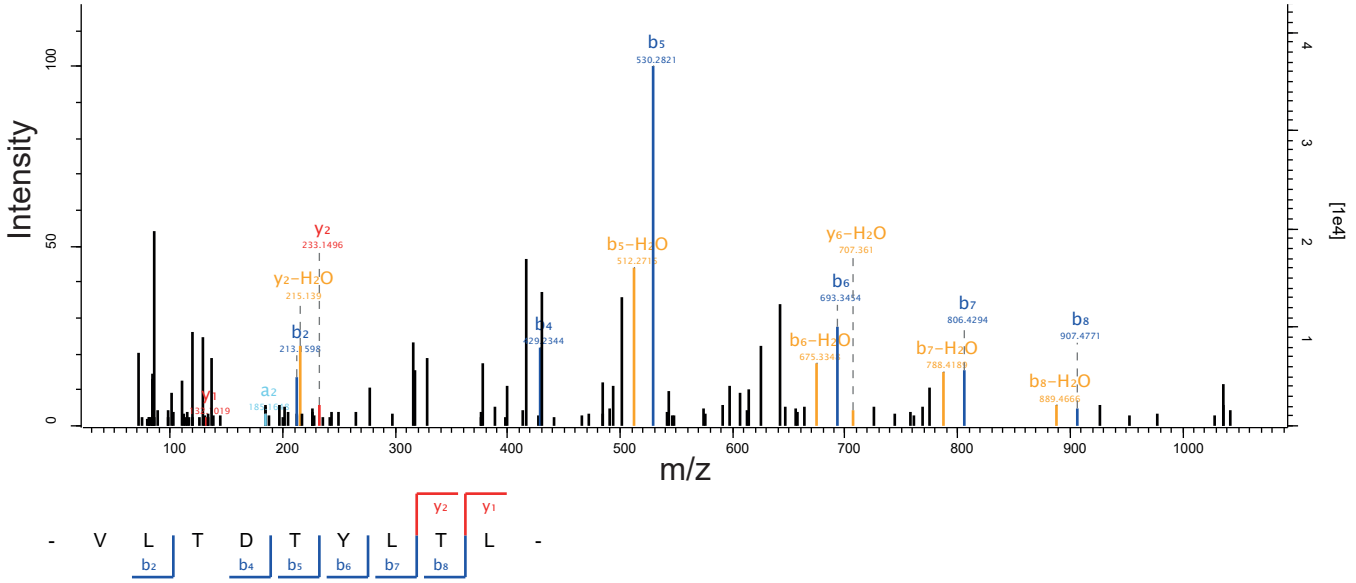
51BR-2 Fusobacterium nucleatim HLA-I GVDLGTLTY

Endogenous peptide



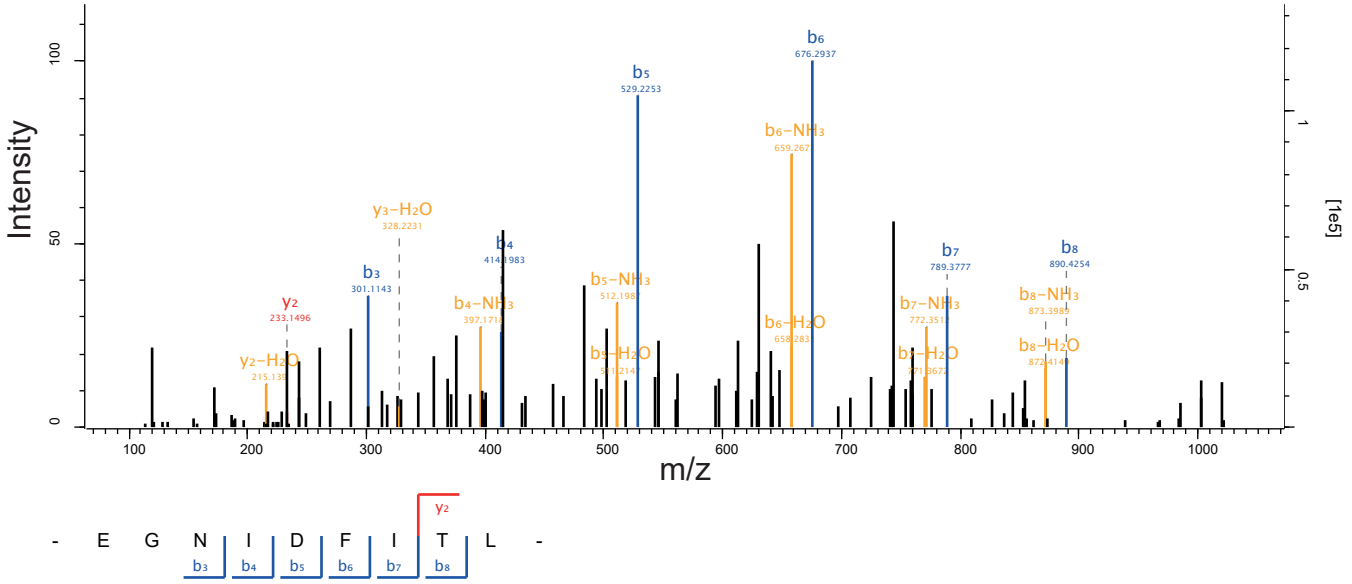
51BR-2 Fusobacterium nucleatim HLA-I VLTDTYLTL

Endogenous peptide



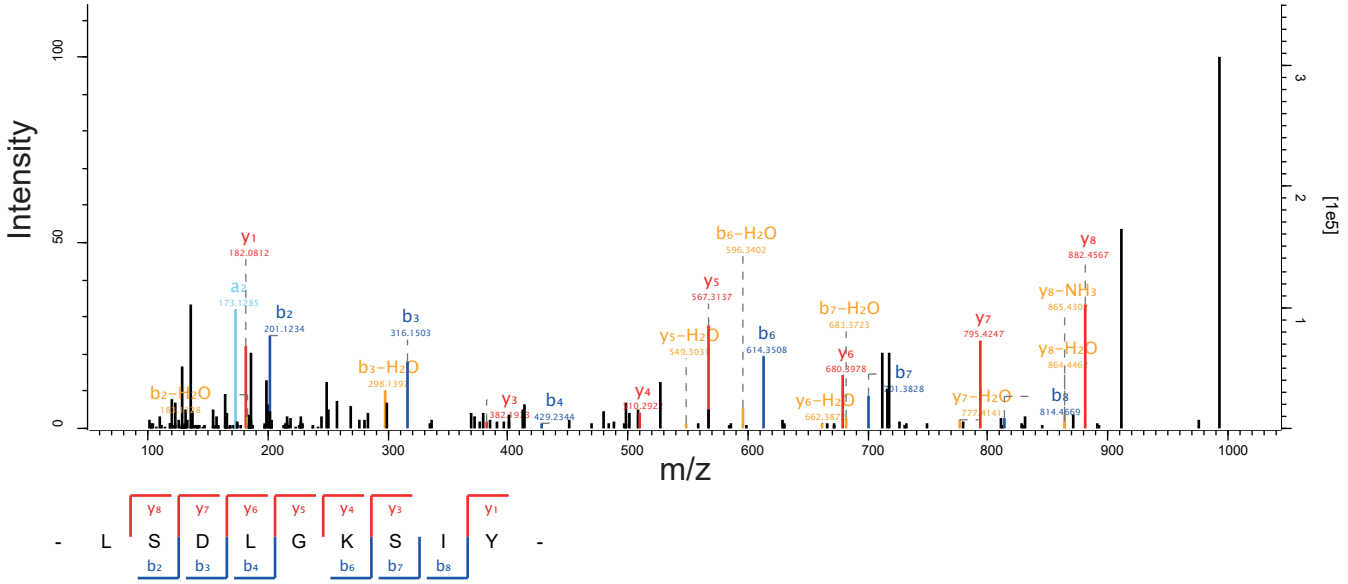
51BR-2 Staphylococcus aureus HLA-I EGNIDFITL

Endogenous peptide



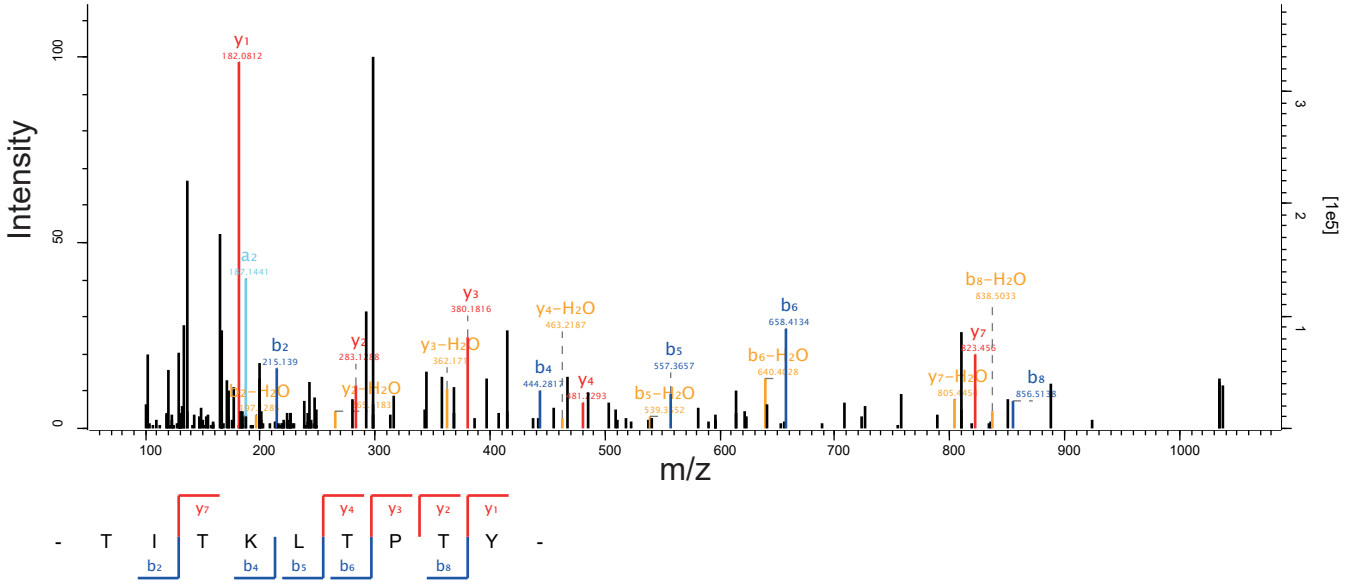
51BR-2 Staphylococcus aureus HLA-I LSDLGKSIY

Endogenous peptide



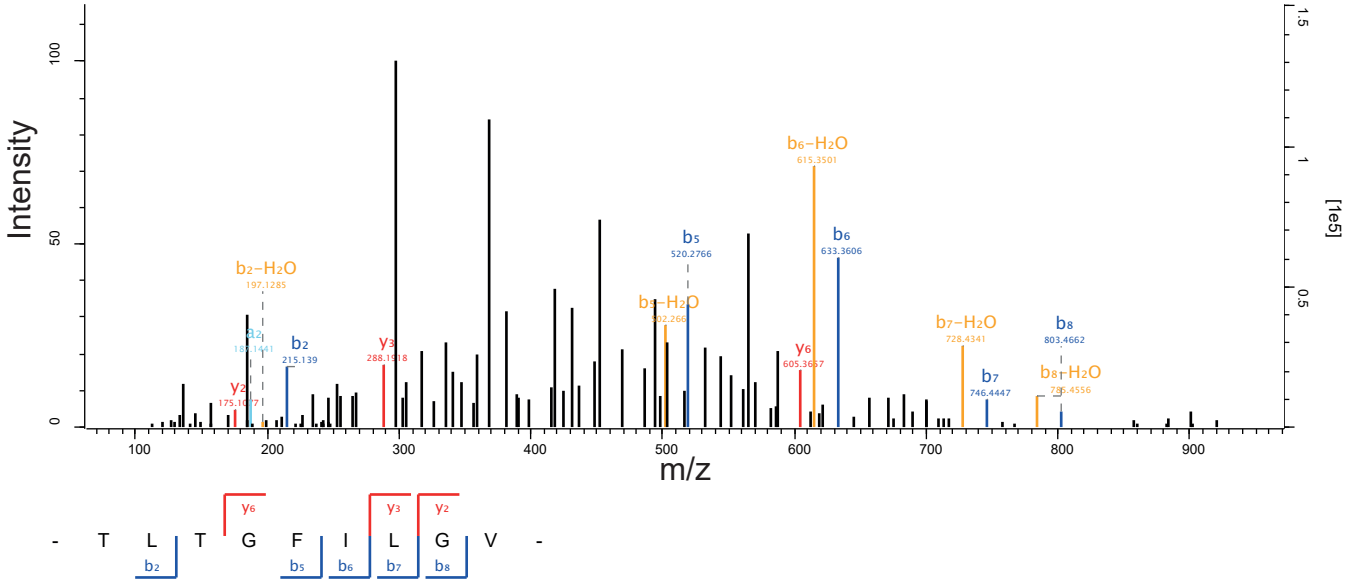
51BR-2 Staphylococcus aureus HLA-I TITKLTPTY

Endogenous peptide



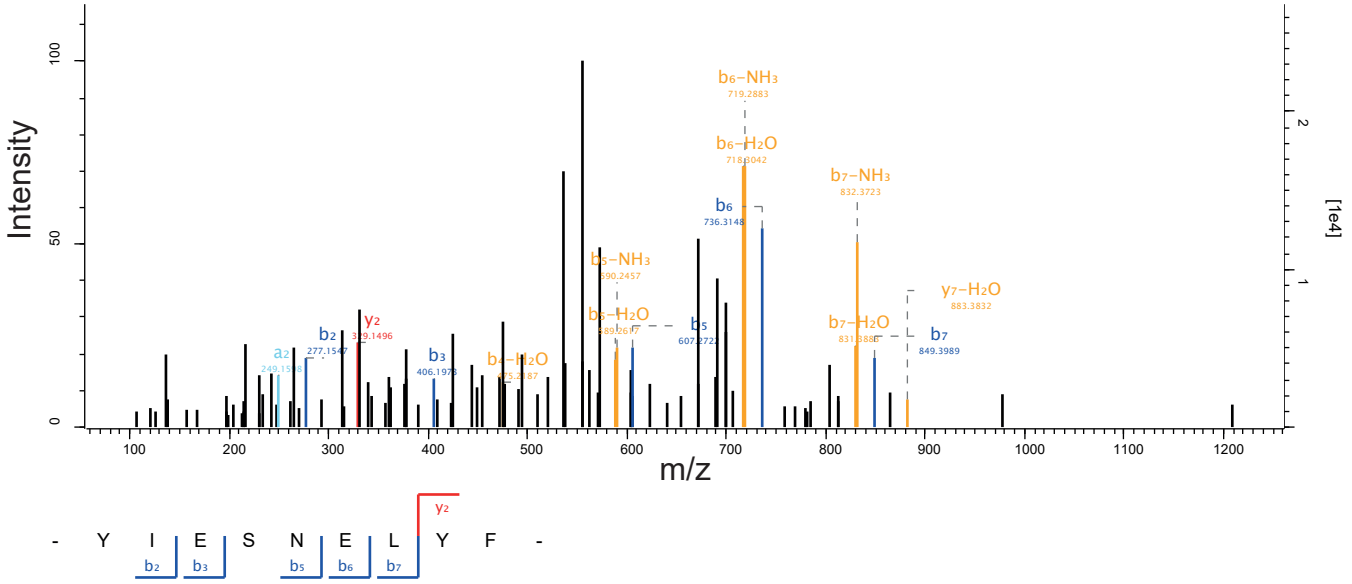
51BR-2 Staphylococcus aureus HLA-I TLTGFILGV

Endogenous peptide



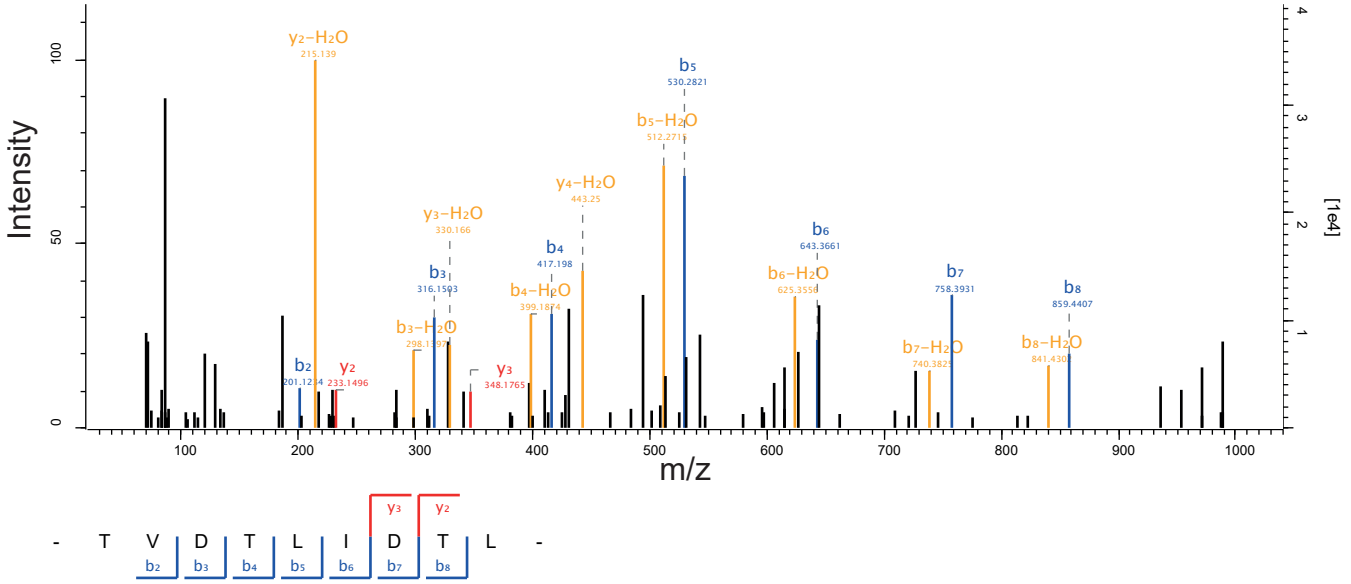
51BR-2 Staphylococcus aureus HLA-I YIESNELYF

Endogenous peptide



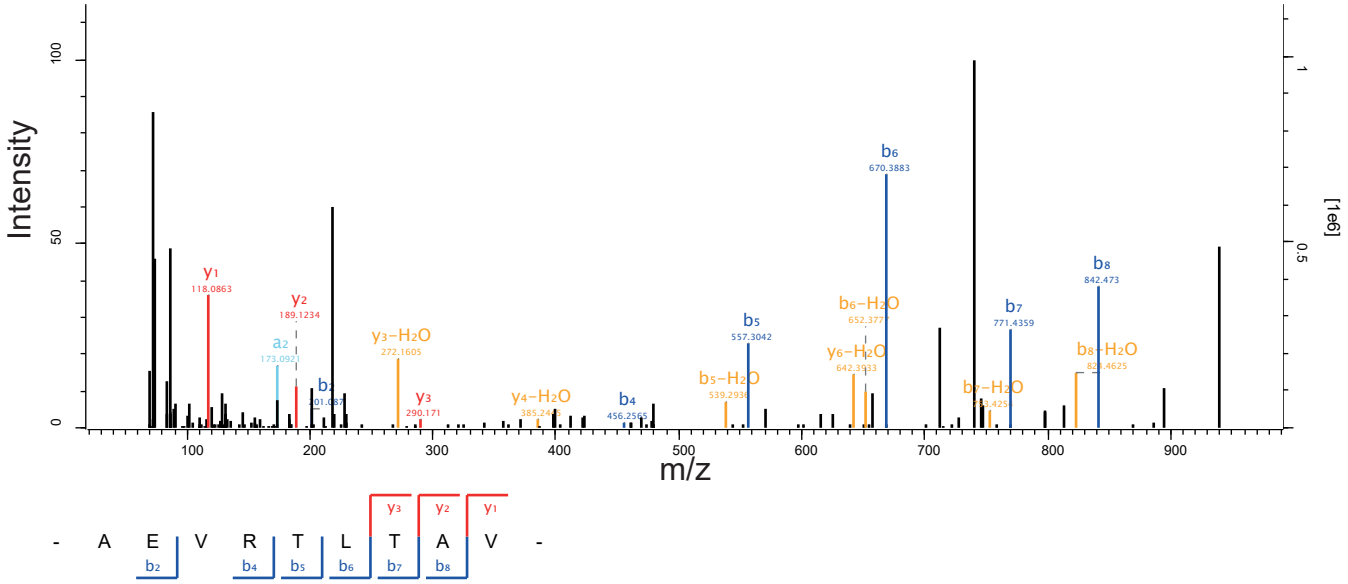
55A3 Kingella denitrificans HLA-I TVDTLIDTL

Endogenous peptide



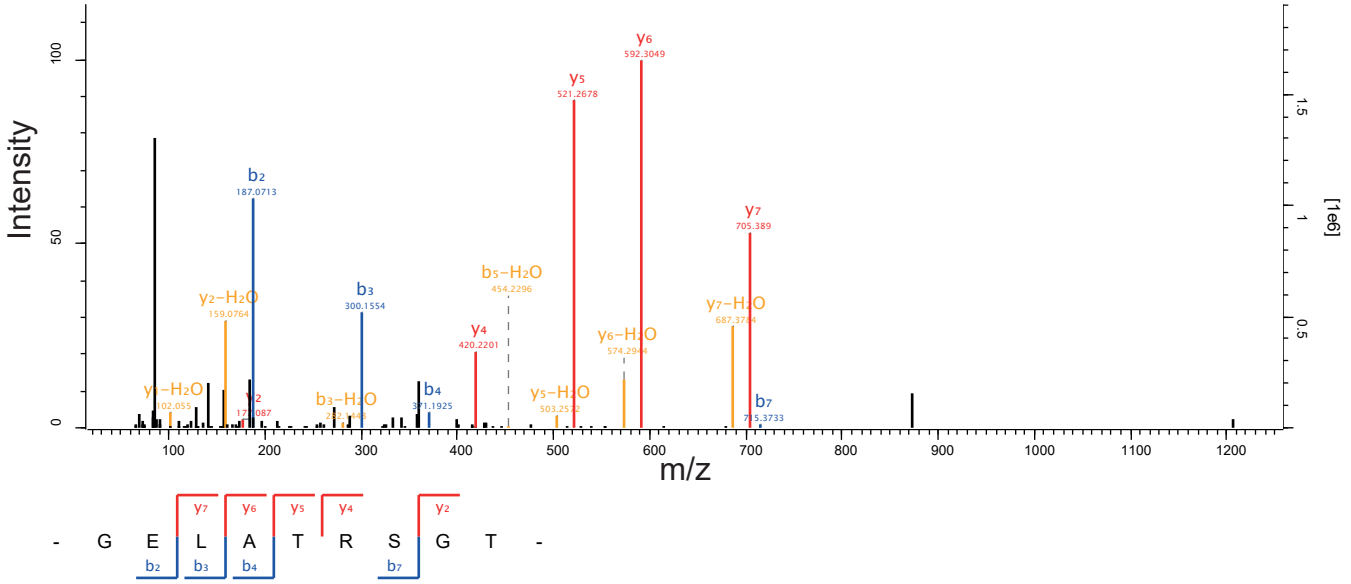
55A3 Staphylococcus capitis HLA-I AEVRTLTAV

Endogenous peptide



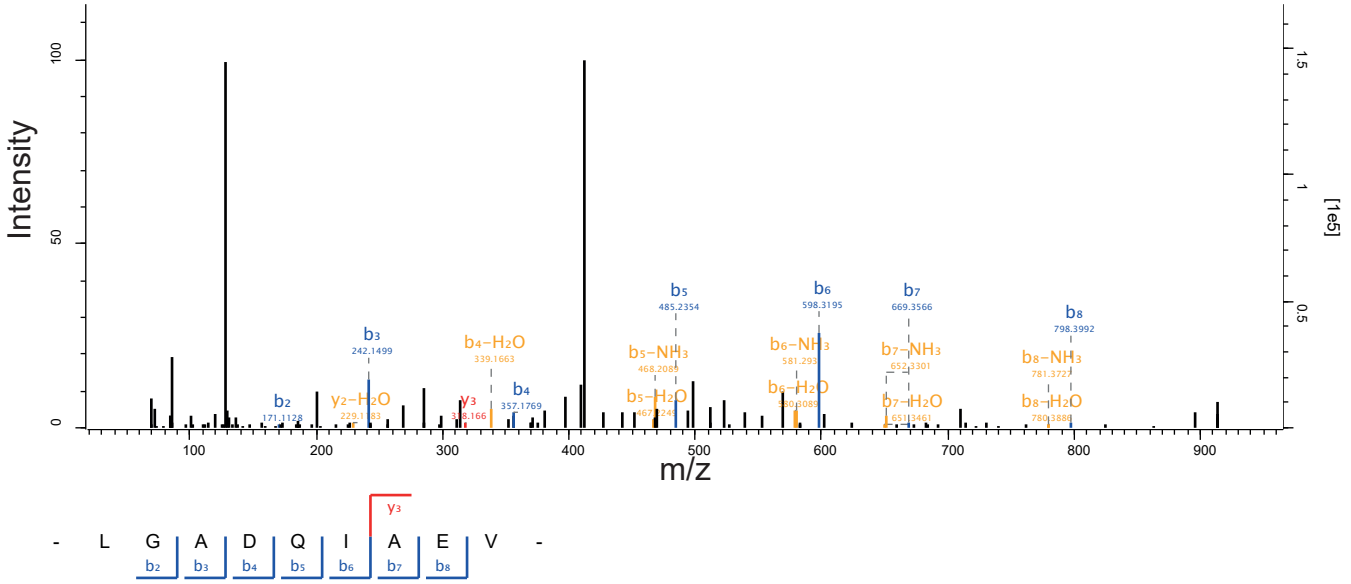
55A3 Staphylococcus capitis HLA-I GELATRSGT

Endogenous peptide



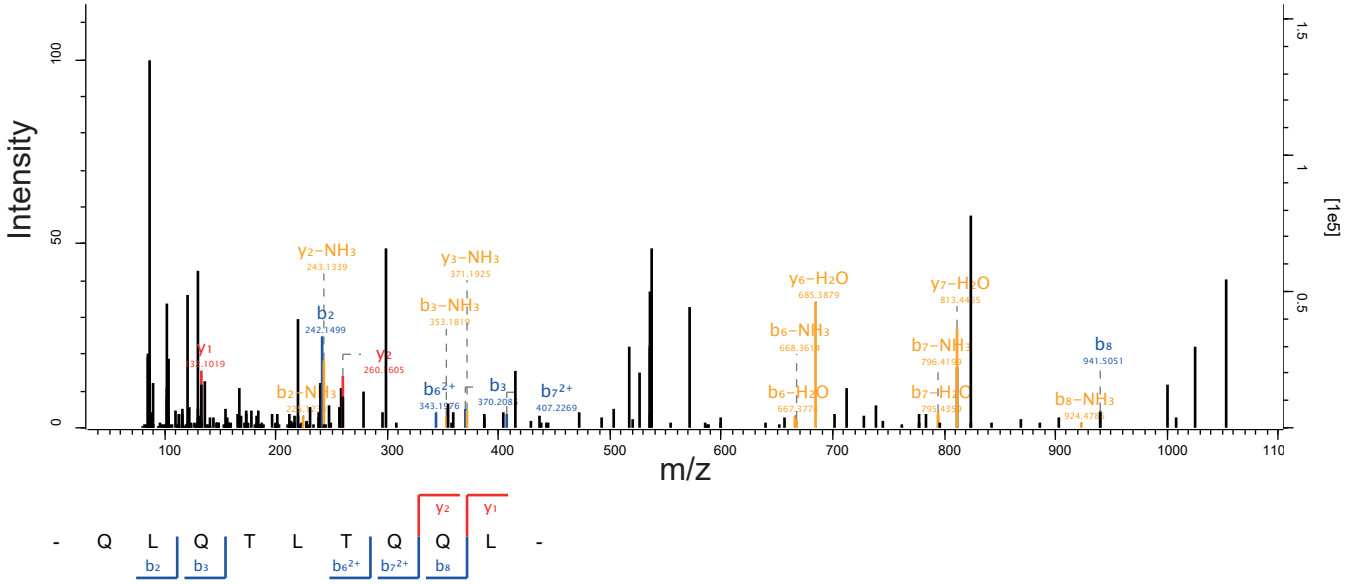
55A3 Staphylococcus capitis HLA-I LGADQIAEV

Endogenous peptide



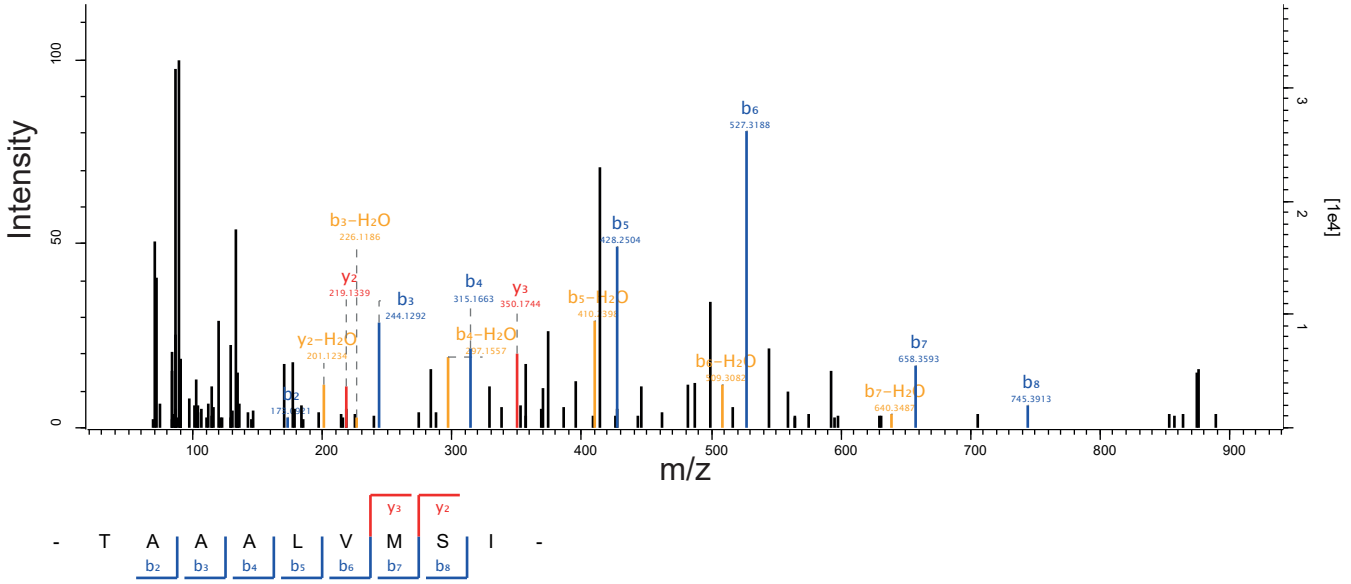
55A3 Staphylococcus capitis HLA-I QLQTLTQQL

Endogenous peptide



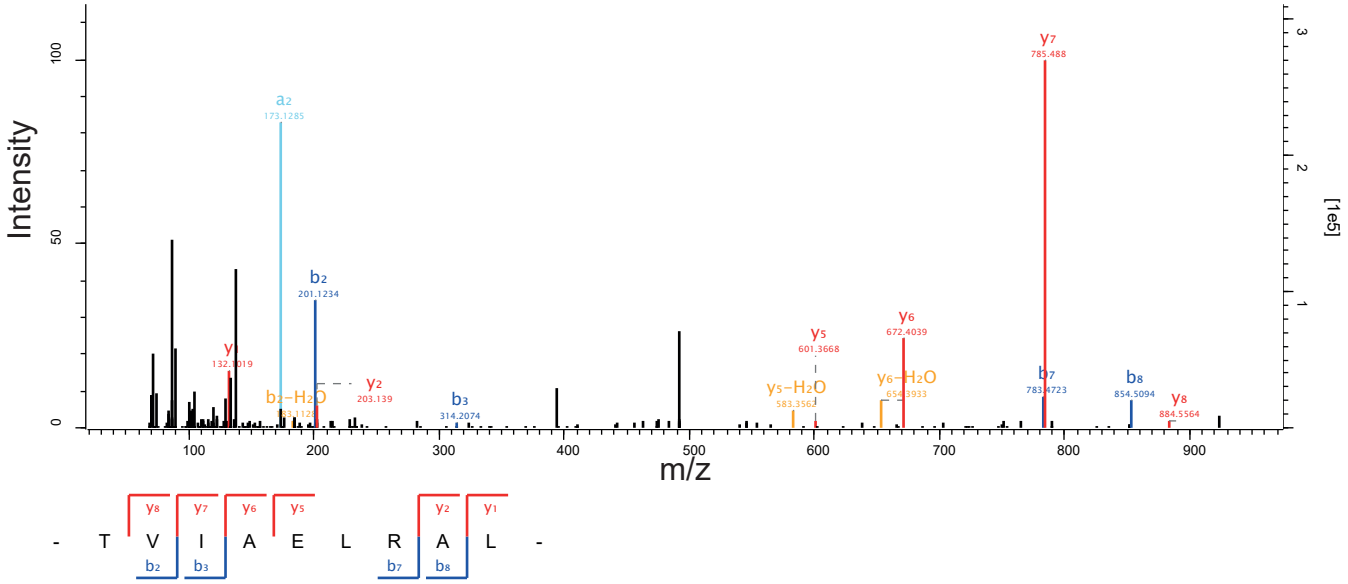
55A3 Staphylococcus capitis HLA-I TAAALVMSI

Endogenous peptide



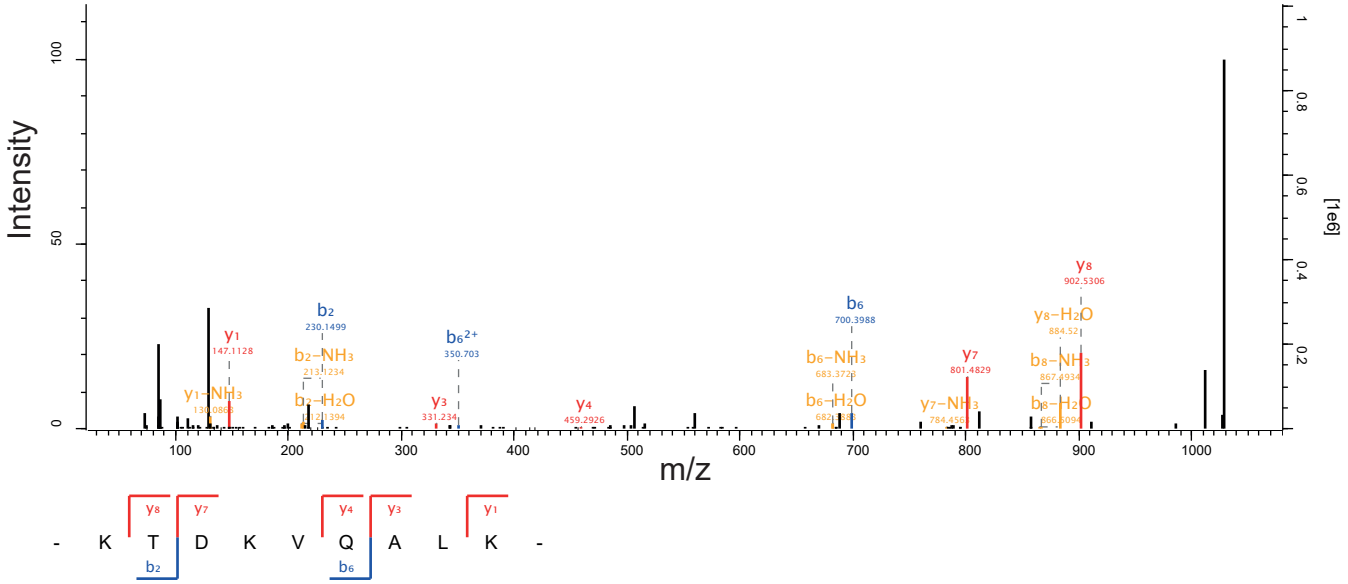
55A3 Staphylococcus capitis HLA-I TVIAELRAL

Endogenous peptide



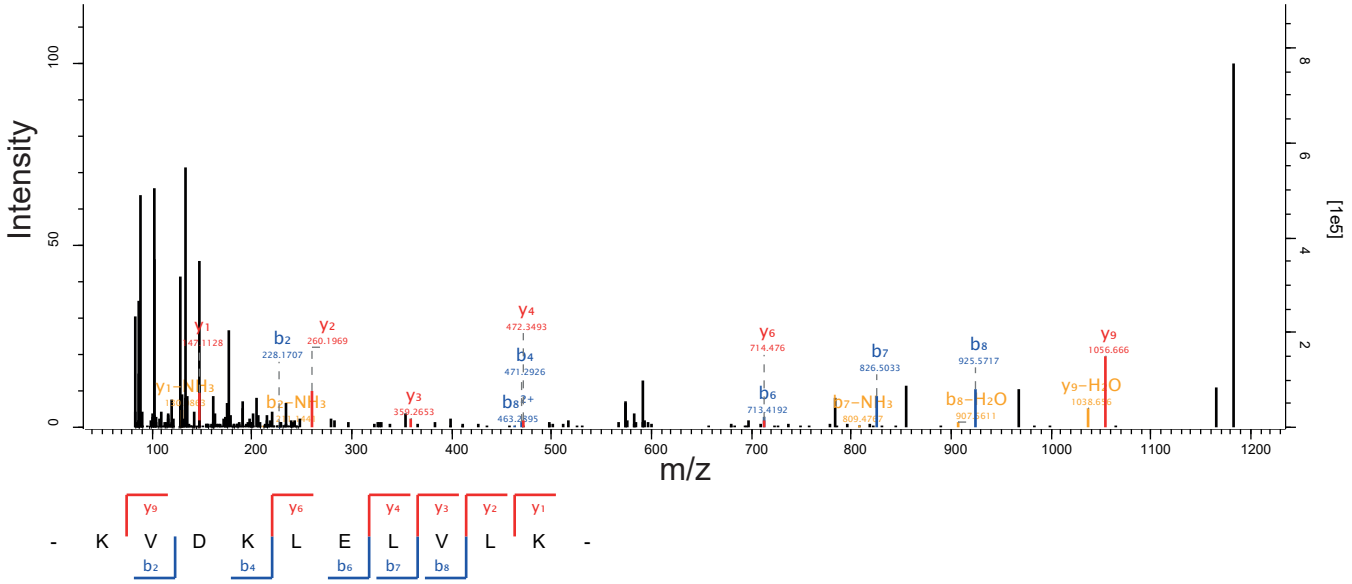
55A3 Streptococcus gordonii HLA-I KTDKVQALK

Endogenous peptide



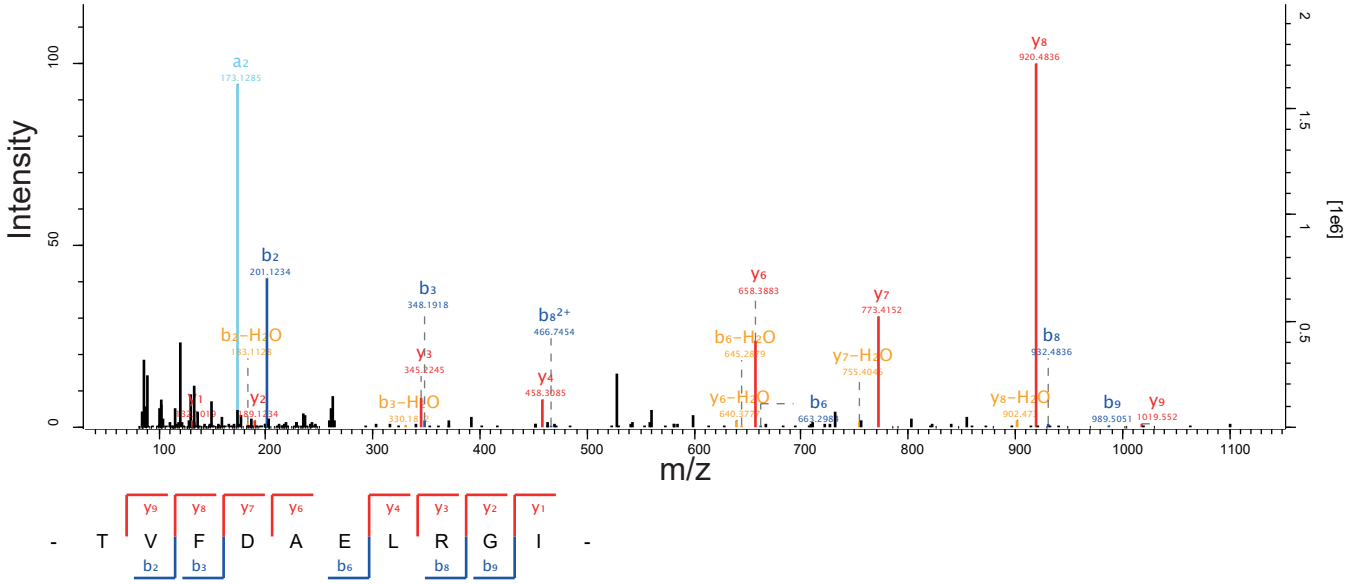
55A7-1 Clostridium clostridioforme HLA-I KVDKLELVLK

Endogenous peptide



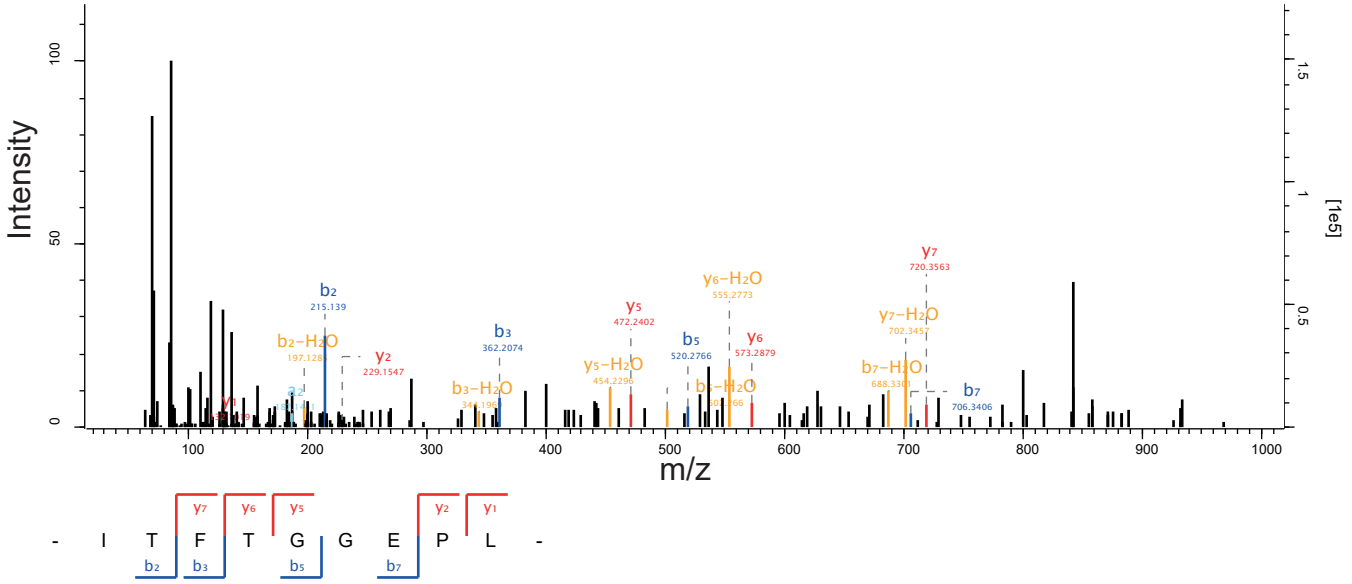
55A7-1 Clostridium clostridioforme HLA-I TVFDAELRGI

Endogenous peptide



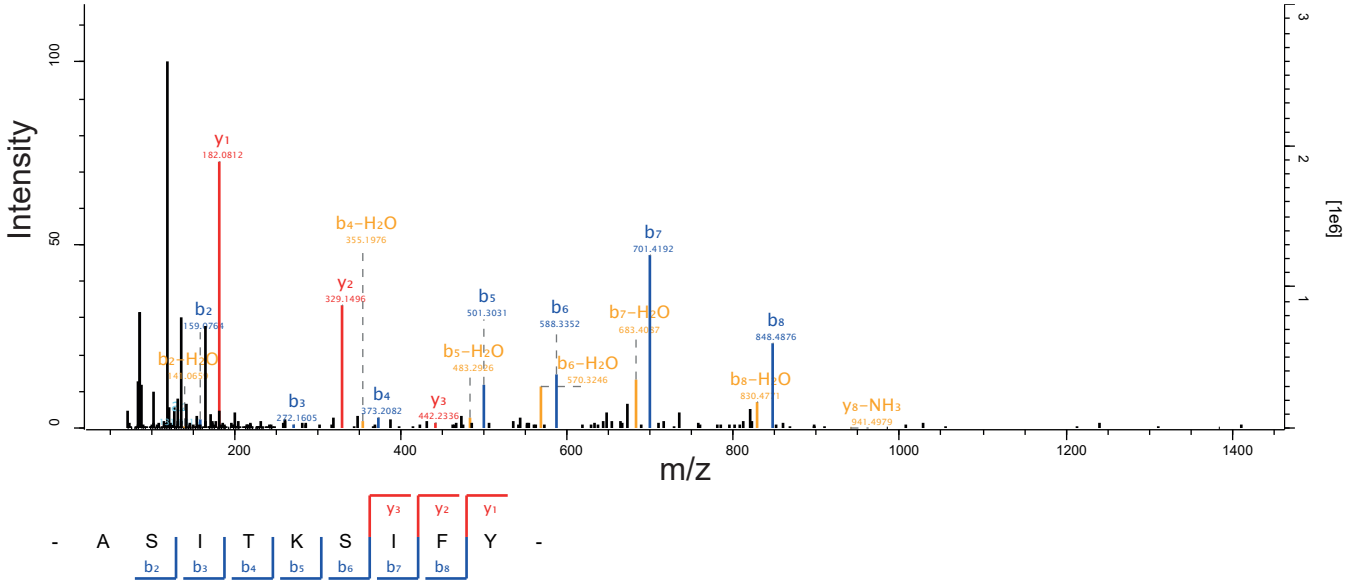
55A7-1 Dialister invisus HLA-I ITFTGGEPL

Endogenous peptide



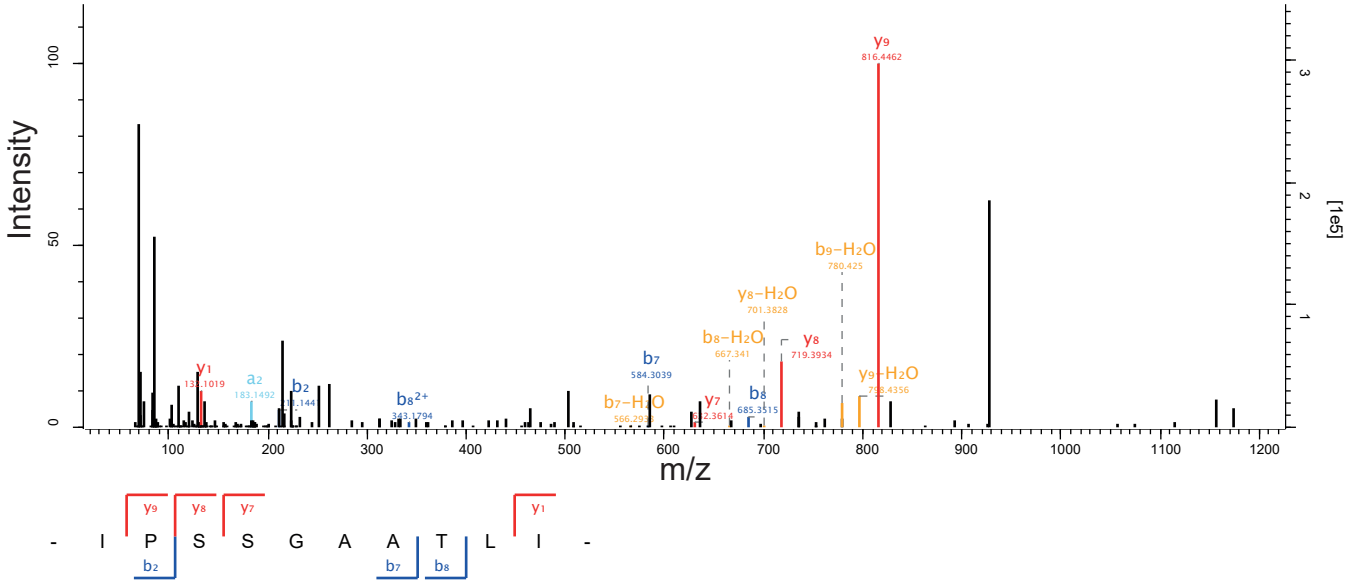
55A7-1 Enterobacter cancerogenus HLA-I ASITKSIFY

Endogenous peptide



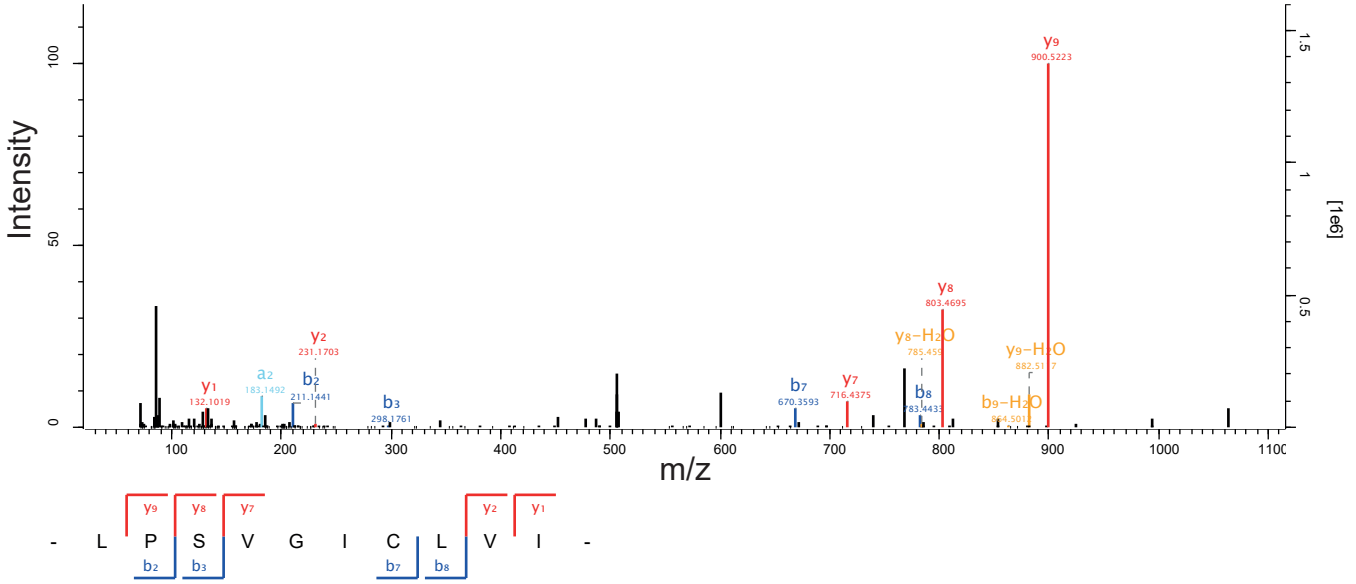
55A7-1 Enterobacter cancerogenus HLA-I IPSSGAATLI

Endogenous peptide



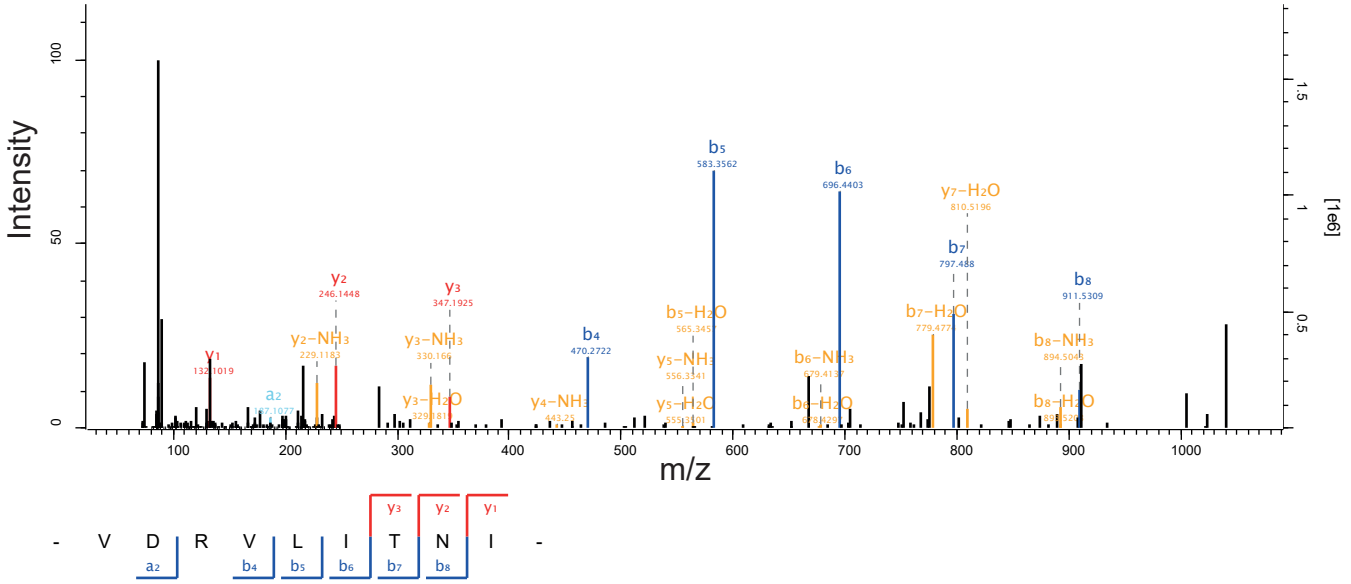
55A7-1 Enterobacter cancerogenus HLA-I LPSVGICLVI

Endogenous peptide



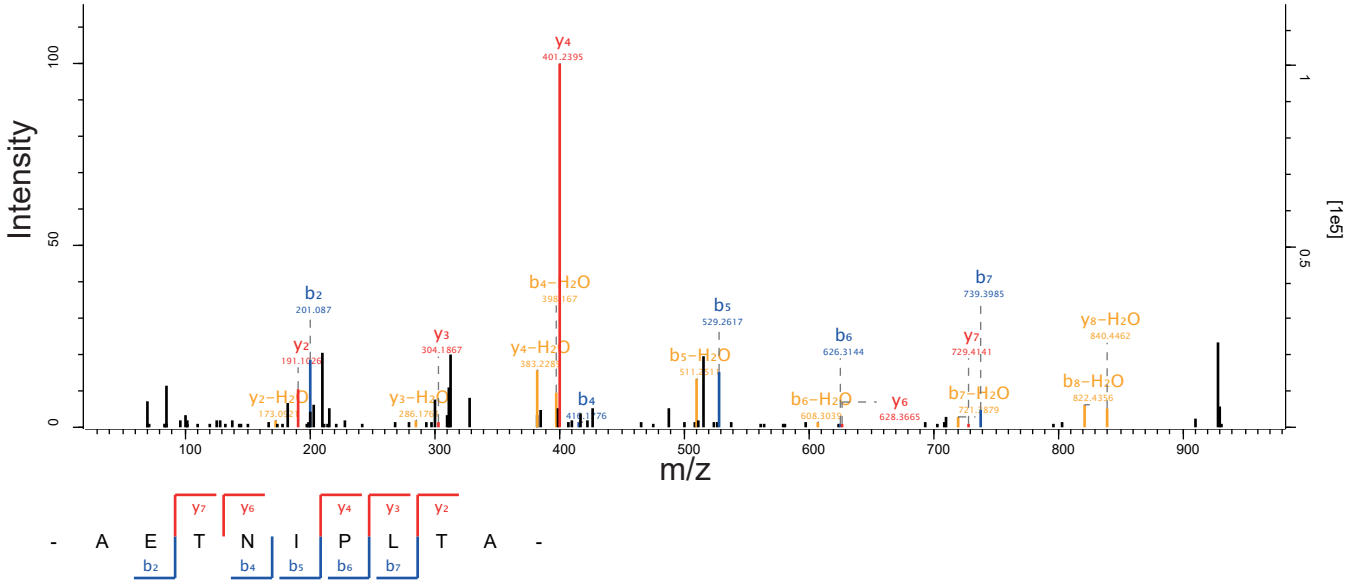
55A7-1 Veillonella dispar HLA-I VDRVLITNI

Endogenous peptide



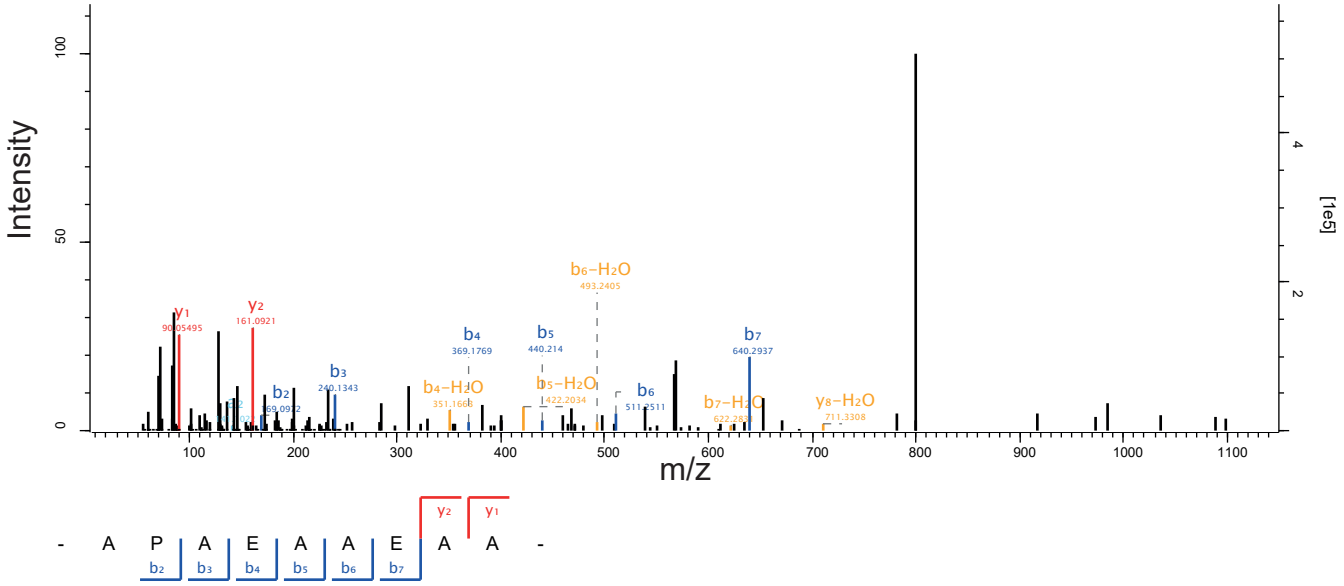
55B3 Gardnerella vaginalis HLA-I AETNIPLTA

Endogenous peptide



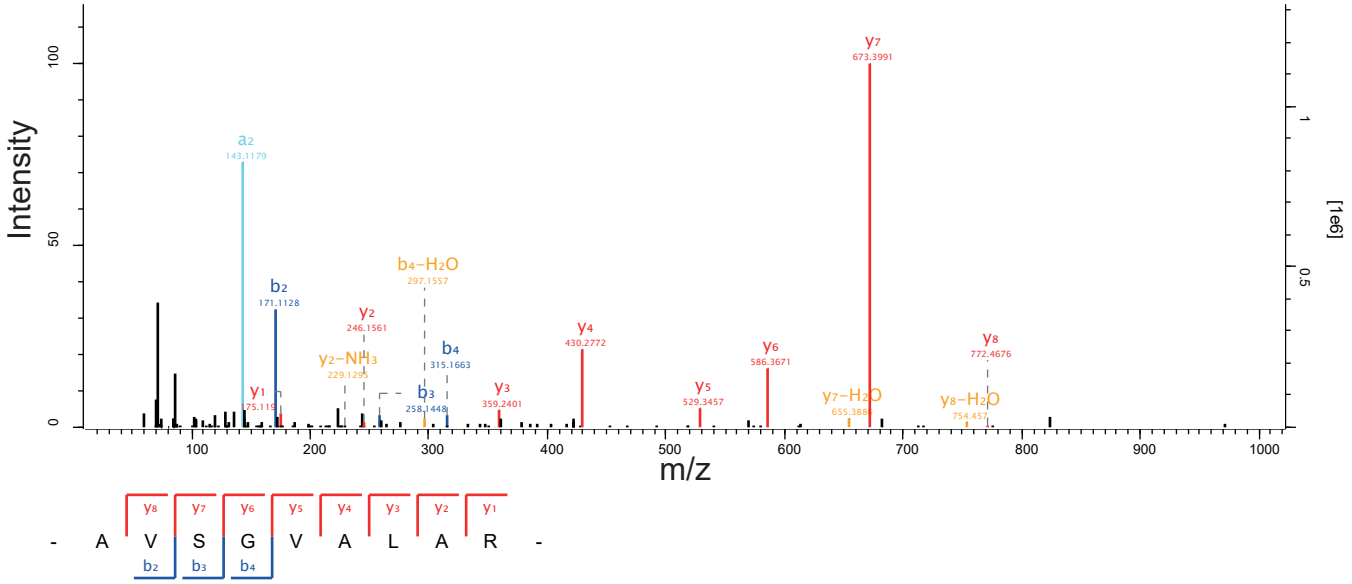
55B3 Gardnerella vaginalis HLA-I APAEAAEAA

Endogenous peptide



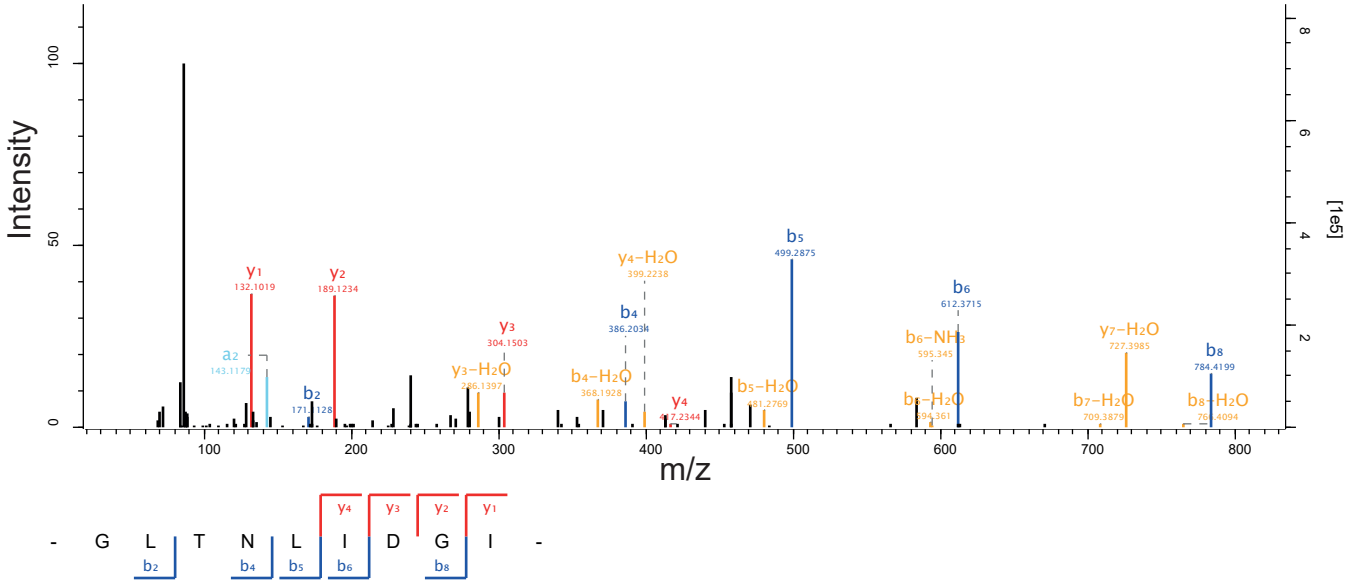
55B3 Gardnerella vaginalis HLA-I AVSGVALAR

Endogenous peptide



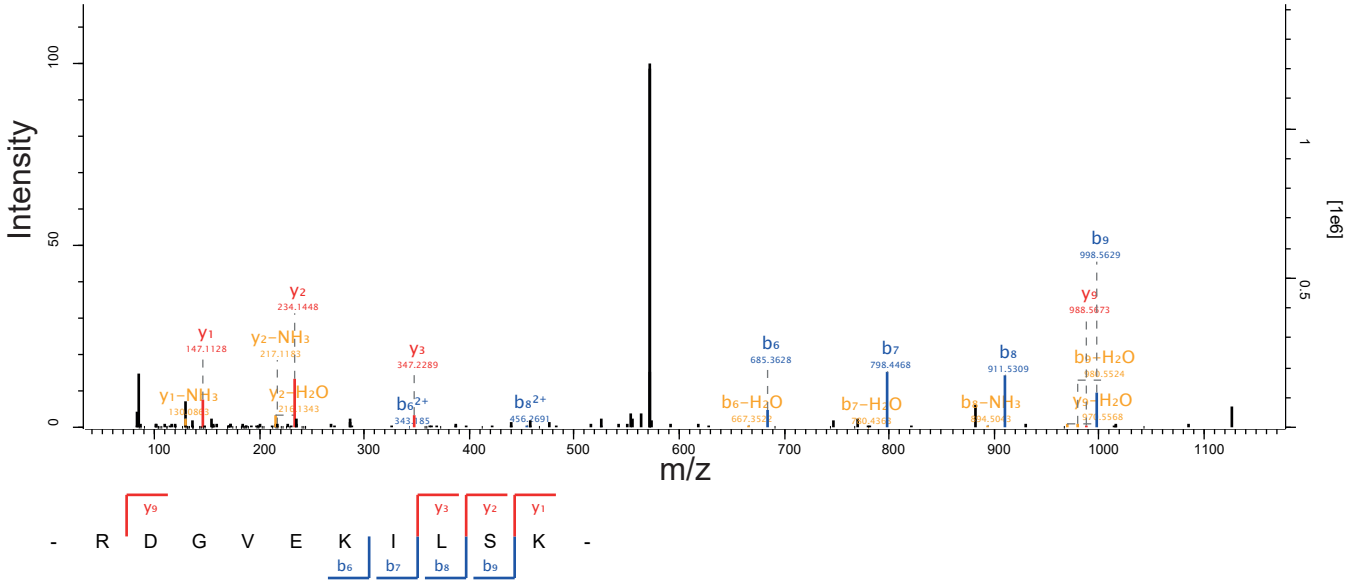
55B3 Gardnerella vaginalis HLA-I GLTNLIDGI

Endogenous peptide



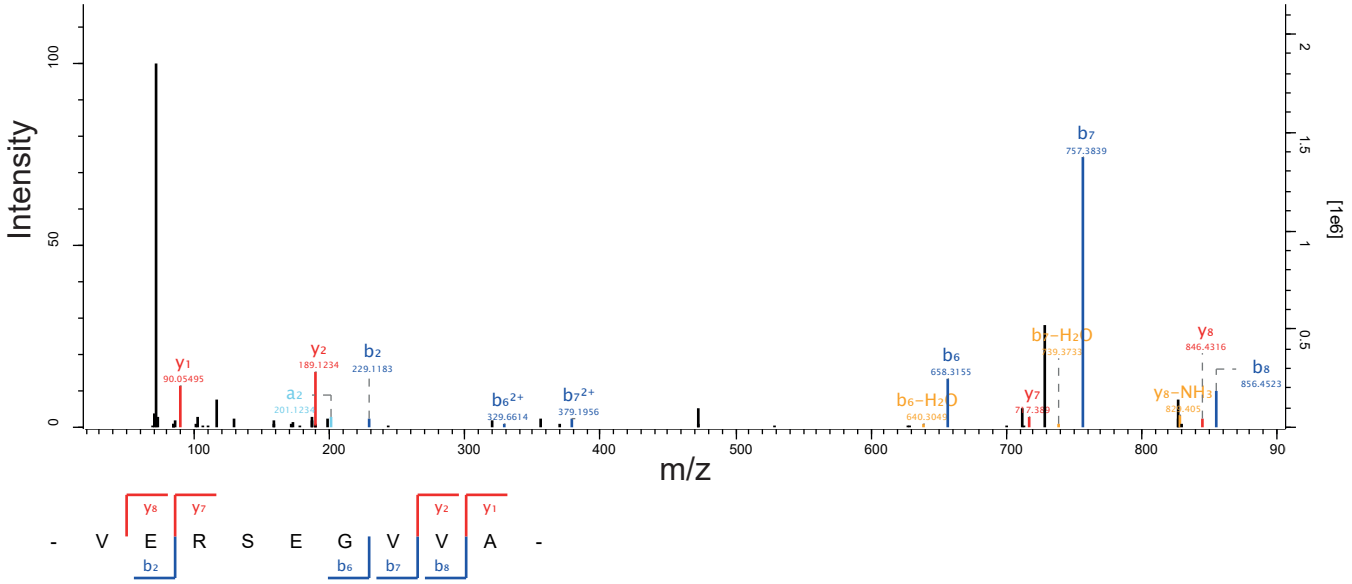
55B3 Gardnerella vaginalis HLA-I RDGVEKILSK

Endogenous peptide



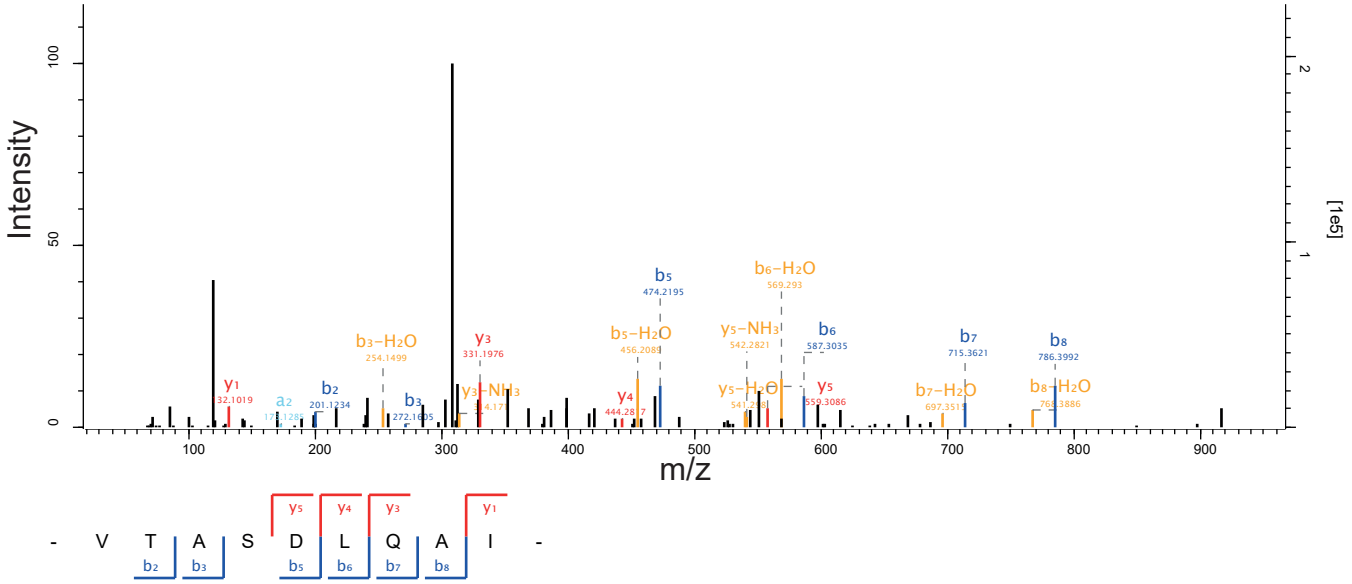
55B3 Gardnerella vaginalis HLA-I VERSEGVVA

Endogenous peptide



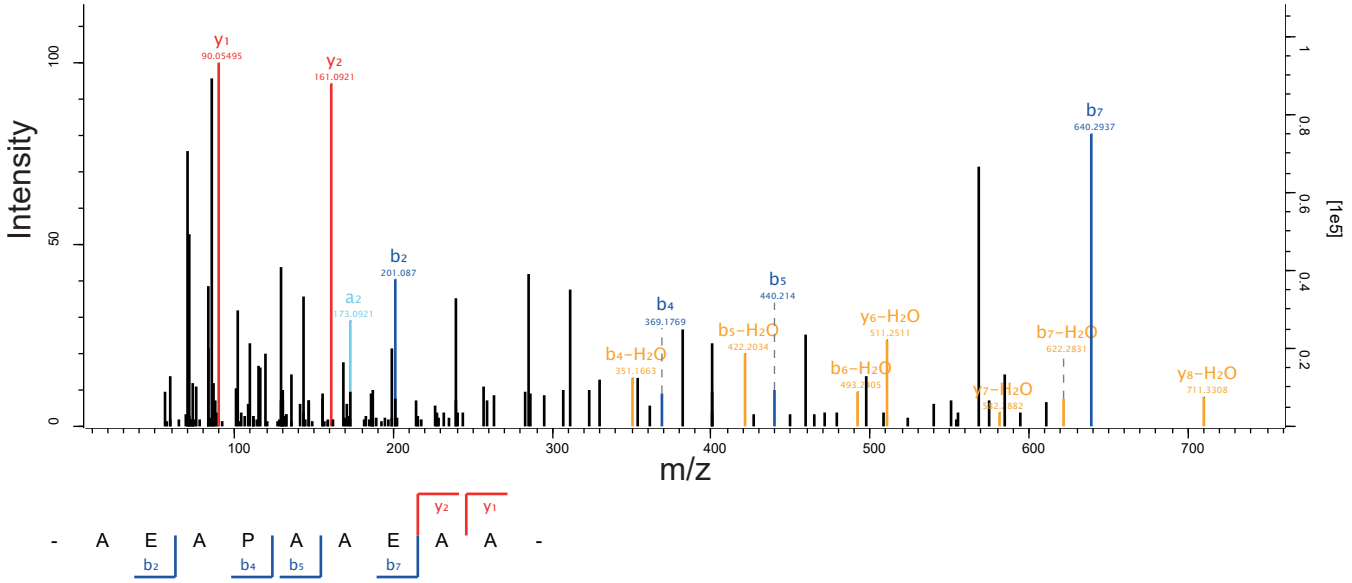
55B3 Gardnerella vaginalis HLA-I VTASDLQAI

Endogenous peptide



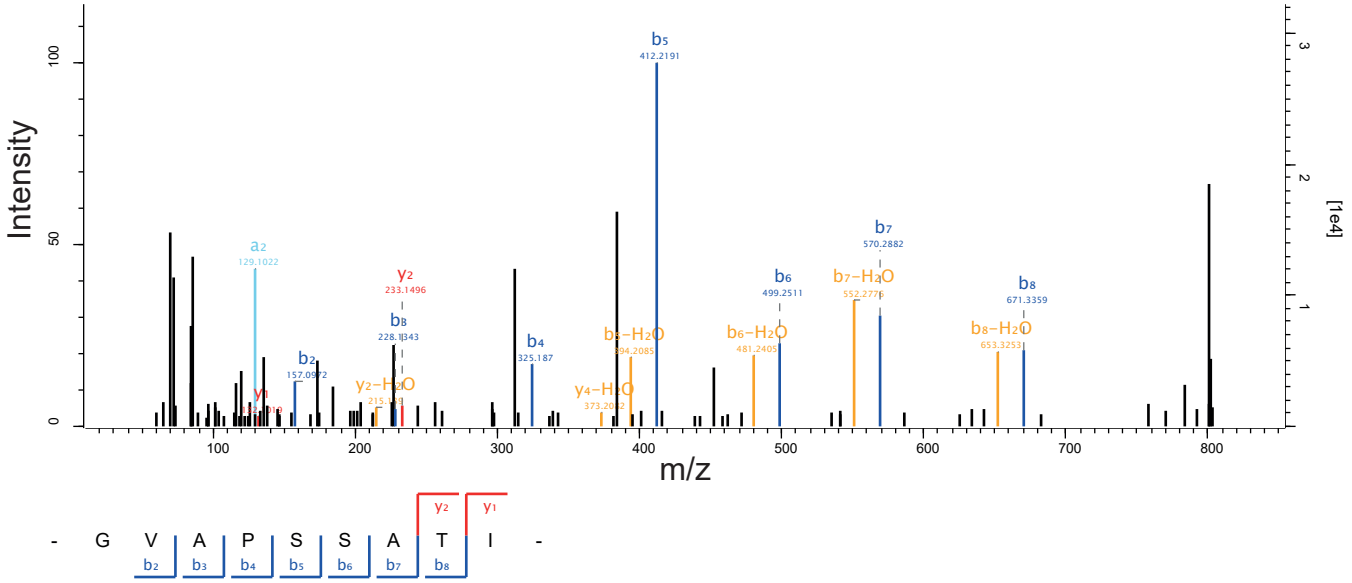
55B3 Sphingomonas dokdonensis HLA-I AEAPAAEAA

Endogenous peptide



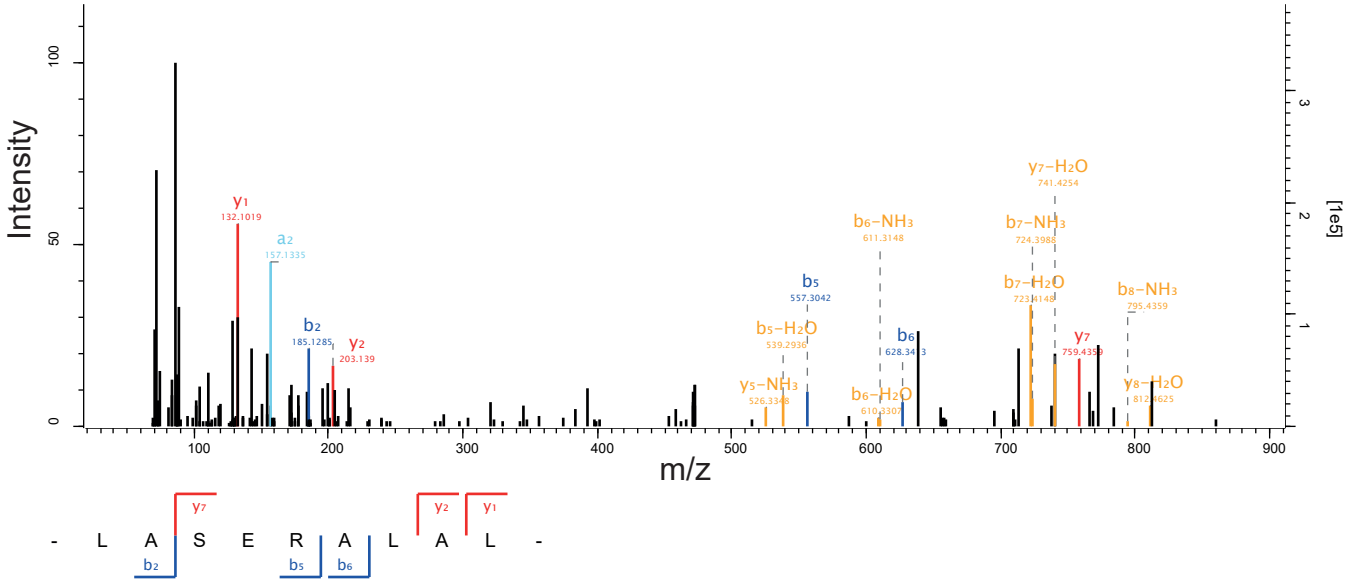
55B3 Sphingomonas dokdonensis HLA-I GVAPSSATI

Endogenous peptide



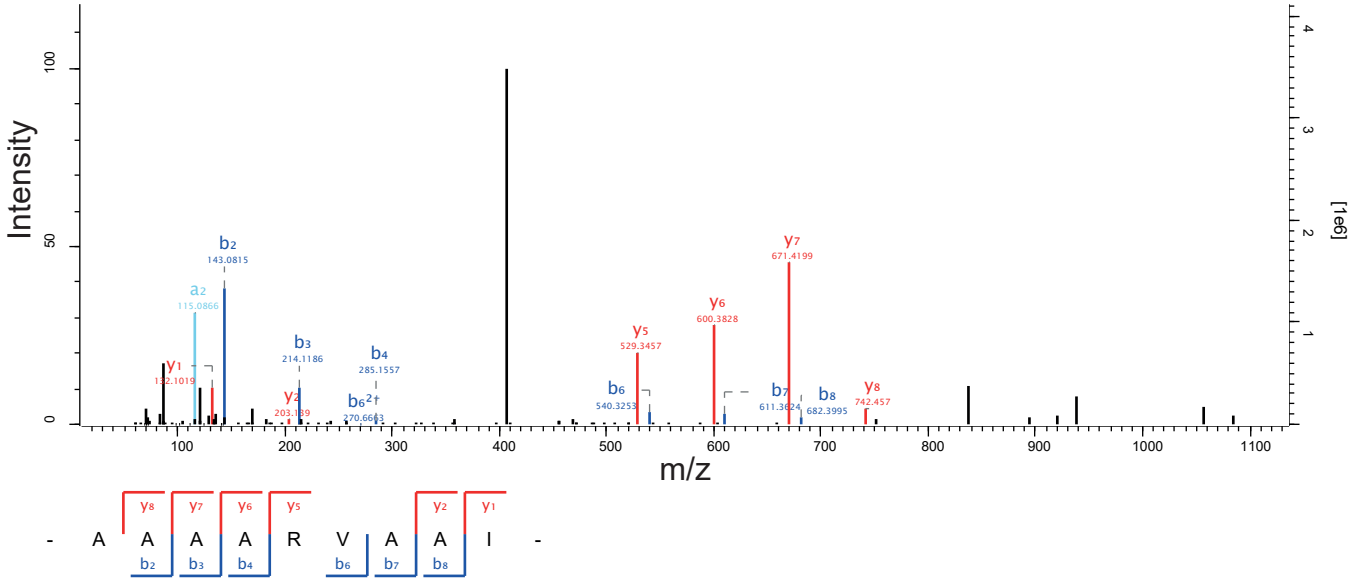
55B3 Sphingomonas dokdonensis HLA-I LASERALAL

Endogenous peptide



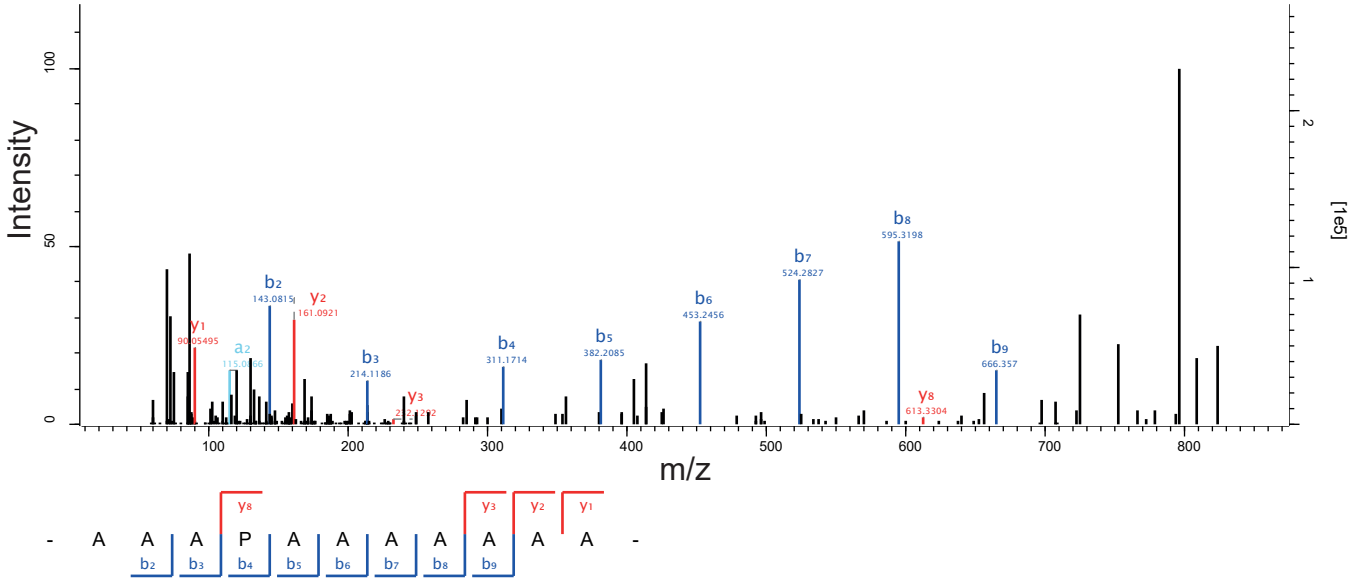
55B3 Sphingomonas melonis HLA-I AAAARVAAI

Endogenous peptide



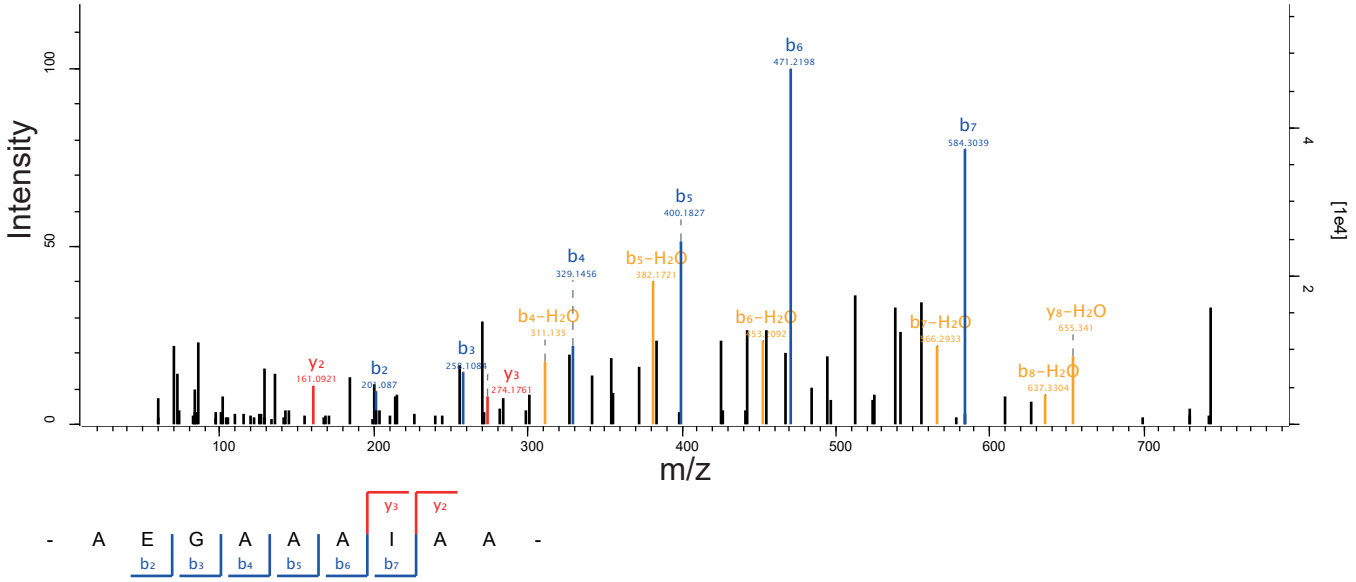
55B3 Sphingomonas melonis HLA-I AAAPAAAAAAA

Endogenous peptide



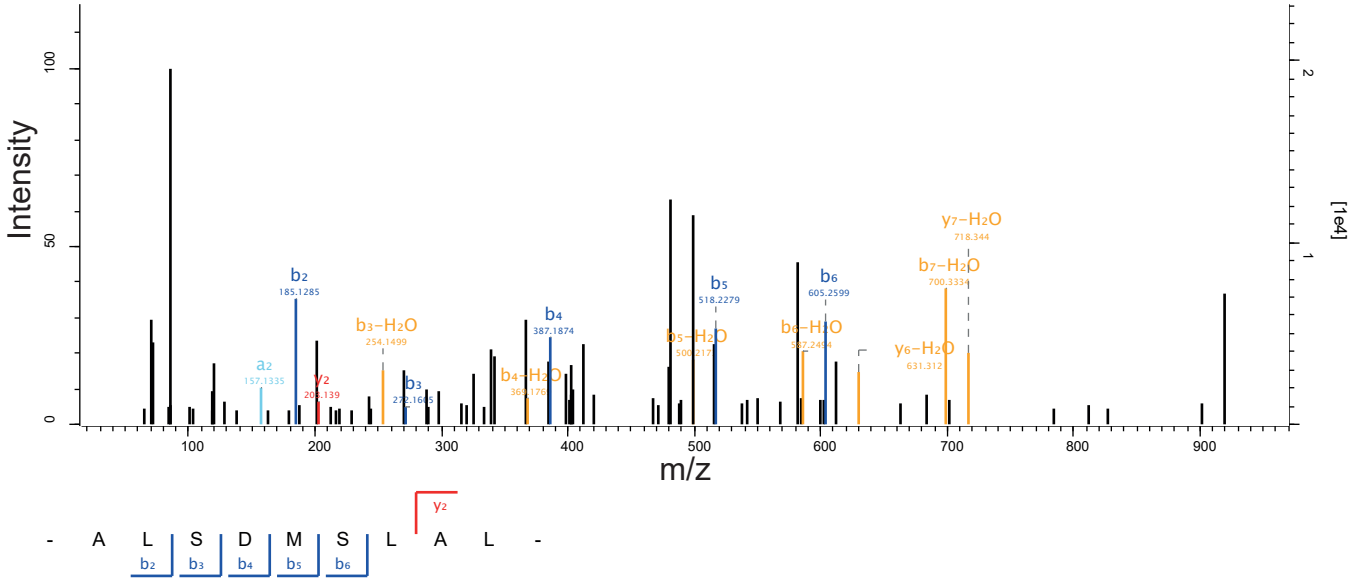
55B3 Sphingomonas melonis HLA-I AEGAAAIAA

Endogenous peptide



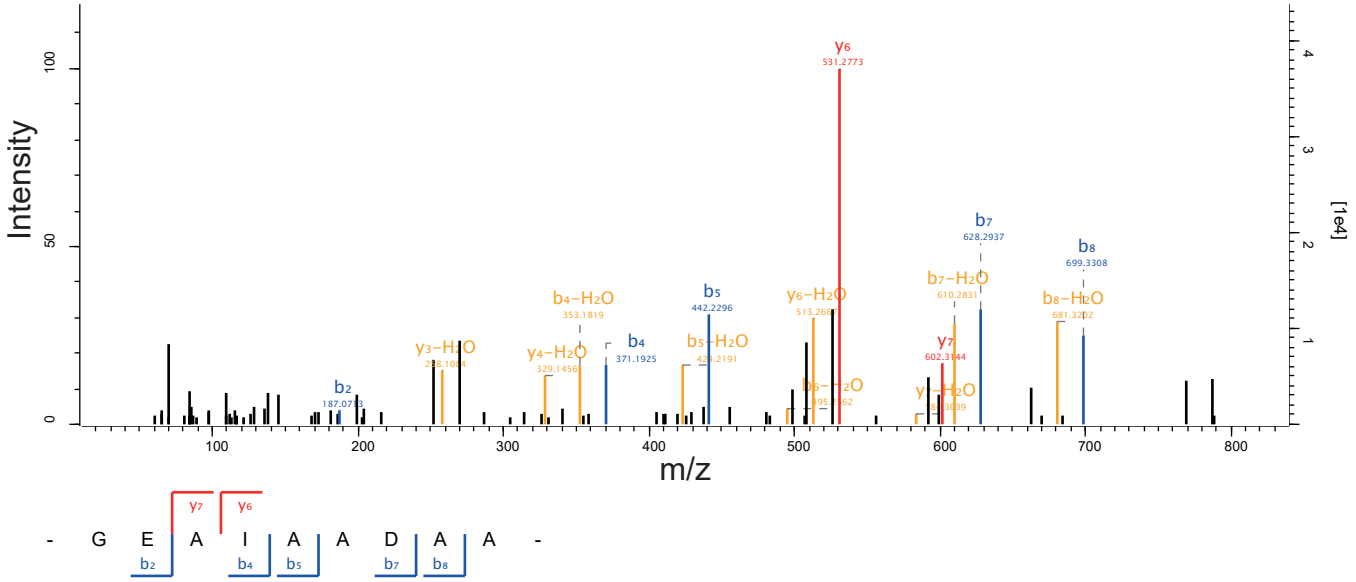
55B3 Sphingomonas melonis HLA-I ALSDMSLAL

Endogenous peptide



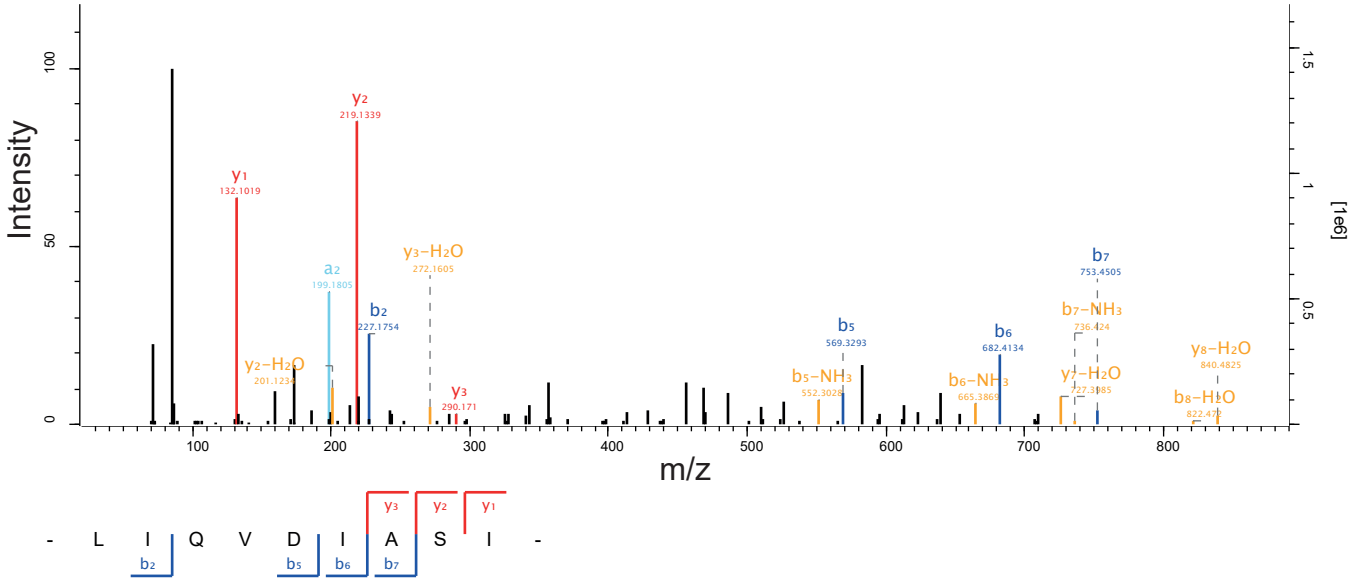
55B3 Sphingomonas melonis HLA-I GEAIAADAA

Endogenous peptide



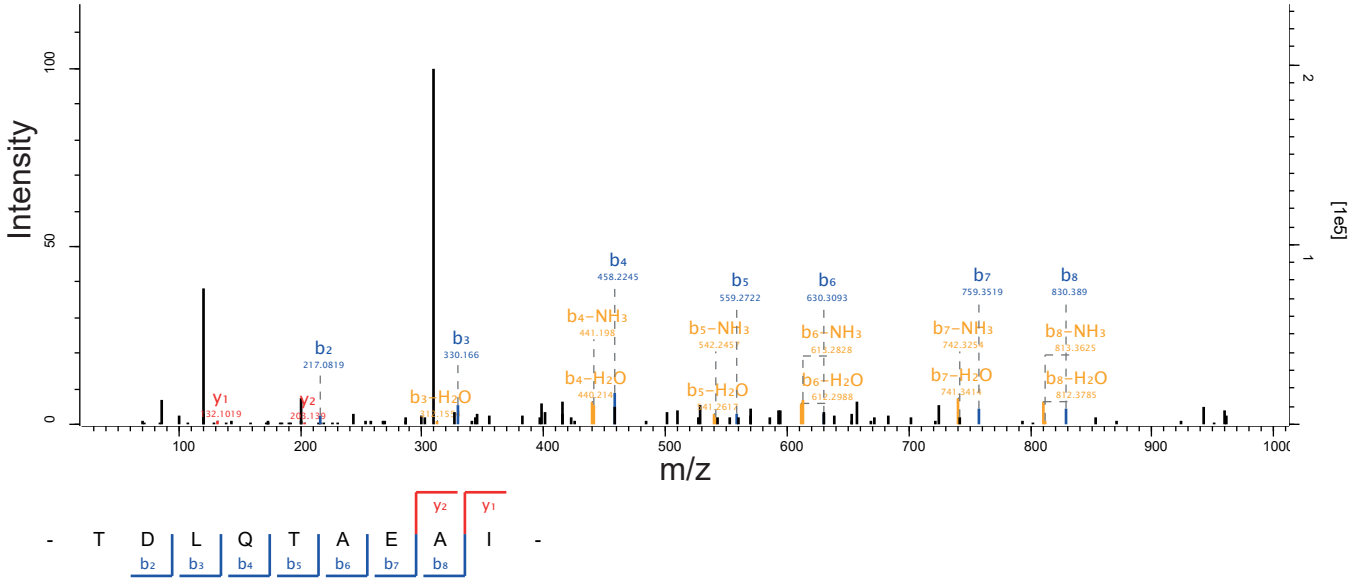
55B3 Sphingomonas melonis HLA-I LIQVDIASI

Endogenous peptide



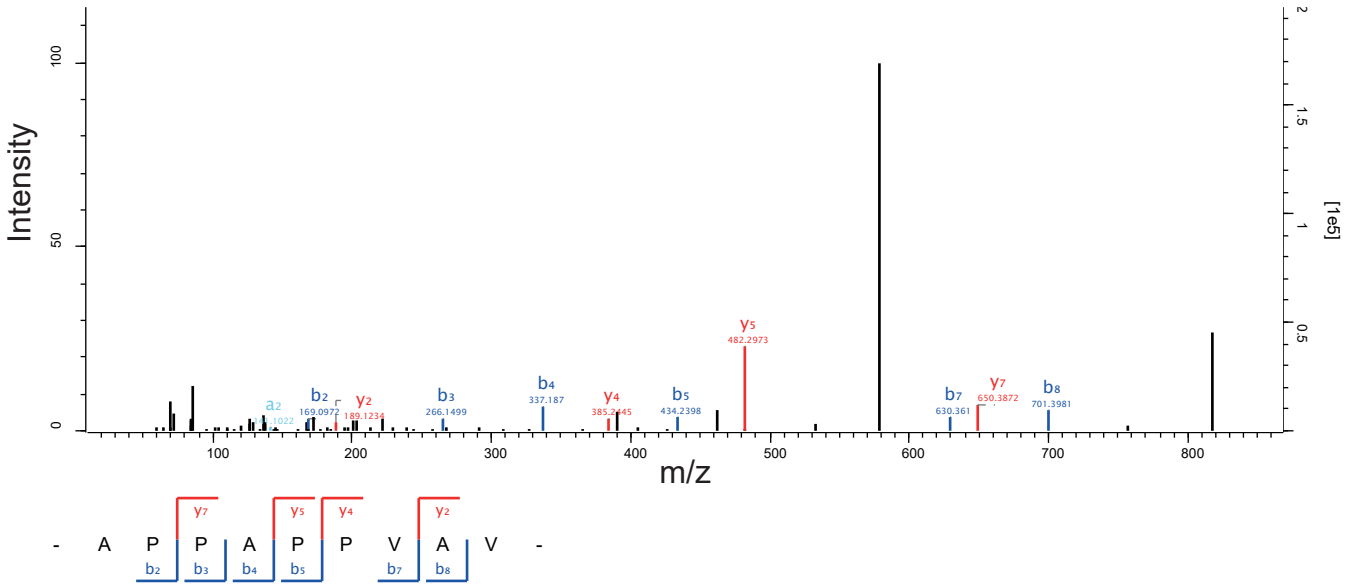
55B3 Sphingomonas melonis HLA-I TDLQTAEAI

Endogenous peptide



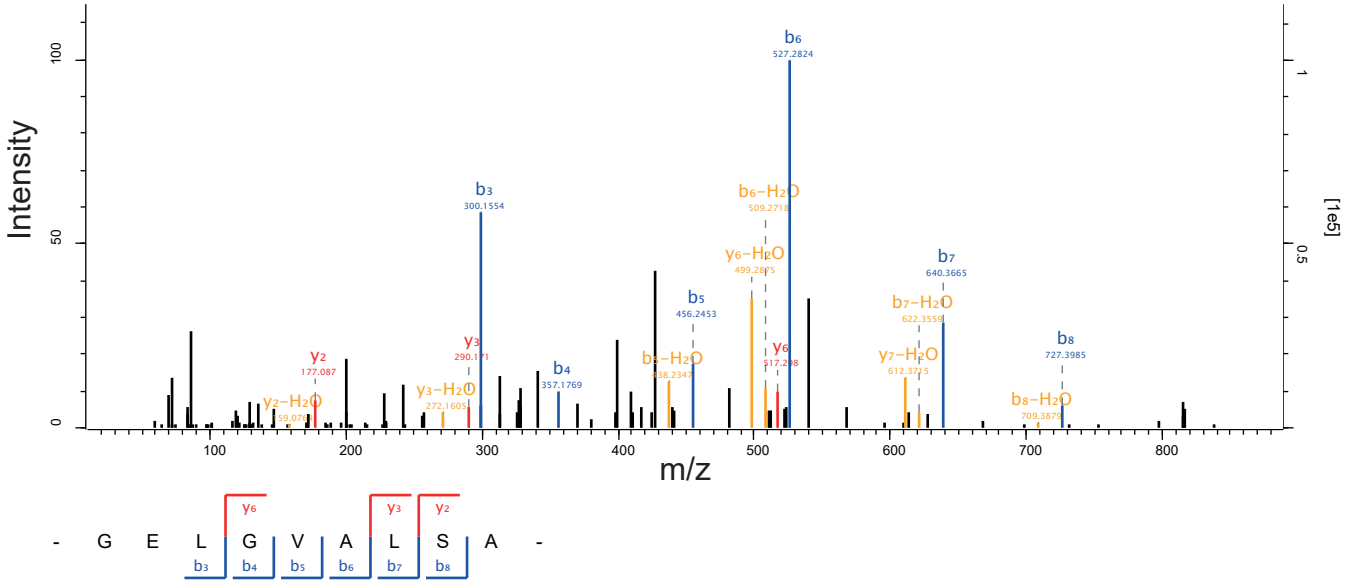
55B3 Staphylococcus capitis HLA-I APPAPPVAV

Endogenous peptide



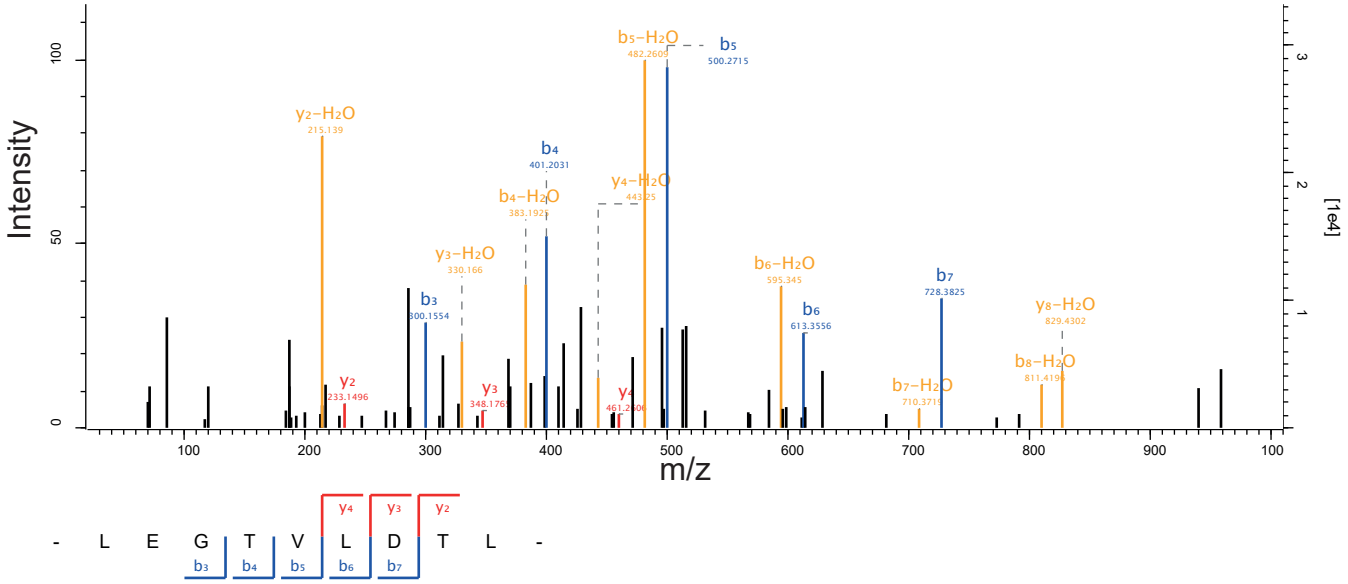
55B3 Staphylococcus capitis HLA-I GELGVALSA

Endogenous peptide



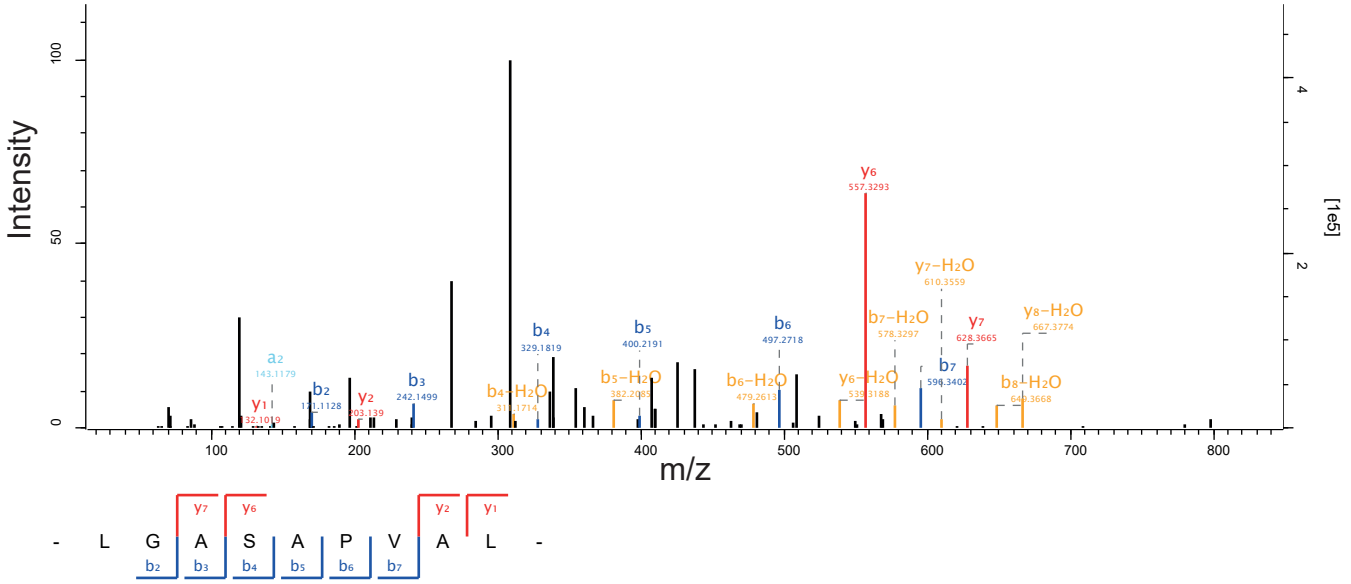
55B3 Staphylococcus capitis HLA-I LEGTVLDTL

Endogenous peptide



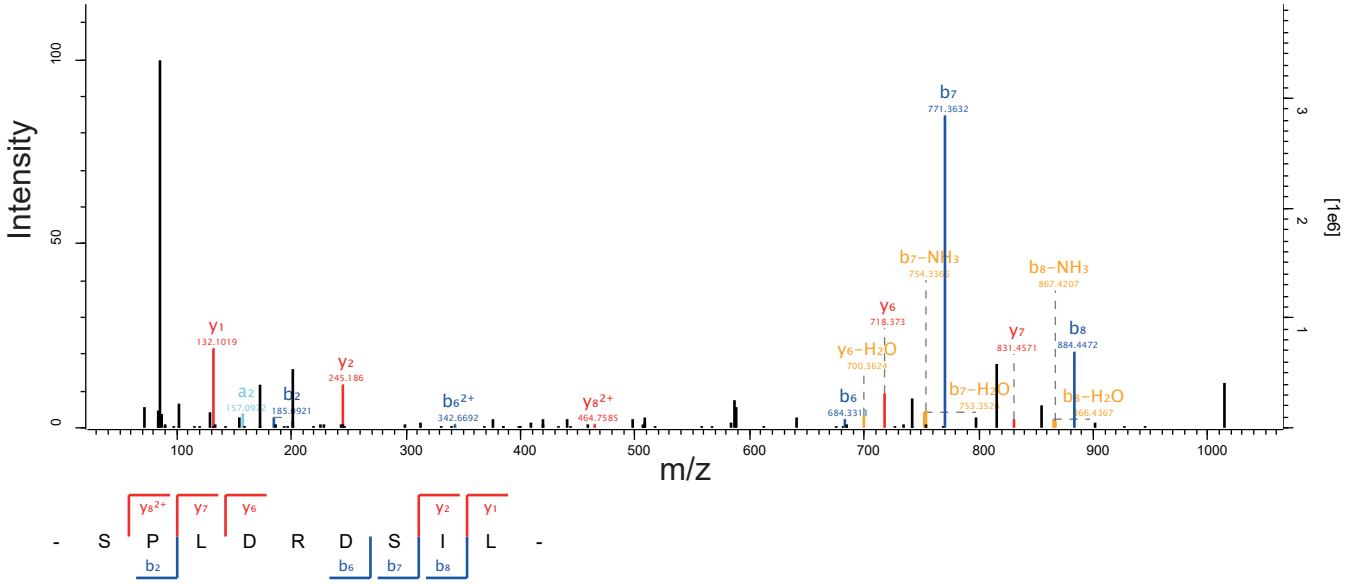
55B3 Staphylococcus capitis HLA-I LGASAPVAL

Endogenous peptide



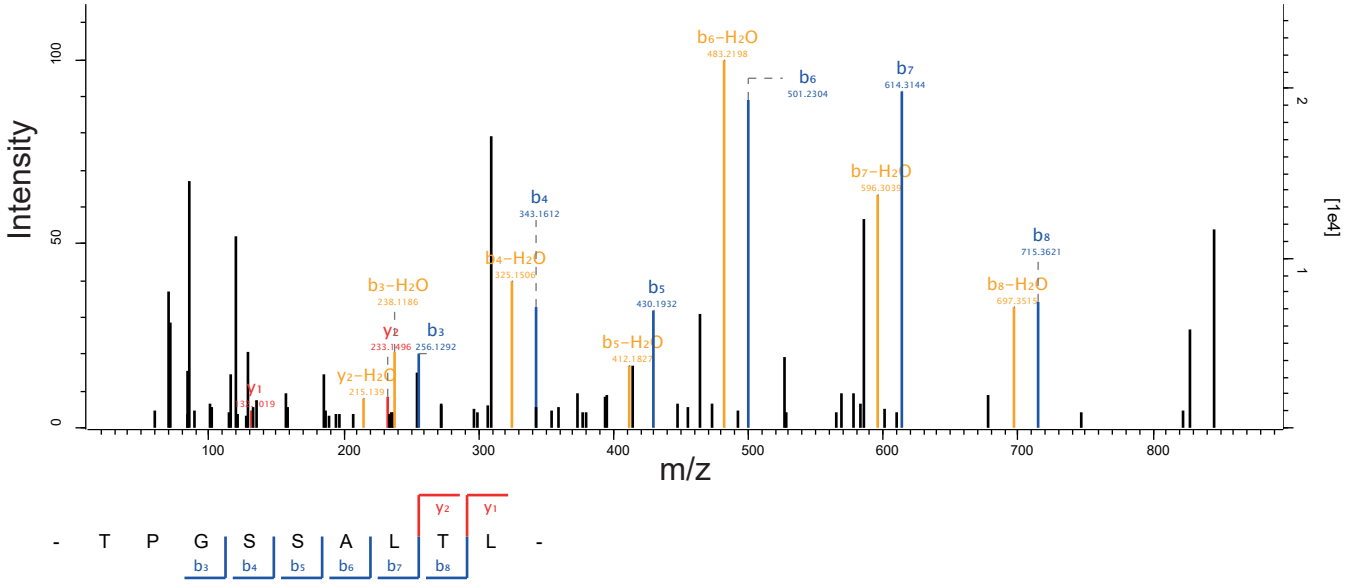
55B3 Staphylococcus capitis HLA-I SPLDRDSIL

Endogenous peptide



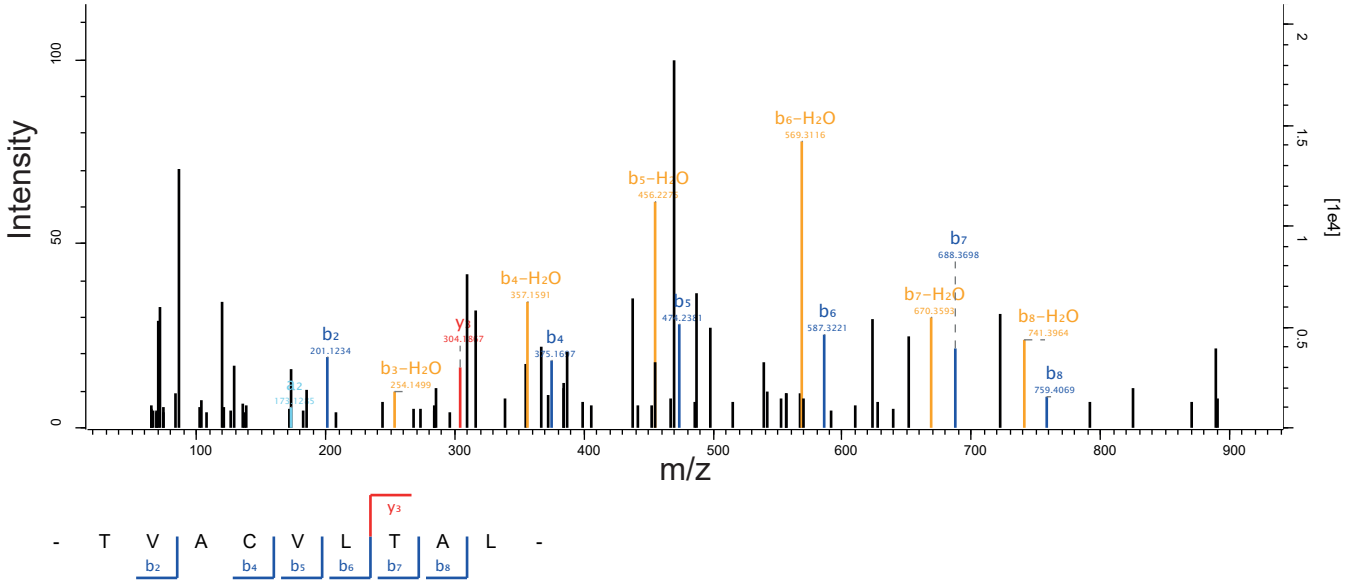
55B3 Staphylococcus capitis HLA-I TPGSSALTL

Endogenous peptide



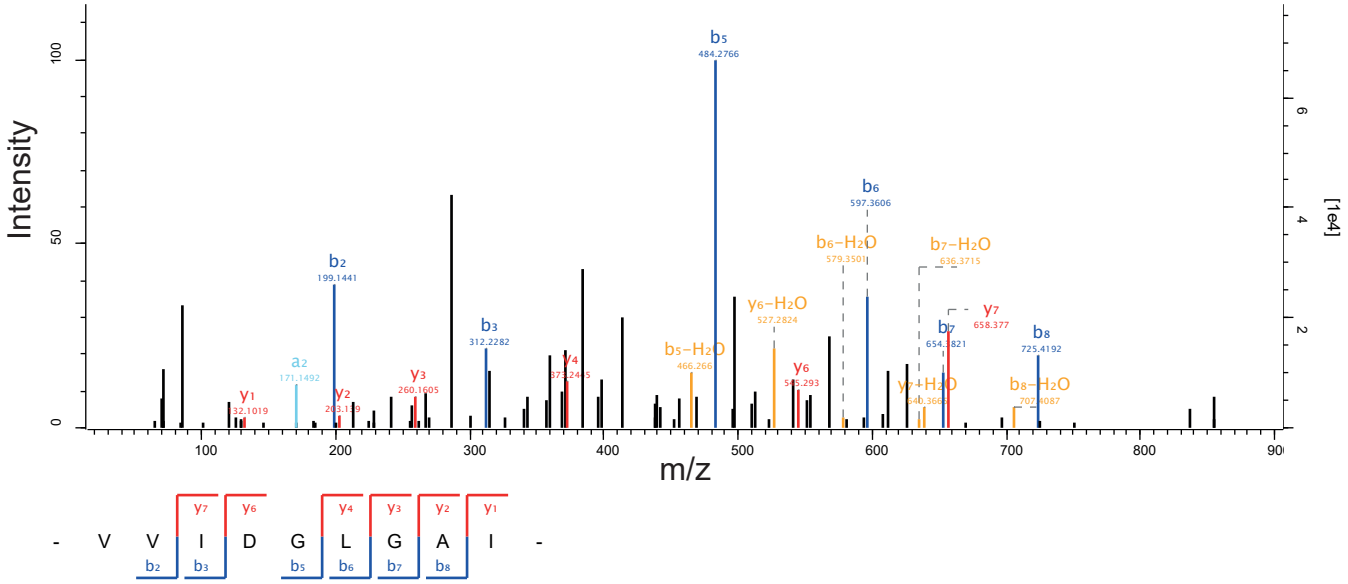
55B3 Staphylococcus capitis HLA-I TVACVLTAL

Endogenous peptide



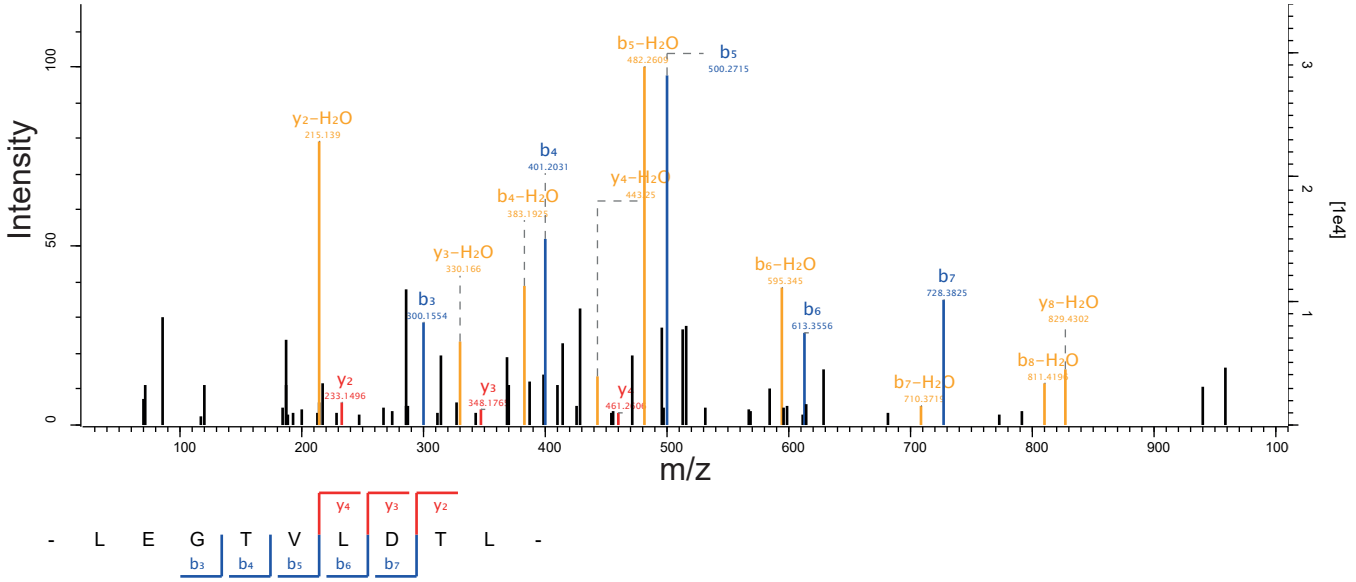
55B3 Staphylococcus capitis HLA-I VVIDGLGAI

Endogenous peptide



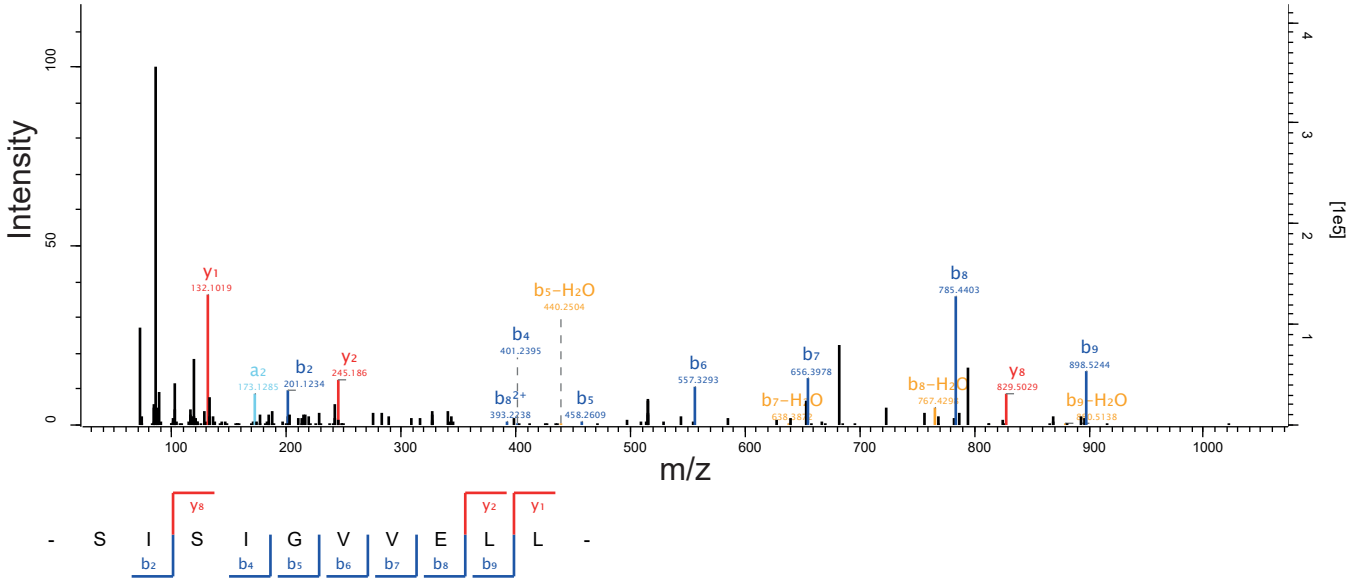
55B3 Staphylococcus caprae HLA-I LEGTVLDTL

Endogenous peptide



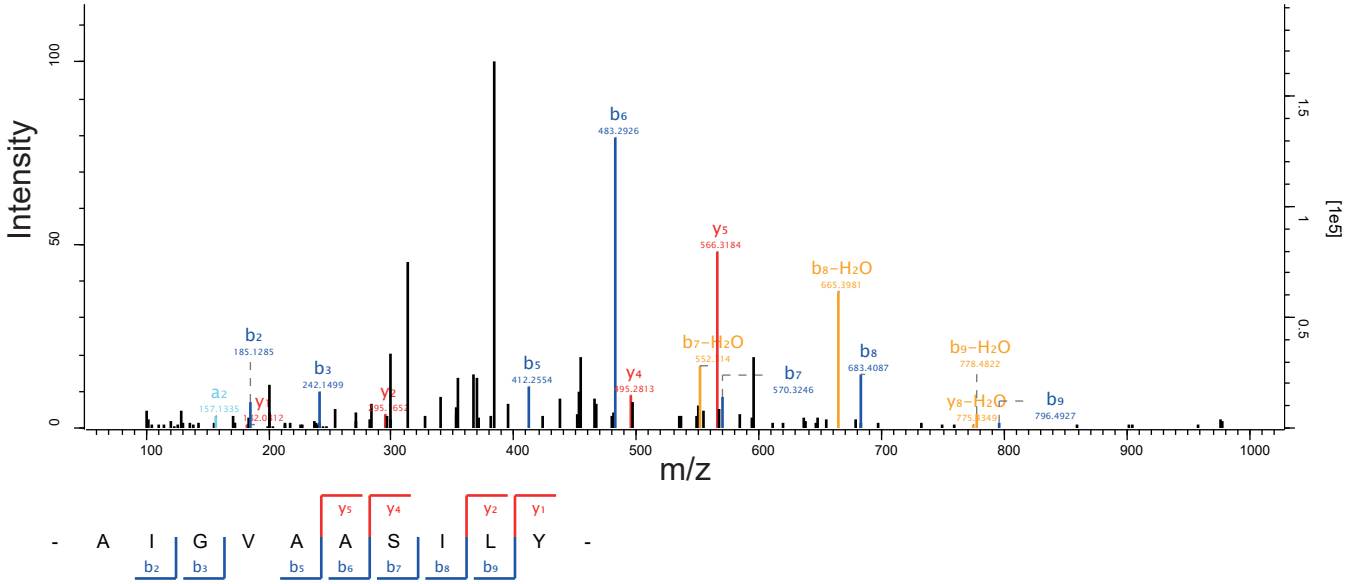
55B3 Staphylococcus caprae HLA-I SISIGVVELL

Endogenous peptide



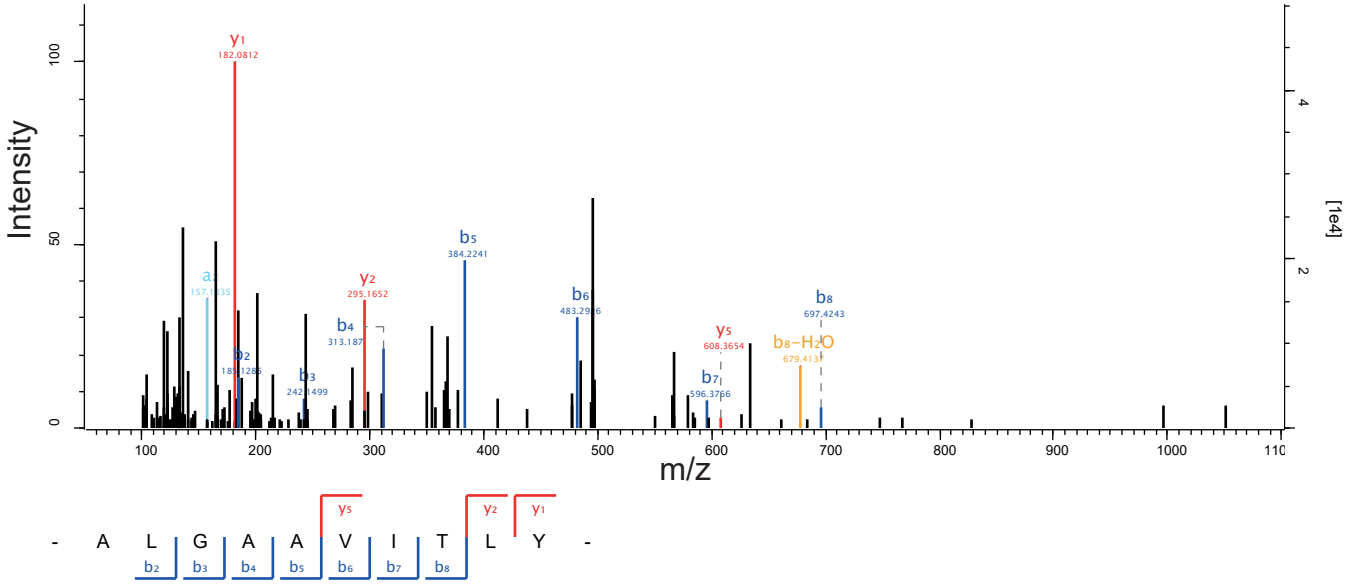
70.1 Staphylococcus capitis HLA-I AIGVAASILY

Endogenous peptide



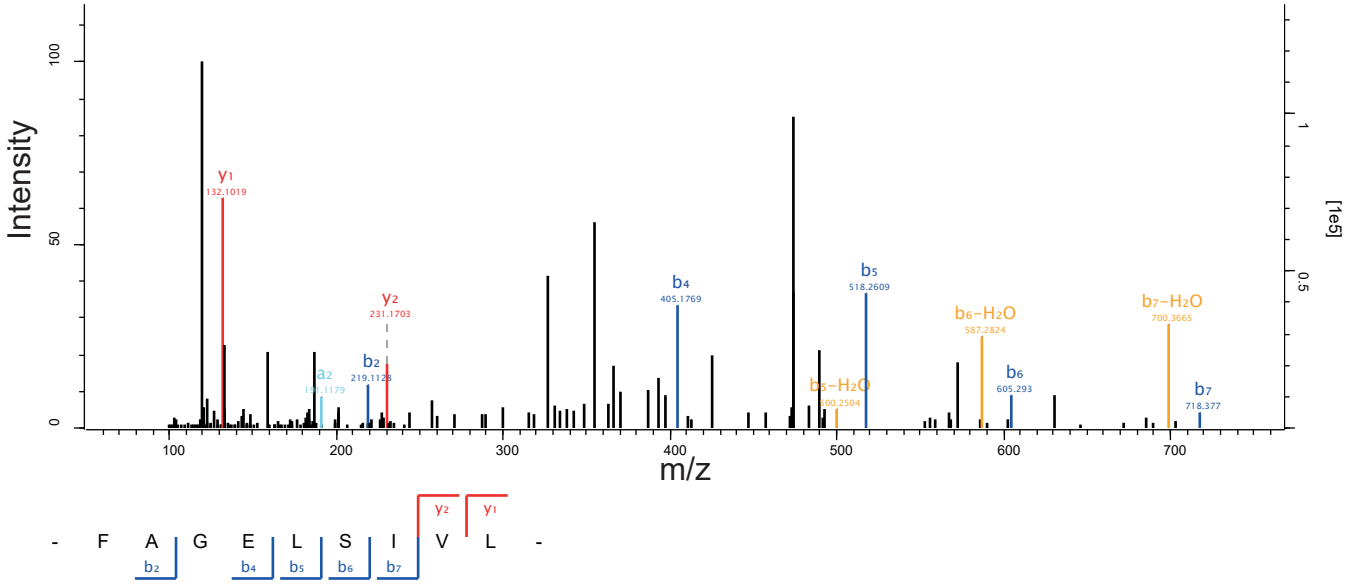
70.1 Staphylococcus capitis HLA-I ALGAAVITLY

Endogenous peptide



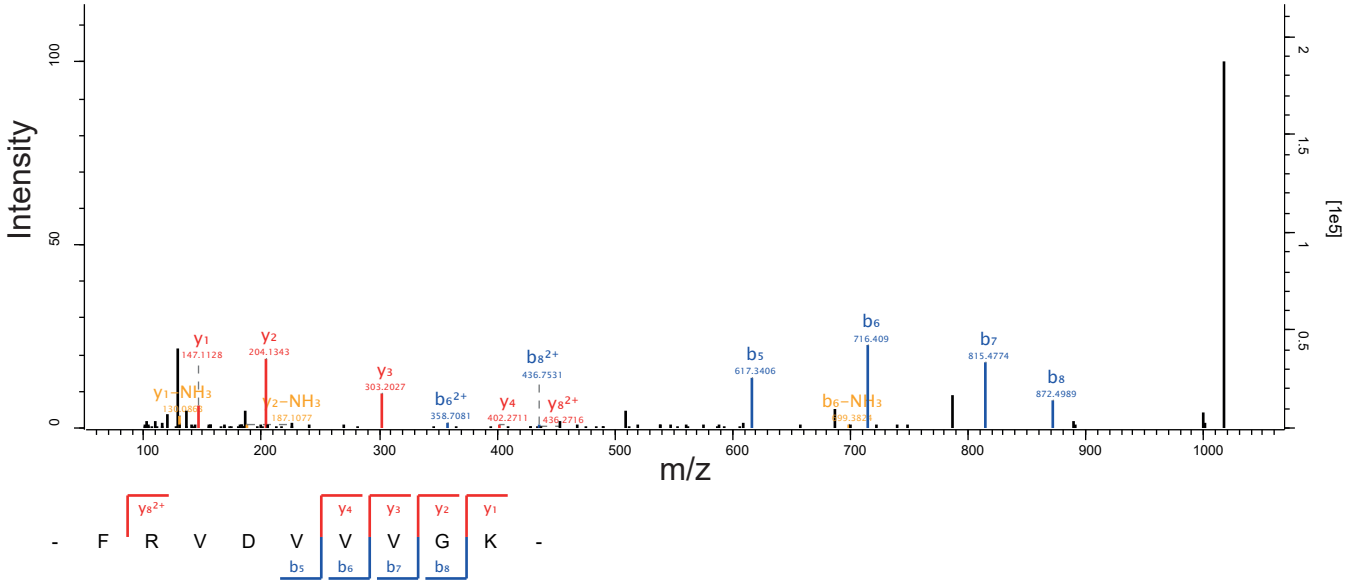
70.1 Staphylococcus capitis HLA-I FAGELSIVL

Endogenous peptide



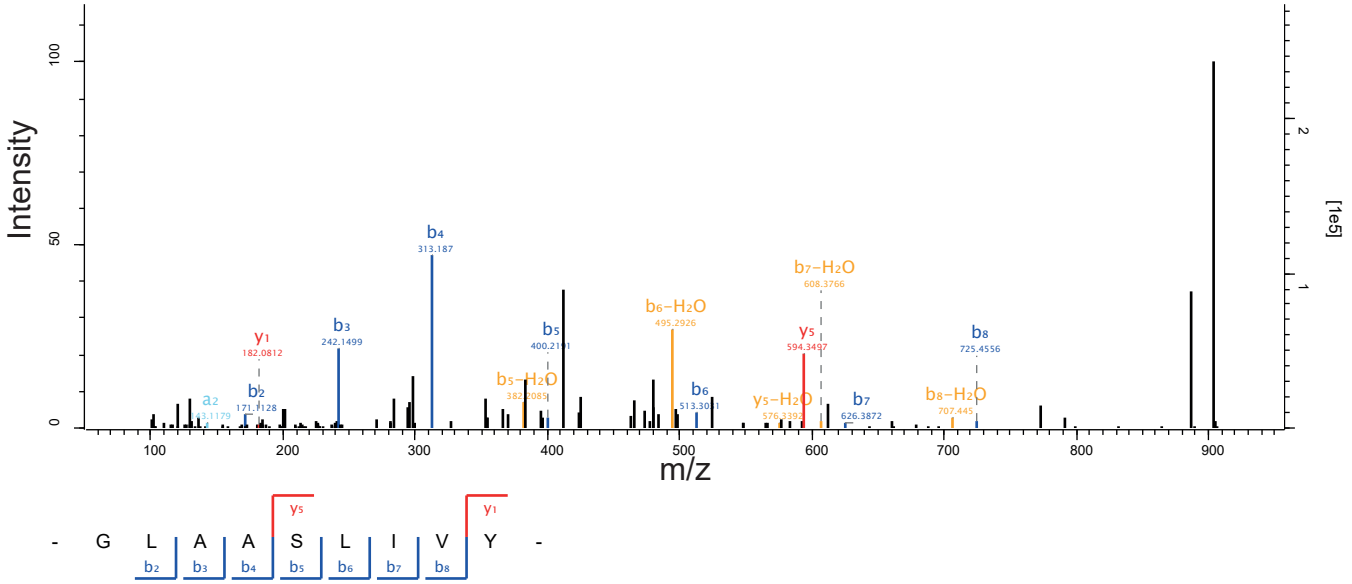
70.1 Staphylococcus capitis HLA-I FRVDVVVGK

Endogenous peptide



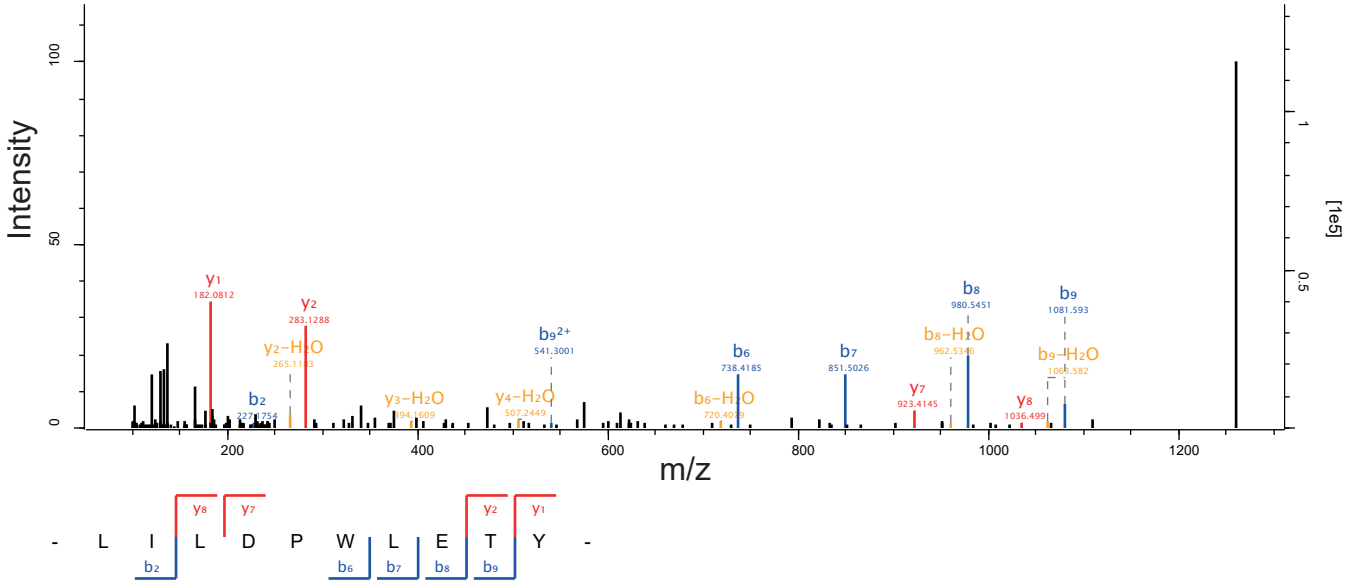
70.1 Staphylococcus capitis HLA-I GLAASLIVY

Endogenous peptide



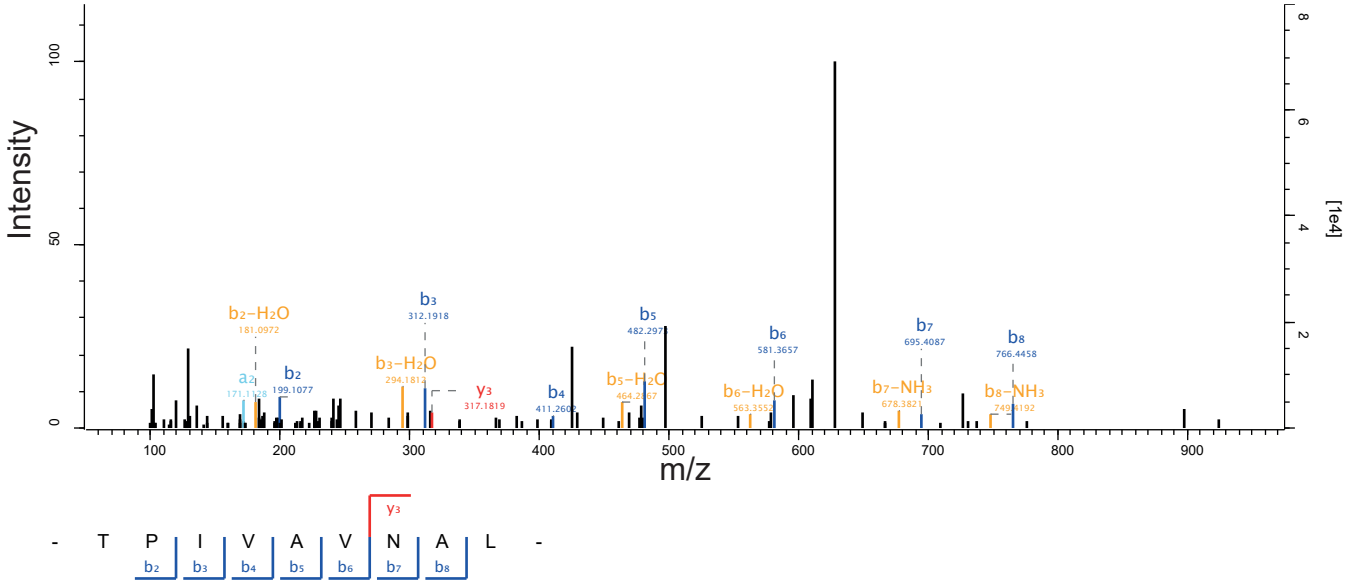
70.1 Staphylococcus capitis HLA-I LILDPWLETY

Endogenous peptide



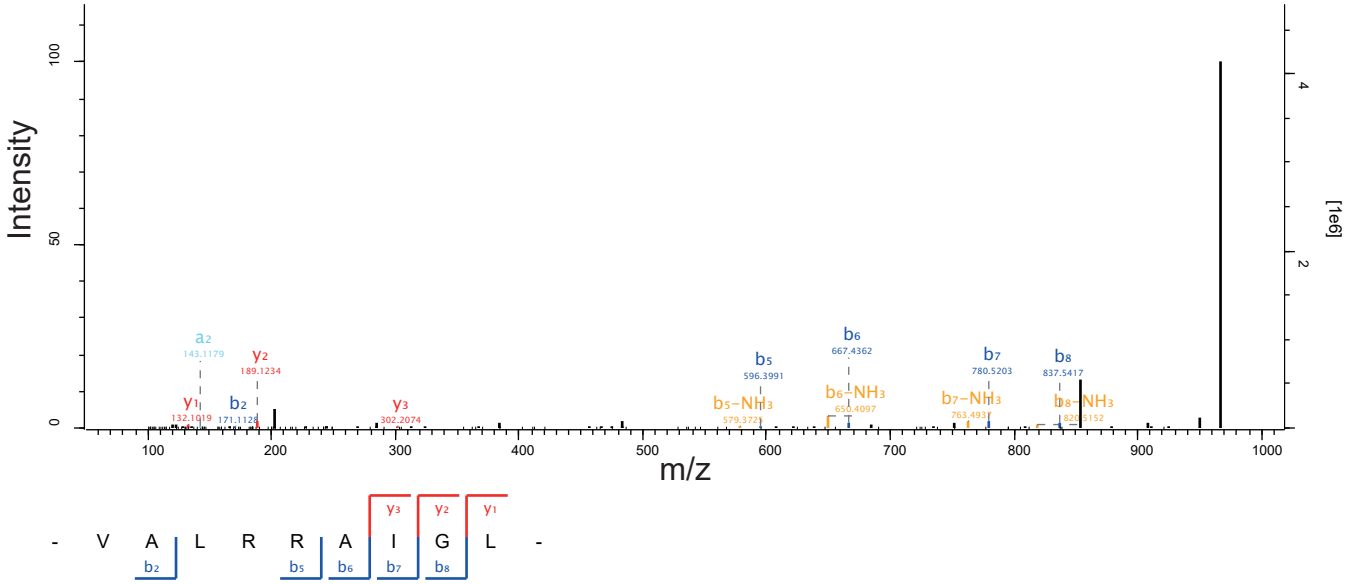
70.1 Staphylococcus capitis HLA-I TPIVAVNAL

Endogenous peptide



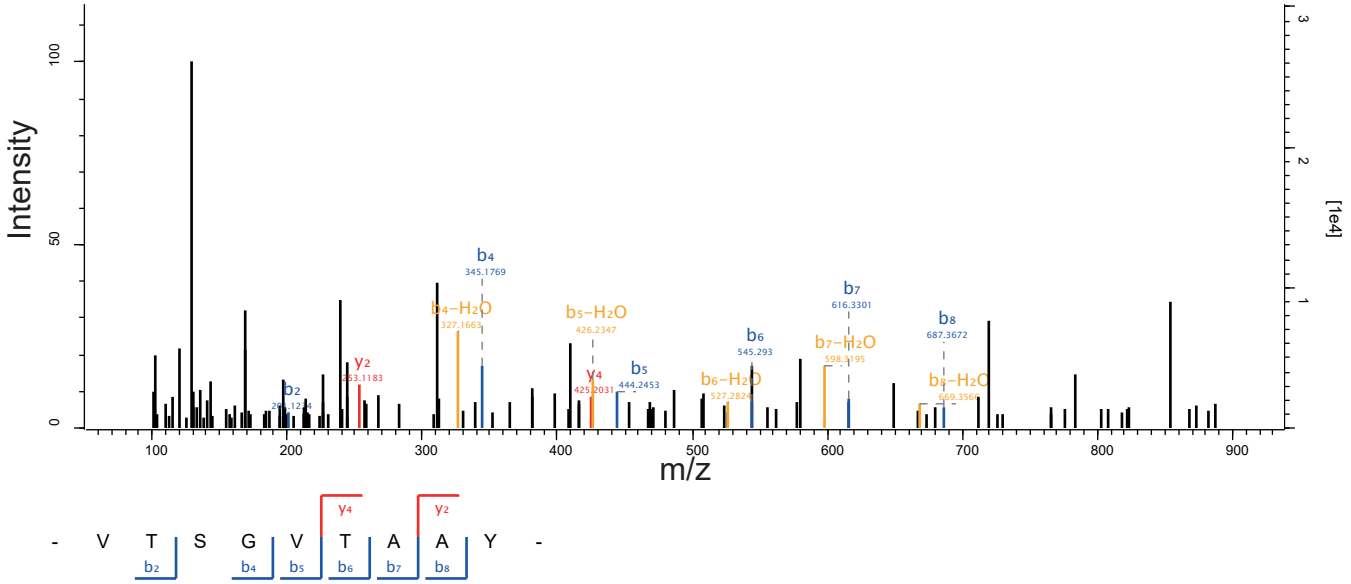
70.1 Staphylococcus capitis HLA-I VALRRAIGL

Endogenous peptide



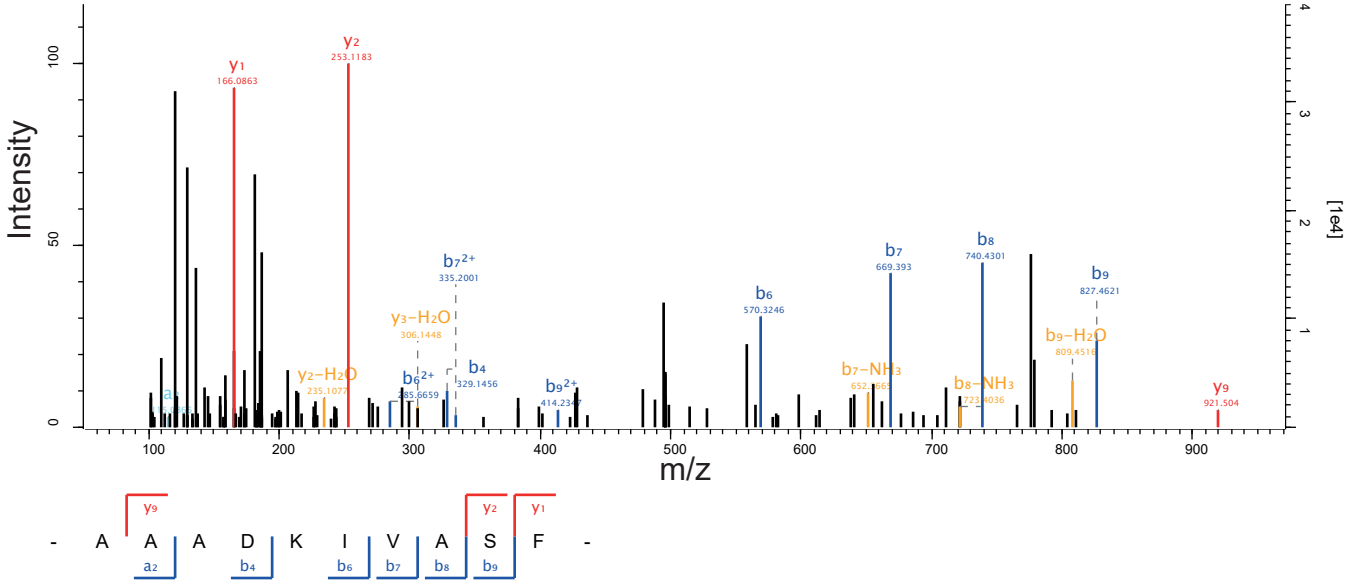
70.1 Staphylococcus capitis HLA-I VTSGVTAAY

Endogenous peptide



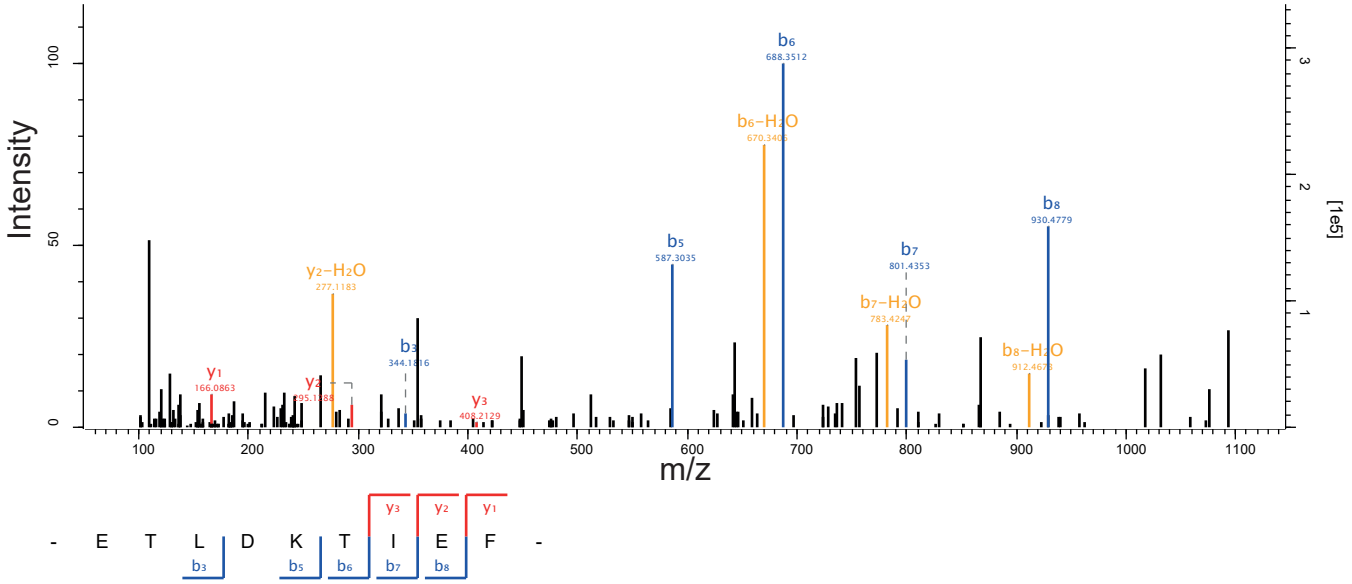
70.1 Streptococcus gordonii HLA-I AAADKIVASF

Endogenous peptide



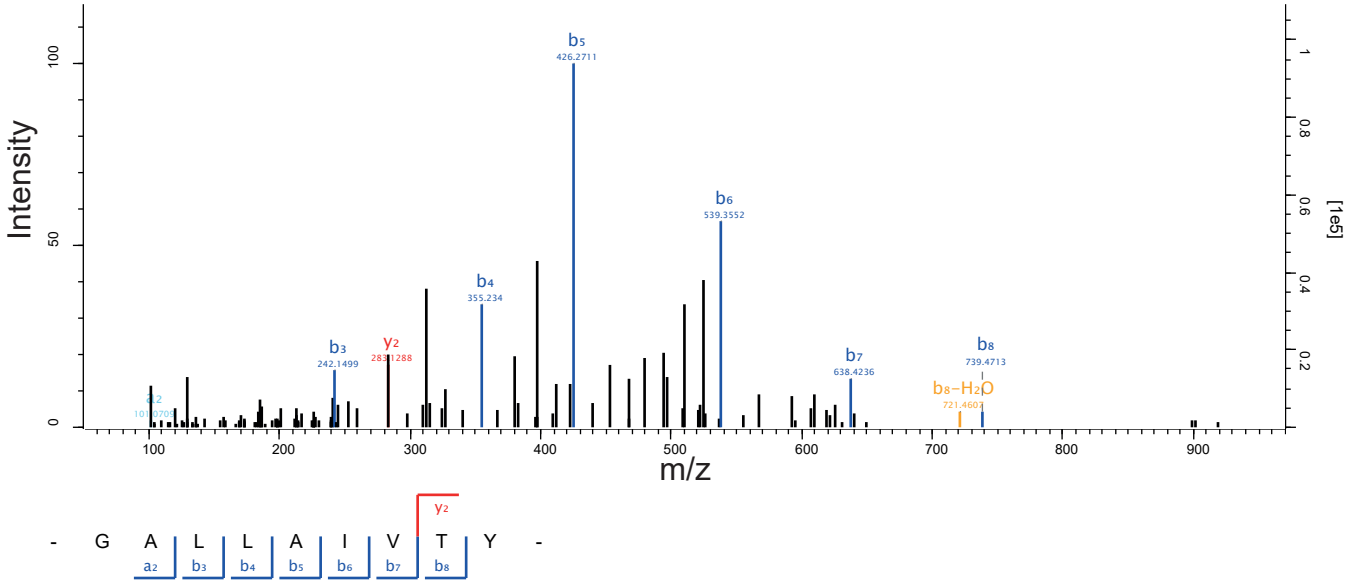
70.1 Streptococcus gordonii HLA-I ETLDKTIEF

Endogenous peptide



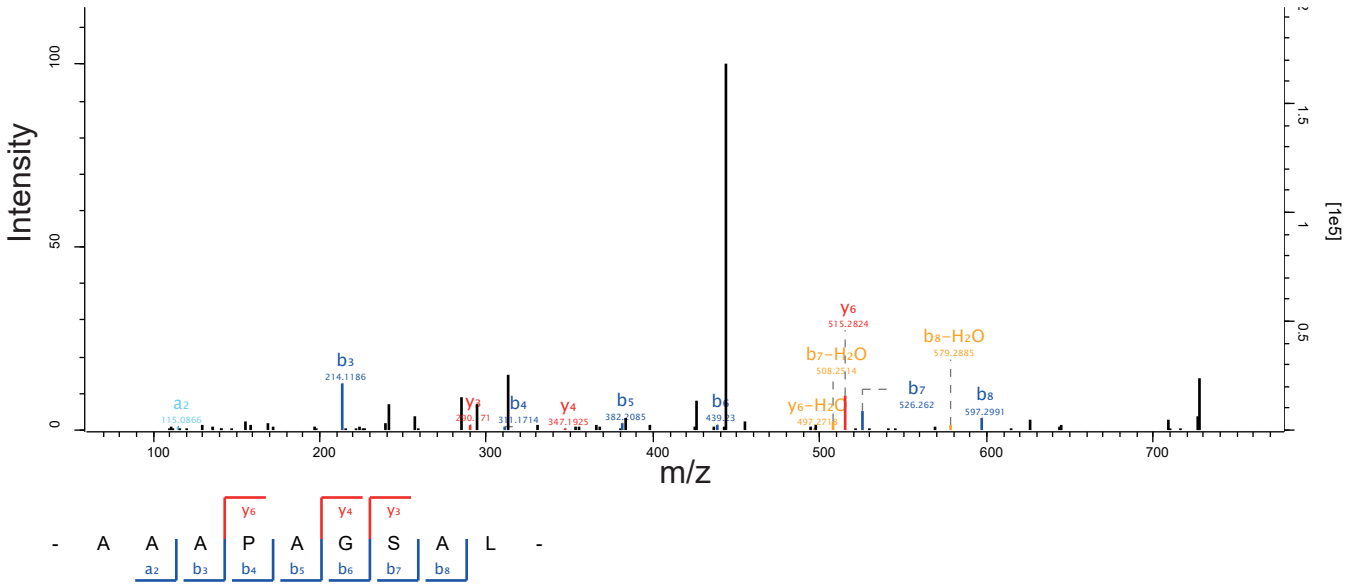
70.1 Streptococcus gordonii HLA-I GALLAIVTY

Endogenous peptide



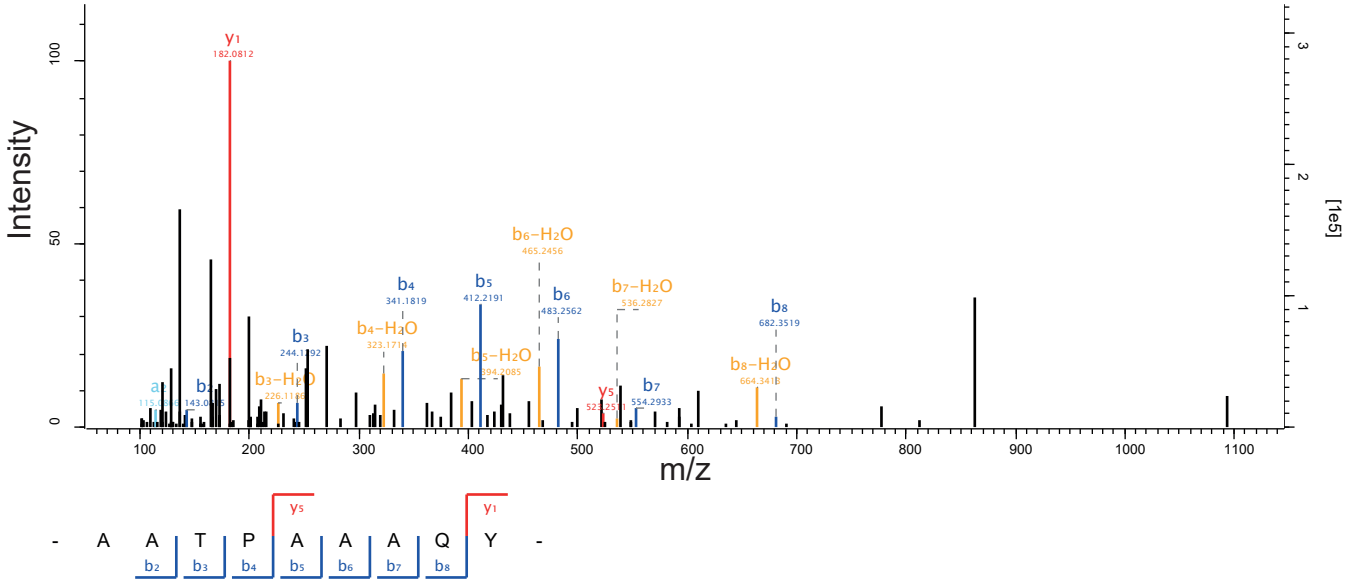
86B Corynebacterium humireducens HLA-I AAAPAGSAL

Endogenous peptide



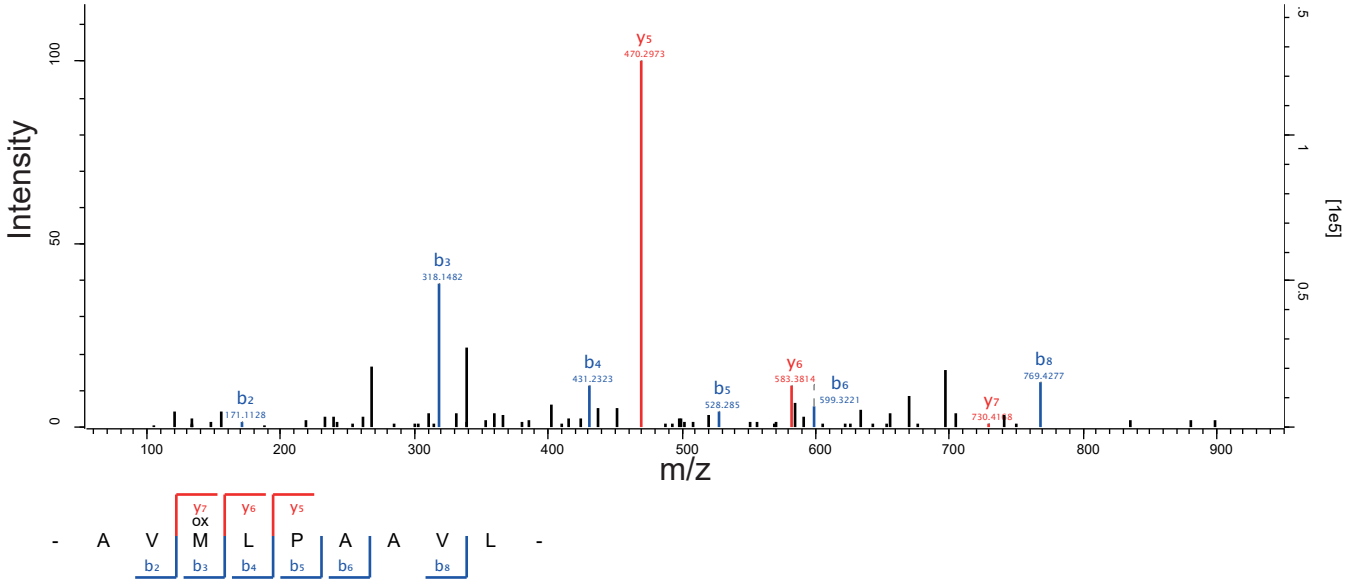
86B Corynebacterium humireducens HLA-I AATPAAAQY

Endogenous peptide



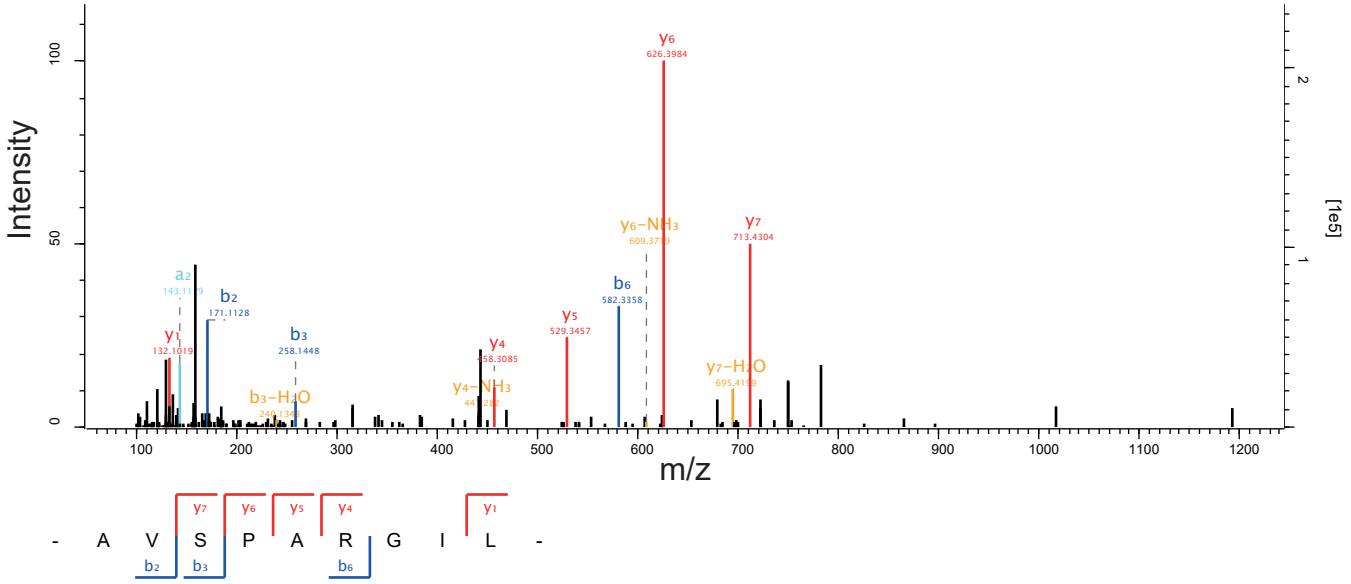
86B Corynebacterium humireducens HLA-I AVMLPAAVL

Endogenous peptide



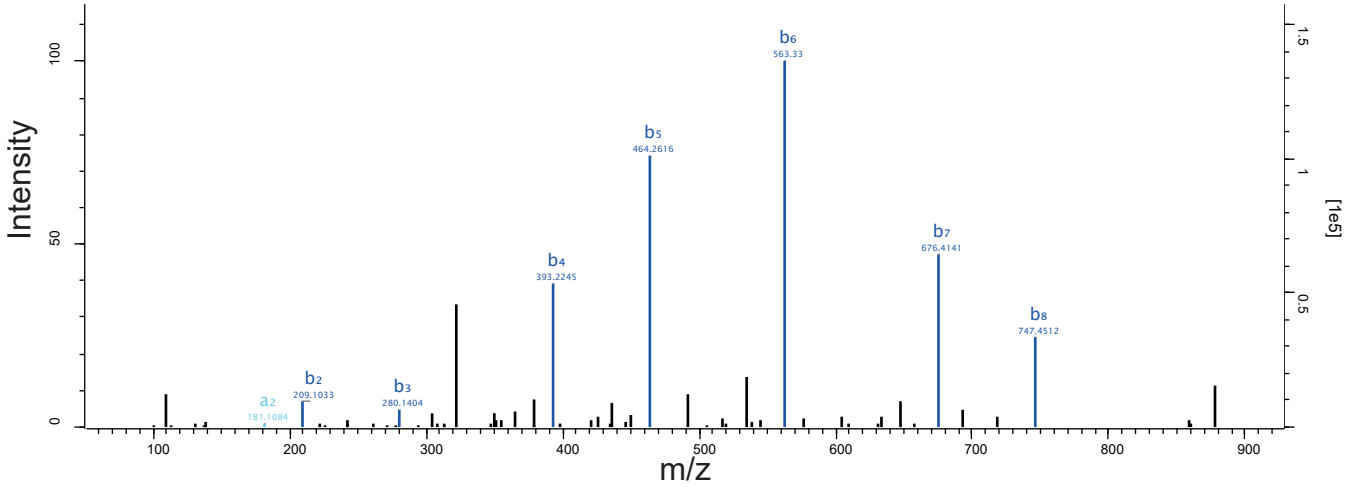
86B Corynebacterium humireducens HLA-I AVSPARGIL

Endogenous peptide



86B Corynebacterium humireducens HLA-I HAALAVIAL

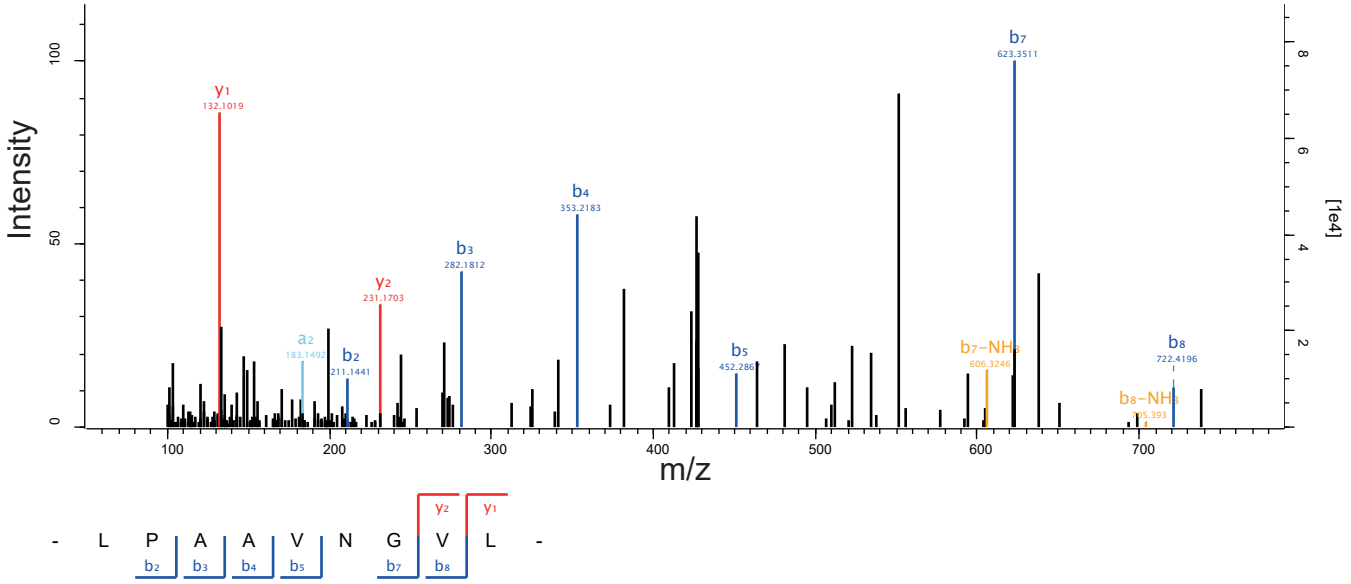
Endogenous peptide



- H A A L A V I A L -
b2 b3 b4 b5 b6 b7 b8

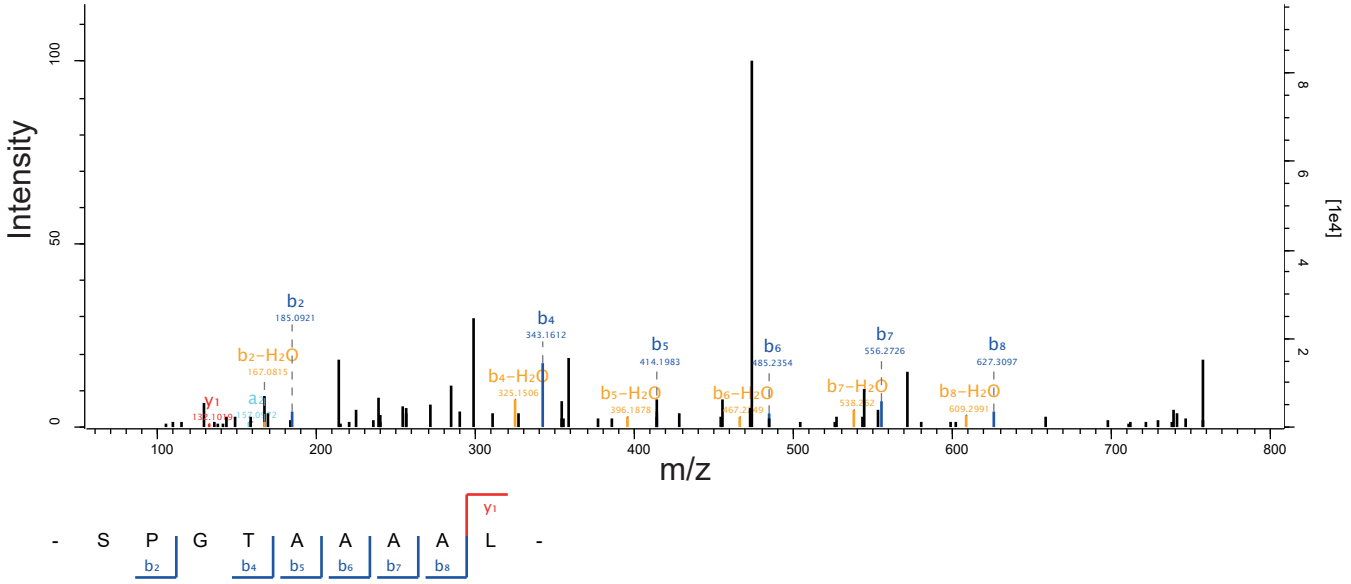
86B Corynebacterium humireducens HLA-I LPAAVNGVL

Endogenous peptide



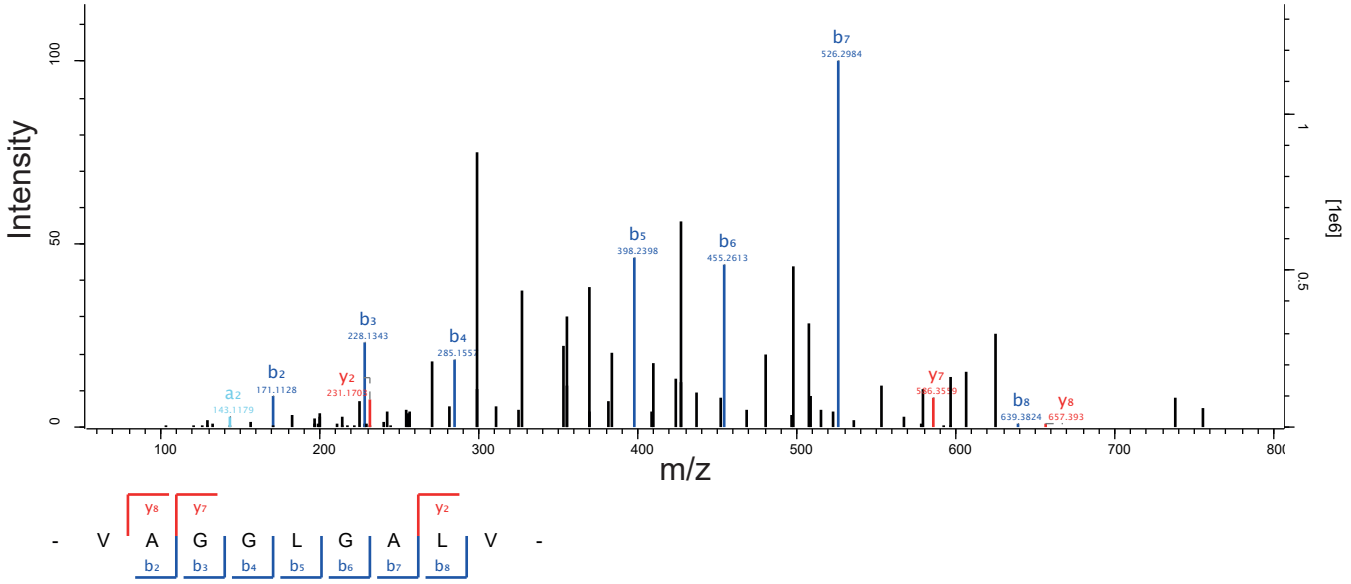
86B Corynebacterium humireducens HLA-I SPGTAAAAL

Endogenous peptide



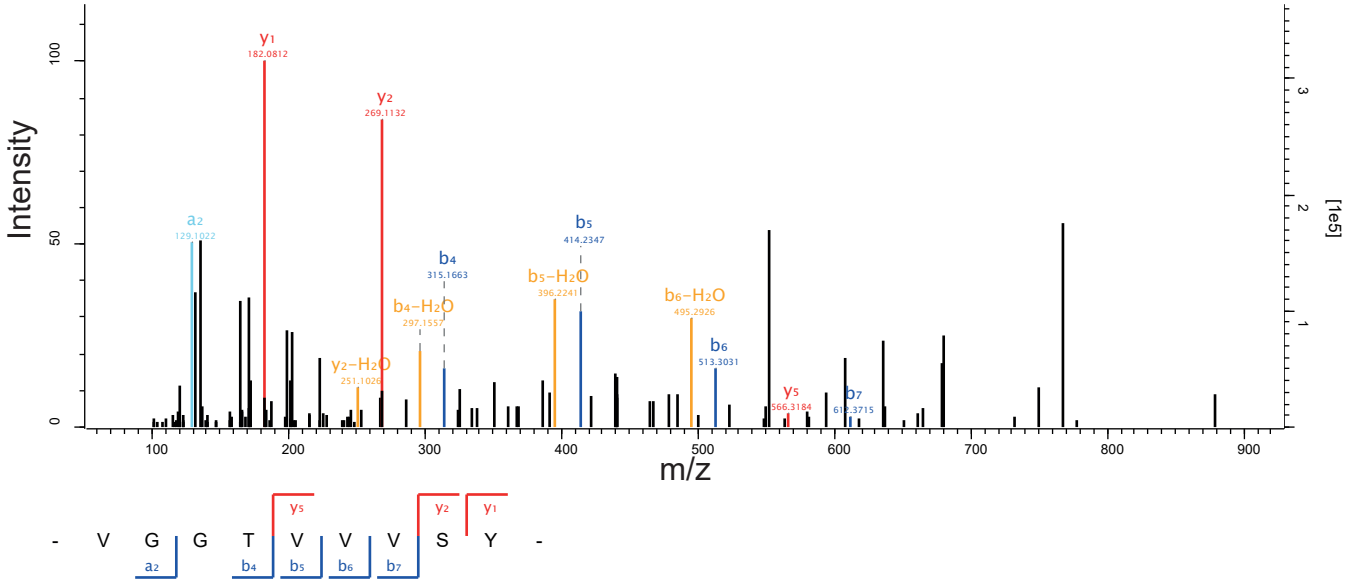
86B Corynebacterium humireducens HLA-I VAGGLGALV

Endogenous peptide



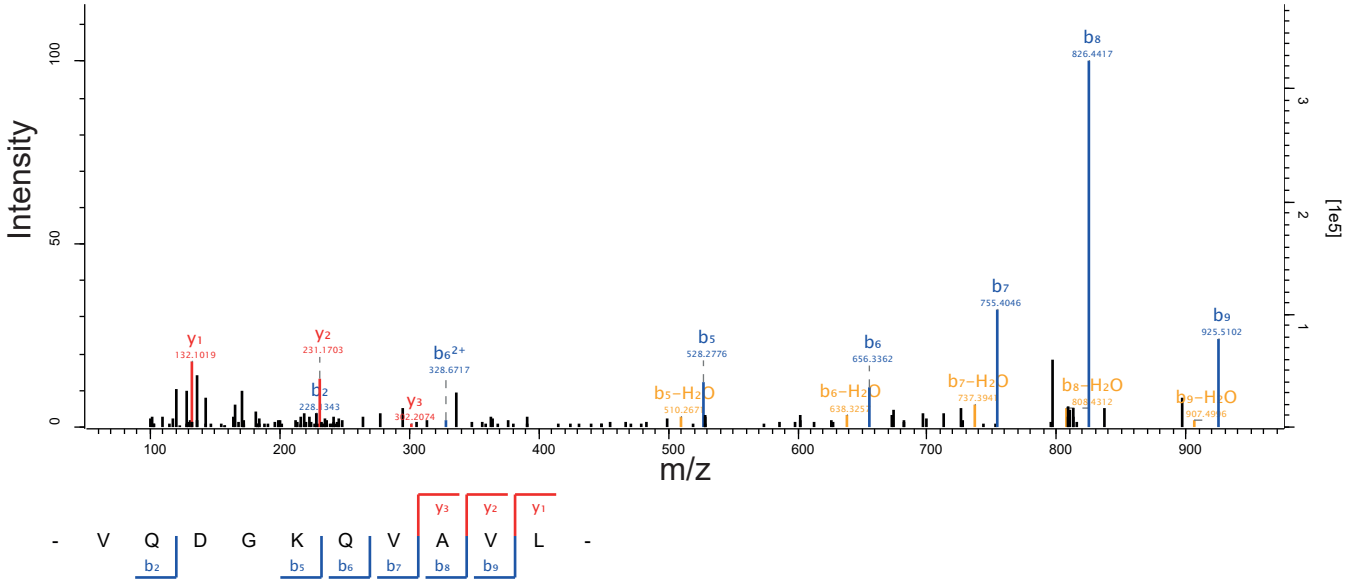
86B Corynebacterium humireducens HLA-I VGGTVVVSY

Endogenous peptide



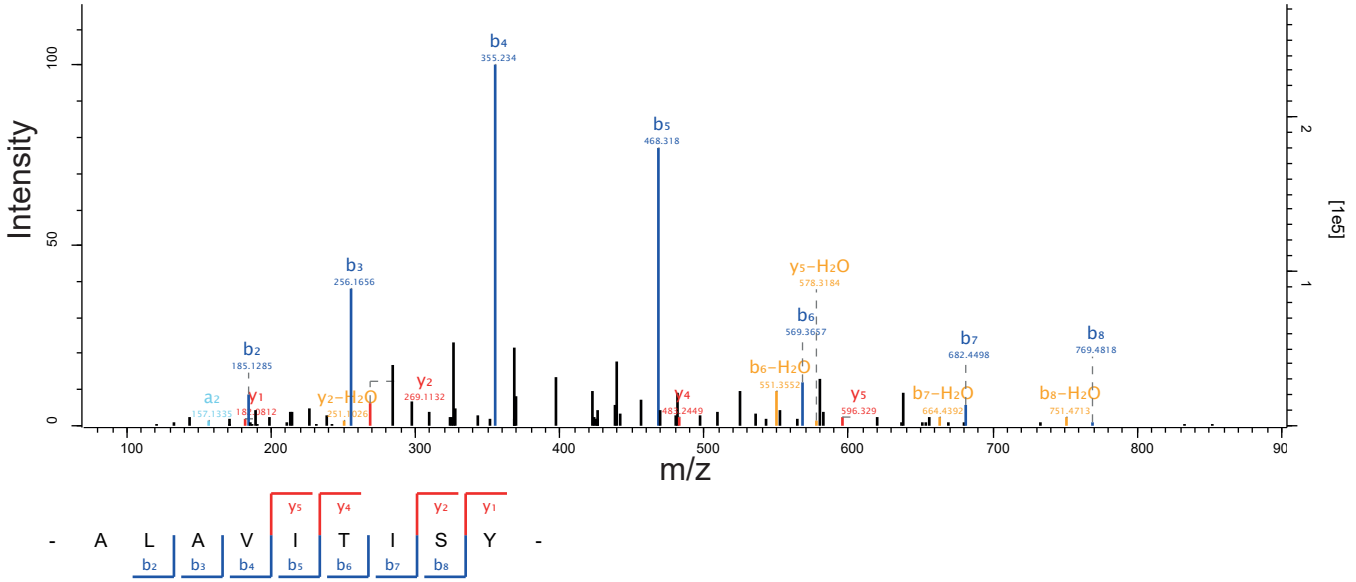
86B Corynebacterium humireducens HLA-I VQDGKQVAVL

Endogenous peptide



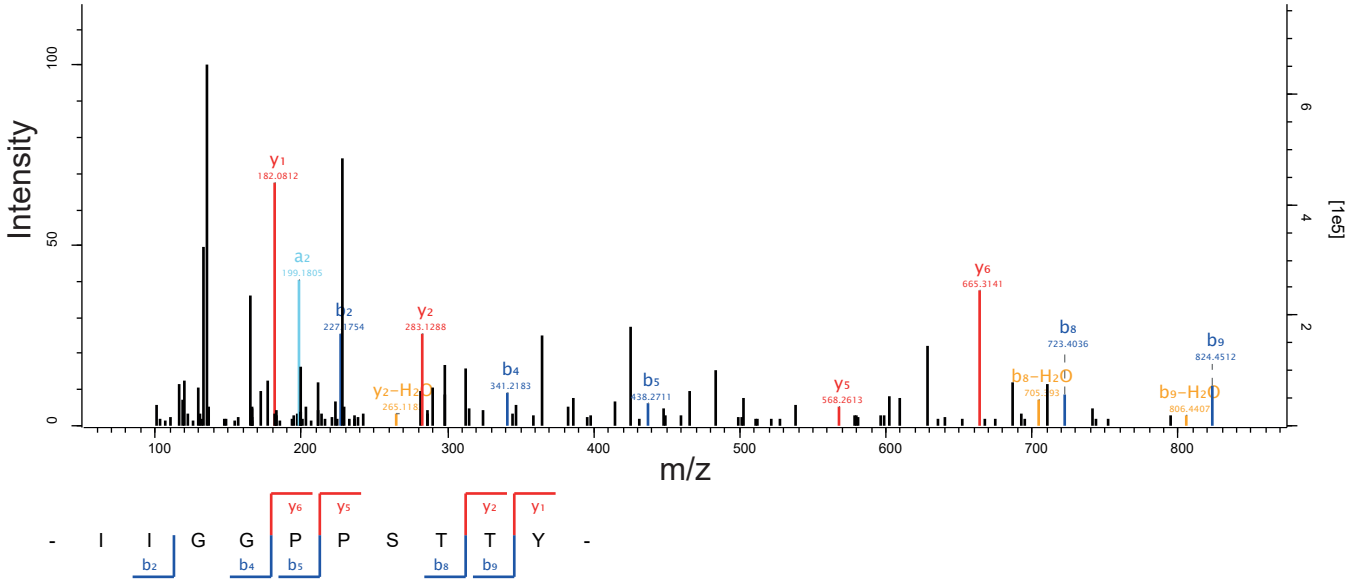
86B Staphylococcus lugdunensis HLA-I ALAVITISY

Endogenous peptide



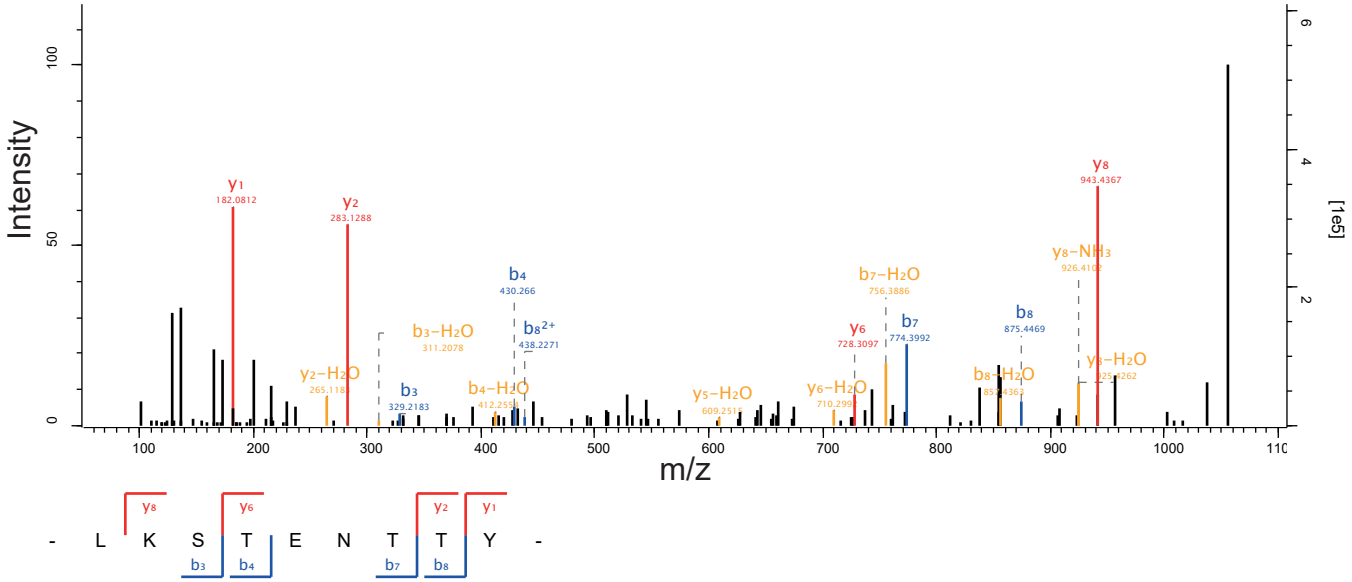
86B Staphylococcus lugdunensis HLA-I IIGGPPSTTY

Endogenous peptide



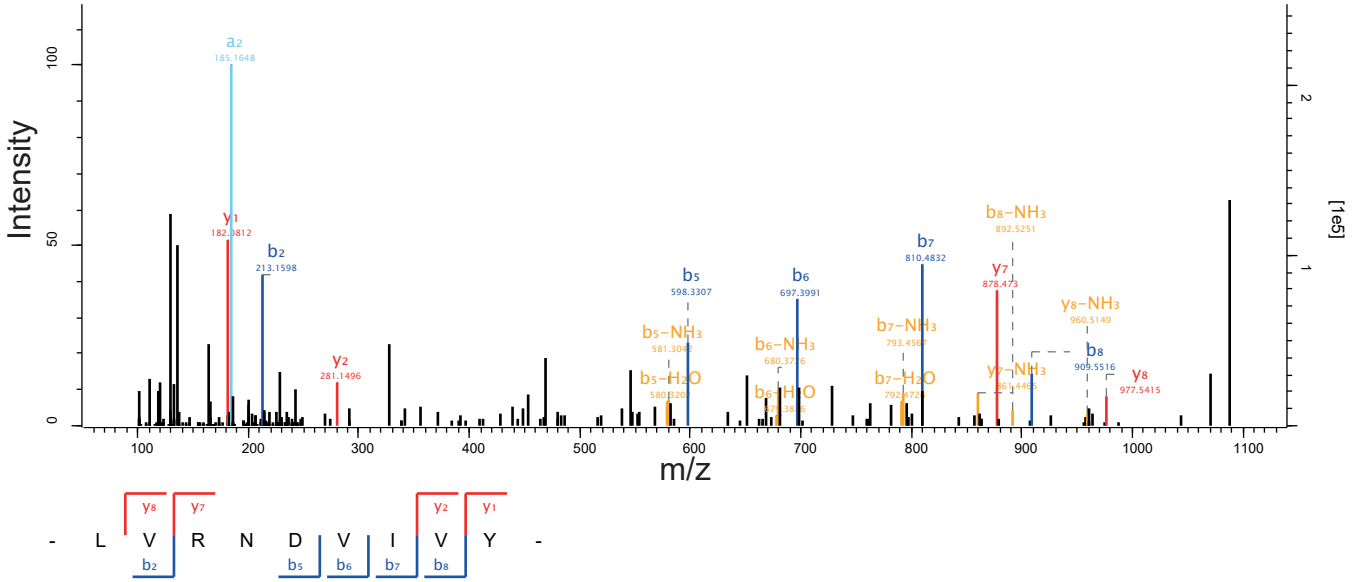
86B Staphylococcus lugdunensis HLA-I LKSTENTTY

Endogenous peptide



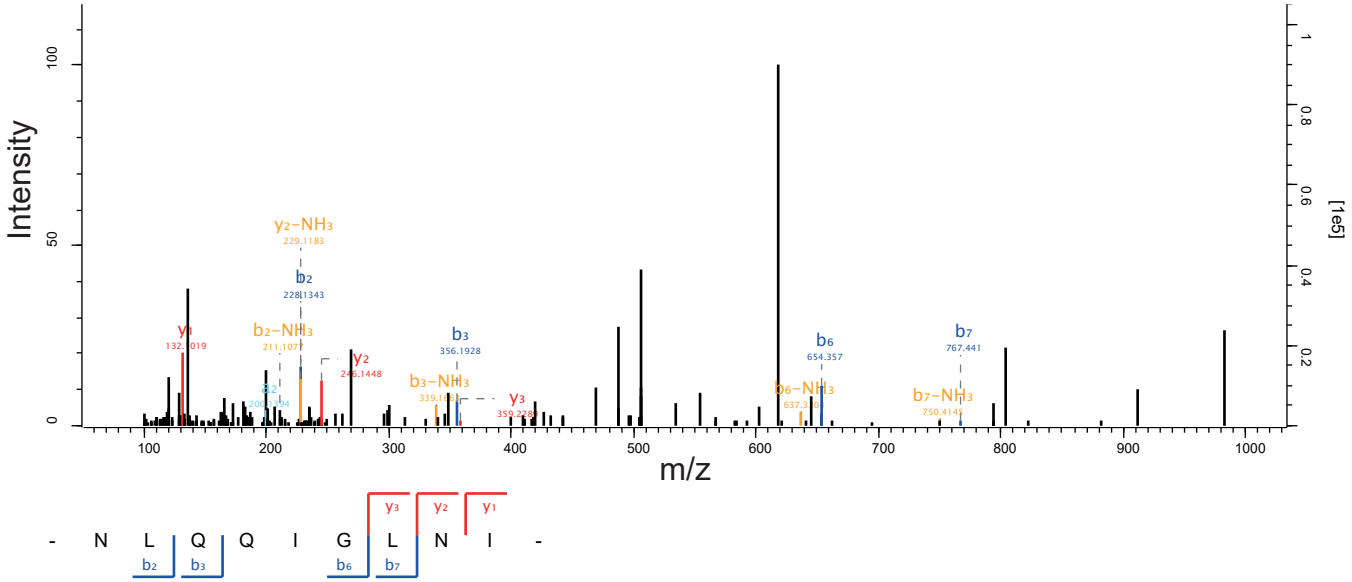
86B Staphylococcus lugdunensis HLA-I LVRNDVIVY

Endogenous peptide



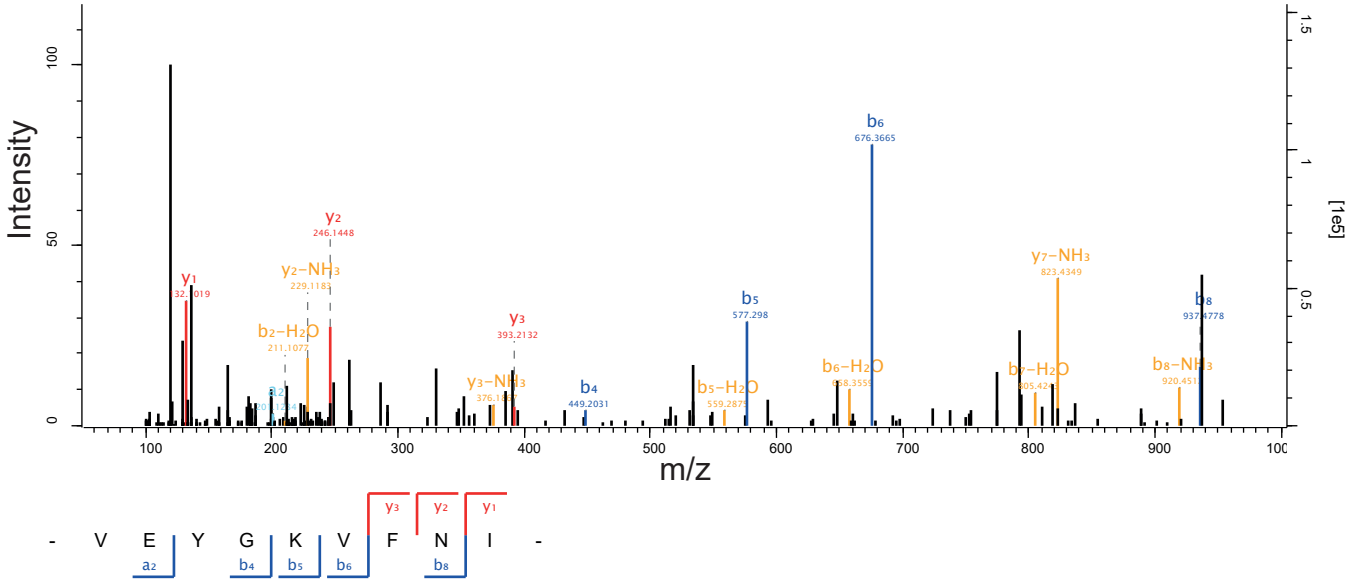
86B Staphylococcus lugdunensis HLA-I NLQQIGLNI

Endogenous peptide



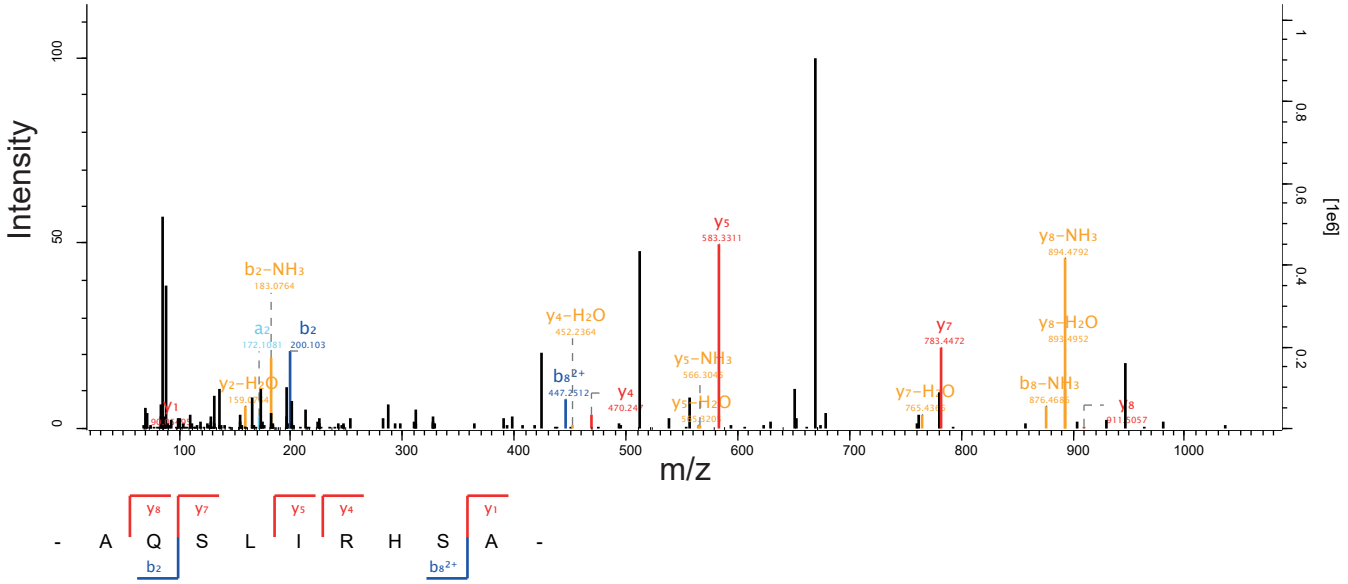
86B Staphylococcus lugdunensis HLA-I VEYGKVFNI

Endogenous peptide



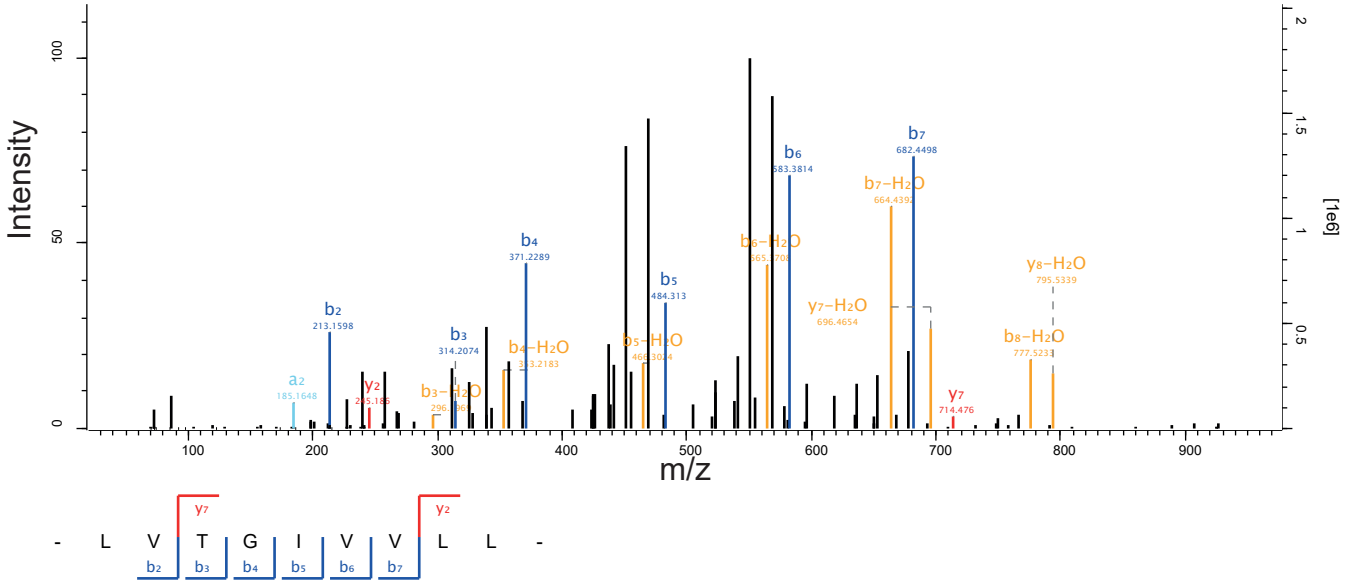
86B2 Acinetobacter cumulans HLA-I AQSLIRHSA

Endogenous peptide



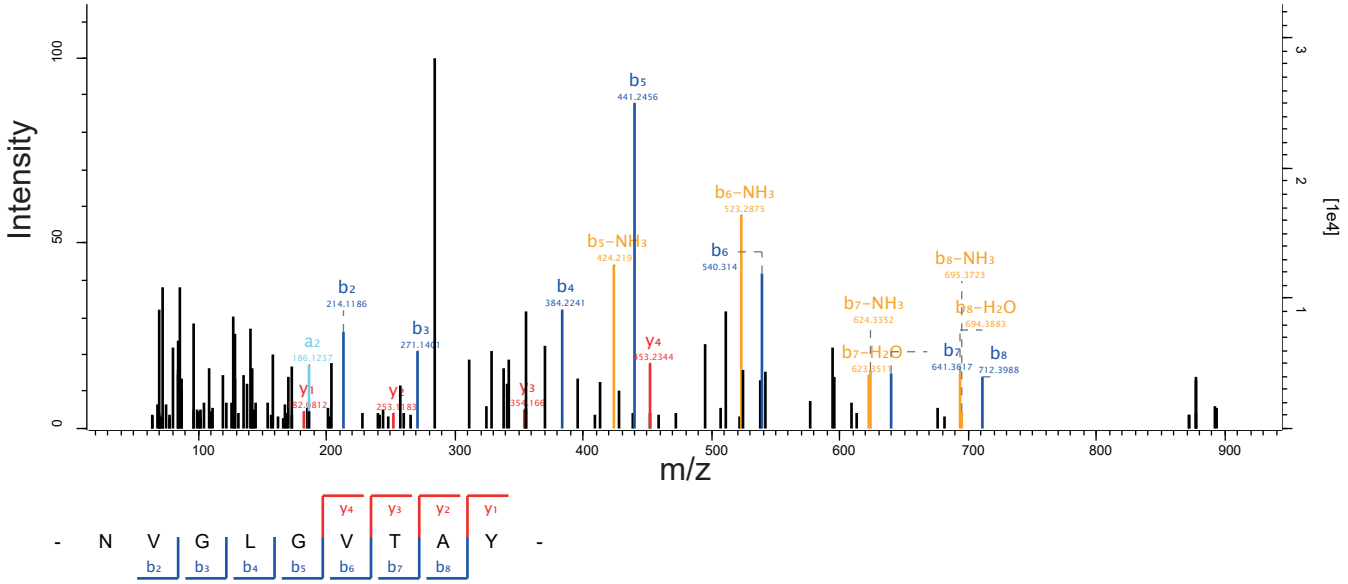
86B2 Acinetobacter cumulans HLA-I LVTGIVVLL

Endogenous peptide



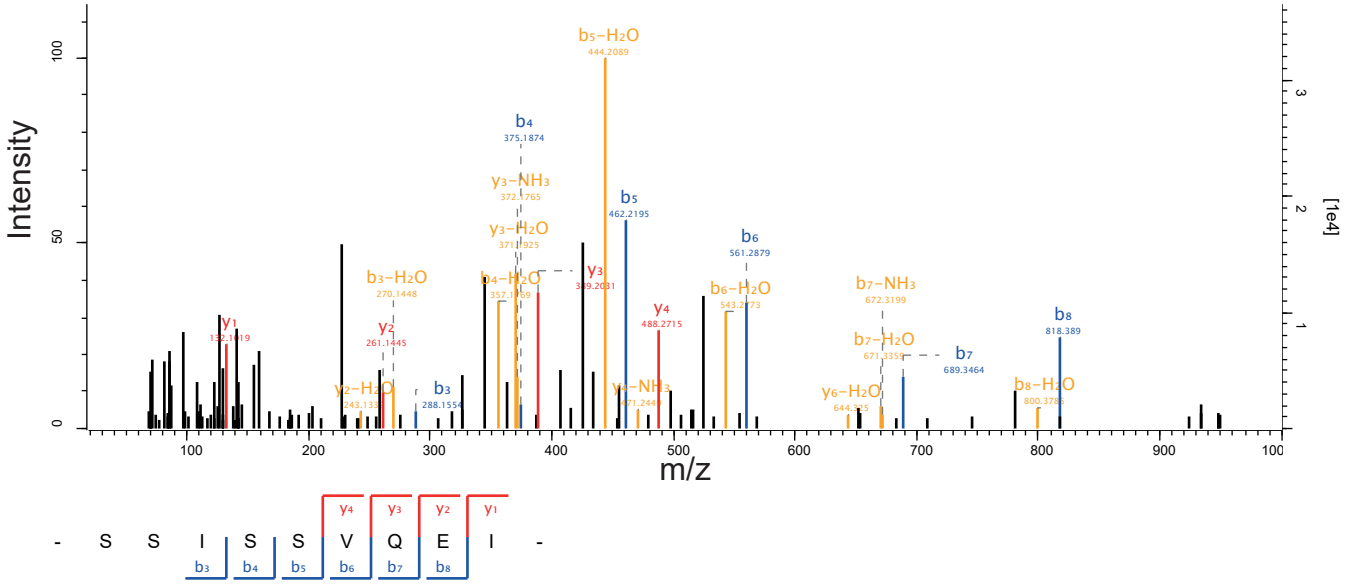
86B2 Acinetobacter cumulans HLA-I NVGLGVTAY

Endogenous peptide



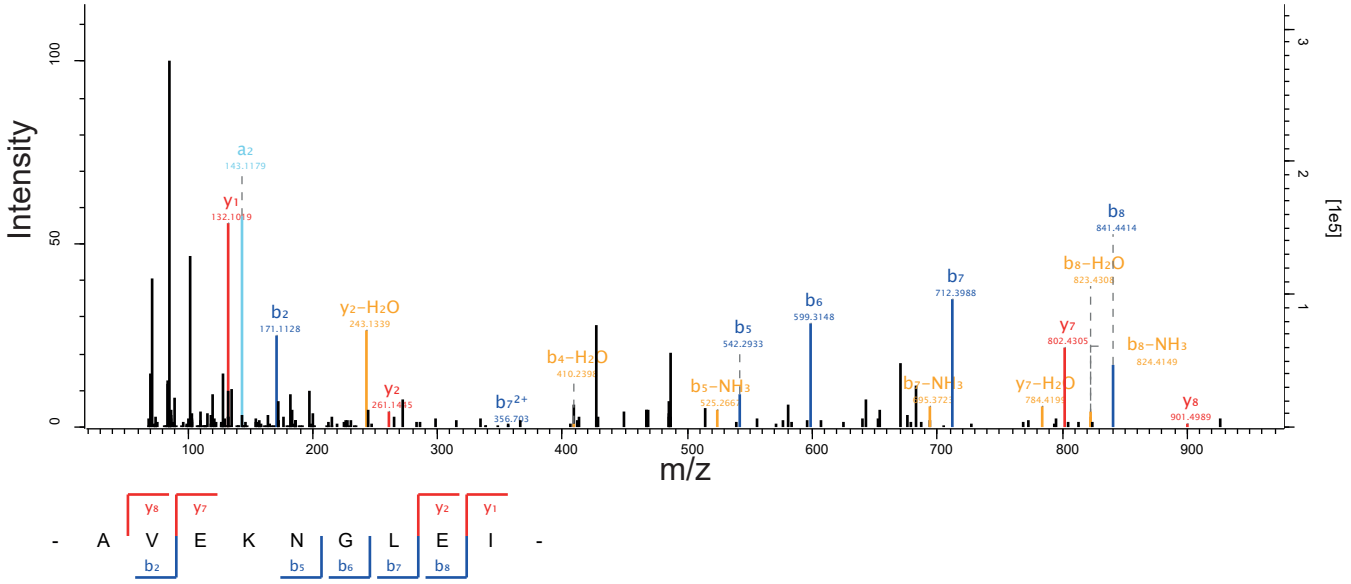
86B2 Acinetobacter cumulans HLA-I SSISSVQEI

Endogenous peptide



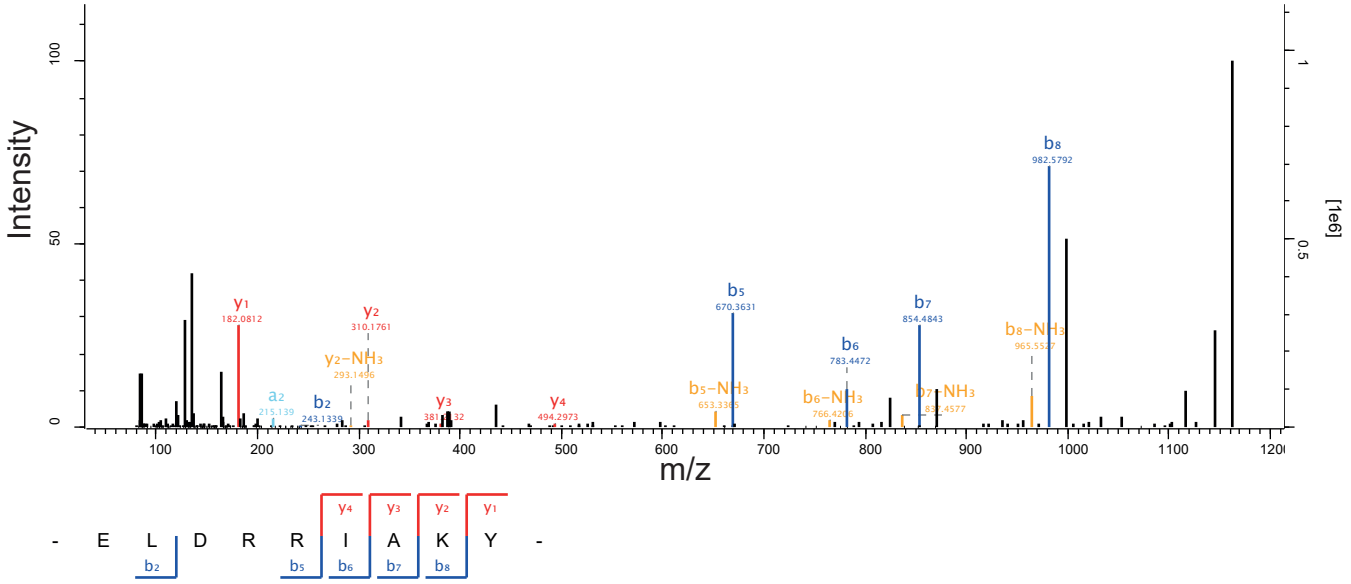
86B2 Bacteroides vulgatus HLA-I AVEKNGLEI

Endogenous peptide



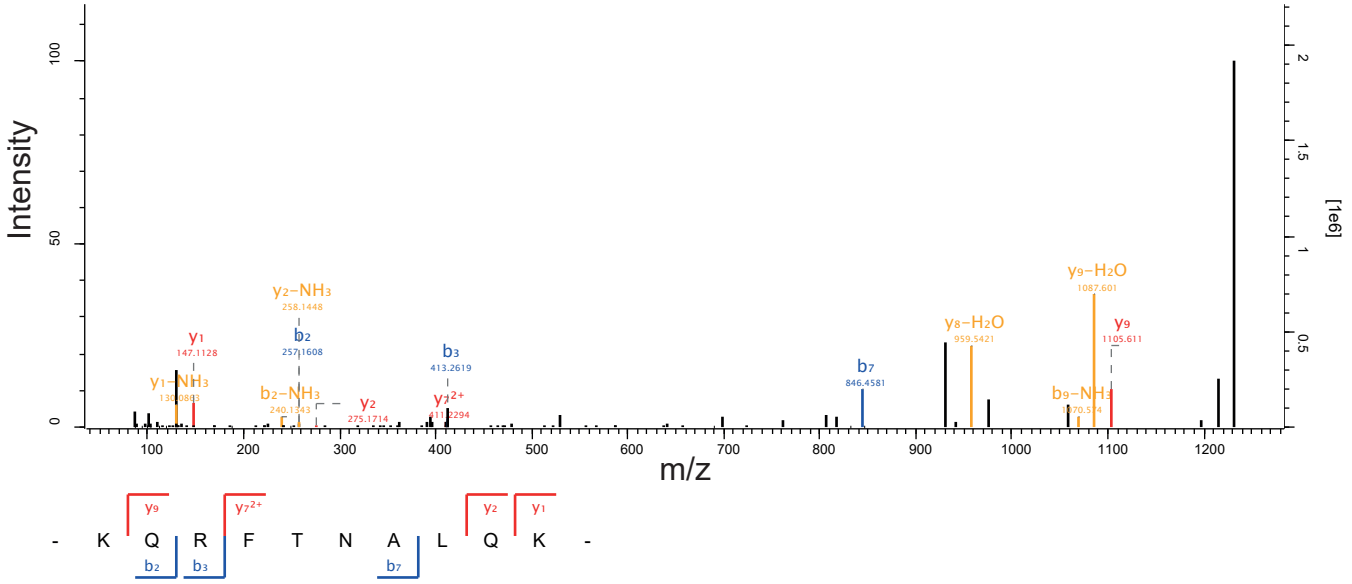
86B2 Bacteroides vulgatus HLA-I ELDRRIAKY

Endogenous peptide



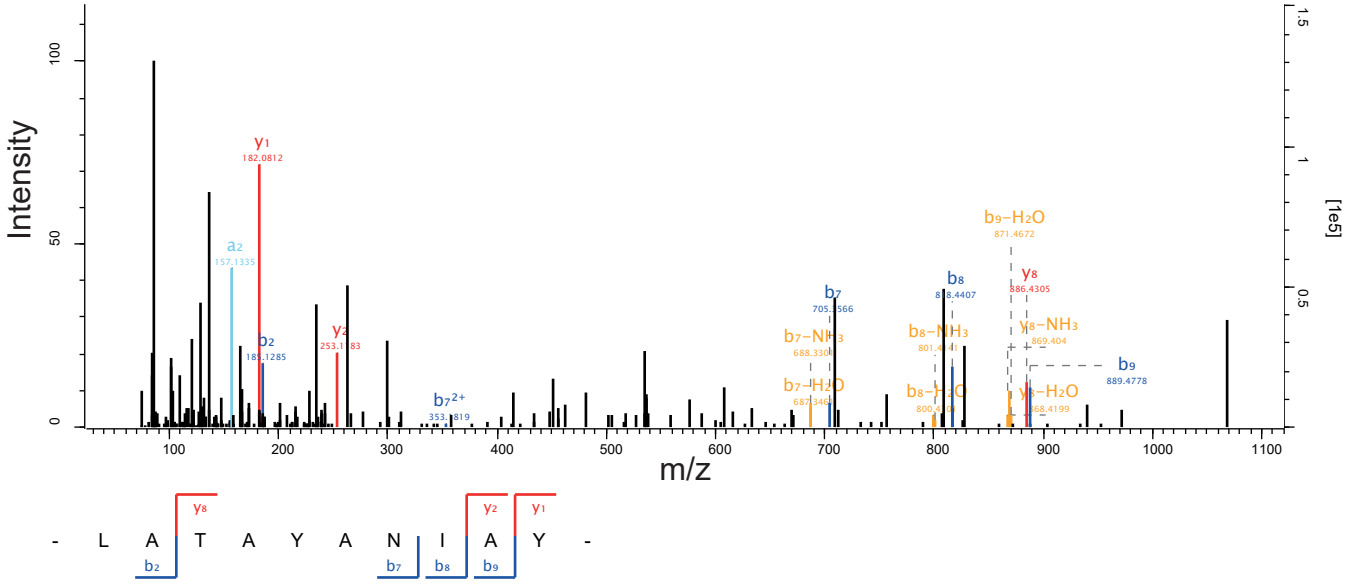
86B2 Bacteroides vulgatus HLA-I KQRFTNALQK

Endogenous peptide



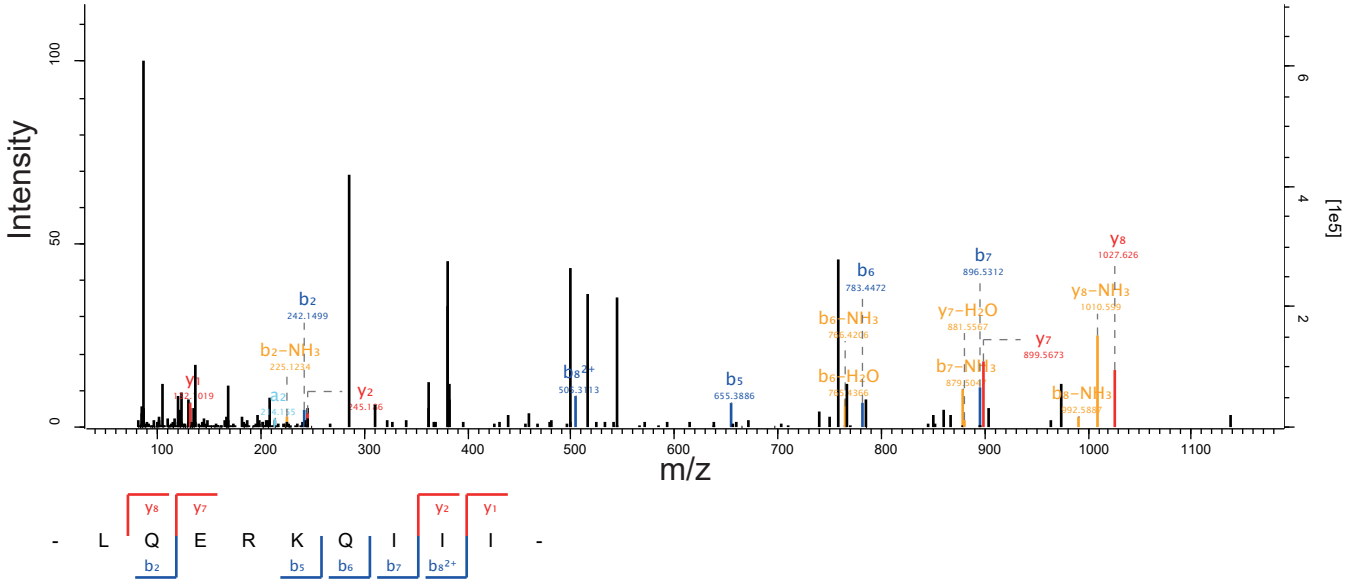
86B2 Bacteroides vulgatus HLA-I LATAYANIAY

Endogenous peptide



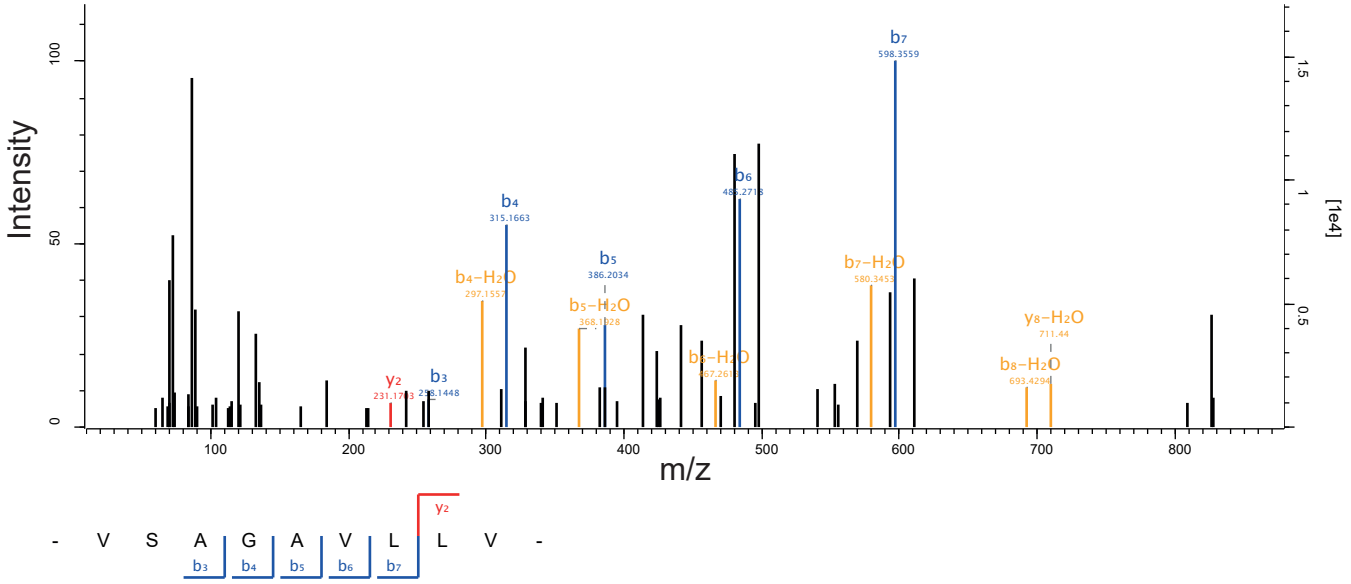
86B2 Bacteroides vulgatus HLA-I LQERKQIII

Endogenous peptide



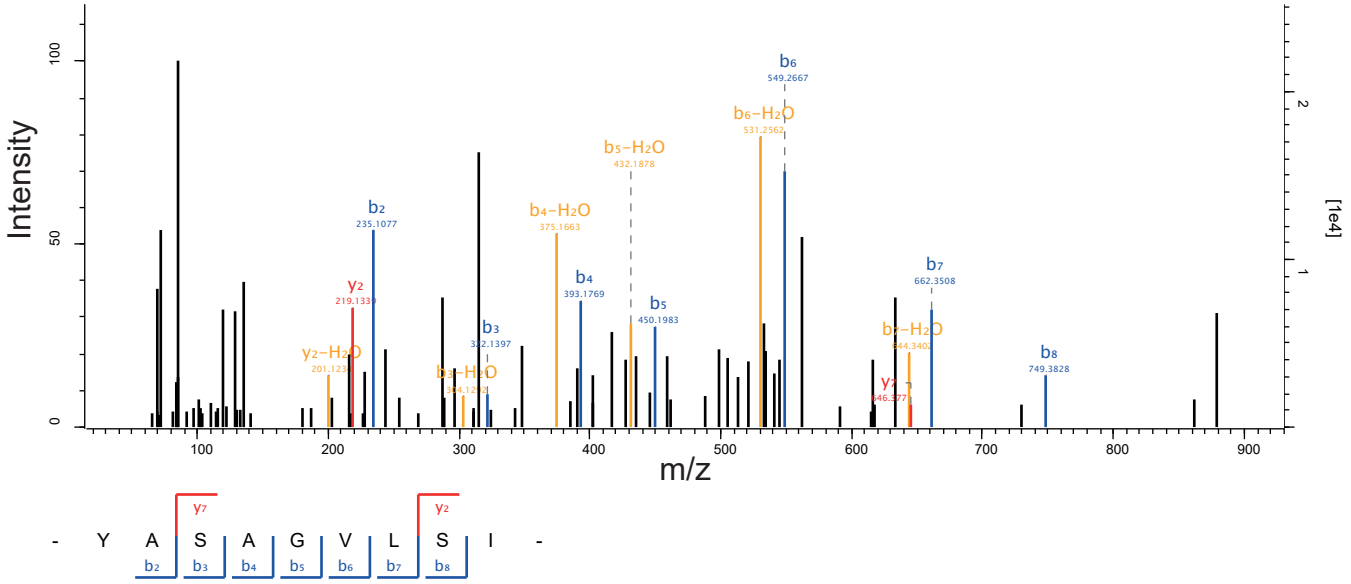
86B2 Bacteroides vulgatus HLA-I VSAGAVLLV

Endogenous peptide



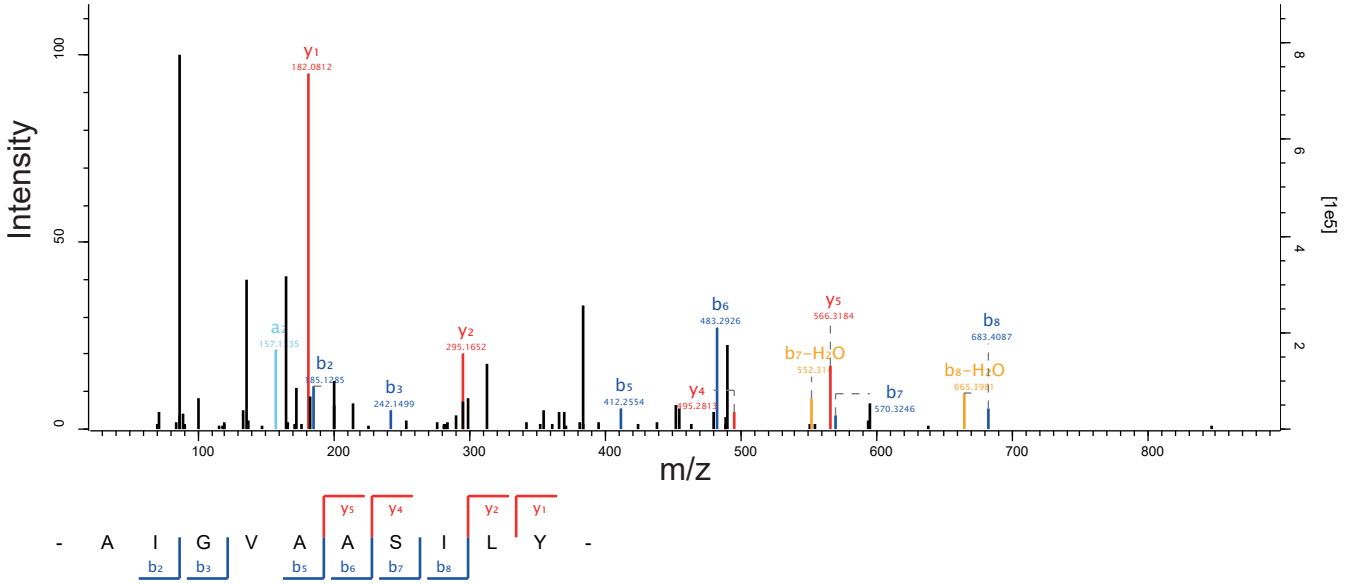
86B2 Bacteroides vulgatus HLA-I YASAGVLSI

Endogenous peptide



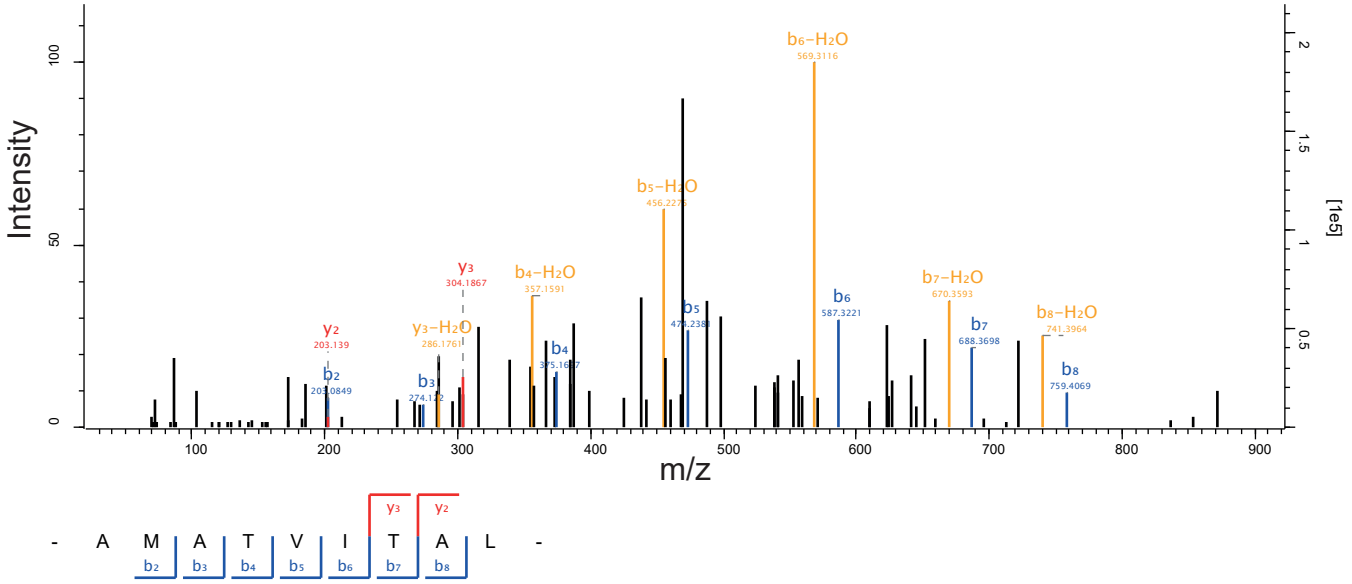
86B2 Staphylococcus capitis HLA-I AIGVAASILY

Endogenous peptide



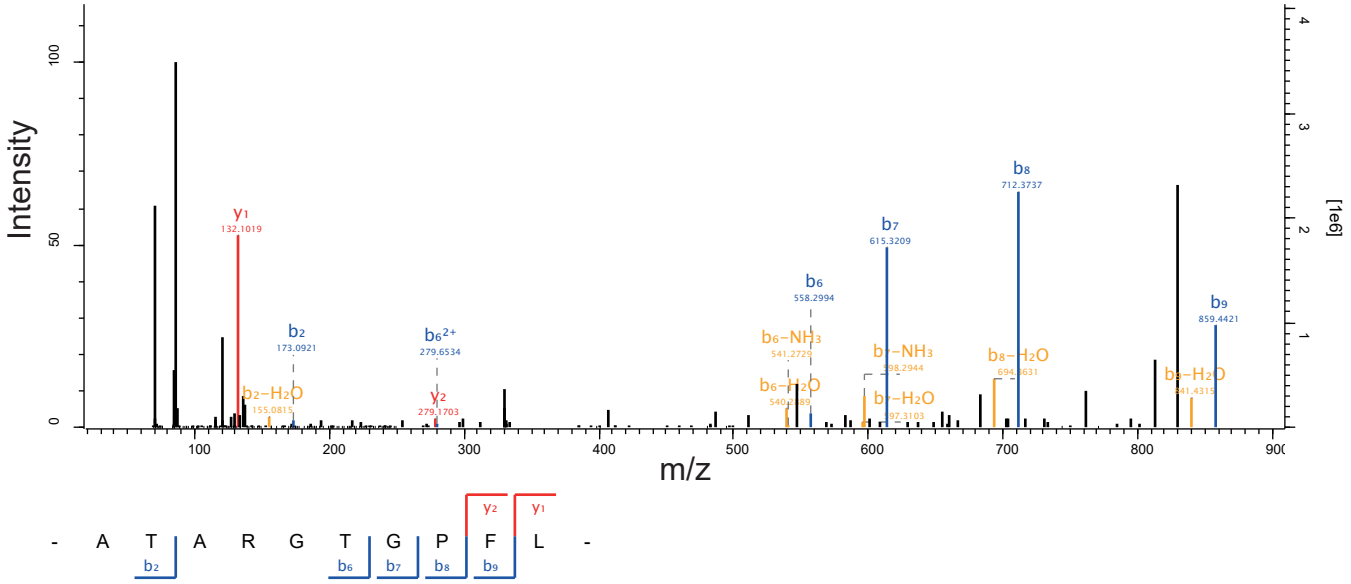
86B2 Staphylococcus capitis HLA-I AMATVITAL

Endogenous peptide



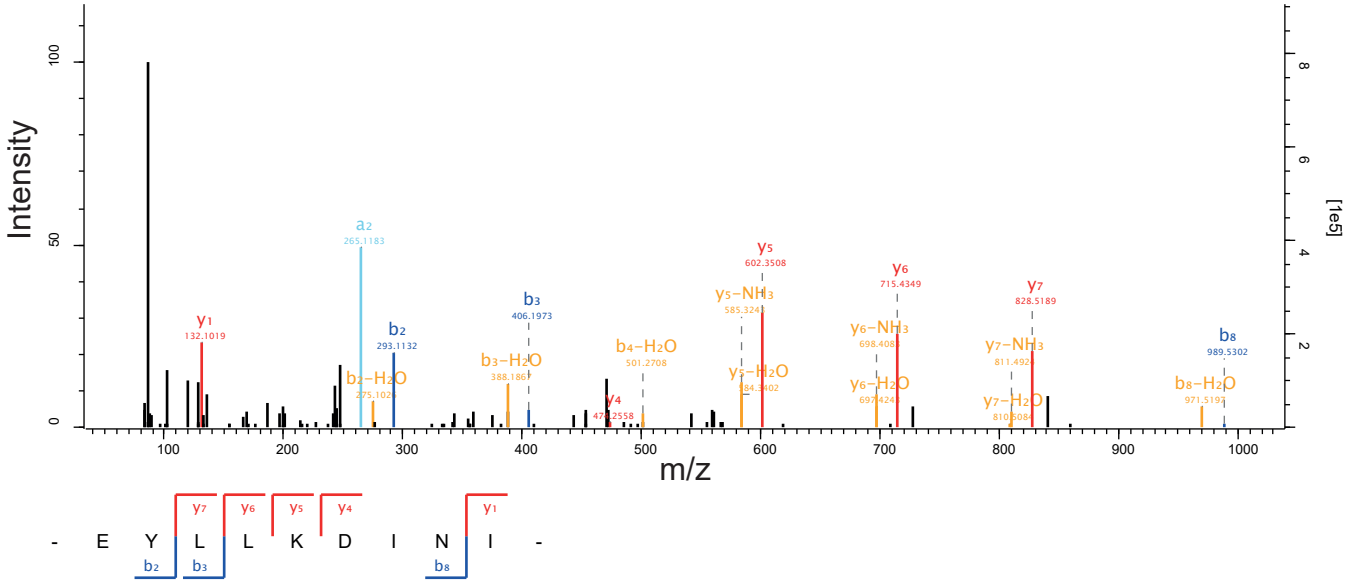
86B2 Staphylococcus capitis HLA-I ATARGTGPFL

Endogenous peptide



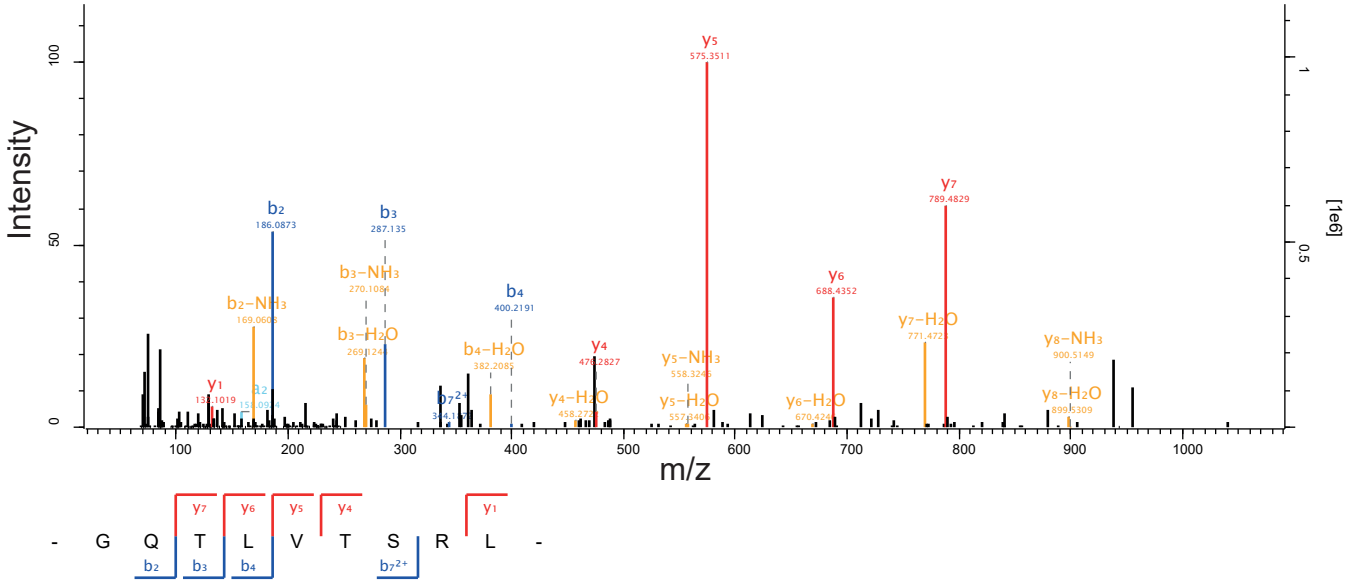
86B2 Staphylococcus capitis HLA-I EYLLKDINI

Endogenous peptide



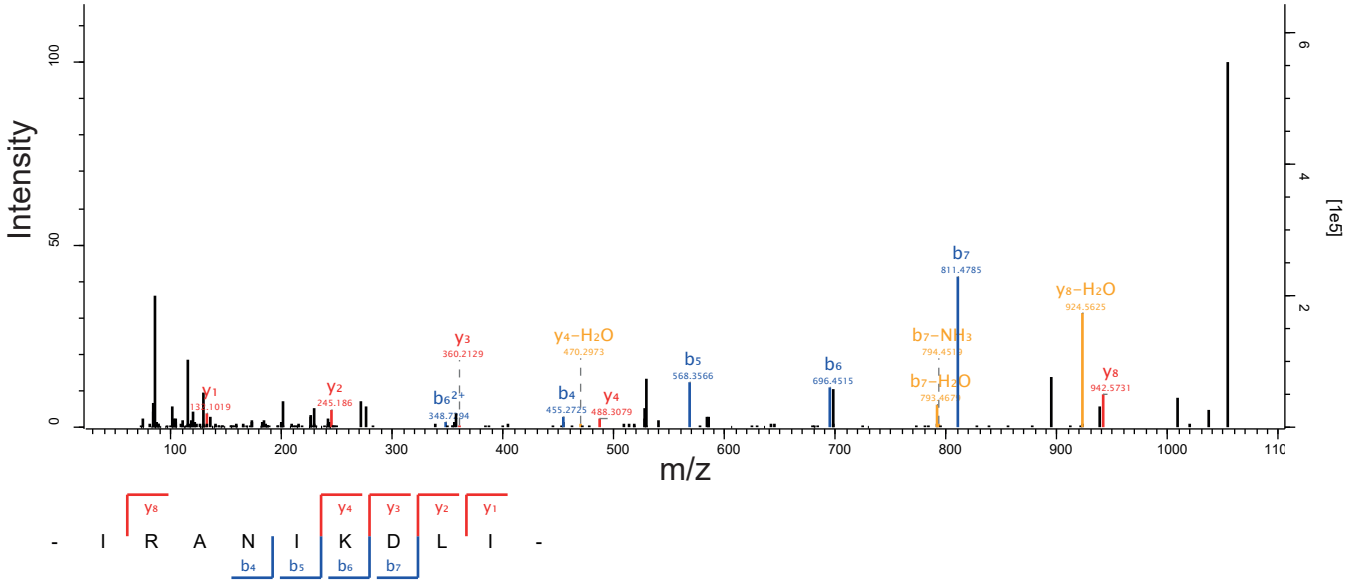
86B2 Staphylococcus capitis HLA-I GQTLVTSRL

Endogenous peptide



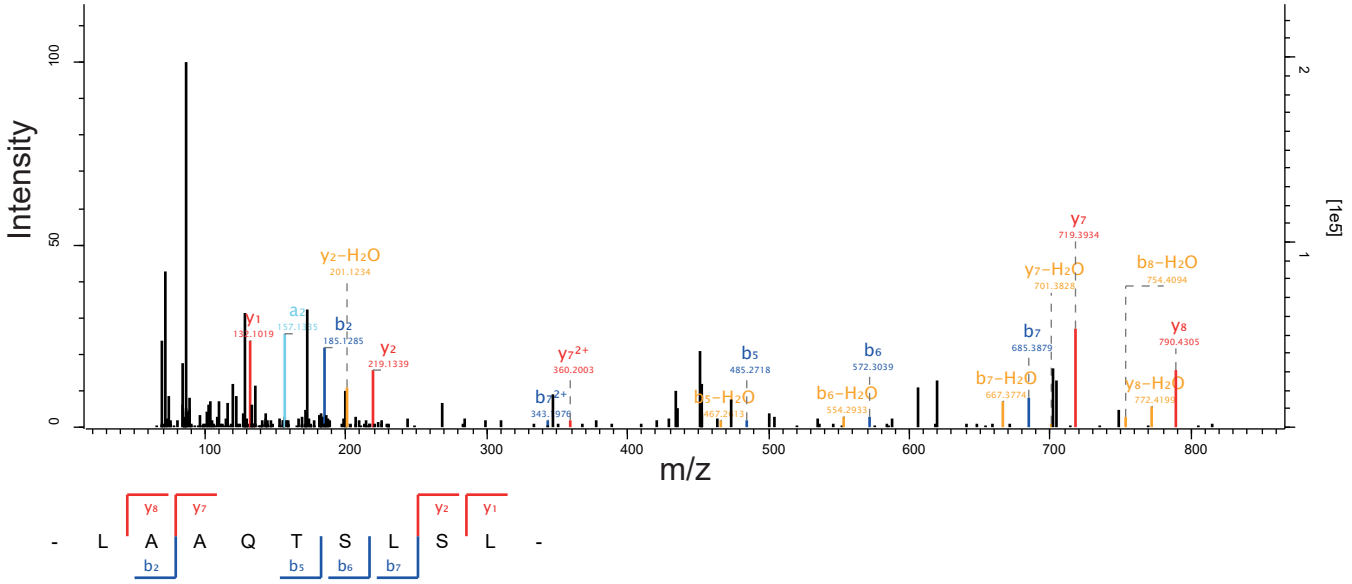
86B2 Staphylococcus capitis HLA-I IRANIKDLI

Endogenous peptide



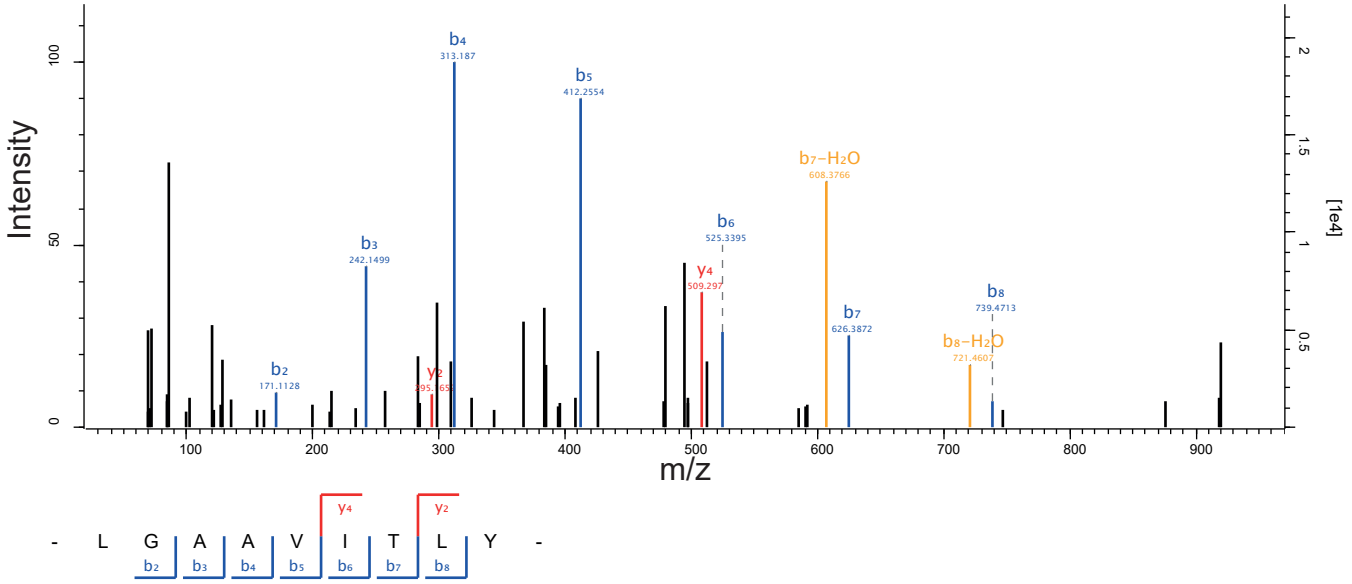
86B2 Staphylococcus capitis HLA-I LAAQTSLSL

Endogenous peptide



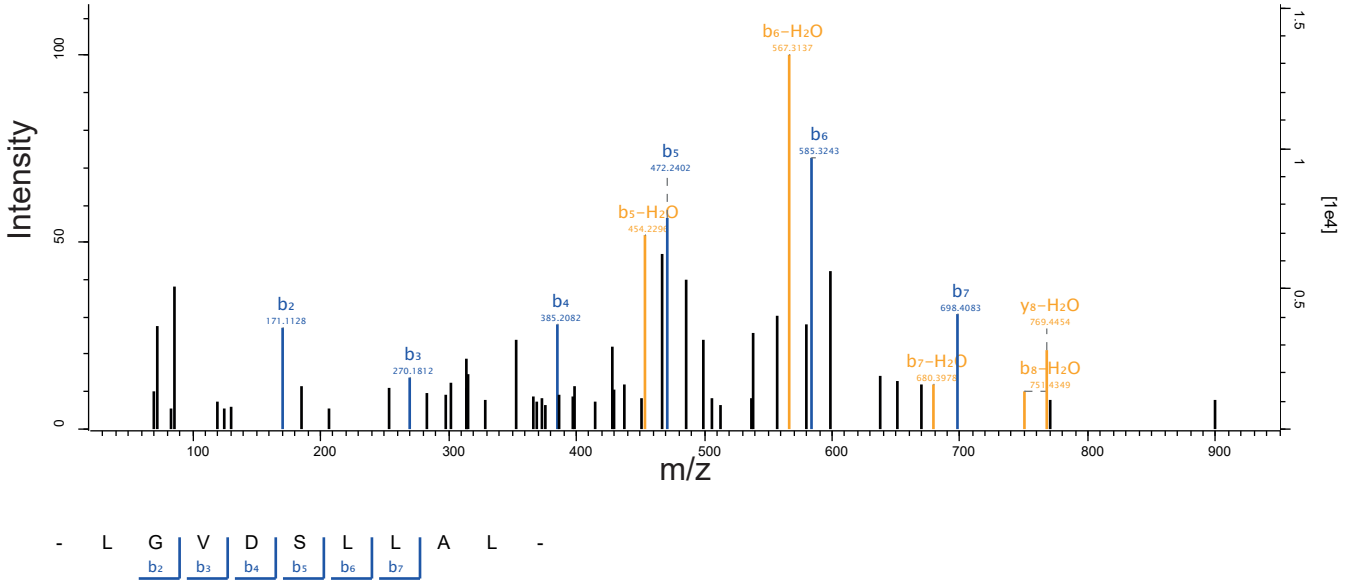
86B2 Staphylococcus capitis HLA-I LGAAVITLY

Endogenous peptide



86B2 Staphylococcus capitis HLA-I LGVDSLLAL

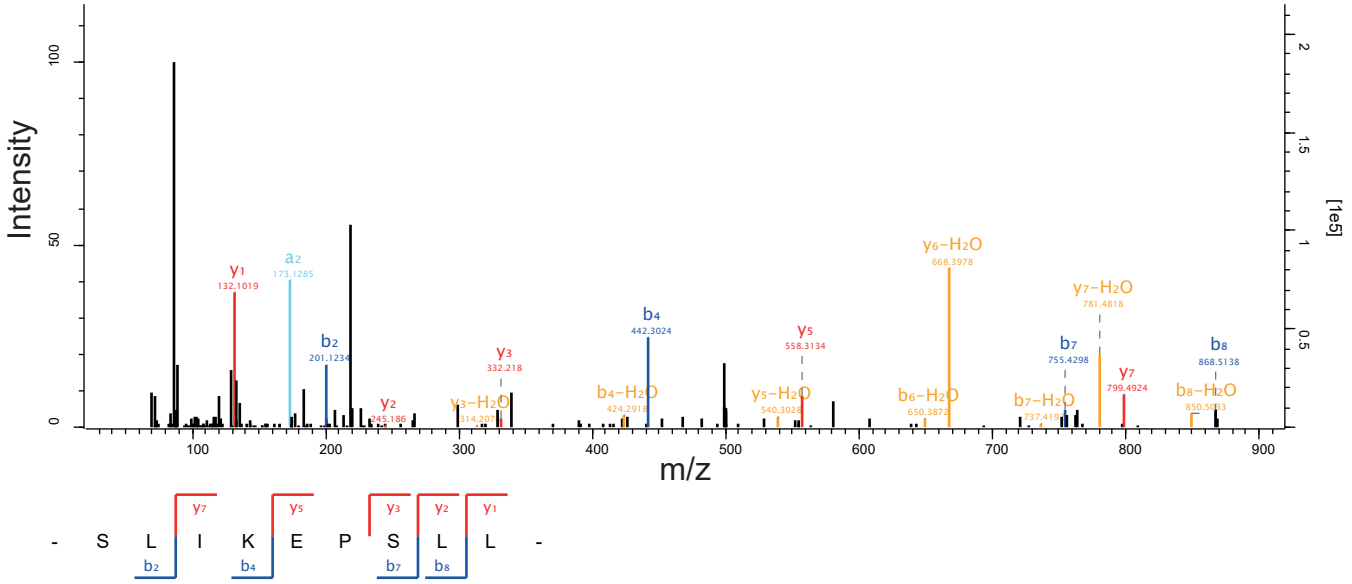
Endogenous peptide



- L G V D S L L A L -
b2 b3 b4 b5 b6 b7

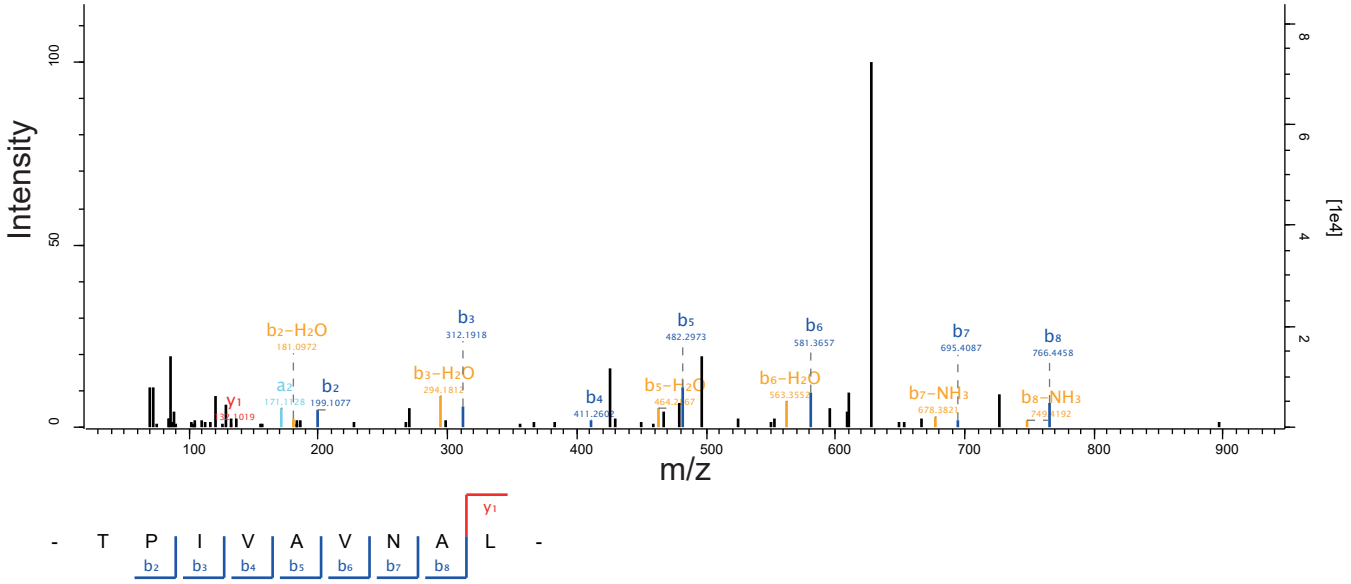
86B2 Staphylococcus capitis HLA-I SLIKEPSLL

Endogenous peptide



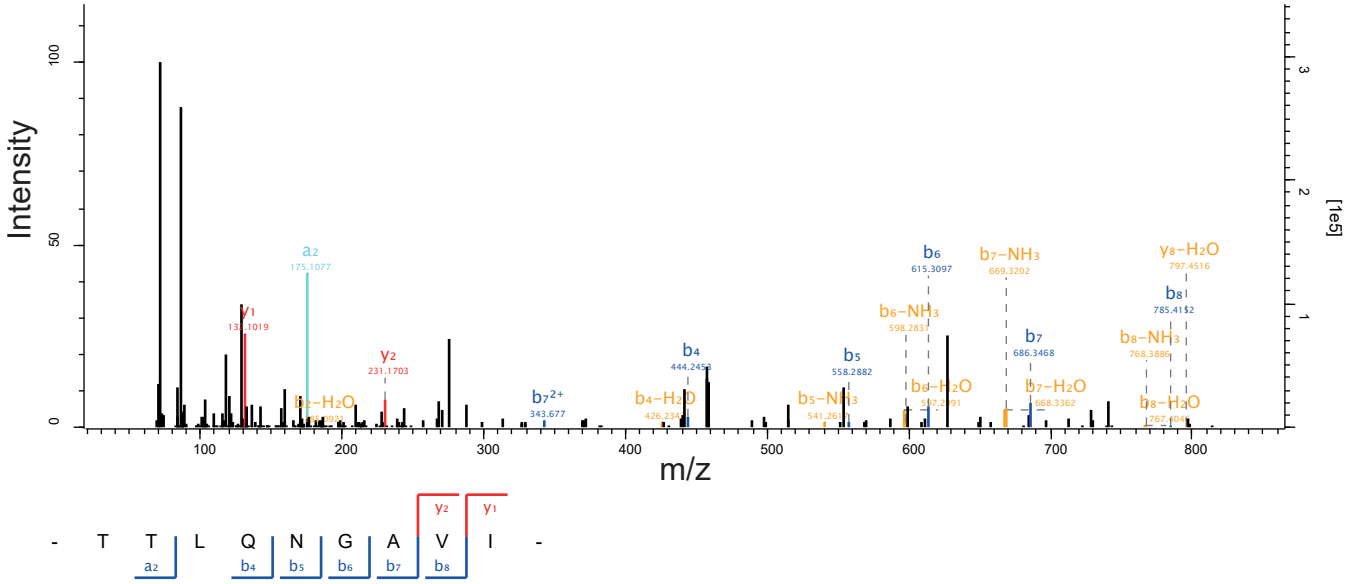
86B2 Staphylococcus capitis HLA-I TPIVAVNAL

Endogenous peptide



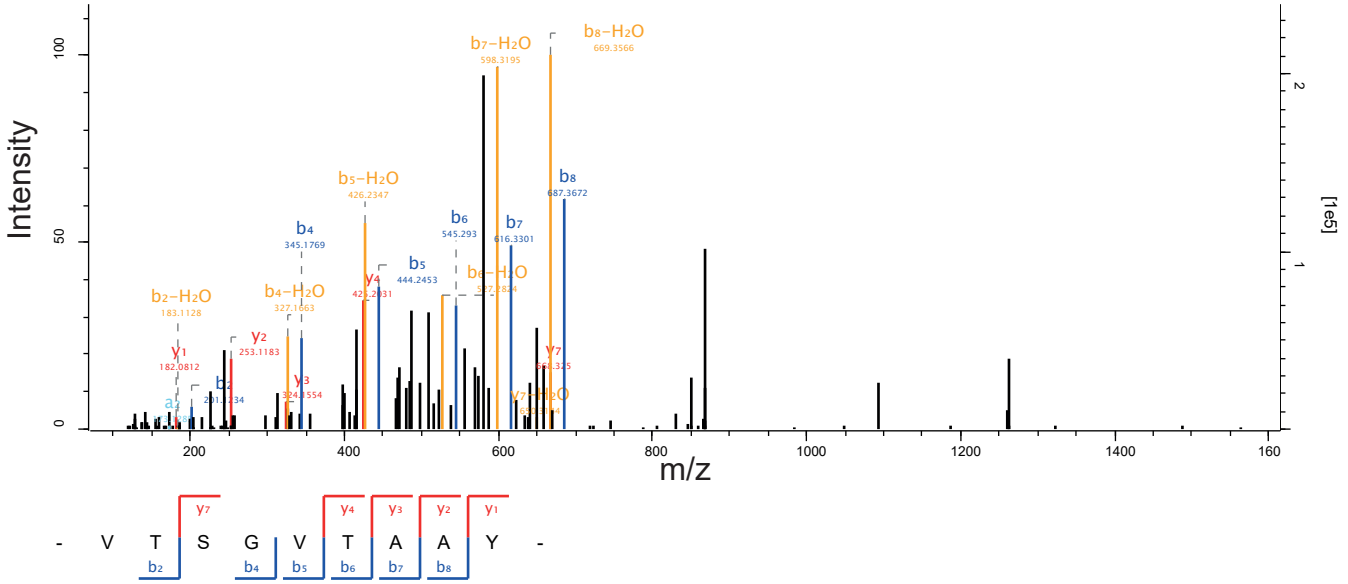
86B2 Staphylococcus capitis HLA-I TTLQNGAVI

Endogenous peptide



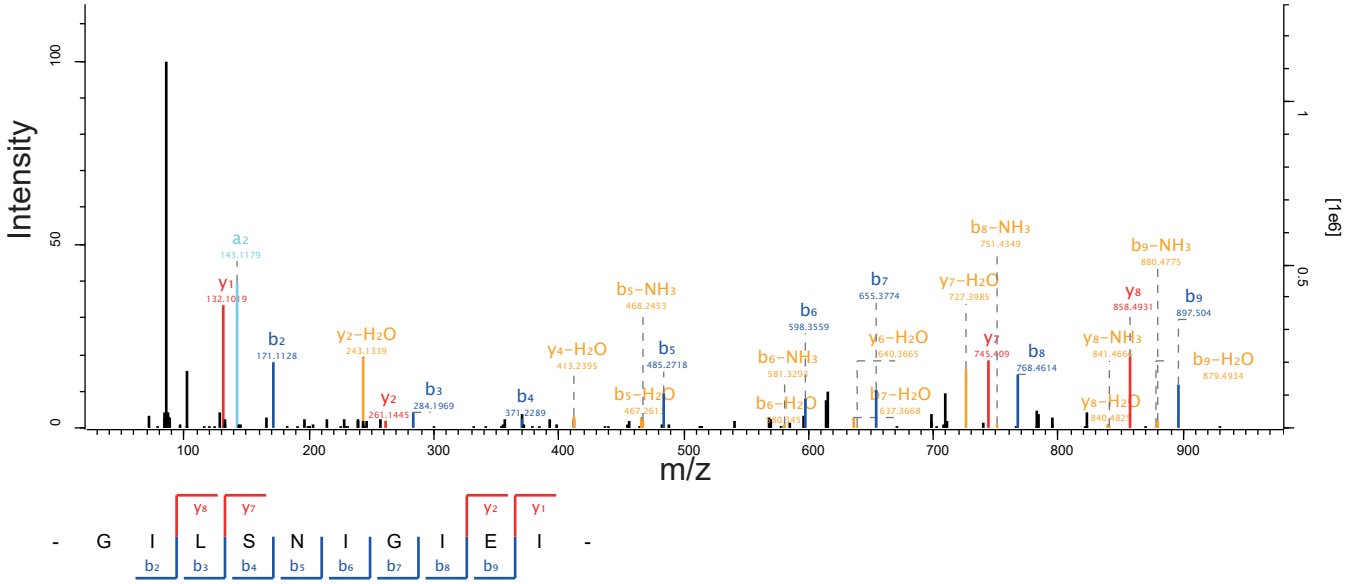
86B2 Staphylococcus capitis HLA-I VTSGVTAAY

Endogenous peptide



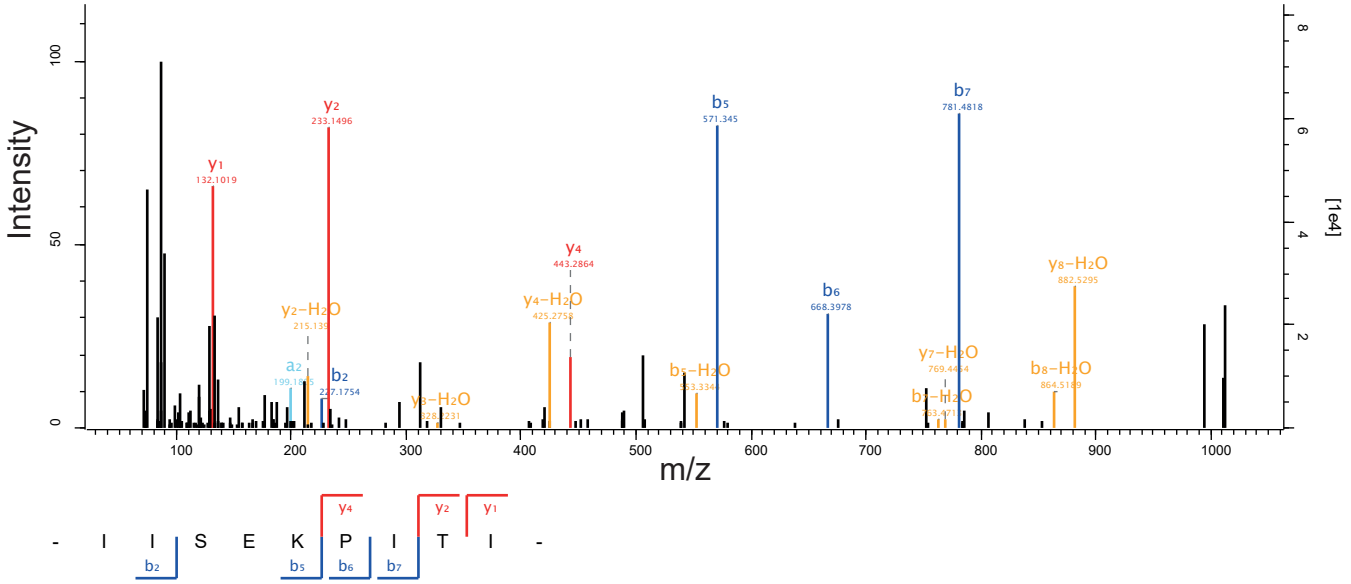
92B3 Fusobacterium nucleatum HLA-I GILSNIGIEI

Endogenous peptide



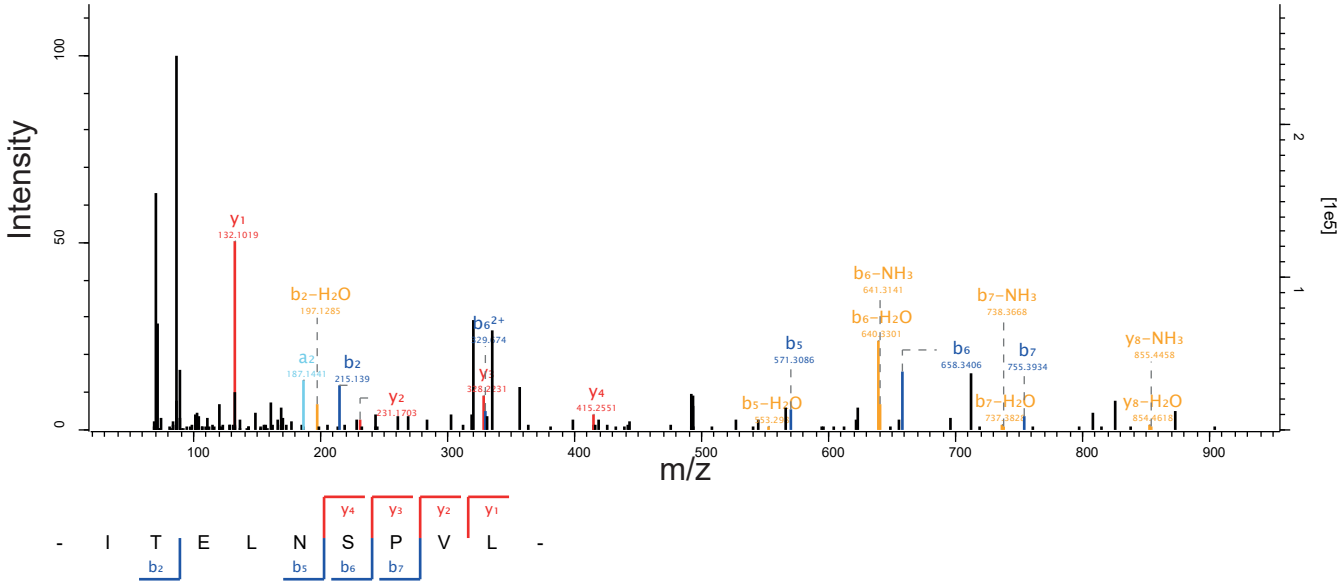
92B3 Fusobacterium nucleatum HLA-I IISEKPITI

Endogenous peptide



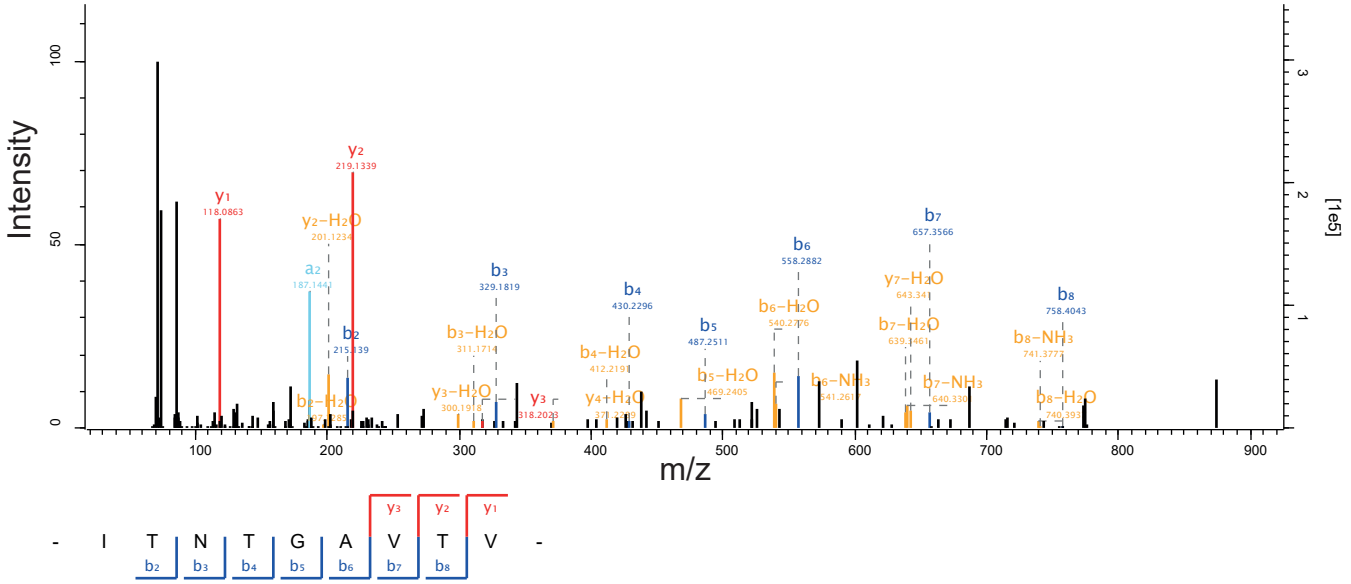
92B3 Fusobacterium nucleatum HLA-I ITELNSPVL

Endogenous peptide



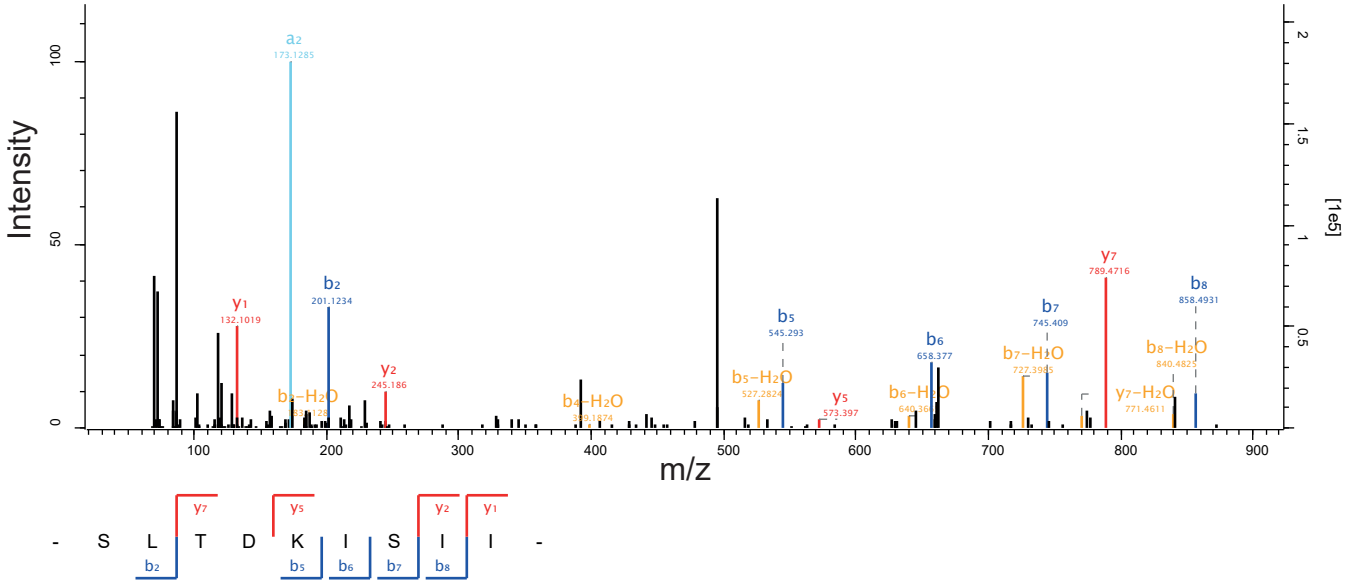
92B3 Fusobacterium nucleatum HLA-I ITNTGAVTV

Endogenous peptide



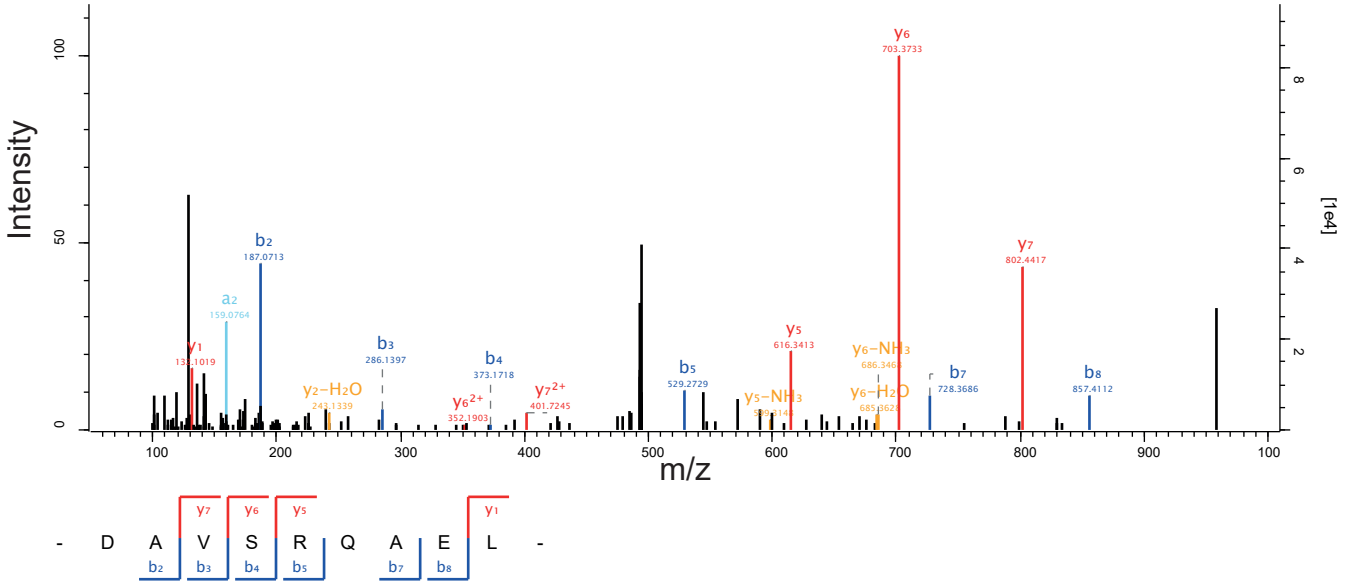
92B3 Fusobacterium nucleatum HLA-I SLTDKISII

Endogenous peptide



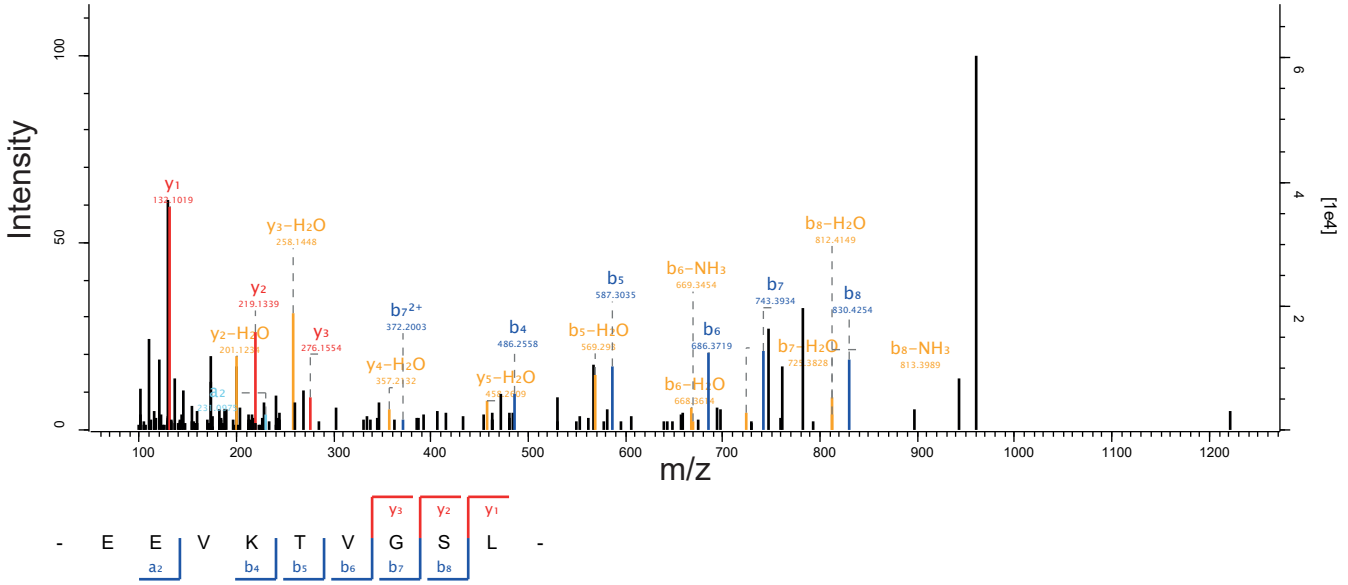
112 Staphylococcus aureus HLA-I DAVSRQAEL

Endogenous peptide



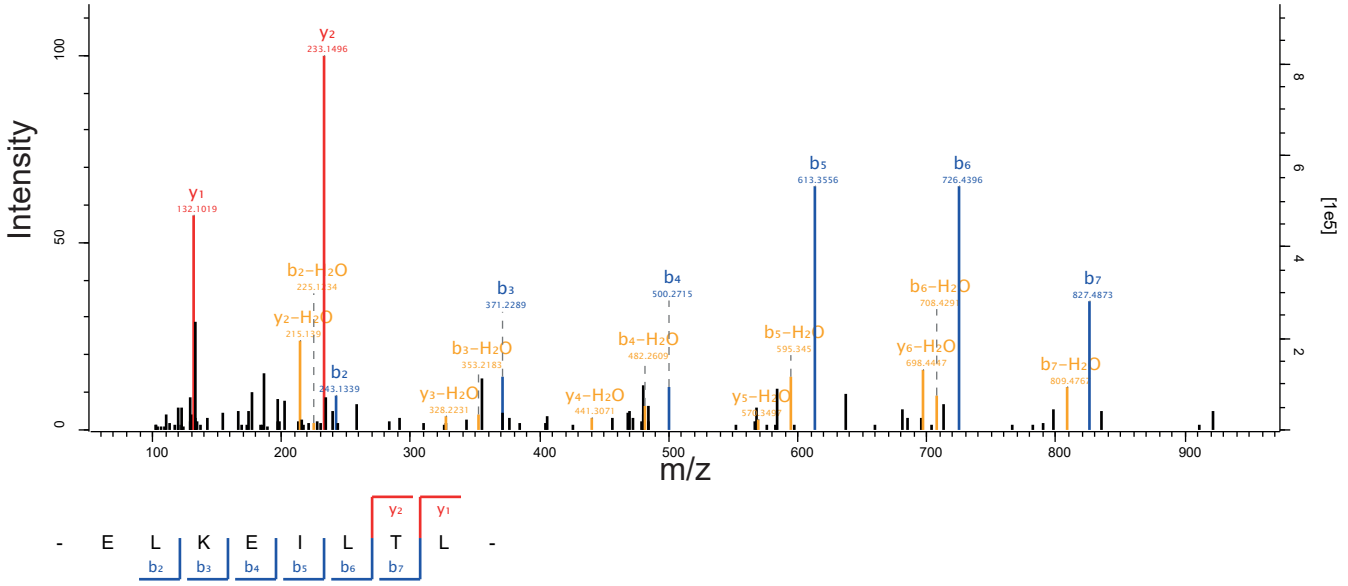
112 Staphylococcus aureus HLA-I EEVKTVGSL

Endogenous peptide



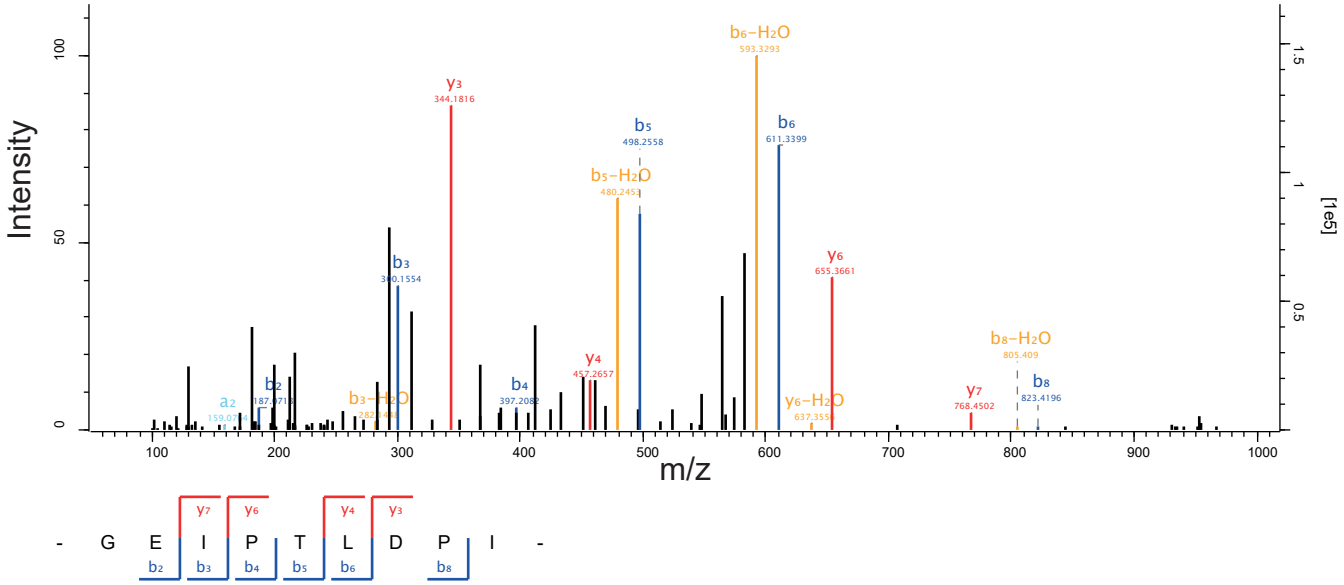
112 Staphylococcus aureus HLA-I ELKEILTL

Endogenous peptide



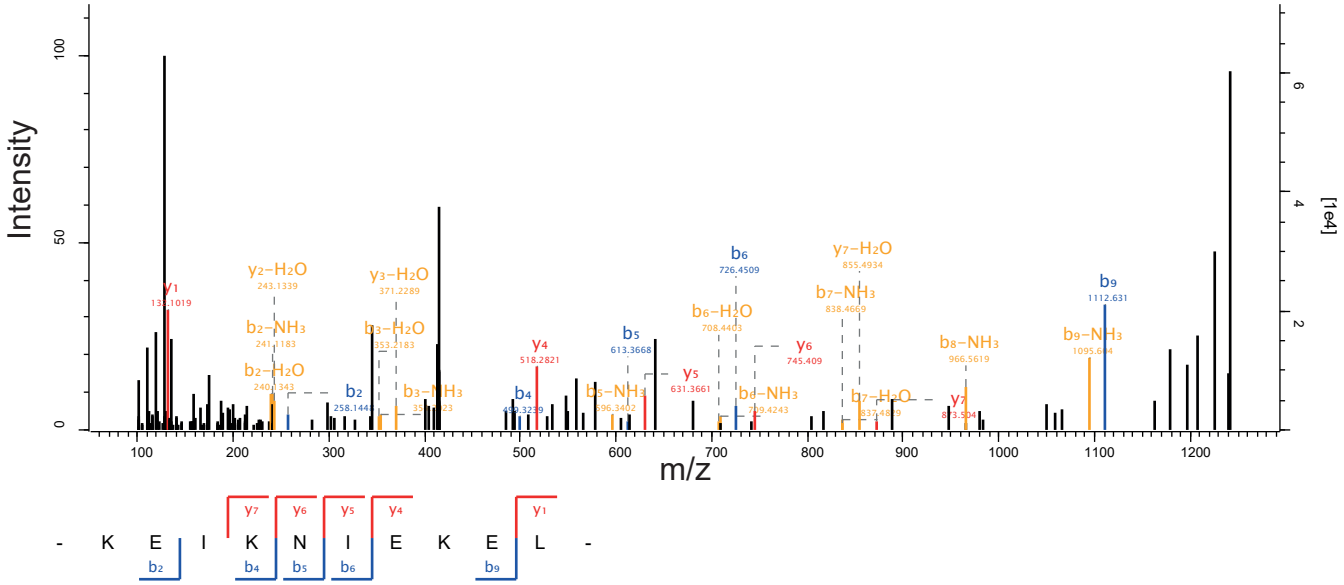
112 Staphylococcus aureus HLA-I GEIPTLDPI

Endogenous peptide



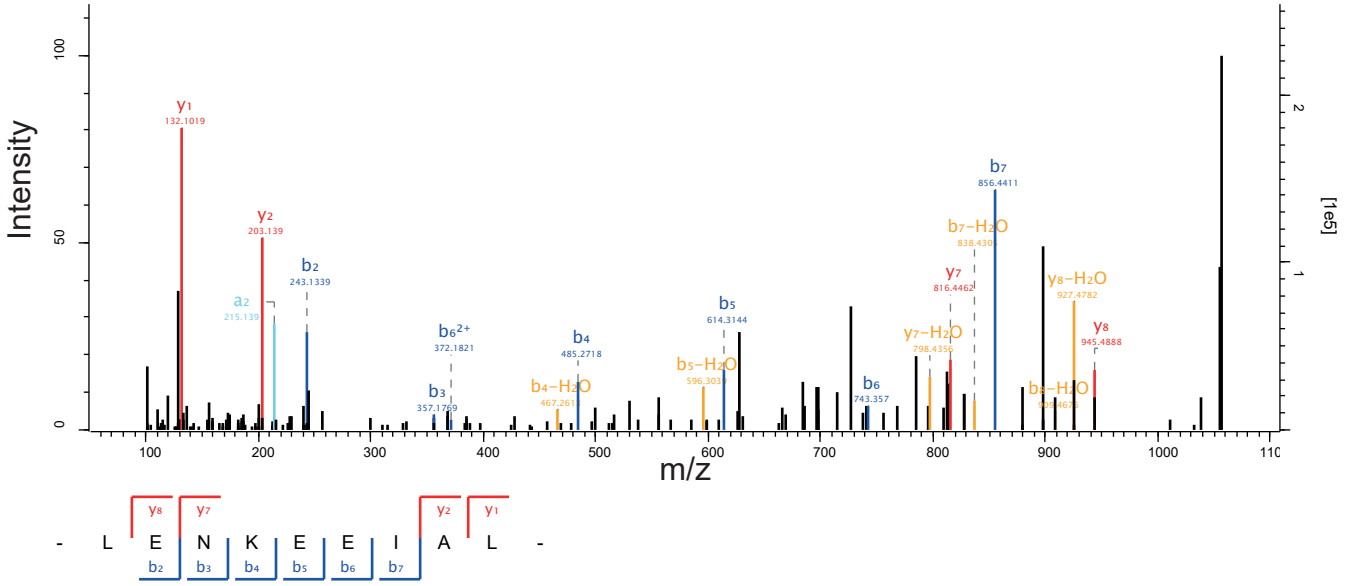
112 Staphylococcus aureus HLA-I KEIKNIEKEL

Endogenous peptide



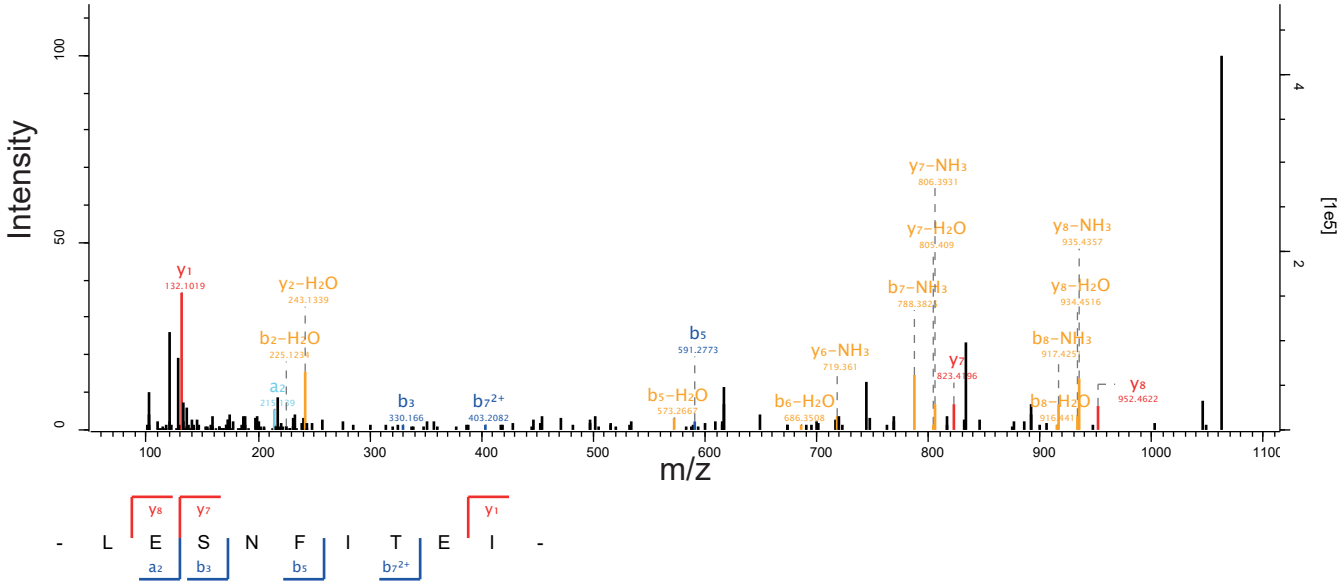
112 Staphylococcus aureus HLA-I LENKEEIAL

Endogenous peptide



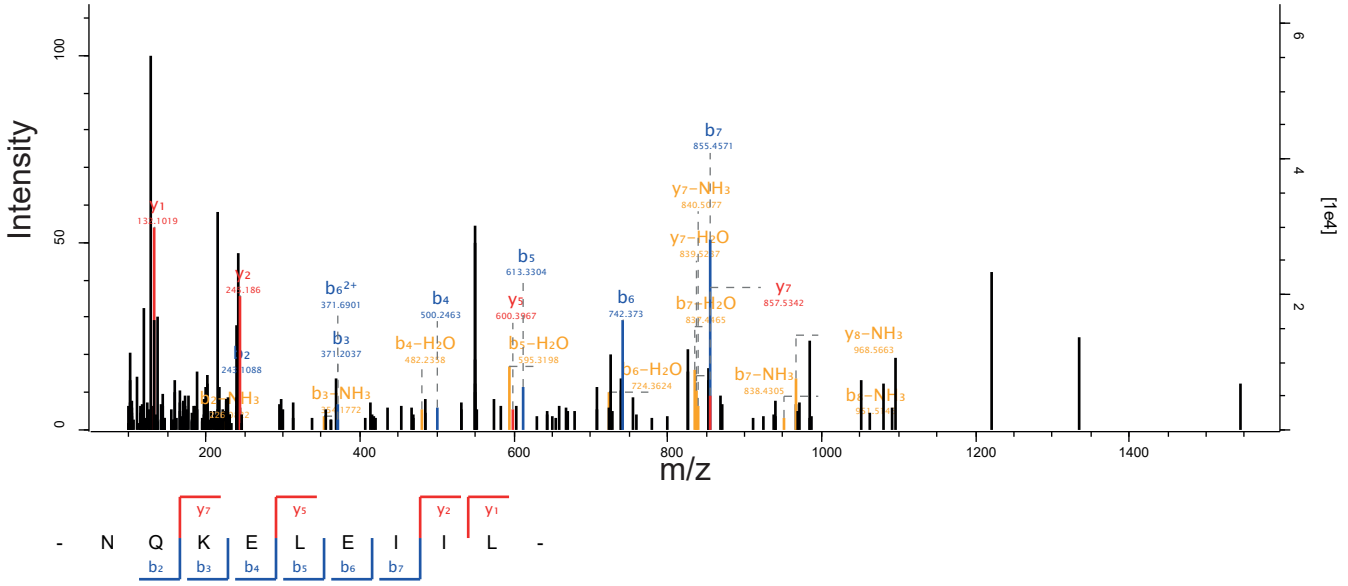
112 Staphylococcus aureus HLA-I LESNFITEI

Endogenous peptide



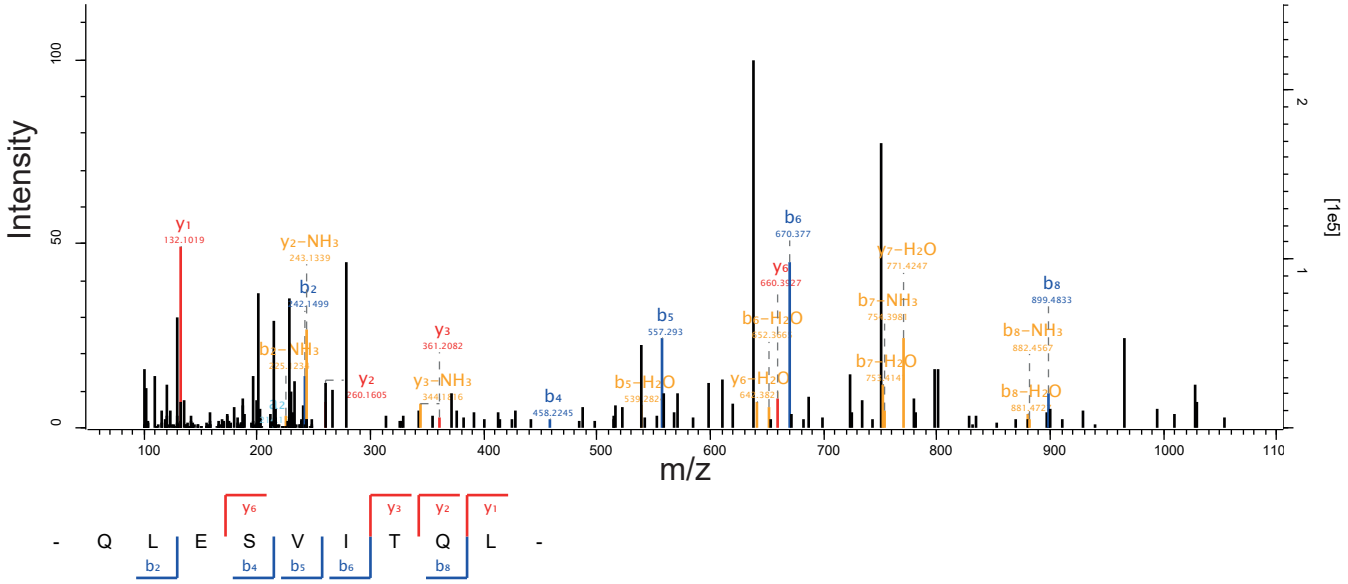
112 Staphylococcus aureus HLA-I NQKELEIIL

Endogenous peptide



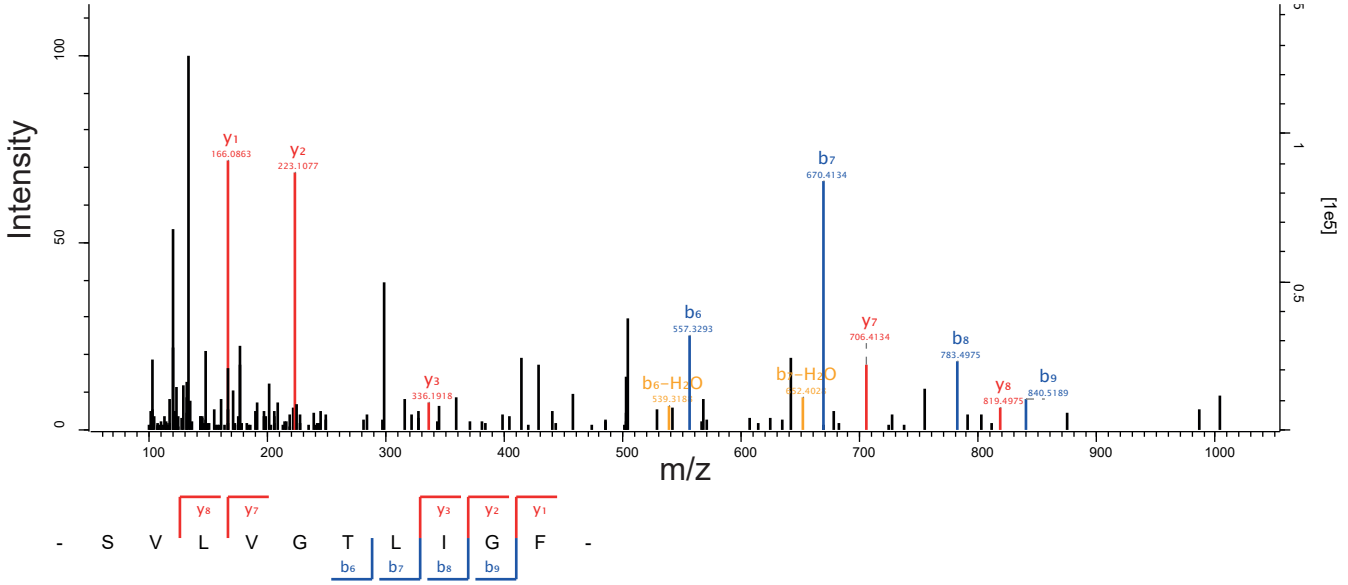
112 Staphylococcus aureus HLA-I QLESVITQL

Endogenous peptide



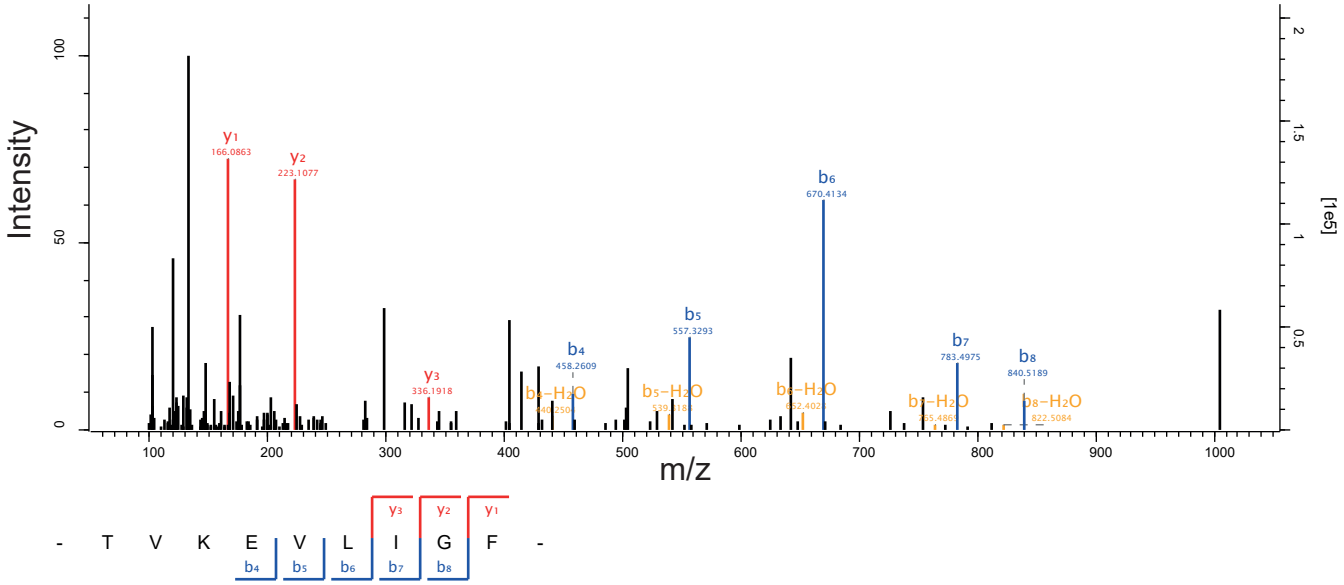
112 Staphylococcus aureus HLA-I SVLVGTLIGF

Endogenous peptide



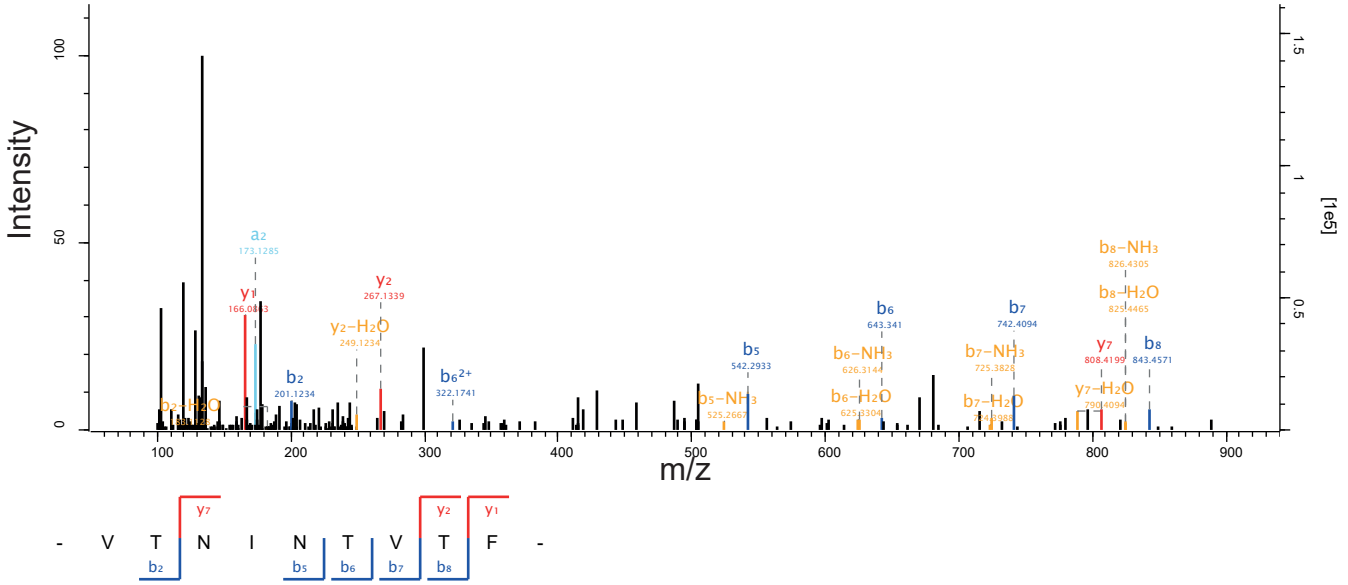
112 Staphylococcus aureus HLA-I TVKEVLIGF

Endogenous peptide



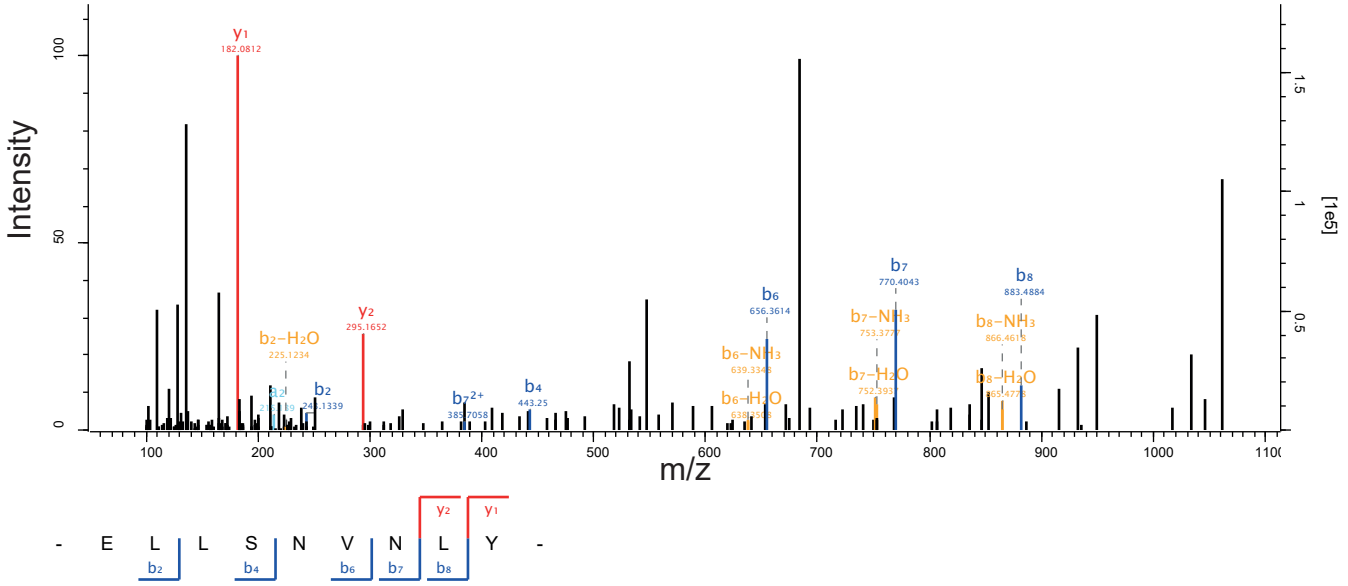
112 Staphylococcus aureus HLA-I VTNINTVTF

Endogenous peptide



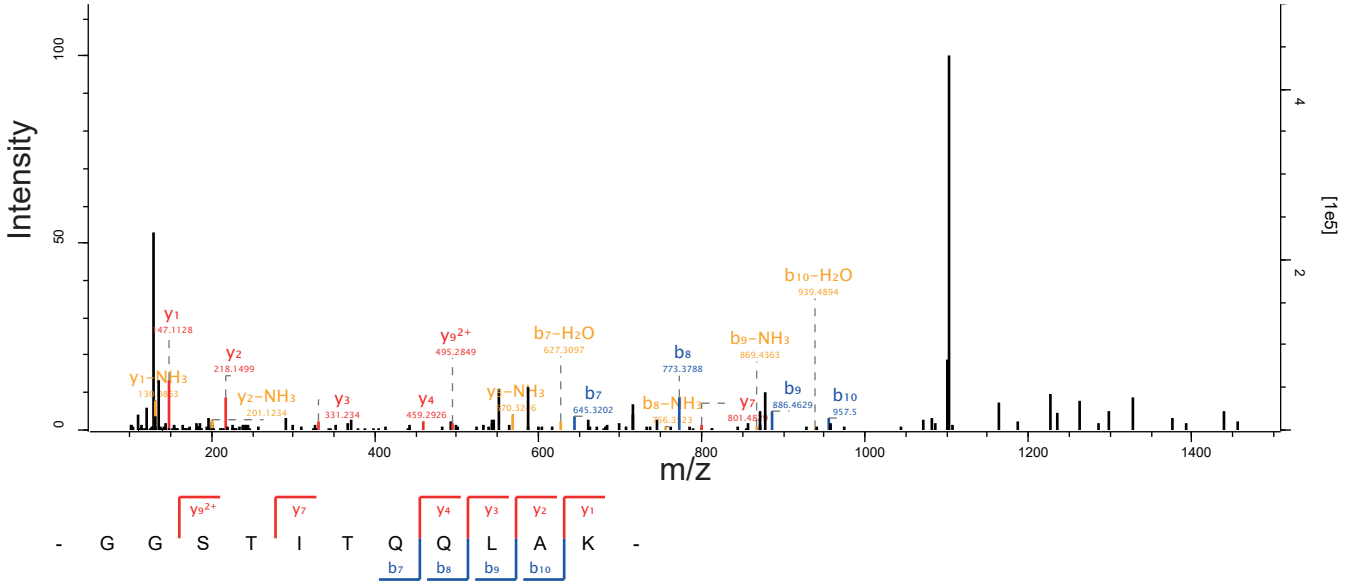
152A2 Bacteroides dorei HLA-I ELLSNVNLY

Endogenous peptide



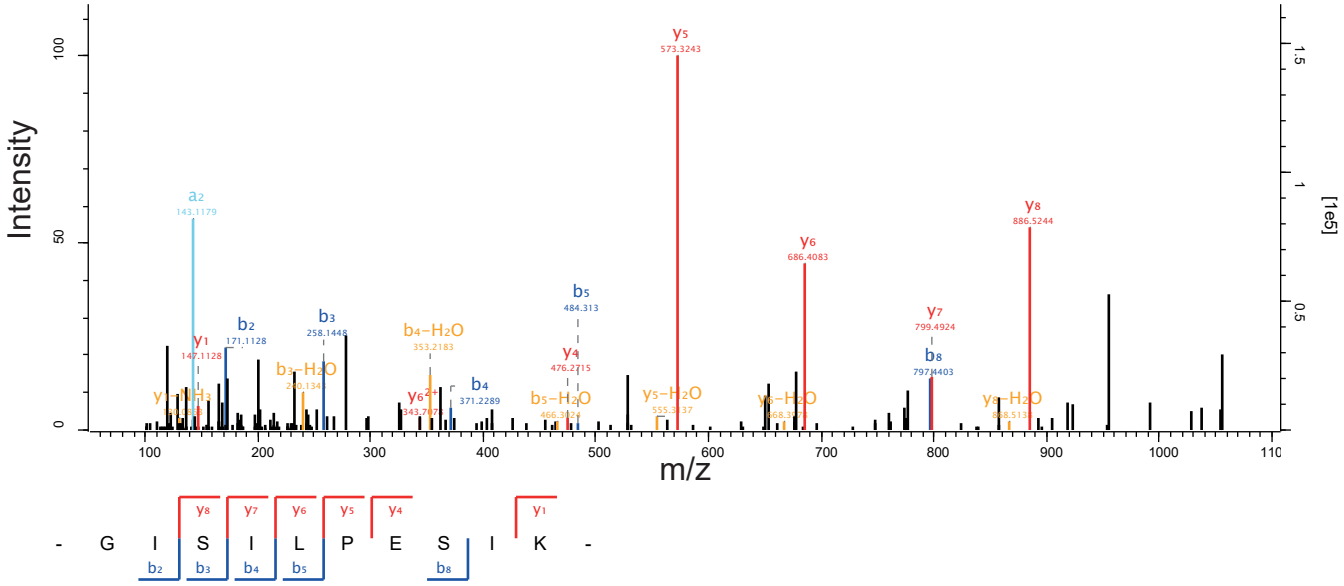
152A2 Bacteroides dorei HLA-I GGSTITQQLAK

Endogenous peptide



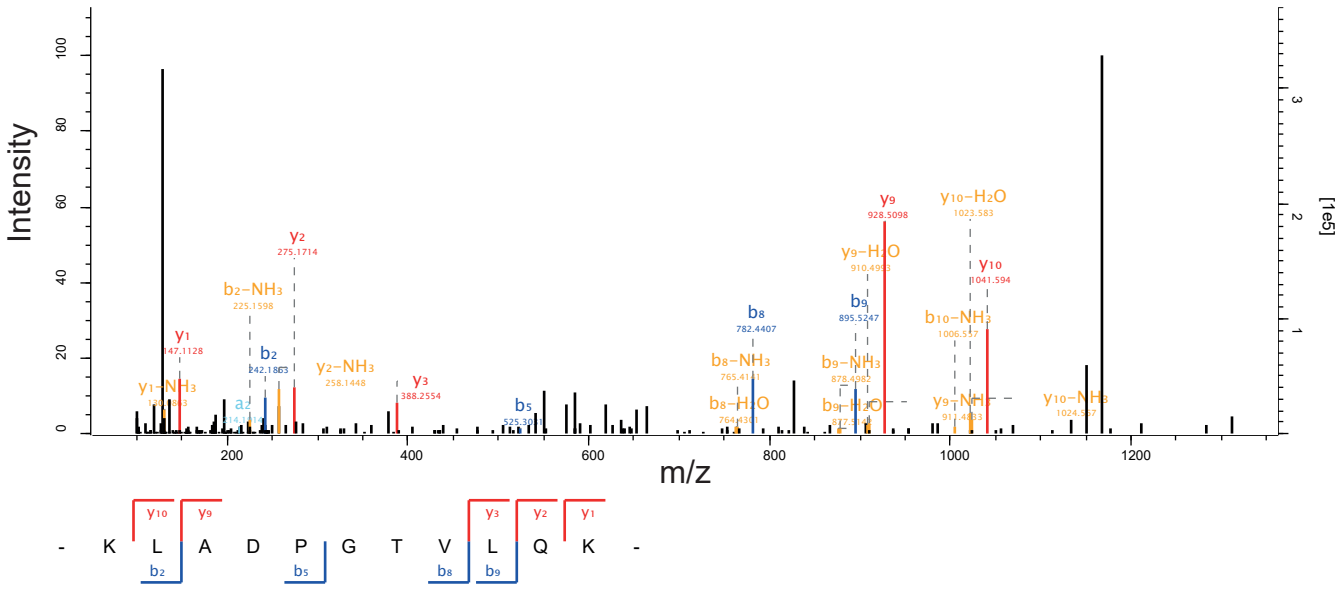
152A2 Bacteroides dorei HLA-I GISILPESIK

Endogenous peptide



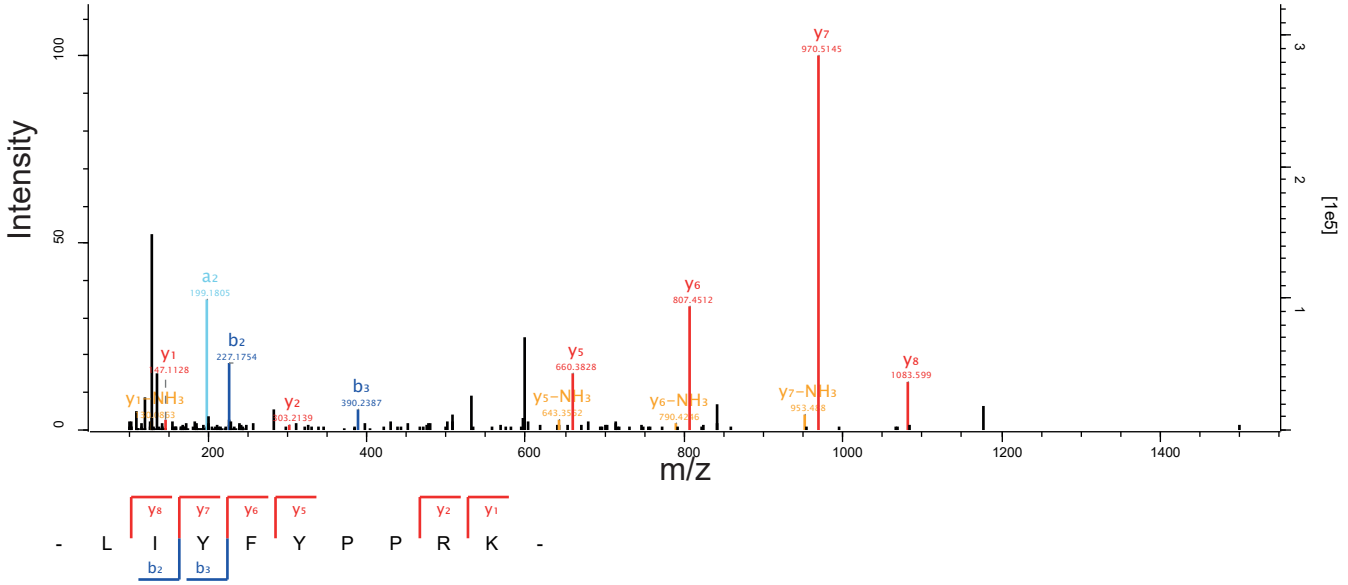
152A2 Bacteroides dorei HLA-I KLADPGTVLQK

Endogenous peptide



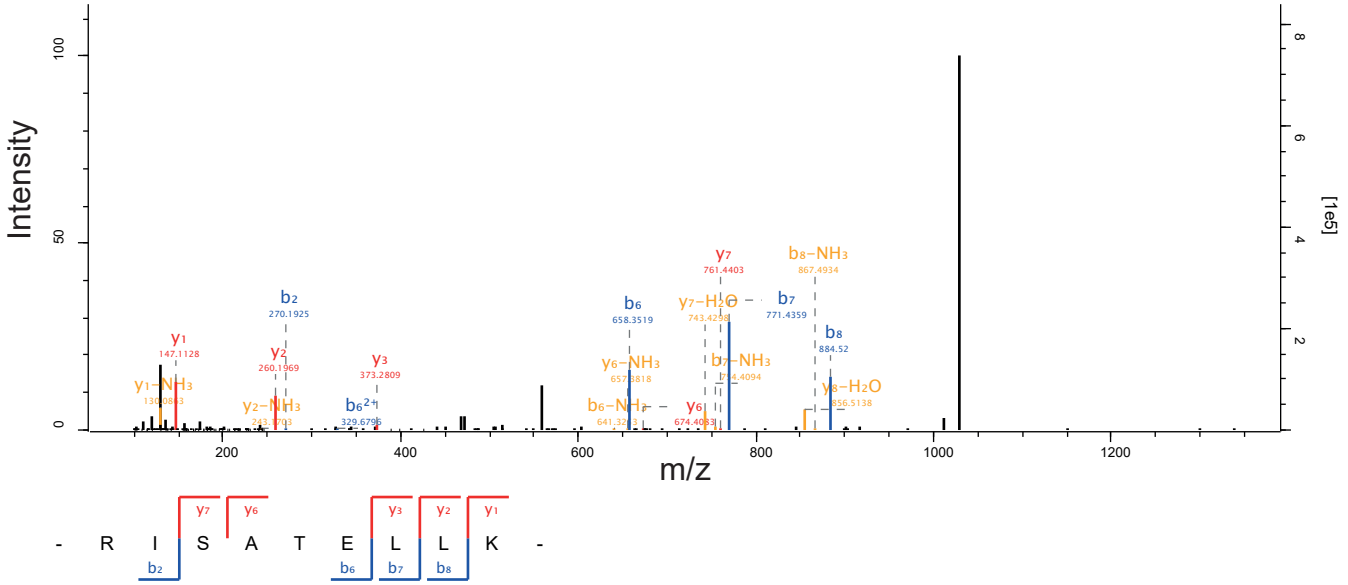
152A2 Bacteroides dorei HLA-I LIYFYPPRK

Endogenous peptide



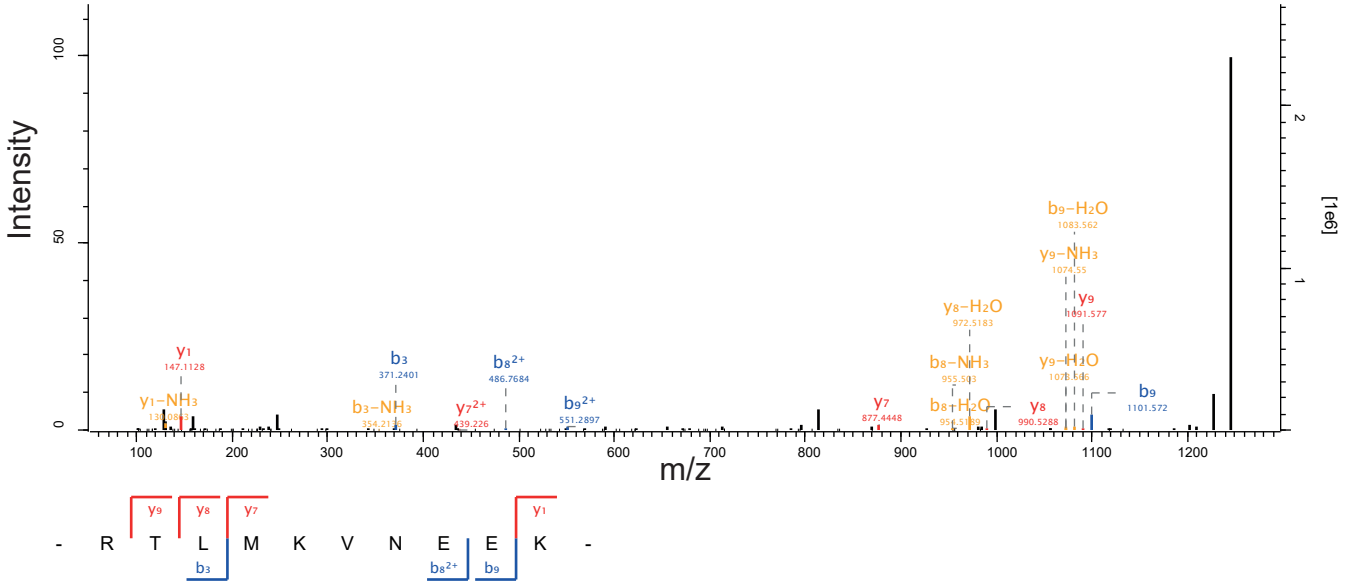
152A2 Bacteroides dorei HLA-I RISATELLK

Endogenous peptide



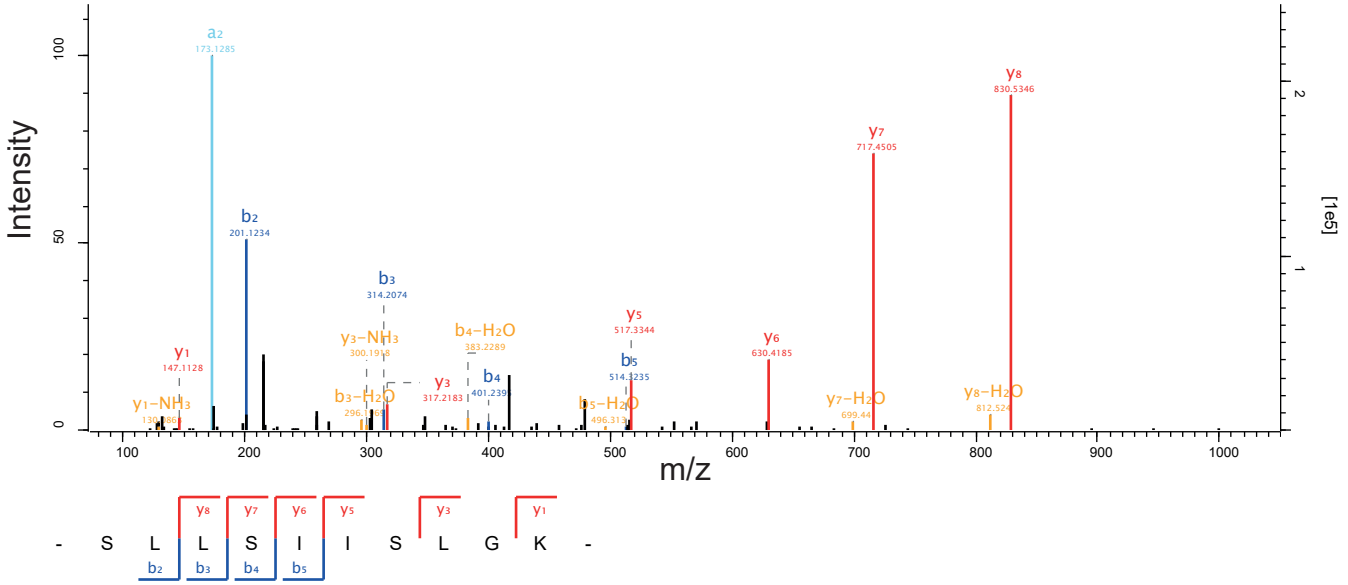
152A2 Bacteroides dorei HLA-I RTLMKVNEEK

Endogenous peptide



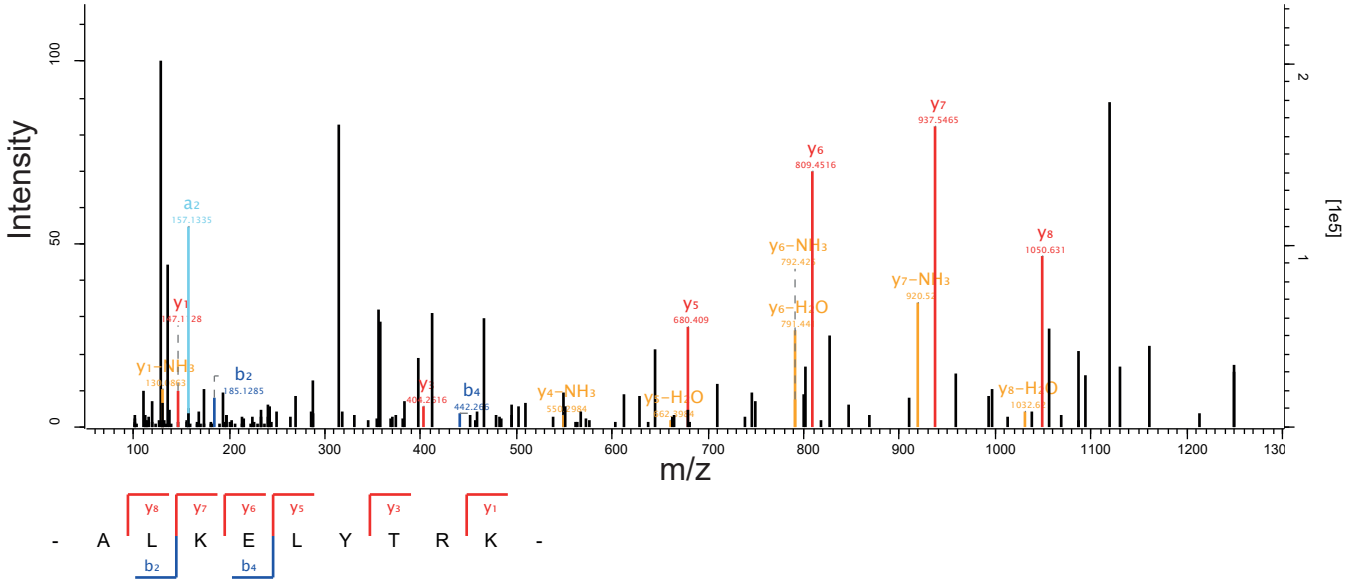
152A2 Bacteroides dorei HLA-I SLLSIISLGK

Endogenous peptide



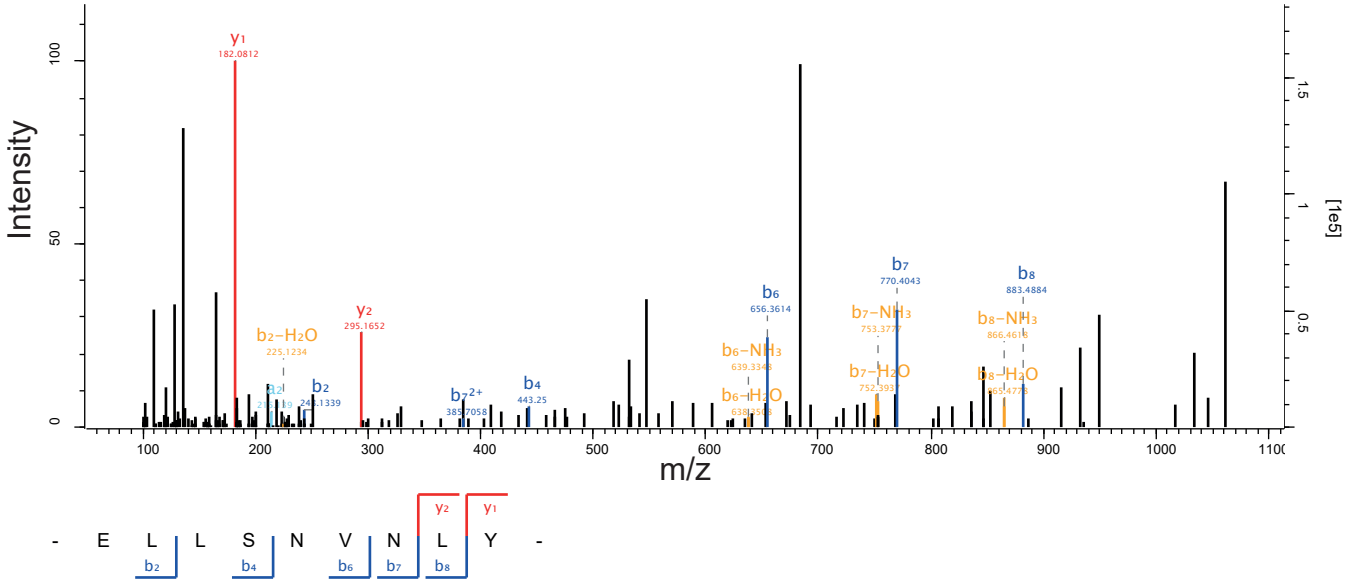
152A2 Bacteroides ovatus HLA-I ALKELYTRK

Endogenous peptide



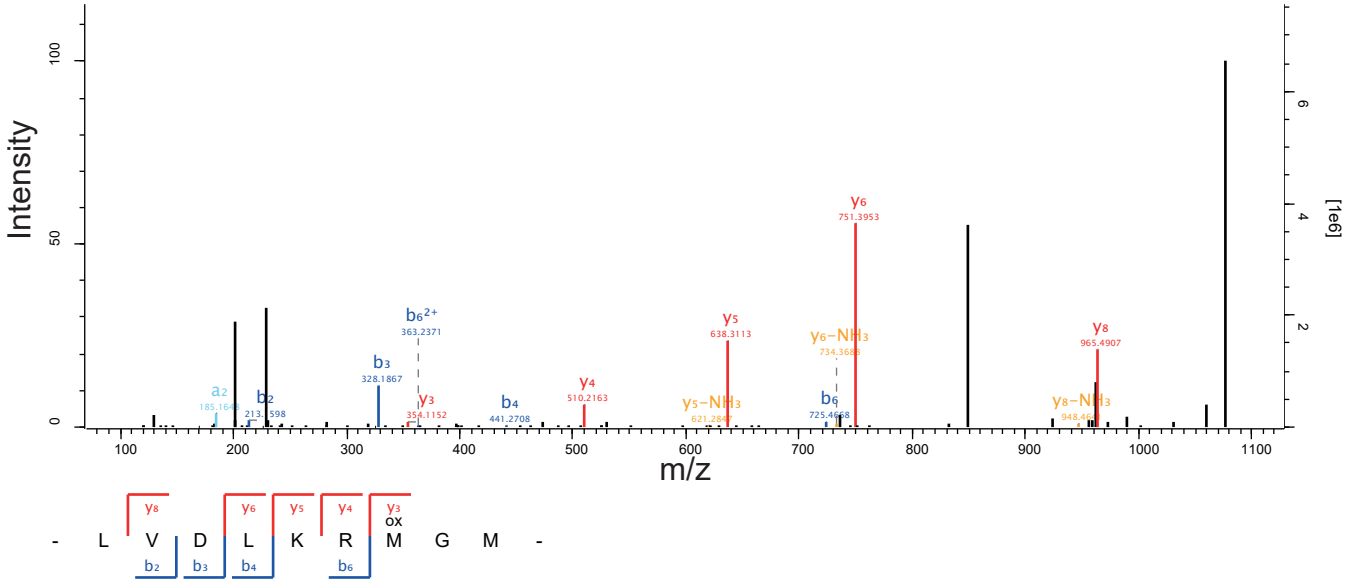
152A2 Bacteroides ovatus HLA-I ELLSNVNLY

Endogenous peptide



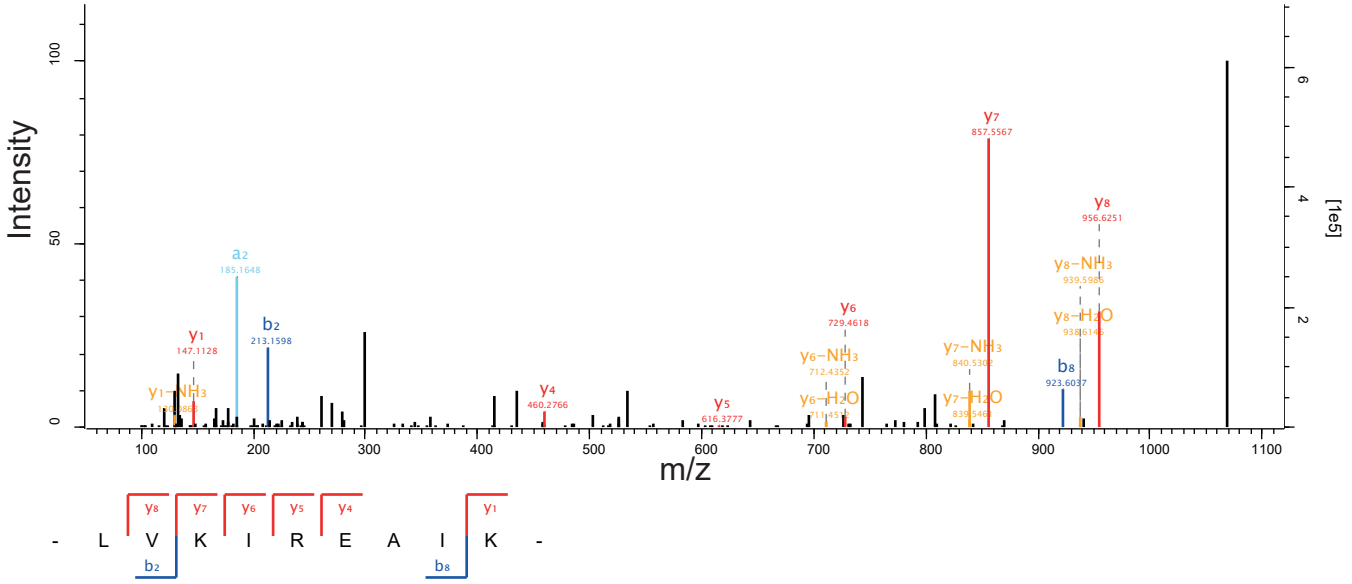
152A2 Bacteroides ovatus HLA-I LVDLKRMGM

Endogenous peptide



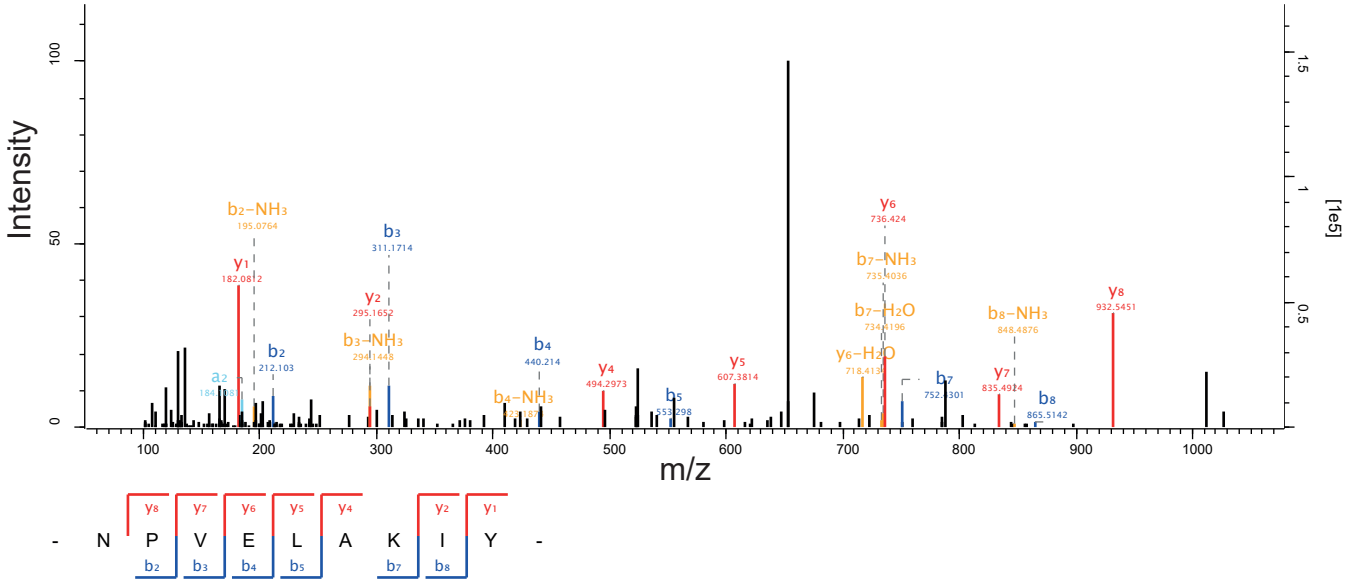
152A2 Bacteroides ovatus HLA-I LVKIREAIK

Endogenous peptide



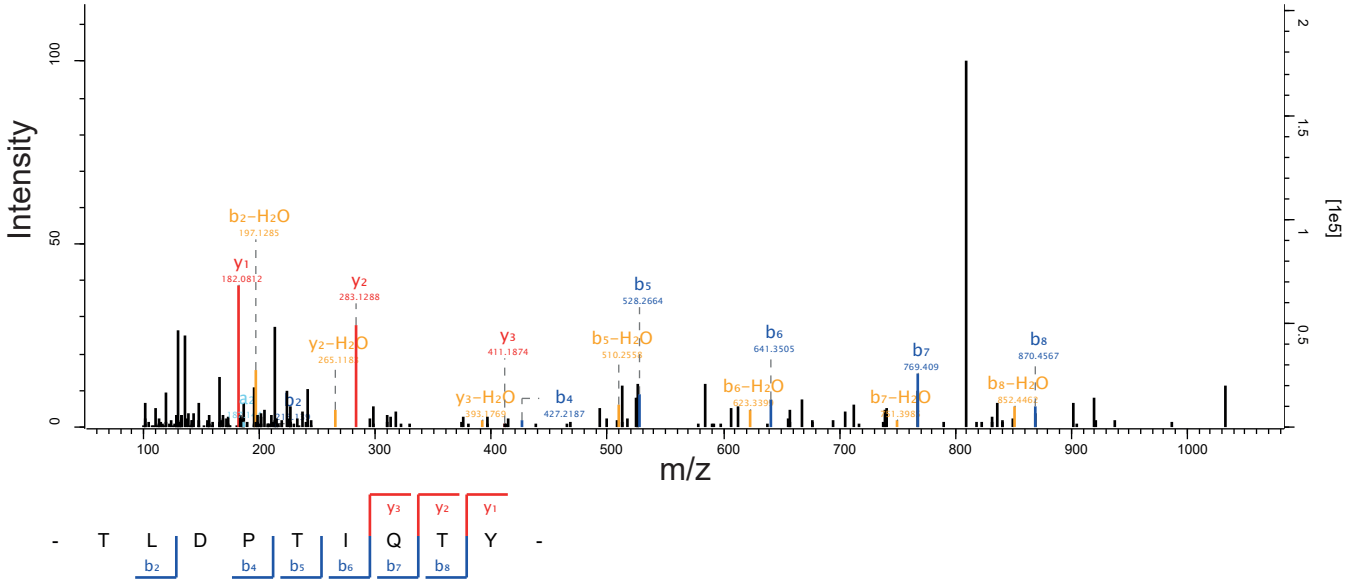
152A2 Bacteroides ovatus HLA-I NPVELAKIY

Endogenous peptide



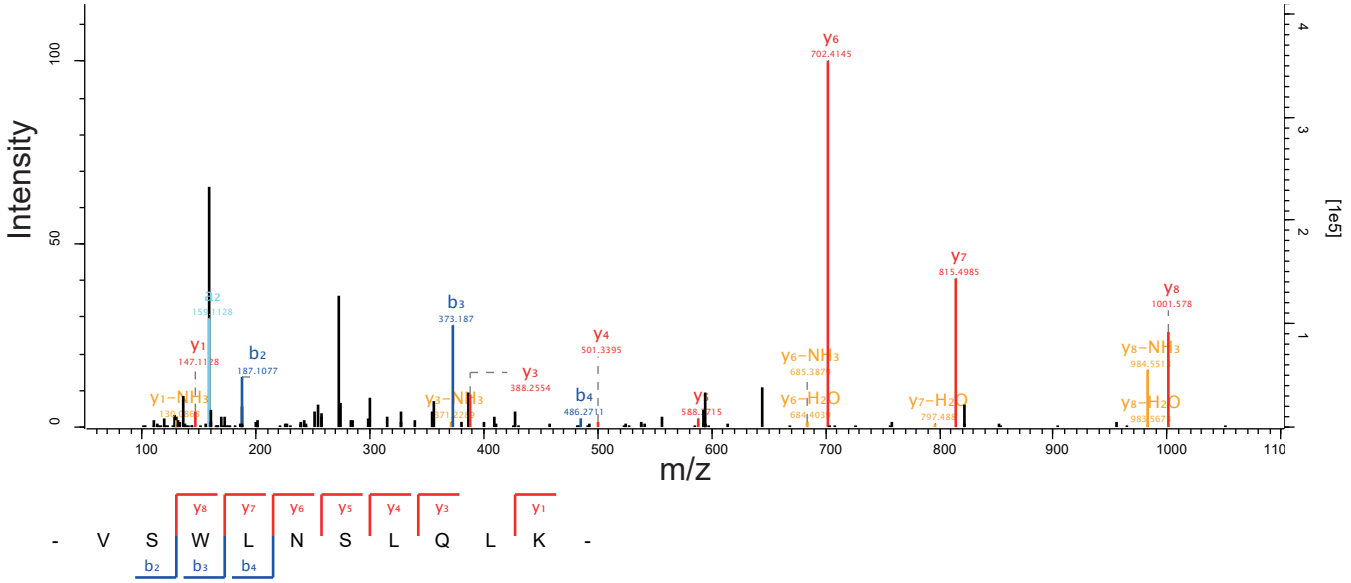
152A2 Bacteroides ovatus HLA-I TLDPTIQTY

Endogenous peptide



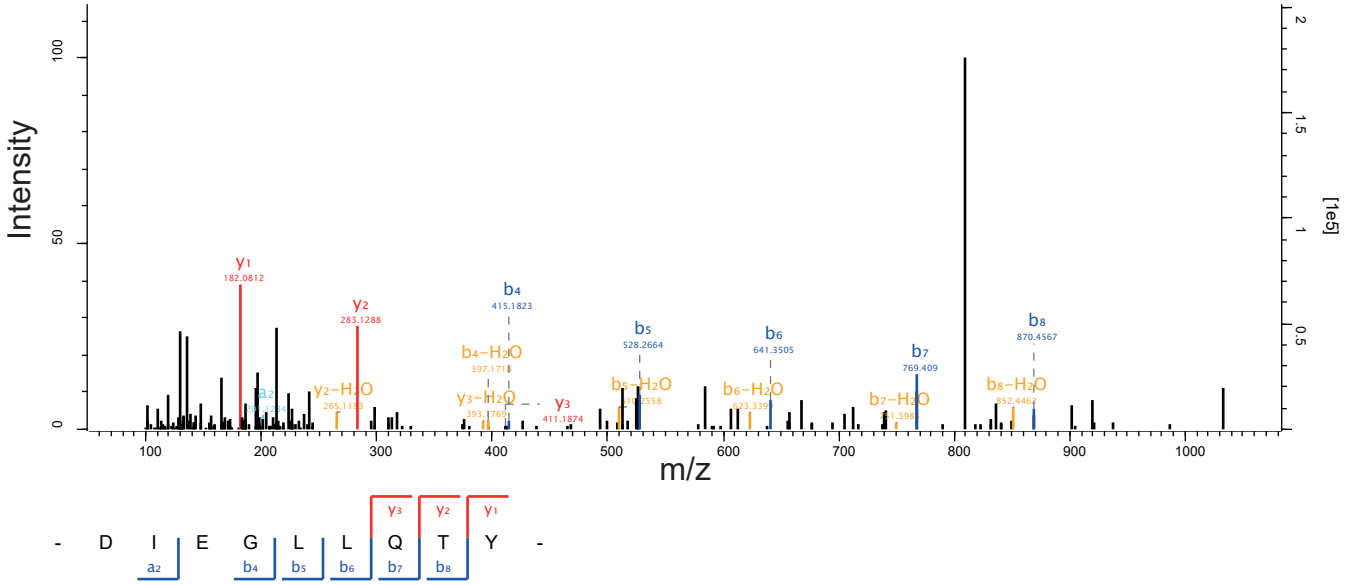
152A2 Bacteroides ovatus HLA-I VSWLNSLQLK

Endogenous peptide



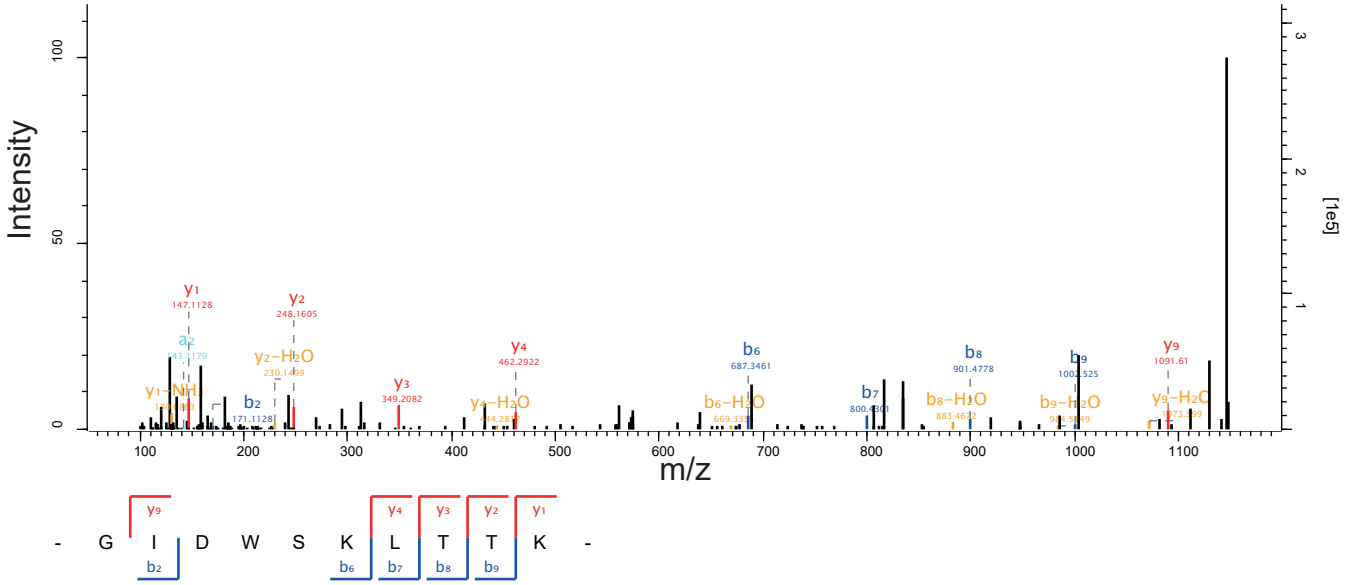
152A2 Bacteroides vulgatus HLA-I DIEGLLQTY

Endogenous peptide



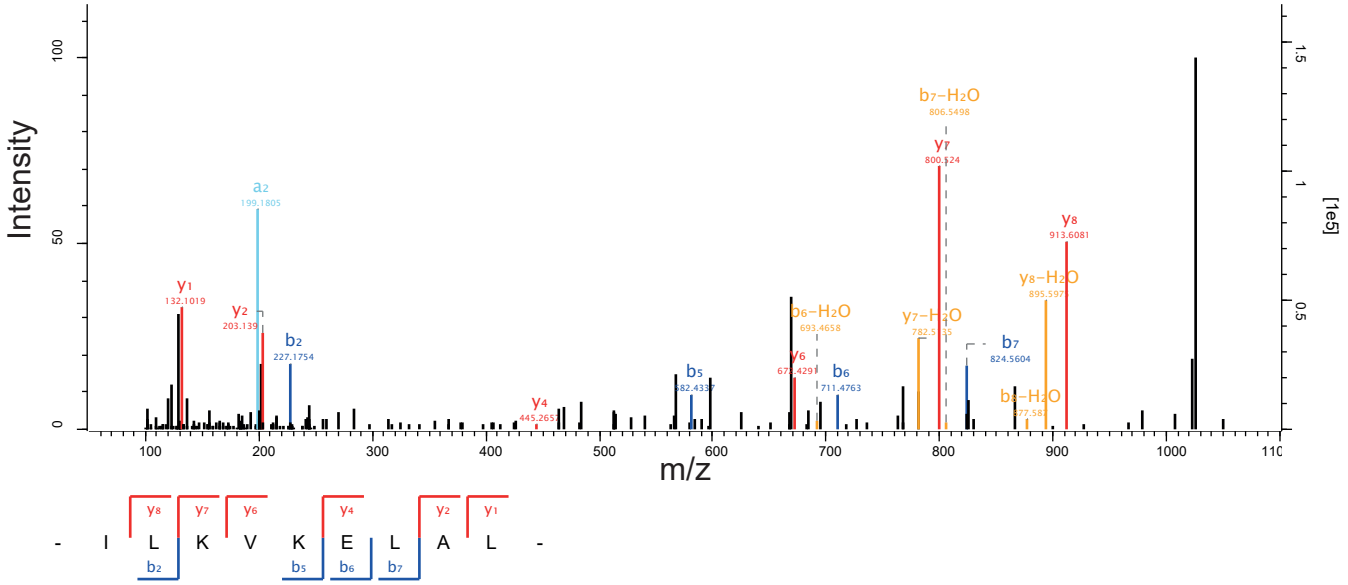
152A2 Bacteroides vulgatus HLA-I GIDWSKLTTK

Endogenous peptide



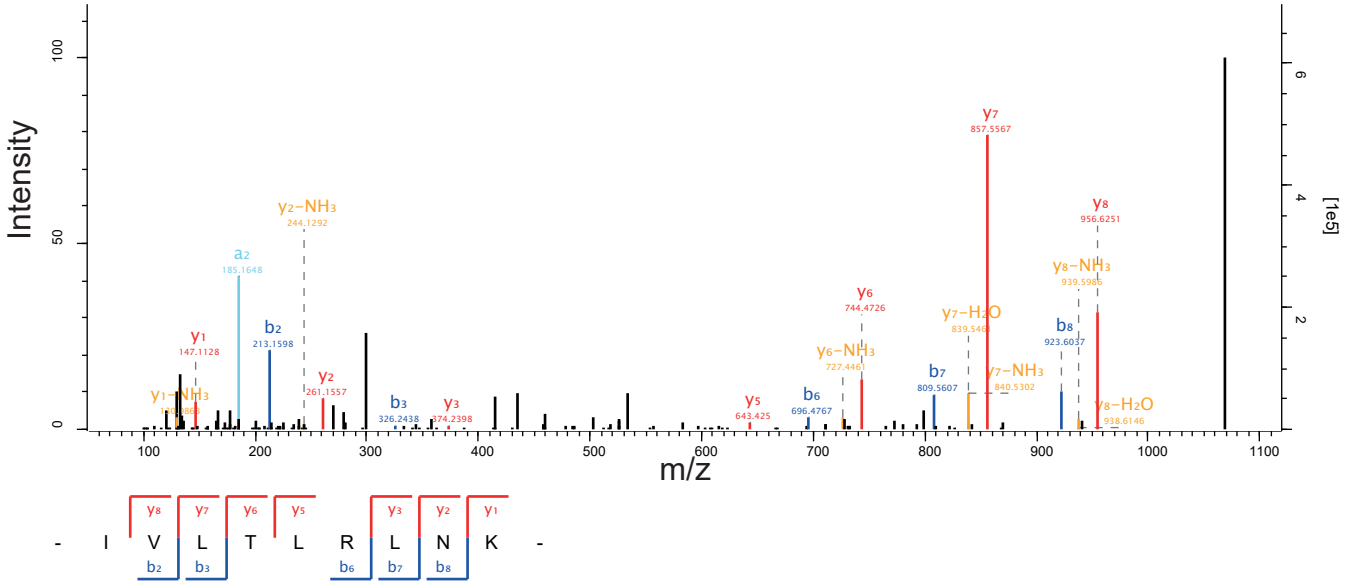
152A2 Bacteroides vulgatus HLA-I ILKVKELAL

Endogenous peptide



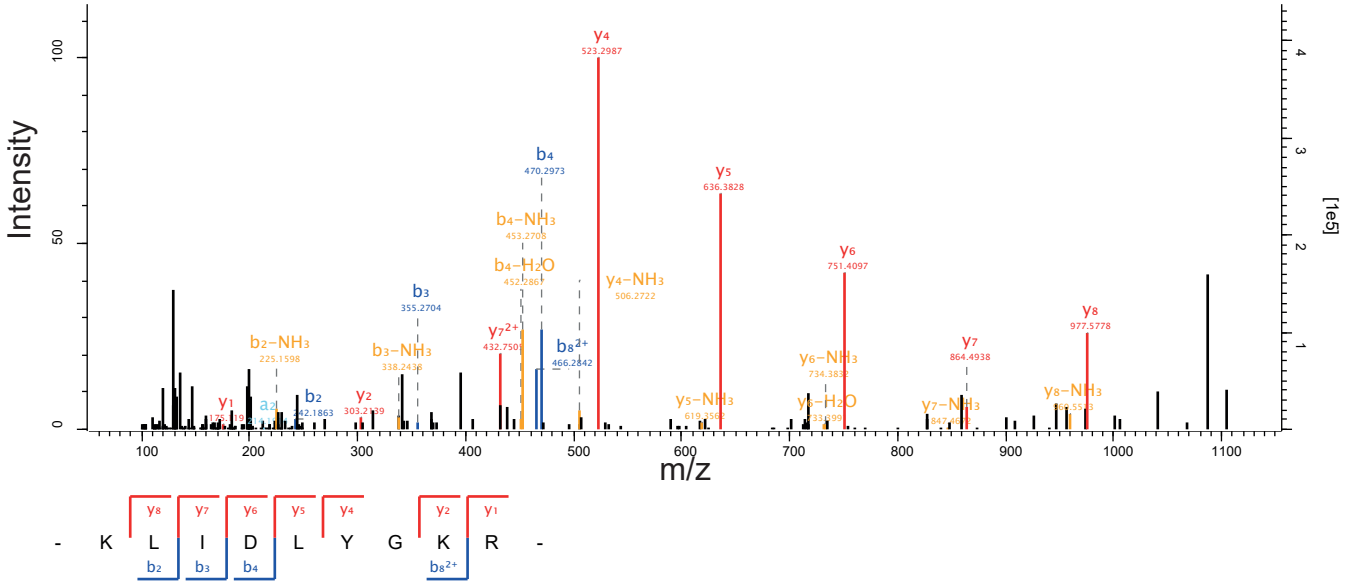
152A2 Bacteroides vulgatus HLA-I IVLTLRLNK

Endogenous peptide



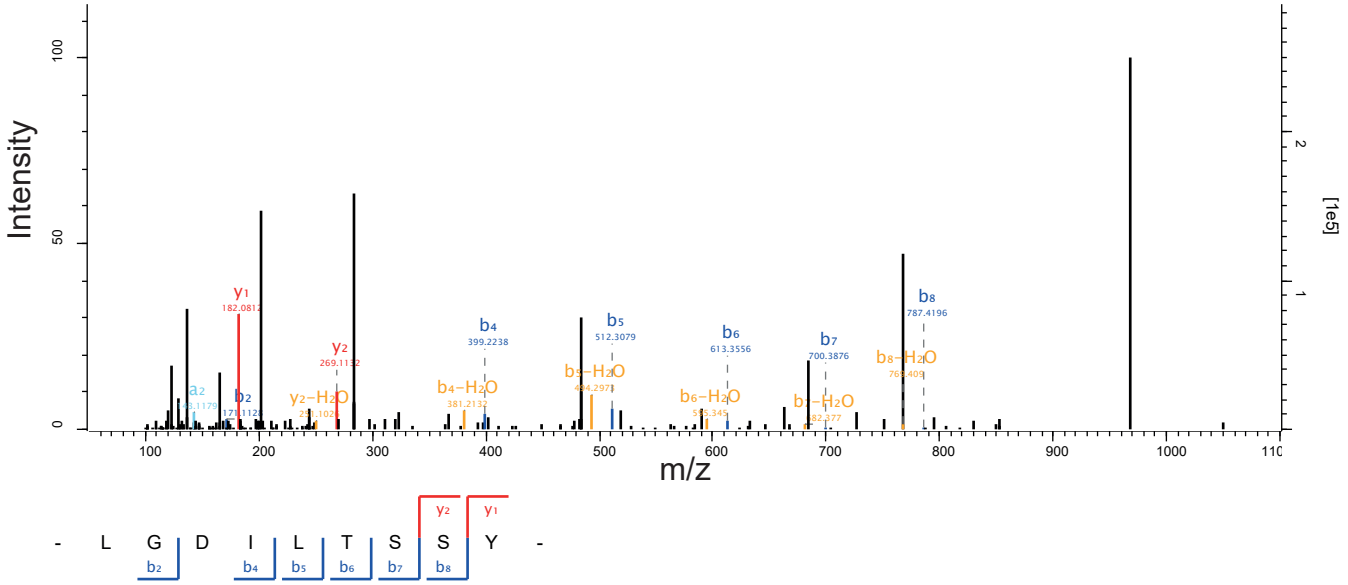
152A2 Bacteroides vulgatus HLA-I KLIDLYGKR

Endogenous peptide



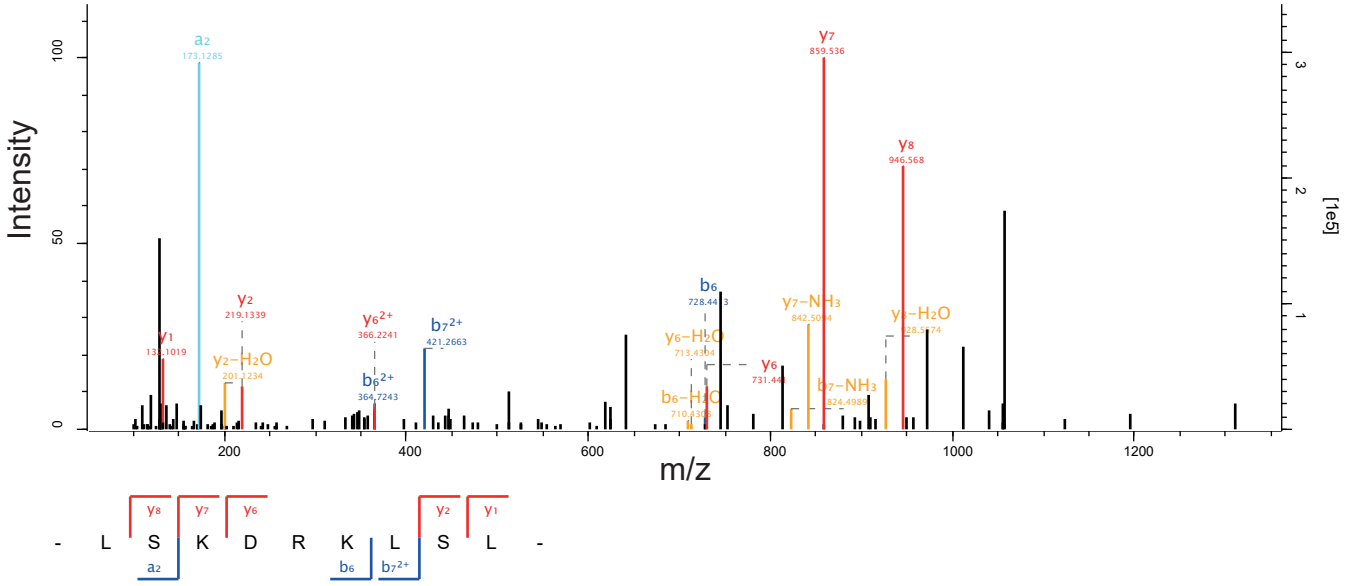
152A2 Bacteroides vulgatus HLA-I LGDILTSSY

Endogenous peptide



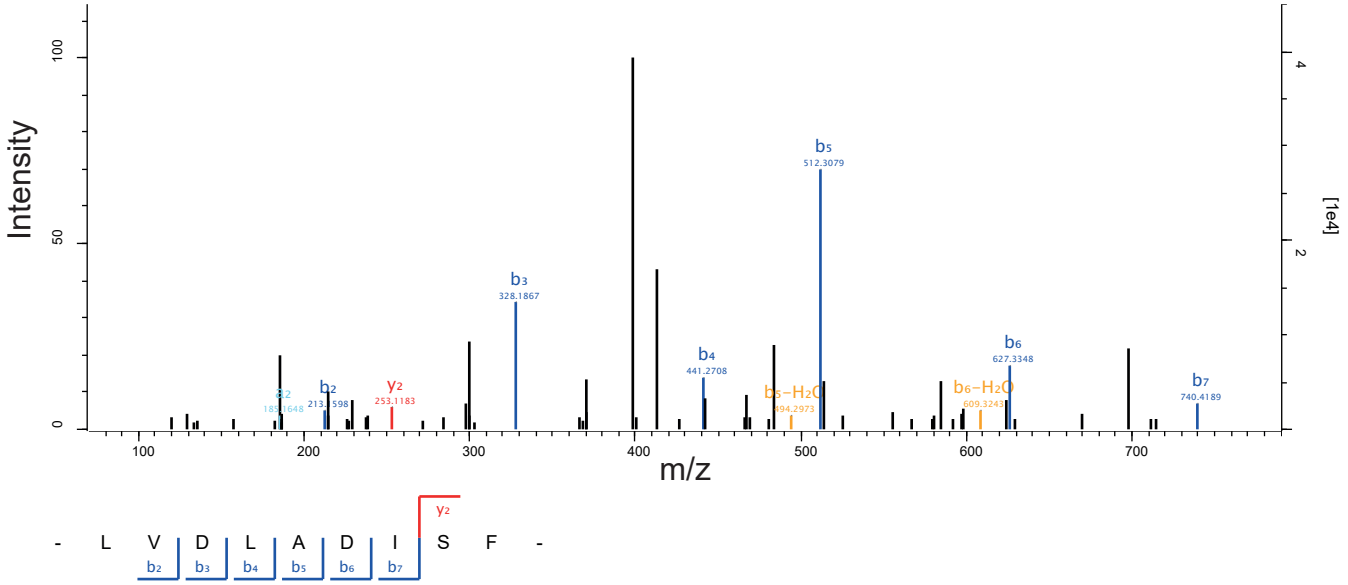
152A2 Bacteroides vulgatus HLA-I LSKDRKLSL

Endogenous peptide



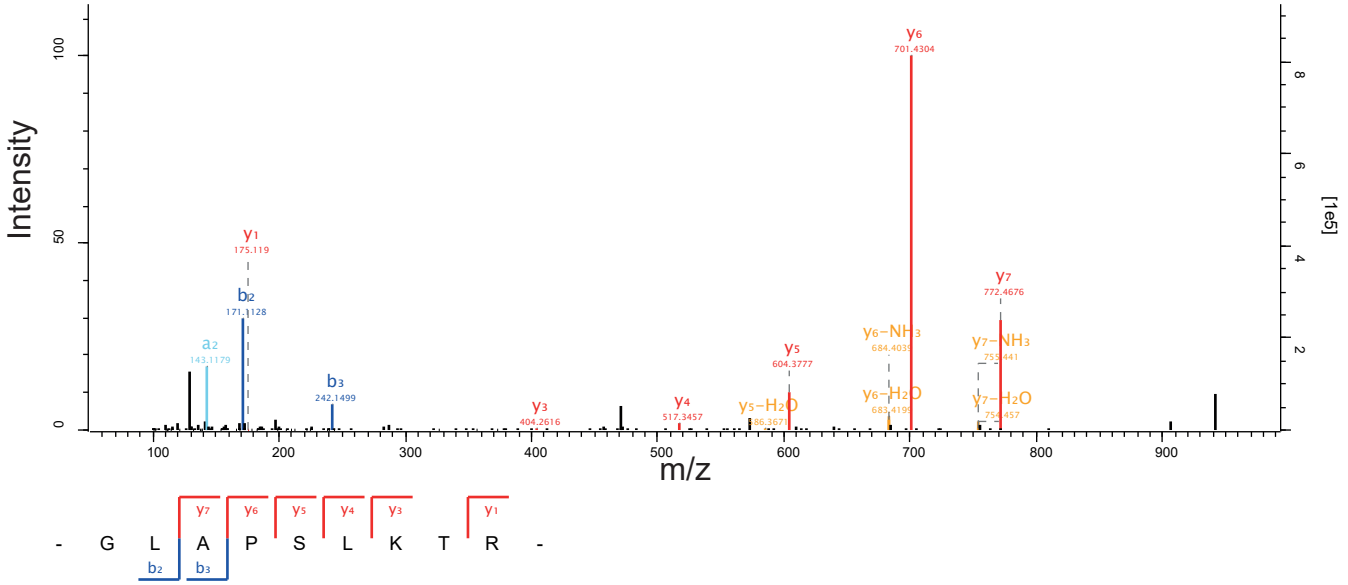
152A2 Bacteroides vulgatus HLA-I LVDLADISF

Endogenous peptide



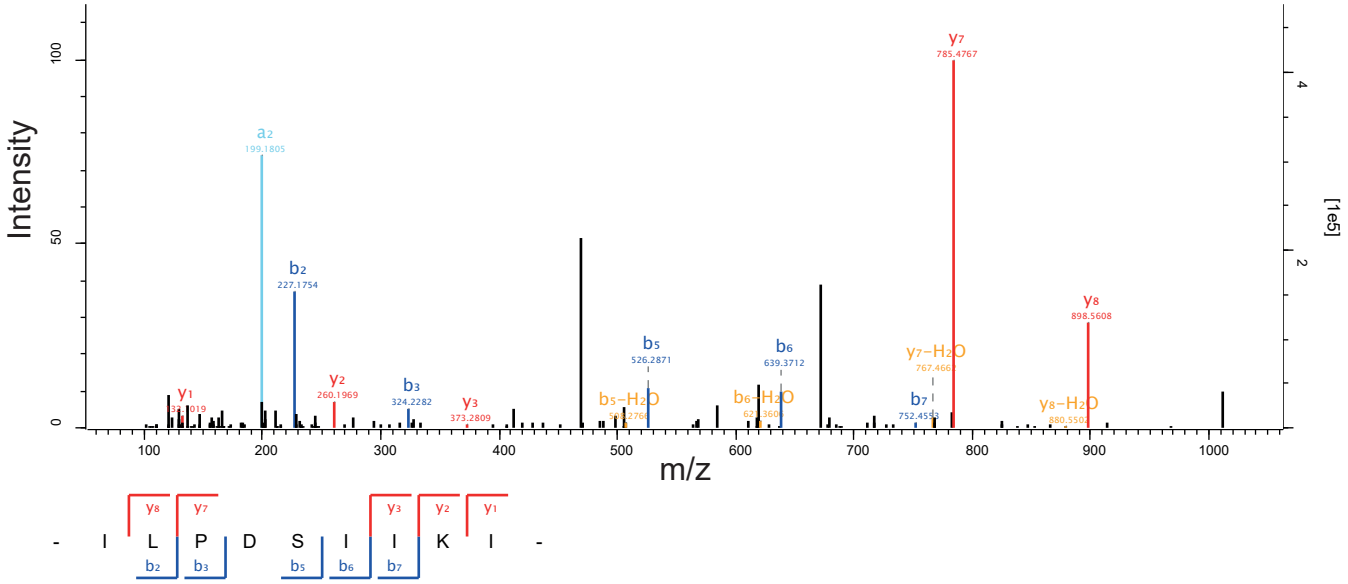
152A2 Campylobacter concisus HLA-I GLAPSLKTR

Endogenous peptide



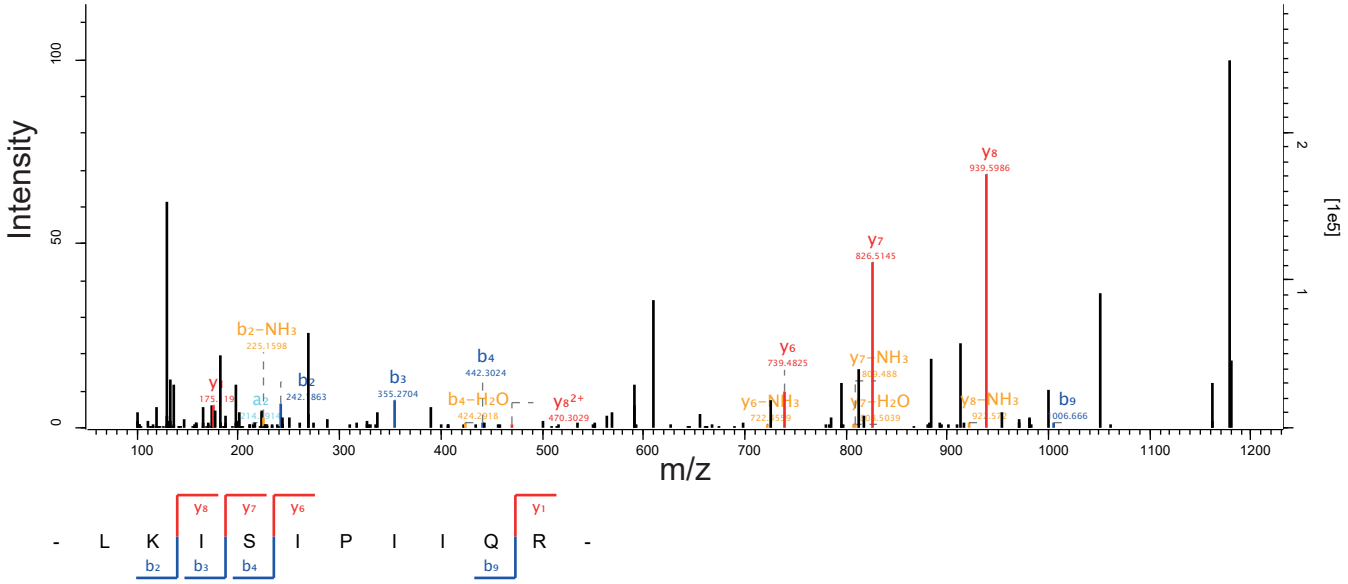
152A2 Campylobacter concisus HLA-I ILPDSIIKI

Endogenous peptide



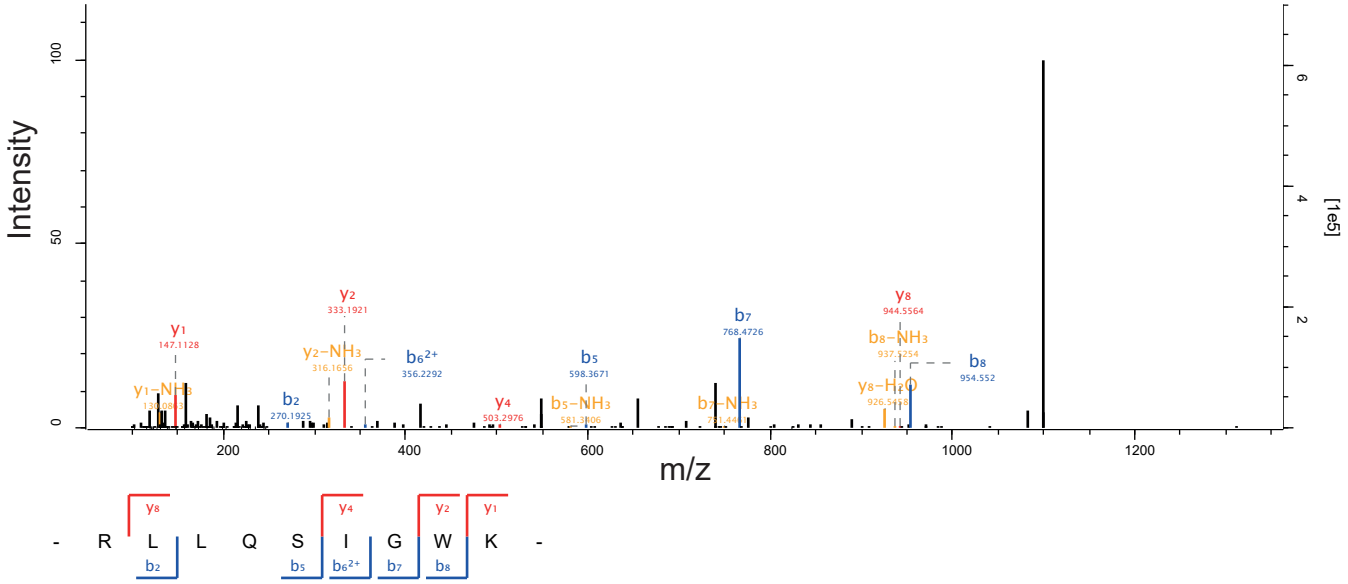
152A2 Campylobacter concisus HLA-I LKISIPIIQR

Endogenous peptide



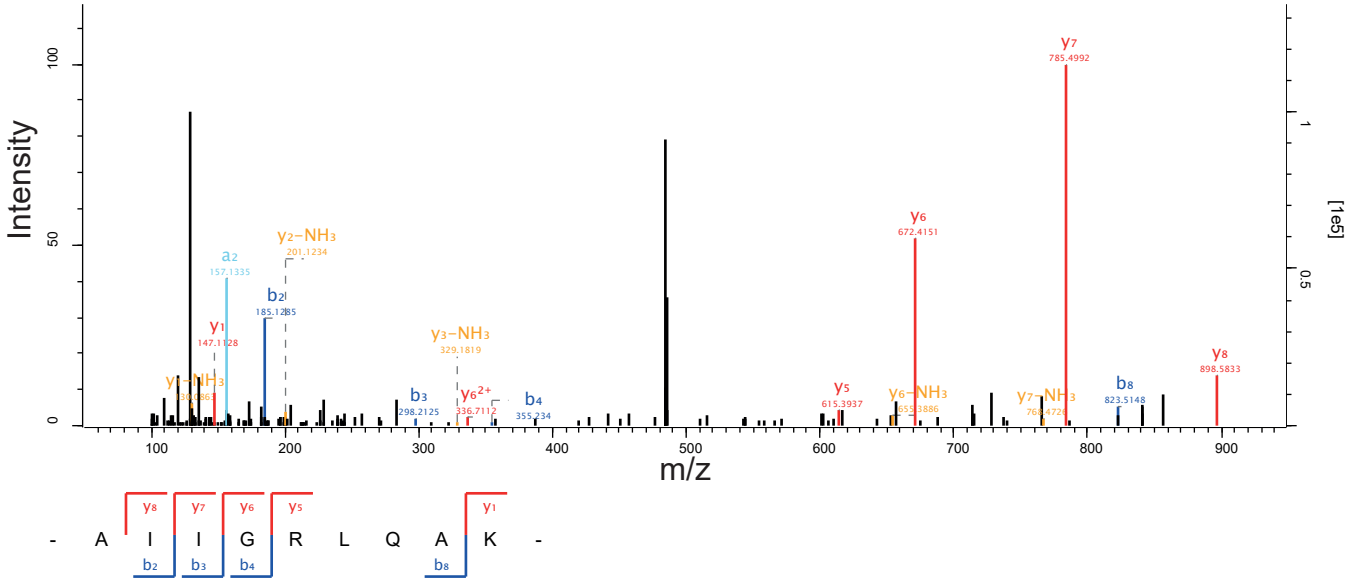
152A2 Campylobacter concisus HLA-I RLLQSIGWK

Endogenous peptide



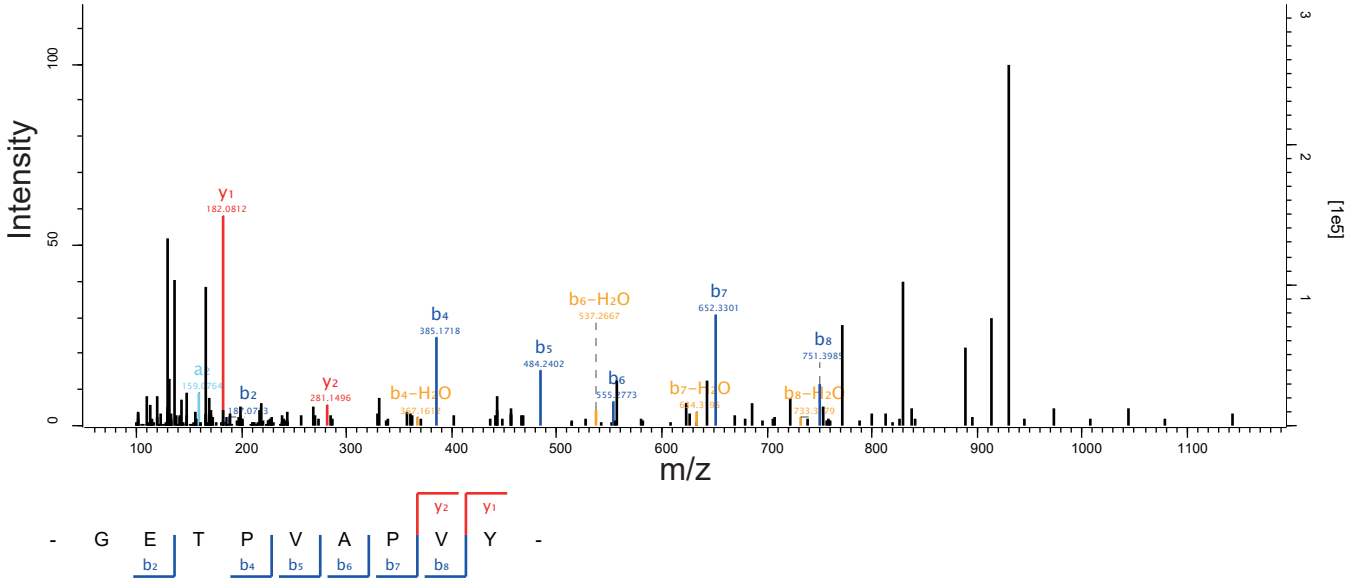
152A2 Eikenella corrodens HLA-I AIIGRLQAK

Endogenous peptide



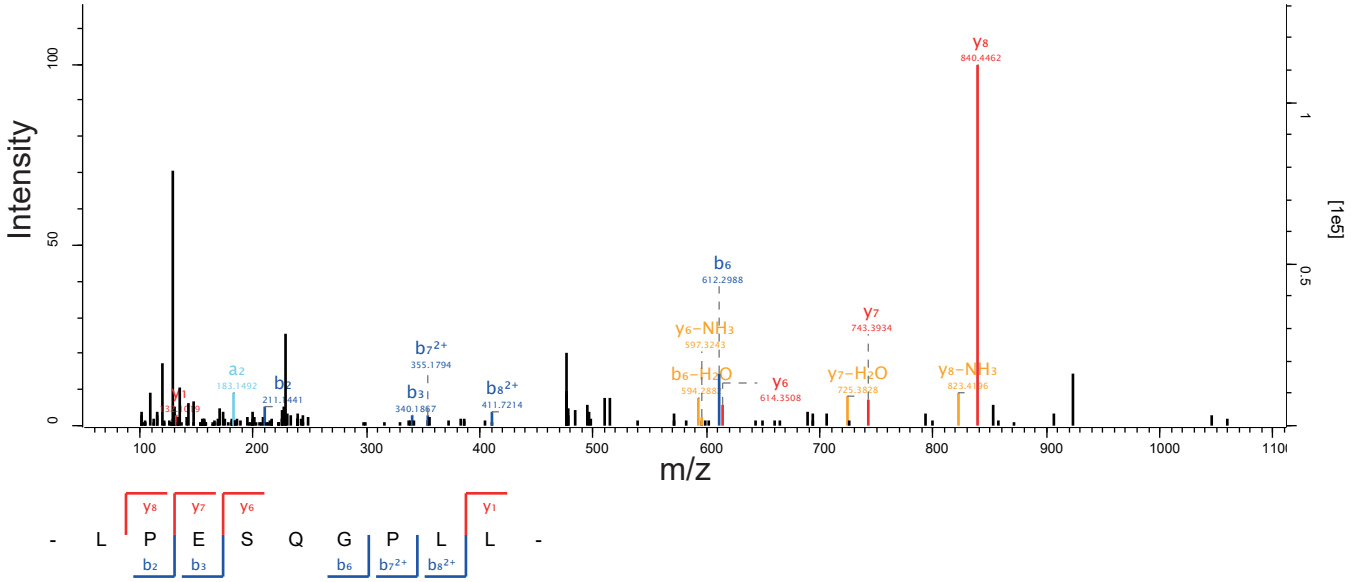
152A2 Eikenella corrodens HLA-I GETPVAPVY

Endogenous peptide



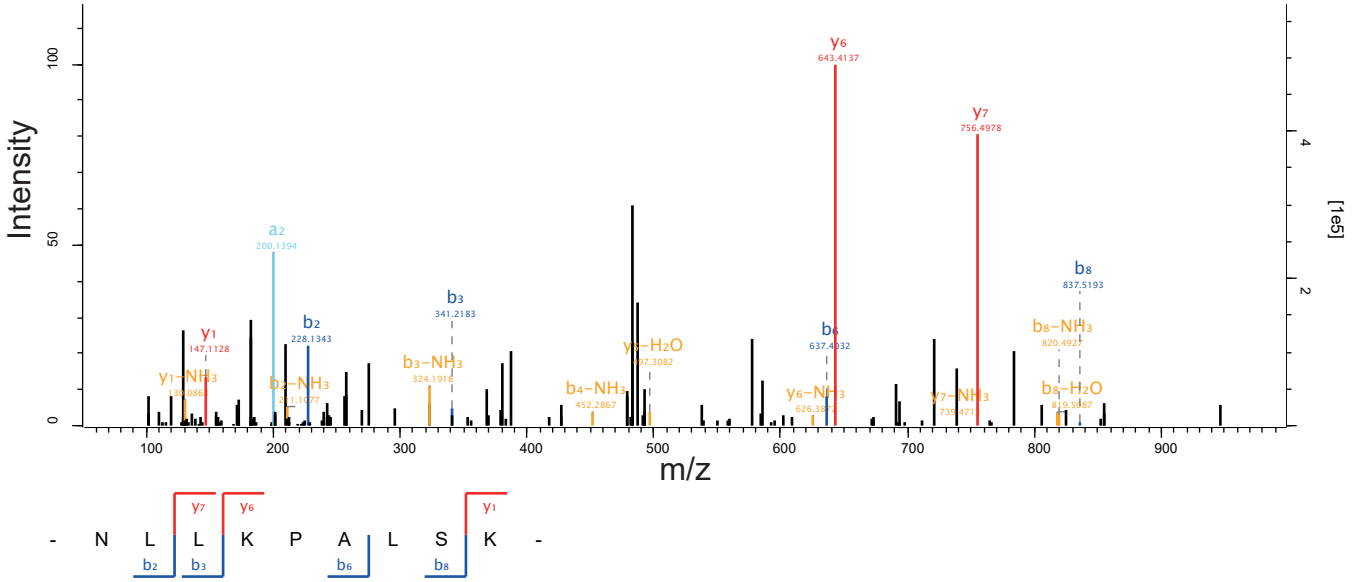
152A2 Eikenella corrodens HLA-I LPESQGPLL

Endogenous peptide



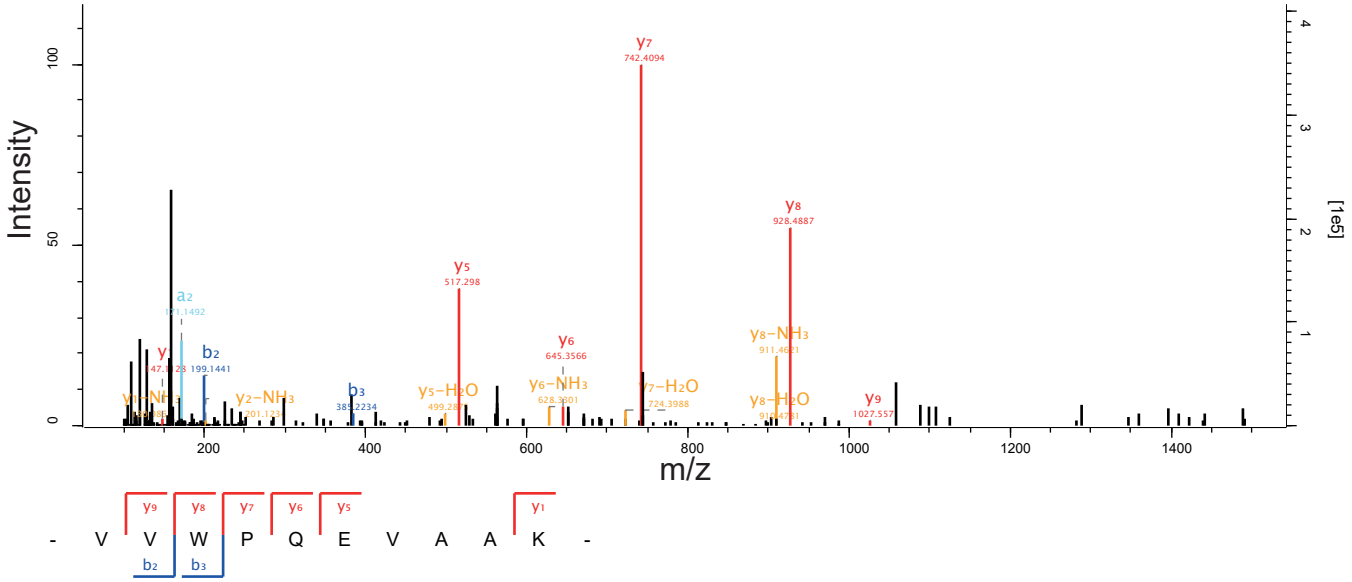
152A2 Eikenella corrodens HLA-I NLLKPALSK

Endogenous peptide



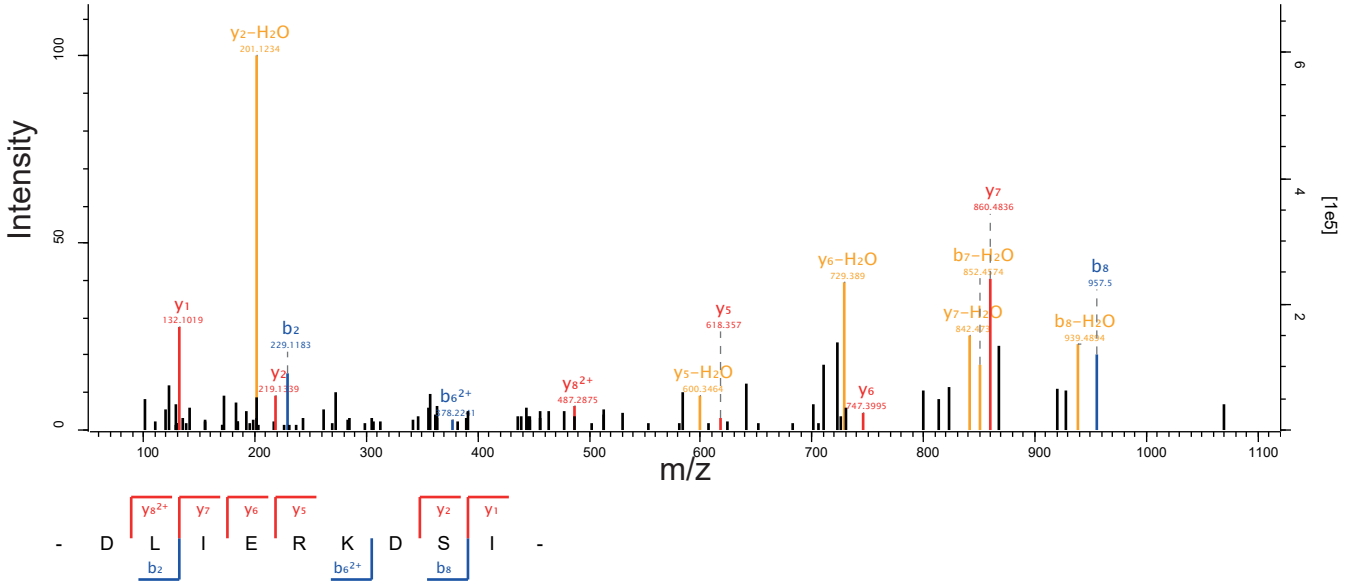
152A2 Eikenella corrodens HLA-I VVWPQEVAAK

Endogenous peptide



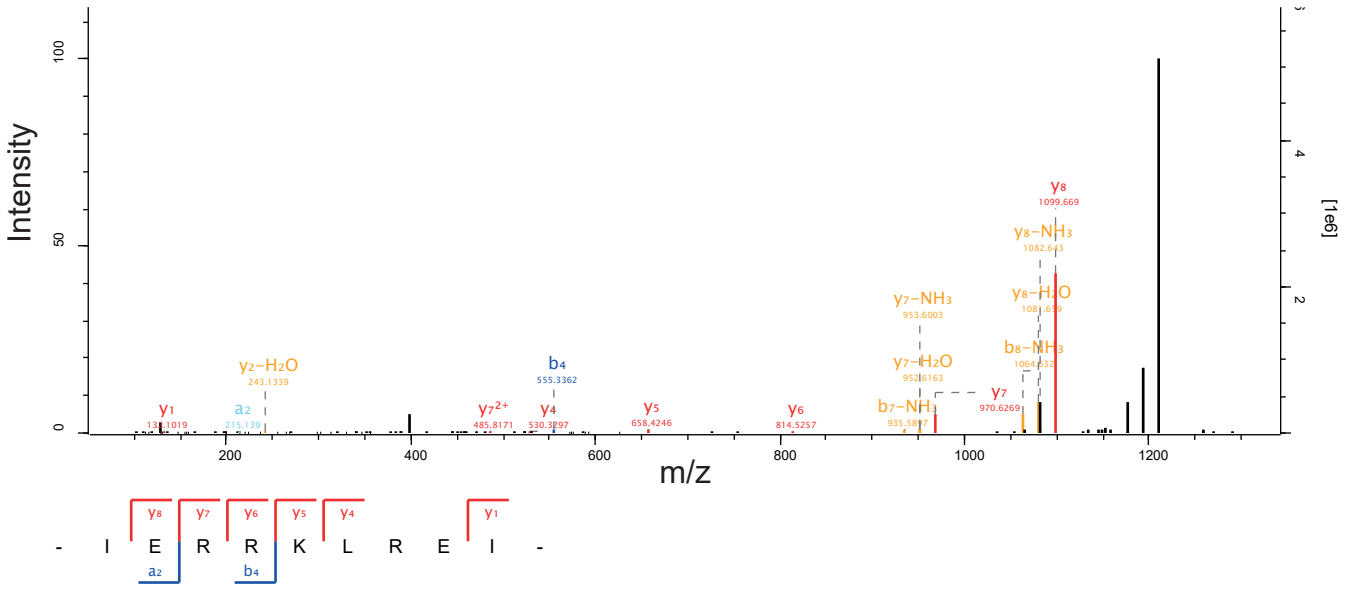
152A2 Fusobacterium nucleatum HLA-I DLIERKDSI

Endogenous peptide



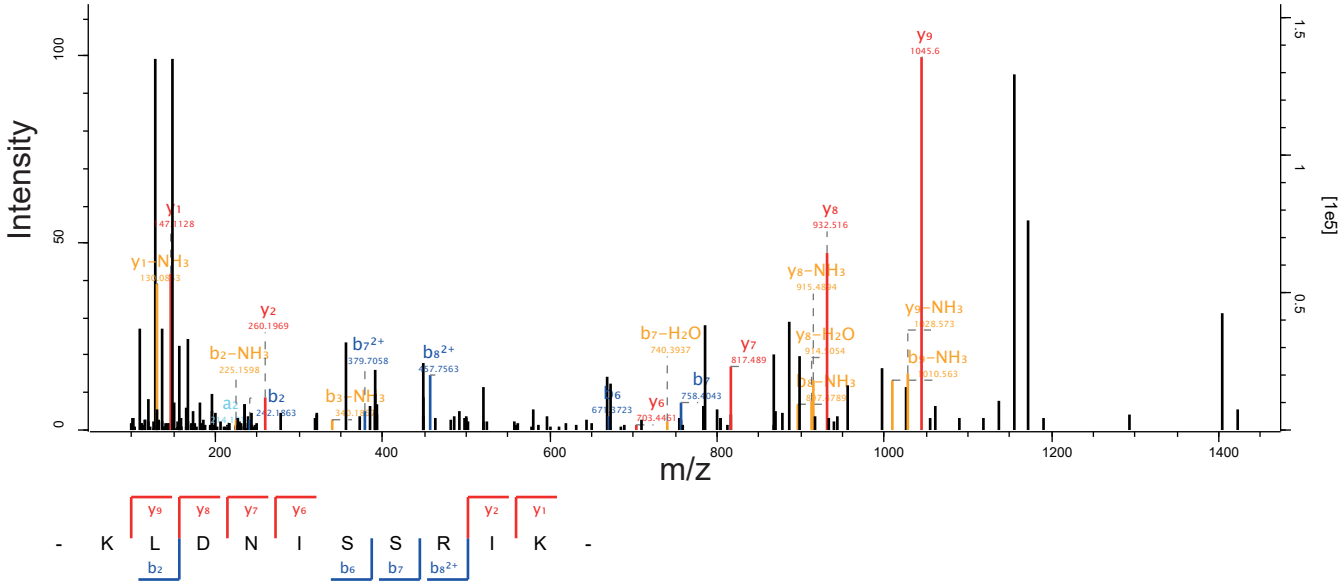
152A2 Fusobacterium nucleatum HLA-I IERRKLREI

Endogenous peptide



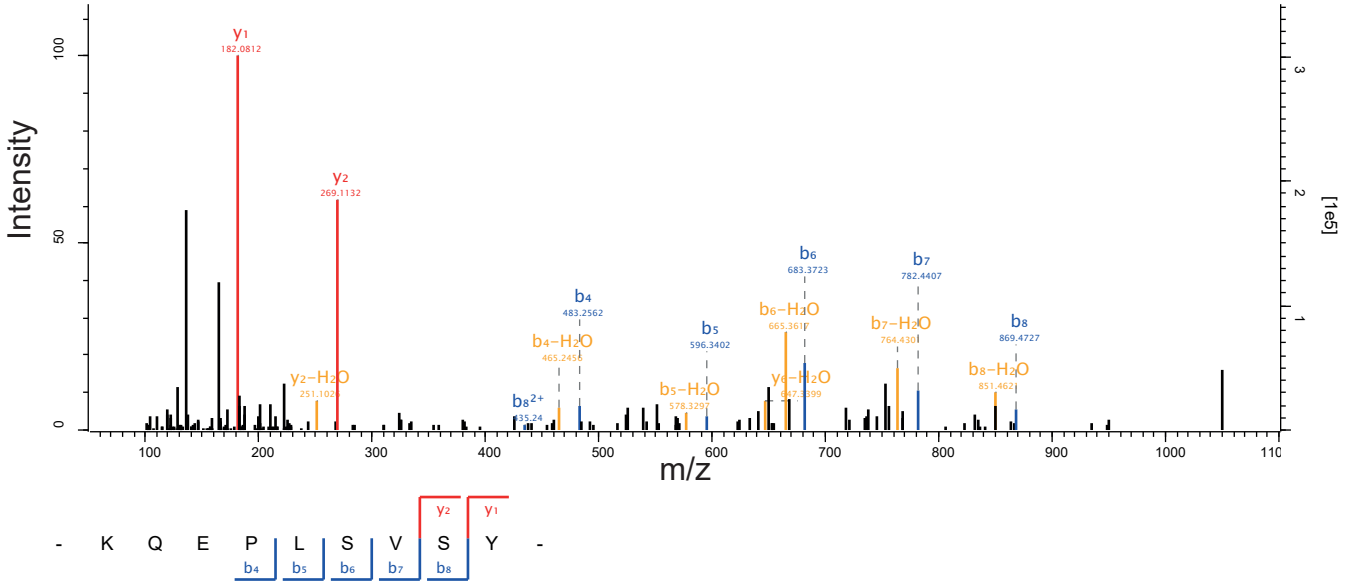
152A2 Fusobacterium nucleatum HLA-I KLDNISSRIK

Endogenous peptide



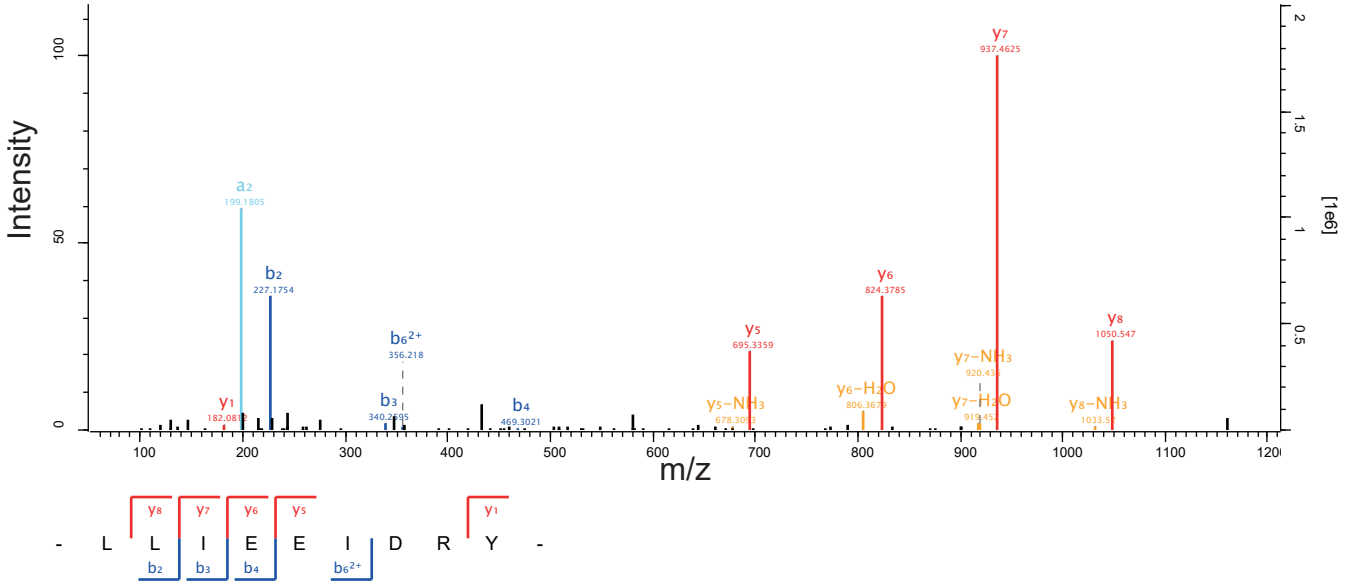
152A2 Fusobacterium nucleatum HLA-I KQEPLSVSY

Endogenous peptide



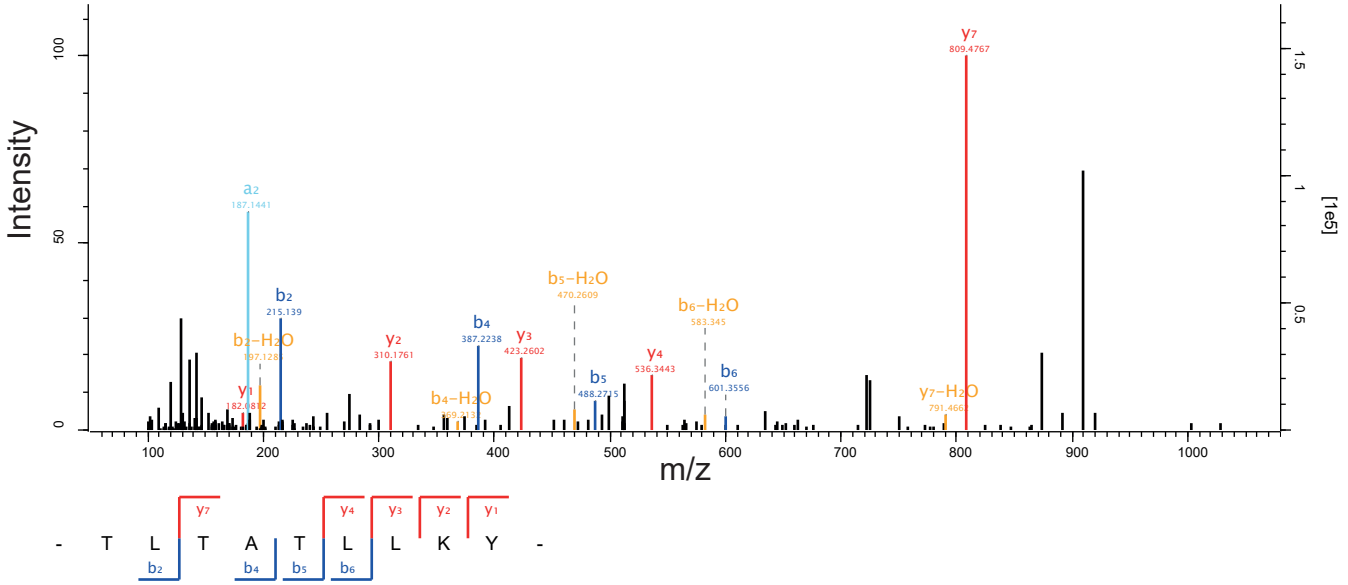
152A2 Fusobacterium nucleatum HLA-I LLIEEIDRY

Endogenous peptide



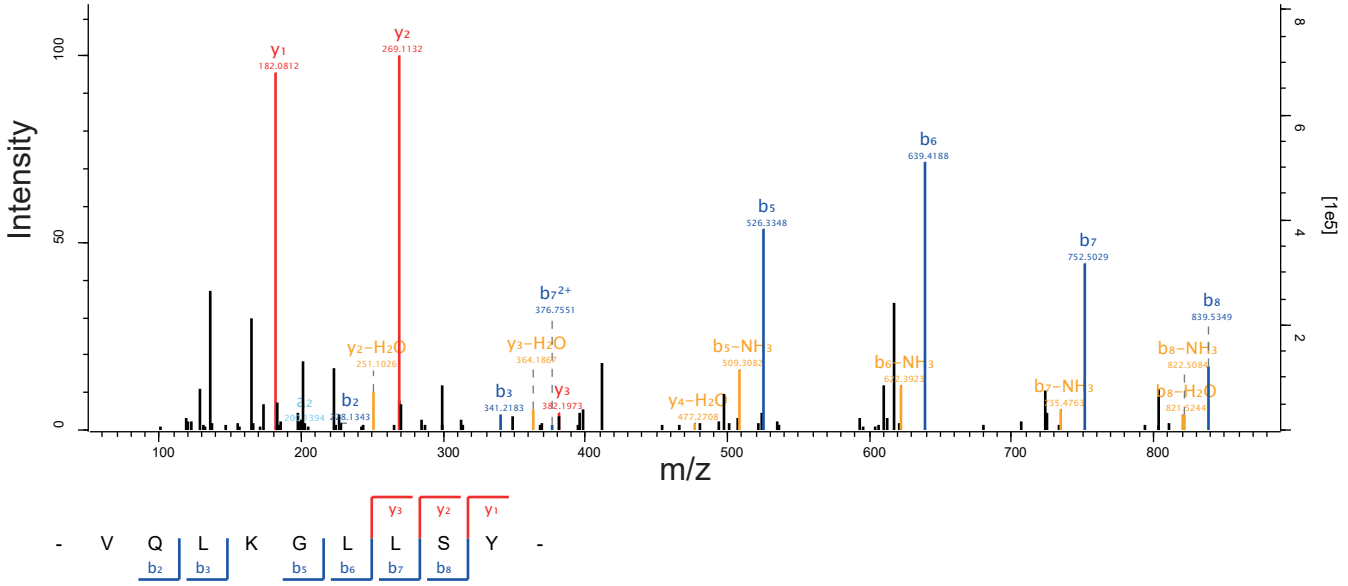
152A2 Fusobacterium nucleatum HLA-I TLTATLLKY

Endogenous peptide



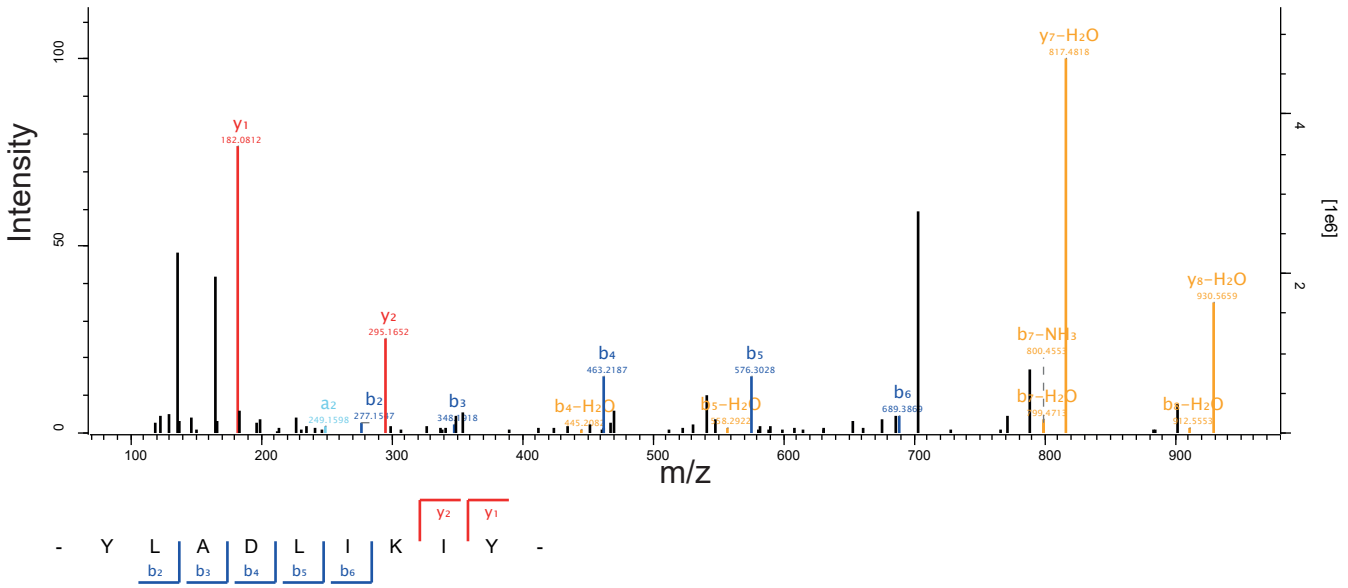
152A2 Fusobacterium nucleatum HLA-I VQLKGLLSY

Endogenous peptide



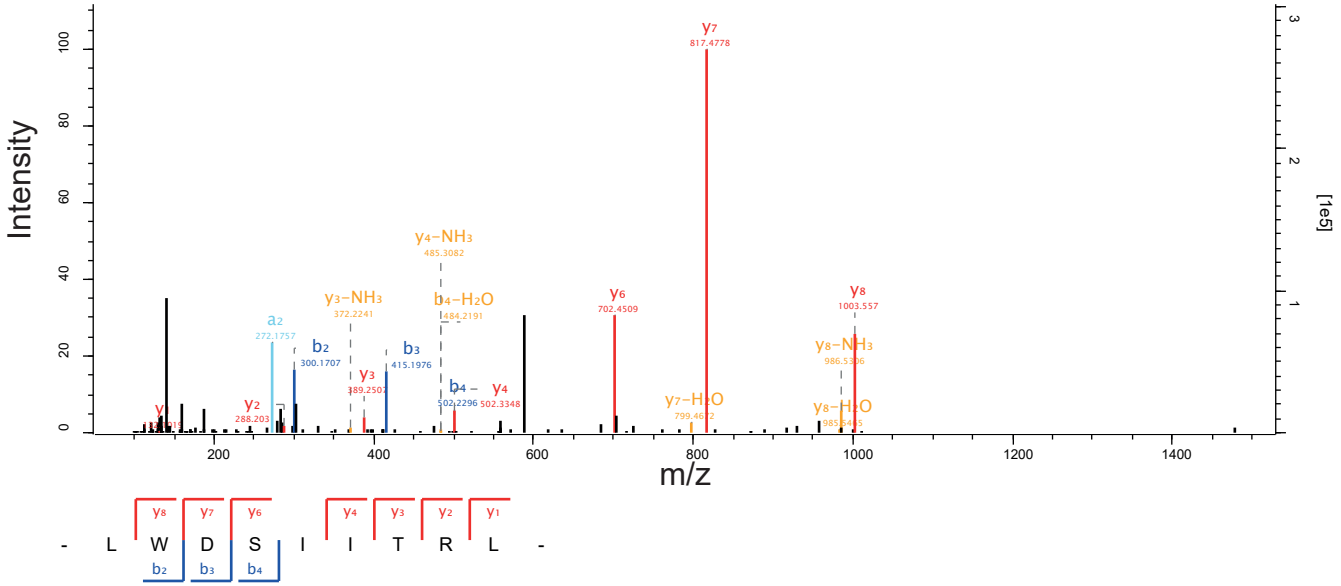
152A2 Fusobacterium nucleatum HLA-I YLADLIKIY

Endogenous peptide



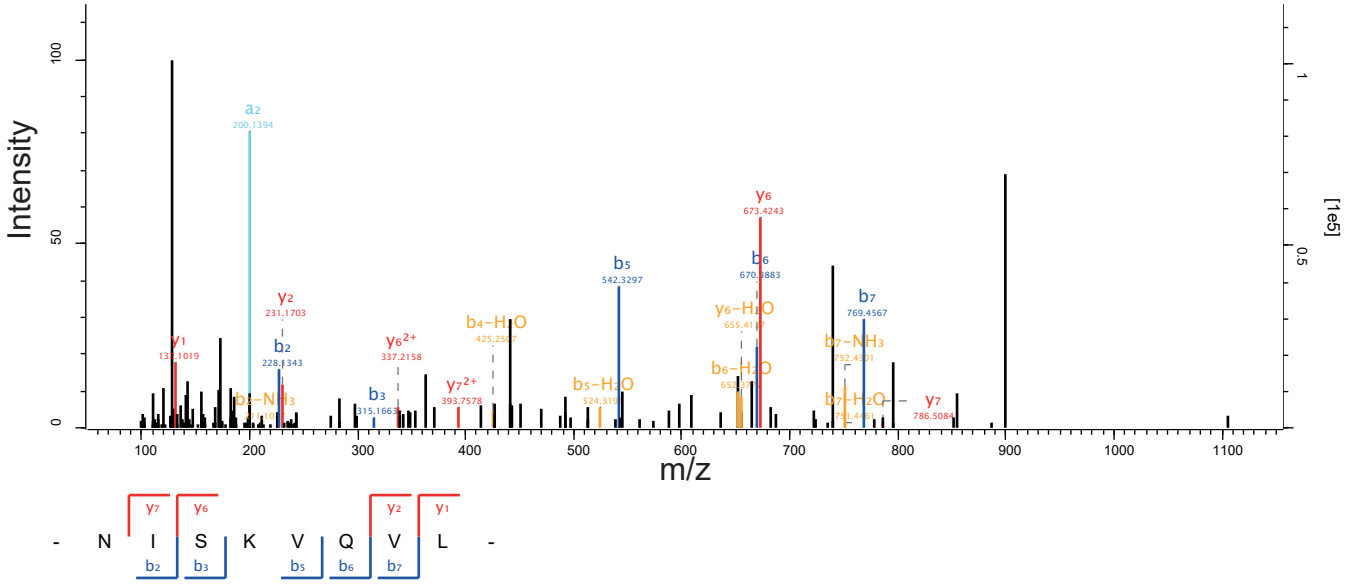
152A2 Veillonella dispar HLA-I LWDSIITRL

Endogenous peptide



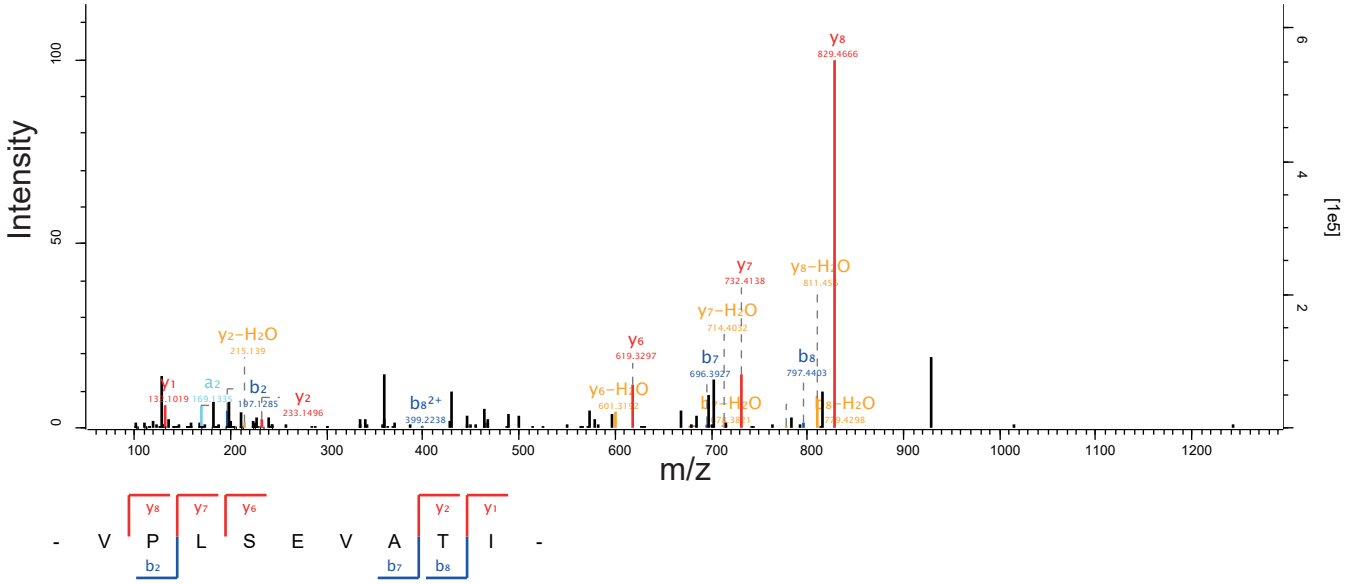
152A2 Veillonella dispar HLA-I NISKVQVL

Endogenous peptide



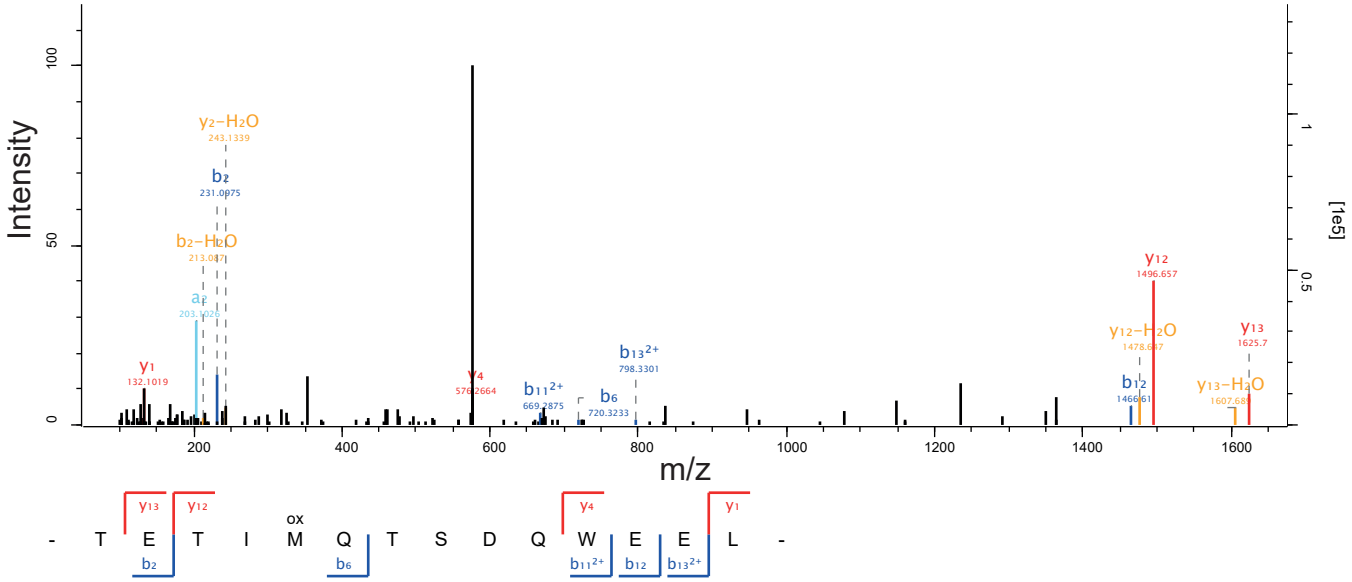
152A2 Veillonella dispar HLA-I VPLSEVATI

Endogenous peptide



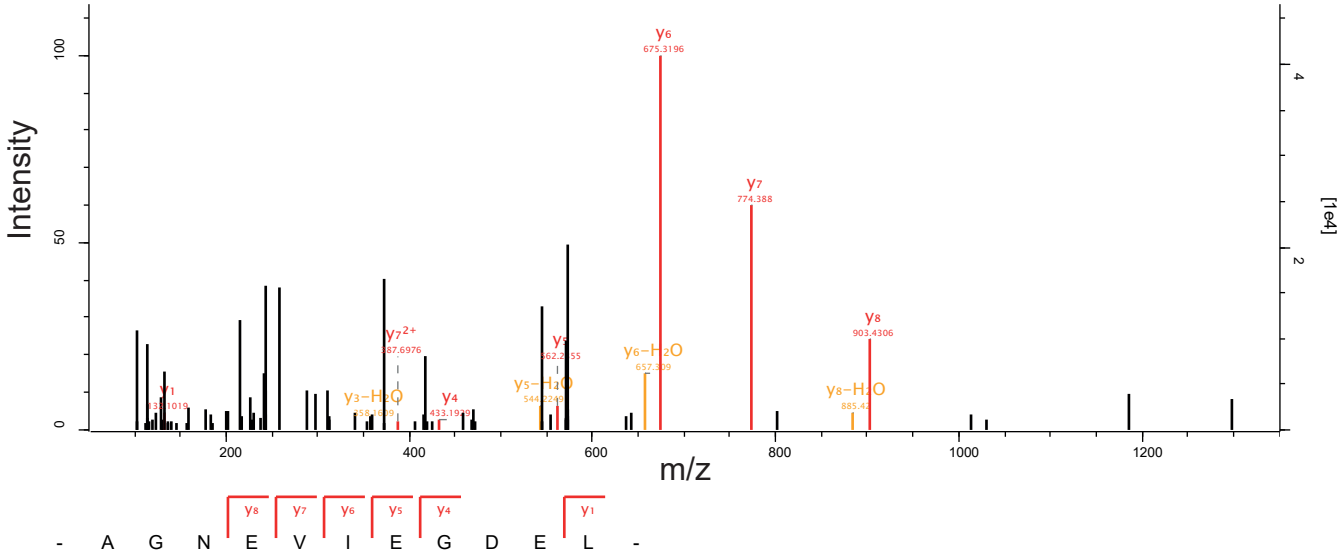
19T Enterococcus faecium HLA-II TETIMQTSDQWEEL

Endogenous peptide



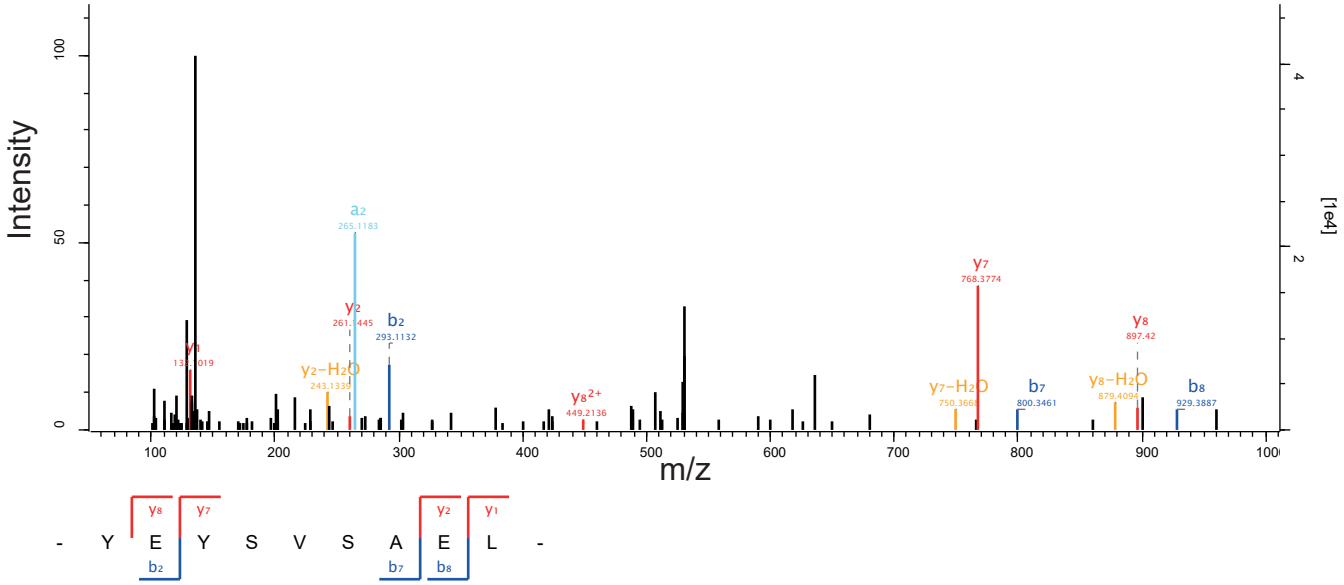
19T Staphylococcus capitis HLA-II AGNEVIEGDEL

Endogenous peptide



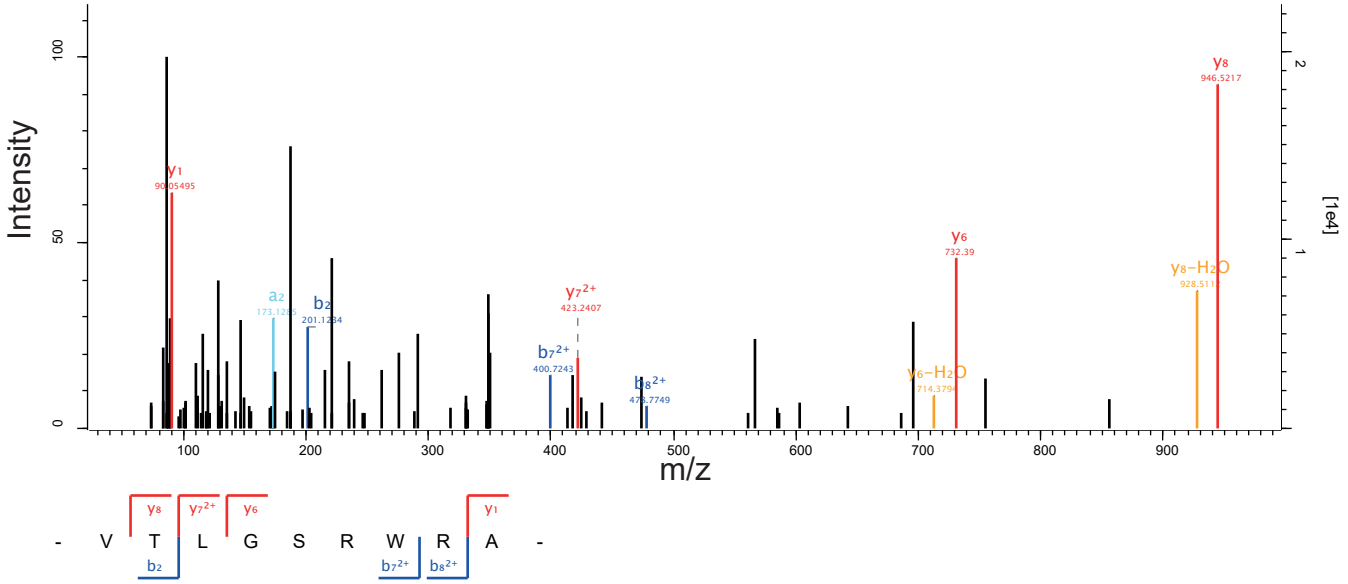
19T Staphylococcus capitis HLA-II YEYSVSAEL

Endogenous peptide



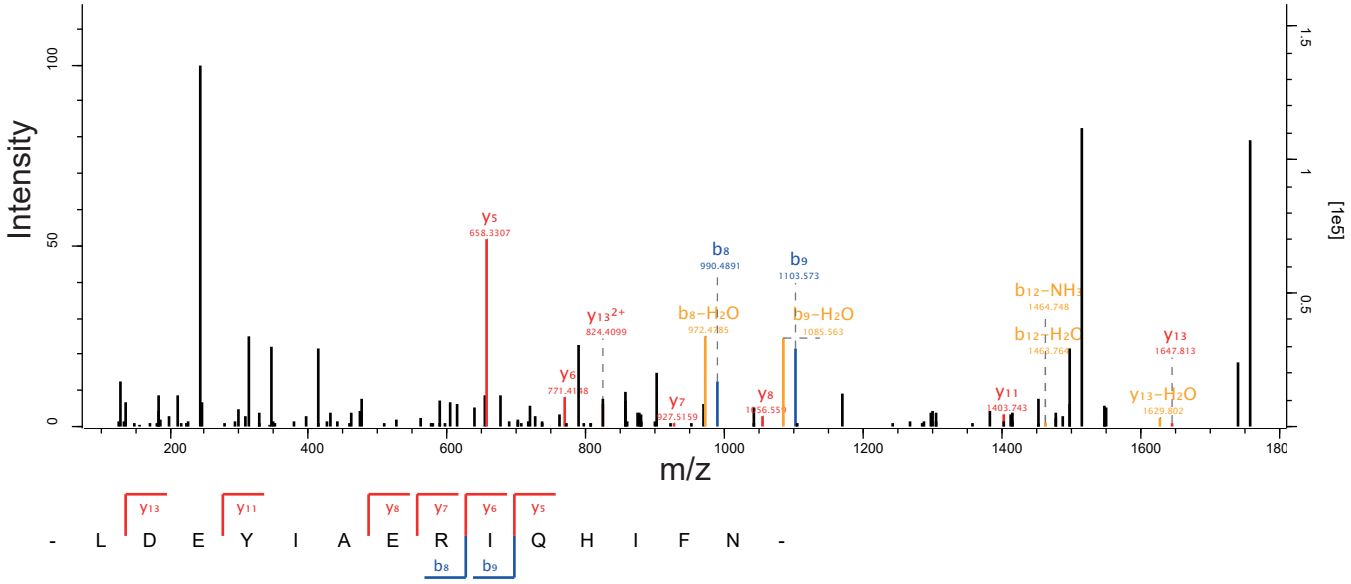
27 Actinomyces odontolyticus HLA-II VTLGSRWRA

Endogenous peptide



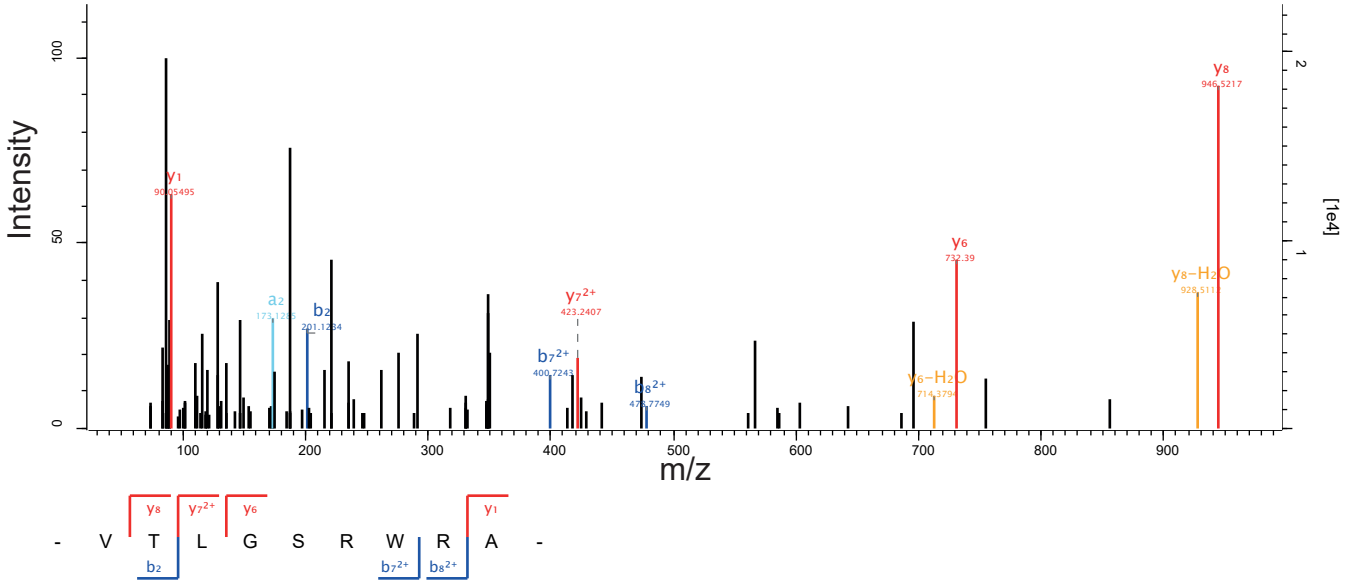
27 Fusobacterium nucleatum HLA-II LDEYIAERIQHIFN

Endogenous peptide



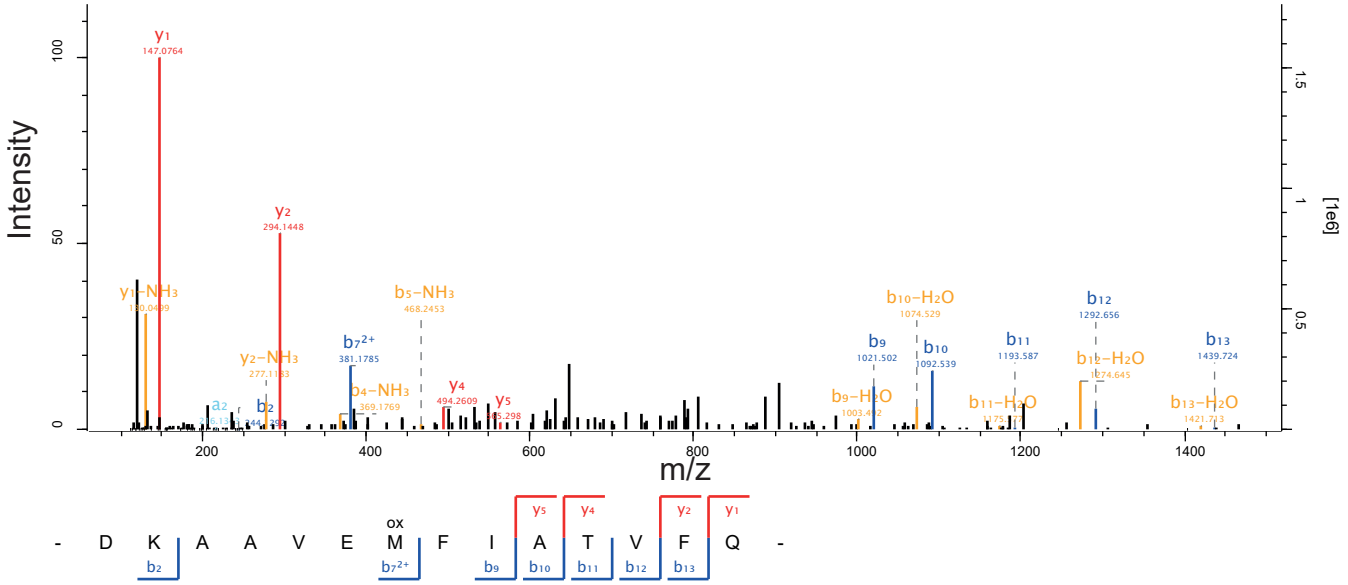
27 Schaalia odontolytica HLA-II VTLGSRWRA

Endogenous peptide



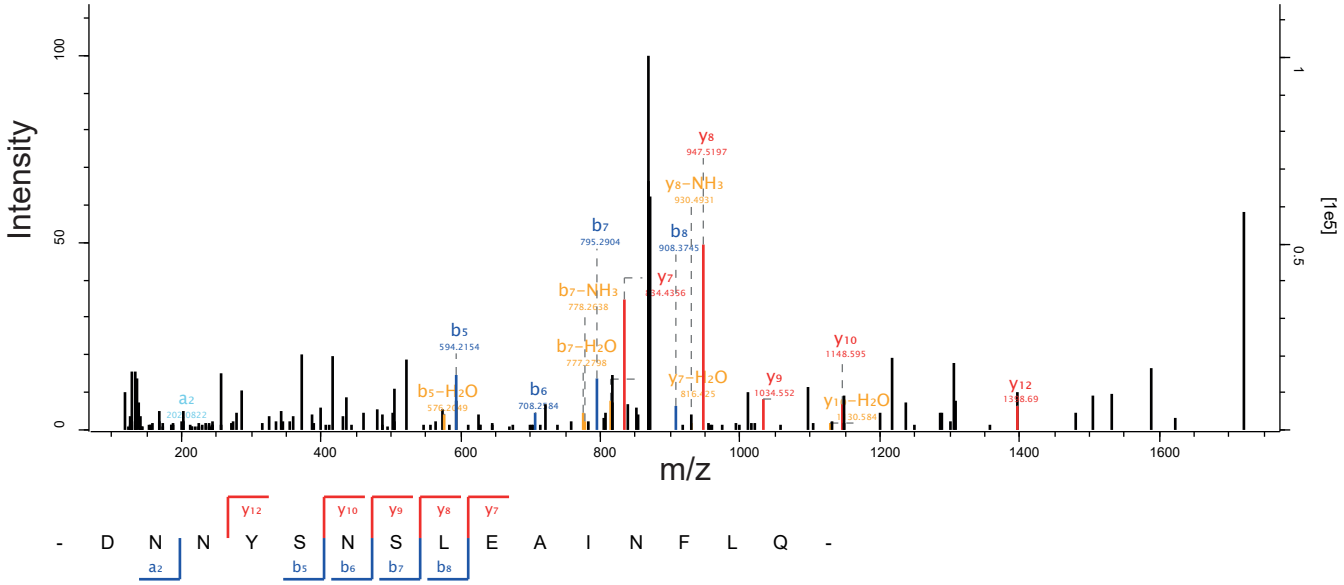
51BR-1 Eikenella corrodens HLA-II DKAAVEMFIATVFQ

Endogenous peptide



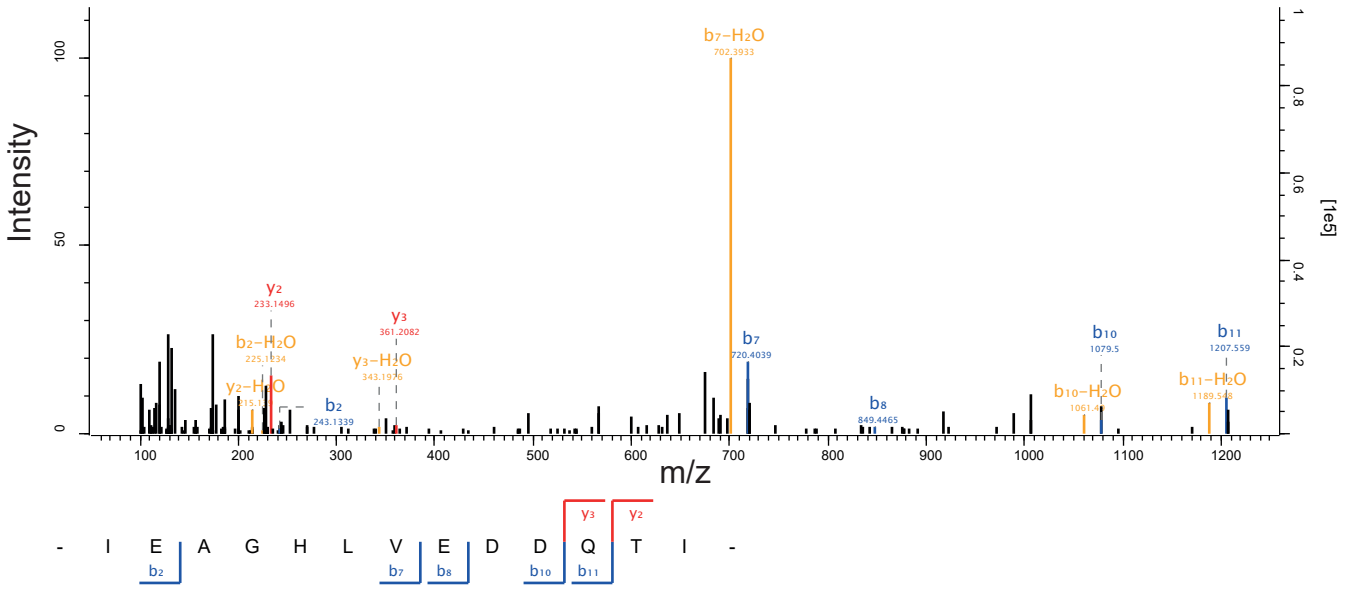
51BR-1 Fusobacterium nucleatum HLA-II DNNYSNSLEAINFLQ

Endogenous peptide



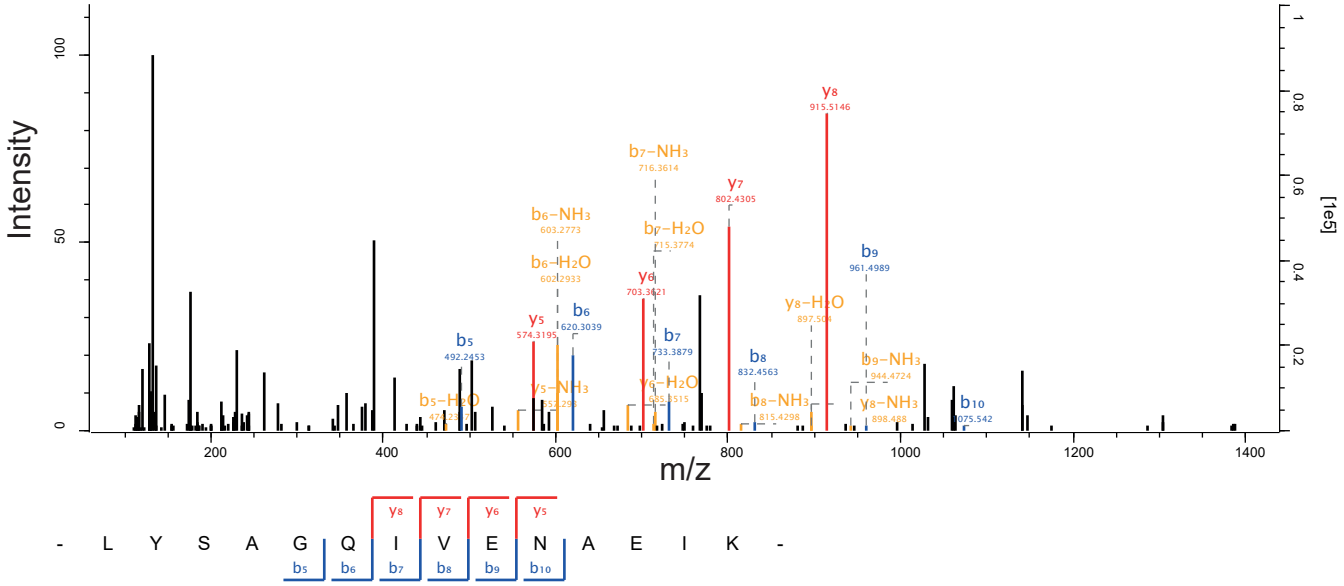
51BR-1 Staphylococcus aureus HLA-II IEAGHLVEDDQTI

Endogenous peptide



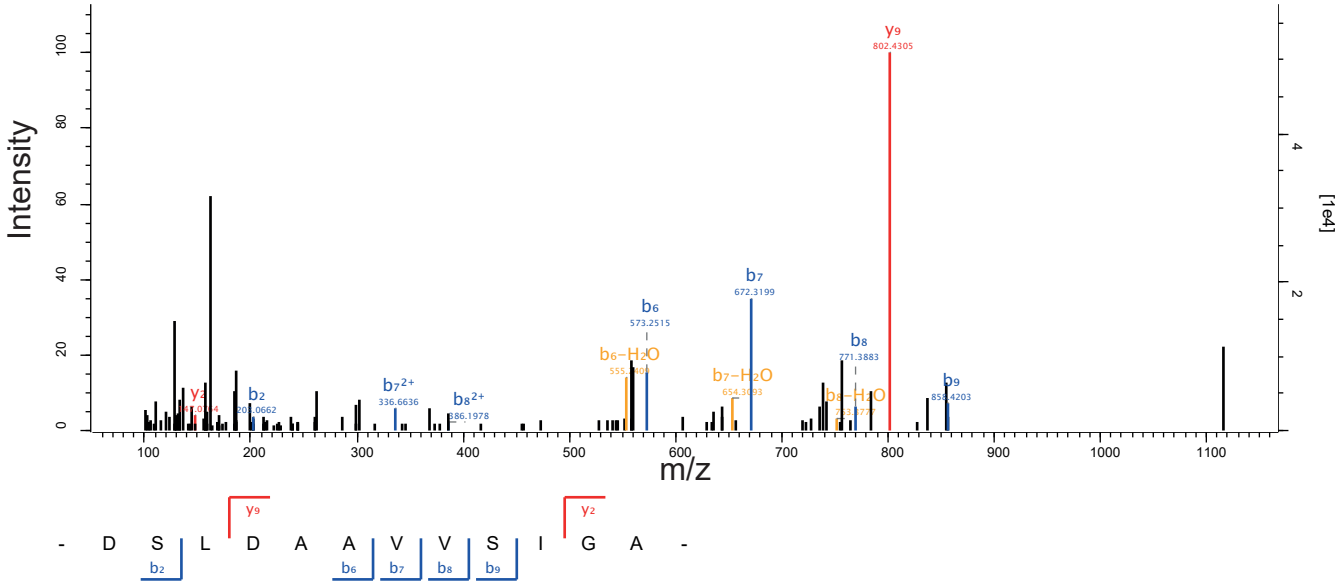
51BR-1 Staphylococcus aureus HLA-II LYSAGQIVENAEIK

Endogenous peptide



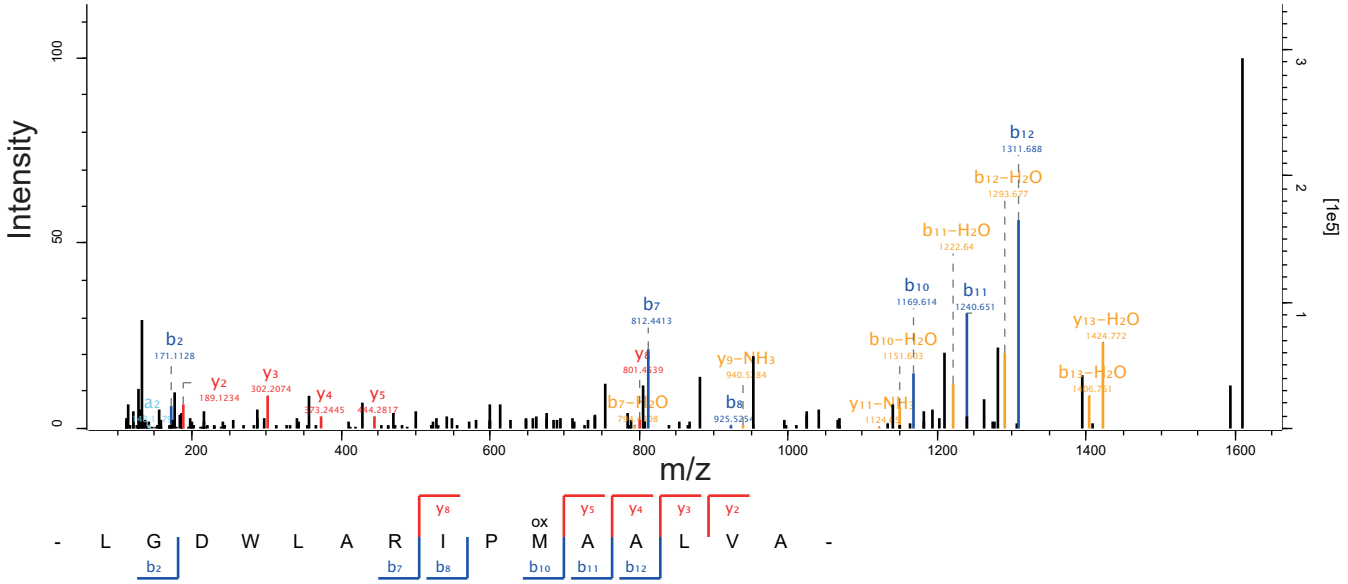
51BR-2 Brachybacterium alimentarium HLA-II DSLDAAVVSIGA

Endogenous peptide



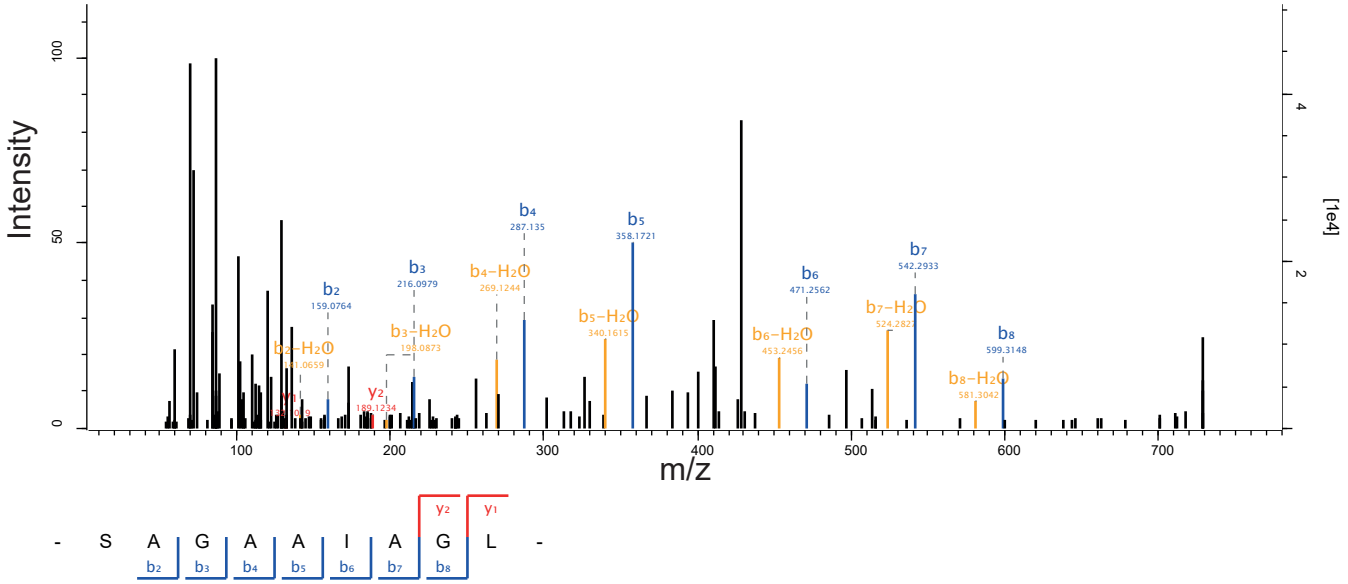
55A3 Staphylococcus capitis HLA-II LGDWLARIPMAALVA

Endogenous peptide



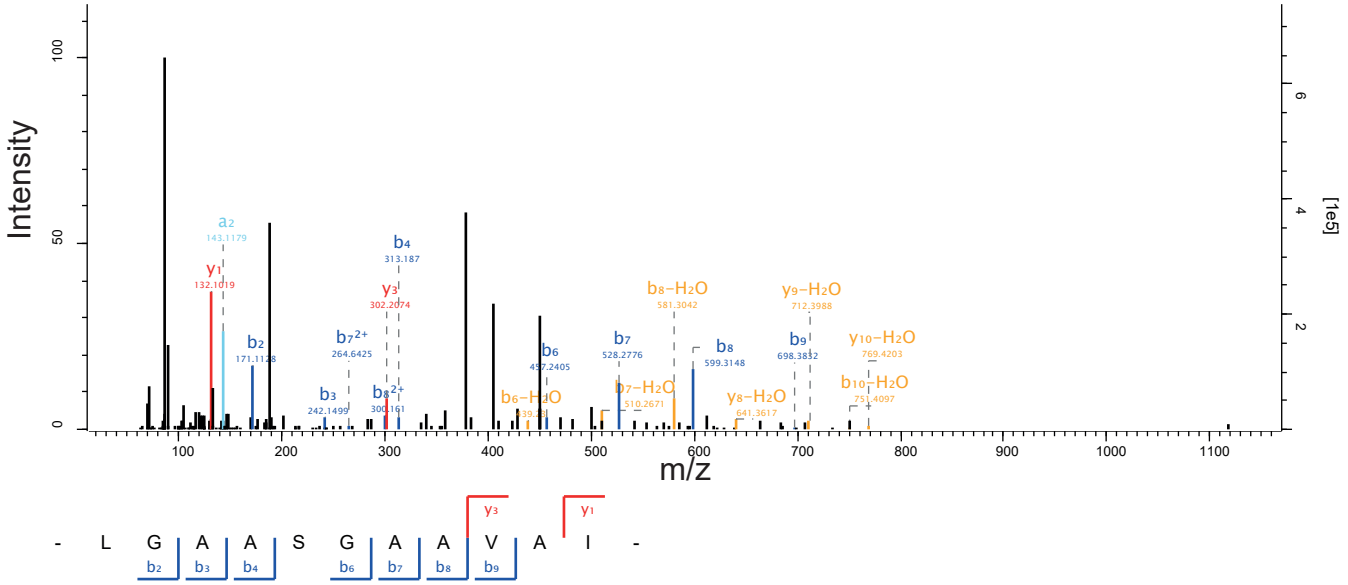
55A3 Staphylococcus capitis HLA-II SAGAAIAGL

Endogenous peptide



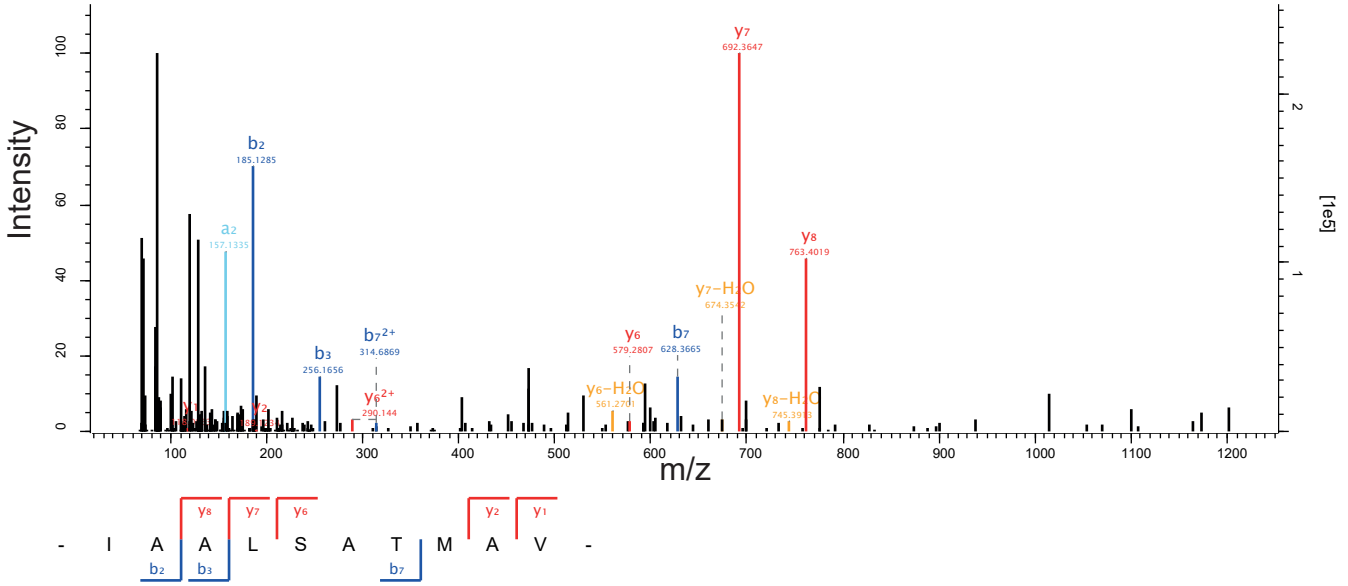
55A3 Streptococcus gordonii HLA-II LGAASGAAVAI

Endogenous peptide



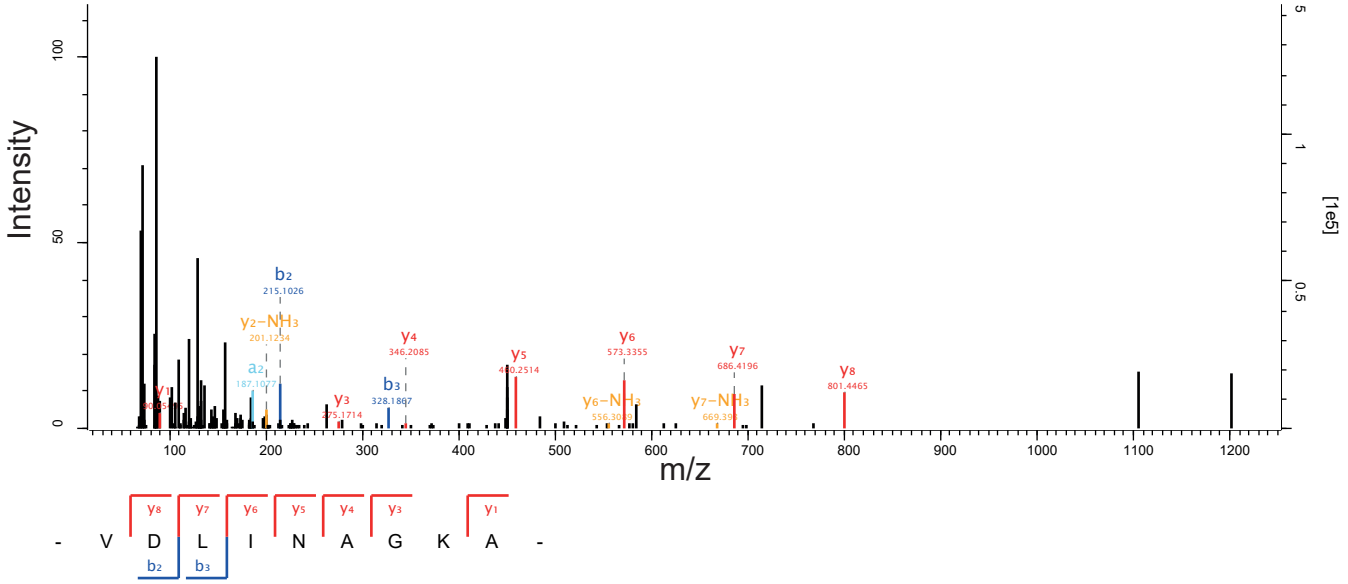
55A7-1 Enterobacter cancerogenus HLA-II IAALSATMAV

Endogenous peptide



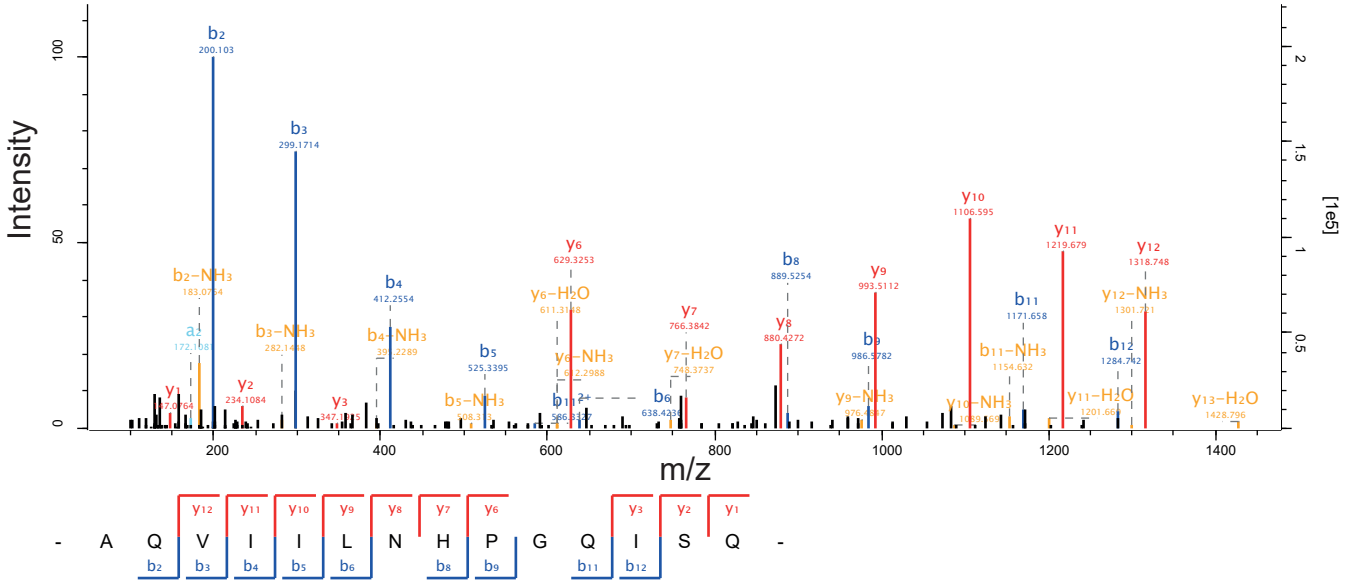
55A7-1 Enterobacter cancerogenus HLA-II VDLINAGKA

Endogenous peptide



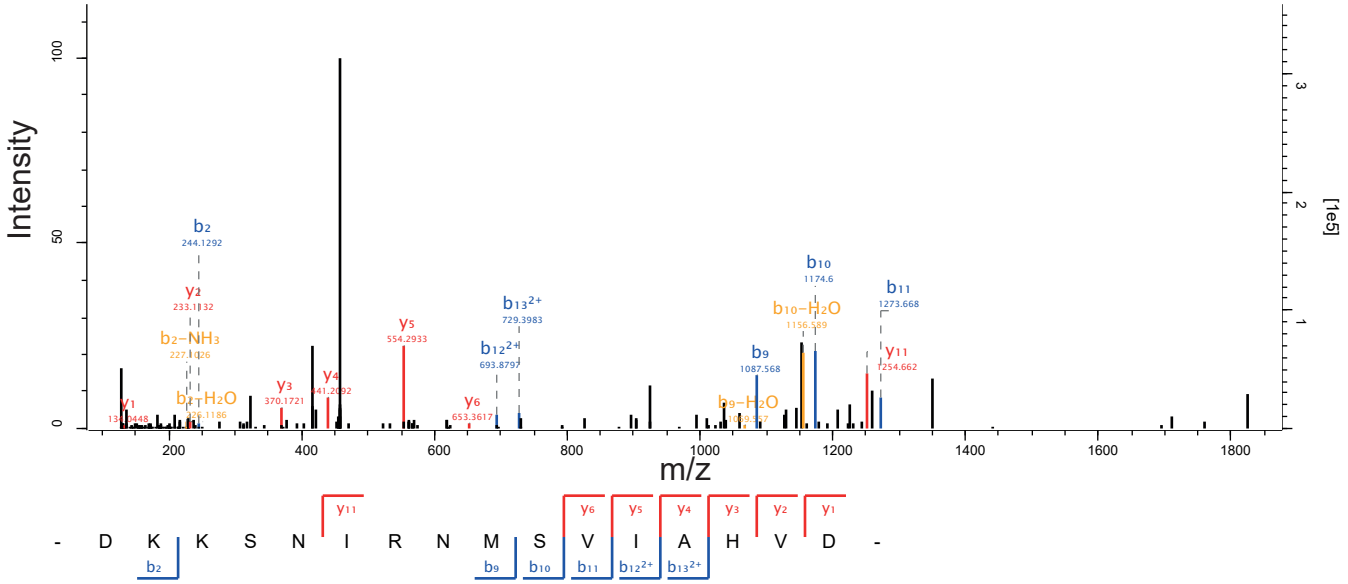
55A7-1 Enterobacter hormaechei HLA-II AQVIILNHPGQISQ

Endogenous peptide



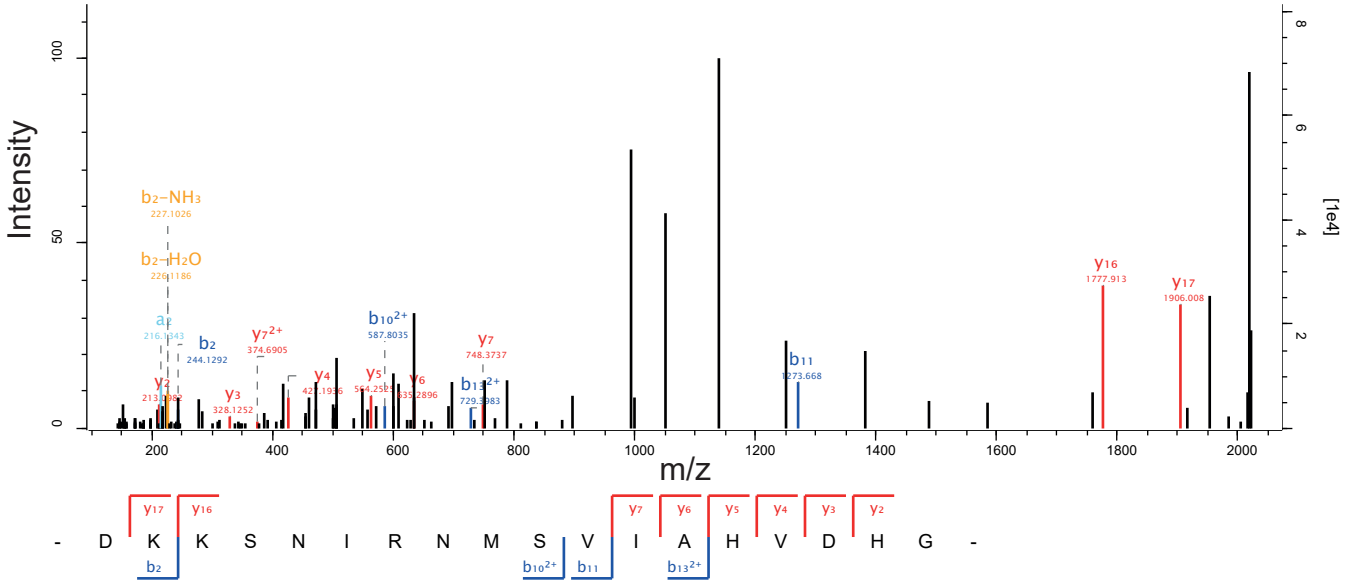
55A7-1 Enterobacter hormaechei HLA-II DKKSNIRNMSVIAHVD

Endogenous peptide



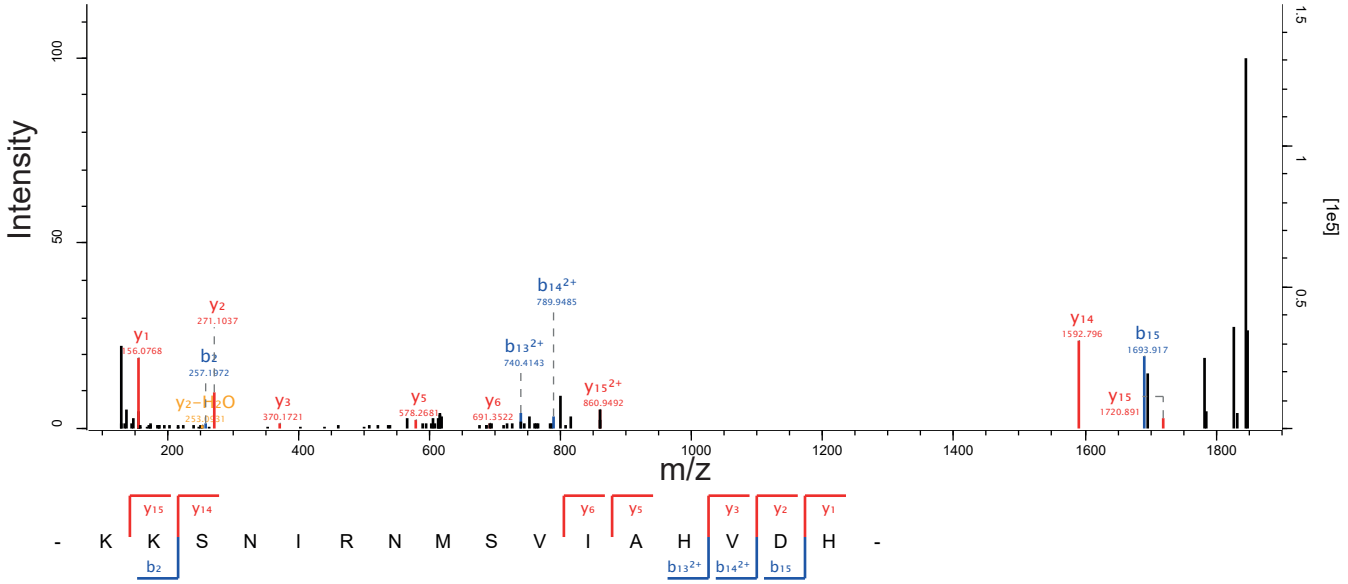
55A7-1 Enterobacter hormaechei HLA-II DKKSNIRNMSVIAHVDHG

Endogenous peptide



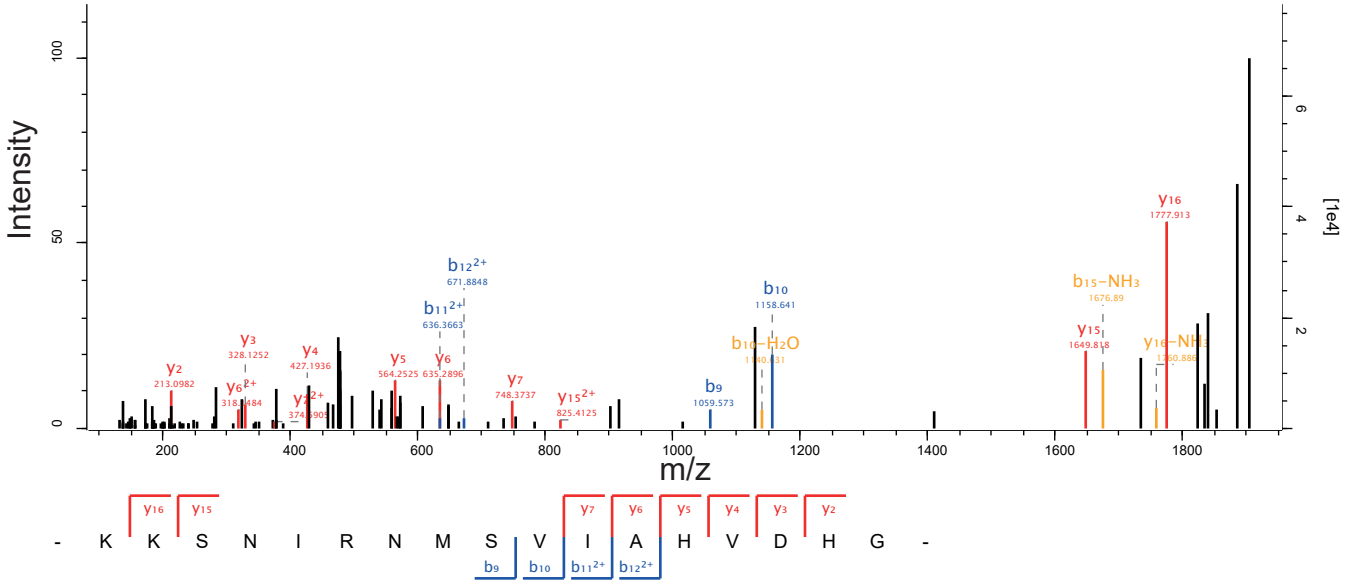
55A7-1 Enterobacter hormaechei HLA-II KKSNIRNMSVIAHVDH

Endogenous peptide



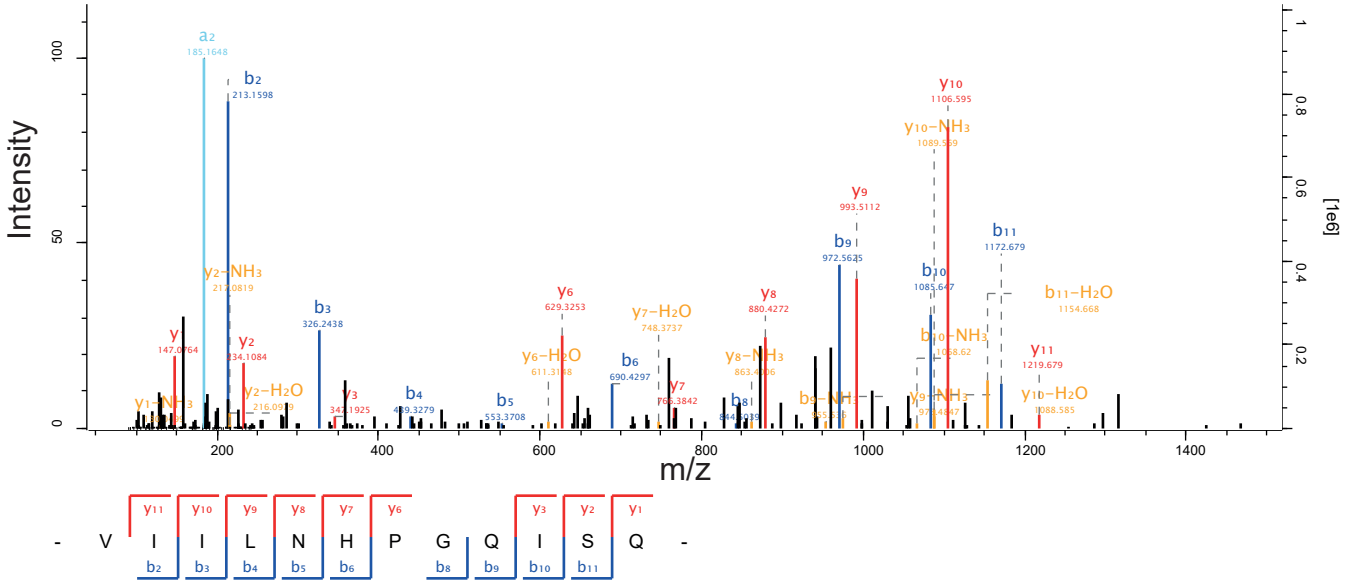
55A7-1 Enterobacter hormaechei HLA-II KKSNIRNMSVIAHVDHG

Endogenous peptide



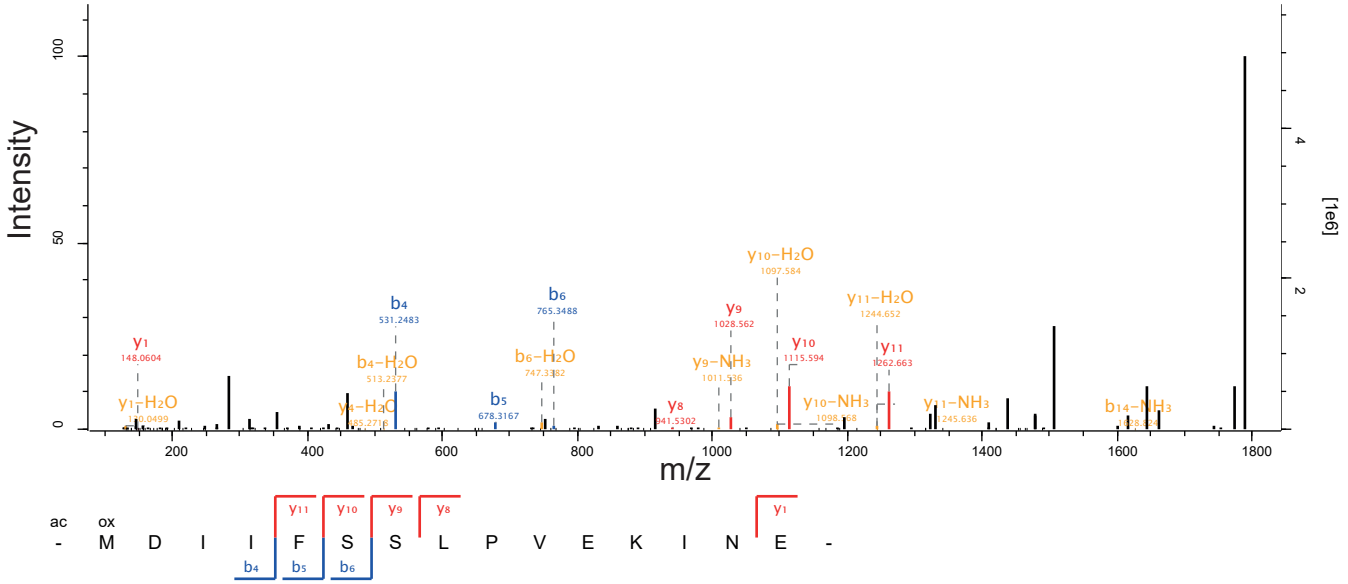
55A7-1 Enterobacter hormaechei HLA-II VIILNHPGQISQ

Endogenous peptide



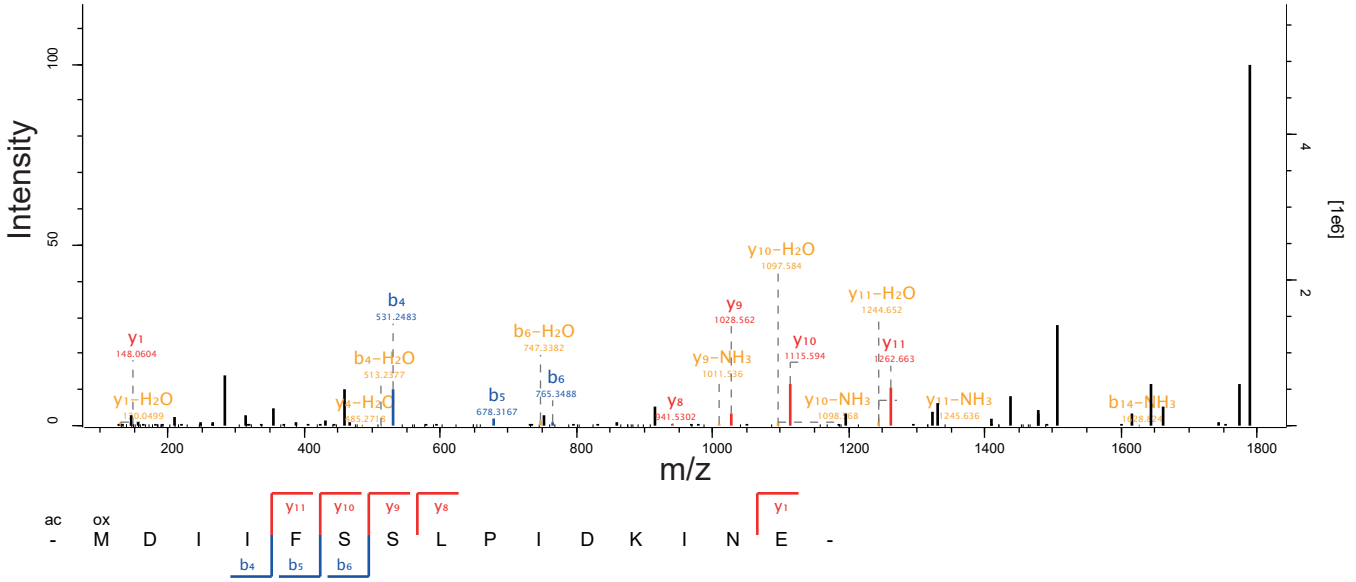
55A7-1 Veillonella dispar HLA-II MDIIFSSLPVEKINE

Endogenous peptide



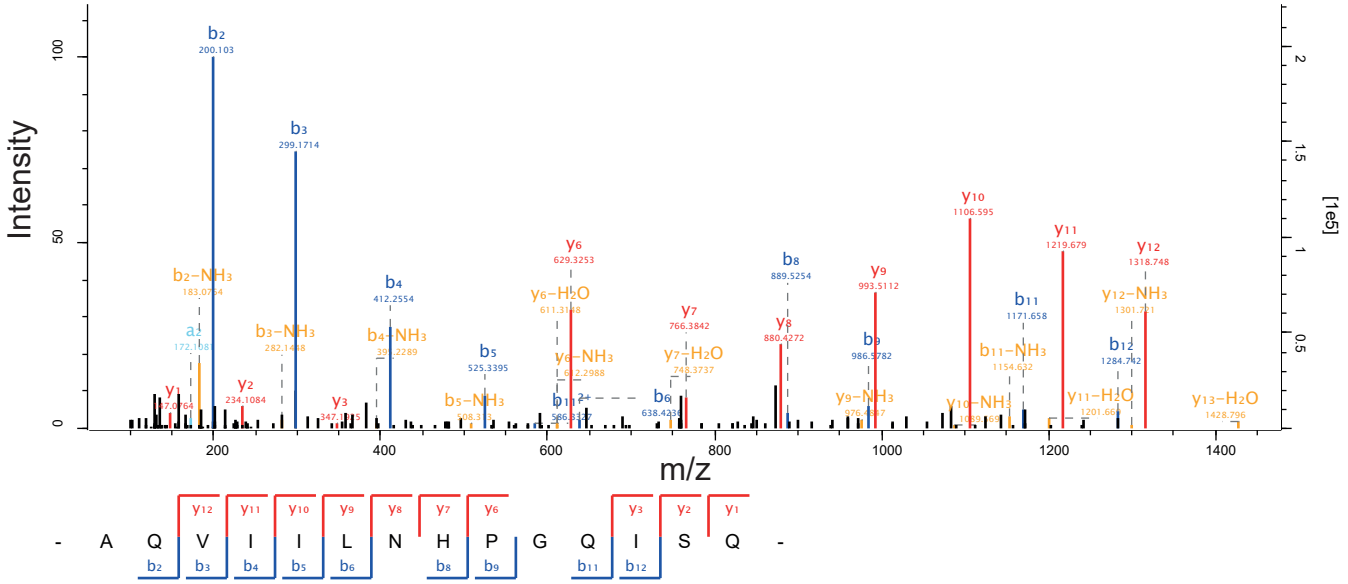
55A7-1 Veillonella parvula HLA-II MDIIFSSLPIDKINE

Endogenous peptide



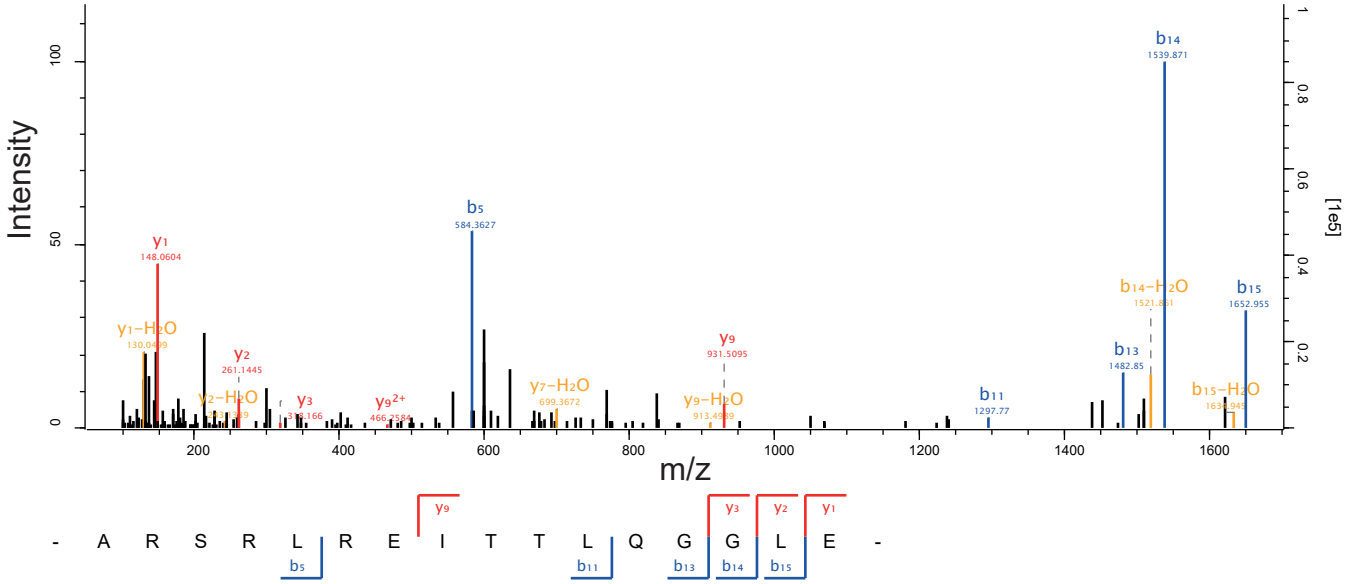
55A7-2 Enterobacter hormaechei HLA-II AQVIILNHPGQISQ

Endogenous peptide



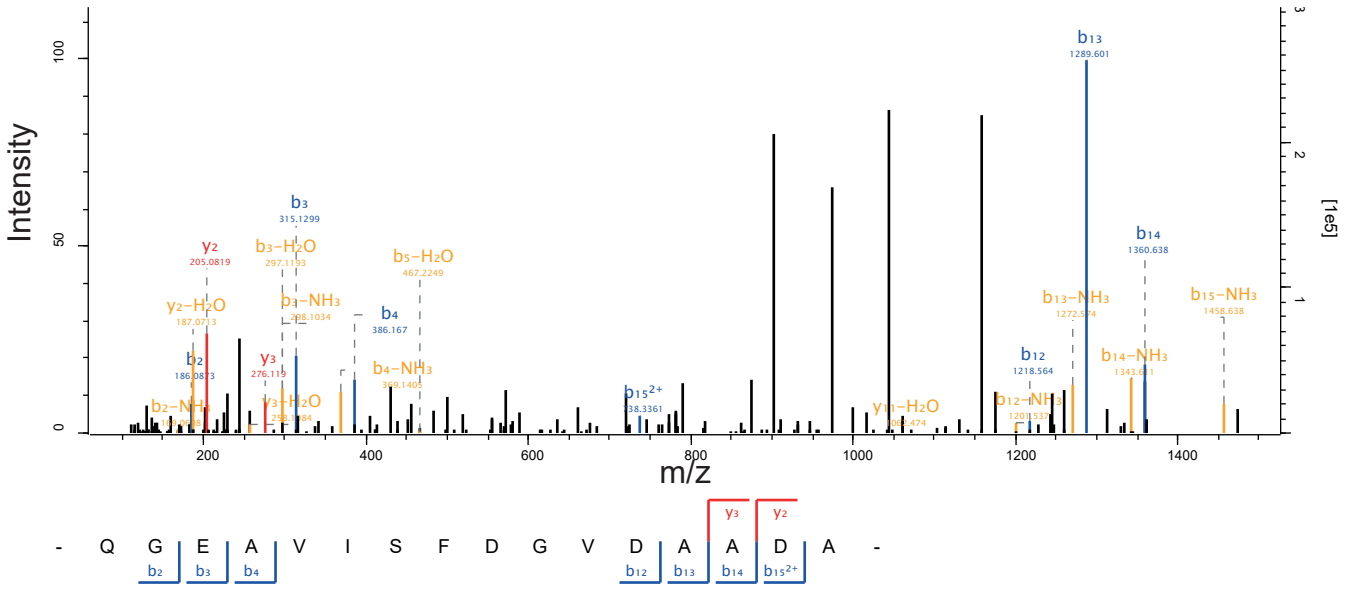
55A7-2 Enterobacter hormaechei HLA-II ARSRLREITTLQGGLE

Endogenous peptide



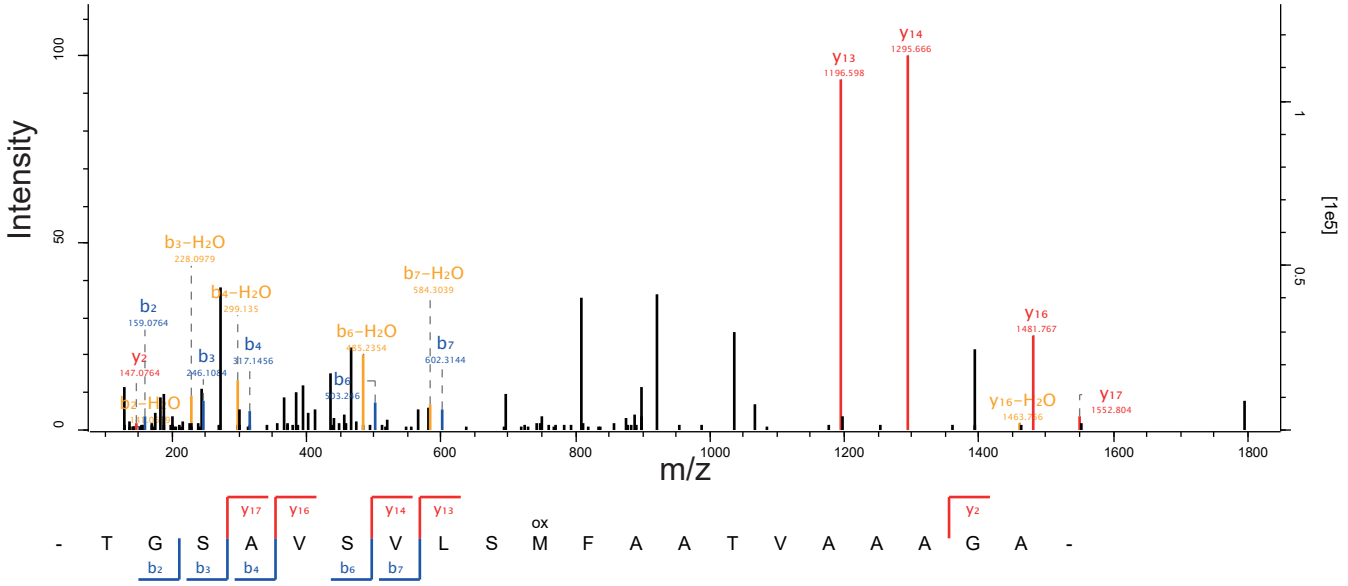
55B3 Gardnerella vaginalis HLA-II QGEAVISFDGVDAADA

Endogenous peptide



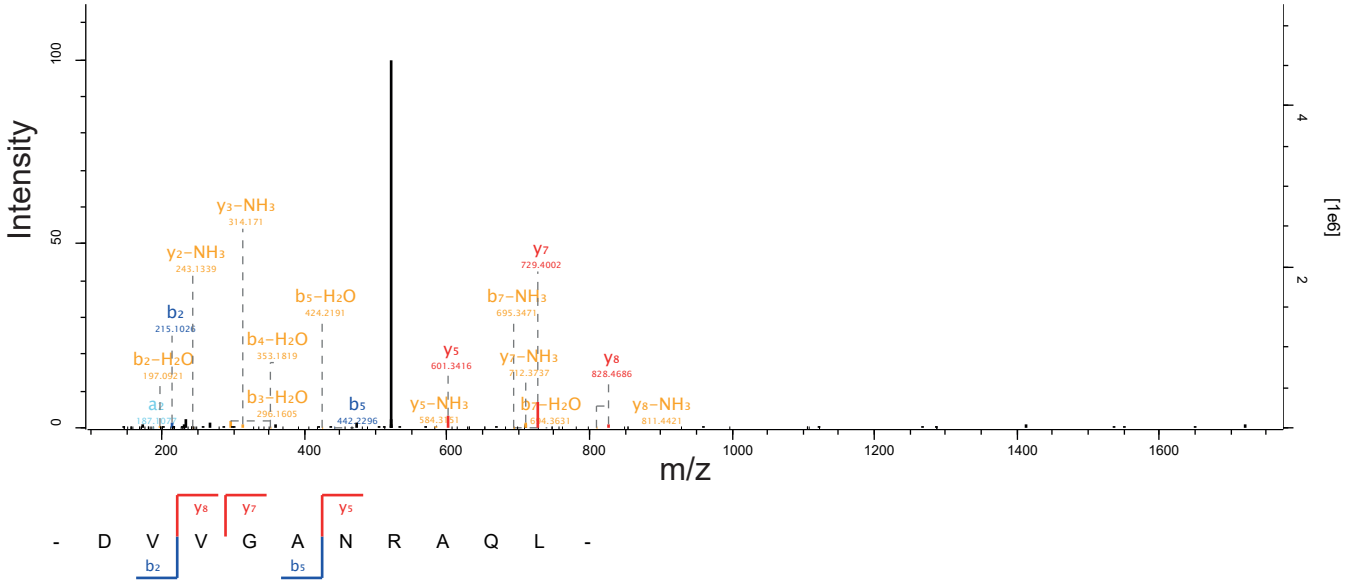
55B3 Gardnerella vaginalis HLA-II TGSAVSVLSMFAATVAAAGA

Endogenous peptide



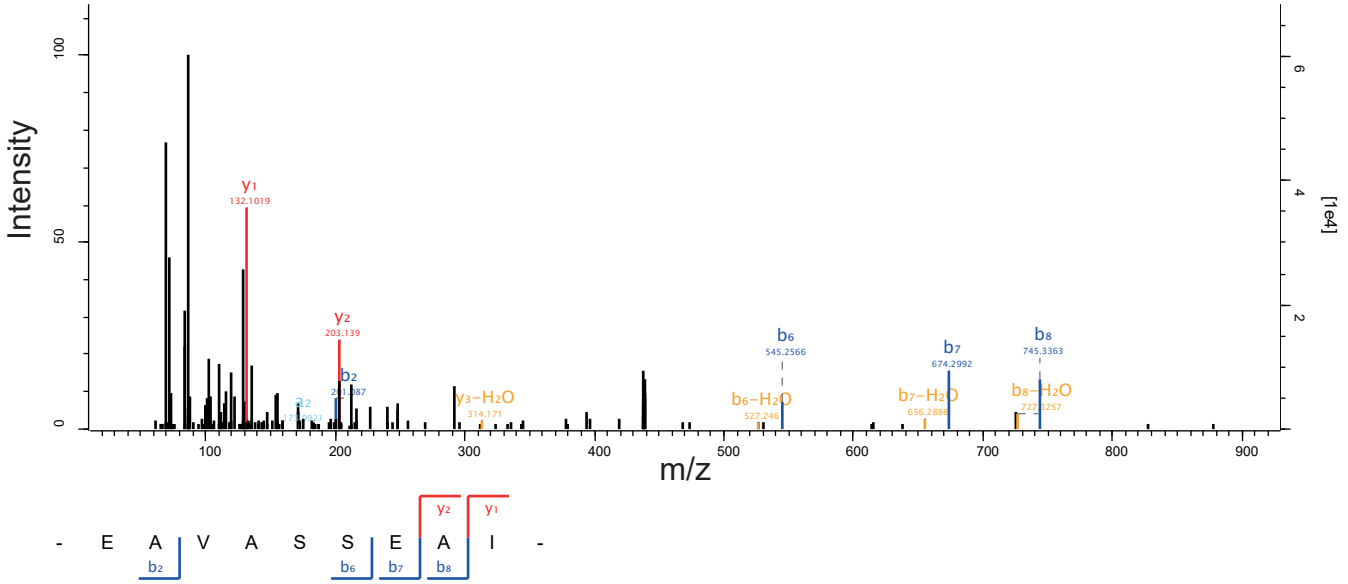
55B3 Sphingomonas dokdonensis HLA-II DVVGANRAQL

Endogenous peptide



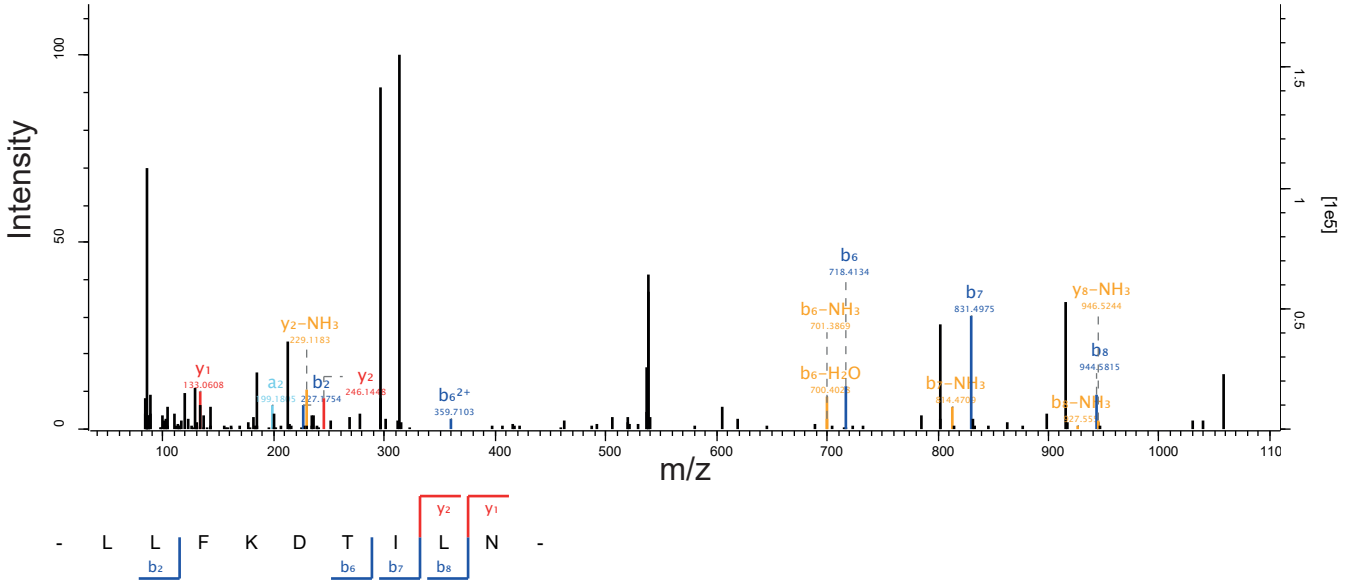
55B3 Sphingomonas melonis HLA-II EAVASSEAI

Endogenous peptide



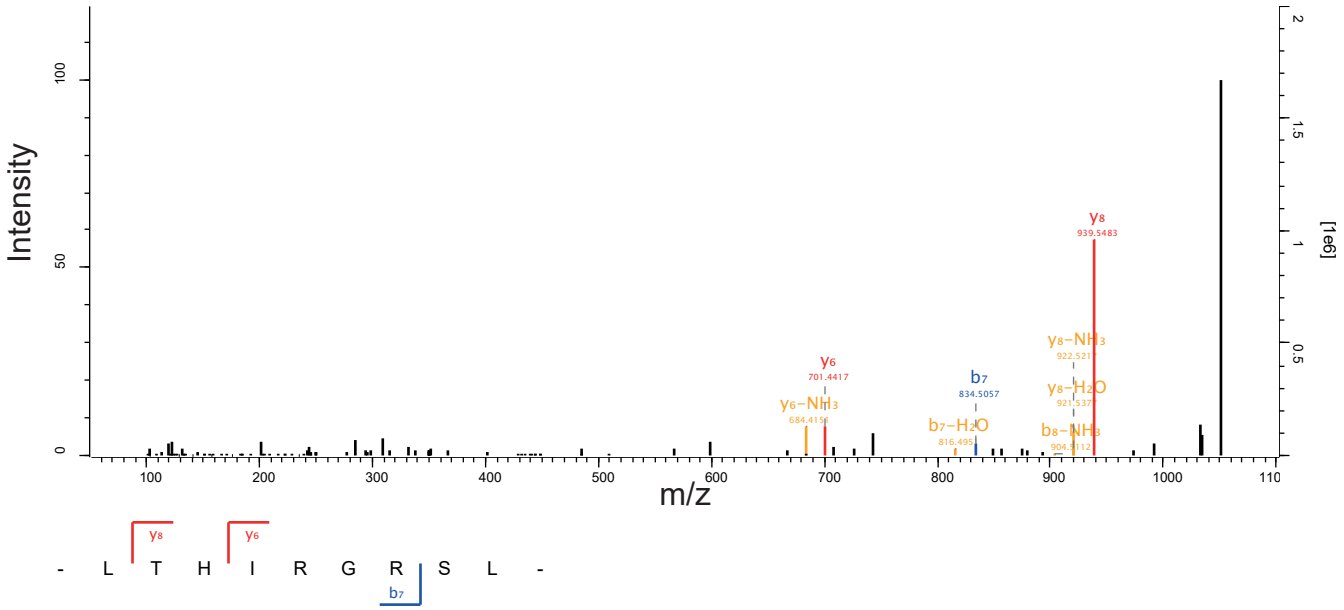
55B3 Sphingomonas melonis HLA-II LLFKDTILN

Endogenous peptide



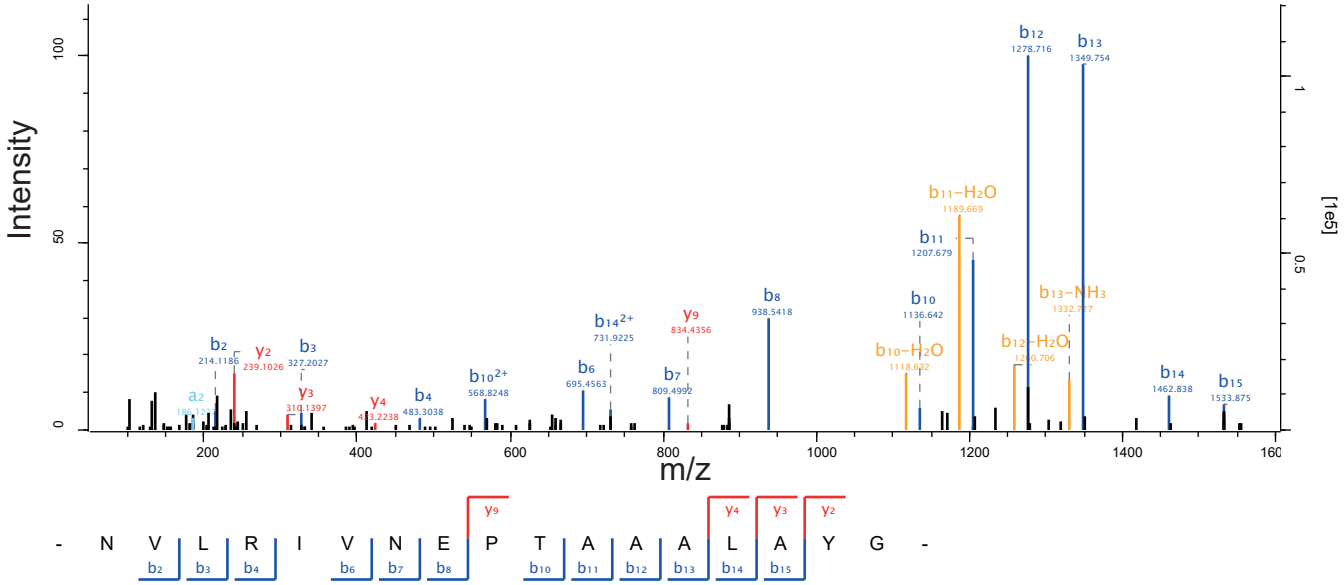
70.1 Staphylococcus capitis HLA-II LTHIRGRSL

Endogenous peptide



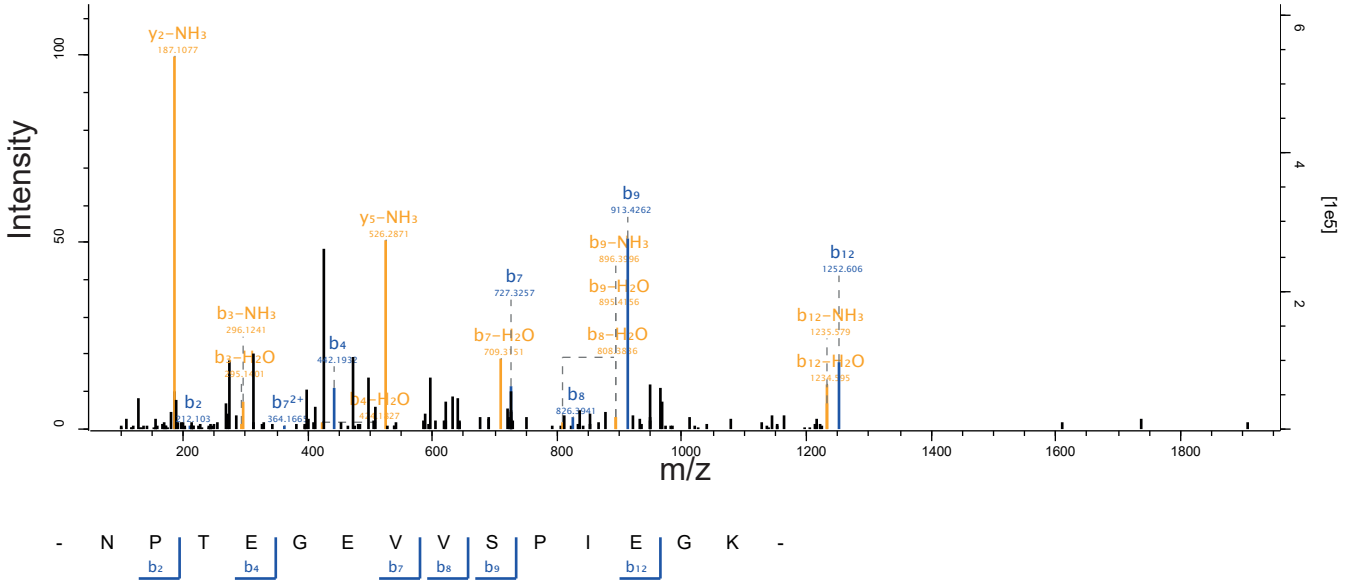
86B Corynebacterium humireducens HLA-II NVLRIVNEPTAAALAYG

Endogenous peptide



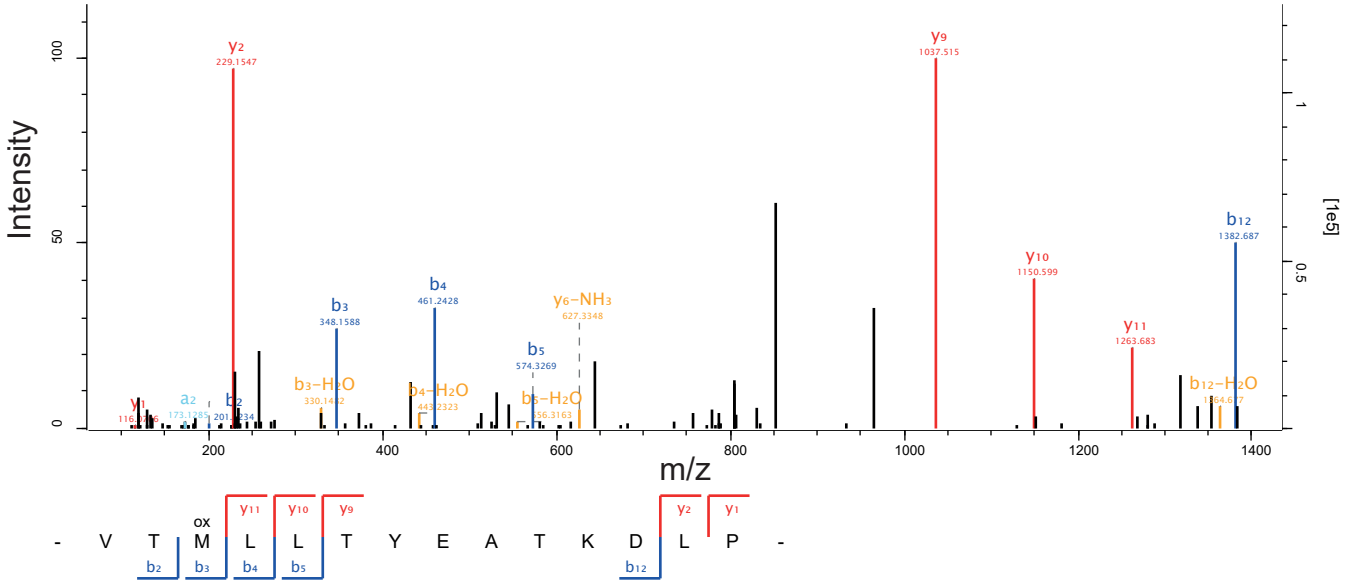
86B Staphylococcus lugdunensis HLA-II NPTEGEVVSPIEGK

Endogenous peptide



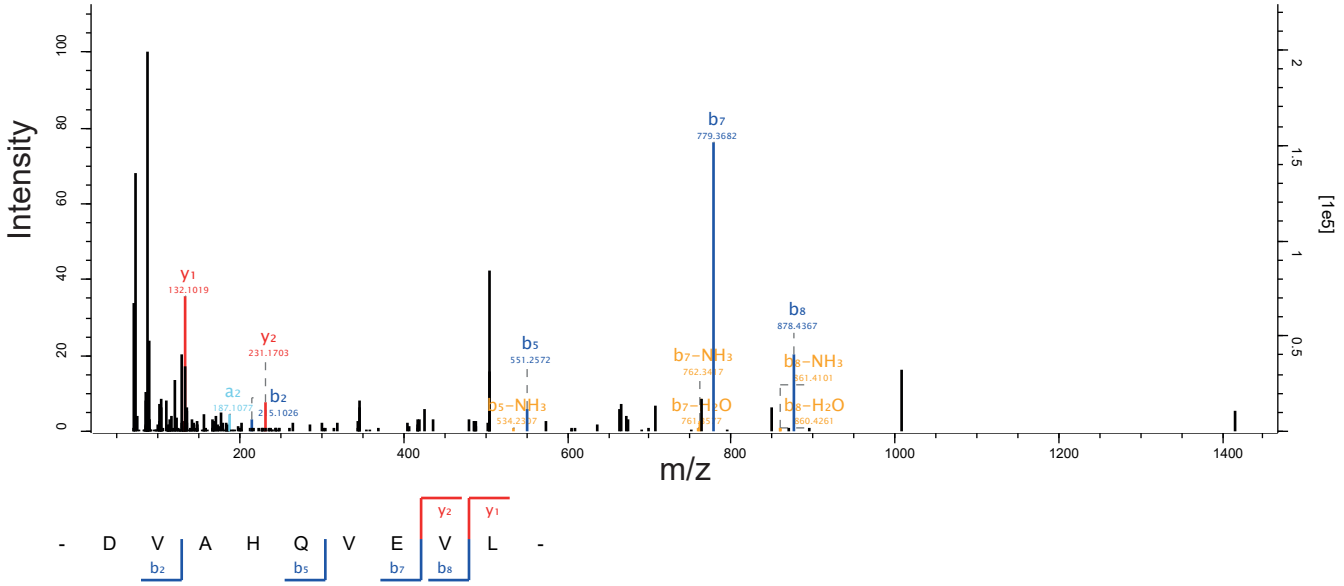
86B2 Acinetobacter cumulans HLA-II VTMLLTYEATKDLP

Endogenous peptide



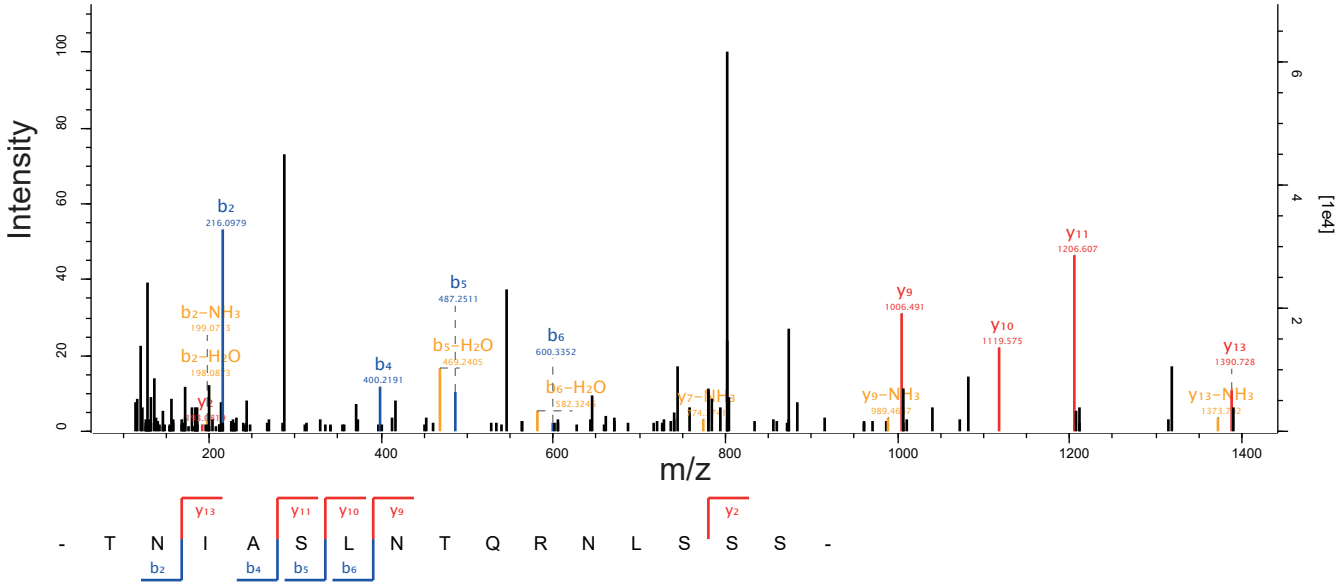
86B2 Staphylococcus capitis HLA-II DVAHQVEVL

Endogenous peptide



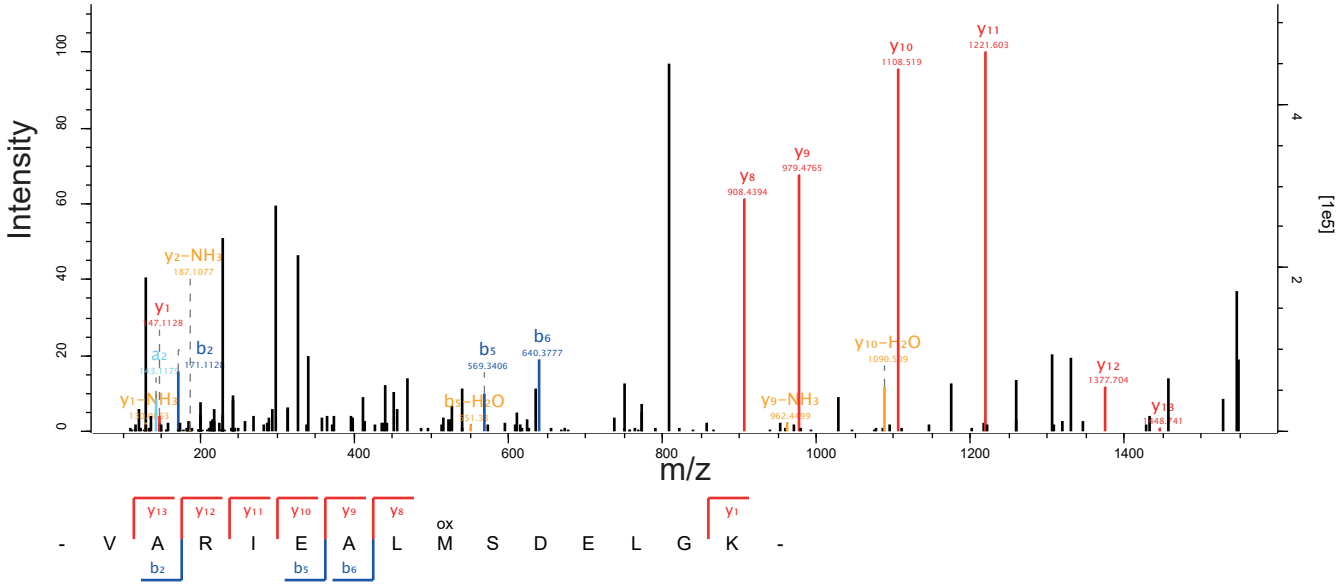
86B2 Staphylococcus capitis HLA-II TNIASLNTQRNLSSS

Endogenous peptide



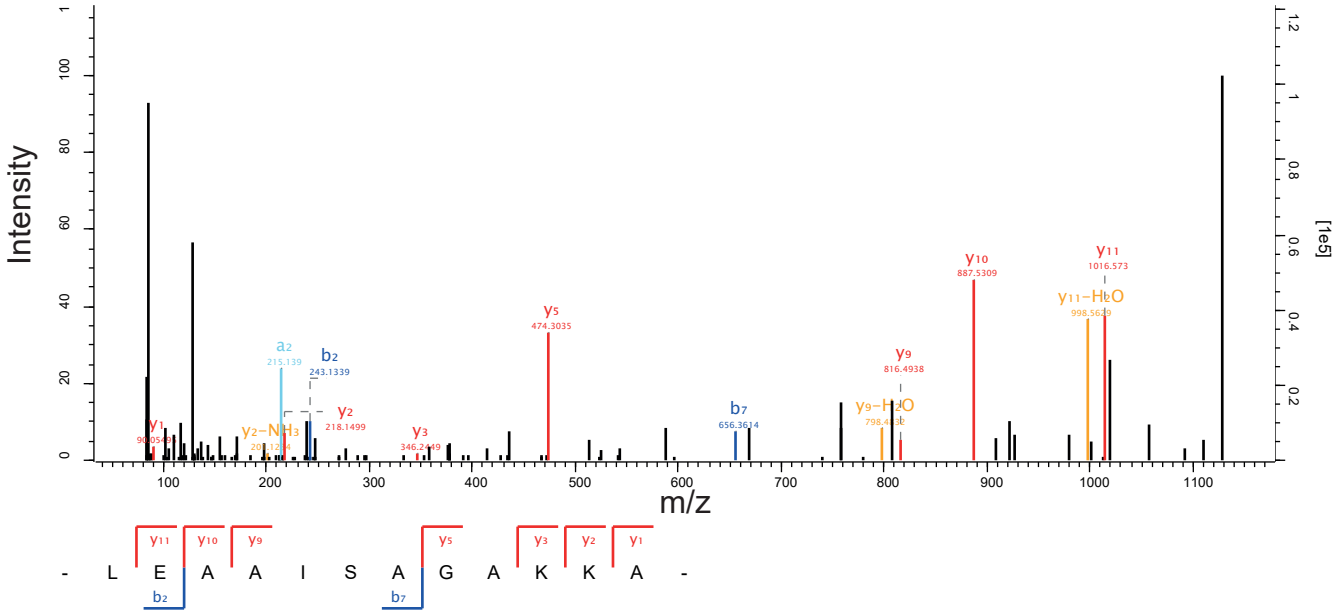
86B2 Staphylococcus capitis HLA-II VARIEALMSDELGK

Endogenous peptide



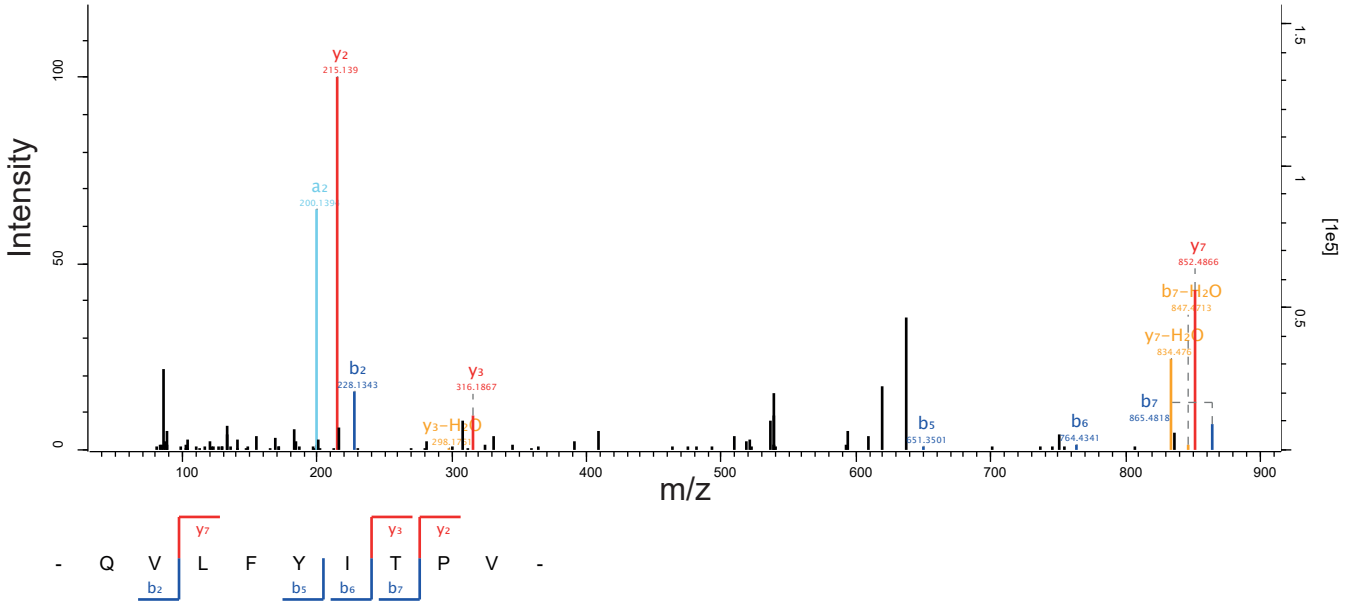
92B3 Fusobacterium nucleatum HLA-II LEAAISAGAKKA

Endogenous peptide



92B3 Fusobacterium nucleatum HLA-II QVLFYITPV

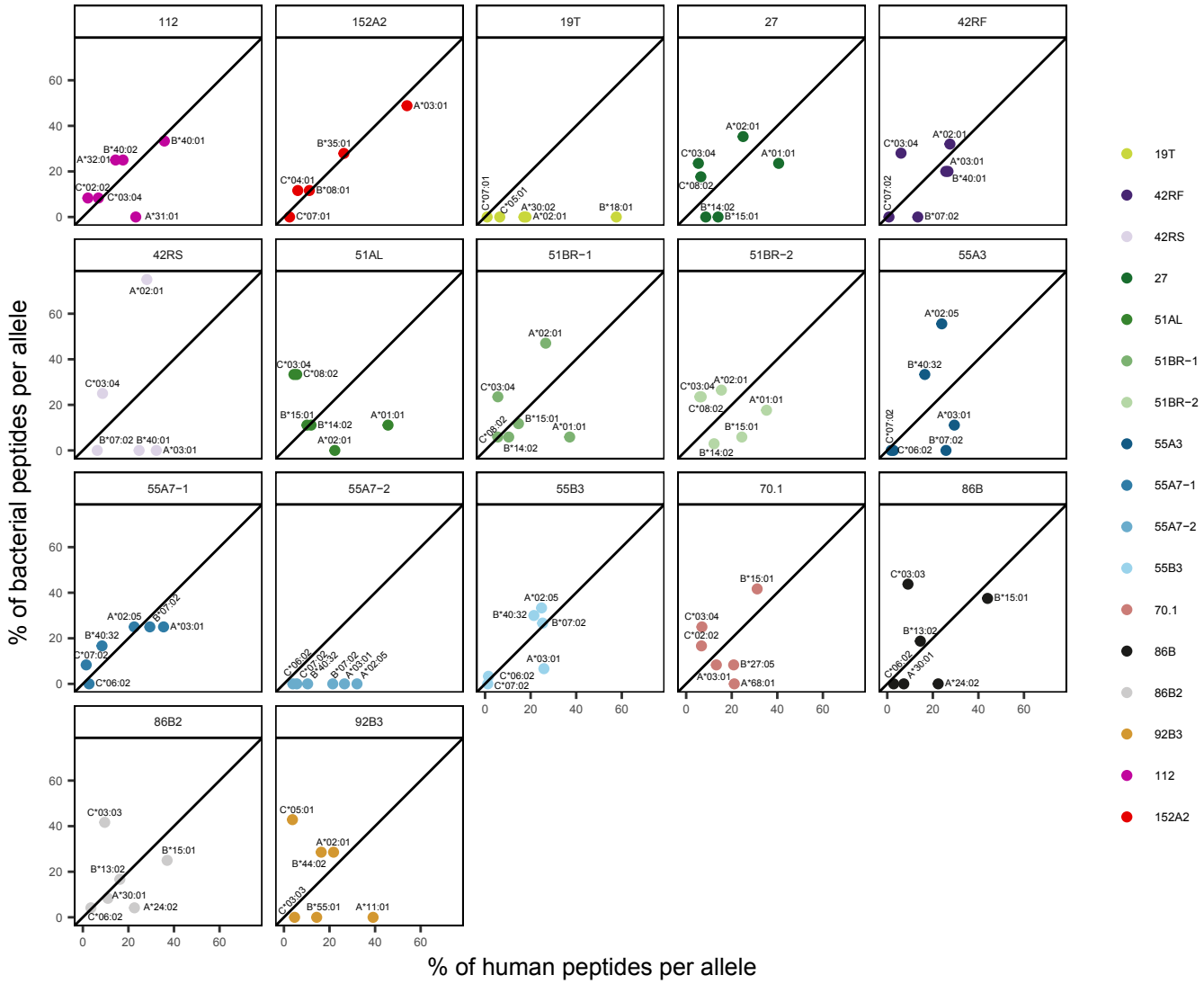
Endogenous peptide

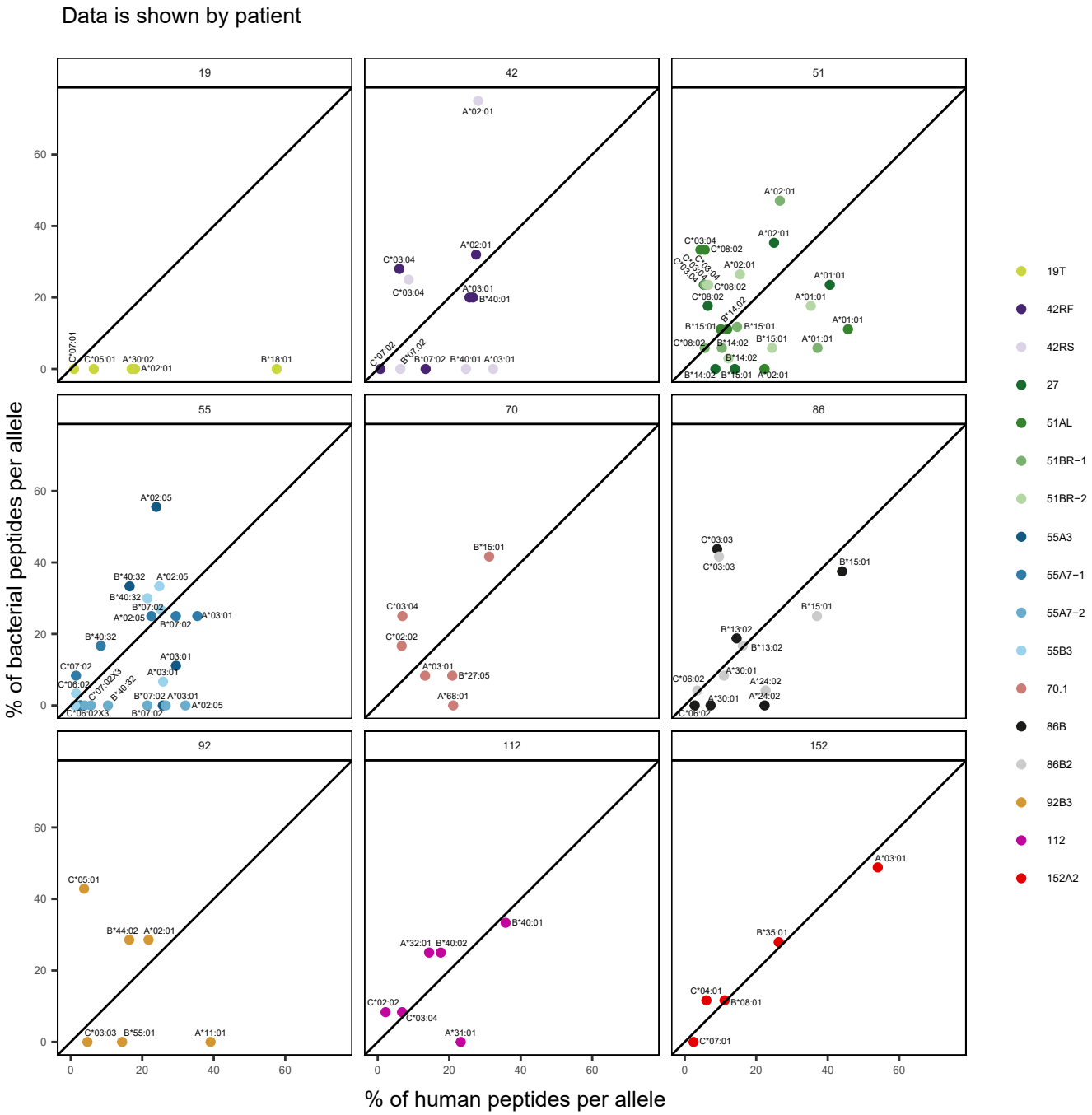


Supplementary Figure 3: Tandem mass spectra of the bacteria peptides identified in melanoma tumors and their corresponding synthetic peptides. Fregmantaion images of endogenous and synthetic peptides were extracted from MaxQuant. A head to tail figure was plotted to compare between the fragmentation of the endogenous and synthetic peptides. The Pearson correlation and the dot product score are indicated for each comparison.

Supplementary Figure 4

Data is shown by metastasis

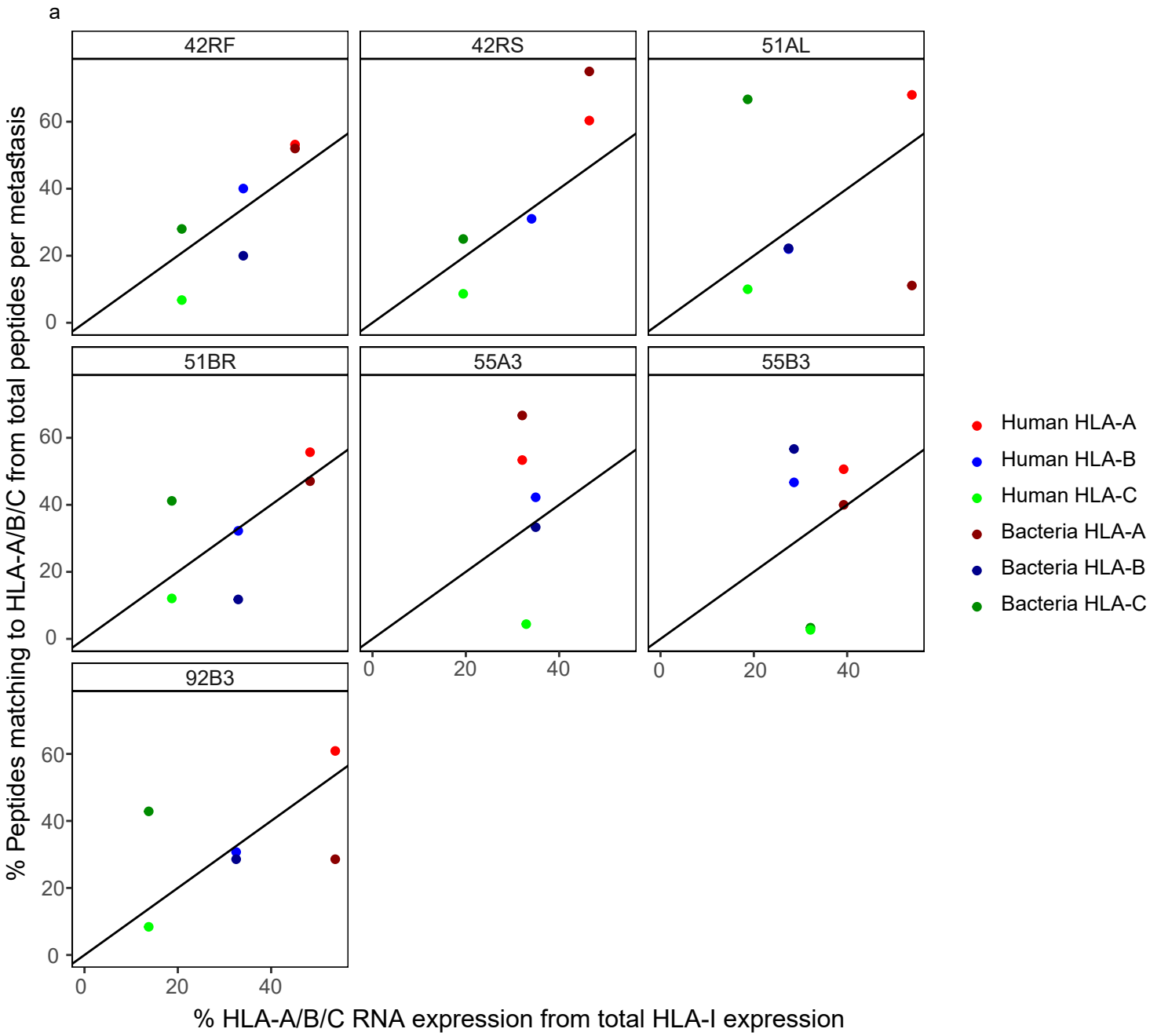




Supplementary Figure 4: Frequency of the HLA alleles among bacterial and human peptides. The percentages of the bacterial (y-axis) and human (x-axis) peptides that match each of the HLA alleles of the patients. Data is shown for each metastasis separately or for all the metastases of each patient together.

Supplementary Figure 5

a



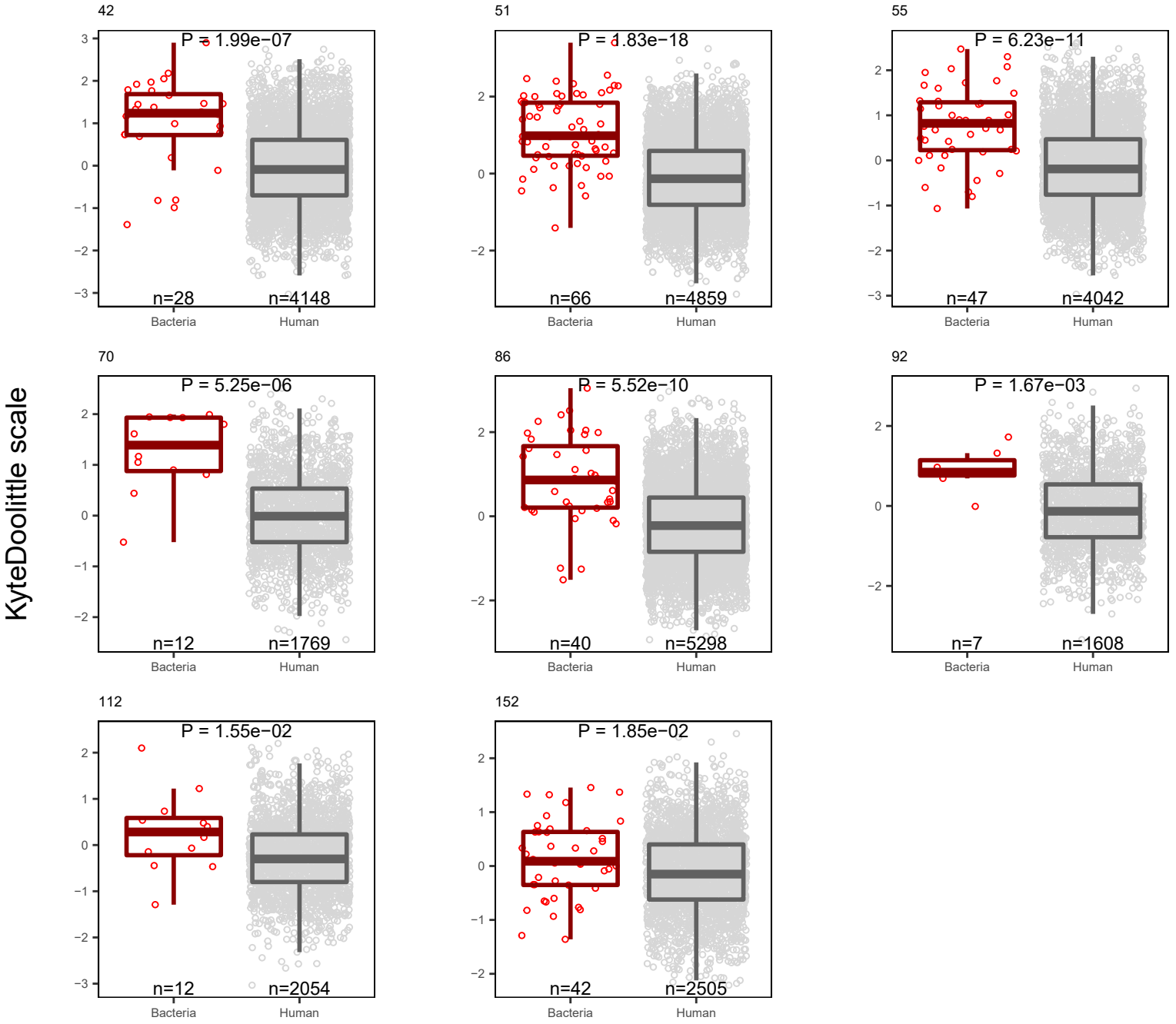
b

Gene	42RF	42RS	51AL	51BR	55A3	55B3	92B3
HLA-A	2691	2600	2781	3294	2268	689	1476
HLA-B	2030	1907	1417	2245	2471	502	893
HLA-C	1244	1088	965	1275	2328	564	378

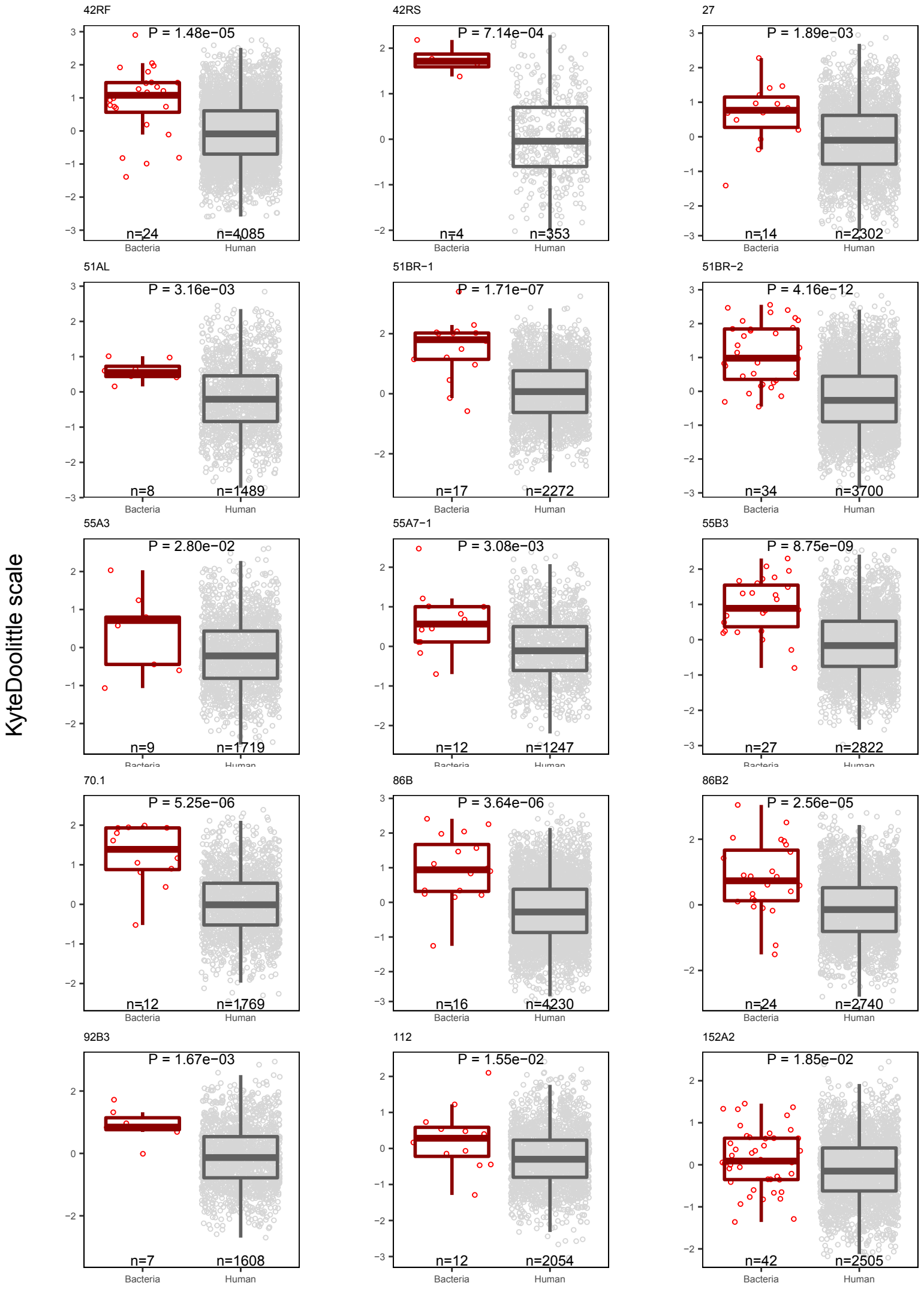
Supplementary Figure 5: Correlation between the percentage of identified peptides per HLA-I allele to the respective HLA-I's RNA expression level. a, The percentage of human and bacterial peptides matching HLA-A/B/C from the sum of peptides as calculated for each metastasis. Each peptide was assigned an allele with the best binding prediction score (%rank). This was compared to the relative RNA expression of HLA-A/B/C (i.e., percentage of total HLA-I expression). HLA-A/B/C were marked red, blue and green, respectively; bacteria are marked by a darker shade, human by a lighter shade. b, The TPM values of HLA-A/B/C expression by RNA sequencing.

Supplementary Figure 6

Data is shown by patient



Data is shown by metastasis

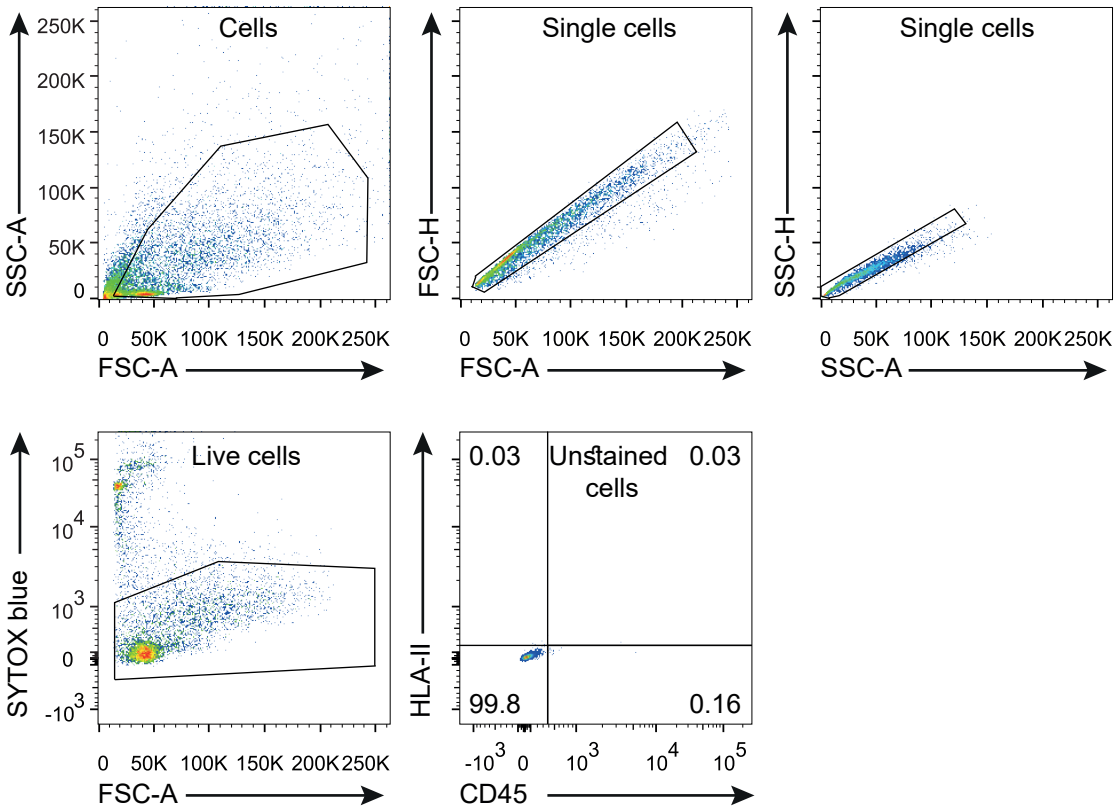


Supplementary Figure 6: Hydrophobicity comparison between bacterial and human peptides presented on HLA-I melanoma tumors. Bacterial peptides from the same patient (combining all the patients' metastases) or from each metastasis were more hydrophobic than human peptides (the statistical analysis used is found in the Material and Methods under "Peptide hydrophobicity"). In box plots, the centre lines represent the medians, the boxes the range between the 25th and 75th percentile, and the whiskers the range between the smallest and largest data point.

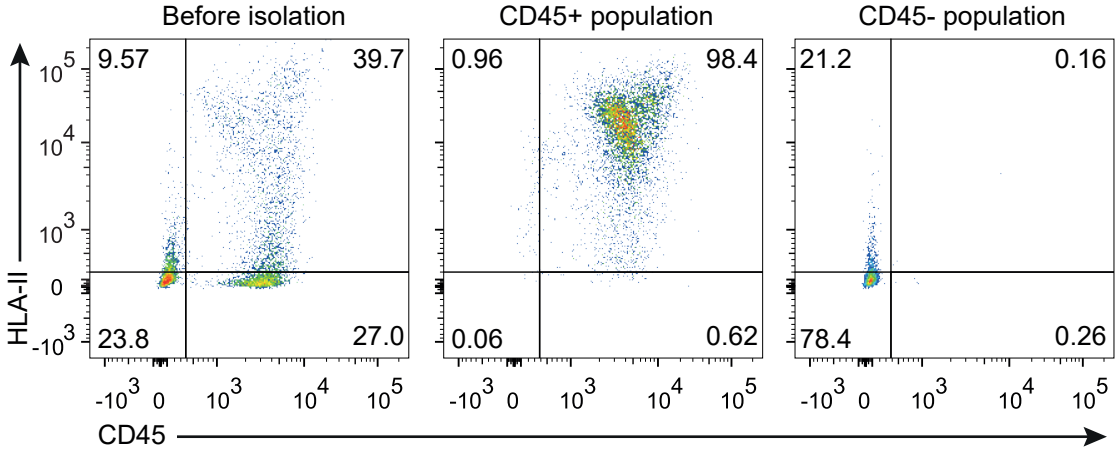
Supplementary Figure 7

a **Tumor HG38**

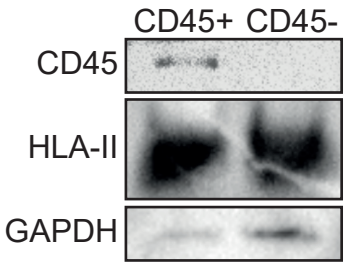
Gating strategy



Stained populations before and after seperation by CD45 marker

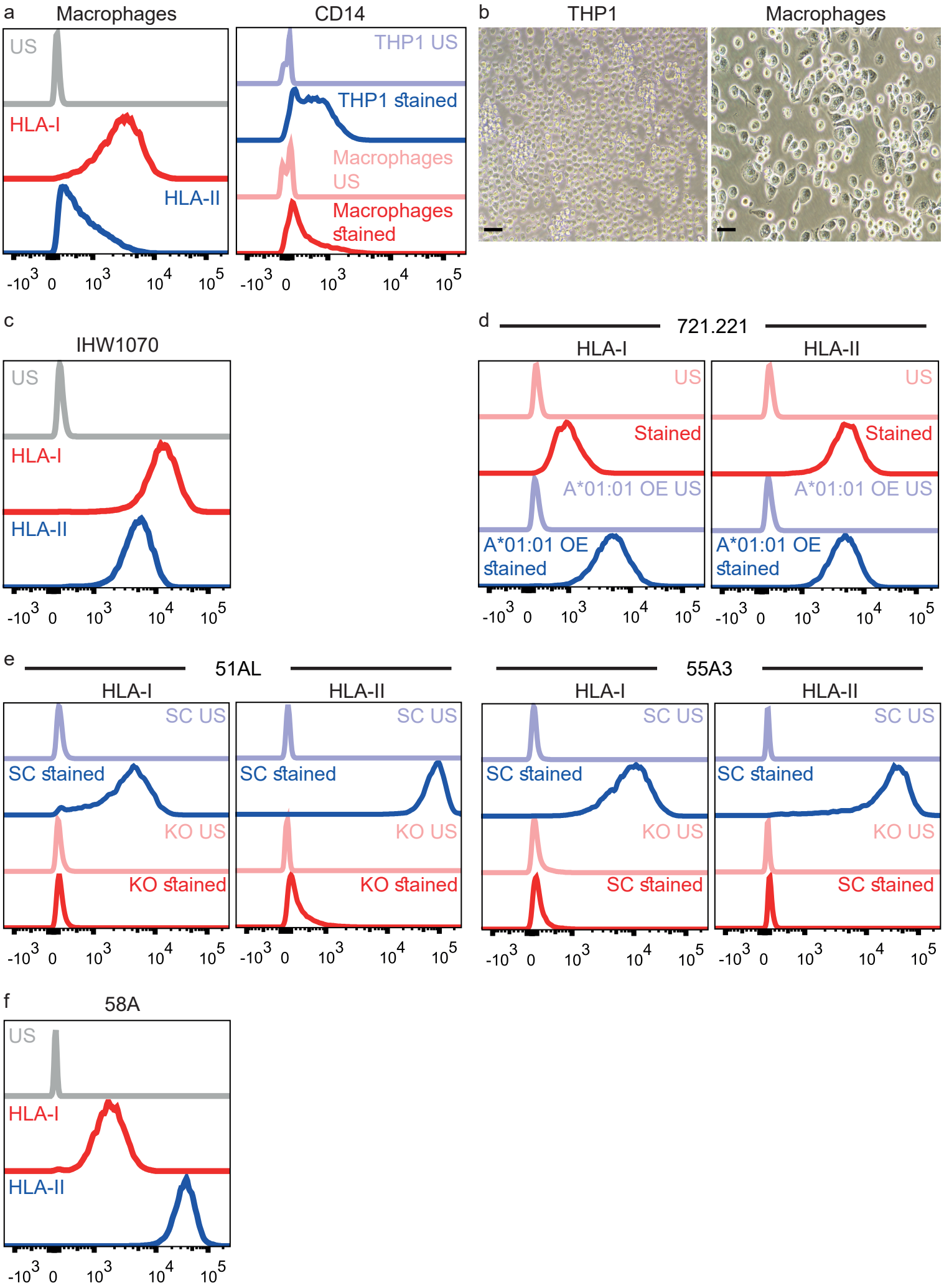


b **Tumor 422**



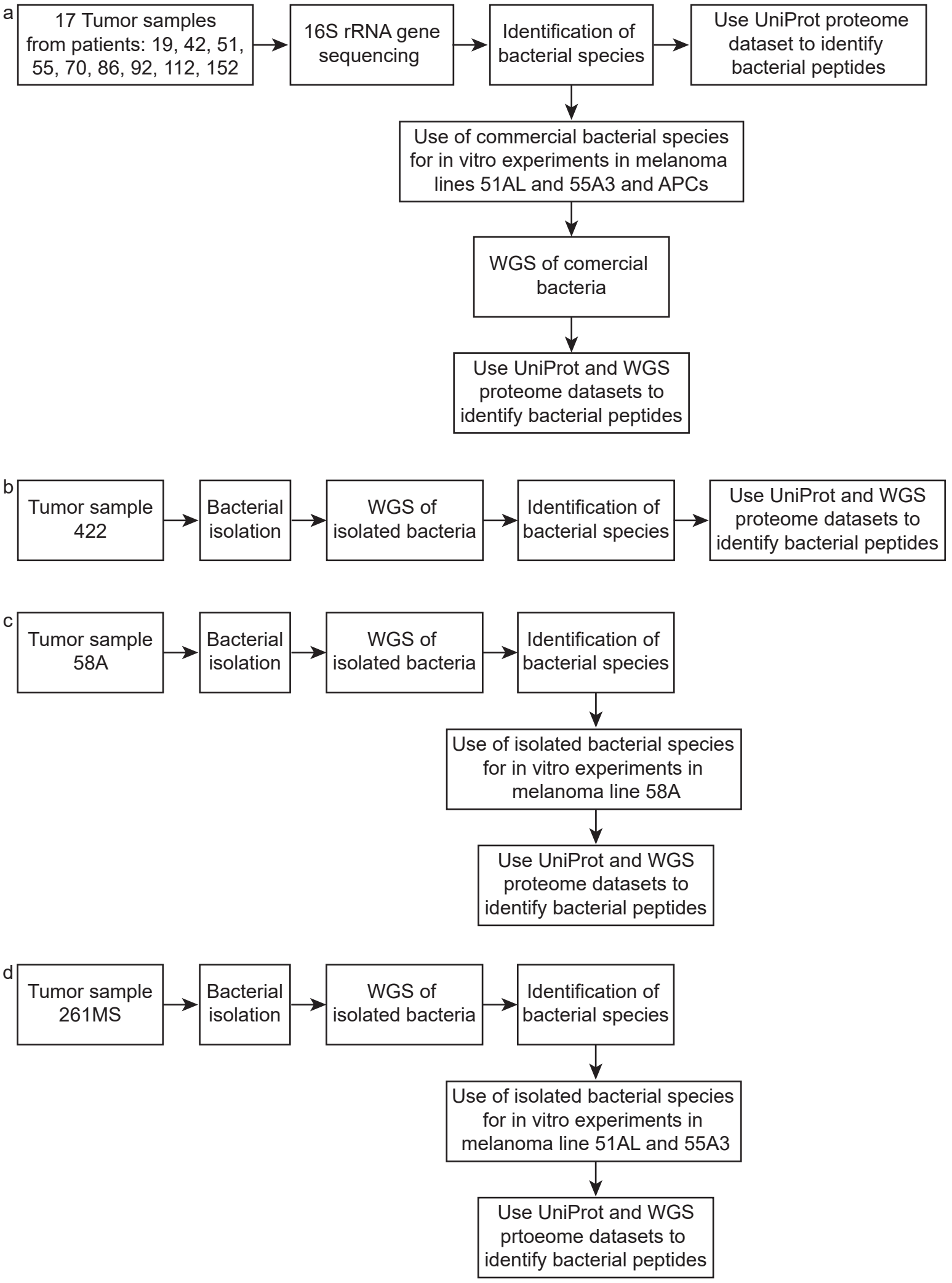
Supplementary Figure 7: CD45+/- population isolation and HLA-II expression in melanoma tumors. a, Flow cytometry analysis of tumor HG38, from which CD45+/- populations were isolated using MACS CD45 MicroBeads and stained with anti-HLA-II and anti-CD45 antibodies, indicating the efficiency of separation and HLA-II expression by the CD45- population. The top panel indicates the gating strategy used for the selection of live single cells. The last figure in this panel describes the unstained cells before separation, which were used as a negative control for the CD45 and HLA-II staining. The panel below shows the stained cells before isolation and the CD45+/- populations after the isolation based on the CD45 marker. b, Immunoblot analysis of CD45 and HLA-II expression in CD45+/- populations isolated from tumor 422.

Supplementary Figure 8



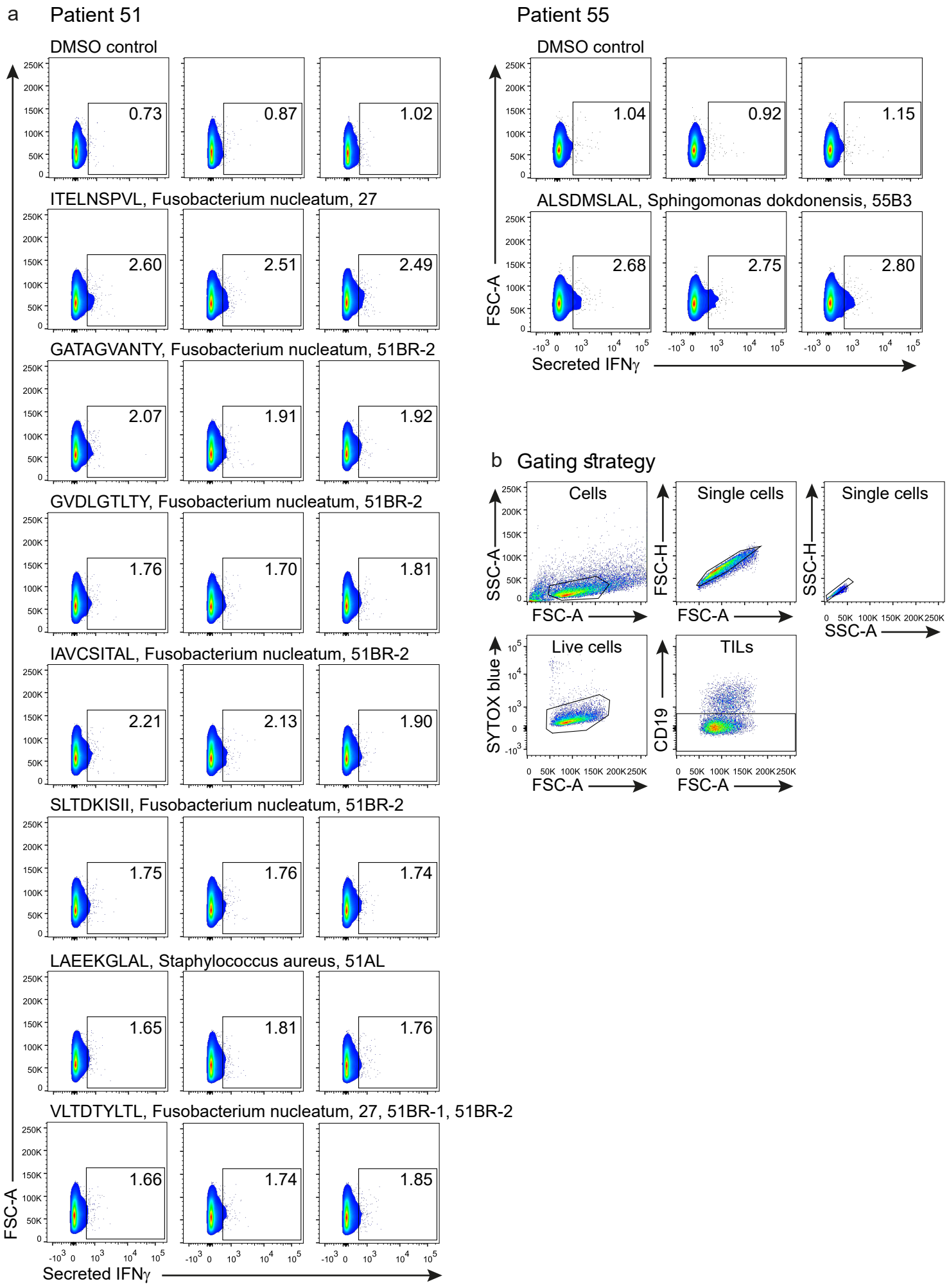
Supplementary Figure 8: THP1 differentiation into macrophages and their HLA-I and HLA-II expression in B cells and melanoma cell lines. a, THP1 cells that differentiated into macrophages, using PMA, were stained with anti-HLA-I and anti-HLA-II and analyzed using flow cytometry. Macrophages show decreased anti-CD14 staining compared to THP1 cells. b, Images of THP1 cells and THP1 differentiated into macrophages. Images are presented at 40x magnification, scale 10 μm. Figures are representative of at least three independent experiments. c,d, Flow cytometry analysis of IHW1070 B cells and 721.221 cells with A*01:01 OE stained with anti-HLA-I and anti-HLA-II. e, f, Flow-cytometry analysis of 51AL and 55A3 KO of B2M or CIITA and 58A melanoma cell lines stained with anti-HLA-I and anti-HLA-II. Unstained cells were used as control.

Supplementary Figure 9



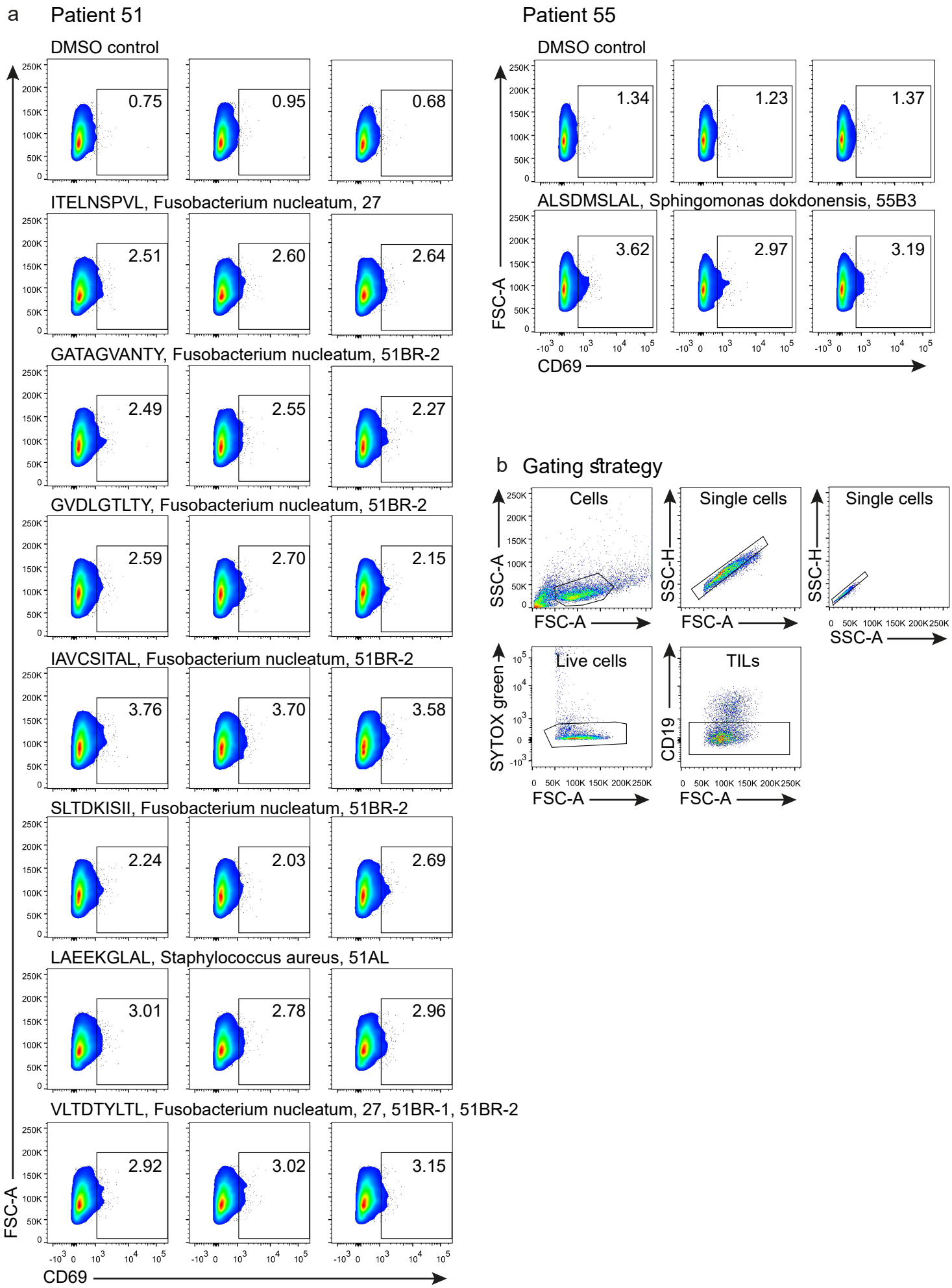
Supplementary Figure 9: Schematic overview of bacteria sequencing prior to the HLA-I and HLA-II peptidomics workflows. a, 17 tumor samples were used for 16S rRNA gene sequencing to identify their intratumor bacteria and in parallel were used for HLA peptidomics. The tumors' bacterial peptides were identified by analyzing the HLA peptidomics data with the Uniprot proteome of the identified bacterial species. In addition, commercial bacterial species were used for in vitro experiments to show entry of the bacteria and presentation of bacterial peptides by melanoma lines derived from the same tumors and APCs. We also performed WGS of the bacteria, and used the proteome assembled from it or the bacteria proteome from Uniprot to analyze the HLA peptidomics data. b, To show that both melanoma cells and APCs from the same tumor can present HLA-I and HLA-II peptides, we used a fresh tumor that was separated based on the CD45 marker. The bacteria isolated from the tumor was identified using WGS. The bacteria proteome to be used to analyze the HLA peptidomics data of the tumor was assembled from the bacteria WGS as well as its Uniprot. c-d, To show that bacterial strains that were isolated from the tumor can enter melanoma cells, we isolated bacteria from tumors 58A and 261MS, and used the bacteria for co-culture experiments with melanoma cells and HLA peptidomics. We also performed WGS of the isolated bacteria, and used the proteome assembled from it or proteome from Uniprot to analyze the HLA peptidomics data.

Supplementary Figure 10



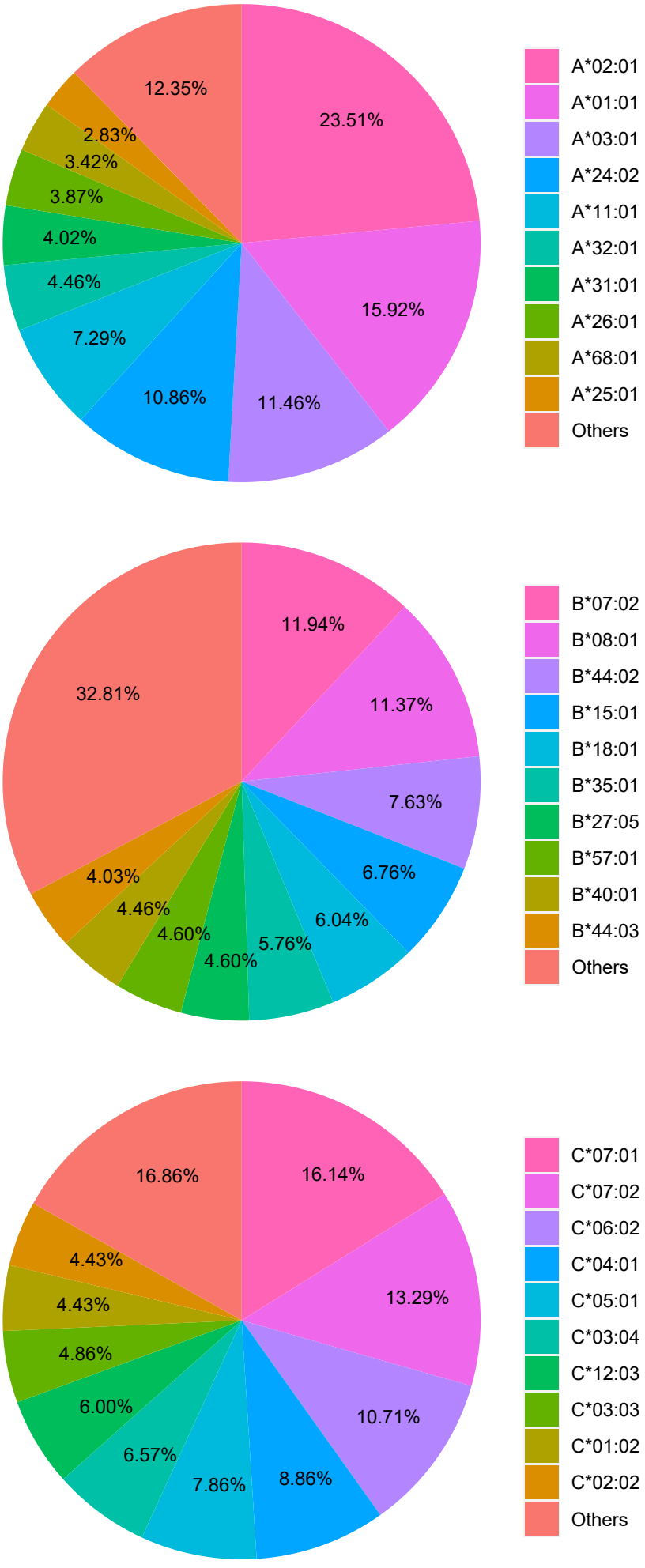
Supplementary Figure 10: Immunogenic bacterial peptides in IFNγ secretion assay. a, Flow cytometry analysis of IFNγ–secreting TILs after the co-culture of 51AL and 55A TILs with B cells loaded with the indicated bacterial peptide or DMSO control, as described in Extended Data Fig. 12. The figure contains peptides in which the ratio of IFNγ secretion to DMSO control was two-fold or above. b, TILs were gated as single, live and CD19-negative cells.

Supplementary Figure 11



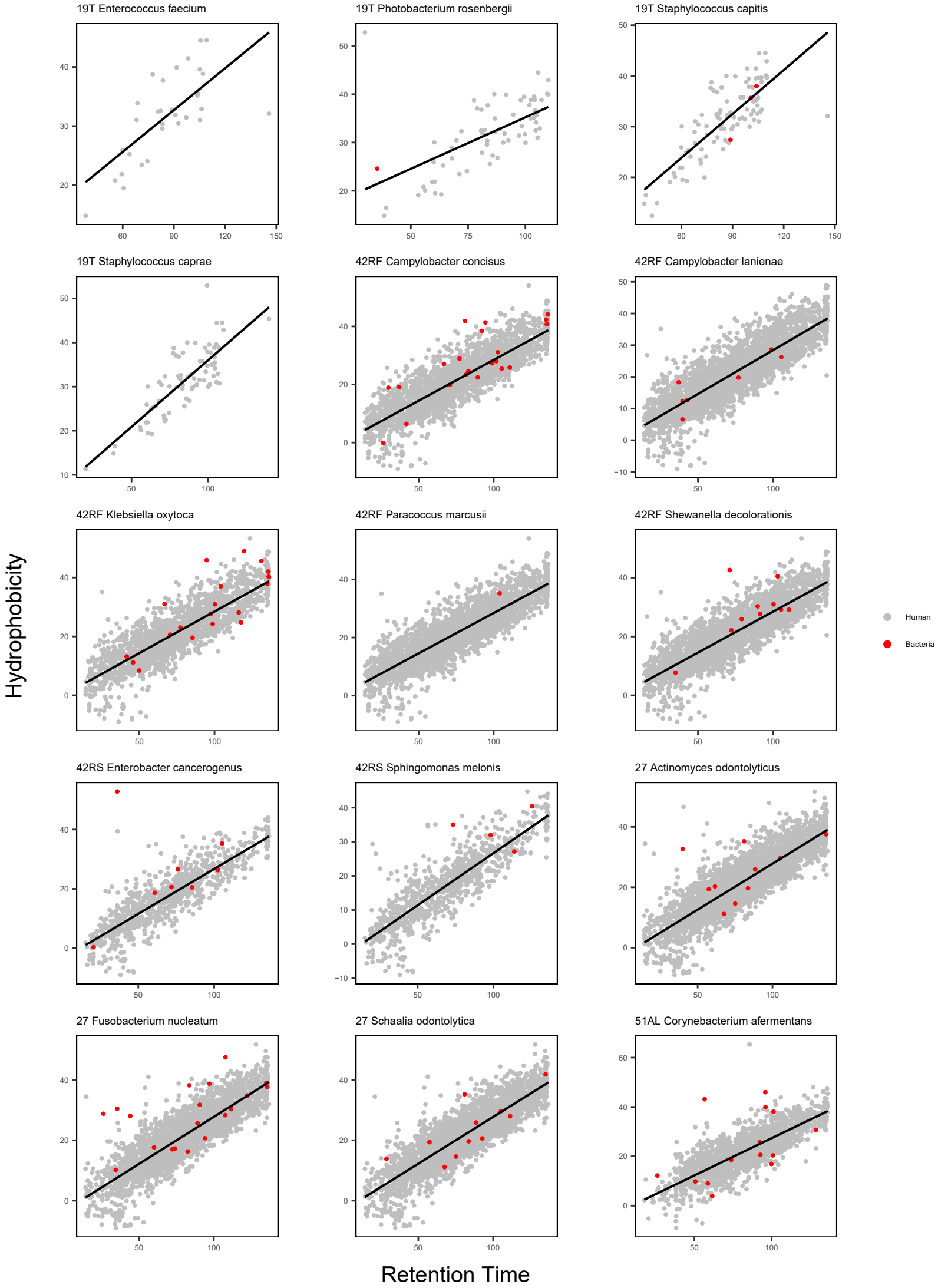
Supplementary Figure 11: CD69 expression by TILs reactive towards immunogenic bacteria-derived antigens. a, Flow-cytometry analysis of CD69-positive TILs after co-culture of 51AL and 55A TILs with B cells loaded with the bacterial peptides shown to increase IFNγ secretion in Supplementary Fig. 10 or DMSO control. b, TILs were gated as single, live and CD19-negative cells.

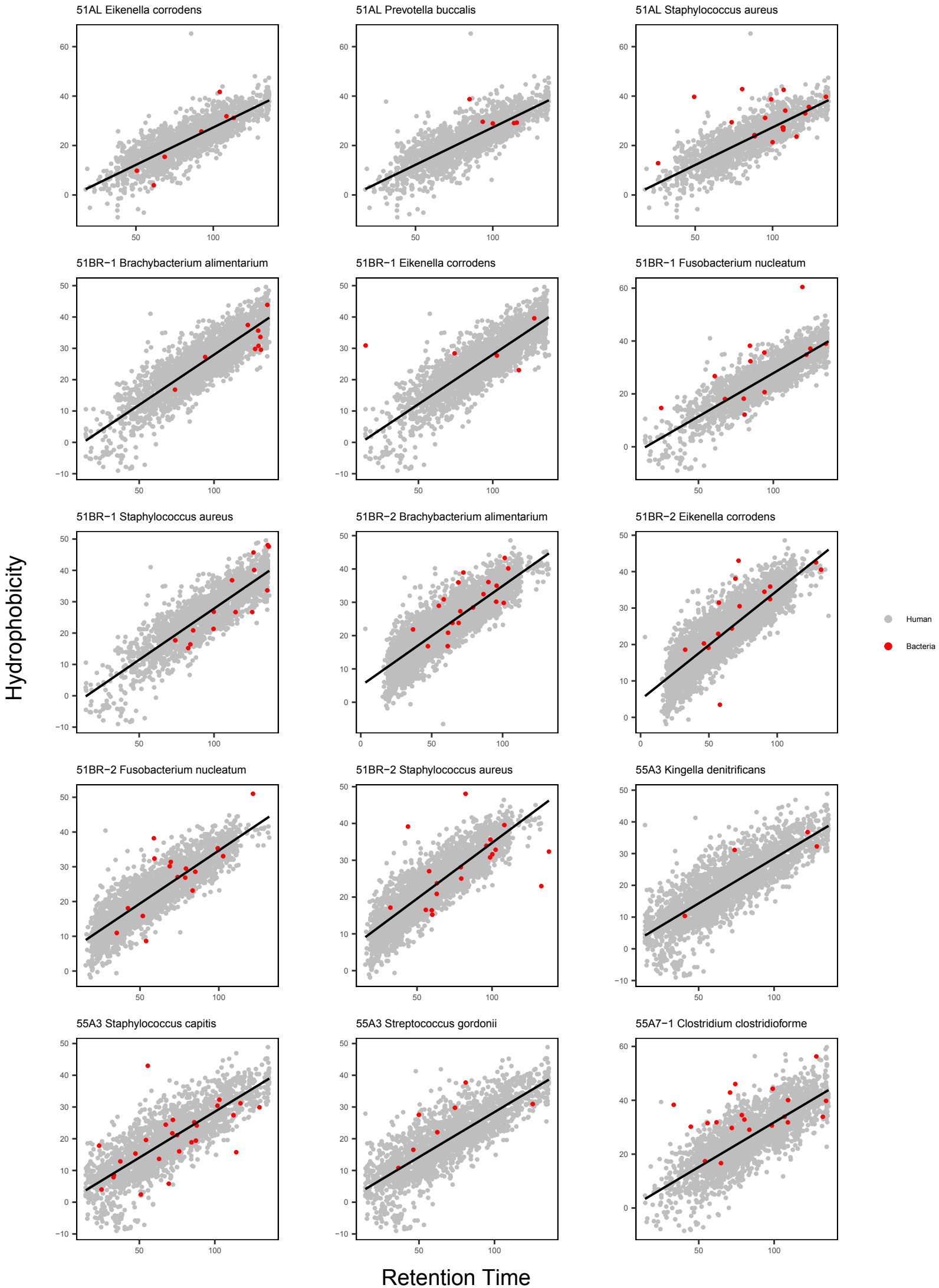
Supplementary Figure 12

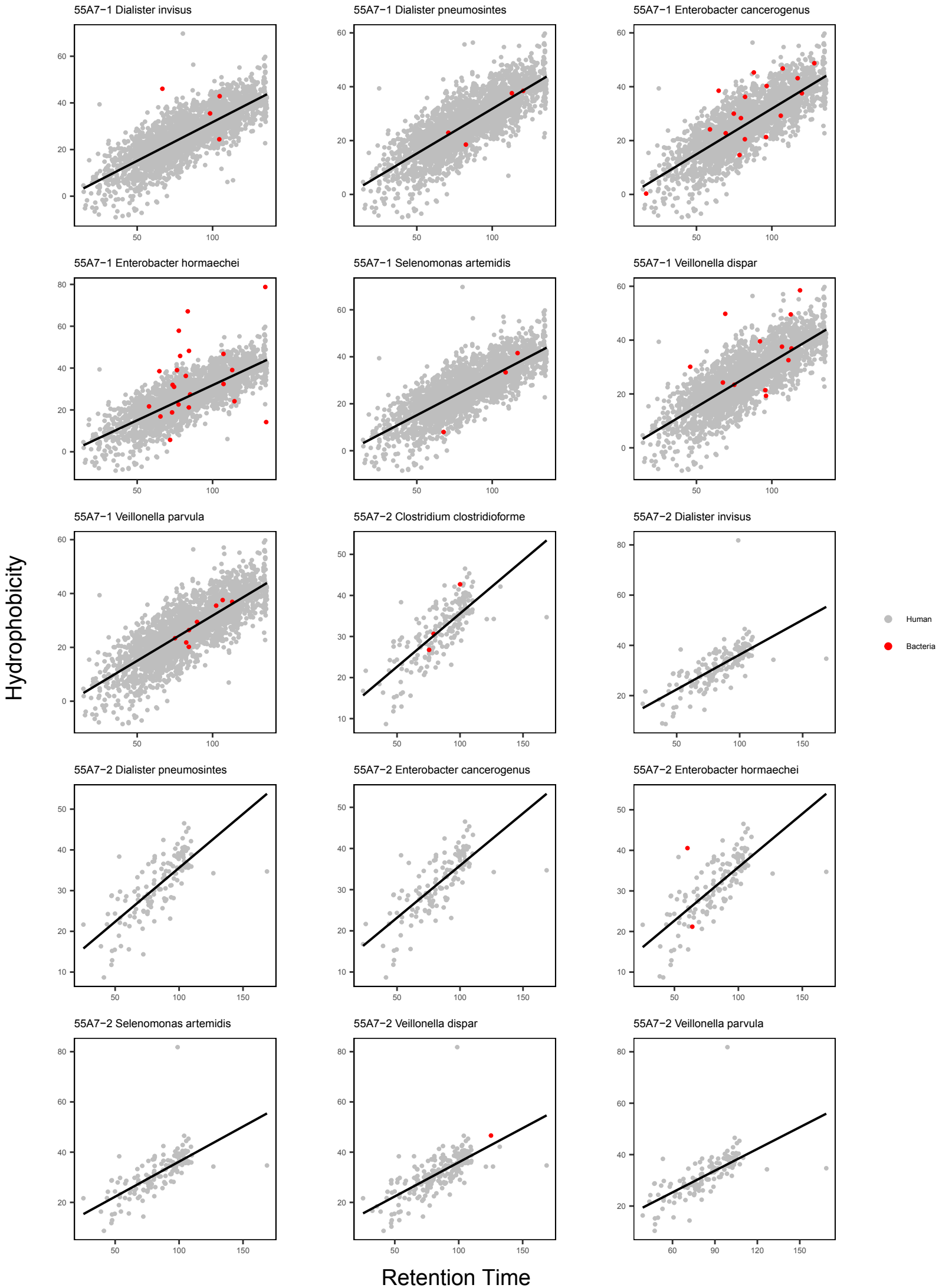


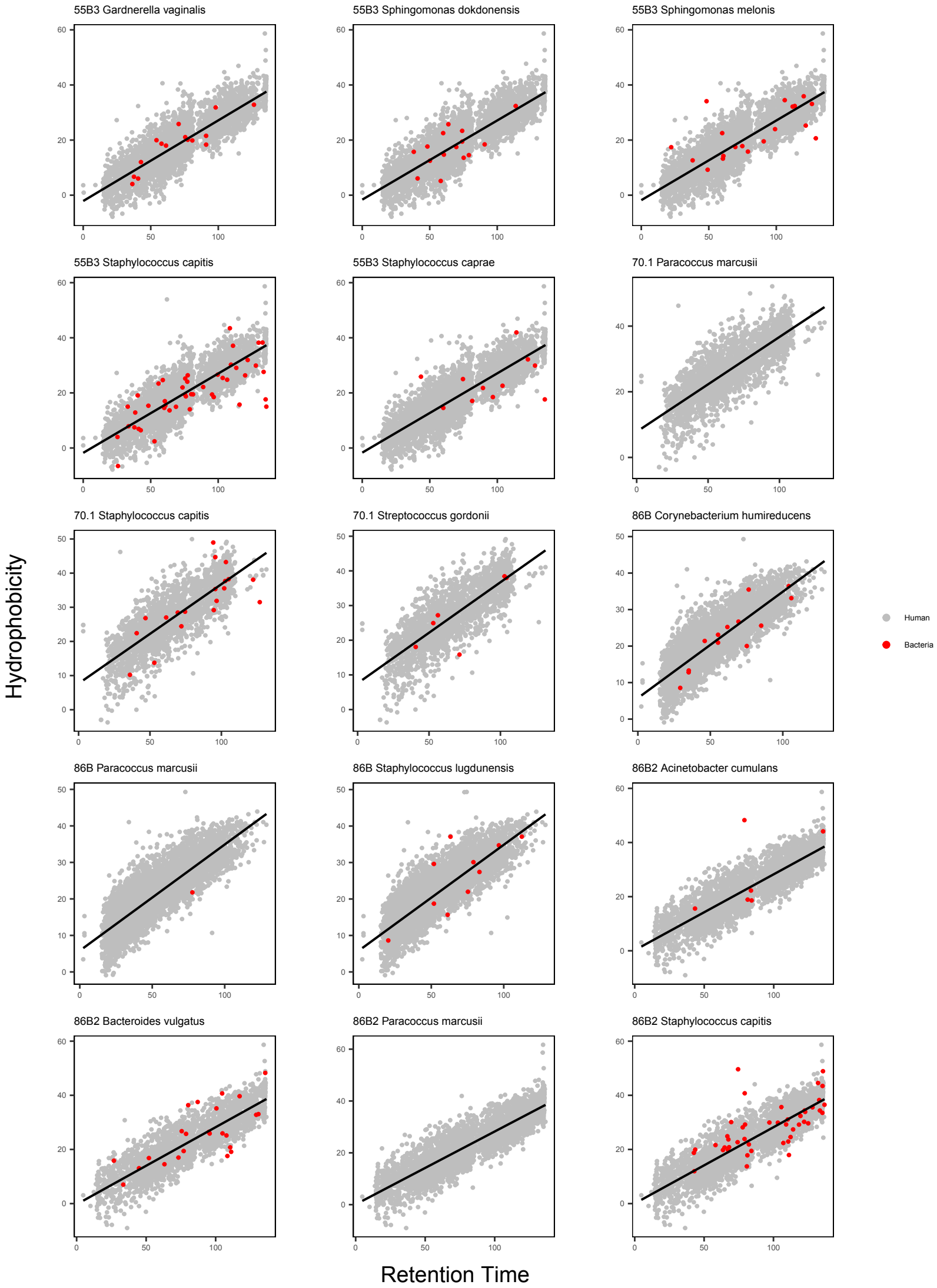
Supplementary Figure 12: Recurrent and immunogenic bacterial peptides were derived from prevalent HLA alleles in melanoma TCGA patients. The distribution of the 10 most abundant HLA-A/B/C alleles were plotted in pie chart. All other alleles were grouped together. Of the 10 most abundant HLA alleles for each of the HLA-A/B/C molecules, we identified recurrent and immunogenic bacterial peptides from the A*02:01, A*01:01, A*11:01, A*03:01, A*24:02, A*68:01, B*44:02, B*15:01, B*07:02, B*27:05, C*03:03, C*05:01, C*03:04, C*06:02, C*07:02 and C*02:02 alleles.

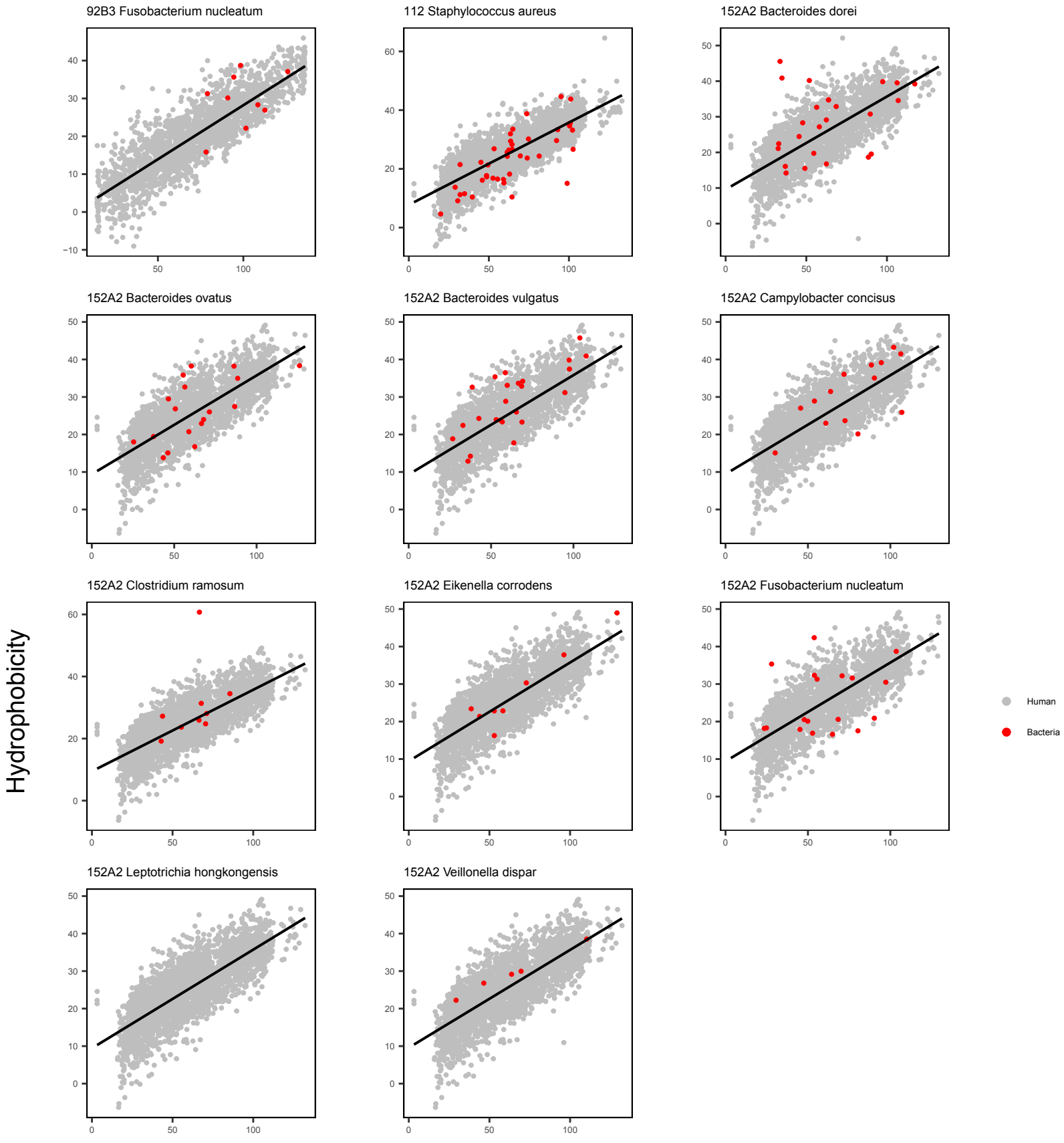
Supplementary Figure 13
Correlation between retention time to predicted hydrophobicity HLA-I peptides

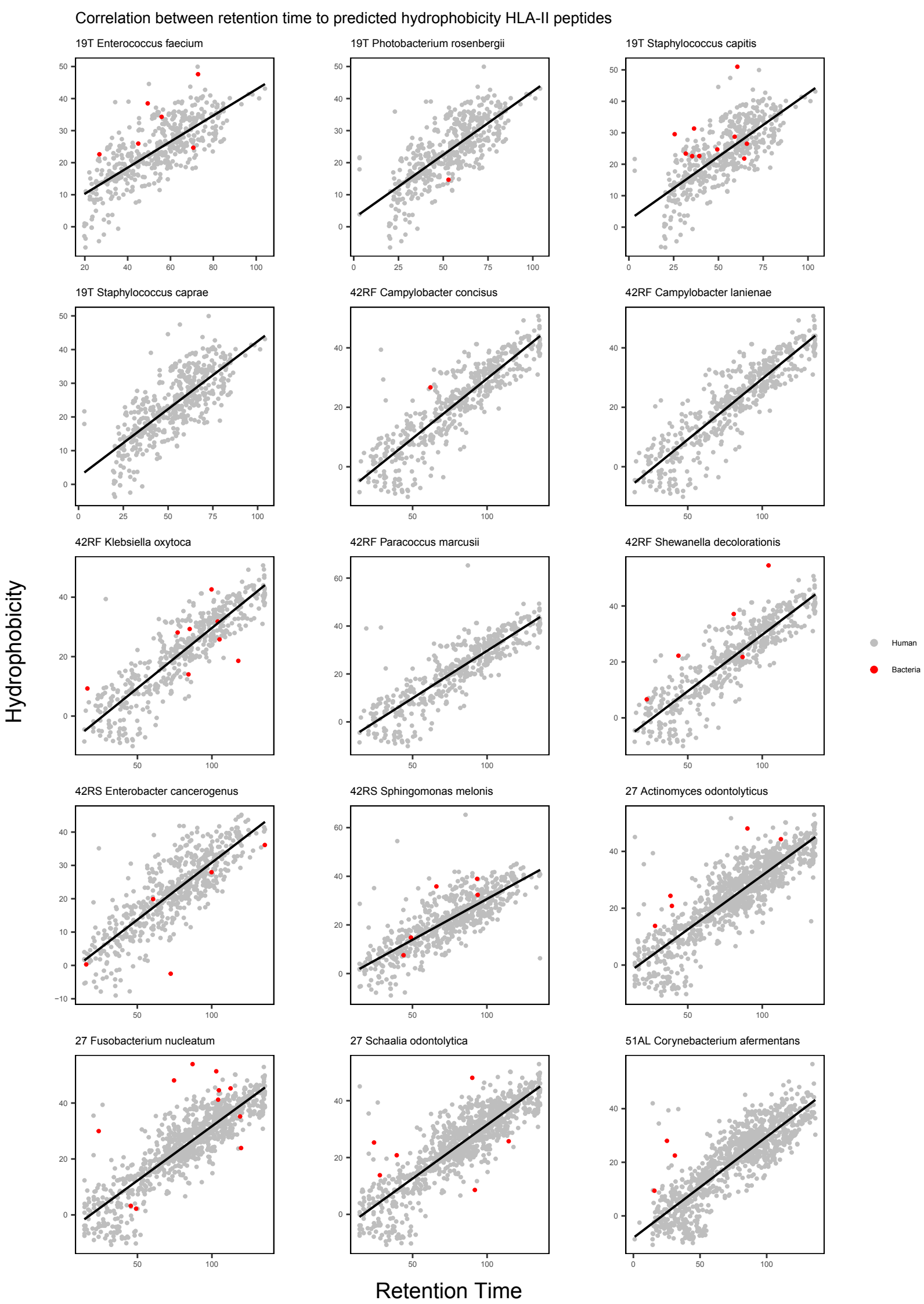


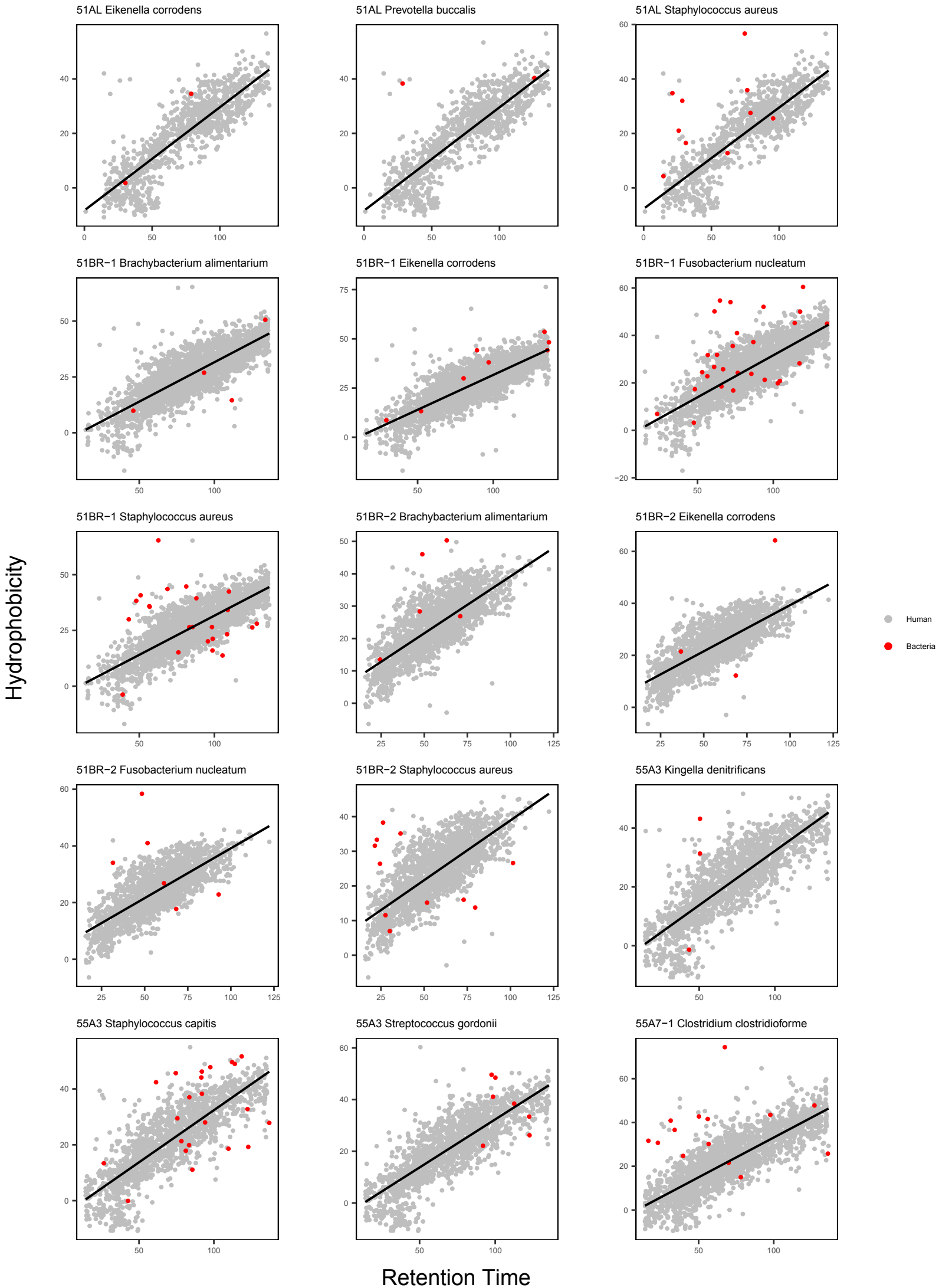

51AL Eikenella corrodens
51AL Prevotella buccalis
51AL Staphylococcus aureus
51BR−1 Brachybacterium alimentarium
51BR−1 Eikenella corrodens
51BR−1 Fusobacterium nucleatum
51BR−1 Staphylococcus aureus
51BR−2 Brachybacterium alimentarium
51BR−2 Eikenella corrodens
51BR−2 Fusobacterium nucleatum
51BR−2 Staphylococcus aureus
55A3 Kingella denitrificans
55A3 Staphylococcus capitis
55A3 Streptococcus gordonii
55A7−1 Clostridium clostridioforme
Hydrophobicity
Retention Time
Human
Bacteria

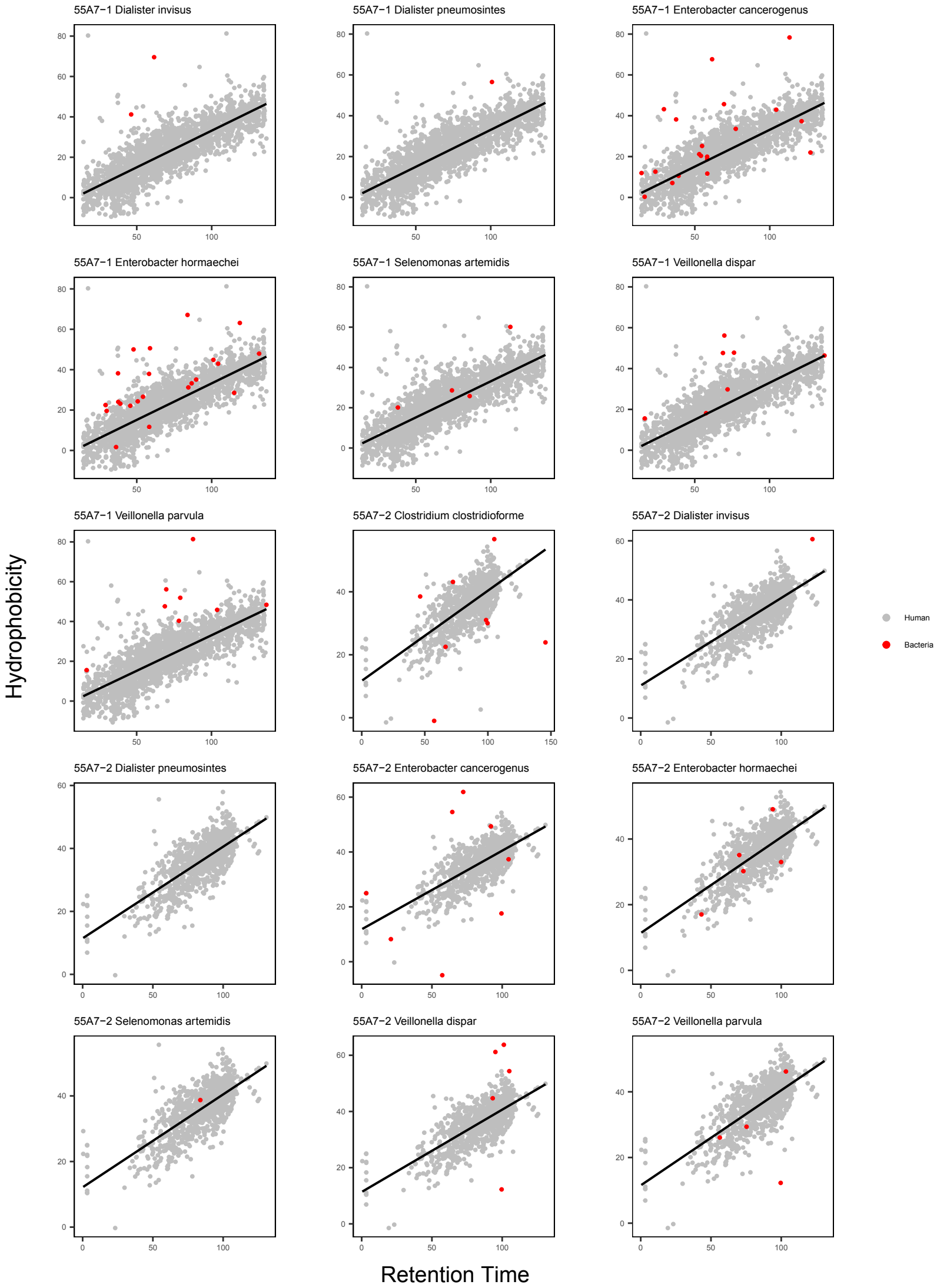

55A7−1 Dialister invisus
55A7−1 Dialister pneumosintes
55A7−1 Enterobacter cancerogenus
55A7−1 Enterobacter hormaechei
55A7−1 Selenomonas artemidis
55A7−1 Veillonella dispar
55A7−1 Veillonella parvula
55A7−2 Clostridium clostridioforme
55A7−2 Dialister invisus
55A7−2 Dialister pneumosintes
55A7−2 Enterobacter cancerogenus
55A7−2 Enterobacter hormaechei
55A7−2 Selenomonas artemidis
55A7−2 Veillonella dispar
55A7−2 Veillonella parvula
Hydrophobicity
Retention Time
Human
Bacteria

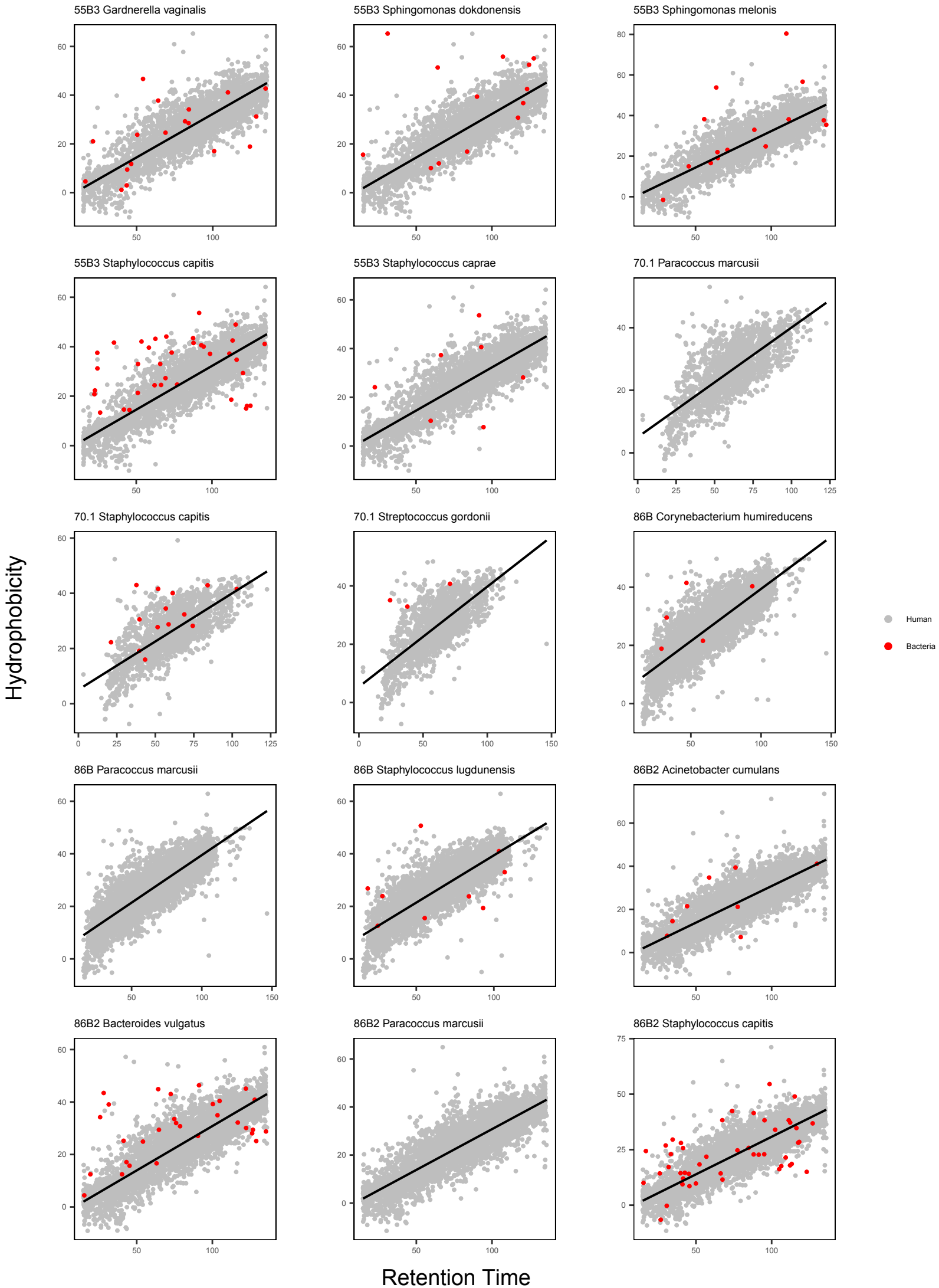

55B3 Gardnerella vaginalis
55B3 Sphingomonas dokdonensis
55B3 Sphingomonas melonis
55B3 Staphylococcus capitis
55B3 Staphylococcus caprae
70.1 Paracoccus marcusii
70.1 Staphylococcus capitis
70.1 Streptococcus gordonii
86B Corynebacterium humireducens
86B Paracoccus marcusii
86B Staphylococcus lugdunensis
86B2 Acinetobacter cumulans
86B2 Bacteroides vulgatus
86B2 Paracoccus marcusii
86B2 Staphylococcus capitis
Hydrophobicity
Retention Time
Human
Bacteria

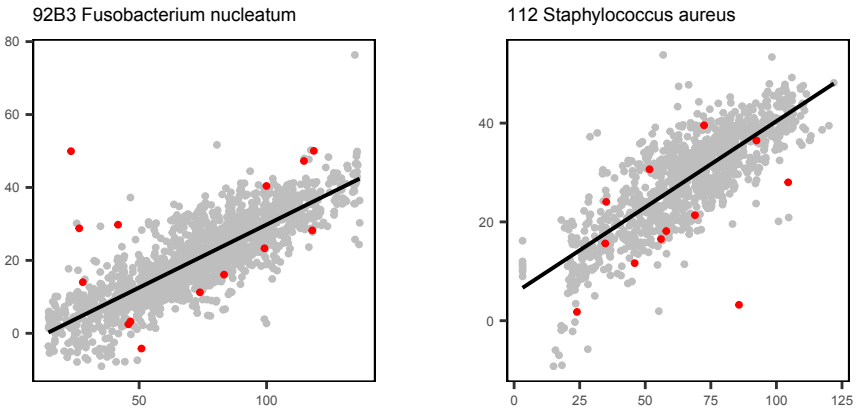

92B3 Fusobacterium nucleatum
112 Staphylococcus aureus
152A2 Bacteroides dorei
152A2 Bacteroides ovatus
152A2 Bacteroides vulgatus
152A2 Campylobacter concisus
152A2 Clostridium ramosum
152A2 Eikenella corrodens
152A2 Fusobacterium nucleatum
152A2 Leptotrichia hongkongensis
152A2 Veillonella dispar
Hydrophobicity
Retention Time
Human
Bacteria


Correlation between retention time to predicted hydrophobicity HLA-II peptides
19T Enterococcus faecium
19T Photobacterium rosenbergii
19T Staphylococcus capitis
19T Staphylococcus caprae
42RF Campylobacter concisus
42RF Campylobacter lanienae
42RF Klebsiella oxytoca
42RF Paracoccus marcusii
42RF Shewanella decolorationis
42RS Enterobacter cancerogenus
42RS Sphingomonas melonis
27 Actinomyces odontolyticus
27 Fusobacterium nucleatum
27 Schaalia odontolytica
51AL Corynebacterium afermentans
Hydrophobicity
Retention Time
Human
Bacteria


51AL Eikenella corrodens
51AL Prevotella buccalis
51AL Staphylococcus aureus
51BR−1 Brachybacterium alimentarium
51BR−1 Eikenella corrodens
51BR−1 Fusobacterium nucleatum
51BR−1 Staphylococcus aureus
51BR−2 Brachybacterium alimentarium
51BR−2 Eikenella corrodens
51BR−2 Fusobacterium nucleatum
51BR−2 Staphylococcus aureus
55A3 Kingella denitrificans
55A3 Staphylococcus capitis
55A3 Streptococcus gordonii
55A7−1 Clostridium clostridioforme
Hydrophobicity
Retention Time
Human
Bacteria


55A7−1 Dialister invisus
55A7−1 Dialister pneumosintes
55A7−1 Enterobacter cancerogenus
55A7−1 Enterobacter hormaechei
55A7−1 Selenomonas artemidis
55A7−1 Veillonella dispar
55A7−1 Veillonella parvula
55A7−2 Clostridium clostridioforme
55A7−2 Dialister invisus
55A7−2 Dialister pneumosintes
55A7−2 Enterobacter cancerogenus
55A7−2 Enterobacter hormaechei
55A7−2 Selenomonas artemidis
55A7−2 Veillonella dispar
55A7−2 Veillonella parvula
Hydrophobicity
Retention Time
Human
Bacteria


55B3 Gardnerella vaginalis
55B3 Sphingomonas dokdonensis
55B3 Sphingomonas melonis
55B3 Staphylococcus capitis
55B3 Staphylococcus caprae
70.1 Paracoccus marcusii
70.1 Staphylococcus capitis
70.1 Streptococcus gordonii
86B Corynebacterium humireducens
86B Paracoccus marcusii
86B Staphylococcus lugdunensis
86B2 Acinetobacter cumulans
86B2 Bacteroides vulgatus
86B2 Paracoccus marcusii
86B2 Staphylococcus capitis
Hydrophobicity
Retention Time
Human
Bacteria

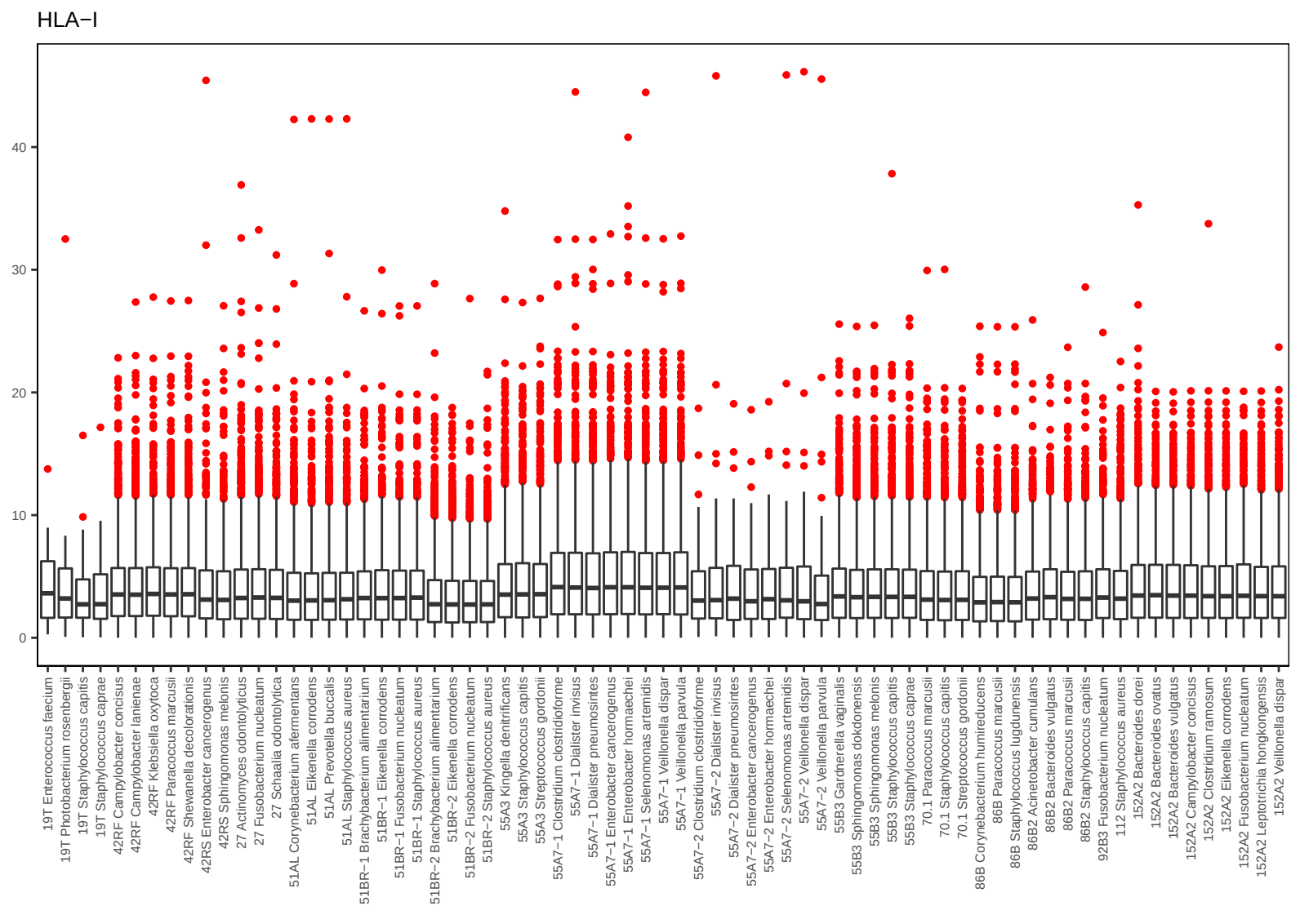


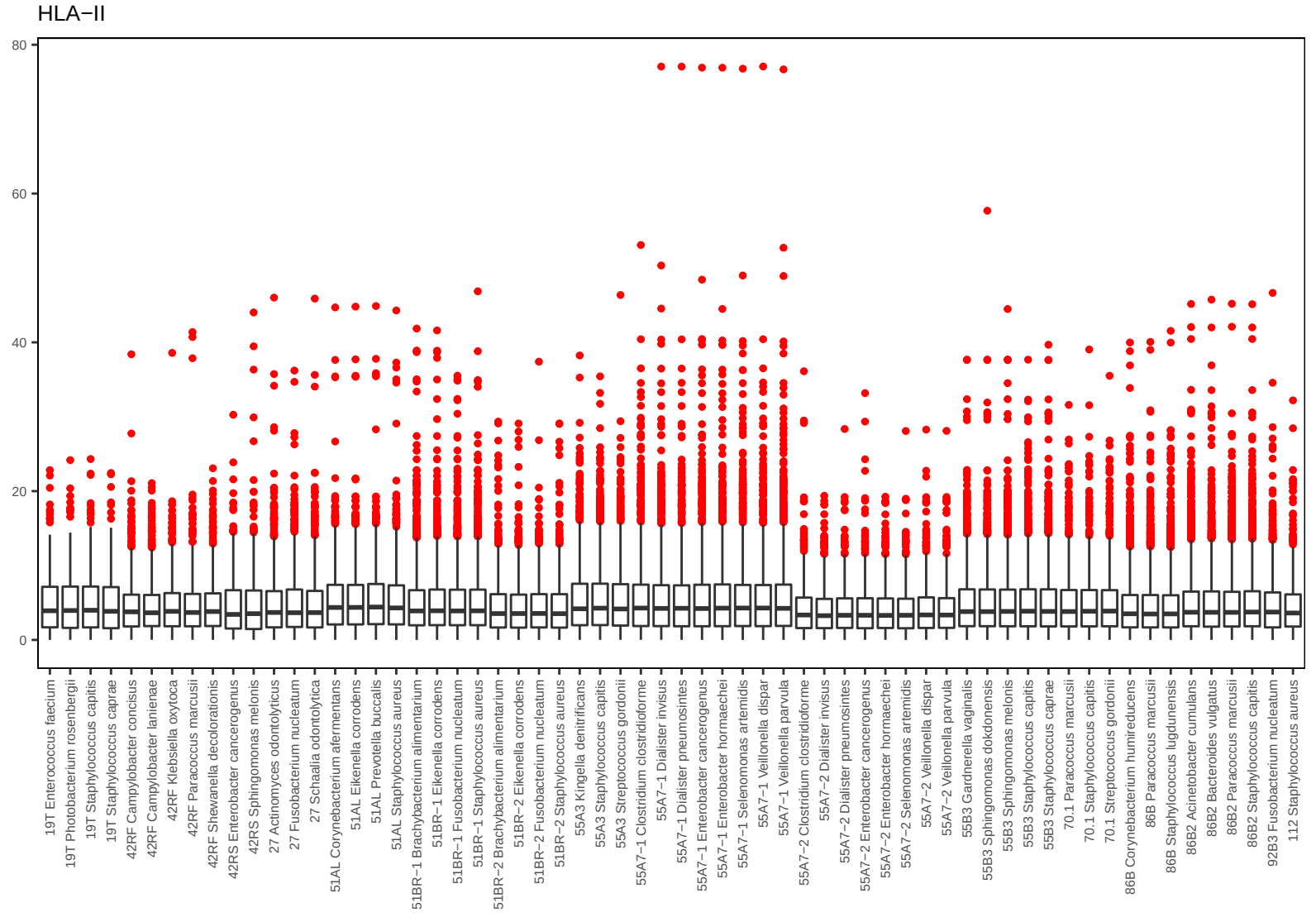
Hydrophobicity

Human

Bacteria

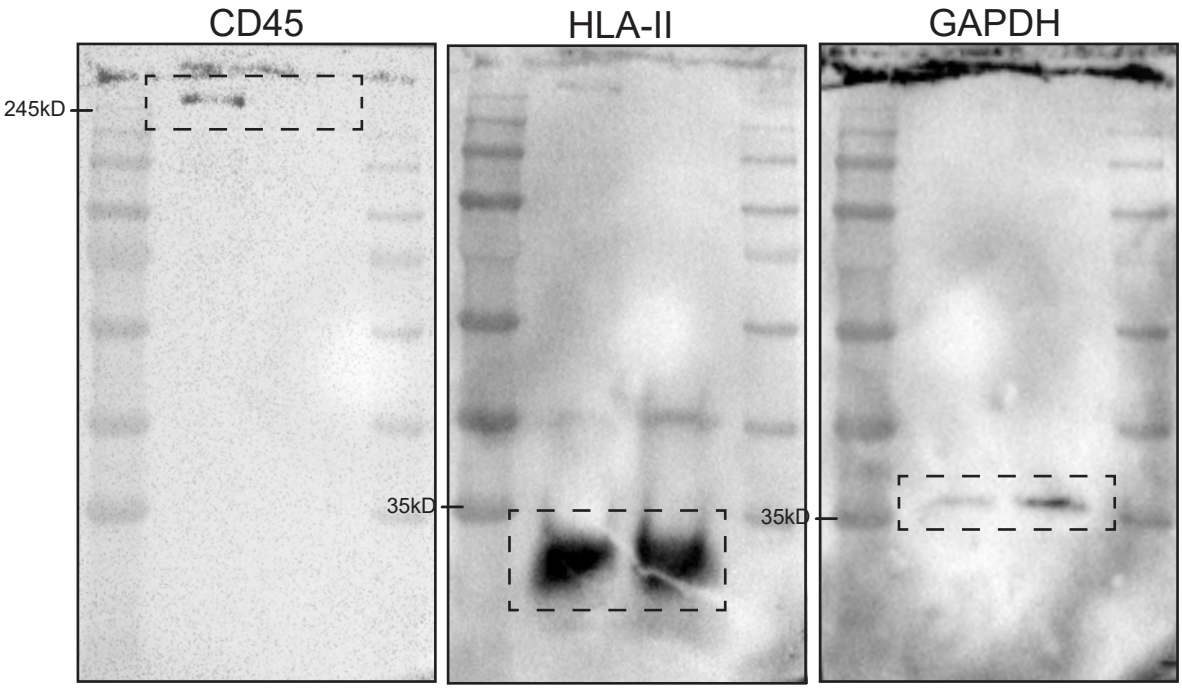
Retention Time


HLA−I
0
10
20
30
40
19T Enterococcus faecium
19T Photobacterium rosenbergii
19T Staphylococcus capitis
19T Staphylococcus caprae
42RF Campylobacter concisus
42RF Campylobacter lanienae
42RF Klebsiella oxytoca
42RF Paracoccus marcusii
42RF Shewanella decolorationis
42RS Enterobacter cancerogenus
42RS Sphingomonas melonis
27 Actinomyces odontolyticus
27 Fusobacterium nucleatum
27 Schaalia odontolytica
51AL Corynebacterium afermentans
51AL Eikenella corrodens
51AL Prevotella buccalis
51AL Staphylococcus aureus
51BR−1 Brachybacterium alimentarium
51BR−1 Eikenella corrodens
51BR−1 Fusobacterium nucleatum
51BR−1 Staphylococcus aureus
51BR−2 Brachybacterium alimentarium
51BR−2 Eikenella corrodens
51BR−2 Fusobacterium nucleatum
51BR−2 Staphylococcus aureus
55A3 Kingella denitrificans
55A3 Staphylococcus capitis
55A3 Streptococcus gordonii
55A7−1 Clostridium clostridioforme
55A7−1 Dialister invisus
55A7−1 Dialister pneumosintes
55A7−1 Enterobacter cancerogenus
55A7−1 Enterobacter hormaechei
55A7−1 Selenomonas artemidis
55A7−1 Veillonella dispar
55A7−1 Veillonella parvula
55A7−2 Clostridium clostridioforme
55A7−2 Dialister invisus
55A7−2 Dialister pneumosintes
55A7−2 Enterobacter cancerogenus
55A7−2 Enterobacter hormaechei
55A7−2 Selenomonas artemidis
55A7−2 Veillonella dispar
55A7−2 Veillonella parvula
55B3 Gardnerella vaginalis
55B3 Sphingomonas dokdonensis
55B3 Sphingomonas melonis
55B3 Staphylococcus capitis
55B3 Staphylococcus caprae
70.1 Paracoccus marcusii
70.1 Staphylococcus capitis
70.1 Streptococcus gordonii
86B Corynebacterium humireducens
86B Paracoccus marcusii
86B Staphylococcus lugdunensis
86B2 Acinetobacter cumulans
86B2 Bacteroides vulgatus
86B2 Paracoccus marcusii
86B2 Staphylococcus capitis
92B3 Fusobacterium nucleatum
112 Staphylococcus aureus
152A2 Bacteroides dorei
152A2 Bacteroides ovatus
152A2 Bacteroides vulgatus
152A2 Campylobacter concisus
152A2 Clostridium ramosum
152A2 Eikenella corrodens
152A2 Fusobacterium nucleatum
152A2 Leptotrichia hongkongensis
152A2 Veillonella dispar


HLA−II
0
20
40
60
80
19T Enterococcus faecium
19T Photobacterium rosenbergii
19T Staphylococcus capitis
19T Staphylococcus caprae
42RF Campylobacter concisus
42RF Campylobacter lanienae
42RF Klebsiella oxytoca
42RF Paracoccus marcusii
42RF Shewanella decolorationis
42RS Enterobacter cancerogenus
42RS Sphingomonas melonis
27 Actinomyces odontolyticus
27 Fusobacterium nucleatum
27 Schaalia odontolytica
51AL Corynebacterium afermentans
51AL Eikenella corrodens
51AL Prevotella buccalis
51AL Staphylococcus aureus
51BR−1 Brachybacterium alimentarium
51BR−1 Eikenella corrodens
51BR−1 Fusobacterium nucleatum
51BR−1 Staphylococcus aureus
51BR−2 Brachybacterium alimentarium
51BR−2 Eikenella corrodens
51BR−2 Fusobacterium nucleatum
51BR−2 Staphylococcus aureus
55A3 Kingella denitrificans
55A3 Staphylococcus capitis
55A3 Streptococcus gordonii
55A7−1 Clostridium clostridioforme
55A7−1 Dialister invisus
55A7−1 Dialister pneumosintes
55A7−1 Enterobacter cancerogenus
55A7−1 Enterobacter hormaechei
55A7−1 Selenomonas artemidis
55A7−1 Veillonella dispar
55A7−1 Veillonella parvula
55A7−2 Clostridium clostridioforme
55A7−2 Dialister invisus
55A7−2 Dialister pneumosintes
55A7−2 Enterobacter cancerogenus
55A7−2 Enterobacter hormaechei
55A7−2 Selenomonas artemidis
55A7−2 Veillonella dispar
55A7−2 Veillonella parvula
55B3 Gardnerella vaginalis
55B3 Sphingomonas dokdonensis
55B3 Sphingomonas melonis
55B3 Staphylococcus capitis
55B3 Staphylococcus caprae
70.1 Paracoccus marcusii
70.1 Staphylococcus capitis
70.1 Streptococcus gordonii
86B Corynebacterium humireducens
86B Paracoccus marcusii
86B Staphylococcus lugdunensis
86B2 Acinetobacter cumulans
86B2 Bacteroides vulgatus
86B2 Paracoccus marcusii
86B2 Staphylococcus capitis
92B3 Fusobacterium nucleatum
112 Staphylococcus aureus

Supplementary Figure 13: Correlation between peptides' experimental retention time to the predicted peptides' hydrophobicity. The peptides' RTs were plotted against the calculated hydrophobicity index for each sample and identified bacteria. The measured RTs were regressed against the calculated hydrophobicity index and the residual absolute errors of the lm-regression were plotted using a boxplot. In box plots, the center lines represent the medians, the boxes the range between the 25th and 75th percentile, and the whiskers the range between the smallest and largest data point. In the scatter plots, the red dots represent the bacterial peptides, with the human peptides marked in gray. In the box plots, the red dots represent the outlier human and bacterial peptides of the lm-regression.

Supplemenatry Figure 14. Uncropped gels used for Supplementary Figure 7



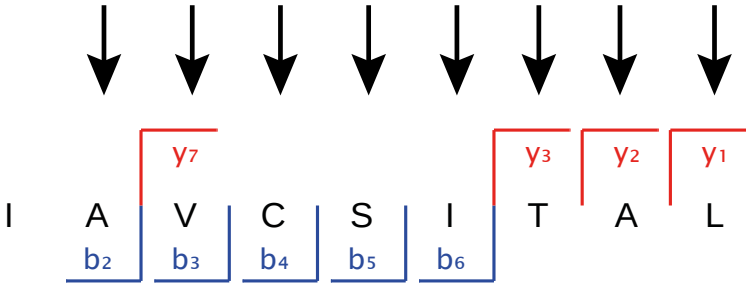
Supplementary Figure 14: Uncropped Immunoblot described in Supplementary Figure 7b.

Supplementary Information

MS/MS coverage calculation and threshold selection

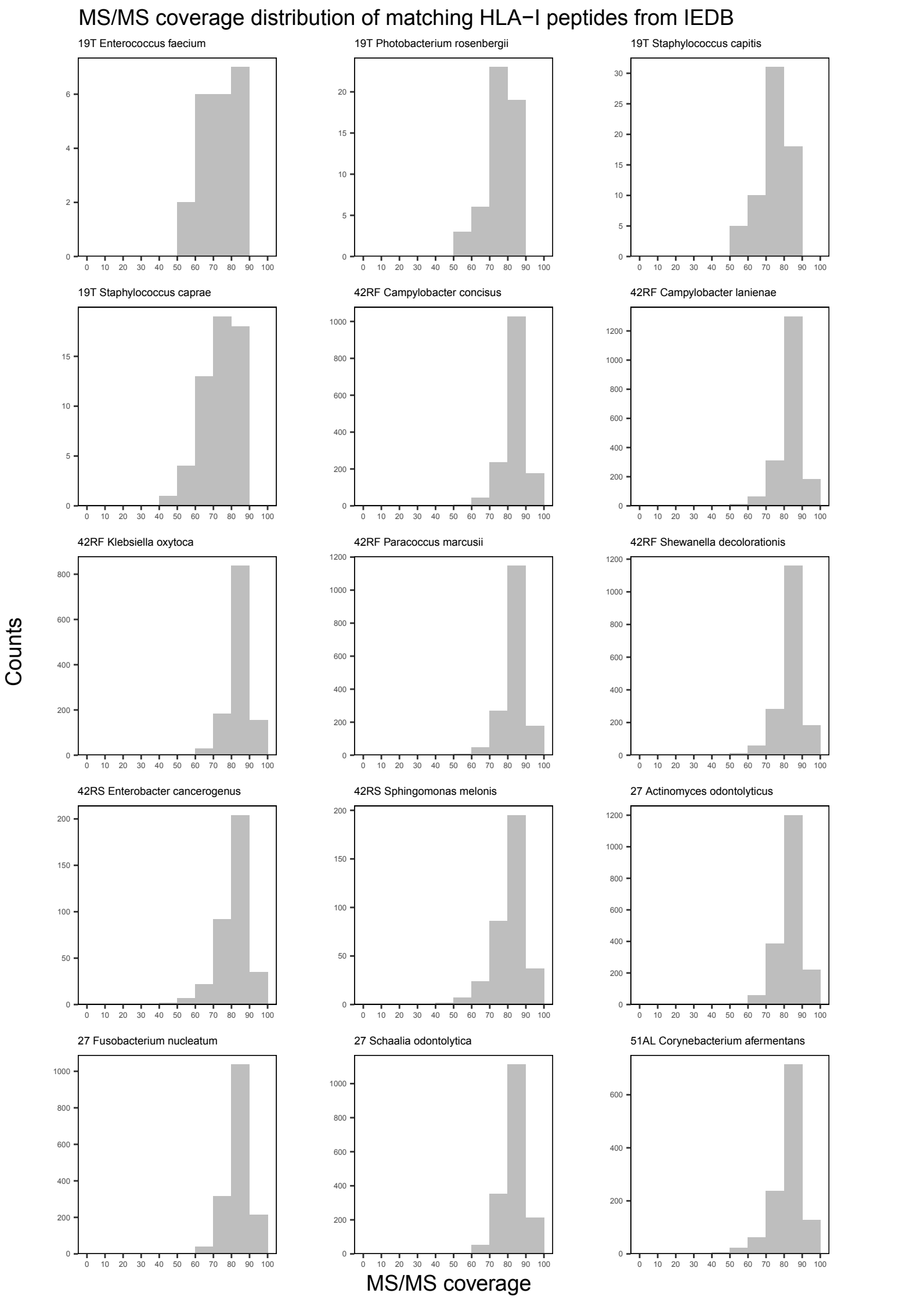
The identified bacteria-derived peptides were filtered to ensure they are not of human origin (wild-type proteins as well as possible non-coding sequences or pseudogenes). Filtering was done by assessing the peptides' predicted ability to bind the patient's HLA alleles (NetMHCpan predictions) and two quality attributes: (i) We used a linear regression model of the peptides' experimental retention time and predicted hydrophobicity. ; and (ii) we determined the peptide's MS/MS coverage, by calculating the value that represents the number of b or y ions measured for the MS/MS spectra relative to the peptide's length. To this end, we extracted for each peptide its ion information (from the msms.txt MaxQuant output), counted the number of amino acids that start either by b or y ions, and then divided this number by the peptide's length; this value is referred to as "MS/MS coverage".

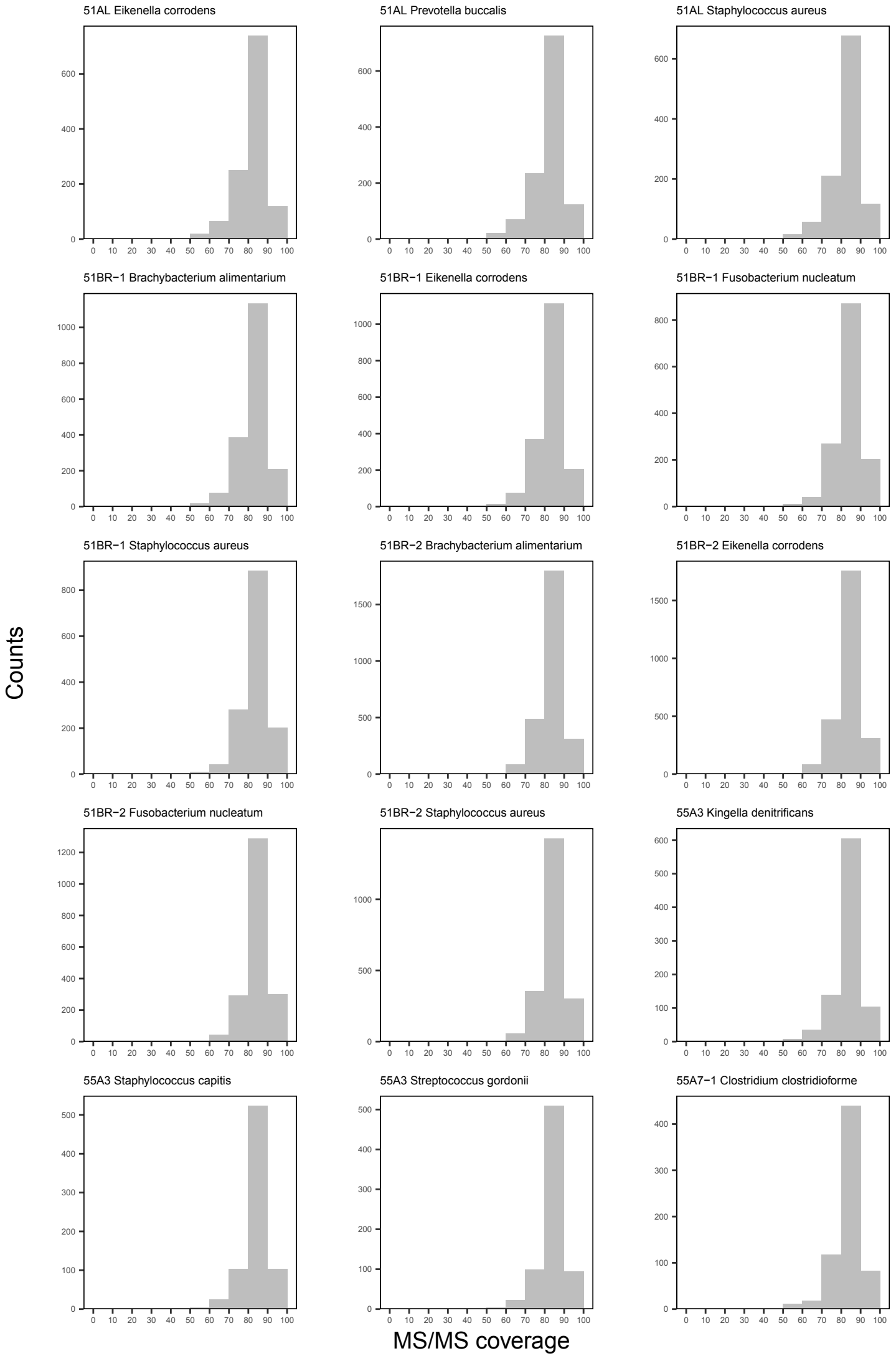
For example:
Peptide sequence: IAVCSITAL
Peptide length: 9 aa
Amino acids included in count: 8 aa (marked with arrow)
MS/MS coverage: 88.88%

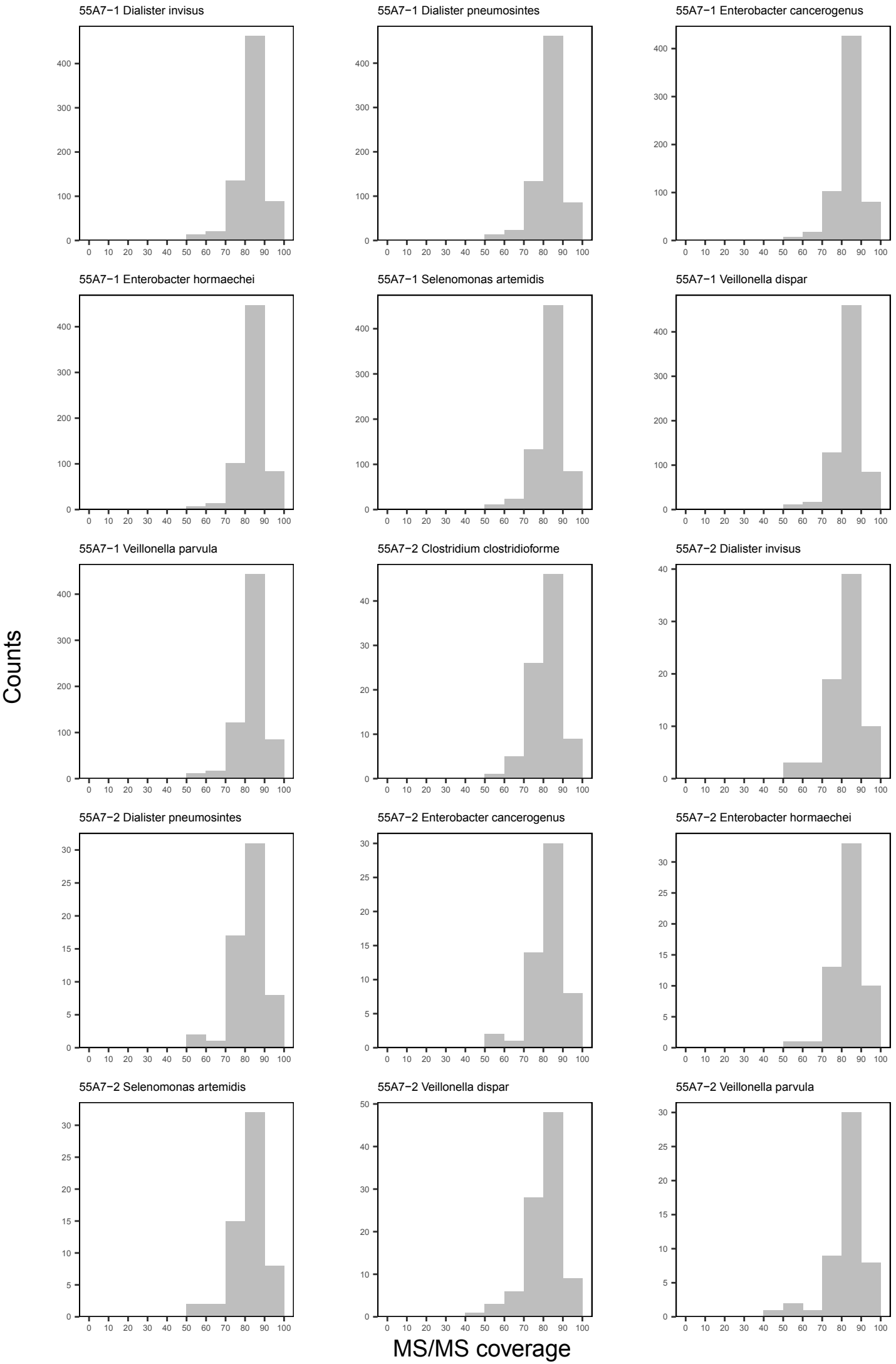


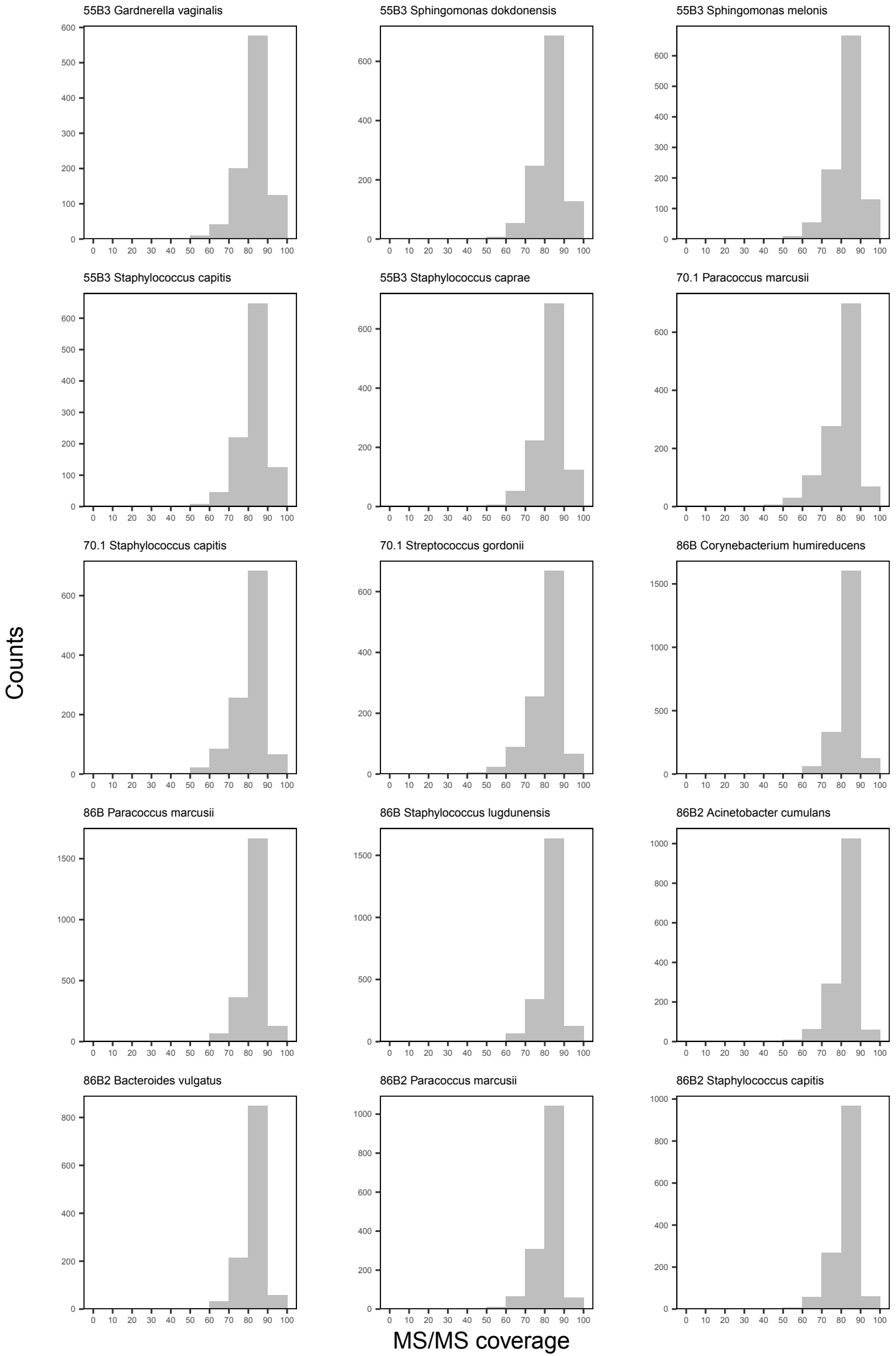
To determine a sufficient coverage threshold, we checked the expected coverage value of peptides with a high certainty of being correct HLA ligands, such as human-derived peptides found in the IEDB database or tumor-associated antigens that were found to be immunogenic in our samples.

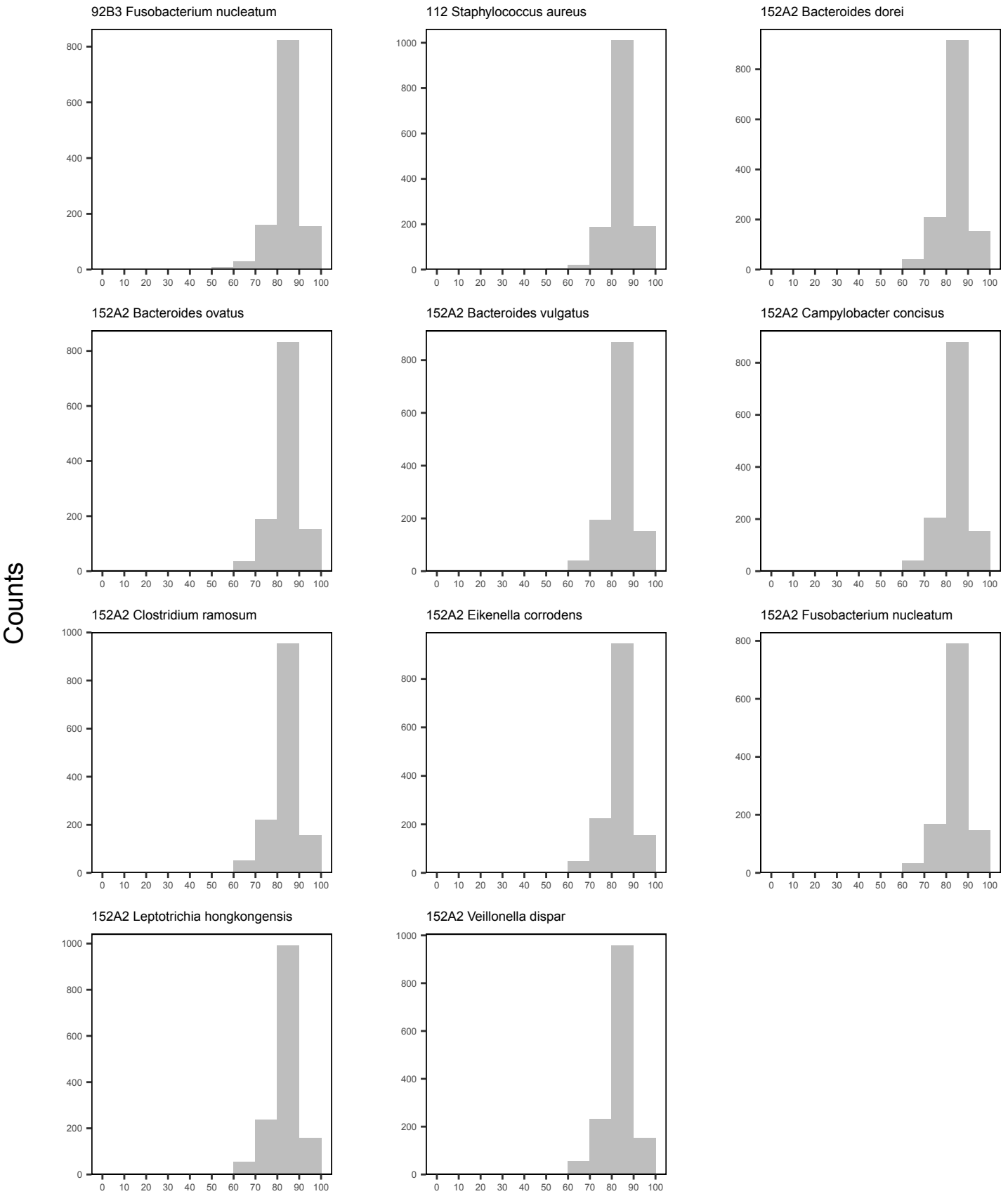
For each patient, we downloaded from the IEDB all the HLA peptides that matched the patient's HLA alleles. The parameters used for the search were: Epitope: "Any Epitopes"; Assay: "Positive assays only"; MHC restriction: the 4-digit HLA alleles of the patient; Host: "Humans". We then searched for these peptides in our list and kept only those that were also found in the IEDB list. We preformed this analysis separately for HLA-I and HLA-II peptides. We then calculated the MS/MS coverage for each of these peptides and plotted them, as shown below:

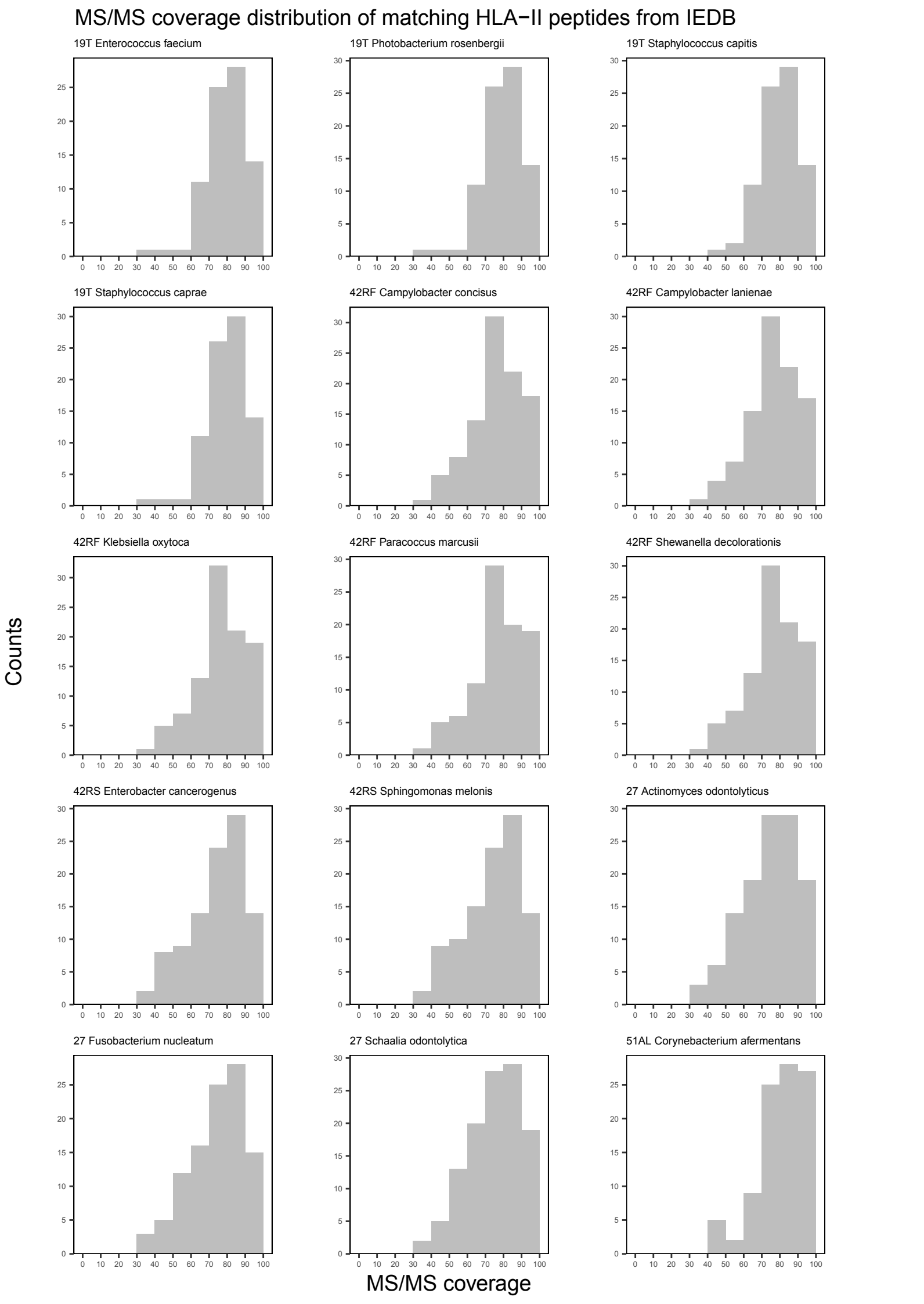

MS/MS coverage distribution of matching HLA−I peptides from IEDB
19T Enterococcus faecium
19T Photobacterium rosenbergii
19T Staphylococcus capitis
19T Staphylococcus caprae
42RF Campylobacter concisus
42RF Campylobacter lanienae
42RF Klebsiella oxytoca
42RF Paracoccus marcusii
42RF Shewanella decolorationis
42RS Enterobacter cancerogenus
42RS Sphingomonas melonis
27 Actinomyces odontolyticus
27 Fusobacterium nucleatum
27 Schaalia odontolytica
51AL Corynebacterium afermentans
Counts
MS/MS coverage

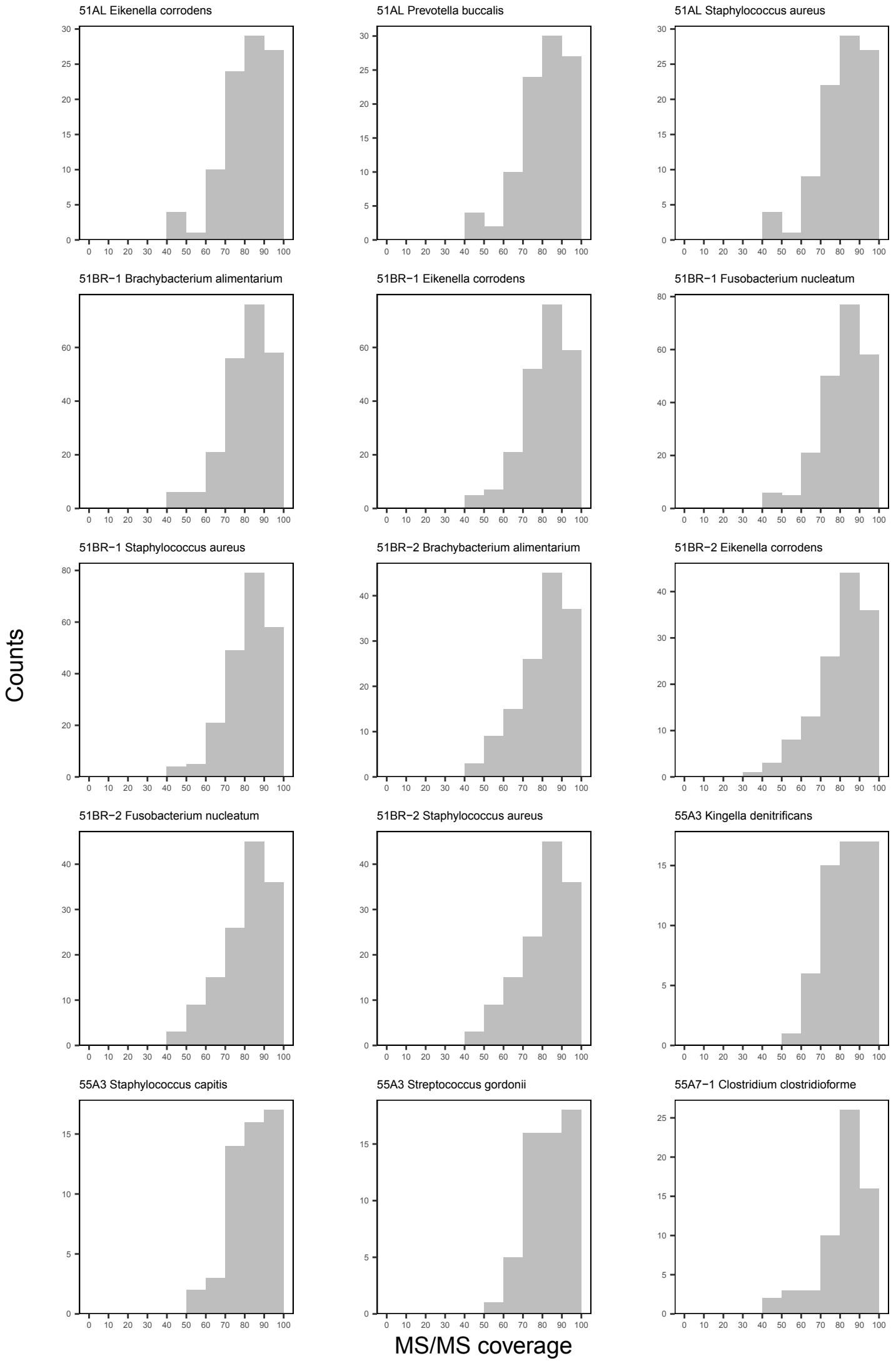

51AL Eikenella corrodens
51AL Prevotella buccalis
51AL Staphylococcus aureus
51BR−1 Brachybacterium alimentarium
51BR−1 Eikenella corrodens
51BR−1 Fusobacterium nucleatum
51BR−1 Staphylococcus aureus
51BR−2 Brachybacterium alimentarium
51BR−2 Eikenella corrodens
51BR−2 Fusobacterium nucleatum
51BR−2 Staphylococcus aureus
55A3 Kingella denitrificans
55A3 Staphylococcus capitis
55A3 Streptococcus gordonii
55A7−1 Clostridium clostridioforme
Counts
MS/MS coverage

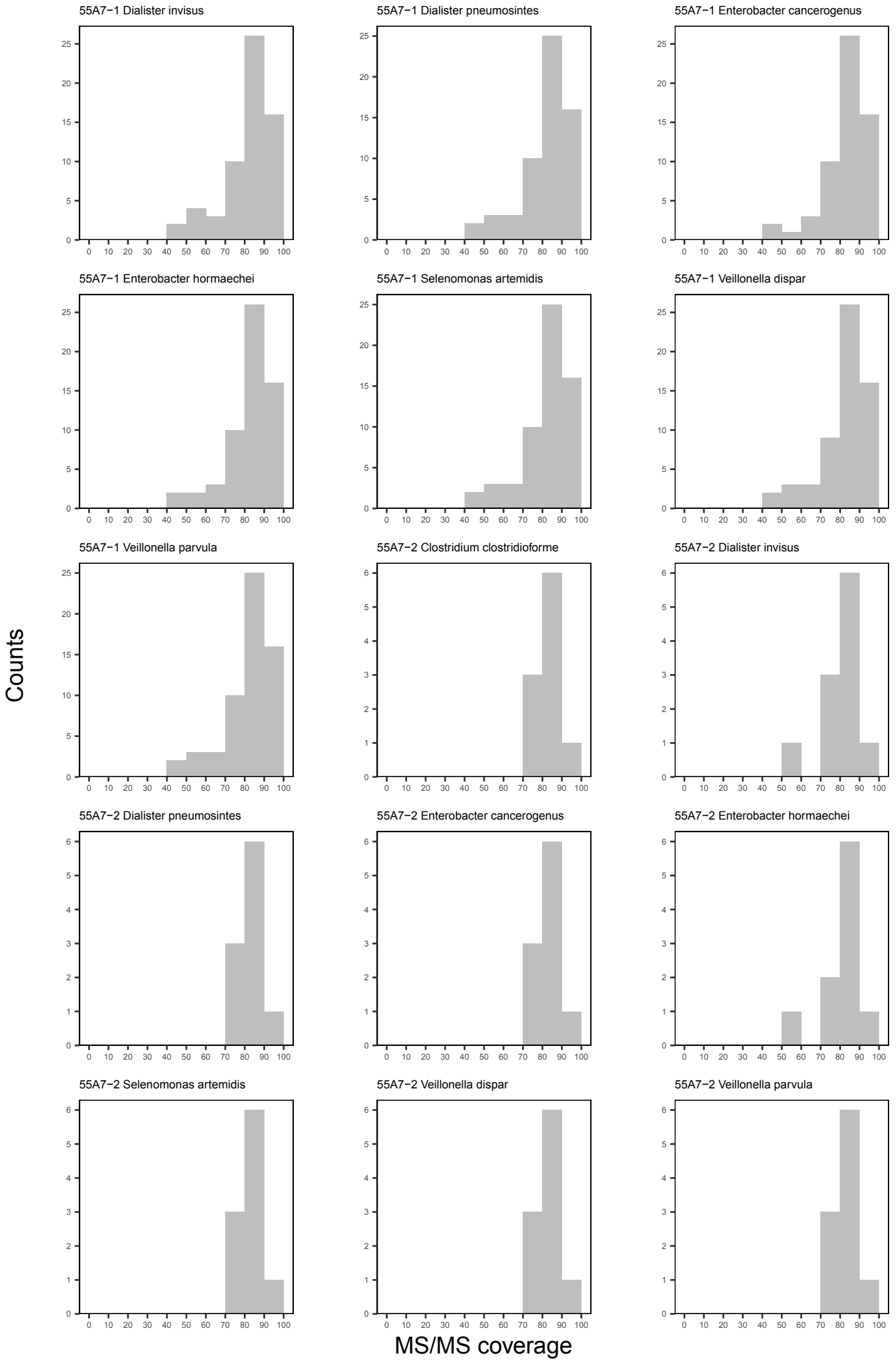

55A7−1 Dialister invisus
55A7−1 Dialister pneumosintes
55A7−1 Enterobacter cancerogenus
55A7−1 Enterobacter hormaechei
55A7−1 Selenomonas artemidis
55A7−1 Veillonella dispar
55A7−1 Veillonella parvula
55A7−2 Clostridium clostridioforme
55A7−2 Dialister invisus
55A7−2 Dialister pneumosintes
55A7−2 Enterobacter cancerogenus
55A7−2 Enterobacter hormaechei
55A7−2 Selenomonas artemidis
55A7−2 Veillonella dispar
55A7−2 Veillonella parvula
Counts
MS/MS coverage

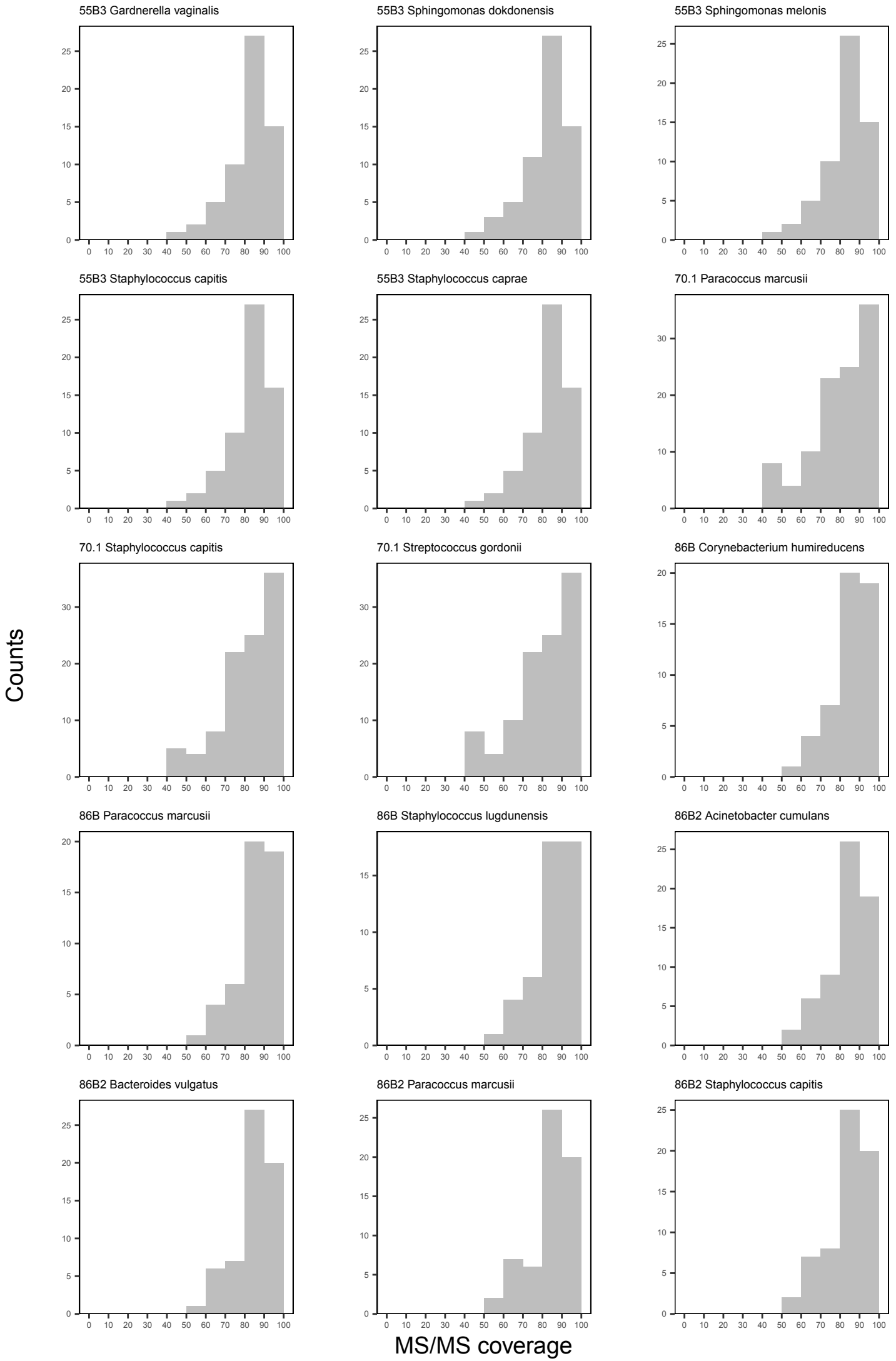

55B3 Gardnerella vaginalis
55B3 Sphingomonas dokdonensis
55B3 Sphingomonas melonis
55B3 Staphylococcus capitis
55B3 Staphylococcus caprae
70.1 Paracoccus marcusii
70.1 Staphylococcus capitis
70.1 Streptococcus gordonii
86B Corynebacterium humireducens
86B Paracoccus marcusii
86B Staphylococcus lugdunensis
86B2 Acinetobacter cumulans
86B2 Bacteroides vulgatus
86B2 Paracoccus marcusii
86B2 Staphylococcus capitis
Counts
MS/MS coverage

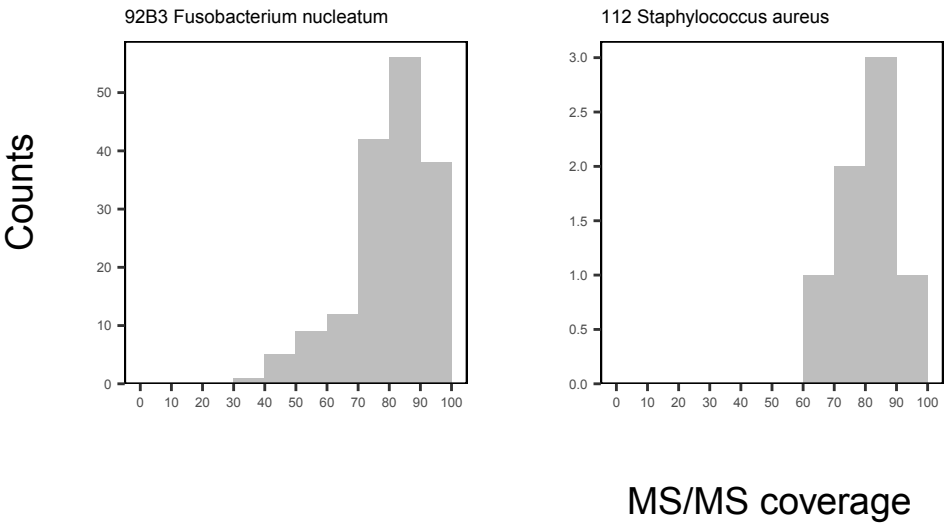

92B3 Fusobacterium nucleatum
112 Staphylococcus aureus
152A2 Bacteroides dorei
152A2 Bacteroides ovatus
152A2 Bacteroides vulgatus
152A2 Campylobacter concisus
152A2 Clostridium ramosum
152A2 Eikenella corrodens
152A2 Fusobacterium nucleatum
152A2 Leptotrichia hongkongensis
152A2 Veillonella dispar
Counts
MS/MS coverage


MS/MS coverage distribution of matching HLA−II peptides from IEDB
19T Enterococcus faecium
19T Photobacterium rosenbergii
19T Staphylococcus capitis
19T Staphylococcus caprae
42RF Campylobacter concisus
42RF Campylobacter lanienae
42RF Klebsiella oxytoca
42RF Paracoccus marcusii
42RF Shewanella decolorationis
42RS Enterobacter cancerogenus
42RS Sphingomonas melonis
27 Actinomyces odontolyticus
27 Fusobacterium nucleatum
27 Schaalia odontolytica
51AL Corynebacterium afermentans
Counts
MS/MS coverage


51AL Eikenella corrodens
51AL Prevotella buccalis
51AL Staphylococcus aureus
51BR−1 Brachybacterium alimentarium
51BR−1 Eikenella corrodens
51BR−1 Fusobacterium nucleatum
51BR−1 Staphylococcus aureus
51BR−2 Brachybacterium alimentarium
51BR−2 Eikenella corrodens
51BR−2 Fusobacterium nucleatum
51BR−2 Staphylococcus aureus
55A3 Kingella denitrificans
55A3 Staphylococcus capitis
55A3 Streptococcus gordonii
55A7−1 Clostridium clostridioforme
Counts
MS/MS coverage

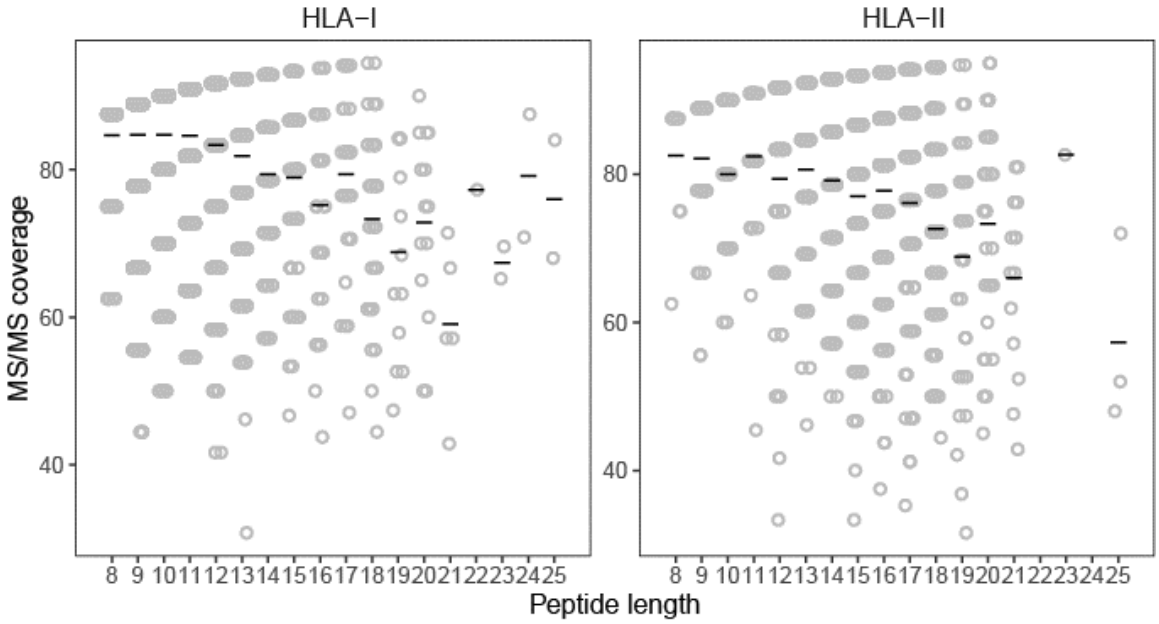

55A7−1 Dialister invisus
55A7−1 Dialister pneumosintes
55A7−1 Enterobacter cancerogenus
55A7−1 Enterobacter hormaechei
55A7−1 Selenomonas artemidis
55A7−1 Veillonella dispar
55A7−1 Veillonella parvula
55A7−2 Clostridium clostridioforme
55A7−2 Dialister invisus
55A7−2 Dialister pneumosintes
55A7−2 Enterobacter cancerogenus
55A7−2 Enterobacter hormaechei
55A7−2 Selenomonas artemidis
55A7−2 Veillonella dispar
55A7−2 Veillonella parvula
Counts
MS/MS coverage


55B3 Gardnerella vaginalis
55B3 Sphingomonas dokdonensis
55B3 Sphingomonas melonis
55B3 Staphylococcus capitis
55B3 Staphylococcus caprae
70.1 Paracoccus marcusii
70.1 Staphylococcus capitis
70.1 Streptococcus gordonii
86B Corynebacterium humireducens
86B Paracoccus marcusii
86B Staphylococcus lugdunensis
86B2 Acinetobacter cumulans
86B2 Bacteroides vulgatus
86B2 Paracoccus marcusii
86B2 Staphylococcus capitis
Counts
MS/MS coverage

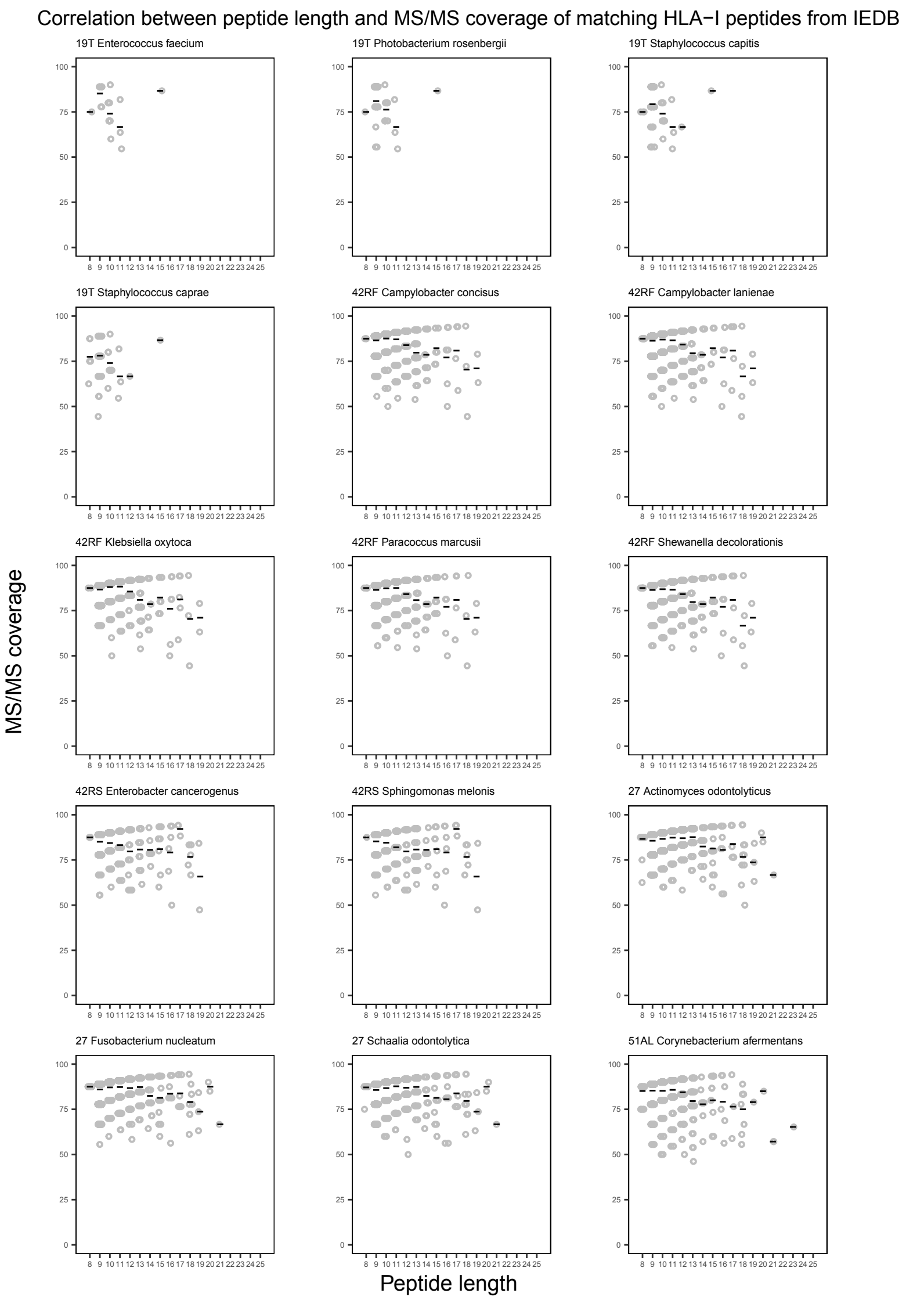

92B3 Fusobacterium nucleatum
112 Staphylococcus aureus
Counts
MS/MS coverage
0
10
20
30
40
50
60
70
80
90
100
0.0
0.5
1.0
1.5
2.0
2.5
3.0

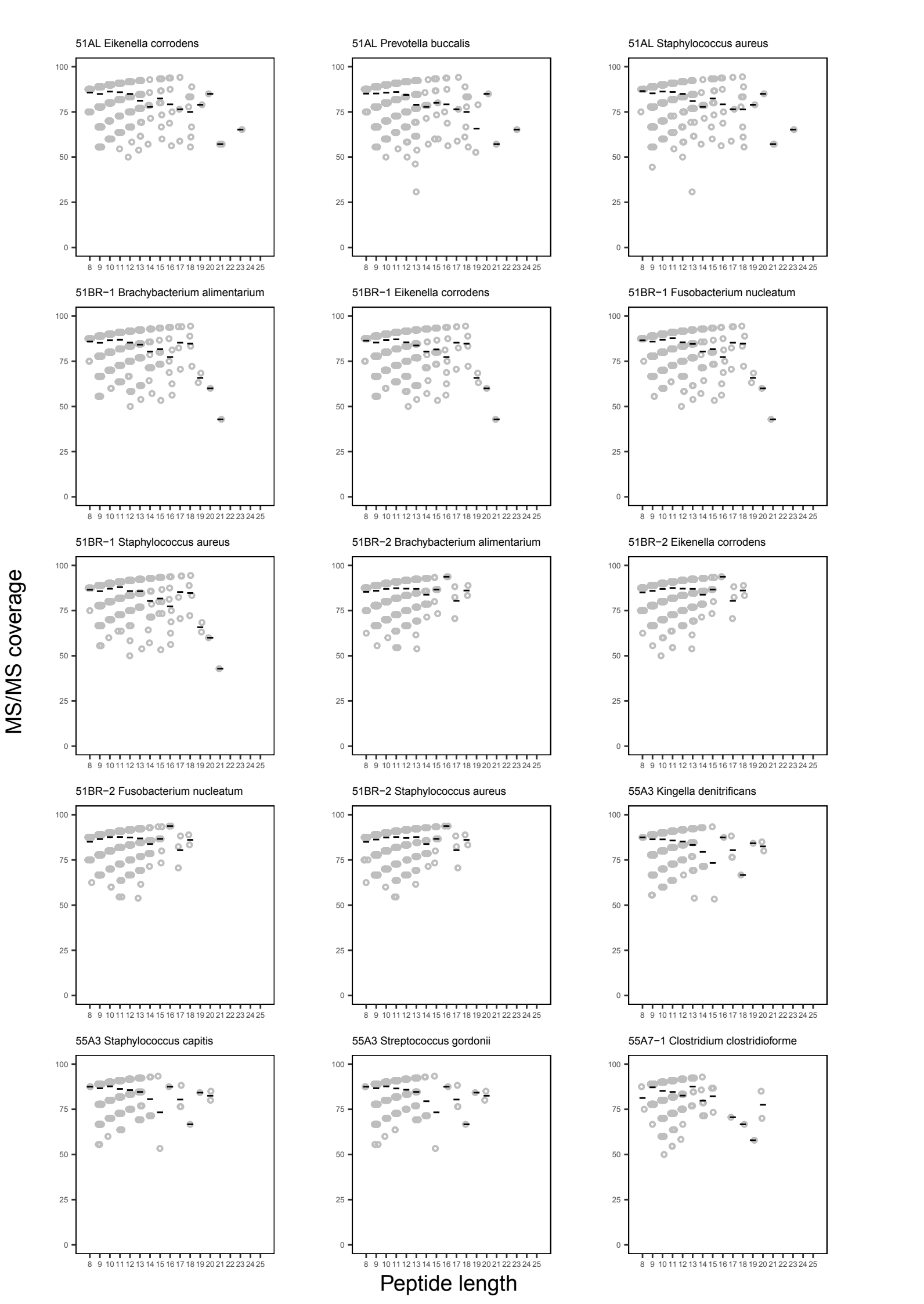
According to this analysis, the threshold for HLA-I peptides is a coverage of >60%, and for HLA-II peptides, >40%. The difference in the observed thresholds for HLA-I and HLA-II peptides is probably due to variations in peptide length. Longer peptides, such as HLA-II peptides, probably produce fewer ions compared to their length. As can be seen below, MS/MS coverage correlates with peptide length:

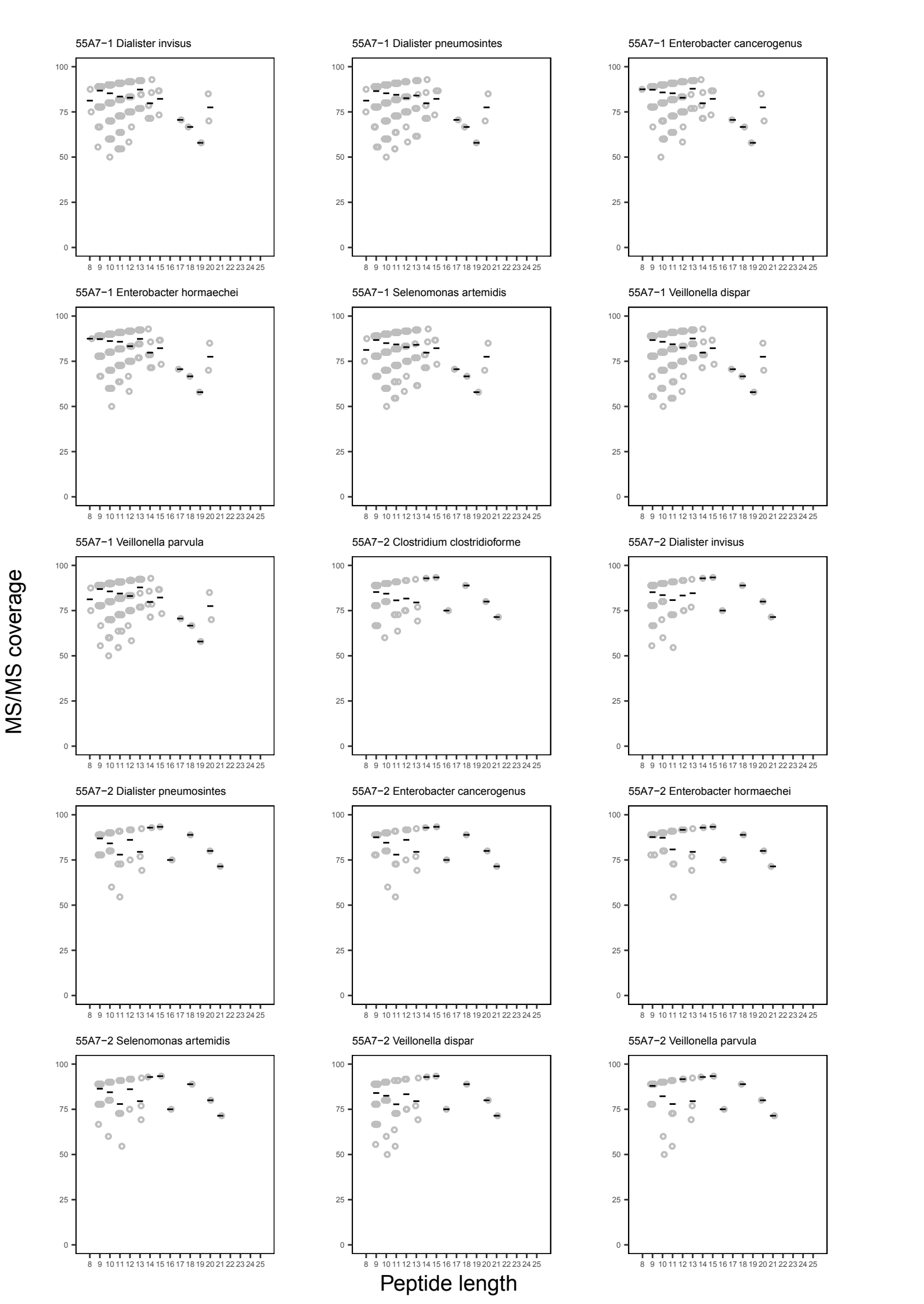
All the identified bacterial peptides together:

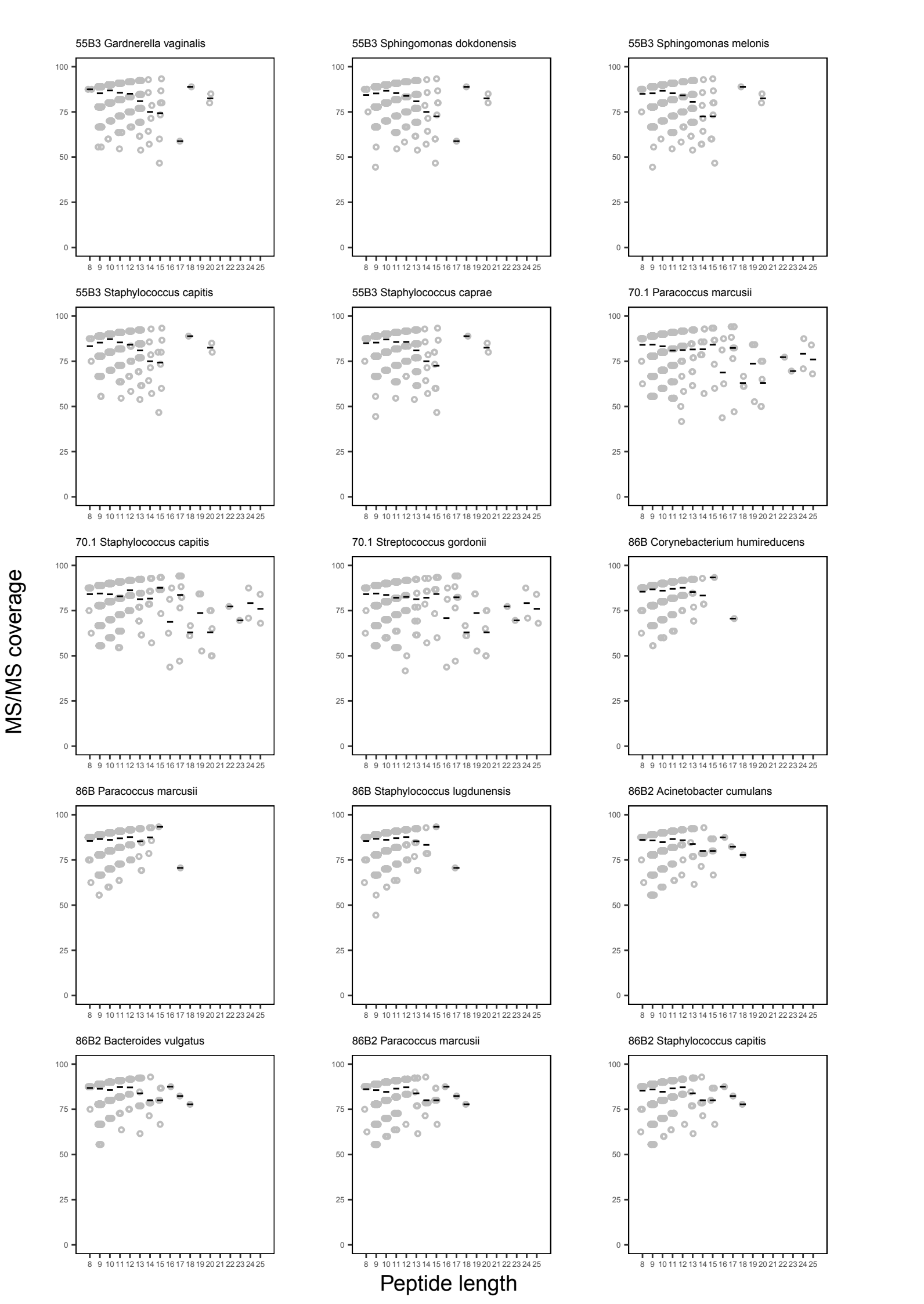


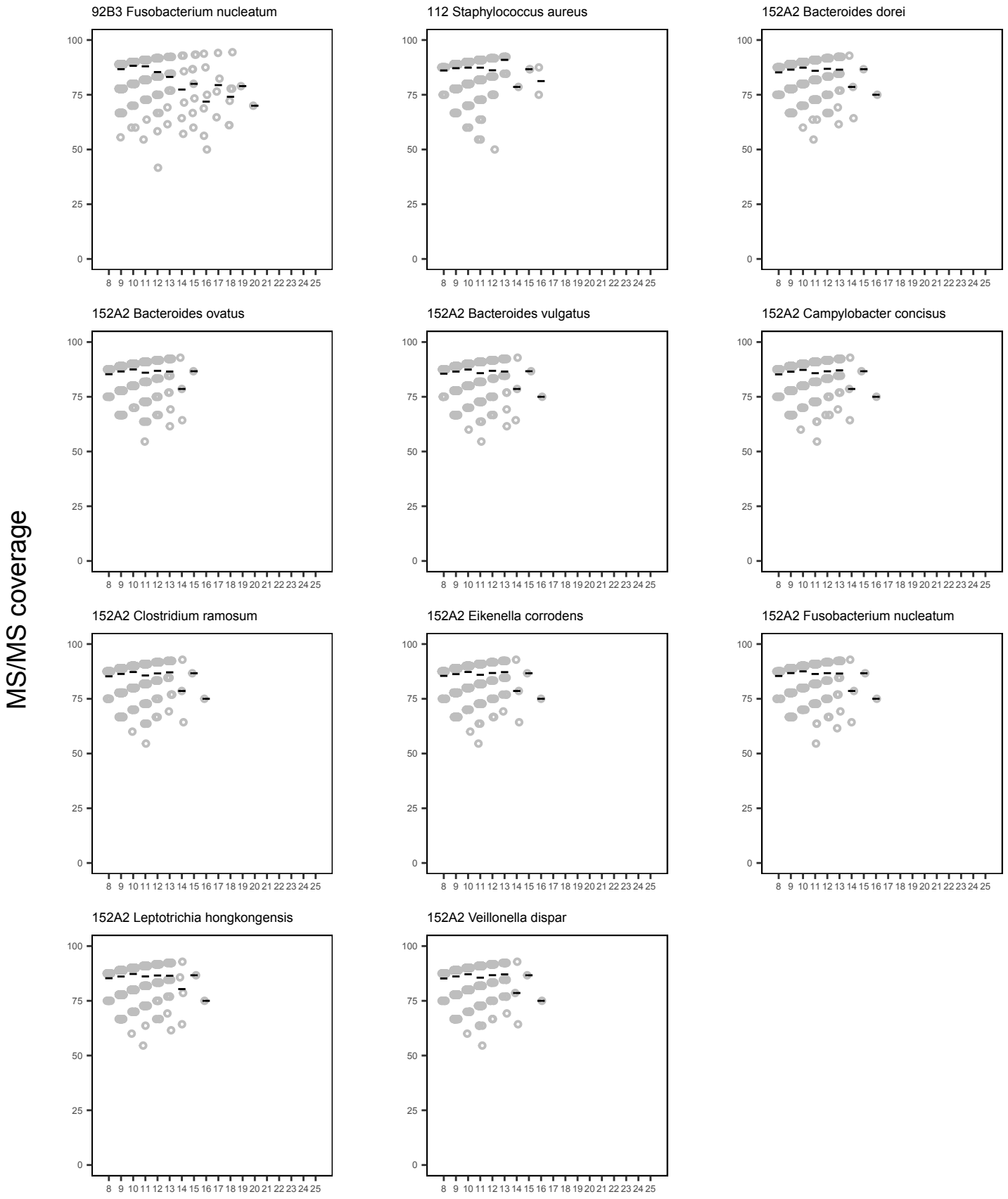
As well as for each sample separately:

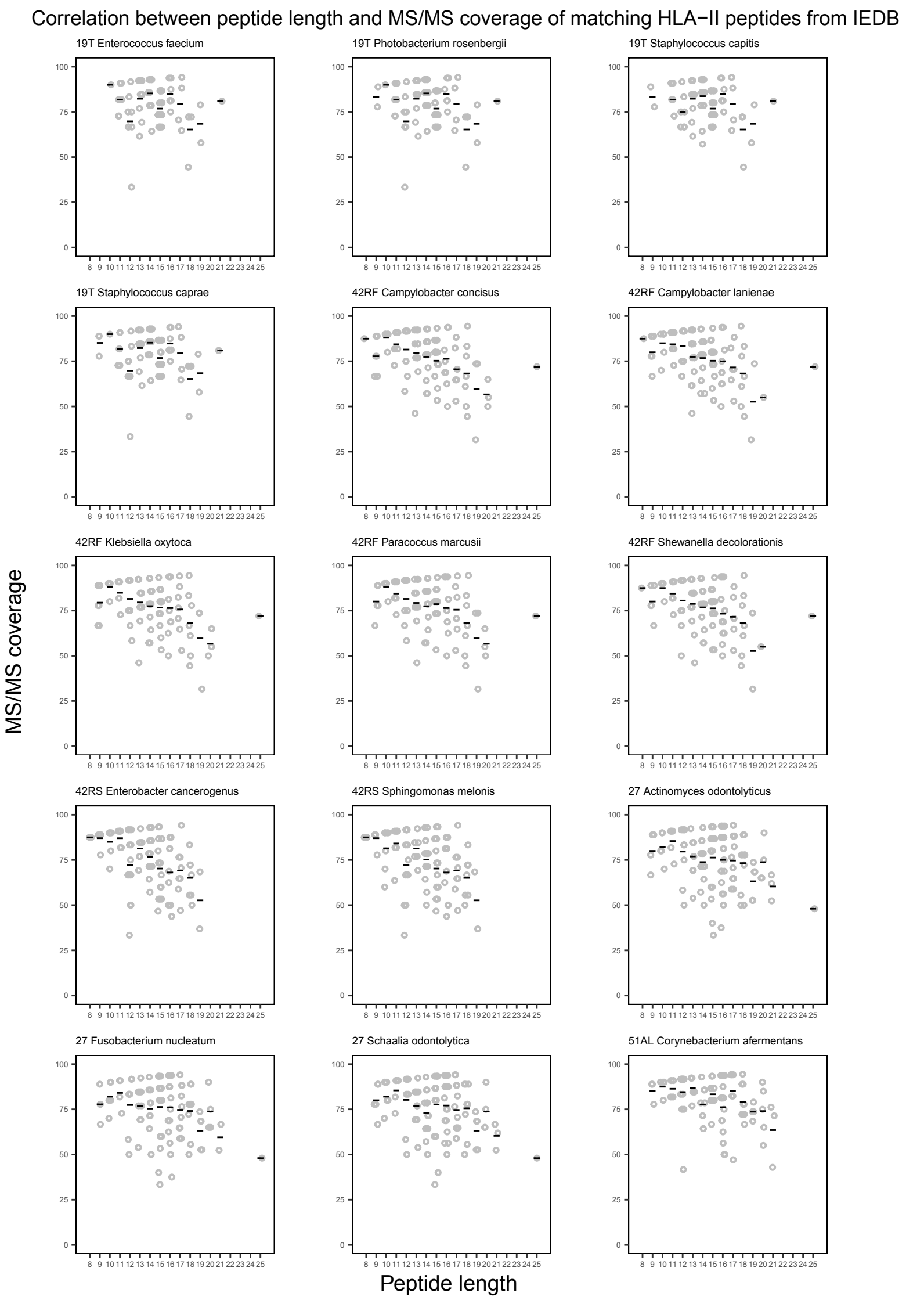

Correlation between peptide length and MS/MS coverage of matching HLA−I peptides from IEDB
19T Enterococcus faecium
19T Photobacterium rosenbergii
19T Staphylococcus capitis
19T Staphylococcus caprae
42RF Campylobacter concisus
42RF Campylobacter lanienae
42RF Klebsiella oxytoca
42RF Paracoccus marcusii
42RF Shewanella decolorationis
42RS Enterobacter cancerogenus
42RS Sphingomonas melonis
27 Actinomyces odontolyticus
27 Fusobacterium nucleatum
27 Schaalia odontolytica
51AL Corynebacterium afermentans
MS/MS coverage
Peptide length

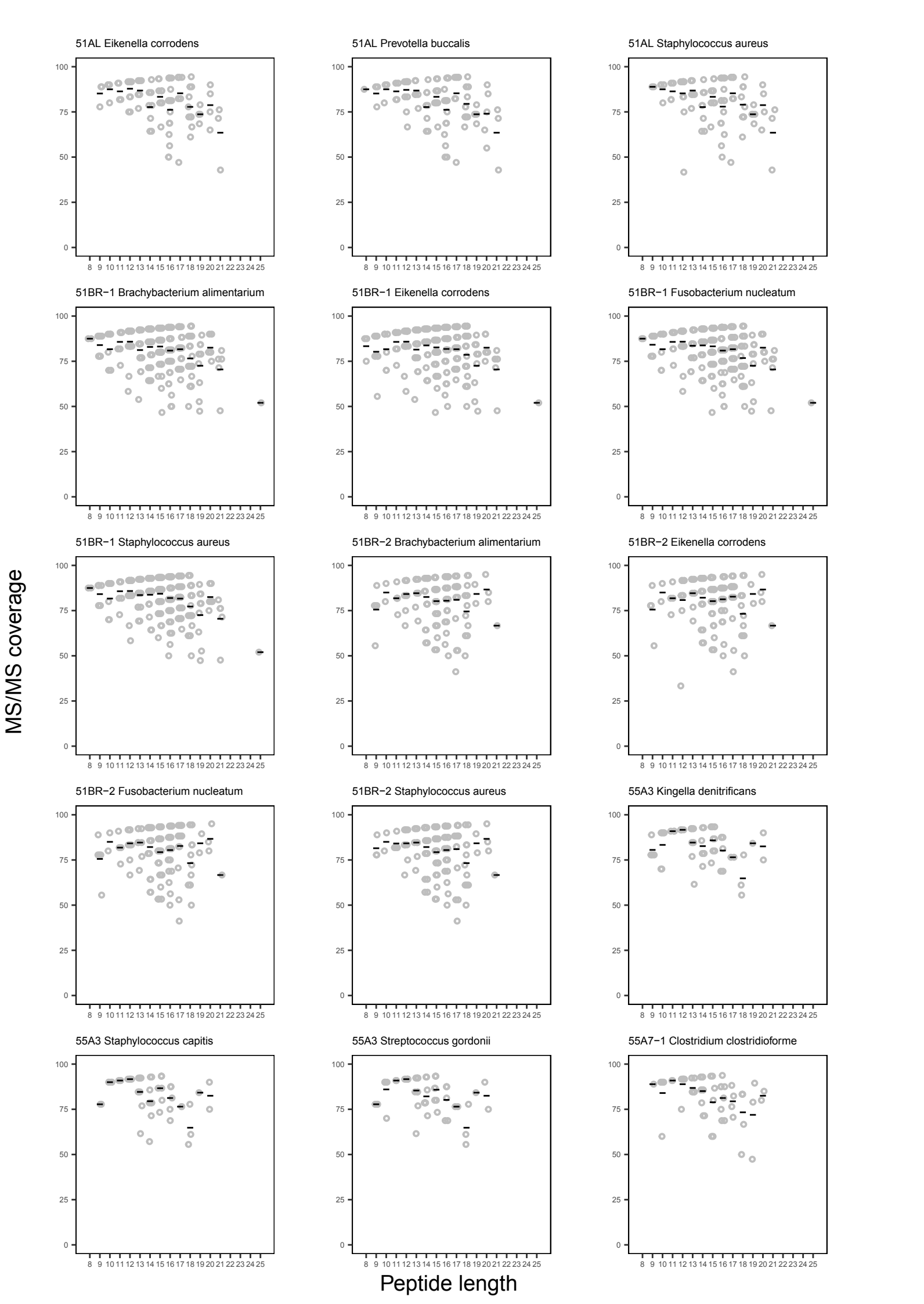

51AL Eikenella corrodens
51AL Prevotella buccalis
51AL Staphylococcus aureus
51BR−1 Brachybacterium alimentarium
51BR−1 Eikenella corrodens
51BR−1 Fusobacterium nucleatum
51BR−1 Staphylococcus aureus
51BR−2 Brachybacterium alimentarium
51BR−2 Eikenella corrodens
51BR−2 Fusobacterium nucleatum
51BR−2 Staphylococcus aureus
55A3 Kingella denitrificans
55A3 Staphylococcus capitis
55A3 Streptococcus gordonii
55A7−1 Clostridium clostridioforme
MS/MS coverage
Peptide length

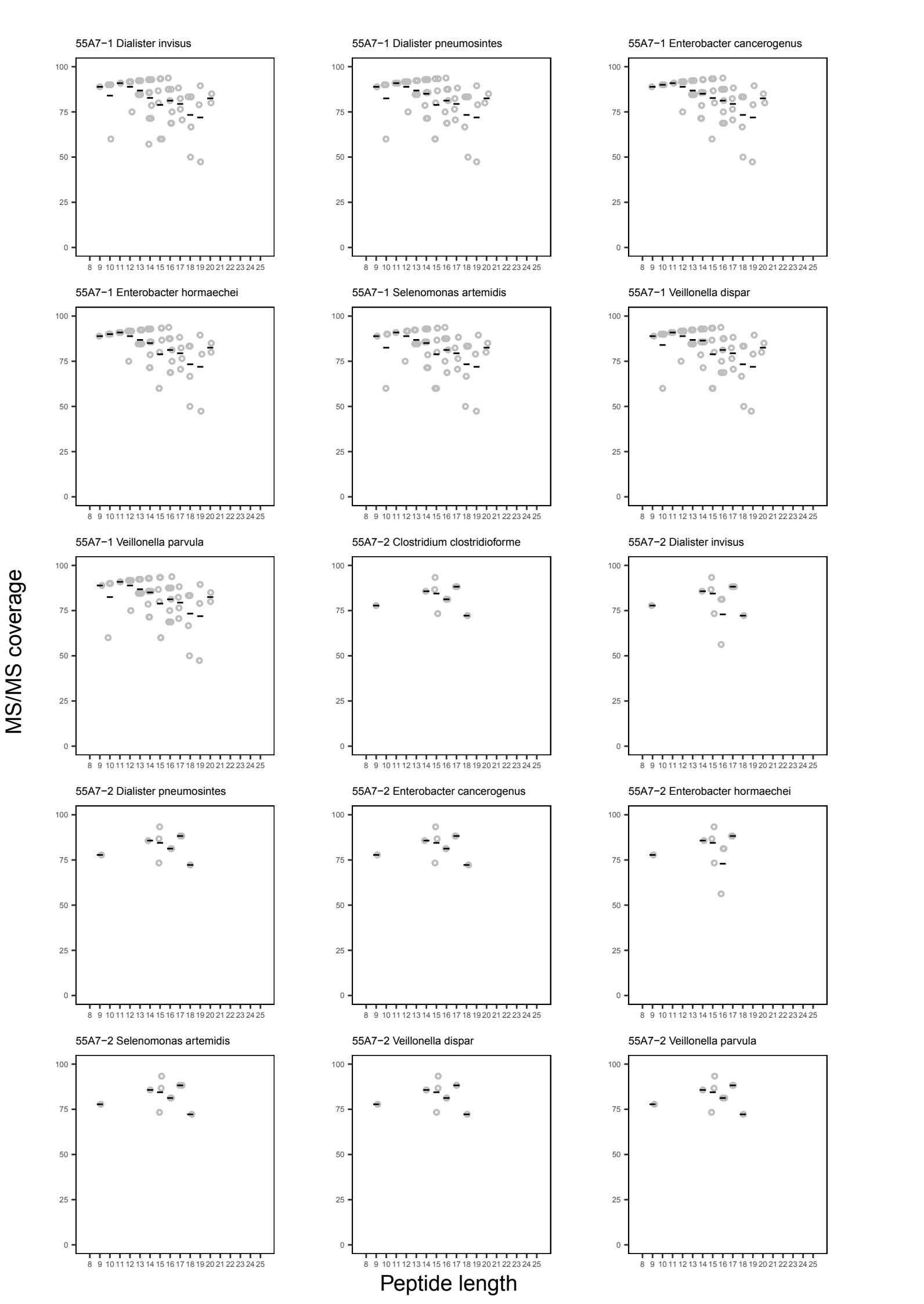

55A7−1 Dialister invisus
55A7−1 Dialister pneumosintes
55A7−1 Enterobacter cancerogenus
55A7−1 Enterobacter hormaechei
55A7−1 Selenomonas artemidis
55A7−1 Veillonella dispar
55A7−1 Veillonella parvula
55A7−2 Clostridium clostridioforme
55A7−2 Dialister invisus
55A7−2 Dialister pneumosintes
55A7−2 Enterobacter cancerogenus
55A7−2 Enterobacter hormaechei
55A7−2 Selenomonas artemidis
55A7−2 Veillonella dispar
55A7−2 Veillonella parvula
MS/MS coverage
Peptide length

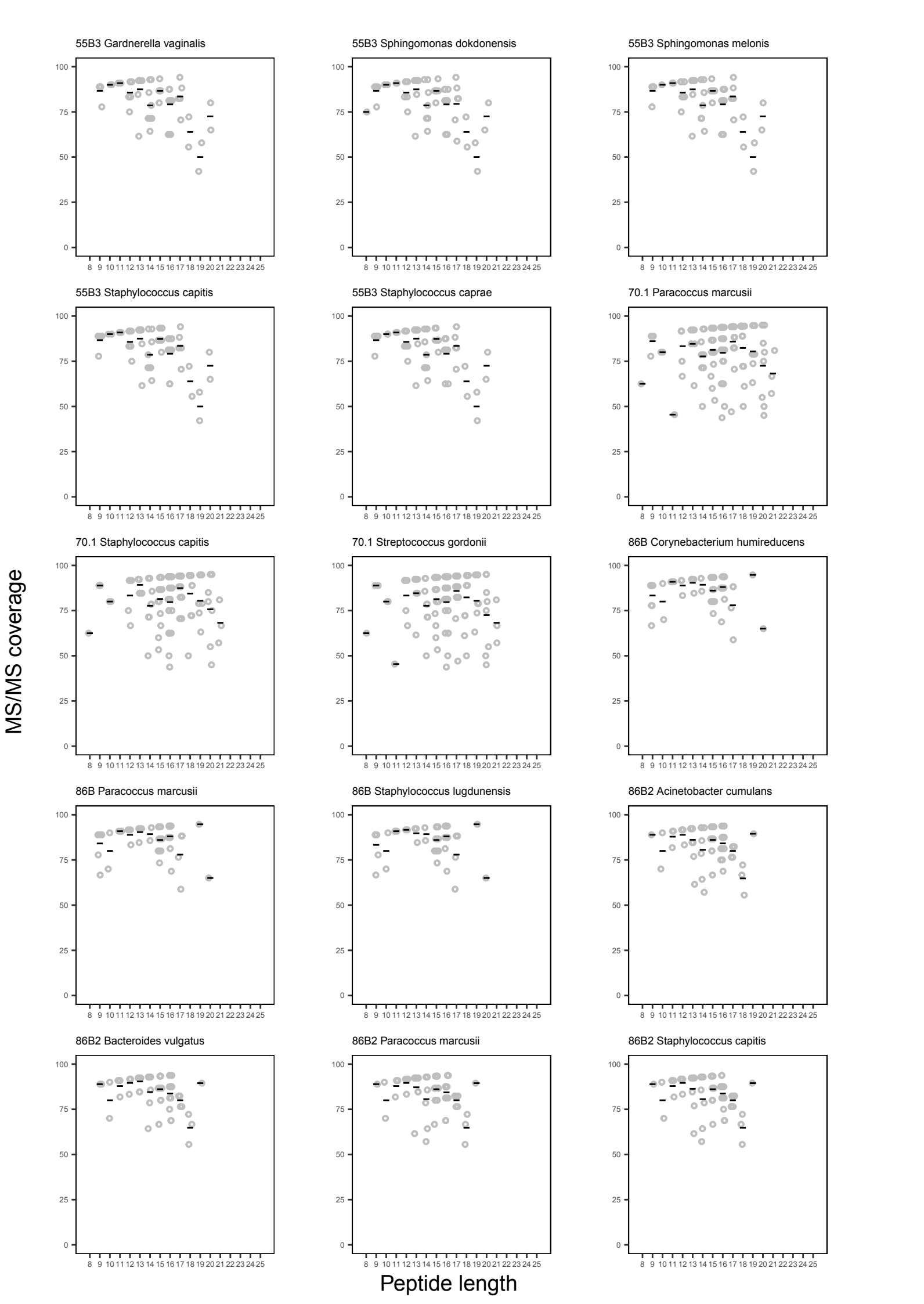

55B3 Gardnerella vaginalis
55B3 Sphingomonas dokdonensis
55B3 Sphingomonas melonis
55B3 Staphylococcus capitis
55B3 Staphylococcus caprae
70.1 Paracoccus marcusii
70.1 Staphylococcus capitis
70.1 Streptococcus gordonii
86B Corynebacterium humireducens
86B Paracoccus marcusii
86B Staphylococcus lugdunensis
86B2 Acinetobacter cumulans
86B2 Bacteroides vulgatus
86B2 Paracoccus marcusii
86B2 Staphylococcus capitis
MS/MS coverage
Peptide length

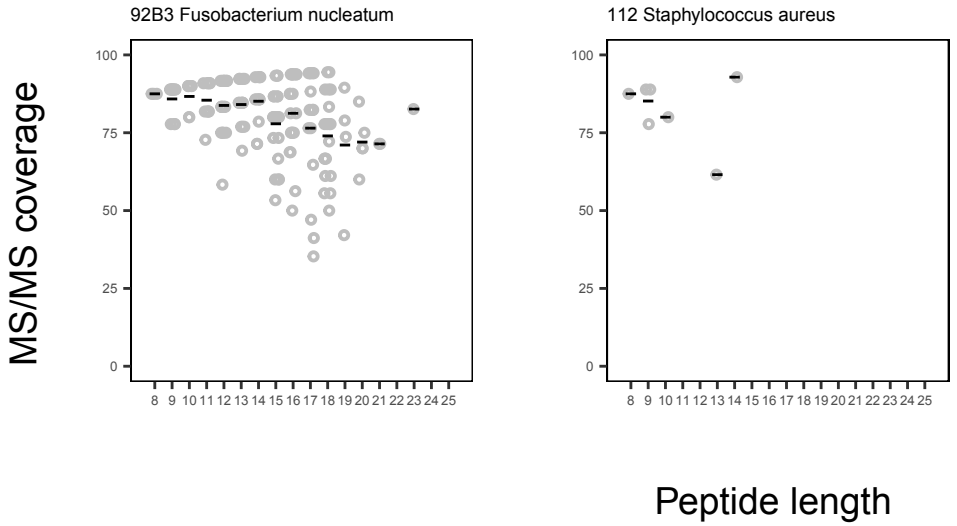

92B3 Fusobacterium nucleatum
112 Staphylococcus aureus
152A2 Bacteroides dorei
152A2 Bacteroides ovatus
152A2 Bacteroides vulgatus
152A2 Campylobacter concisus
152A2 Clostridium ramosum
152A2 Eikenella corrodens
152A2 Fusobacterium nucleatum
152A2 Leptotrichia hongkongensis
152A2 Veillonella dispar
MS/MS coverage
Peptide length


Correlation between peptide length and MS/MS coverage of matching HLA−II peptides from IEDB
19T Enterococcus faecium
19T Photobacterium rosenbergii
19T Staphylococcus capitis
19T Staphylococcus caprae
42RF Campylobacter concisus
42RF Campylobacter lanienae
42RF Klebsiella oxytoca
42RF Paracoccus marcusii
42RF Shewanella decolorationis
42RS Enterobacter cancerogenus
42RS Sphingomonas melonis
27 Actinomyces odontolyticus
27 Fusobacterium nucleatum
27 Schaalia odontolytica
51AL Corynebacterium afermentans
MS/MS coverage
Peptide length


51AL Eikenella corrodens
51AL Prevotella buccalis
51AL Staphylococcus aureus
51BR−1 Brachybacterium alimentarium
51BR−1 Eikenella corrodens
51BR−1 Fusobacterium nucleatum
51BR−1 Staphylococcus aureus
51BR−2 Brachybacterium alimentarium
51BR−2 Eikenella corrodens
51BR−2 Fusobacterium nucleatum
51BR−2 Staphylococcus aureus
55A3 Kingella denitrificans
55A3 Staphylococcus capitis
55A3 Streptococcus gordonii
55A7−1 Clostridium clostridioforme
MS/MS coverage
Peptide length

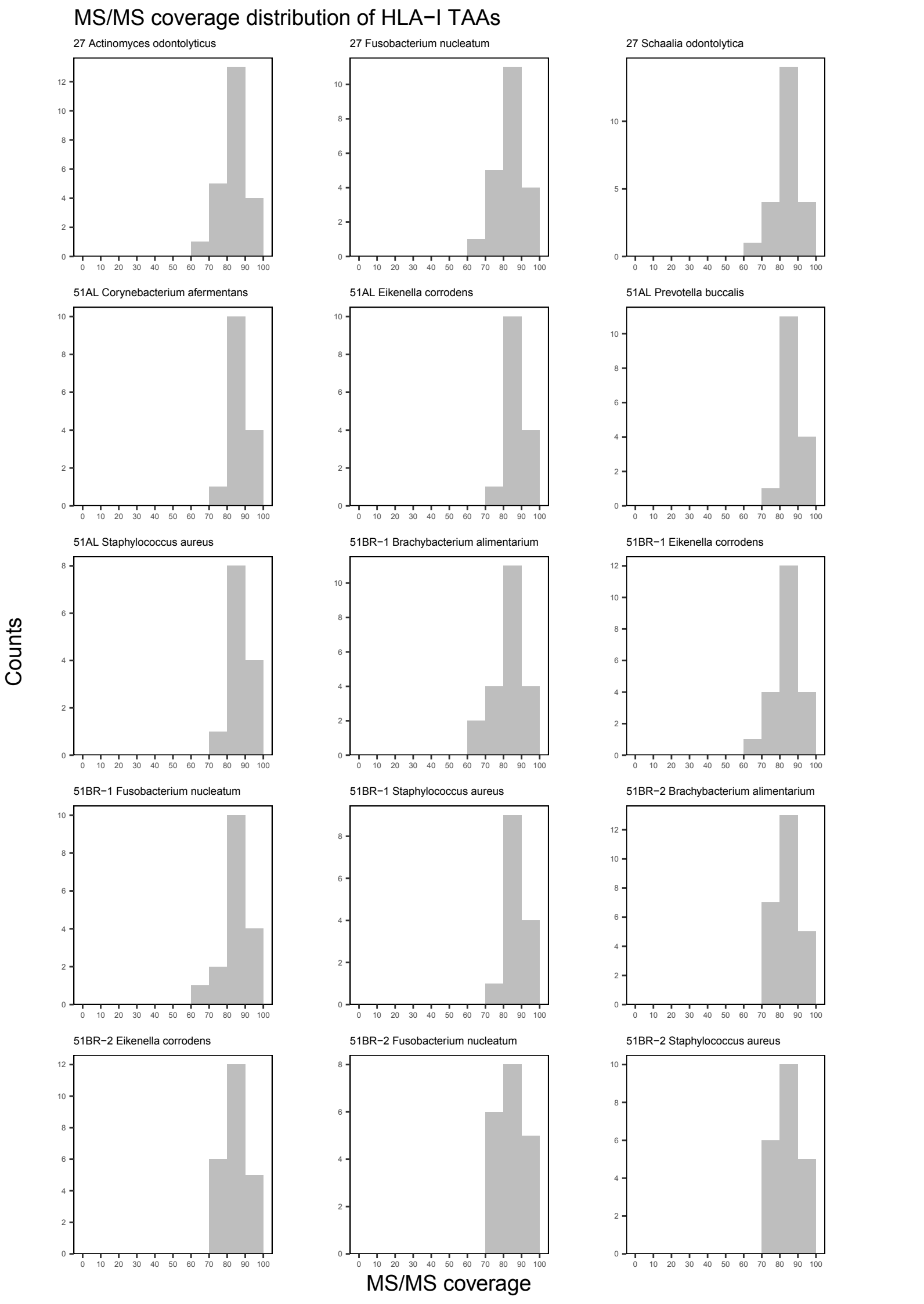

55A7−1 Dialister invisus
55A7−1 Dialister pneumosintes
55A7−1 Enterobacter cancerogenus
55A7−1 Enterobacter hormaechei
55A7−1 Selenomonas artemidis
55A7−1 Veillonella dispar
55A7−1 Veillonella parvula
55A7−2 Clostridium clostridioforme
55A7−2 Dialister invisus
55A7−2 Dialister pneumosintes
55A7−2 Enterobacter cancerogenus
55A7−2 Enterobacter hormaechei
55A7−2 Selenomonas artemidis
55A7−2 Veillonella dispar
55A7−2 Veillonella parvula
MS/MS coverage
Peptide length

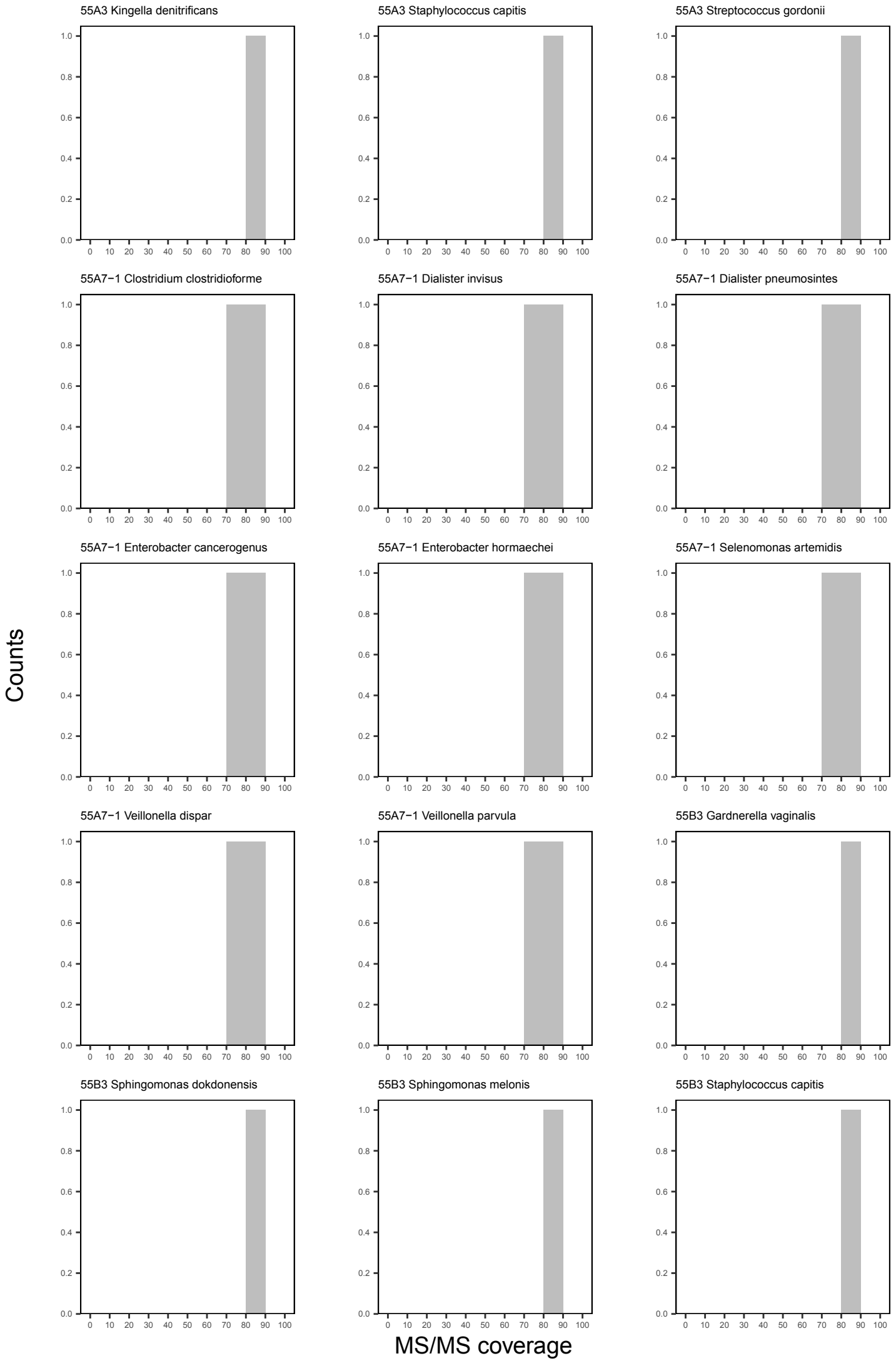

55B3 Gardnerella vaginalis
55B3 Sphingomonas dokdonensis
55B3 Sphingomonas melonis
55B3 Staphylococcus capitis
55B3 Staphylococcus caprae
70.1 Paracoccus marcusii
70.1 Staphylococcus capitis
70.1 Streptococcus gordonii
86B Corynebacterium humireducens
86B Paracoccus marcusii
86B Staphylococcus lugdunensis
86B2 Acinetobacter cumulans
86B2 Bacteroides vulgatus
86B2 Paracoccus marcusii
86B2 Staphylococcus capitis
MS/MS coverage
Peptide length

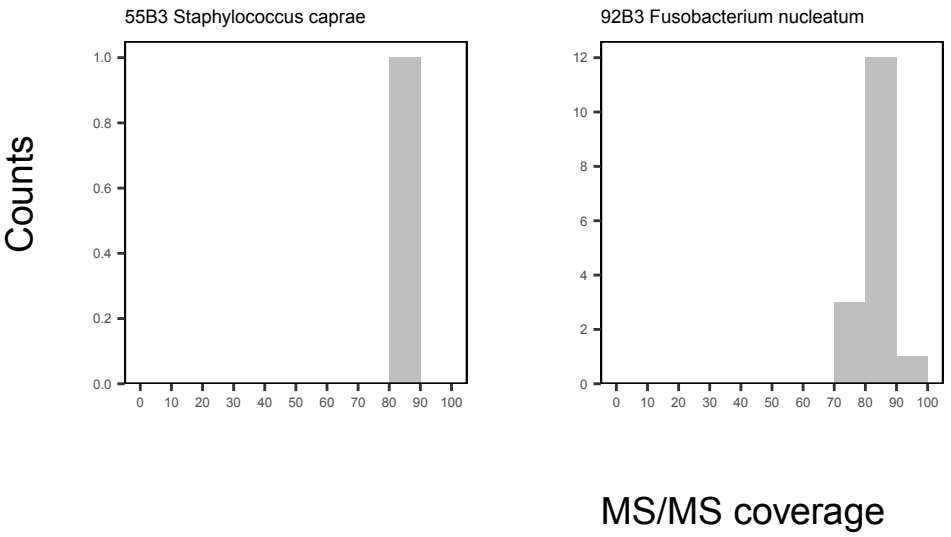

92B3 Fusobacterium nucleatum
112 Staphylococcus aureus
MS/MS coverage
100
75
50
25
0
8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
Peptide length

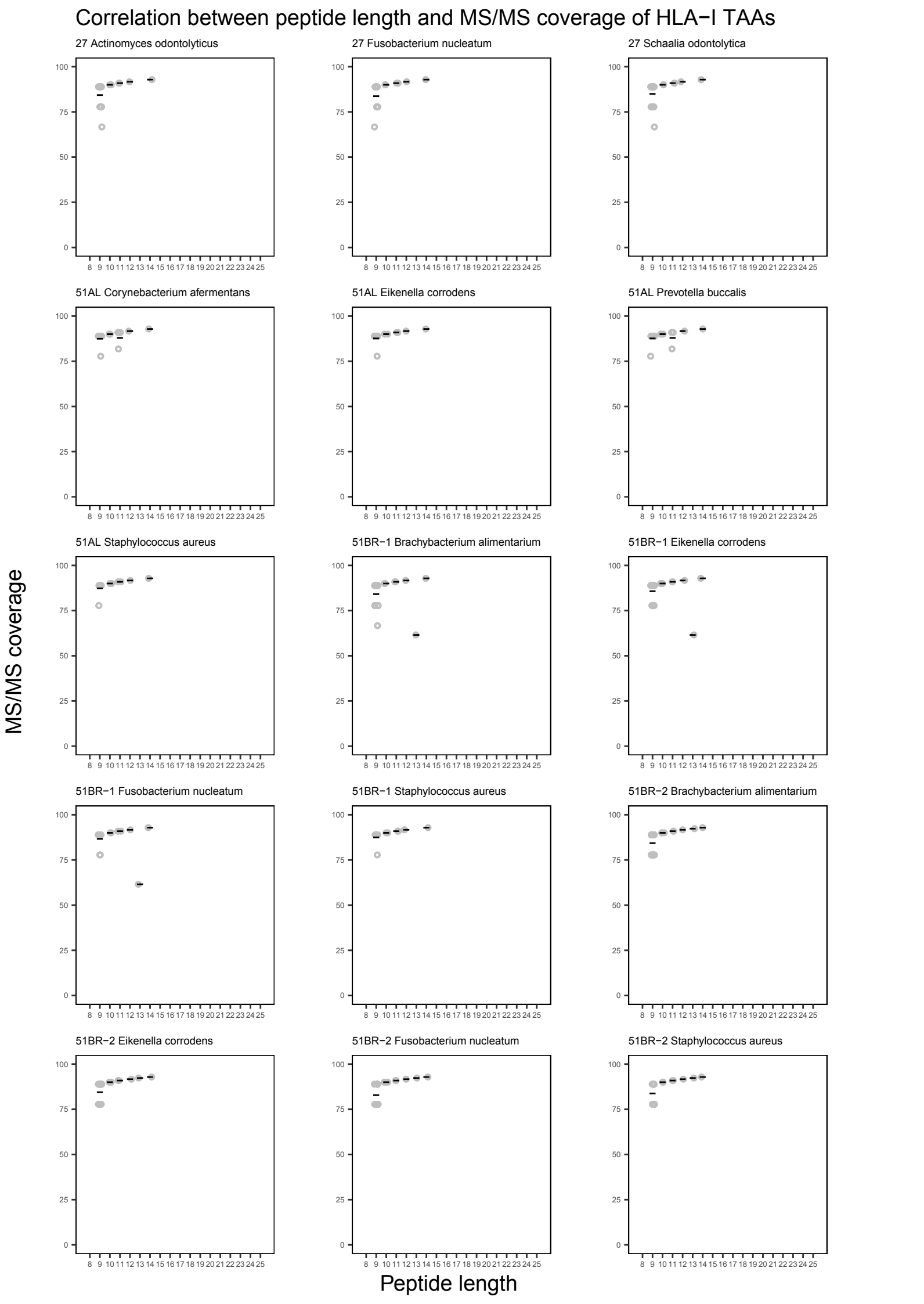
To further validate the selected threshold for HLA-I peptides, we used our prior knowledge of validated tumor associated antigens (TAAs) that were tested for reactivity with the patient's TILs [1]. As some of the samples used in this study overlapped with those identified in our previous study, we knew which of the identified human peptides are immunogenic TAAs. In the previous study, we analyzed the same raw data only with the human proteome. Here, we analyzed the human proteome together with the proteome of the bacterial species identified in the tumor using 16S rRNA sequencing. As these TAA peptides were found to be immunogenic when tested with the patient's autologous TILs, we knew they were accurately identified and, therefore, had a higher certainty of being correct. We found their MS/MS coverage to be above >60%, strengthening our threshold selection of >60% coverage for HLA-I peptides.

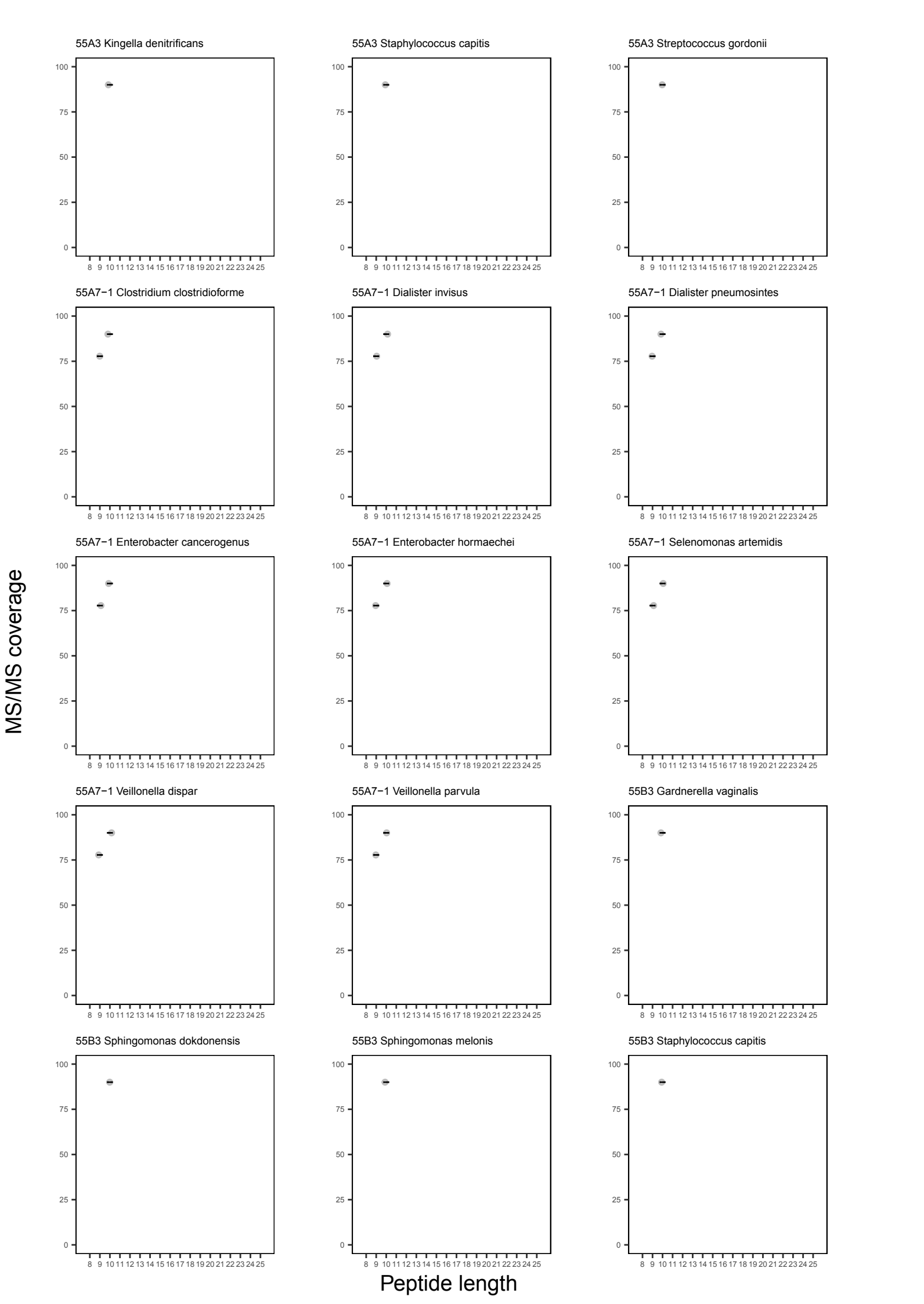
The MS/MS distribution of these peptides in each analysis is as follows:

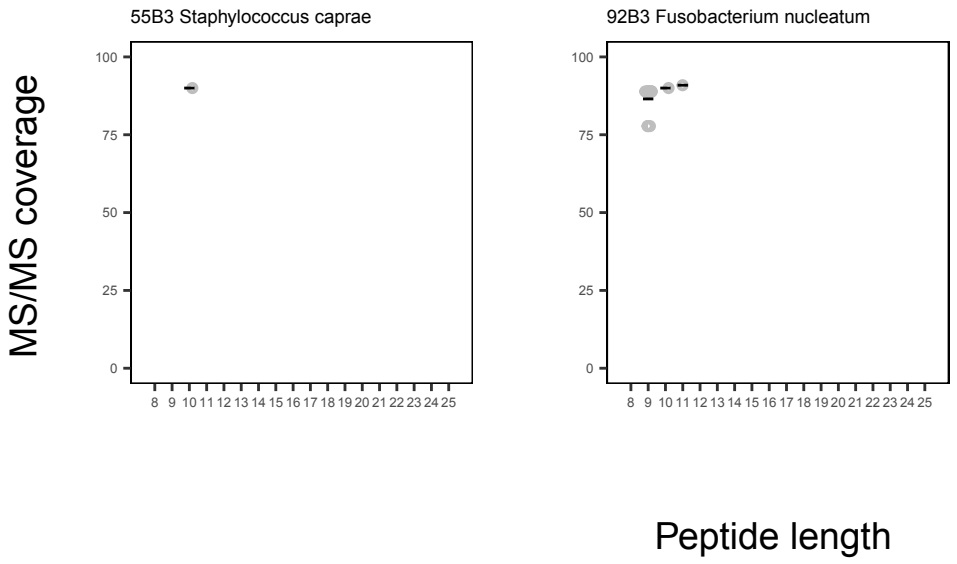

MS/MS coverage distribution of HLA−I TAAs
27 Actinomyces odontolyticus
27 Fusobacterium nucleatum
27 Schaalia odontolytica
51AL Corynebacterium afermentans
51AL Eikenella corrodens
51AL Prevotella buccalis
51AL Staphylococcus aureus
51BR−1 Brachybacterium alimentarium
51BR−1 Eikenella corrodens
51BR−1 Fusobacterium nucleatum
51BR−1 Staphylococcus aureus
51BR−2 Brachybacterium alimentarium
51BR−2 Eikenella corrodens
51BR−2 Fusobacterium nucleatum
51BR−2 Staphylococcus aureus
Counts
MS/MS coverage


55A3 Kingella denitrificans
55A3 Staphylococcus capitis
55A3 Streptococcus gordonii
55A7−1 Clostridium clostridioforme
55A7−1 Dialister invisus
55A7−1 Dialister pneumosintes
55A7−1 Enterobacter cancerogenus
55A7−1 Enterobacter hormaechei
55A7−1 Selenomonas artemidis
55A7−1 Veillonella dispar
55A7−1 Veillonella parvula
55B3 Gardnerella vaginalis
55B3 Sphingomonas dokdonensis
55B3 Sphingomonas melonis
55B3 Staphylococcus capitis
Counts
MS/MS coverage

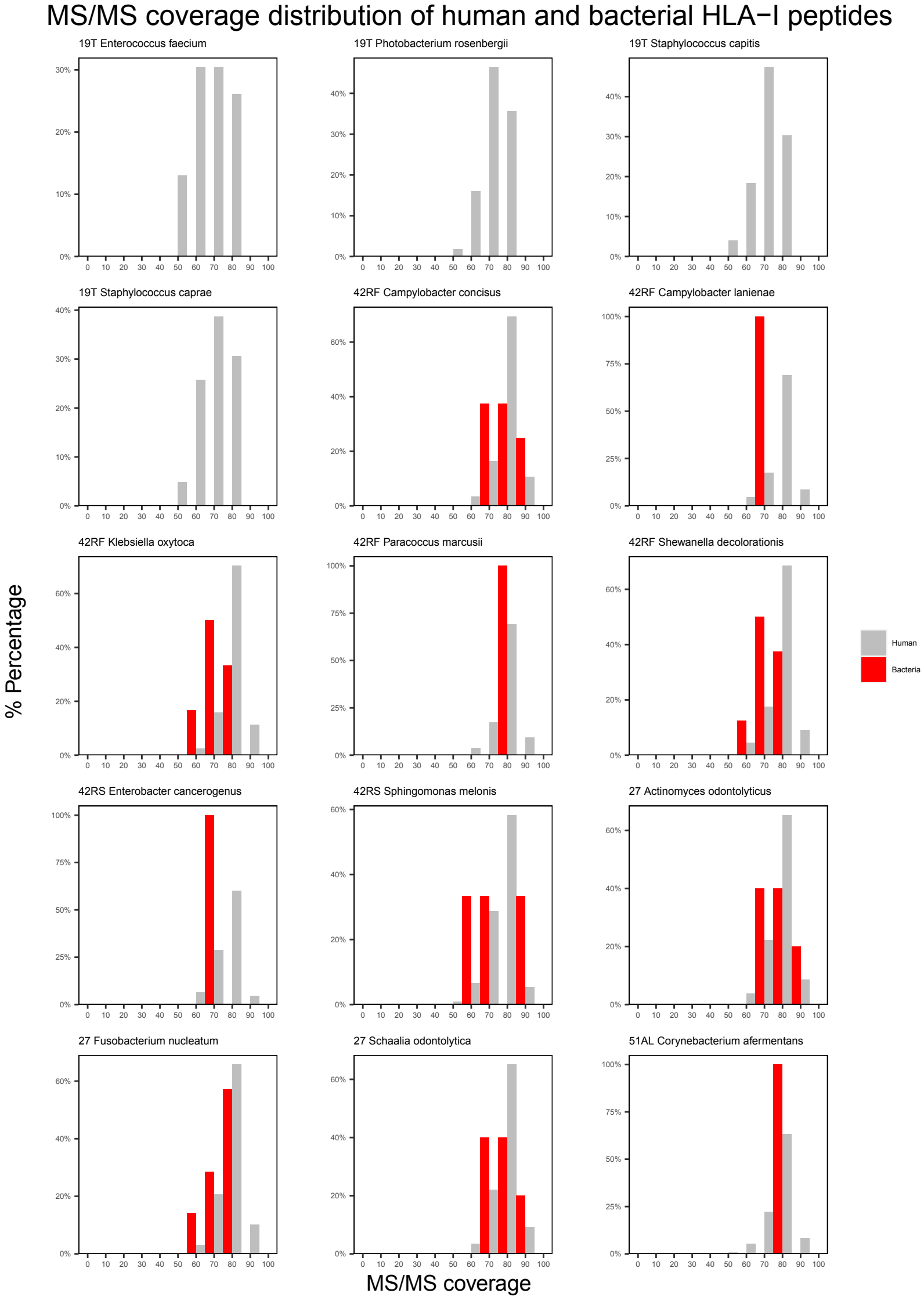

55B3 Staphylococcus caprae
92B3 Fusobacterium nucleatum
Counts
MS/MS coverage
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10
12
0 10 20 30 40 50 60 70 80 90 100

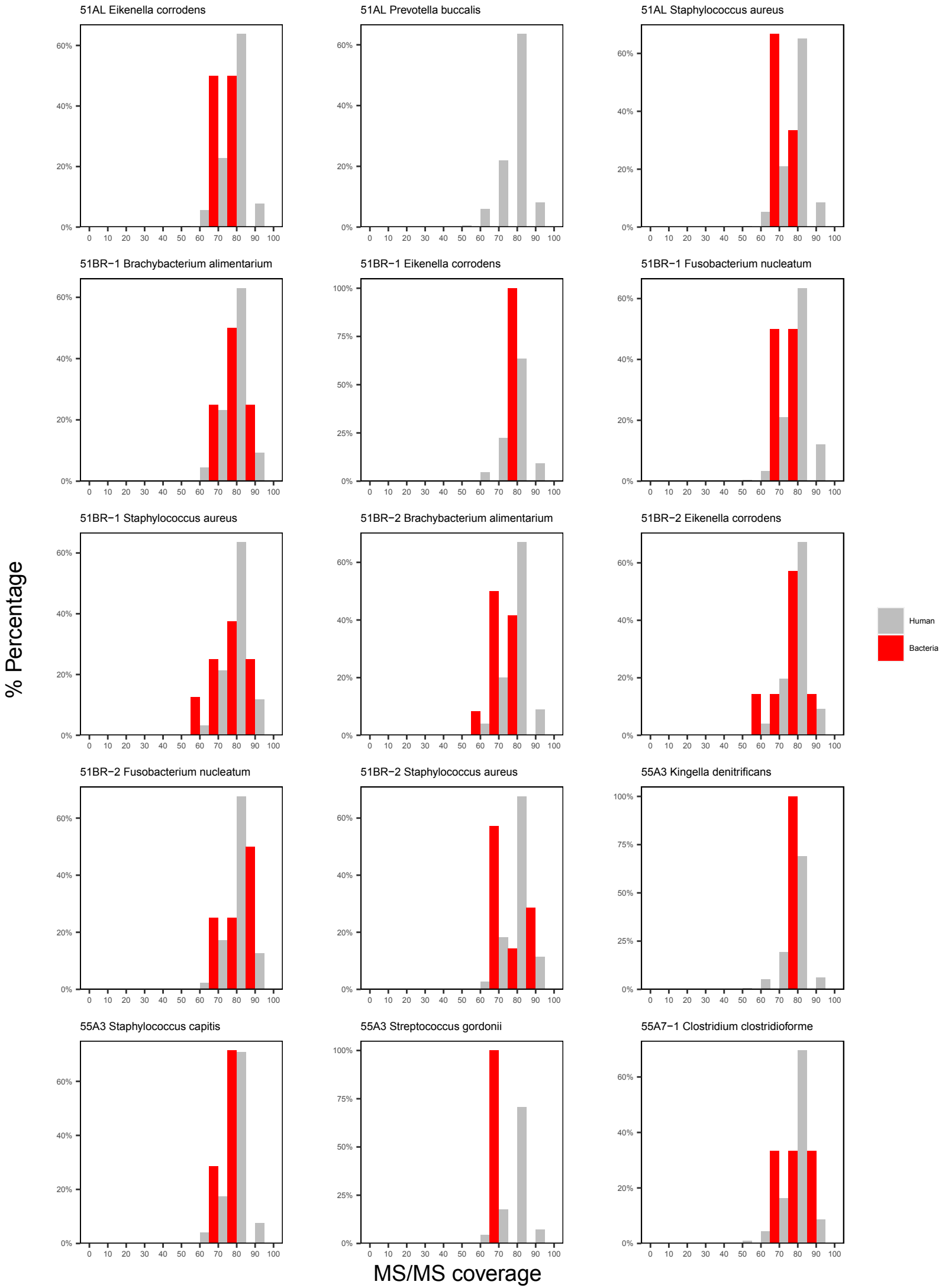

Correlation between peptide length and MS/MS coverage of HLA−I TAAs
27 Actinomyces odontolyticus
27 Fusobacterium nucleatum
27 Schaalia odontolytica
51AL Corynebacterium afermentans
51AL Eikenella corrodens
51AL Prevotella buccalis
51AL Staphylococcus aureus
51BR−1 Brachybacterium alimentarium
51BR−1 Eikenella corrodens
51BR−1 Fusobacterium nucleatum
51BR−1 Staphylococcus aureus
51BR−2 Brachybacterium alimentarium
51BR−2 Eikenella corrodens
51BR−2 Fusobacterium nucleatum
51BR−2 Staphylococcus aureus
MS/MS coverage
Peptide length

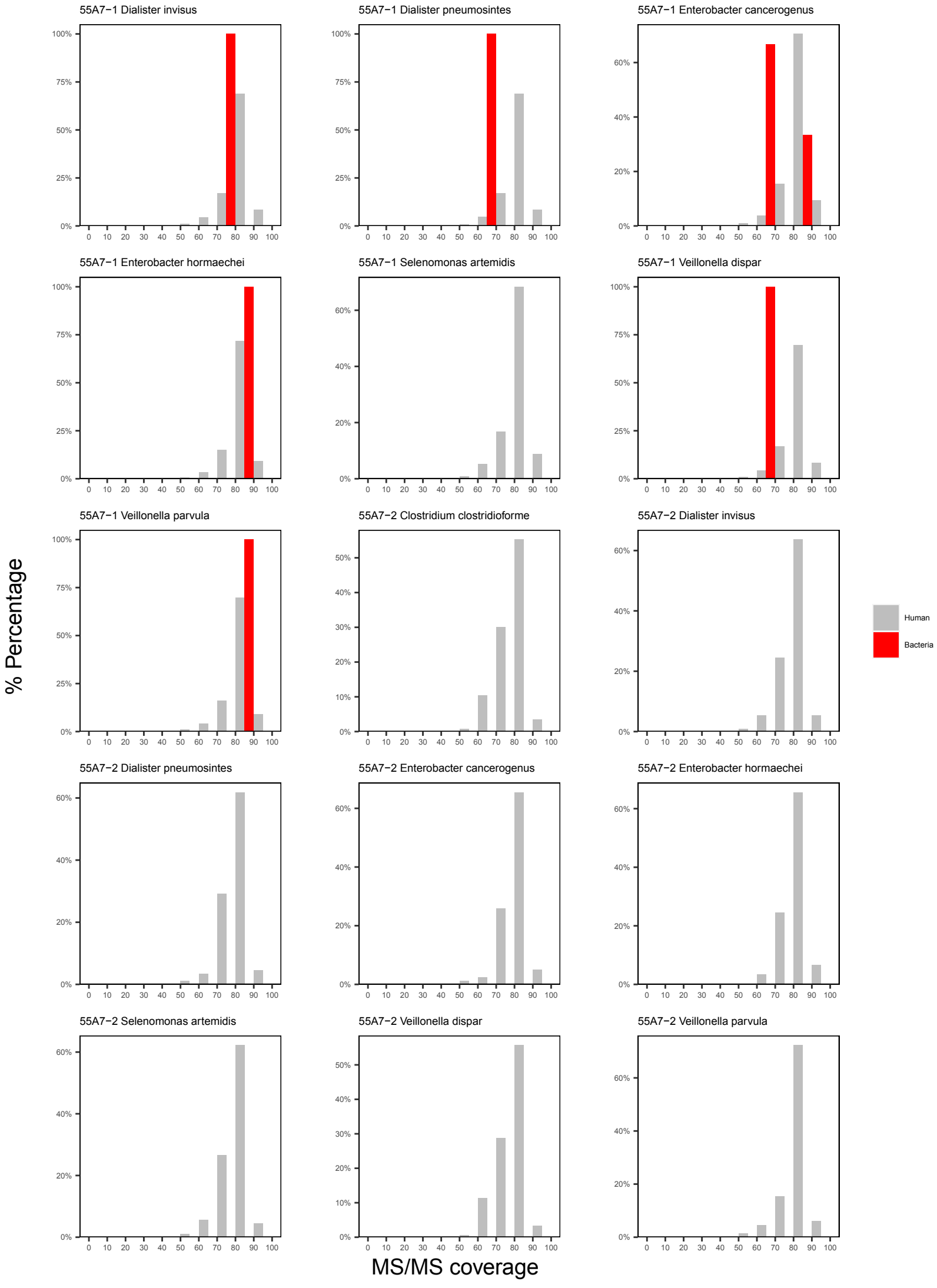

55A3 Kingella denitrificans
55A3 Staphylococcus capitis
55A3 Streptococcus gordonii
55A7−1 Clostridium clostridioforme
55A7−1 Dialister invisus
55A7−1 Dialister pneumosintes
55A7−1 Enterobacter cancerogenus
55A7−1 Enterobacter hormaechei
55A7−1 Selenomonas artemidis
55A7−1 Veillonella dispar
55A7−1 Veillonella parvula
55B3 Gardnerella vaginalis
55B3 Sphingomonas dokdonensis
55B3 Sphingomonas melonis
55B3 Staphylococcus capitis
MS/MS coverage
Peptide length

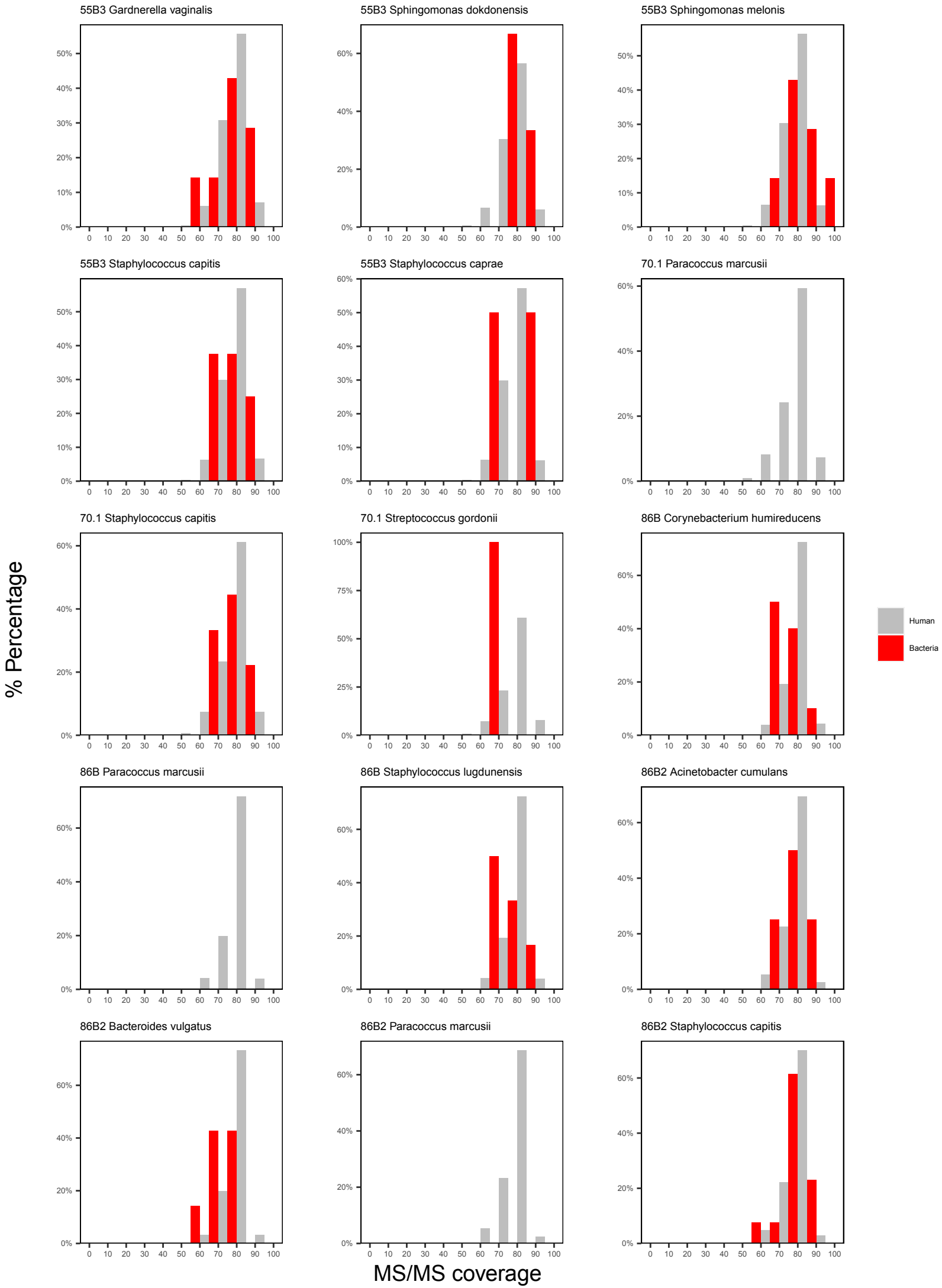

55B3 Staphylococcus caprae
92B3 Fusobacterium nucleatum
MS/MS coverage
100
75
50
25
0
8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
Peptide length

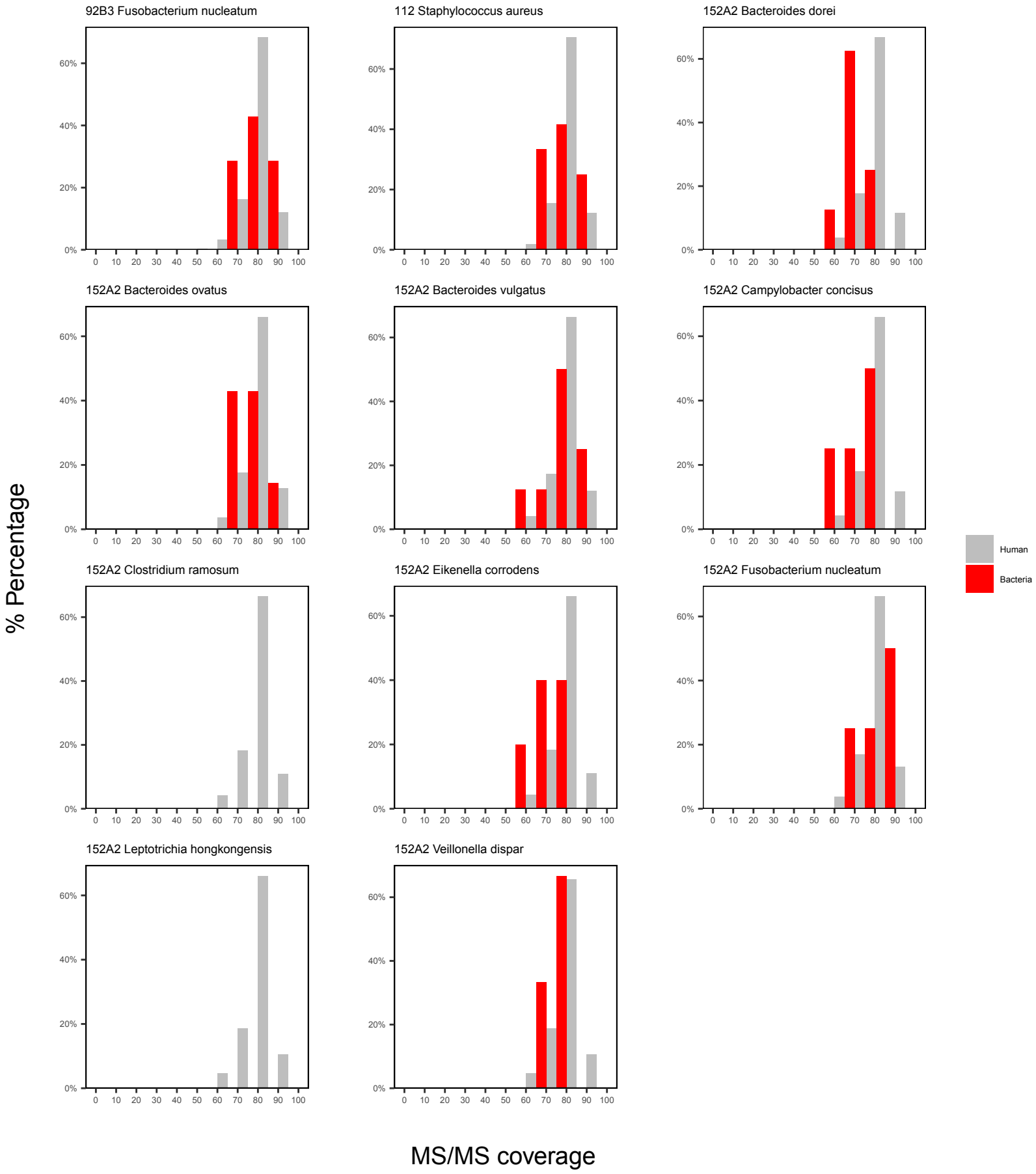
After selecting the threshold and applying our filtration criteria to our identified bacterial peptides, we plotted the distribution of MS/MS coverage of the selected bacteria-derived peptides found in the tumor samples. Also here, we identified a correlation with peptide length. We assume that the bacteria-derived peptides' MS/MS coverage is lower than that of the control human-derived peptides (which was used to determine the threshold) due to less presentation of bacterial proteins compared to human proteins, as there are far more human proteins than bacterial proteins in the tumor (bacteria are smaller than human cells, as is the amount of protein they contain).

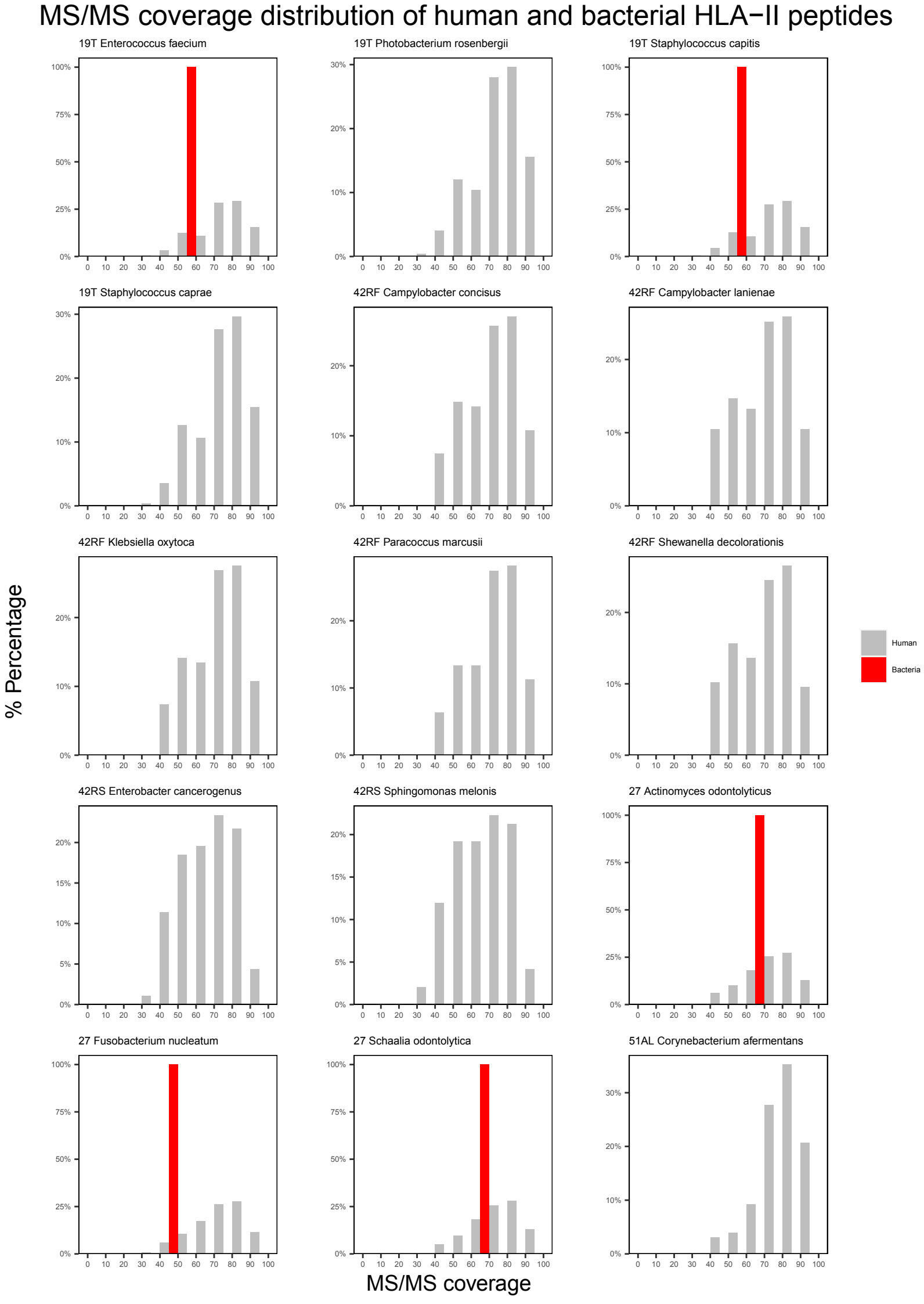

MS/MS coverage distribution of human and bacterial HLA−I peptides
19T Enterococcus faecium
19T Photobacterium rosenbergii
19T Staphylococcus capitis
19T Staphylococcus caprae
42RF Campylobacter concisus
42RF Campylobacter lanienae
42RF Klebsiella oxytoca
42RF Paracoccus marcusii
42RF Shewanella decolorationis
42RS Enterobacter cancerogenus
42RS Sphingomonas melonis
27 Actinomyces odontolyticus
27 Fusobacterium nucleatum
27 Schaalia odontolytica
51AL Corynebacterium afermentans
% Percentage
MS/MS coverage
Human
Bacteria

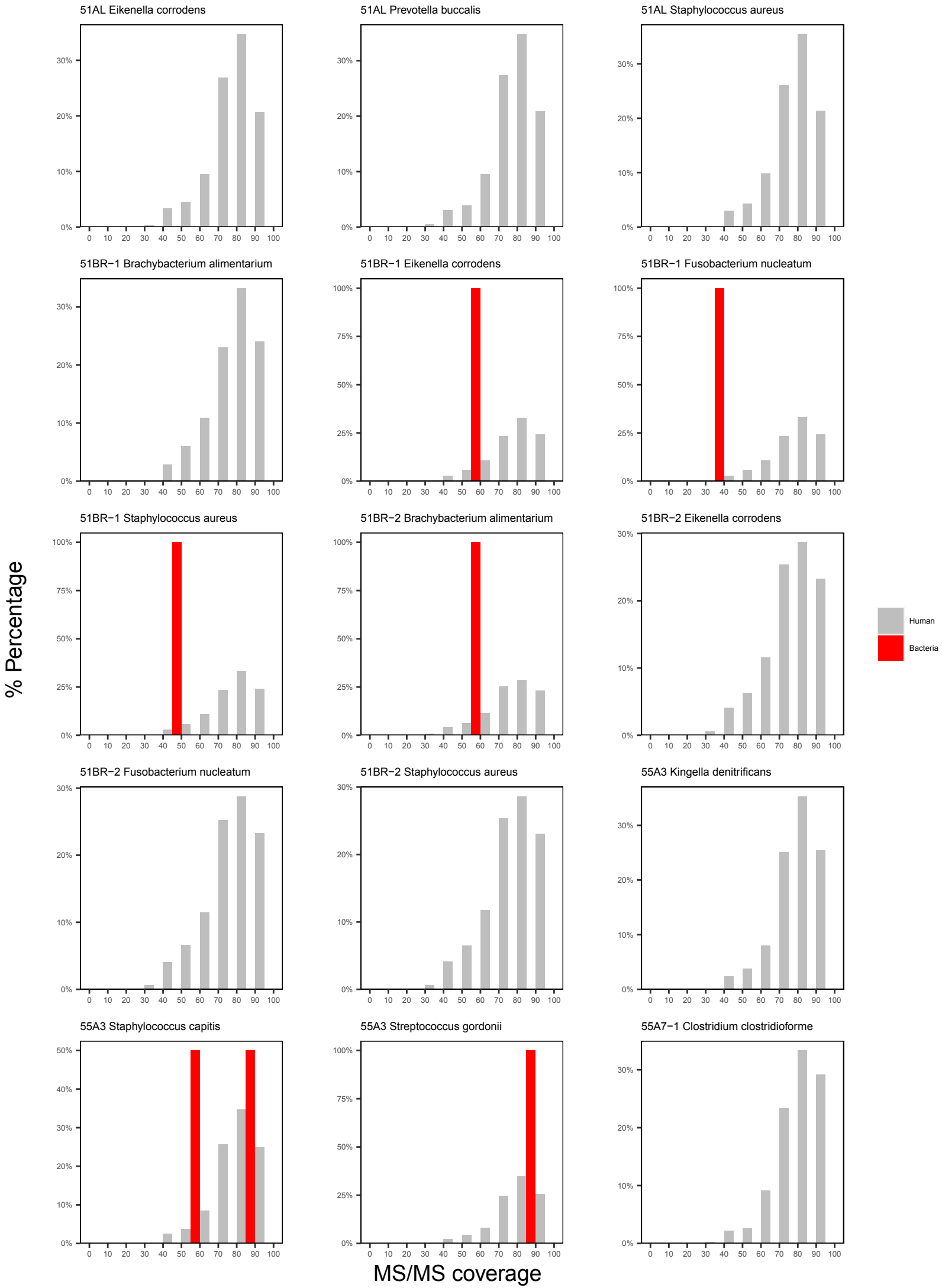

51AL Eikenella corrodens
51AL Prevotella buccalis
51AL Staphylococcus aureus
51BR−1 Brachybacterium alimentarium
51BR−1 Eikenella corrodens
51BR−1 Fusobacterium nucleatum
51BR−1 Staphylococcus aureus
51BR−2 Brachybacterium alimentarium
51BR−2 Eikenella corrodens
51BR−2 Fusobacterium nucleatum
51BR−2 Staphylococcus aureus
55A3 Kingella denitrificans
55A3 Staphylococcus capitis
55A3 Streptococcus gordonii
55A7−1 Clostridium clostridioforme
% Percentage
MS/MS coverage
Human
Bacteria

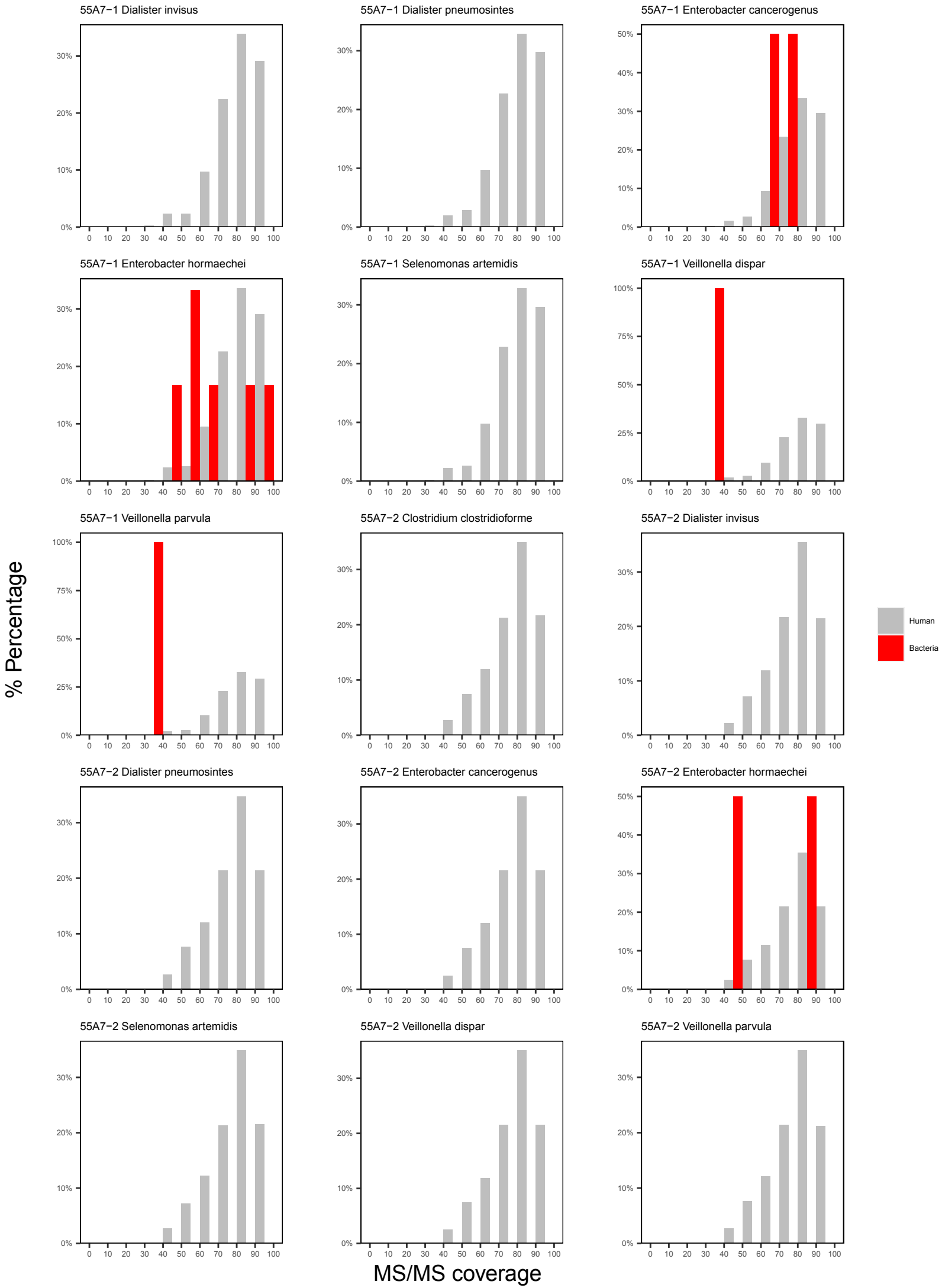

55A7−1 Dialister invisus
55A7−1 Dialister pneumosintes
55A7−1 Enterobacter cancerogenus
55A7−1 Enterobacter hormaechei
55A7−1 Selenomonas artemidis
55A7−1 Veillonella dispar
55A7−1 Veillonella parvula
55A7−2 Clostridium clostridioforme
55A7−2 Dialister invisus
55A7−2 Dialister pneumosintes
55A7−2 Enterobacter cancerogenus
55A7−2 Enterobacter hormaechei
55A7−2 Selenomonas artemidis
55A7−2 Veillonella dispar
55A7−2 Veillonella parvula
% Percentage
MS/MS coverage
Human
Bacteria

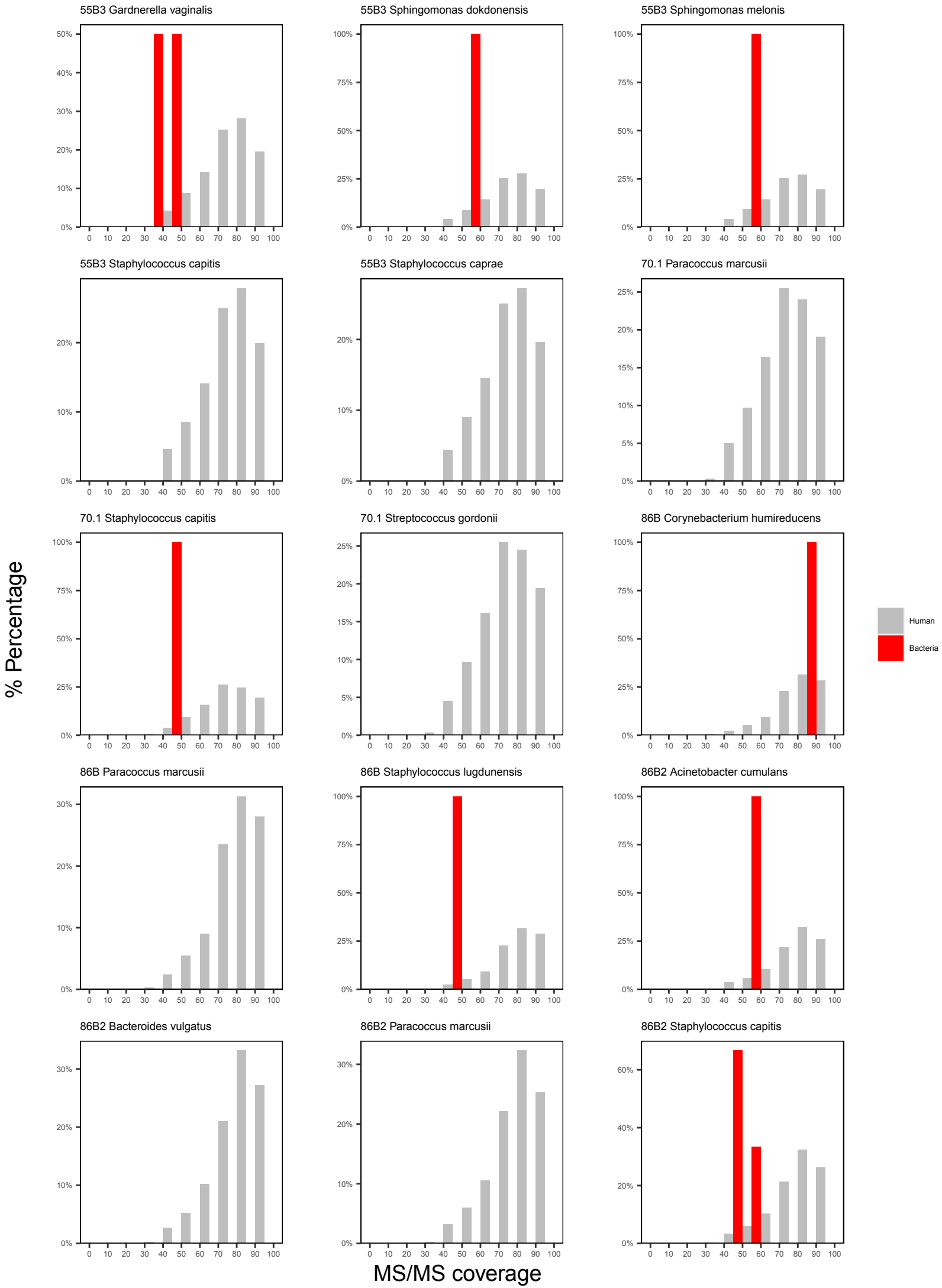

55B3 Gardnerella vaginalis
55B3 Sphingomonas dokdonensis
55B3 Sphingomonas melonis
55B3 Staphylococcus capitis
55B3 Staphylococcus caprae
70.1 Paracoccus marcusii
70.1 Staphylococcus capitis
70.1 Streptococcus gordonii
86B Corynebacterium humireducens
86B Paracoccus marcusii
86B Staphylococcus lugdunensis
86B2 Acinetobacter cumulans
86B2 Bacteroides vulgatus
86B2 Paracoccus marcusii
86B2 Staphylococcus capitis
% Percentage
MS/MS coverage
Human
Bacteria


92B3 Fusobacterium nucleatum
112 Staphylococcus aureus
152A2 Bacteroides dorei
152A2 Bacteroides ovatus
152A2 Bacteroides vulgatus
152A2 Campylobacter concisus
152A2 Clostridium ramosum
152A2 Eikenella corrodens
152A2 Fusobacterium nucleatum
152A2 Leptotrichia hongkongensis
152A2 Veillonella dispar
% Percentage
MS/MS coverage
Human
Bacteria

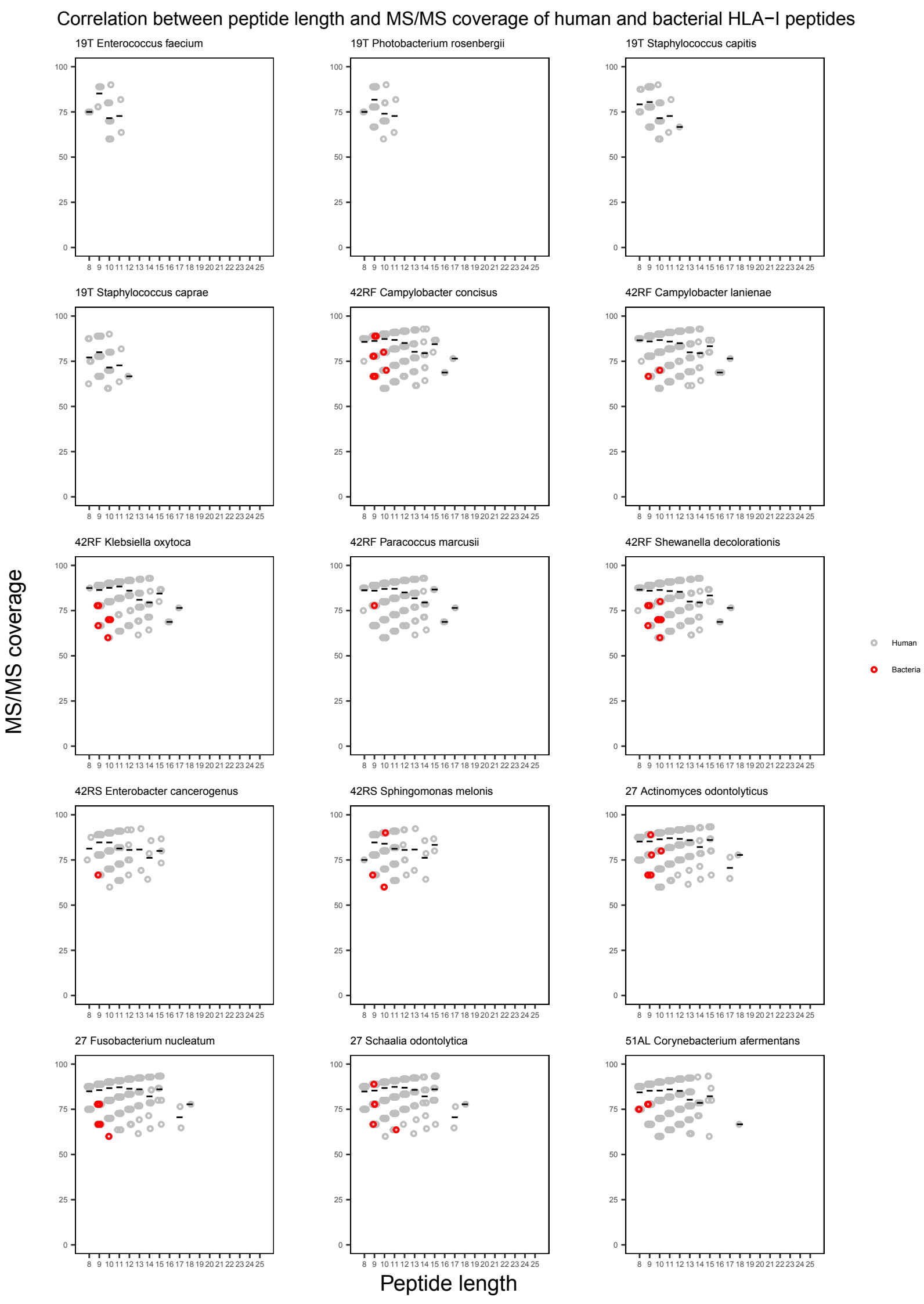

MS/MS coverage distribution of human and bacterial HLA−II peptides
19T Enterococcus faecium
19T Photobacterium rosenbergii
19T Staphylococcus capitis
19T Staphylococcus caprae
42RF Campylobacter concisus
42RF Campylobacter lanienae
42RF Klebsiella oxytoca
42RF Paracoccus marcusii
42RF Shewanella decolorationis
42RS Enterobacter cancerogenus
42RS Sphingomonas melonis
27 Actinomyces odontolyticus
27 Fusobacterium nucleatum
27 Schaalia odontolytica
51AL Corynebacterium afermentans
% Percentage
MS/MS coverage
Human
Bacteria

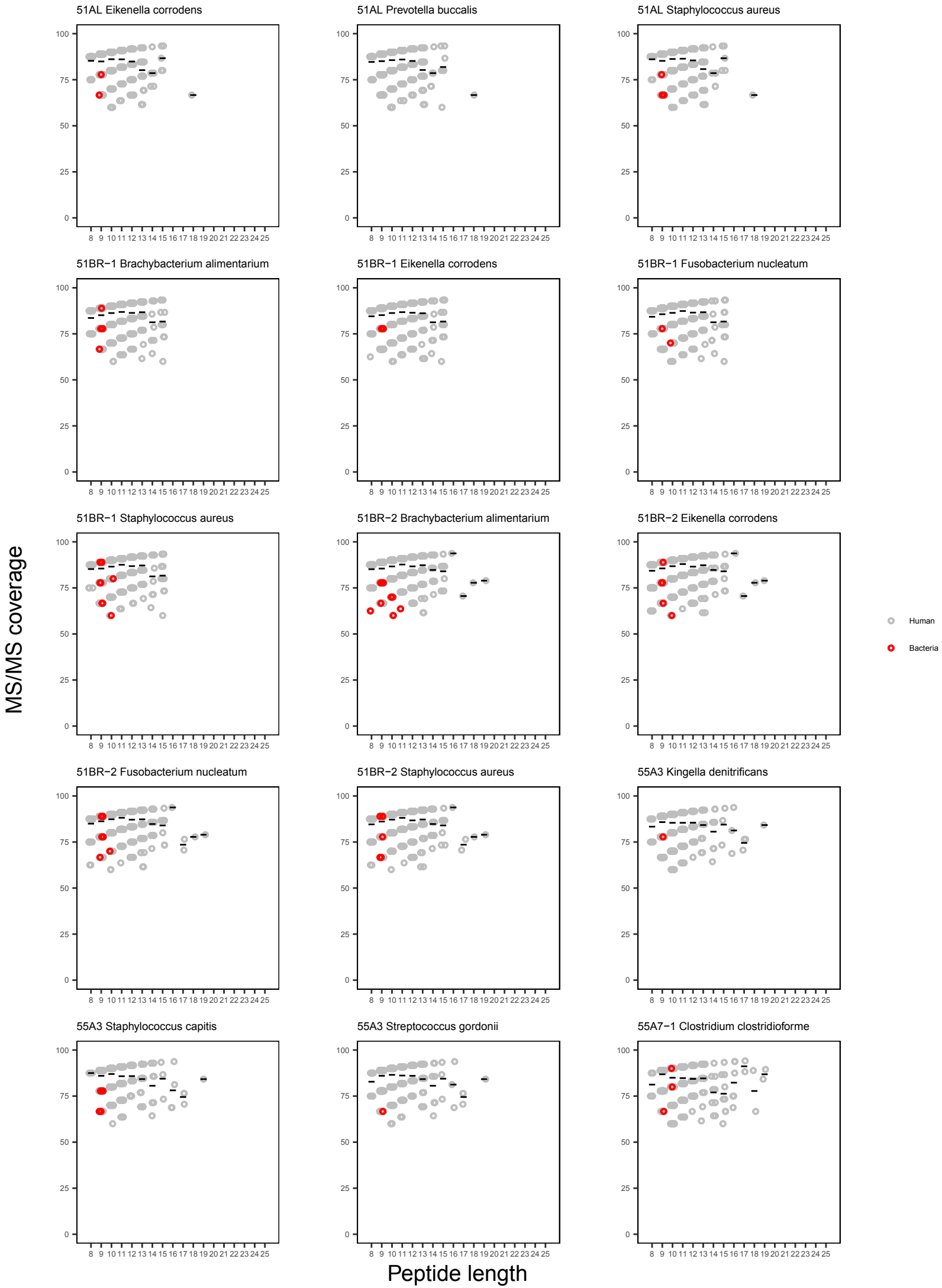

51AL Eikenella corrodens
51AL Prevotella buccalis
51AL Staphylococcus aureus
51BR−1 Brachybacterium alimentarium
51BR−1 Eikenella corrodens
51BR−1 Fusobacterium nucleatum
51BR−1 Staphylococcus aureus
51BR−2 Brachybacterium alimentarium
51BR−2 Eikenella corrodens
51BR−2 Fusobacterium nucleatum
51BR−2 Staphylococcus aureus
55A3 Kingella denitrificans
55A3 Staphylococcus capitis
55A3 Streptococcus gordonii
55A7−1 Clostridium clostridioforme
% Percentage
MS/MS coverage
Human
Bacteria

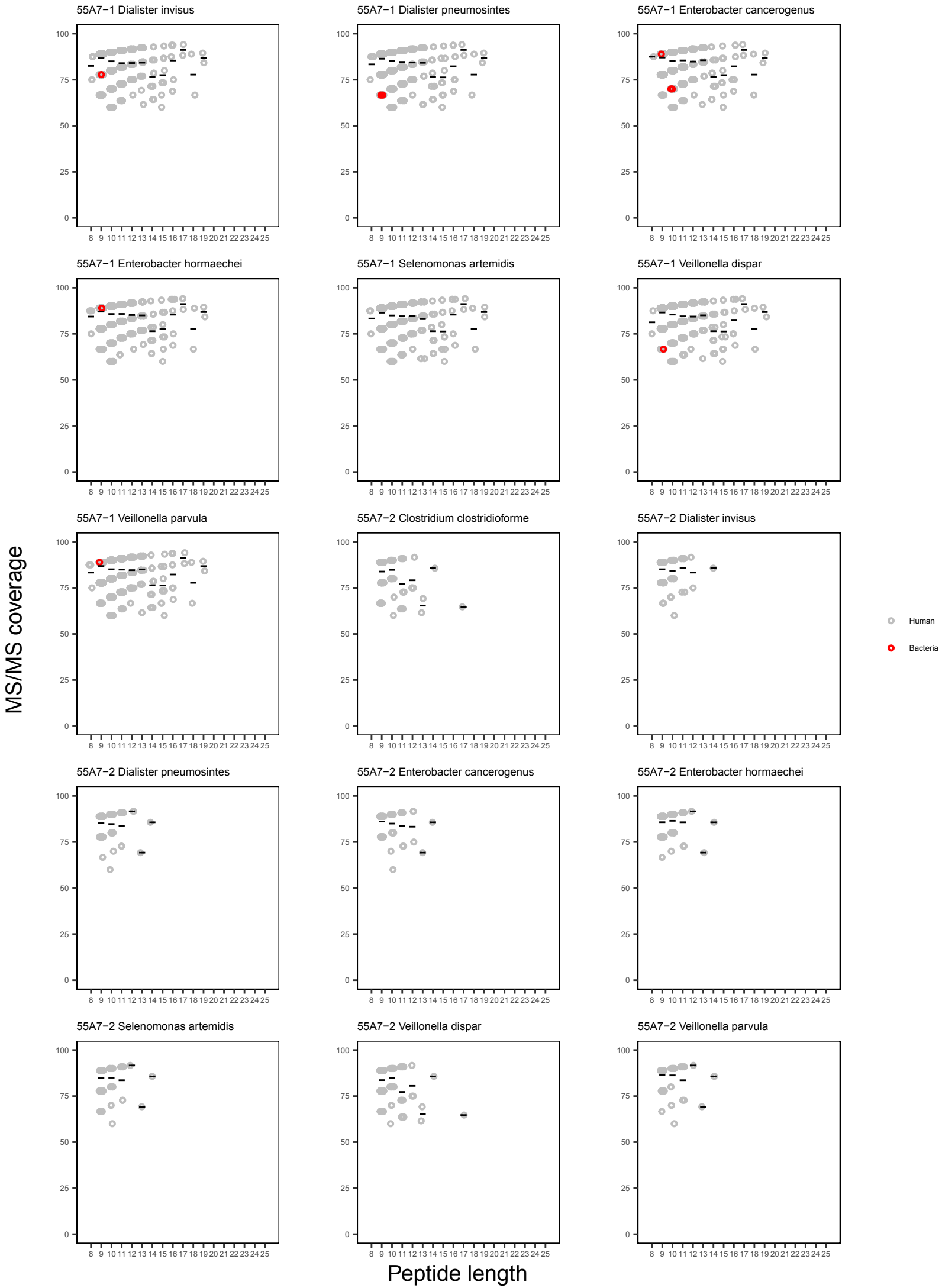

55A7−1 Dialister invisus
55A7−1 Dialister pneumosintes
55A7−1 Enterobacter cancerogenus
55A7−1 Enterobacter hormaechei
55A7−1 Selenomonas artemidis
55A7−1 Veillonella dispar
55A7−1 Veillonella parvula
55A7−2 Clostridium clostridioforme
55A7−2 Dialister invisus
55A7−2 Dialister pneumosintes
55A7−2 Enterobacter cancerogenus
55A7−2 Enterobacter hormaechei
55A7−2 Selenomonas artemidis
55A7−2 Veillonella dispar
55A7−2 Veillonella parvula
% Percentage
MS/MS coverage
Human
Bacteria

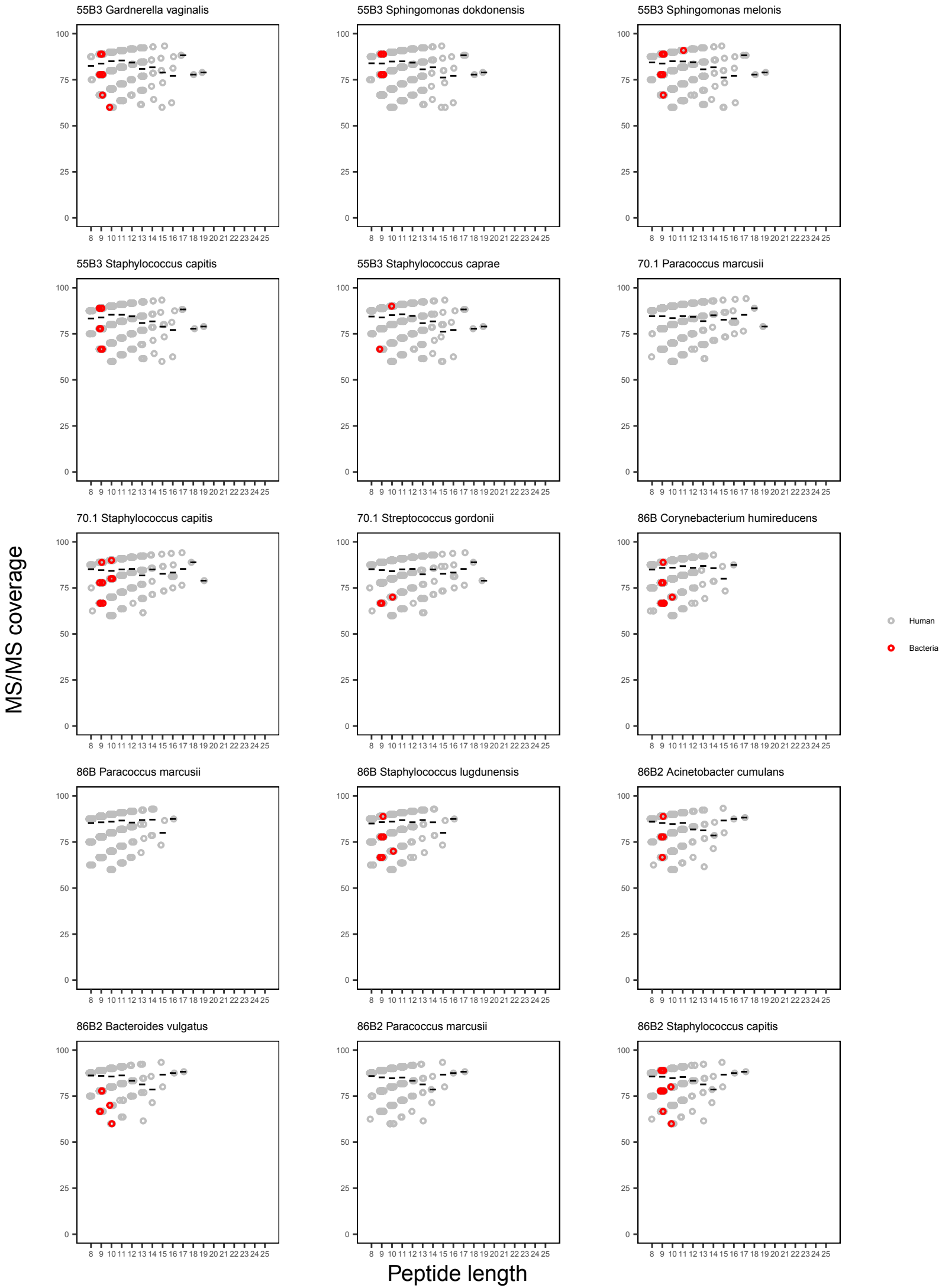

55B3 Gardnerella vaginalis
55B3 Sphingomonas dokdonensis
55B3 Sphingomonas melonis
55B3 Staphylococcus capitis
55B3 Staphylococcus caprae
70.1 Paracoccus marcusii
70.1 Staphylococcus capitis
70.1 Streptococcus gordonii
86B Corynebacterium humireducens
86B Paracoccus marcusii
86B Staphylococcus lugdunensis
86B2 Acinetobacter cumulans
86B2 Bacteroides vulgatus
86B2 Paracoccus marcusii
86B2 Staphylococcus capitis
% Percentage
MS/MS coverage
Human
Bacteria

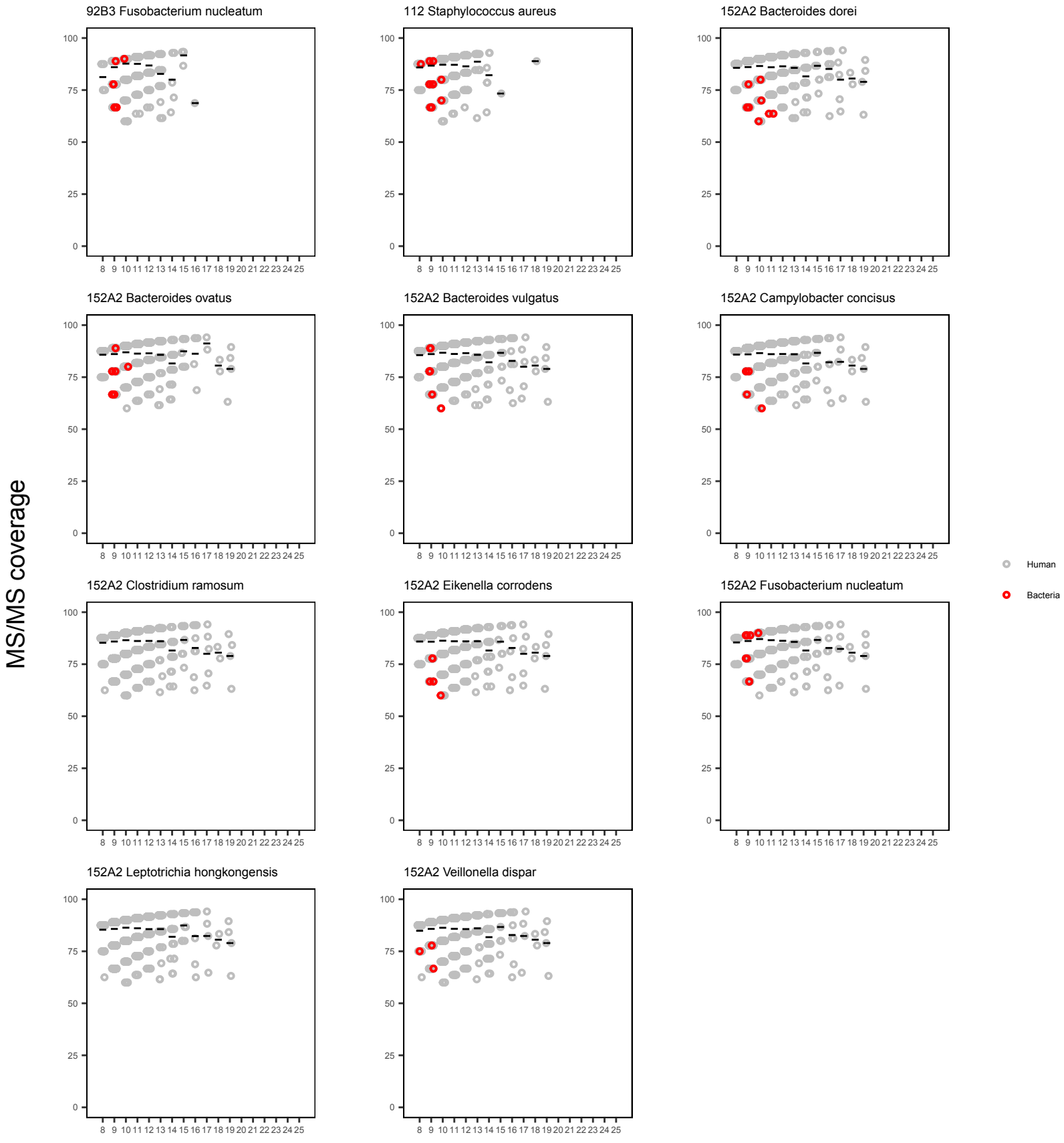

92B3 Fusobacterium nucleatum
% Percentage
100%
75%
50%
25%
0%
0 10 20 30 40 50 60 70 80 90 100
MS/MS coverage
Human
Bacteria

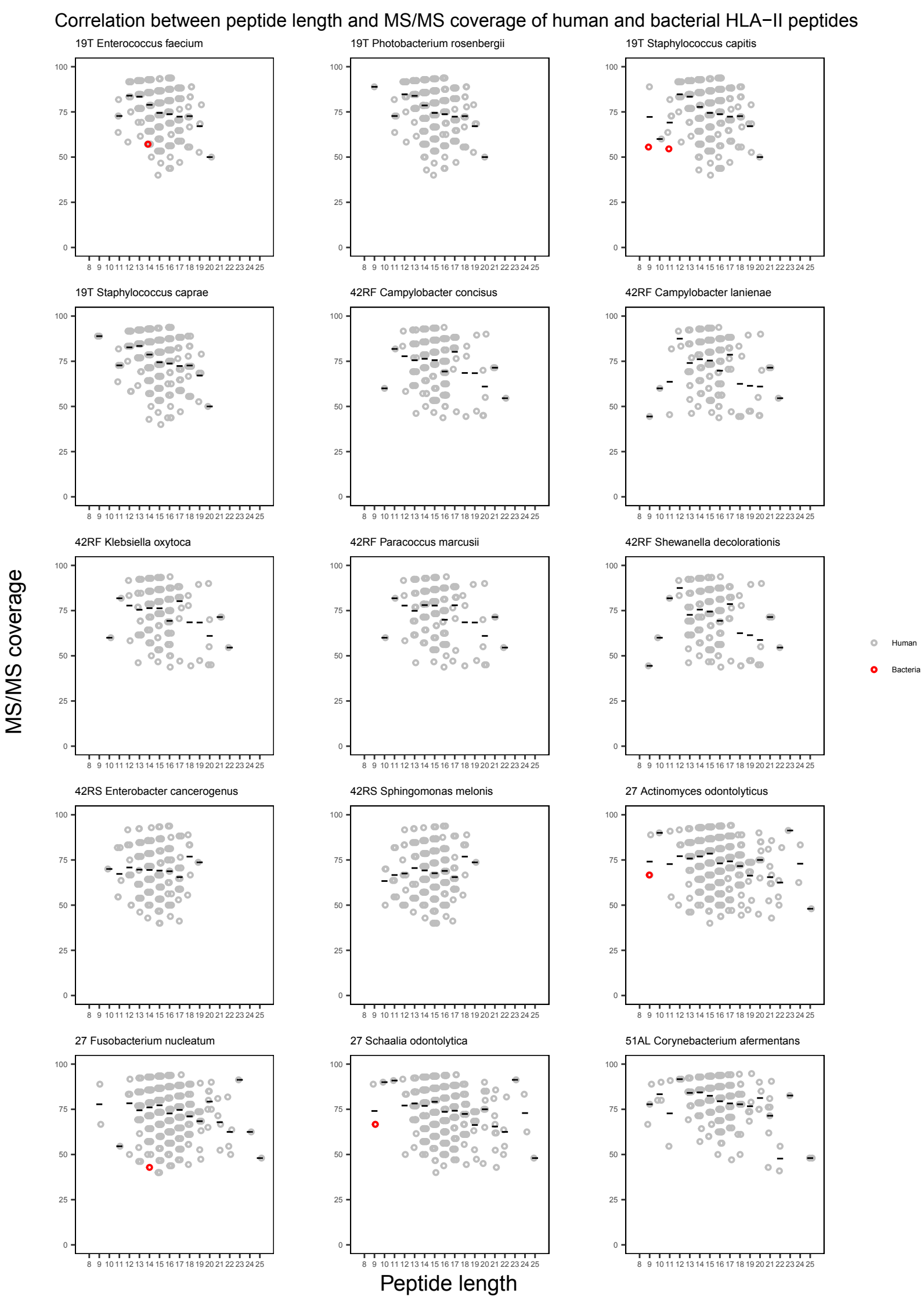

Correlation between peptide length and MS/MS coverage of human and bacterial HLA−I peptides
19T Enterococcus faecium
19T Photobacterium rosenbergii
19T Staphylococcus capitis
19T Staphylococcus caprae
42RF Campylobacter concisus
42RF Campylobacter lanienae
42RF Klebsiella oxytoca
42RF Paracoccus marcusii
42RF Shewanella decolorationis
42RS Enterobacter cancerogenus
42RS Sphingomonas melonis
27 Actinomyces odontolyticus
27 Fusobacterium nucleatum
27 Schaalia odontolytica
51AL Corynebacterium afermentans
MS/MS coverage
Peptide length
Human
Bacteria

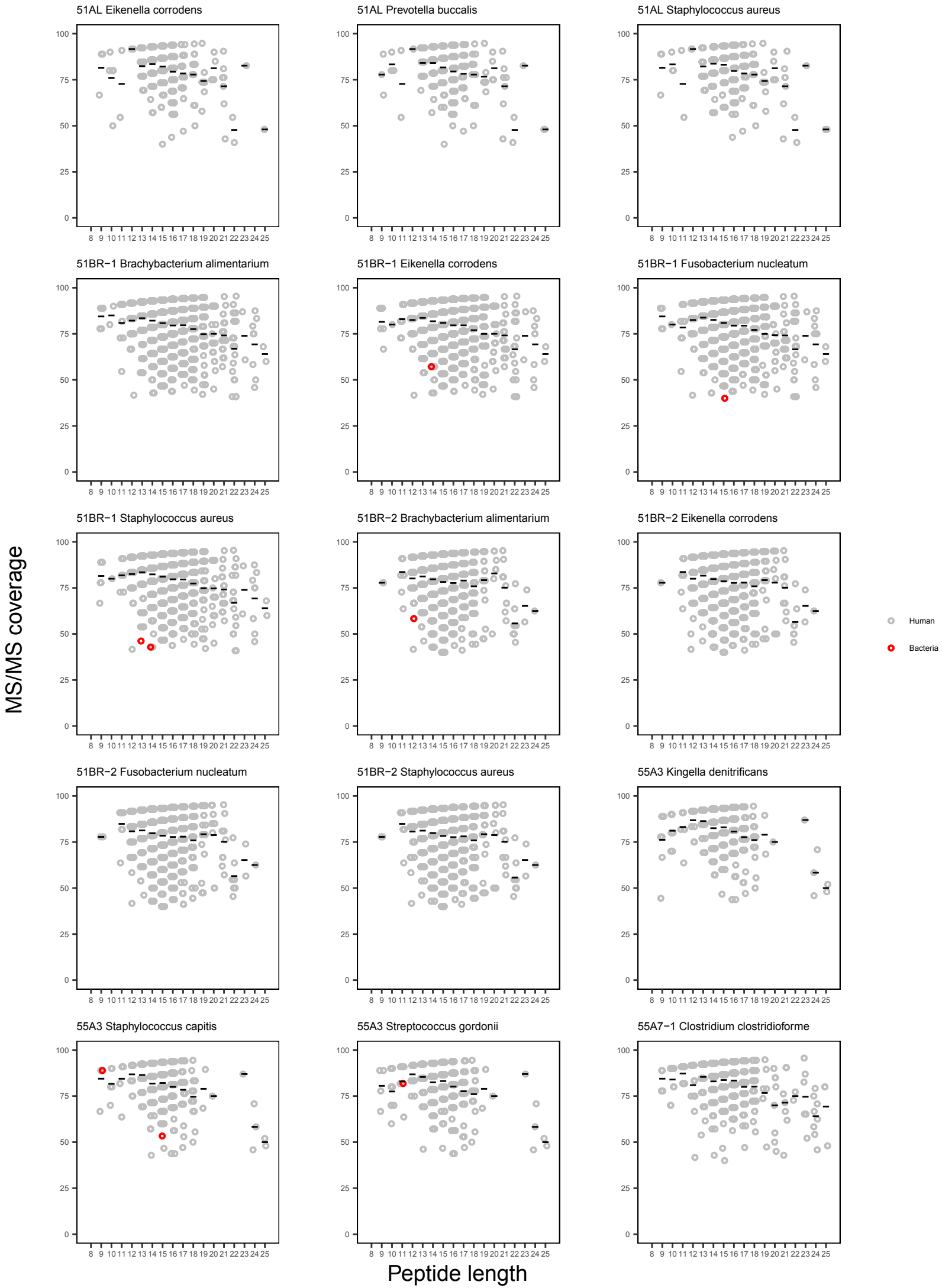

51AL Eikenella corrodens
51AL Prevotella buccalis
51AL Staphylococcus aureus
51BR−1 Brachybacterium alimentarium
51BR−1 Eikenella corrodens
51BR−1 Fusobacterium nucleatum
51BR−1 Staphylococcus aureus
51BR−2 Brachybacterium alimentarium
51BR−2 Eikenella corrodens
51BR−2 Fusobacterium nucleatum
51BR−2 Staphylococcus aureus
55A3 Kingella denitrificans
55A3 Staphylococcus capitis
55A3 Streptococcus gordonii
55A7−1 Clostridium clostridioforme
MS/MS coverage
Peptide length
Human
Bacteria

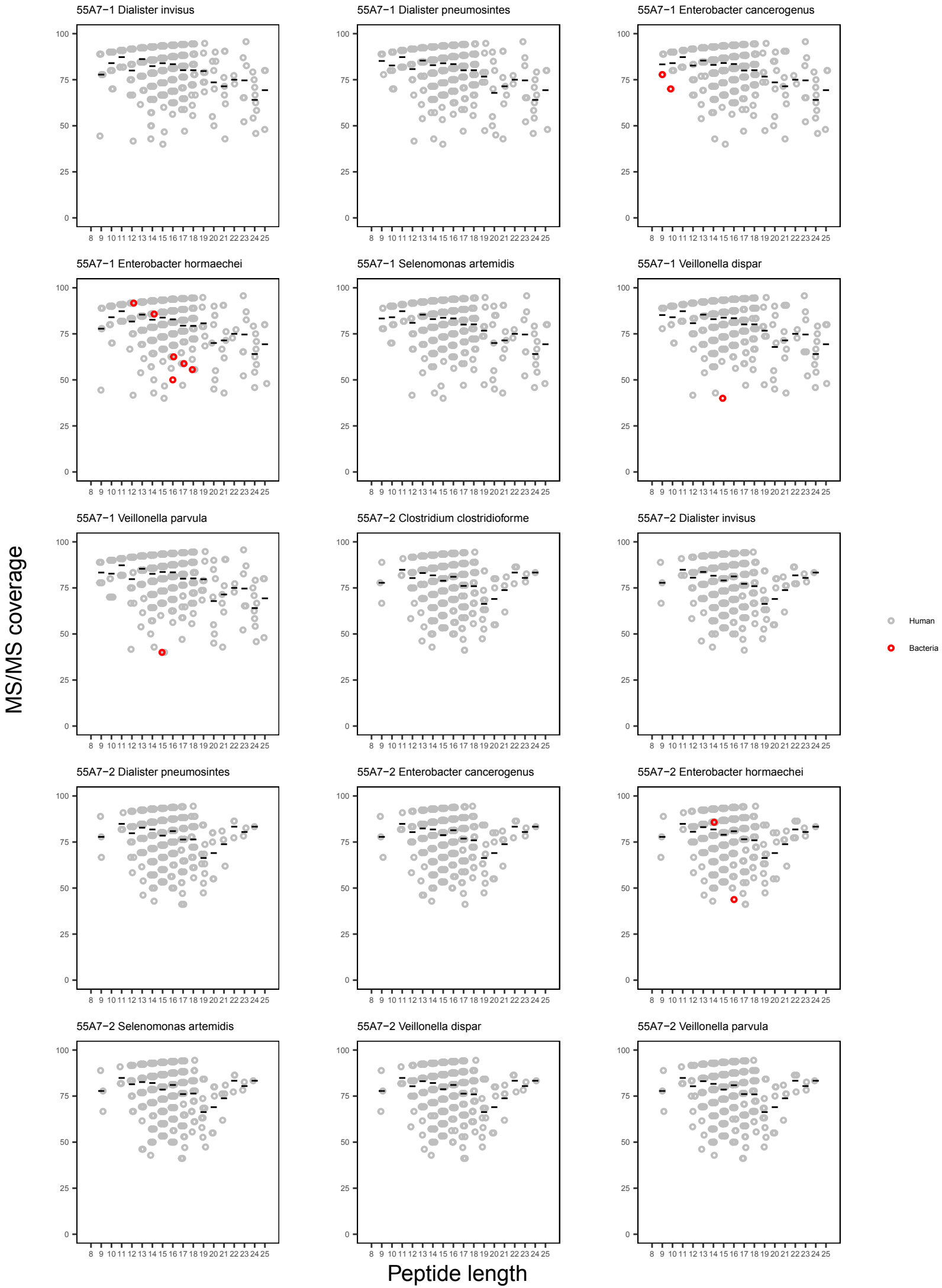

55A7−1 Dialister invisus
55A7−1 Dialister pneumosintes
55A7−1 Enterobacter cancerogenus
55A7−1 Enterobacter hormaechei
55A7−1 Selenomonas artemidis
55A7−1 Veillonella dispar
55A7−1 Veillonella parvula
55A7−2 Clostridium clostridioforme
55A7−2 Dialister invisus
55A7−2 Dialister pneumosintes
55A7−2 Enterobacter cancerogenus
55A7−2 Enterobacter hormaechei
55A7−2 Selenomonas artemidis
55A7−2 Veillonella dispar
55A7−2 Veillonella parvula
MS/MS coverage
Peptide length
Human
Bacteria

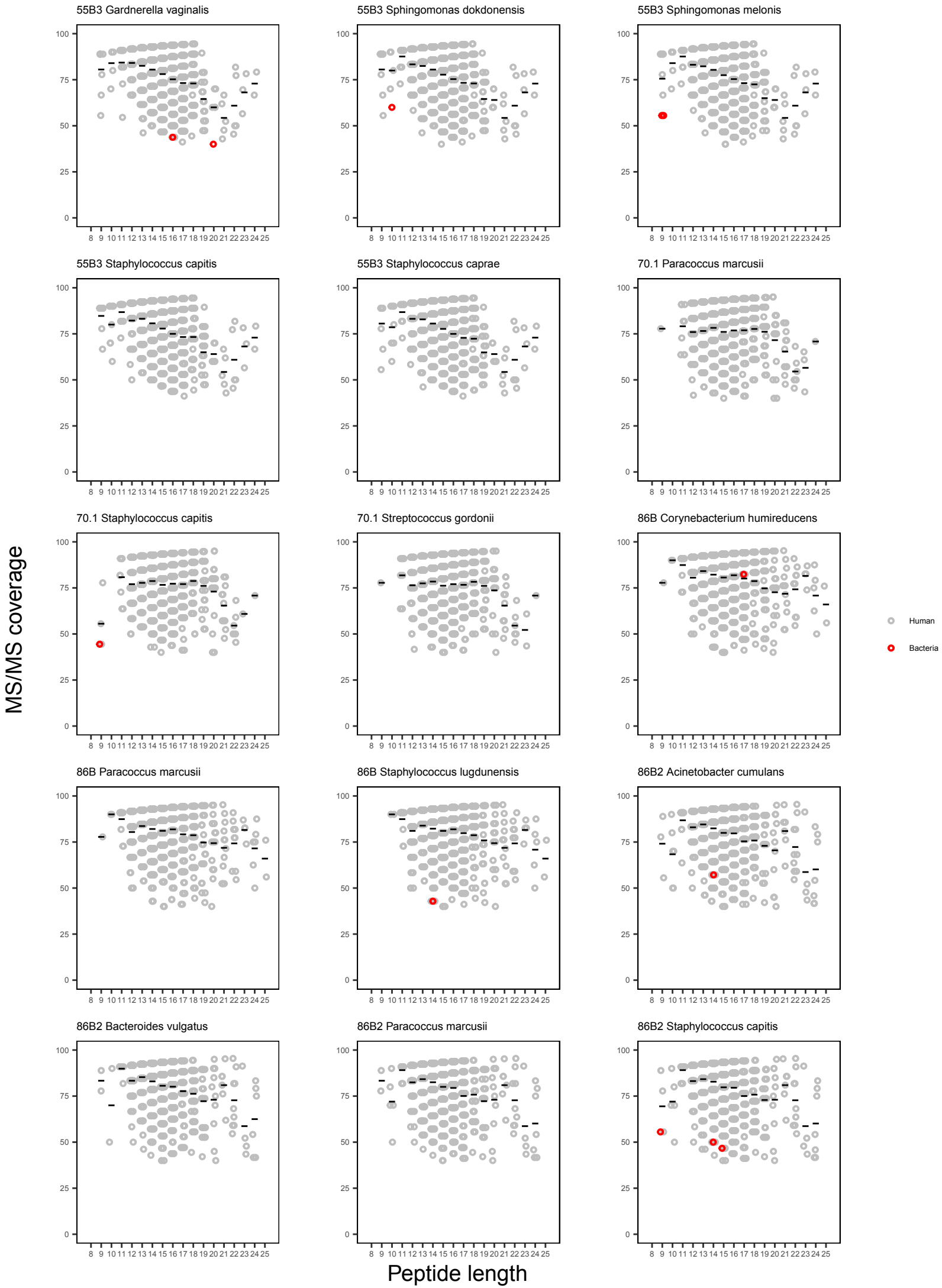

55B3 Gardnerella vaginalis
55B3 Sphingomonas dokdonensis
55B3 Sphingomonas melonis
55B3 Staphylococcus capitis
55B3 Staphylococcus caprae
70.1 Paracoccus marcusii
70.1 Staphylococcus capitis
70.1 Streptococcus gordonii
86B Corynebacterium humireducens
86B Paracoccus marcusii
86B Staphylococcus lugdunensis
86B2 Acinetobacter cumulans
86B2 Bacteroides vulgatus
86B2 Paracoccus marcusii
86B2 Staphylococcus capitis
MS/MS coverage
Peptide length
Human
Bacteria

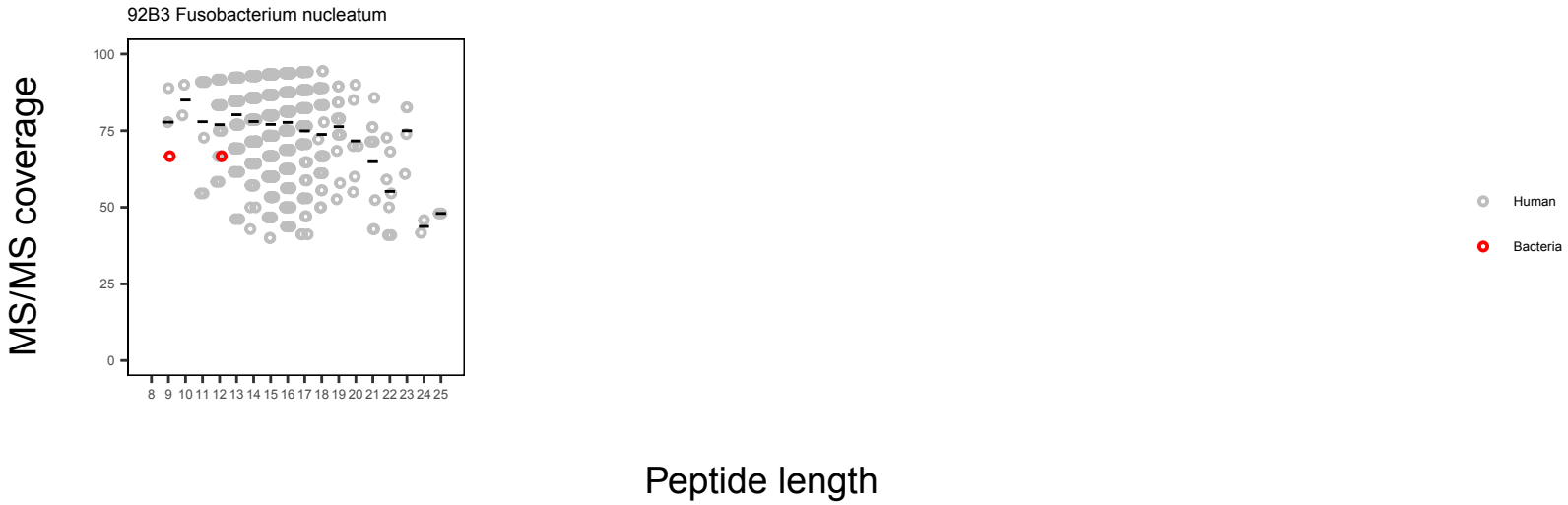

92B3 Fusobacterium nucleatum
112 Staphylococcus aureus
152A2 Bacteroides dorei
152A2 Bacteroides ovatus
152A2 Bacteroides vulgatus
152A2 Campylobacter concisus
152A2 Clostridium ramosum
152A2 Eikenella corrodens
152A2 Fusobacterium nucleatum
152A2 Leptotrichia hongkongensis
152A2 Veillonella dispar
MS/MS coverage
Peptide length
Human
Bacteria


Correlation between peptide length and MS/MS coverage of human and bacterial HLA−II peptides
19T Enterococcus faecium
19T Photobacterium rosenbergii
19T Staphylococcus capitis
19T Staphylococcus caprae
42RF Campylobacter concisus
42RF Campylobacter lanienae
42RF Klebsiella oxytoca
42RF Paracoccus marcusii
42RF Shewanella decolorationis
42RS Enterobacter cancerogenus
42RS Sphingomonas melonis
27 Actinomyces odontolyticus
27 Fusobacterium nucleatum
27 Schaalia odontolytica
51AL Corynebacterium afermentans
MS/MS coverage
Peptide length
Human
Bacteria


51AL Eikenella corrodens
51AL Prevotella buccalis
51AL Staphylococcus aureus
51BR−1 Brachybacterium alimentarium
51BR−1 Eikenella corrodens
51BR−1 Fusobacterium nucleatum
51BR−1 Staphylococcus aureus
51BR−2 Brachybacterium alimentarium
51BR−2 Eikenella corrodens
51BR−2 Fusobacterium nucleatum
51BR−2 Staphylococcus aureus
55A3 Kingella denitrificans
55A3 Staphylococcus capitis
55A3 Streptococcus gordonii
55A7−1 Clostridium clostridioforme
MS/MS coverage
Peptide length
Human
Bacteria

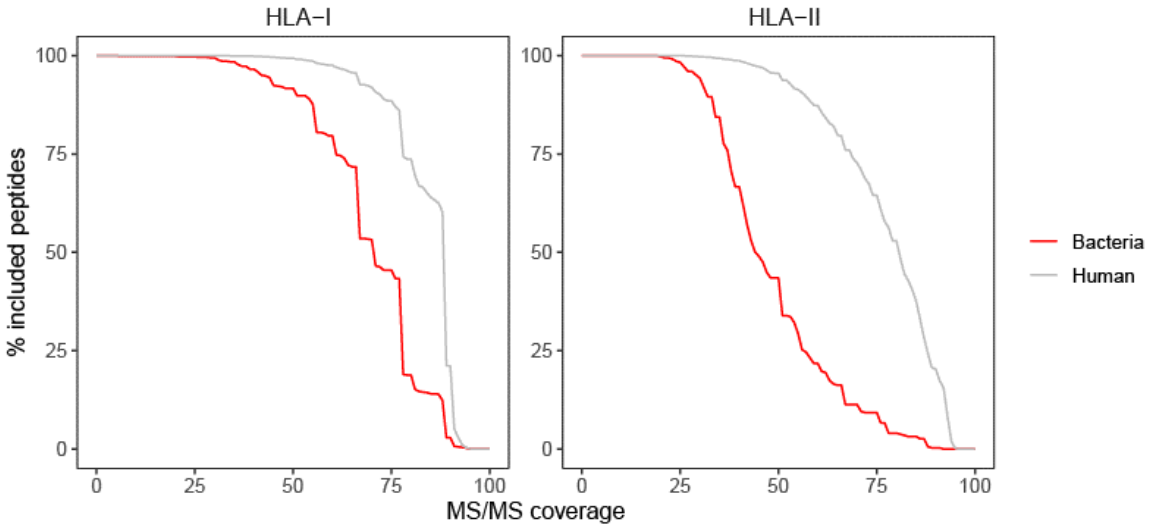

55A7−1 Dialister invisus
55A7−1 Dialister pneumosintes
55A7−1 Enterobacter cancerogenus
55A7−1 Enterobacter hormaechei
55A7−1 Selenomonas artemidis
55A7−1 Veillonella dispar
55A7−1 Veillonella parvula
55A7−2 Clostridium clostridioforme
55A7−2 Dialister invisus
55A7−2 Dialister pneumosintes
55A7−2 Enterobacter cancerogenus
55A7−2 Enterobacter hormaechei
55A7−2 Selenomonas artemidis
55A7−2 Veillonella dispar
55A7−2 Veillonella parvula
MS/MS coverage
Peptide length
Human
Bacteria


55B3 Gardnerella vaginalis
55B3 Sphingomonas dokdonensis
55B3 Sphingomonas melonis
55B3 Staphylococcus capitis
55B3 Staphylococcus caprae
70.1 Paracoccus marcusii
70.1 Staphylococcus capitis
70.1 Streptococcus gordonii
86B Corynebacterium humireducens
86B Paracoccus marcusii
86B Staphylococcus lugdunensis
86B2 Acinetobacter cumulans
86B2 Bacteroides vulgatus
86B2 Paracoccus marcusii
86B2 Staphylococcus capitis
MS/MS coverage
Peptide length
Human
Bacteria

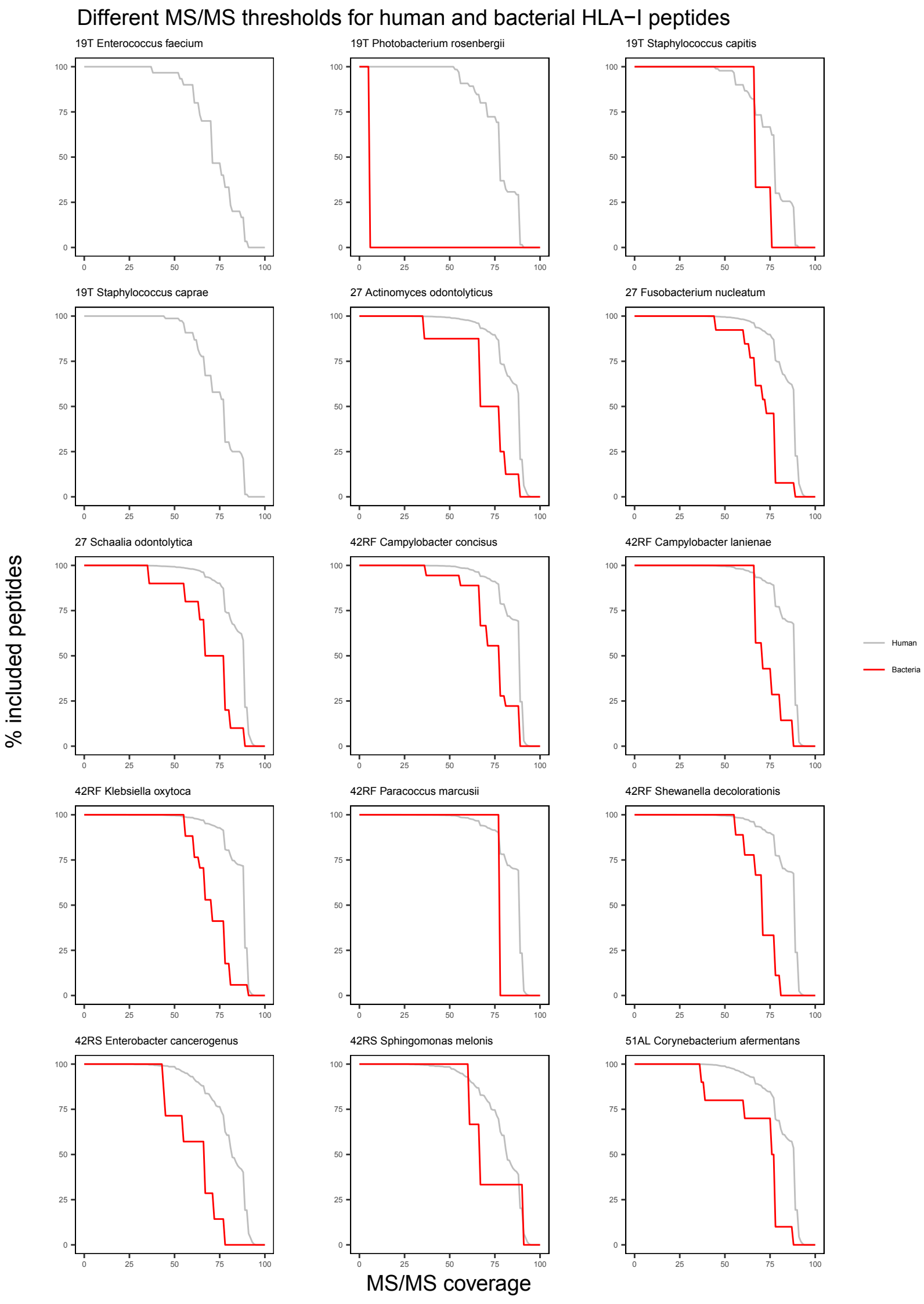

92B3 Fusobacterium nucleatum
MS/MS coverage
100
75
50
25
0
8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
Peptide length
Human
Bacteria

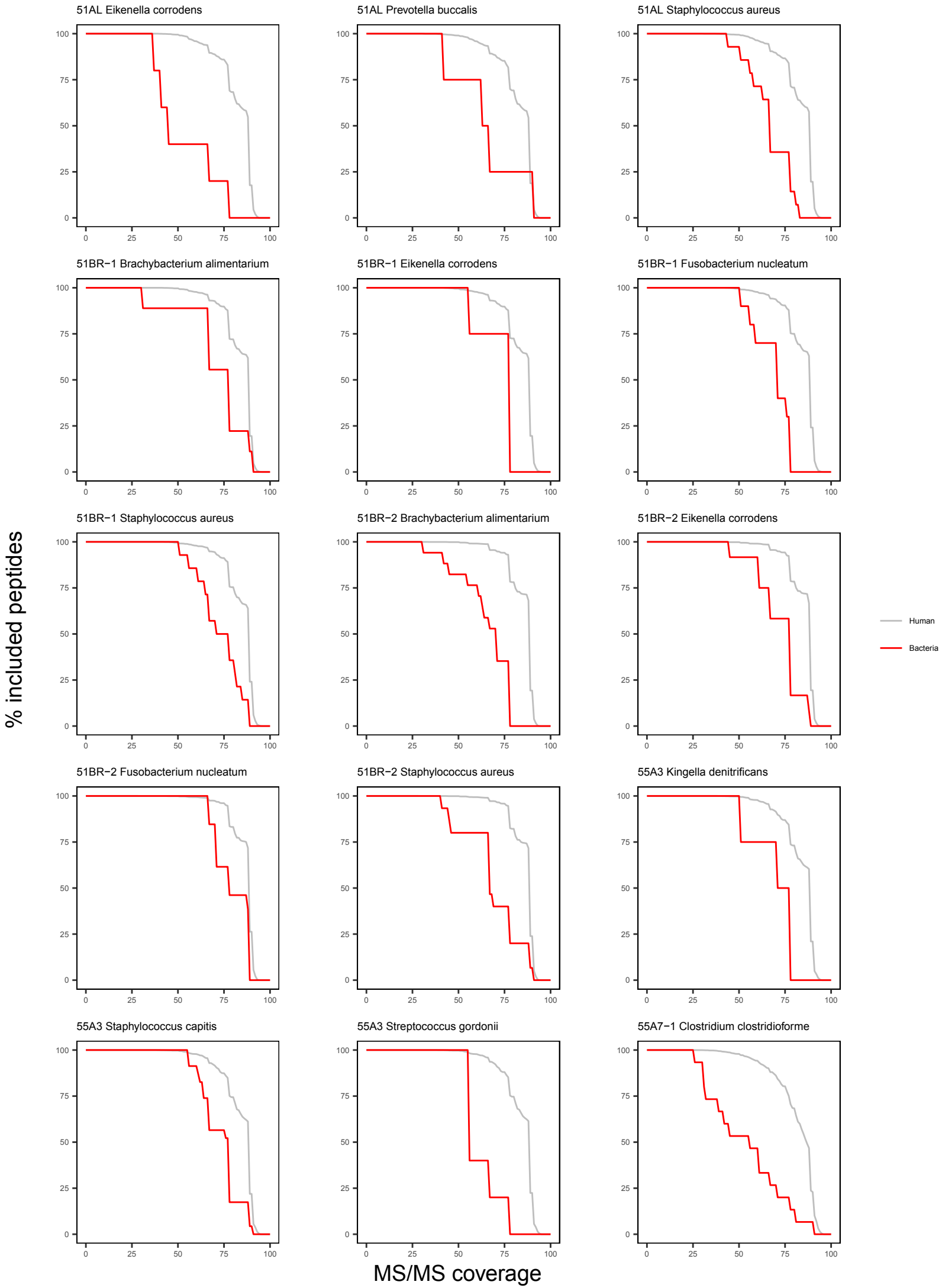
To assess the amount of human and bacterial peptides retained in the different MS/MS thresholds, we plotted the percentage of human (gray) / bacterial (red) peptides retained in each MS/MS coverage threshold:

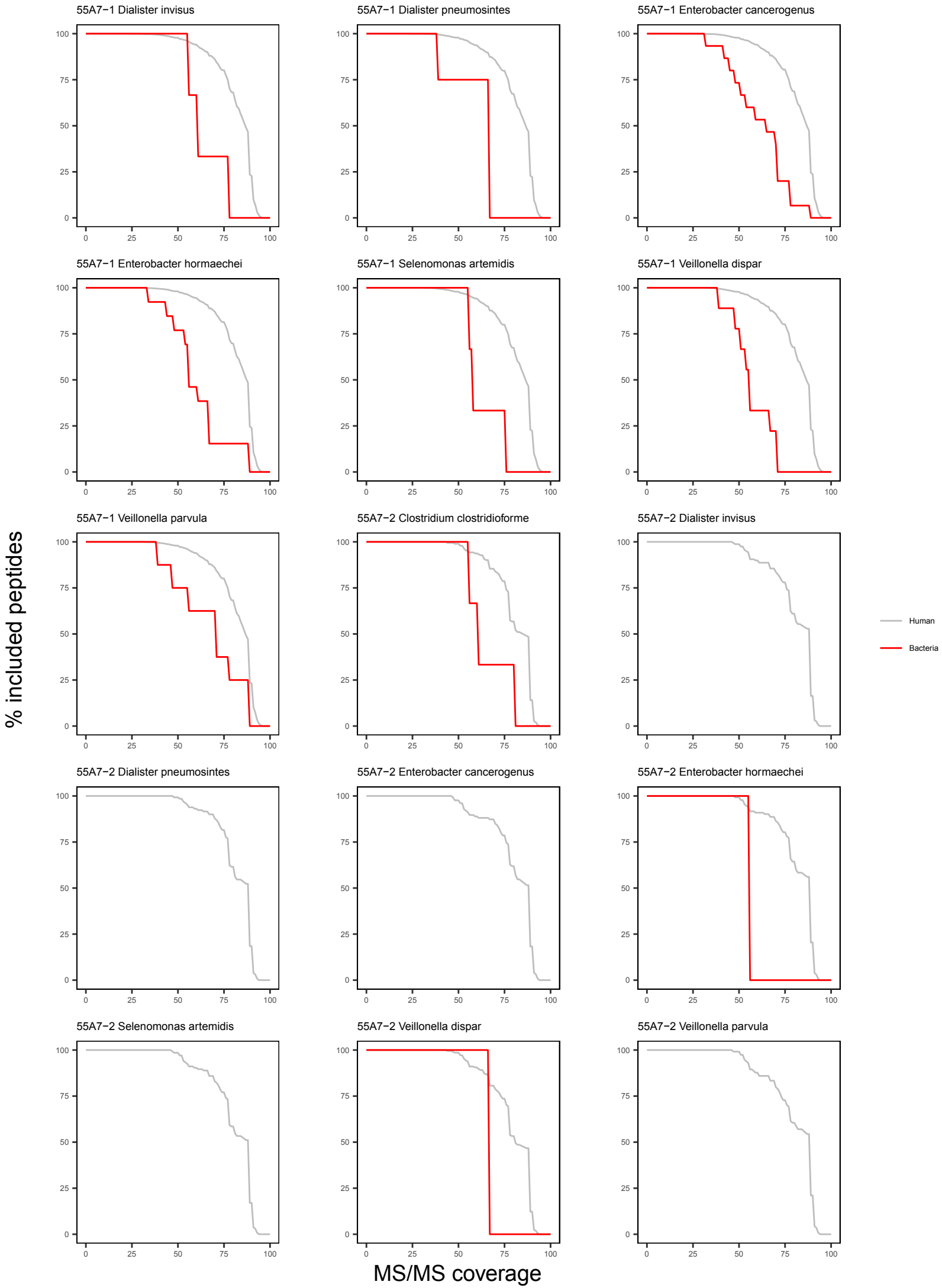
For all the peptides together:

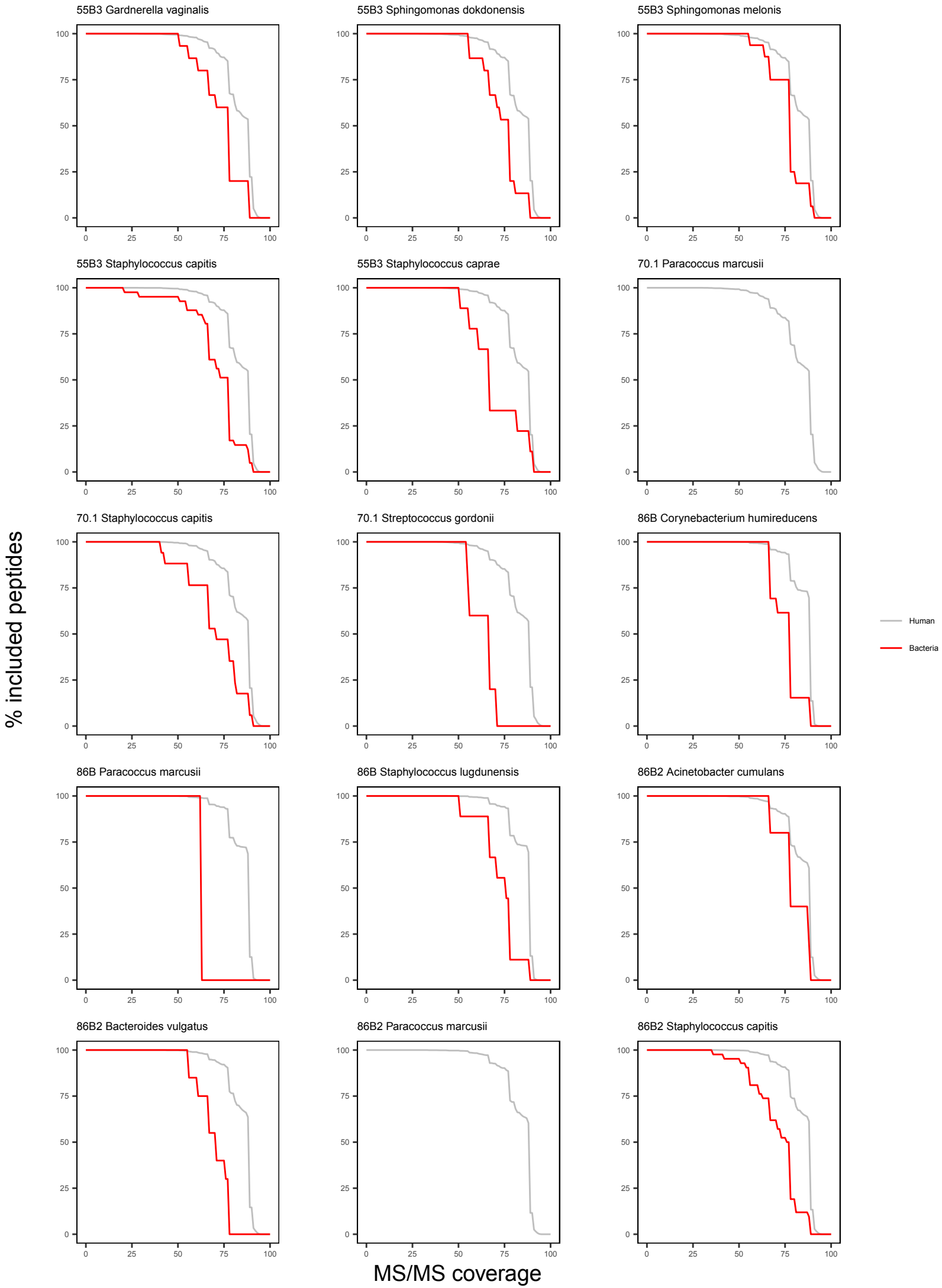


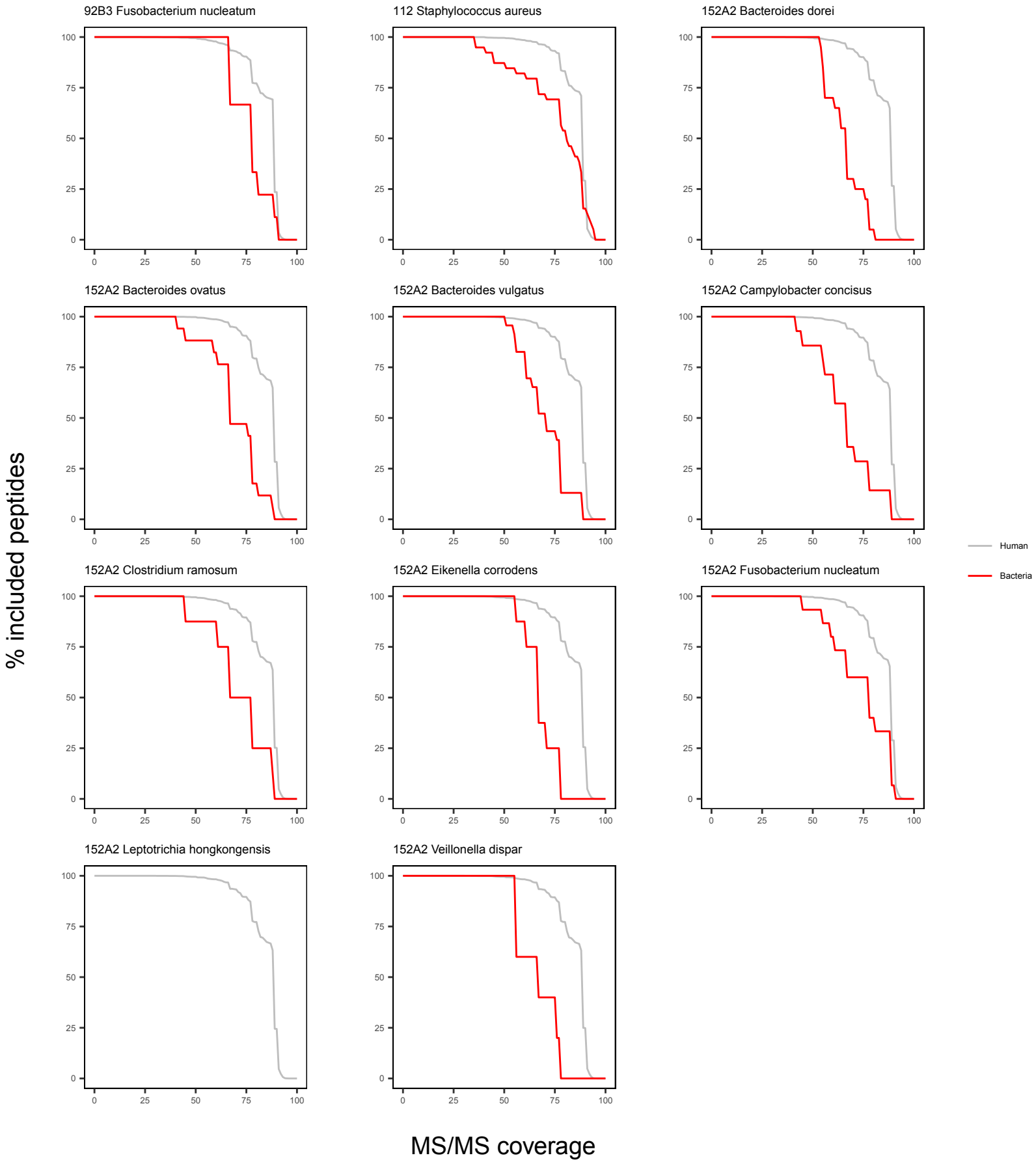
For each sample separately:

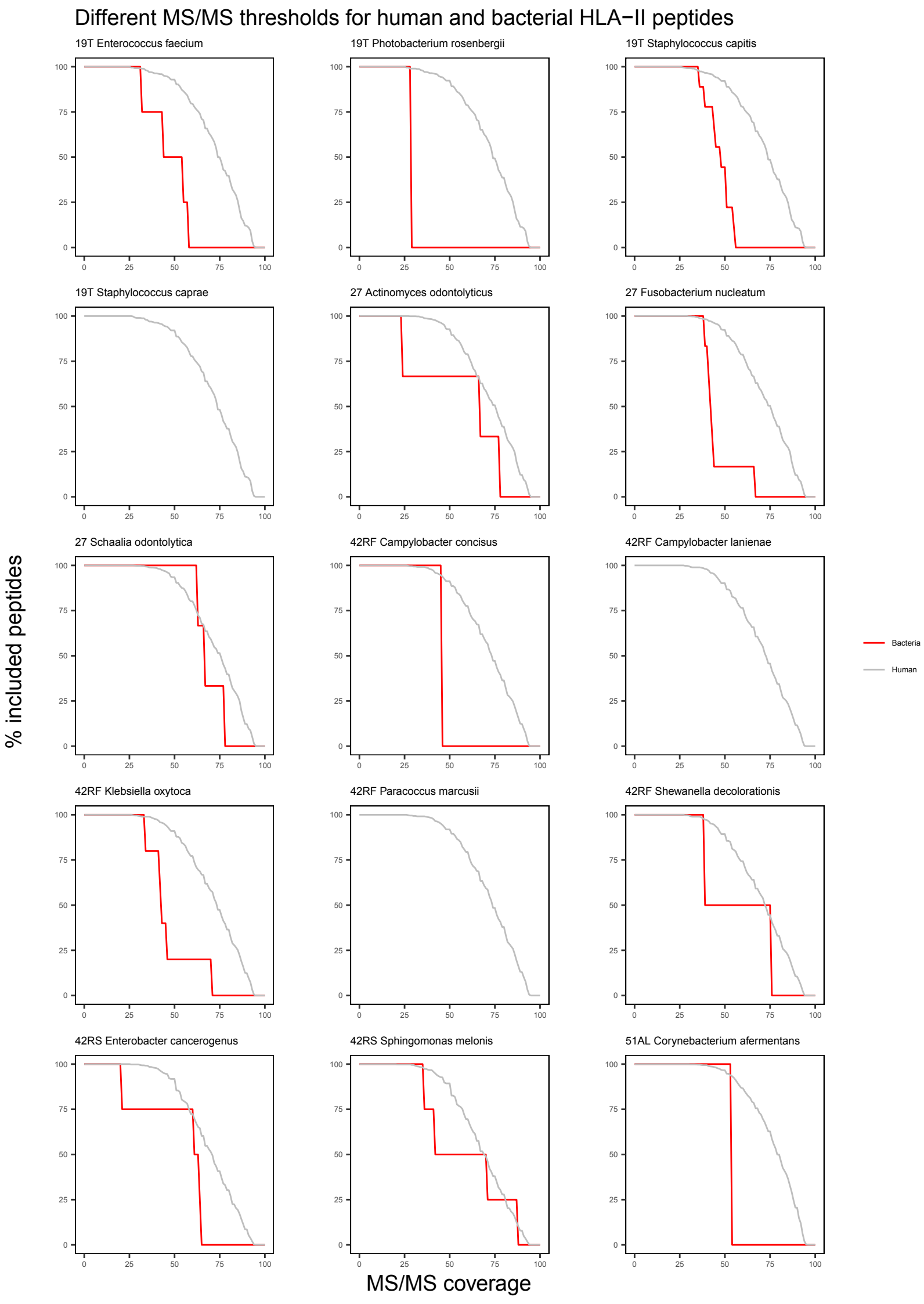

Different MS/MS thresholds for human and bacterial HLA−I peptides
19T Enterococcus faecium
19T Photobacterium rosenbergii
19T Staphylococcus capitis
19T Staphylococcus caprae
27 Actinomyces odontolyticus
27 Fusobacterium nucleatum
27 Schaalia odontolytica
42RF Campylobacter concisus
42RF Campylobacter lanienae
42RF Klebsiella oxytoca
42RF Paracoccus marcusii
42RF Shewanella decolorationis
42RS Enterobacter cancerogenus
42RS Sphingomonas melonis
51AL Corynebacterium afermentans
% included peptides
MS/MS coverage
Human
Bacteria

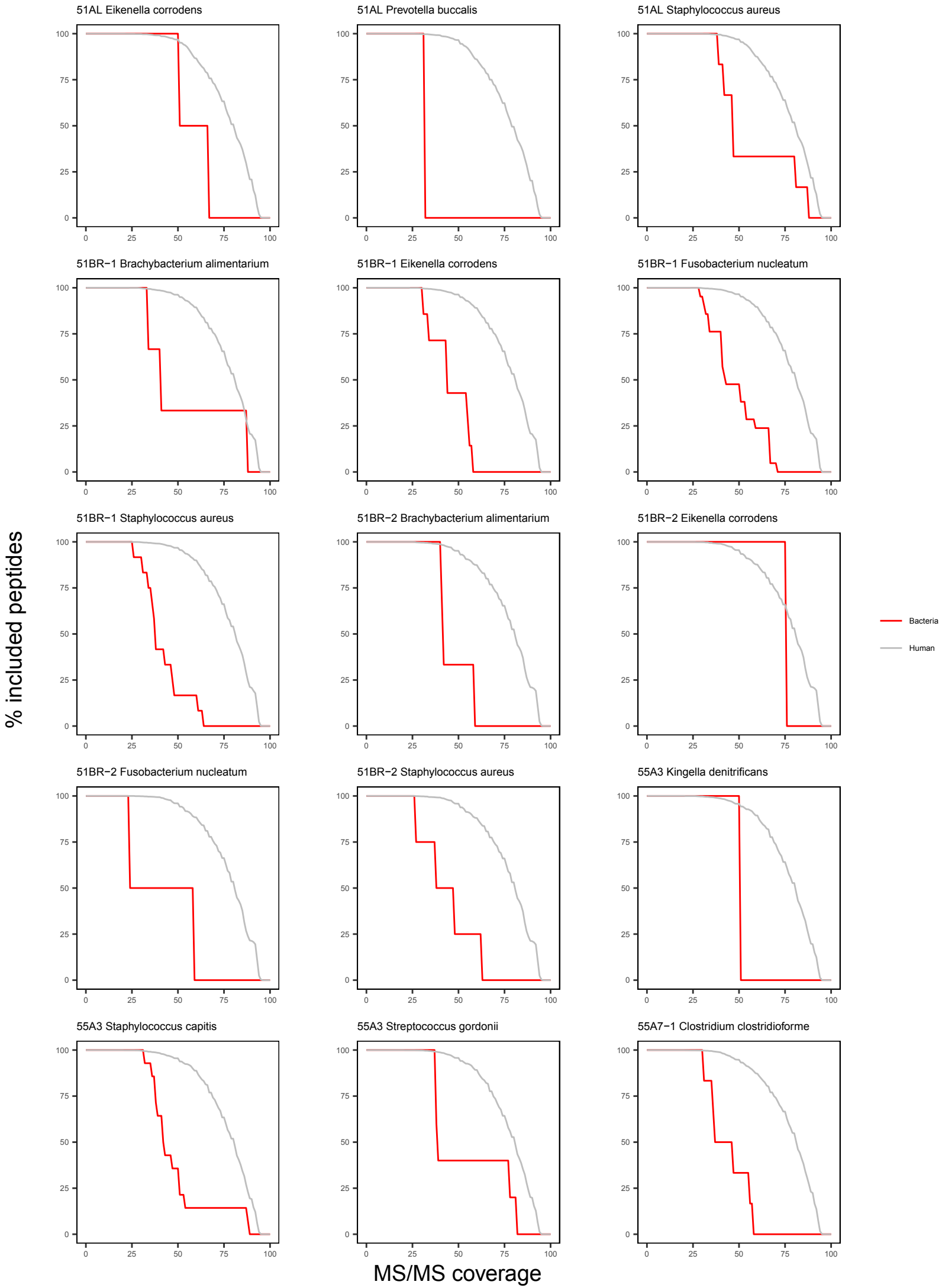

51AL Eikenella corrodens
51AL Prevotella buccalis
51AL Staphylococcus aureus
51BR−1 Brachybacterium alimentarium
51BR−1 Eikenella corrodens
51BR−1 Fusobacterium nucleatum
51BR−1 Staphylococcus aureus
51BR−2 Brachybacterium alimentarium
51BR−2 Eikenella corrodens
51BR−2 Fusobacterium nucleatum
51BR−2 Staphylococcus aureus
55A3 Kingella denitrificans
55A3 Staphylococcus capitis
55A3 Streptococcus gordonii
55A7−1 Clostridium clostridioforme
% included peptides
MS/MS coverage
Human
Bacteria

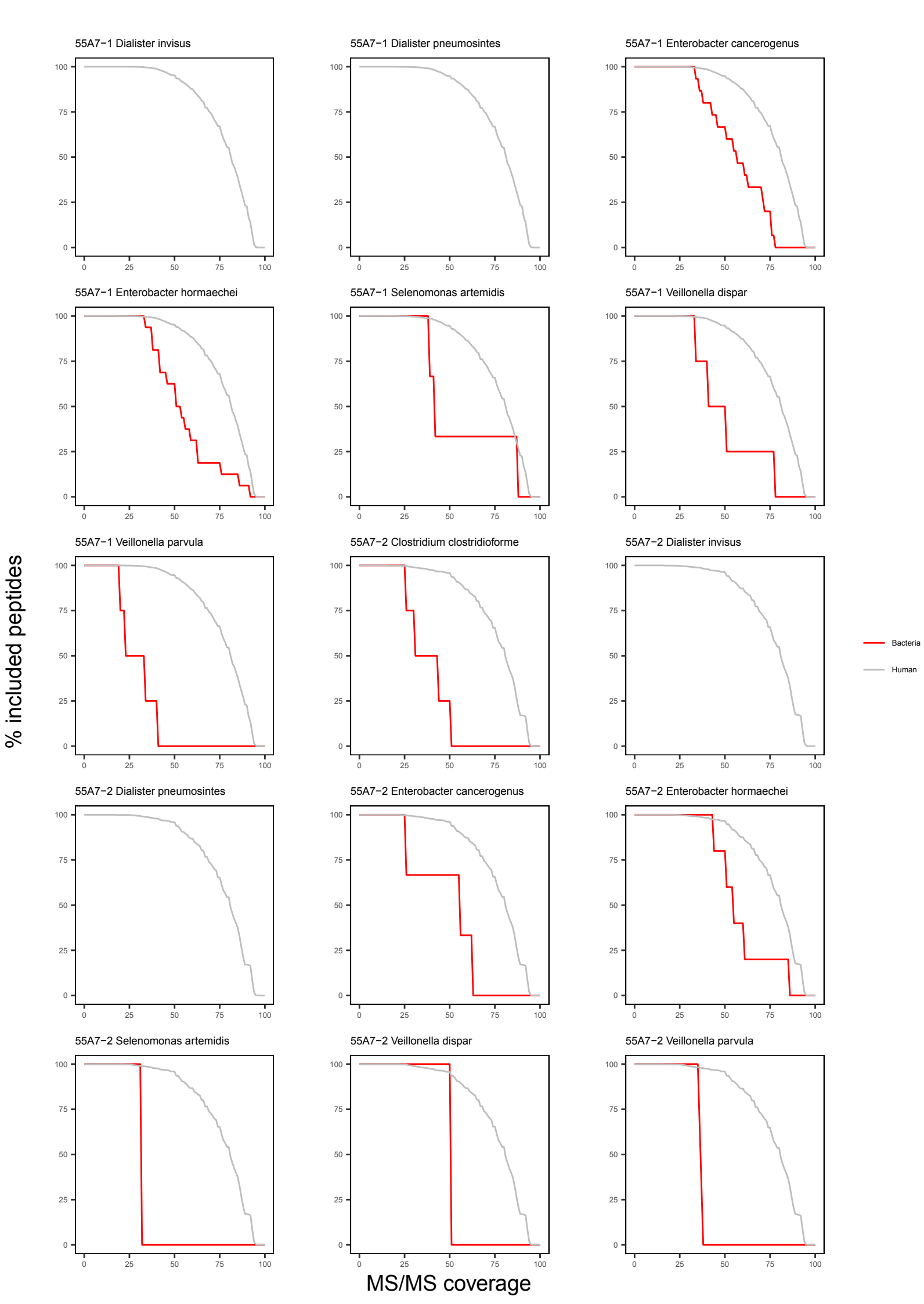

55A7−1 Dialister invisus
55A7−1 Dialister pneumosintes
55A7−1 Enterobacter cancerogenus
55A7−1 Enterobacter hormaechei
55A7−1 Selenomonas artemidis
55A7−1 Veillonella dispar
55A7−1 Veillonella parvula
55A7−2 Clostridium clostridioforme
55A7−2 Dialister invisus
55A7−2 Dialister pneumosintes
55A7−2 Enterobacter cancerogenus
55A7−2 Enterobacter hormaechei
55A7−2 Selenomonas artemidis
55A7−2 Veillonella dispar
55A7−2 Veillonella parvula
% included peptides
MS/MS coverage
Human
Bacteria

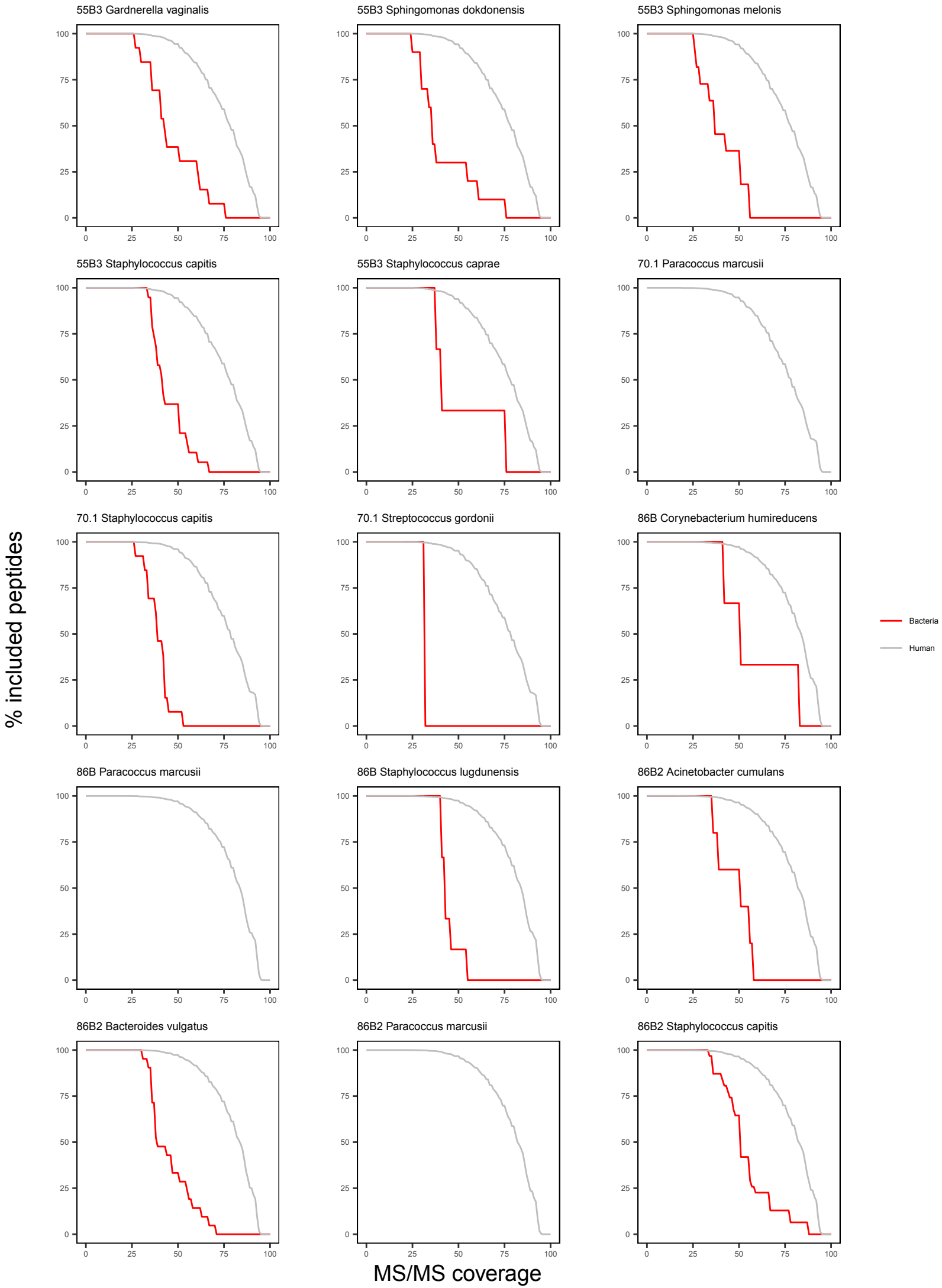

55B3 Gardnerella vaginalis
55B3 Sphingomonas dokdonensis
55B3 Sphingomonas melonis
55B3 Staphylococcus capitis
55B3 Staphylococcus caprae
70.1 Paracoccus marcusii
70.1 Staphylococcus capitis
70.1 Streptococcus gordonii
86B Corynebacterium humireducens
86B Paracoccus marcusii
86B Staphylococcus lugdunensis
86B2 Acinetobacter cumulans
86B2 Bacteroides vulgatus
86B2 Paracoccus marcusii
86B2 Staphylococcus capitis
% included peptides
MS/MS coverage
Human
Bacteria

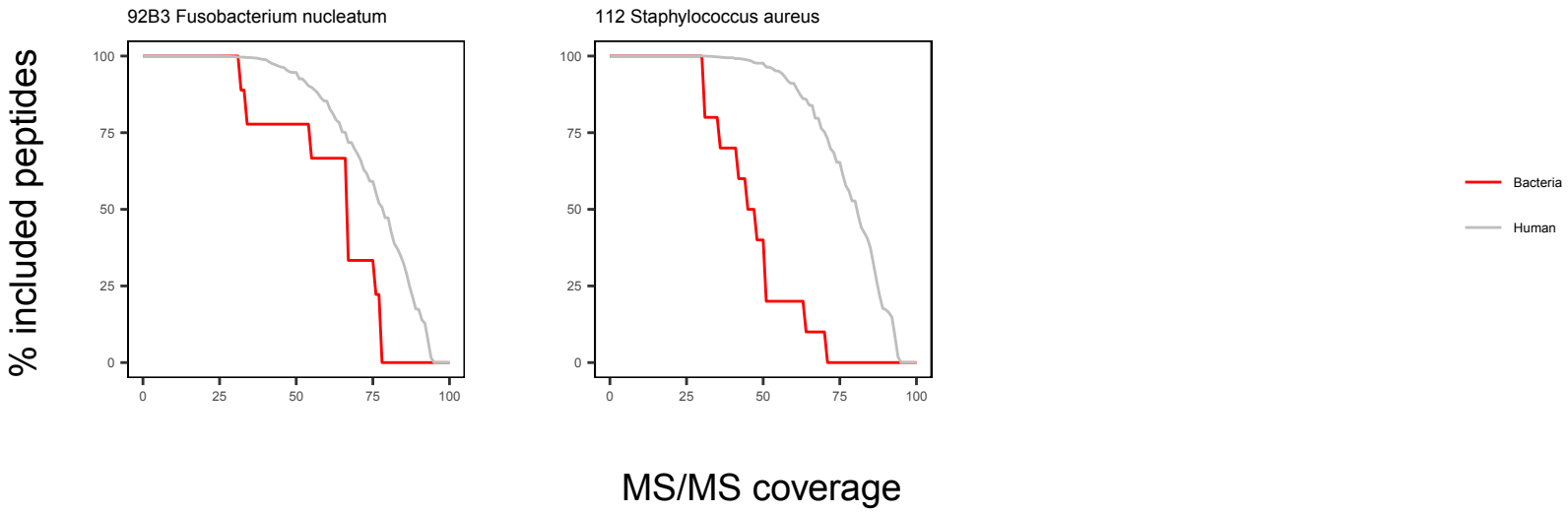

92B3 Fusobacterium nucleatum
112 Staphylococcus aureus
152A2 Bacteroides dorei
152A2 Bacteroides ovatus
152A2 Bacteroides vulgatus
152A2 Campylobacter concisus
152A2 Clostridium ramosum
152A2 Eikenella corrodens
152A2 Fusobacterium nucleatum
152A2 Leptotrichia hongkongensis
152A2 Veillonella dispar
% included peptides
MS/MS coverage
Human
Bacteria


Different MS/MS thresholds for human and bacterial HLA−II peptides
19T Enterococcus faecium
19T Photobacterium rosenbergii
19T Staphylococcus capitis
19T Staphylococcus caprae
27 Actinomyces odontolyticus
27 Fusobacterium nucleatum
27 Schaalia odontolytica
42RF Campylobacter concisus
42RF Campylobacter lanienae
42RF Klebsiella oxytoca
42RF Paracoccus marcusii
42RF Shewanella decolorationis
42RS Enterobacter cancerogenus
42RS Sphingomonas melonis
51AL Corynebacterium afermentans
% included peptides
MS/MS coverage
Bacteria
Human


51AL Eikenella corrodens
51AL Prevotella buccalis
51AL Staphylococcus aureus
51BR−1 Brachybacterium alimentarium
51BR−1 Eikenella corrodens
51BR−1 Fusobacterium nucleatum
51BR−1 Staphylococcus aureus
51BR−2 Brachybacterium alimentarium
51BR−2 Eikenella corrodens
51BR−2 Fusobacterium nucleatum
51BR−2 Staphylococcus aureus
55A3 Kingella denitrificans
55A3 Staphylococcus capitis
55A3 Streptococcus gordonii
55A7−1 Clostridium clostridioforme
% included peptides
MS/MS coverage
Bacteria
Human


55A7−1 Dialister invisus
55A7−1 Dialister pneumosintes
55A7−1 Enterobacter cancerogenus
55A7−1 Enterobacter hormaechei
55A7−1 Selenomonas artemidis
55A7−1 Veillonella dispar
55A7−1 Veillonella parvula
55A7−2 Clostridium clostridioforme
55A7−2 Dialister invisus
55A7−2 Dialister pneumosintes
55A7−2 Enterobacter cancerogenus
55A7−2 Enterobacter hormaechei
55A7−2 Selenomonas artemidis
55A7−2 Veillonella dispar
55A7−2 Veillonella parvula
% included peptides
MS/MS coverage
Bacteria
Human


55B3 Gardnerella vaginalis
55B3 Sphingomonas dokdonensis
55B3 Sphingomonas melonis
55B3 Staphylococcus capitis
55B3 Staphylococcus caprae
70.1 Paracoccus marcusii
70.1 Staphylococcus capitis
70.1 Streptococcus gordonii
86B Corynebacterium humireducens
86B Paracoccus marcusii
86B Staphylococcus lugdunensis
86B2 Acinetobacter cumulans
86B2 Bacteroides vulgatus
86B2 Paracoccus marcusii
86B2 Staphylococcus capitis
% included peptides
MS/MS coverage
Bacteria
Human

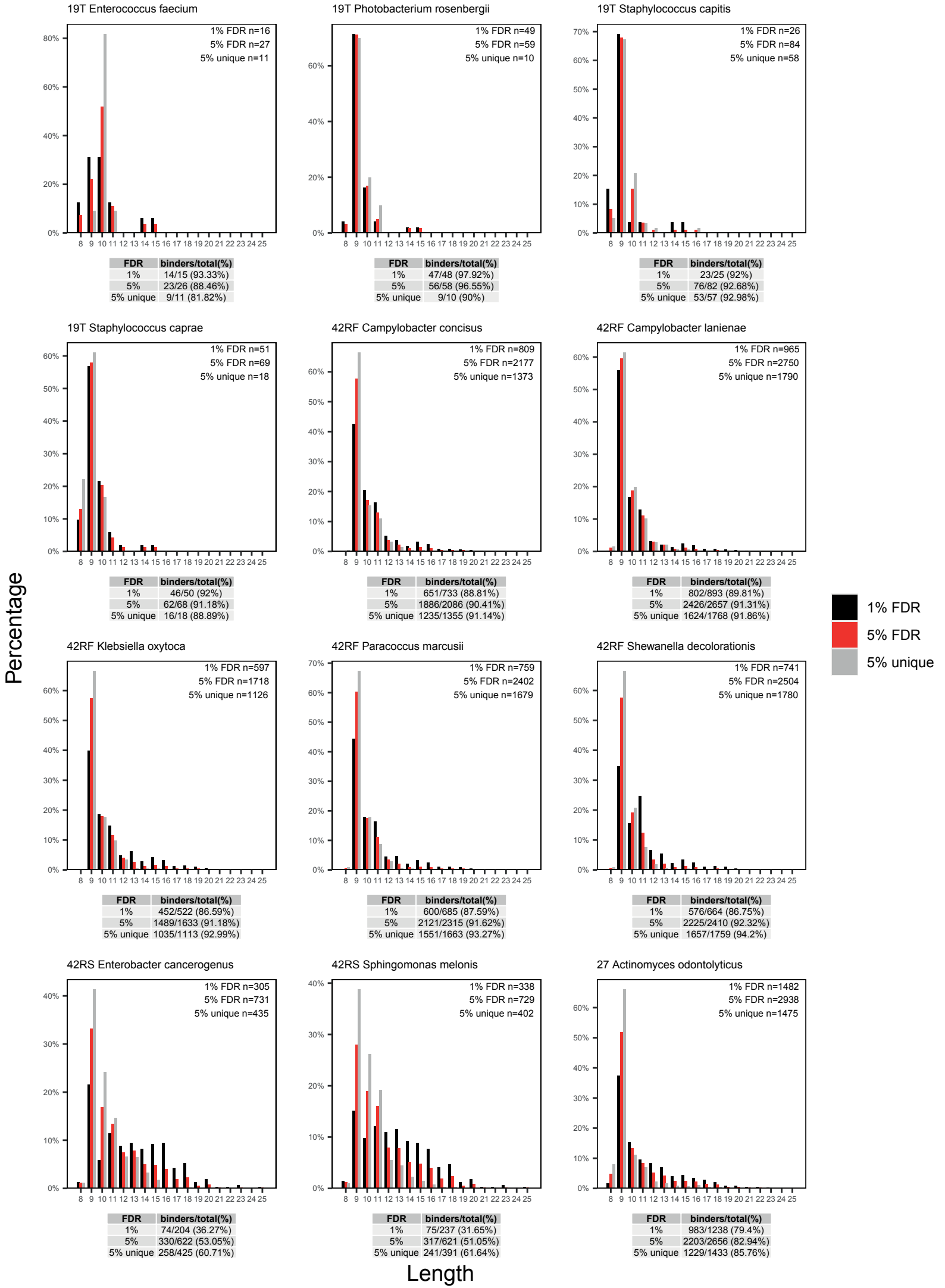

92B3 Fusobacterium nucleatum
112 Staphylococcus aureus
% included peptides
100
75
50
25
0
0
25
50
75
100
MS/MS coverage
Bacteria
Human

We believe that using the quality filters employed in the study is better than analyzing the data with a lower FDR.

While the number of peptides identified when using 1% FDR was significantly lower than when using 5% FDR, there was no difference in terms of quality. Therefore, reducing the FDR in the analysis can result in many false negatives compared to using the quality filtrations used in our study.

Below, we compared the peptide length, HLA allele distribution and percentage of peptides that were predicted to bind the patient's HLA alleles (using NetMHCpan) for peptides that were identified in 1% FDR, 5% FDR and uniquely identified with 5% FDR. We did not find a difference in their length and HLA allele distributions or in their ability to bind the patient's HLA alleles, emphasizing their high quality.

Length distribution of all identified peptides using different FDR



FDR	binders/total(%)
1%	14/15 (93.33%)
5%	23/26 (88.46%)
5% unique	9/11 (81.82%)

FDR	binders/total(%)
1%	47/48 (97.92%)
5%	56/58 (96.55%)
5% unique	9/10 (90%)

FDR	binders/total(%)
1%	23/25 (92%)
5%	76/82 (92.68%)
5% unique	53/57 (92.98%)

FDR	binders/total(%)
1%	46/50 (92%)
5%	62/68 (91.18%)
5% unique	16/18 (88.89%)

FDR	binders/total(%)
1%	651/733 (88.81%)
5%	1886/2086 (90.41%)
5% unique	1235/1355 (91.14%)

FDR	binders/total(%)
1%	802/893 (89.81%)
5%	2426/2657 (91.31%)
5% unique	1624/1768 (91.86%)

FDR	binders/total(%)
1%	452/522 (86.59%)
5%	1489/1633 (91.18%)
5% unique	1035/1113 (92.99%)

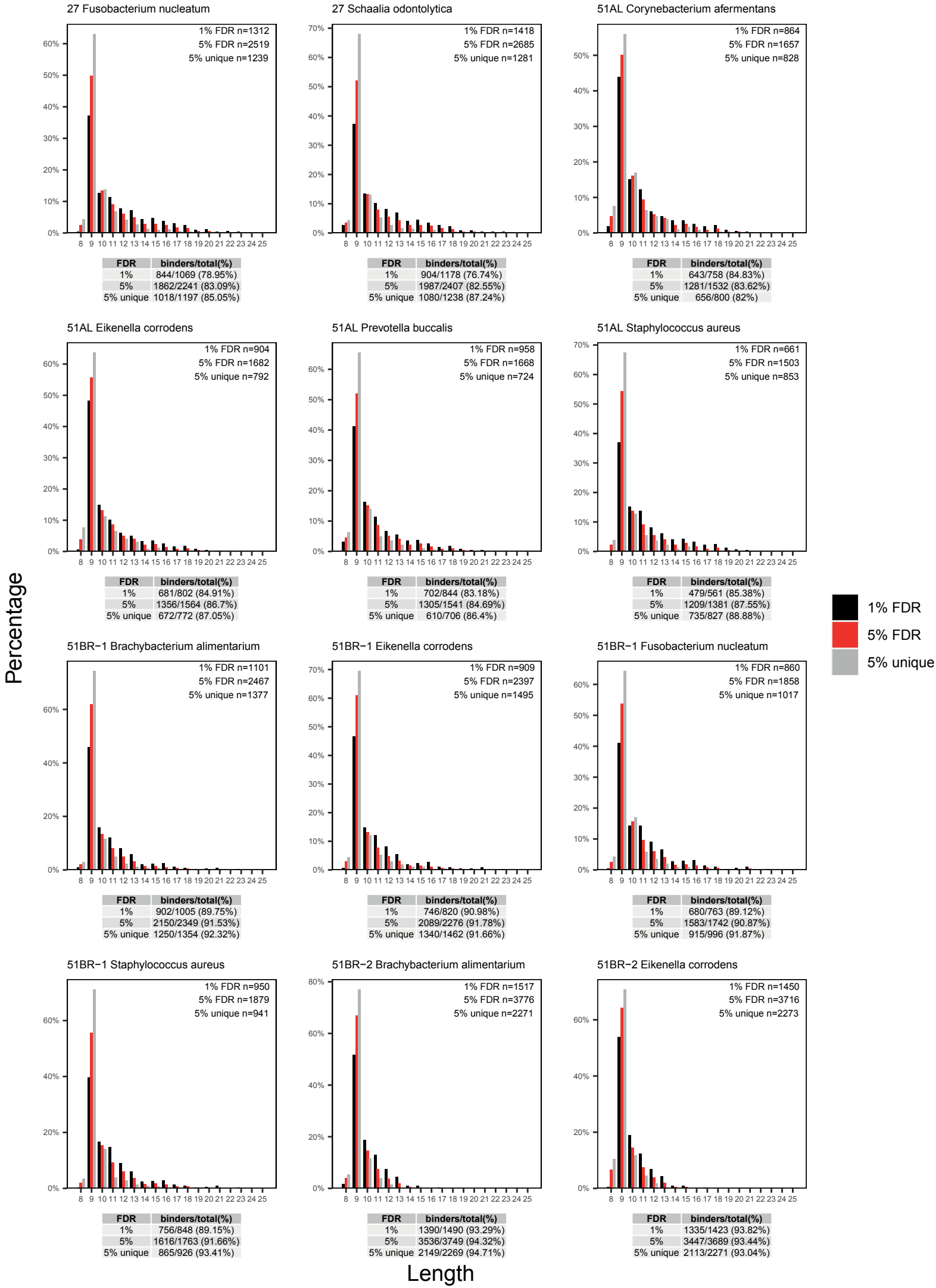
FDR	binders/total(%)
1%	600/685 (87.59%)
5%	2121/2315 (91.62%)
5% unique	1551/1663 (93.27%)

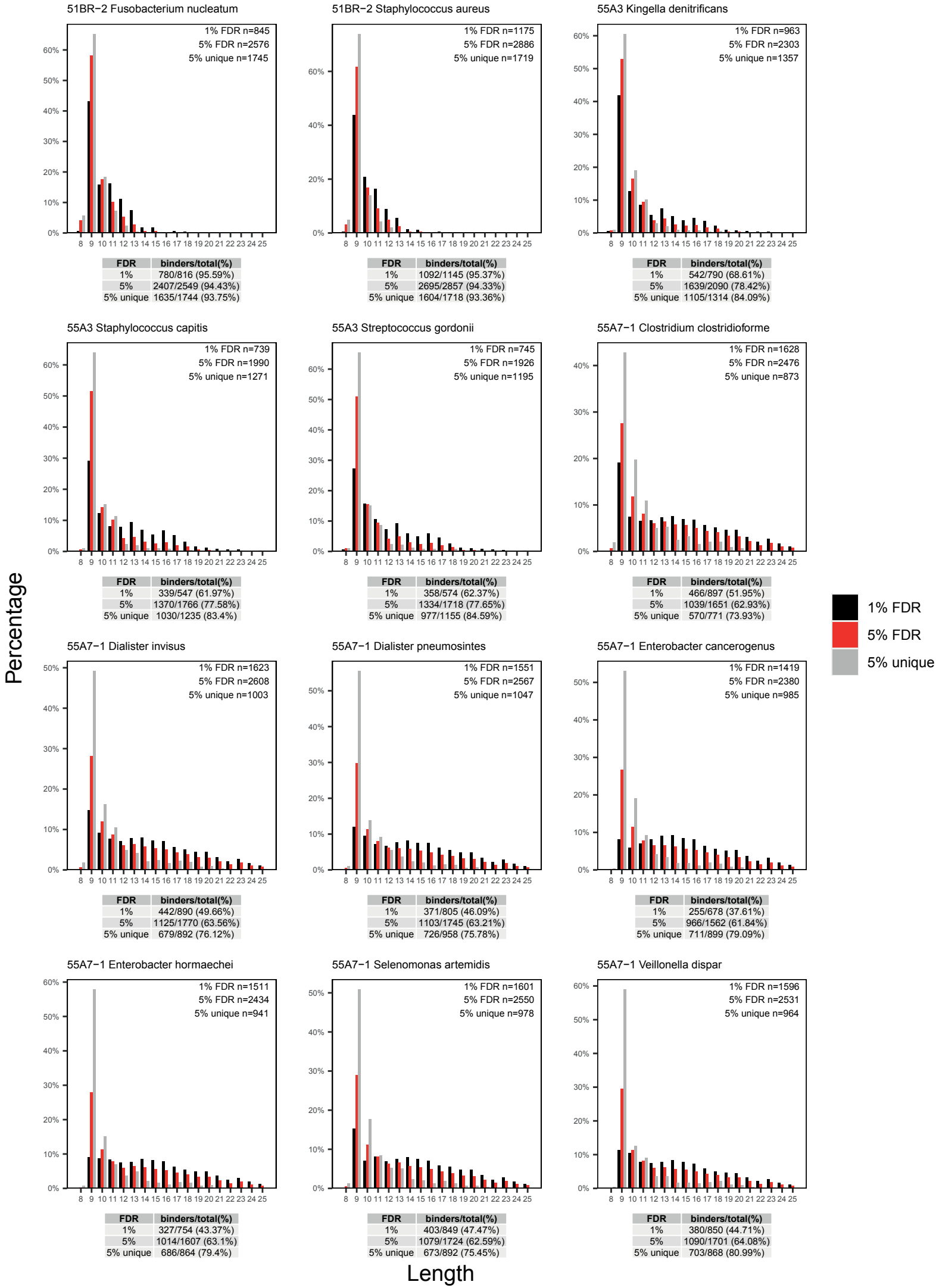
FDR	binders/total(%)
1%	576/664 (86.75%)
5%	2225/2410 (92.32%)
5% unique	1657/1759 (94.2%)

FDR	binders/total(%)
1%	74/204 (36.27%)
5%	330/622 (53.05%)
5% unique	258/425 (60.71%)

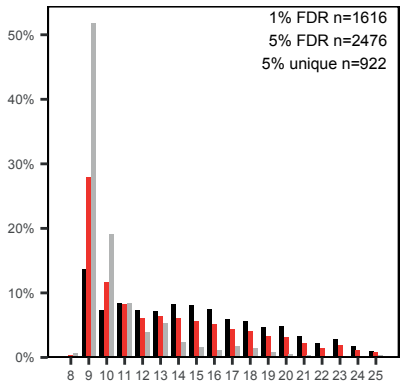
FDR	binders/total(%)
1%	75/237 (31.65%)
5%	317/621 (51.05%)
5% unique	241/391 (61.64%)

FDR	binders/total(%)
1%	983/1238 (79.4%)
5%	2203/2656 (82.94%)
5% unique	1229/1433 (85.76%)


27 Fusobacterium nucleatum
1% FDR n=1312
5% FDR n=2519
5% unique n=1239
FDR | binders/total(%)
1% | 844/1069 (78.95%)
5% | 1862/2241 (83.09%)
5% unique | 1018/1197 (85.05%)
27 Schaalia odontolytica
1% FDR n=1418
5% FDR n=2685
5% unique n=1281
FDR | binders/total(%)
1% | 904/1178 (76.74%)
5% | 1987/2407 (82.55%)
5% unique | 1080/1238 (87.24%)
51AL Corynebacterium afermentans
1% FDR n=864
5% FDR n=1657
5% unique n=828
FDR | binders/total(%)
1% | 643/758 (84.83%)
5% | 1281/1532 (83.62%)
5% unique | 656/800 (82%)
51AL Eikenella corrodens
1% FDR n=904
5% FDR n=1682
5% unique n=792
FDR | binders/total(%)
1% | 681/802 (84.91%)
5% | 1356/1564 (86.7%)
5% unique | 672/772 (87.05%)
51AL Prevotella buccalis
1% FDR n=958
5% FDR n=1668
5% unique n=724
FDR | binders/total(%)
1% | 702/844 (83.18%)
5% | 1305/1541 (84.69%)
5% unique | 610/706 (86.4%)
51AL Staphylococcus aureus
1% FDR n=661
5% FDR n=1503
5% unique n=853
FDR | binders/total(%)
1% | 479/561 (85.38%)
5% | 1209/1381 (87.55%)
5% unique | 735/827 (88.88%)
51BR−1 Brachybacterium alimentarium
1% FDR n=1101
5% FDR n=2467
5% unique n=1377
FDR | binders/total(%)
1% | 902/1005 (89.75%)
5% | 2150/2349 (91.53%)
5% unique | 1250/1354 (92.32%)
51BR−1 Eikenella corrodens
1% FDR n=909
5% FDR n=2397
5% unique n=1495
FDR | binders/total(%)
1% | 746/820 (90.98%)
5% | 2089/2276 (91.78%)
5% unique | 1340/1462 (91.66%)
51BR−1 Fusobacterium nucleatum
1% FDR n=860
5% FDR n=1858
5% unique n=1017
FDR | binders/total(%)
1% | 680/763 (89.12%)
5% | 1583/1742 (90.87%)
5% unique | 915/996 (91.87%)
51BR−1 Staphylococcus aureus
1% FDR n=950
5% FDR n=1879
5% unique n=941
FDR | binders/total(%)
1% | 756/848 (89.15%)
5% | 1616/1763 (91.66%)
5% unique | 865/926 (93.41%)
51BR−2 Brachybacterium alimentarium
1% FDR n=1517
5% FDR n=3776
5% unique n=2271
FDR | binders/total(%)
1% | 1390/1490 (93.29%)
5% | 3536/3749 (94.32%)
5% unique | 2149/2269 (94.71%)
51BR−2 Eikenella corrodens
1% FDR n=1450
5% FDR n=3716
5% unique n=2273
FDR | binders/total(%)
1% | 1335/1423 (93.82%)
5% | 3447/3689 (93.44%)
5% unique | 2113/2271 (93.04%)
Percentage
Length
1% FDR
5% FDR
5% unique

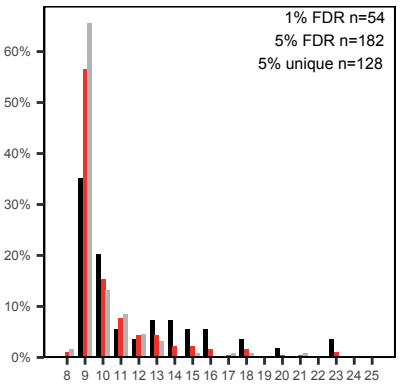

51BR−2 Fusobacterium nucleatum
1% FDR n=845
5% FDR n=2576
5% unique n=1745
FDR | binders/total(%)
1% | 780/816 (95.59%)
5% | 2407/2549 (94.43%)
5% unique | 1635/1744 (93.75%)
51BR−2 Staphylococcus aureus
1% FDR n=1175
5% FDR n=2886
5% unique n=1719
FDR | binders/total(%)
1% | 1092/1145 (95.37%)
5% | 2695/2857 (94.33%)
5% unique | 1604/1718 (93.36%)
55A3 Kingella denitrificans
1% FDR n=963
5% FDR n=2303
5% unique n=1357
FDR | binders/total(%)
1% | 542/790 (68.61%)
5% | 1639/2090 (78.42%)
5% unique | 1105/1314 (84.09%)
55A3 Staphylococcus capitis
1% FDR n=739
5% FDR n=1990
5% unique n=1271
FDR | binders/total(%)
1% | 339/547 (61.97%)
5% | 1370/1766 (77.58%)
5% unique | 1030/1235 (83.4%)
55A3 Streptococcus gordonii
1% FDR n=745
5% FDR n=1926
5% unique n=1195
FDR | binders/total(%)
1% | 358/574 (62.37%)
5% | 1334/1718 (77.65%)
5% unique | 977/1155 (84.59%)
55A7−1 Clostridium clostridioforme
1% FDR n=1628
5% FDR n=2476
5% unique n=873
FDR | binders/total(%)
1% | 466/897 (51.95%)
5% | 1039/1651 (62.93%)
5% unique | 570/771 (73.93%)
55A7−1 Dialister invisus
1% FDR n=1623
5% FDR n=2608
5% unique n=1003
FDR | binders/total(%)
1% | 442/890 (49.66%)
5% | 1125/1770 (63.56%)
5% unique | 679/892 (76.12%)
55A7−1 Dialister pneumosintes
1% FDR n=1551
5% FDR n=2567
5% unique n=1047
FDR | binders/total(%)
1% | 371/805 (46.09%)
5% | 1103/1745 (63.21%)
5% unique | 726/958 (75.78%)
55A7−1 Enterobacter cancerogenus
1% FDR n=1419
5% FDR n=2380
5% unique n=985
FDR | binders/total(%)
1% | 255/678 (37.61%)
5% | 966/1562 (61.84%)
5% unique | 711/899 (79.09%)
55A7−1 Enterobacter hormaechei
1% FDR n=1511
5% FDR n=2434
5% unique n=941
FDR | binders/total(%)
1% | 327/754 (43.37%)
5% | 1014/1607 (63.1%)
5% unique | 686/864 (79.4%)
55A7−1 Selenomonas artemidis
1% FDR n=1601
5% FDR n=2550
5% unique n=978
FDR | binders/total(%)
1% | 403/849 (47.47%)
5% | 1079/1724 (62.59%)
5% unique | 673/892 (75.45%)
55A7−1 Veillonella dispar
1% FDR n=1596
5% FDR n=2531
5% unique n=964
FDR | binders/total(%)
1% | 380/850 (44.71%)
5% | 1090/1701 (64.08%)
5% unique | 703/868 (80.99%)
Percentage
Length
1% FDR
5% FDR
5% unique

55A7−1 Veillonella parvula



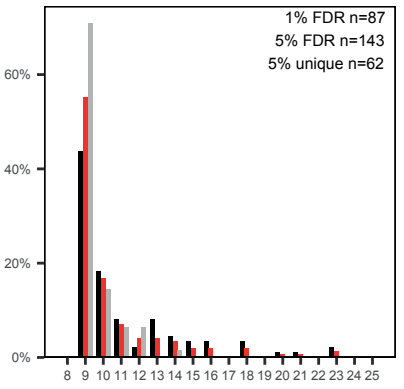
FDR	binders/total(%)
1%	409/843 (48.52%)
5%	1039/1651 (62.93%)
5% unique	643/846 (76%)

55A7−2 Clostridium clostridioforme



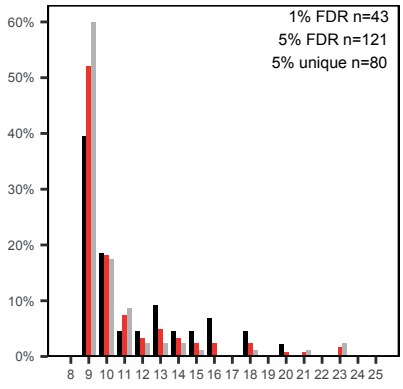
FDR	binders/total(%)
1%	31/43 (72.09%)
5%	142/167 (85.03%)
5% unique	111/124 (89.52%)

55A7−2 Dialister invisus



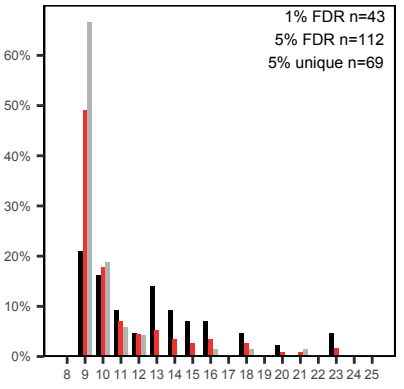
FDR	binders/total(%)
1%	57/74 (77.03%)
5%	110/130 (84.62%)
5% unique	57/62 (91.94%)

55A7−2 Dialister pneumosintes



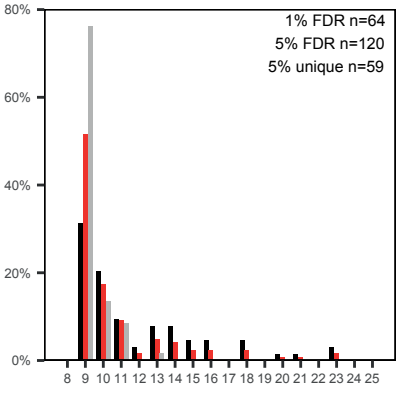
FDR	binders/total(%)
1%	25/35 (71.43%)
5%	89/108 (82.41%)
5% unique	66/75 (88%)

55A7−2 Enterobacter cancerogenus



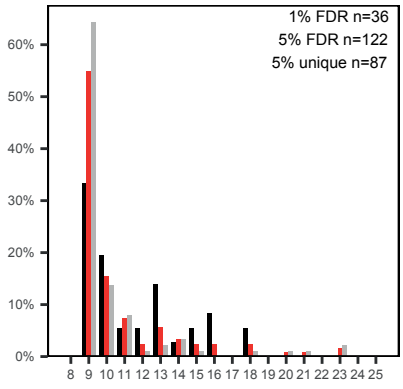
FDR	binders/total(%)
1%	20/32 (62.5%)
5%	81/98 (82.65%)
5% unique	61/66 (92.42%)

55A7−2 Enterobacter hormaechei



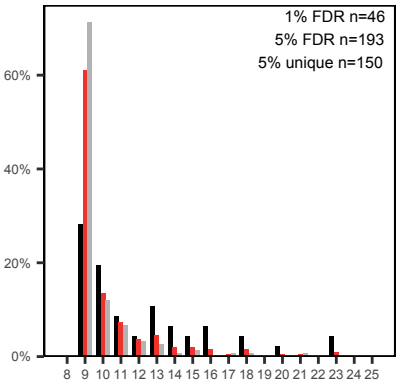
FDR	binders/total(%)
1%	37/51 (72.55%)
5%	90/107 (84.11%)
5% unique	56/59 (94.92%)

55A7−2 Selenomonas artemidis



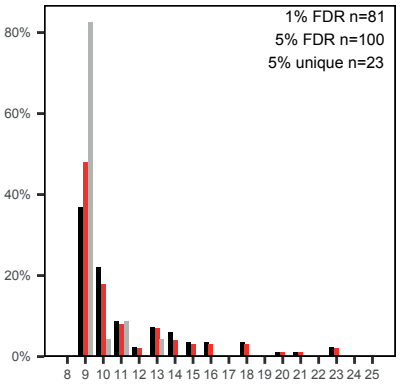
FDR	binders/total(%)
1%	19/29 (65.52%)
5%	90/109 (82.57%)
5% unique	72/81 (88.89%)

55A7−2 Veillonella dispar



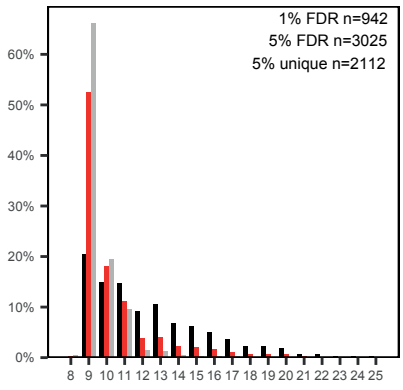
FDR	binders/total(%)
1%	21/36 (58.33%)
5%	148/178 (83.15%)
5% unique	130/145 (89.66%)

55A7−2 Veillonella parvula



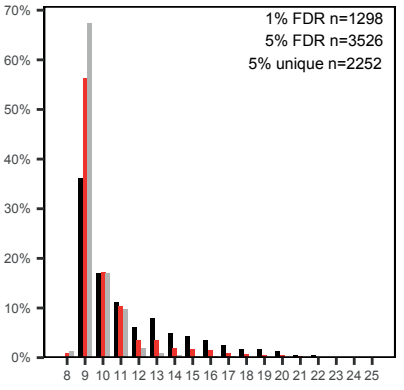
FDR	binders/total(%)
1%	50/68 (73.53%)
5%	65/87 (74.71%)
5% unique	18/23 (78.26%)

55B3 Gardnerella vaginalis



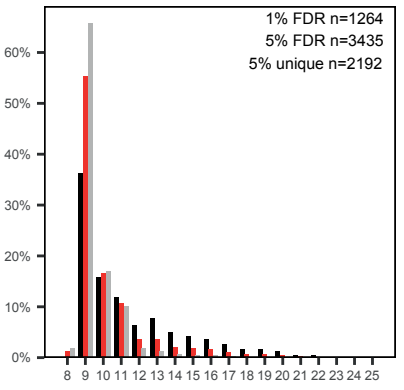
FDR	binders/total(%)
1%	421/723 (58.23%)
5%	2205/2795 (78.89%)
5% unique	1793/2088 (85.87%)

55B3 Sphingomonas dokdonensis



FDR	binders/total(%)
1%	766/1084 (70.66%)
5%	2678/3296 (81.25%)
5% unique	1917/2226 (86.12%)

55B3 Sphingomonas melonis



FDR	binders/total(%)
1%	739/1055 (70.05%)
5%	2541/3205 (79.28%)
5% unique	1806/2162 (83.53%)

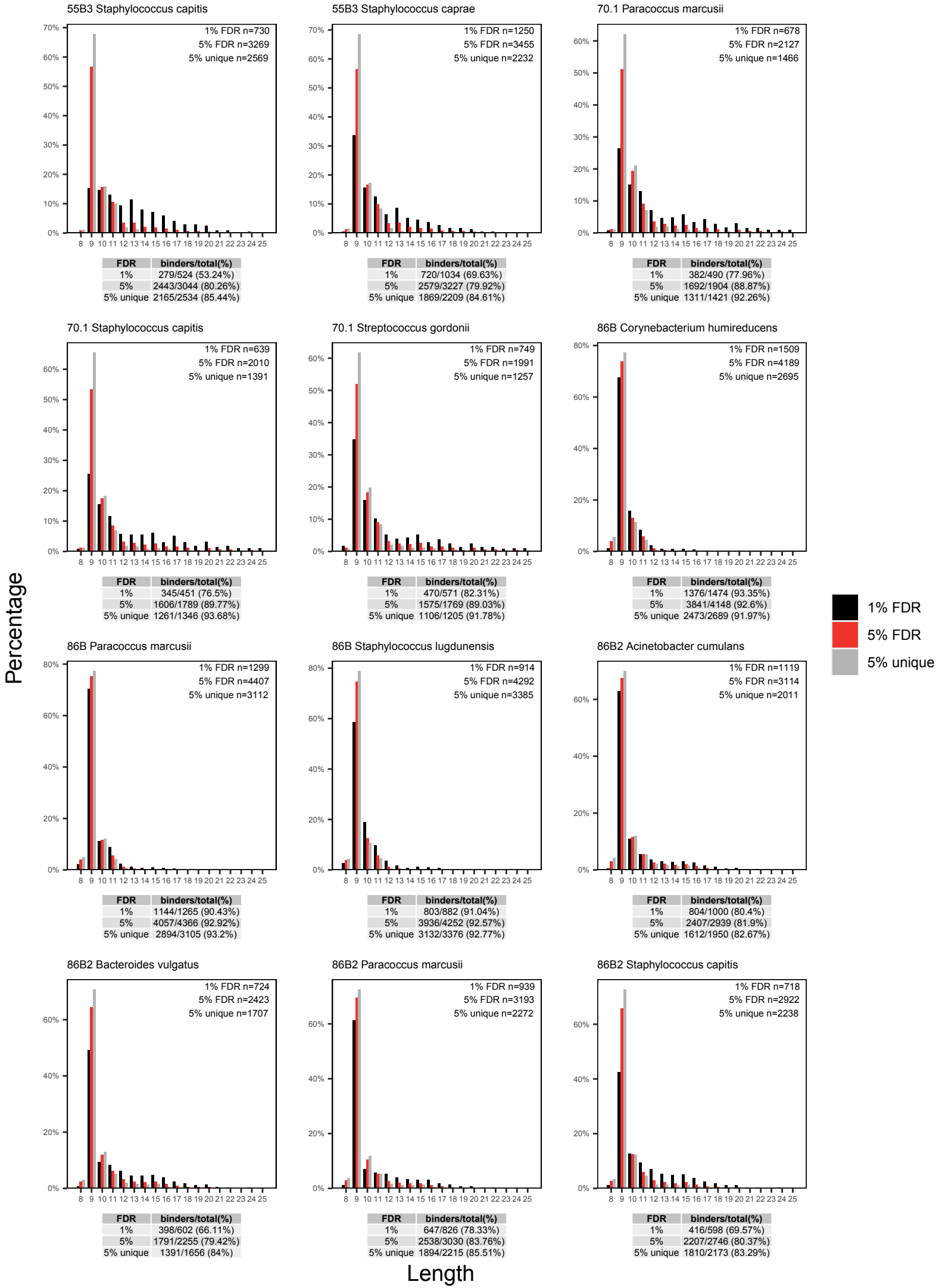
Percentage

Length

1% FDR

5% FDR

5% unique



FDR	binders/total(%)
1%	279/524 (53.24%)
5%	2443/3044 (80.26%)
5% unique	2165/2534 (85.44%)

FDR	binders/total(%)
1%	720/1034 (69.63%)
5%	2579/3227 (79.92%)
5% unique	1869/2209 (84.61%)

FDR	binders/total(%)
1%	382/490 (77.96%)
5%	1692/1904 (88.87%)
5% unique	1311/1421 (92.26%)

FDR	binders/total(%)
1%	345/451 (76.5%)
5%	1606/1789 (89.77%)
5% unique	1261/1346 (93.68%)

FDR	binders/total(%)
1%	470/571 (82.31%)
5%	1575/1769 (89.03%)
5% unique	1106/1205 (91.78%)

FDR	binders/total(%)
1%	1376/1474 (93.35%)
5%	3841/4148 (92.6%)
5% unique	2473/2689 (91.97%)

FDR	binders/total(%)
1%	1144/1265 (90.43%)
5%	4057/4366 (92.92%)
5% unique	2894/3105 (93.2%)

FDR	binders/total(%)
1%	803/882 (91.04%)
5%	3936/4252 (92.57%)
5% unique	3132/3376 (92.77%)

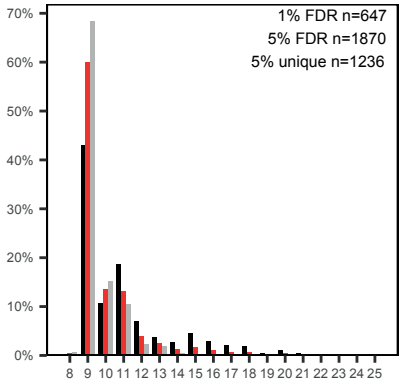
FDR	binders/total(%)
1%	804/1000 (80.4%)
5%	2407/2939 (81.9%)
5% unique	1612/1950 (82.67%)

FDR	binders/total(%)
1%	398/602 (66.11%)
5%	1791/2255 (79.42%)
5% unique	1391/1656 (84%)

FDR	binders/total(%)
1%	647/826 (78.33%)
5%	2538/3030 (83.76%)
5% unique	1894/2215 (85.51%)

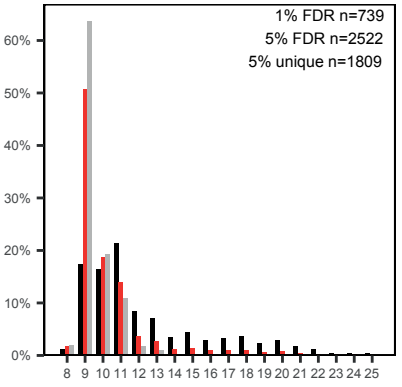
FDR	binders/total(%)
1%	416/598 (69.57%)
5%	2207/2746 (80.37%)
5% unique	1810/2173 (83.29%)

92B3 Fusobacterium nucleatum



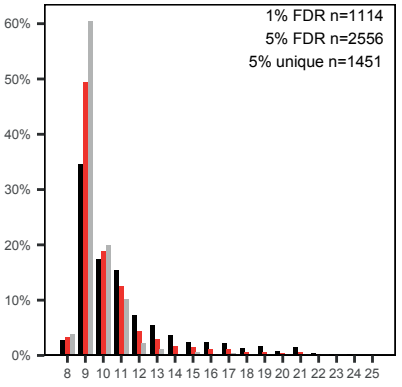
FDR	binders/total(%)
1%	474/558 (84.95%)
5%	1610/1778 (90.55%)
5% unique	1137/1228 (92.59%)

112 Staphylococcus aureus



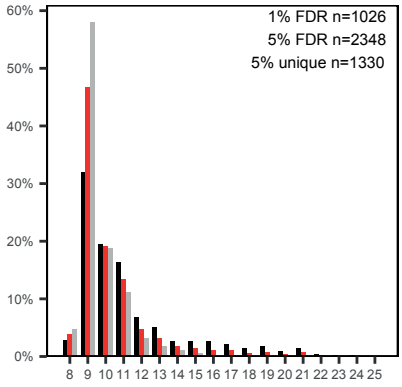
FDR	binders/total(%)
1%	392/558 (70.25%)
5%	2064/2339 (88.24%)
5% unique	1675/1793 (93.42%)

152A2 Bacteroides dorei



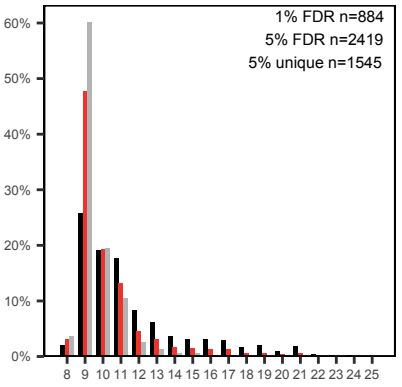
FDR	binders/total(%)
1%	813/967 (84.07%)
5%	2125/2387 (89.02%)
5% unique	1313/1427 (92.01%)

152A2 Bacteroides ovatus



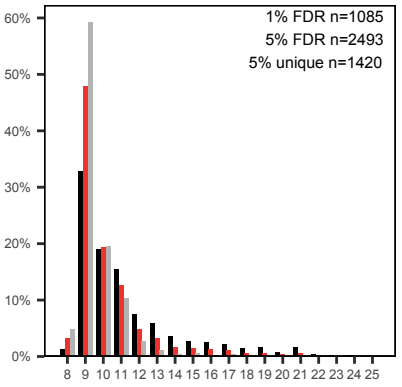
FDR	binders/total(%)
1%	728/879 (82.82%)
5%	1911/2189 (87.3%)
5% unique	1183/1315 (89.96%)

152A2 Bacteroides vulgatus



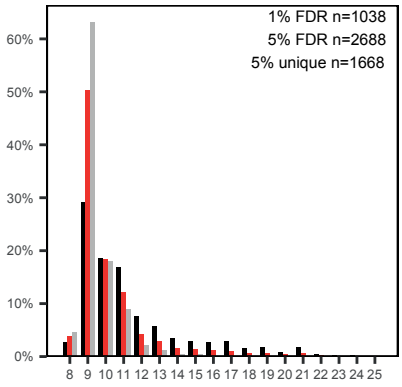
FDR	binders/total(%)
1%	597/734 (81.34%)
5%	1986/2250 (88.27%)
5% unique	1391/1522 (91.39%)

152A2 Campylobacter concisus



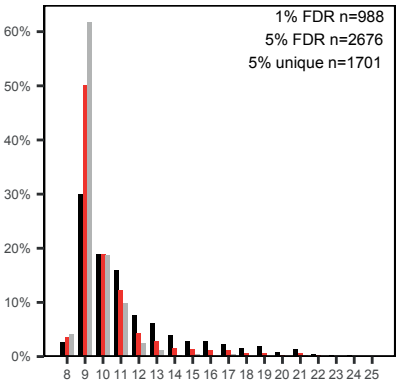
FDR	binders/total(%)
1%	781/934 (83.62%)
5%	2059/2324 (88.6%)
5% unique	1282/1399 (91.64%)

152A2 Clostridium ramosum



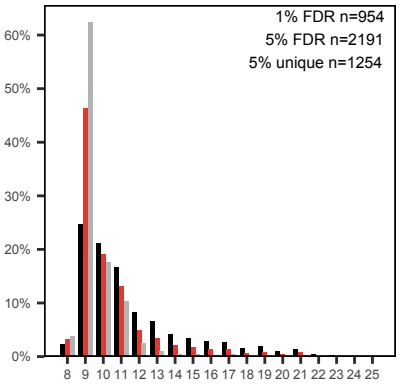
FDR	binders/total(%)
1%	726/876 (82.88%)
5%	2220/2517 (88.2%)
5% unique	1499/1651 (90.79%)

152A2 Eikenella corrodens



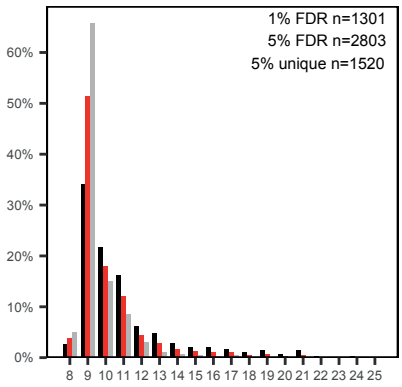
FDR	binders/total(%)
1%	696/842 (82.66%)
5%	2237/2509 (89.16%)
5% unique	1545/1675 (92.24%)

152A2 Fusobacterium nucleatum



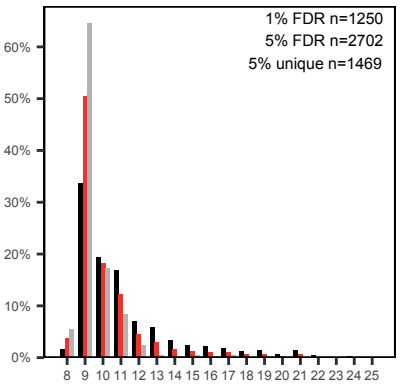
FDR	binders/total(%)
1%	662/803 (82.44%)
5%	1787/2025 (88.25%)
5% unique	1132/1234 (91.73%)

152A2 Leptotrichia hongkongensis



FDR	binders/total(%)
1%	968/1152 (84.03%)
5%	2351/2642 (88.99%)
5% unique	1389/1503 (92.42%)

152A2 Veillonella dispar

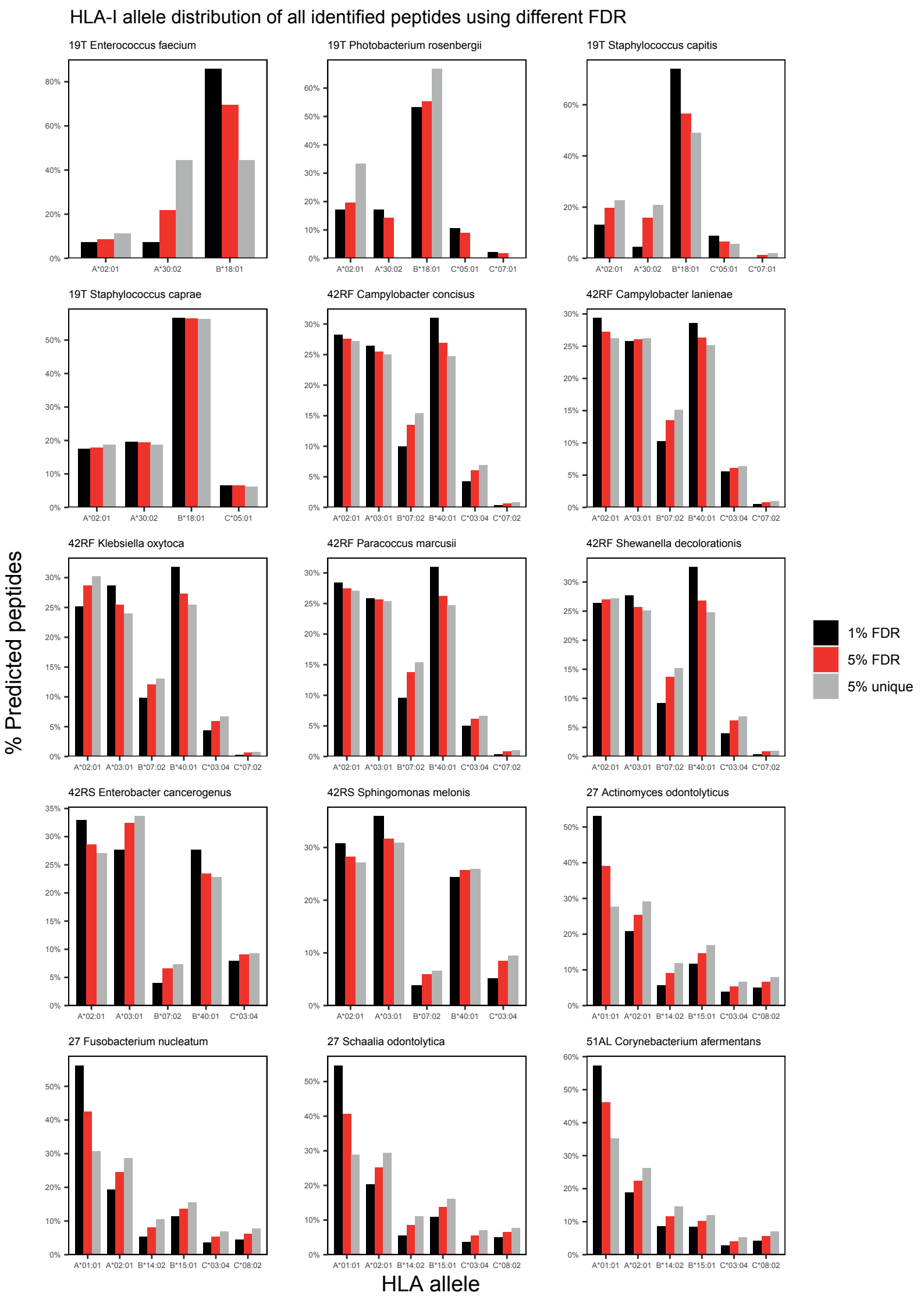


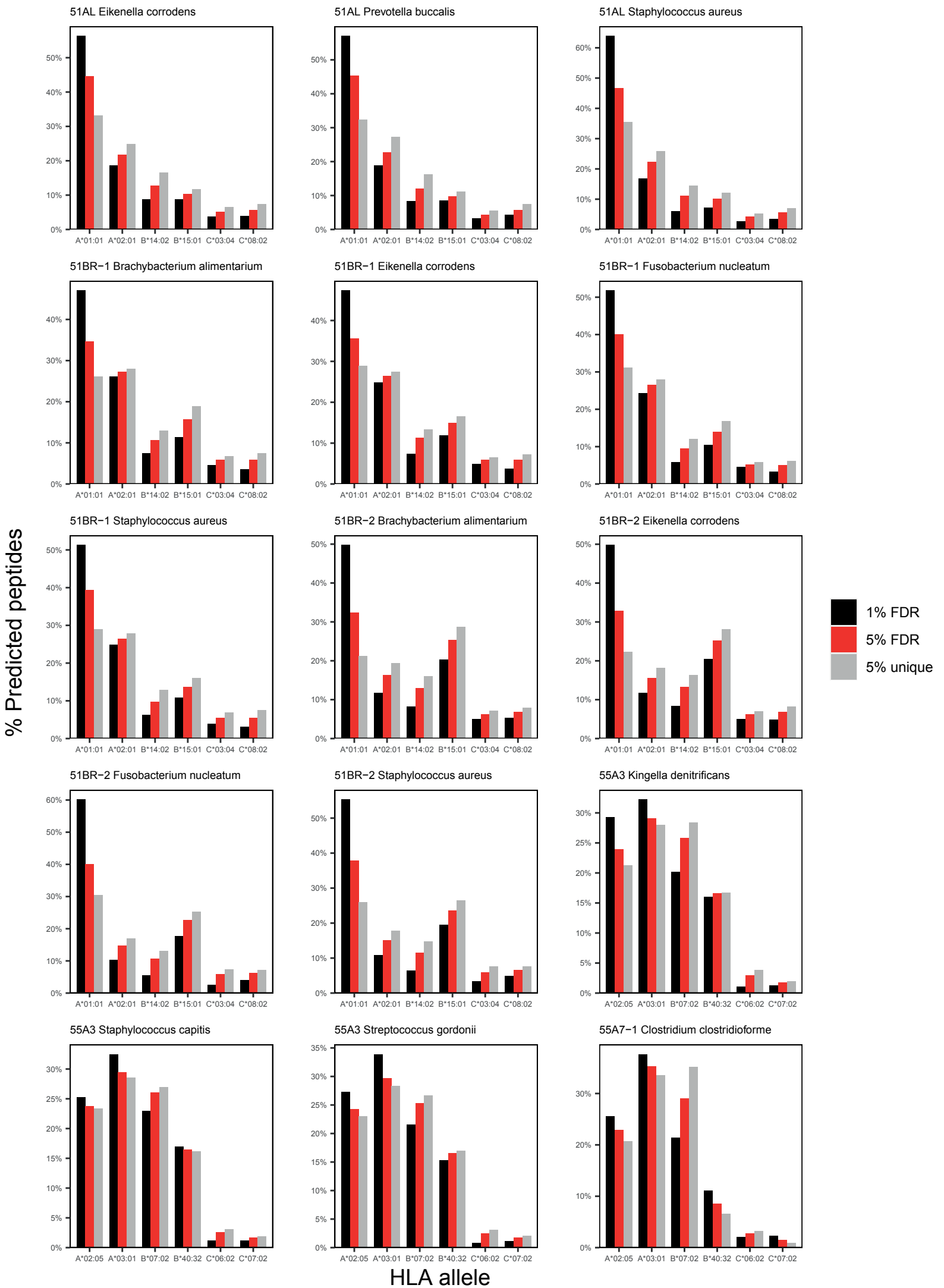
FDR	binders/total(%)
1%	923/1097 (84.14%)
5%	2277/2539 (89.68%)
5% unique	1362/1454 (93.67%)

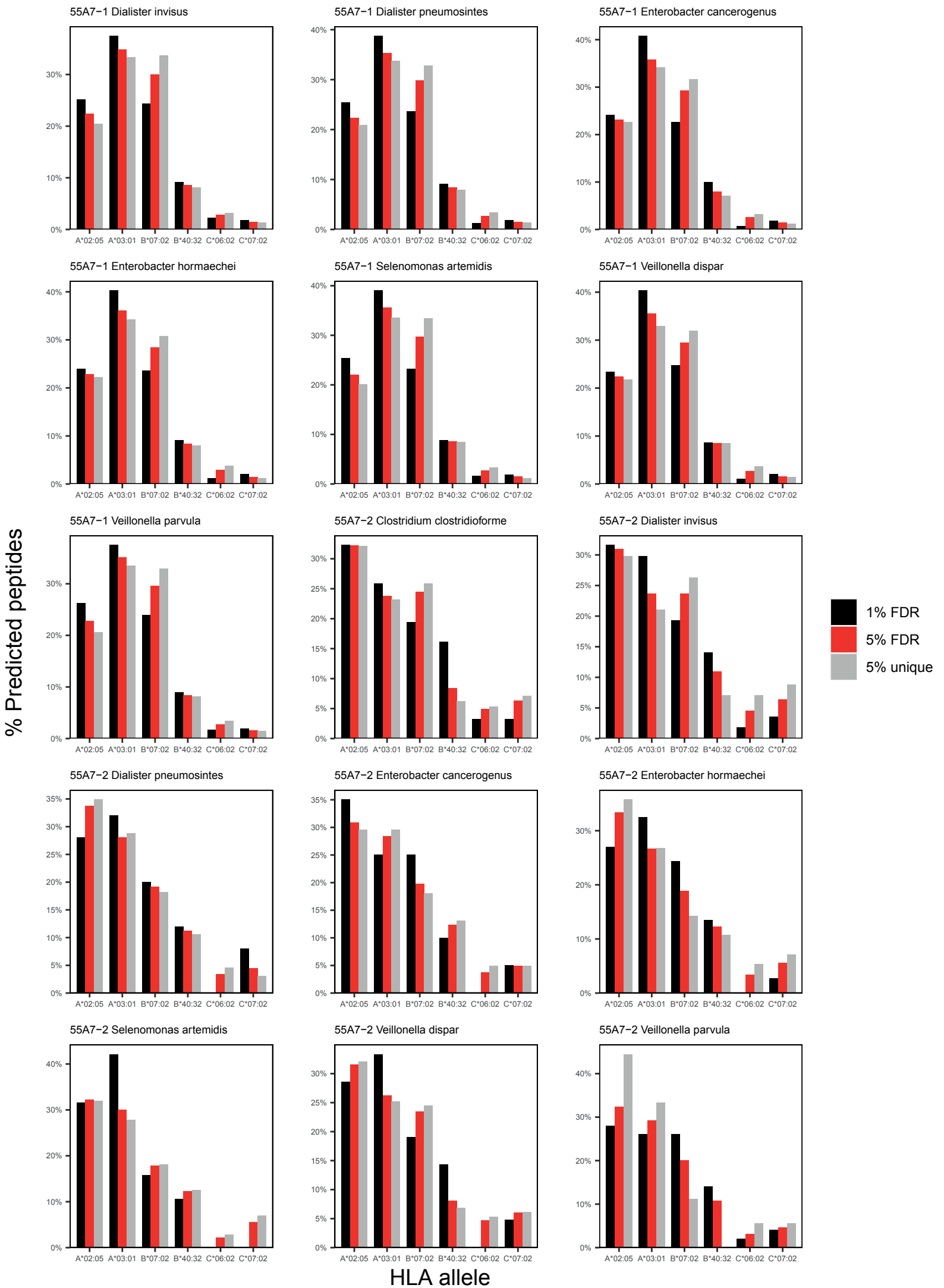
Percentage

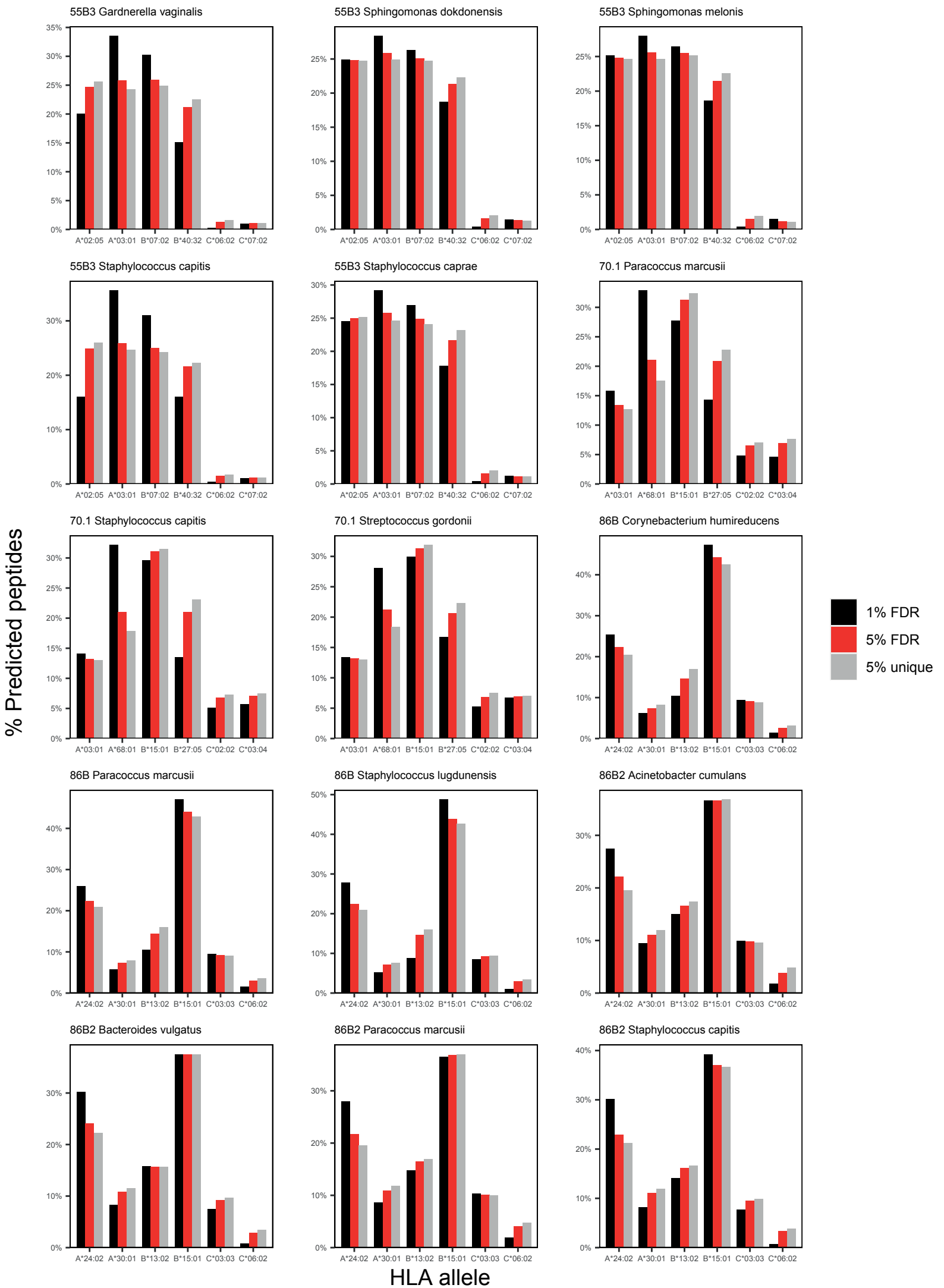
Length

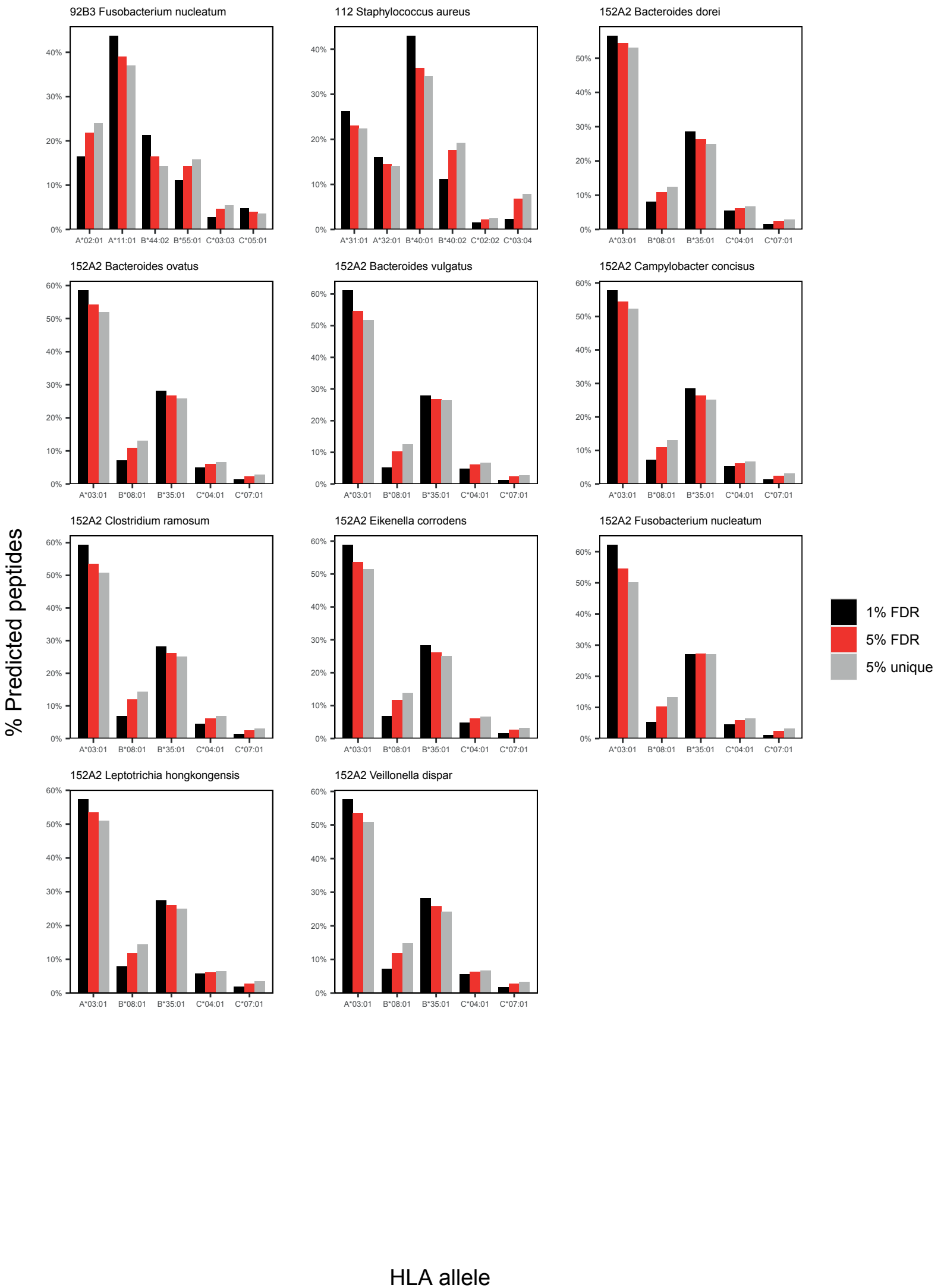
1% FDR
5% FDR
5% unique


HLA-I allele distribution of all identified peptides using different FDR
19T Enterococcus faecium
19T Photobacterium rosenbergii
19T Staphylococcus capitis
19T Staphylococcus caprae
42RF Campylobacter concisus
42RF Campylobacter lanienae
42RF Klebsiella oxytoca
42RF Paracoccus marcusii
42RF Shewanella decolorationis
42RS Enterobacter cancerogenus
42RS Sphingomonas melonis
27 Actinomyces odontolyticus
27 Fusobacterium nucleatum
27 Schaalia odontolytica
51AL Corynebacterium afermentans
% Predicted peptides
HLA allele
1% FDR
5% FDR
5% unique


51AL Eikenella corrodens
51AL Prevotella buccalis
51AL Staphylococcus aureus
51BR−1 Brachybacterium alimentarium
51BR−1 Eikenella corrodens
51BR−1 Fusobacterium nucleatum
51BR−1 Staphylococcus aureus
51BR−2 Brachybacterium alimentarium
51BR−2 Eikenella corrodens
51BR−2 Fusobacterium nucleatum
51BR−2 Staphylococcus aureus
55A3 Kingella denitrificans
55A3 Staphylococcus capitis
55A3 Streptococcus gordonii
55A7−1 Clostridium clostridioforme
A*01:01 A*02:01 B*14:02 B*15:01 C*03:04 C*08:02
A*02:05 A*03:01 B*07:02 B*40:32 C*06:02 C*07:02
% Predicted peptides
HLA allele
1% FDR
5% FDR
5% unique


55A7−1 Dialister invisus
55A7−1 Dialister pneumosintes
55A7−1 Enterobacter cancerogenus
55A7−1 Enterobacter hormaechei
55A7−1 Selenomonas artemidis
55A7−1 Veillonella dispar
55A7−1 Veillonella parvula
55A7−2 Clostridium clostridioforme
55A7−2 Dialister invisus
55A7−2 Dialister pneumosintes
55A7−2 Enterobacter cancerogenus
55A7−2 Enterobacter hormaechei
55A7−2 Selenomonas artemidis
55A7−2 Veillonella dispar
55A7−2 Veillonella parvula
A*02:05 A*03:01 B*07:02 B*40:32 C*06:02 C*07:02
% Predicted peptides
HLA allele
1% FDR
5% FDR
5% unique


55B3 Gardnerella vaginalis
55B3 Sphingomonas dokdonensis
55B3 Sphingomonas melonis
55B3 Staphylococcus capitis
55B3 Staphylococcus caprae
70.1 Paracoccus marcusii
70.1 Staphylococcus capitis
70.1 Streptococcus gordonii
86B Corynebacterium humireducens
86B Paracoccus marcusii
86B Staphylococcus lugdunensis
86B2 Acinetobacter cumulans
86B2 Bacteroides vulgatus
86B2 Paracoccus marcusii
86B2 Staphylococcus capitis
A*02:05 A*03:01 B*07:02 B*40:32 C*06:02 C*07:02
A*03:01 A*68:01 B*15:01 B*27:05 C*02:02 C*03:04
A*24:02 A*30:01 B*13:02 B*15:01 C*03:03 C*06:02
% Predicted peptides
HLA allele
1% FDR
5% FDR
5% unique


92B3 Fusobacterium nucleatum
112 Staphylococcus aureus
152A2 Bacteroides dorei
152A2 Bacteroides ovatus
152A2 Bacteroides vulgatus
152A2 Campylobacter concisus
152A2 Clostridium ramosum
152A2 Eikenella corrodens
152A2 Fusobacterium nucleatum
152A2 Leptotrichia hongkongensis
152A2 Veillonella dispar
A*02:01 A*11:01 B*44:02 B*55:01 C*03:03 C*05:01
A*31:01 A*32:01 B*40:01 B*40:02 C*02:02 C*03:04
A*03:01 B*08:01 B*35:01 C*04:01 C*07:01
% Predicted peptides
HLA allele
1% FDR
5% FDR
5% unique

Below is the list of bacterial peptides identified using 5% FDR, in which we marked the peptides that were identified also using 1% FDR.

HLA class	Patient	metastasis	Bacterial species	Peptide sequence	Identified in 1% FDR
HLA-I	42	42RF	Campylobacter concisus	ALIGGAIELL	+
				IGFDKVVEA	
				KLLKSENISK	
				NEKIDLNEV	+
				NGIGVIAEV	
				SEIASVGVL	
				SVDDITPTL	
				SVVNLSVSK	
			Campylobacter lanienae	GEYQITQTQI	
				VSNVSLSVK	
			Klebsiella oxytoca	ATVEDVITV	
				AVTVESVVAL	
				IVGLMIVSV	
				LAATIIHQV	+
				TKVSVGSIQK	+
				TMALSIIASI	
			Paracoccus marcusii	ELVNYVLEV	
			Shewanella decolorationis	ADDLADAIL	
				ALSRPAITAL	
				AVTVESVVAL	
				IEAMTSVAGI	
				KVERLGSNK	
				LVTAAGSTAL	
				SIGIDEISLL	
				SLGAITVSI	
		42RS	Enterobacter cancerogenus	ALSASDIAL	
			Sphingomonas melonis	AVGGTLLGV	
				GIVSFAVEGV	
				IAATIYALAL	
	51	27	Actinomyces odontolyticus	ALVDSTGIL	
				SAIVAIVEL	
				TMAEDLVTV	
				TVKEINTQL	
				VTIEDDGTVY	
			Fusobacterium nucleatum	ETTLVVTEY	
				GLDLGTLTY	
				IITIEIPEI	
				ITELNSPVL	
				LTDKKLEEY	
				SVVVDELFEV	
				VLTDTYLTL	
			Schaalia odontolytica	ALVDSTGIL	
				FLFAGRFNEAL	
				SAIDGVSAA	+
				TMAEDLVTV	
				TVKEINTQL	
		51AL	Corynebacterium afermentans	IARATLPEL	
				LTDISIEL	+
				LVDTSASL	+
				VQAQSAQVL	
			Eikenella corrodens	IARATLPEL	+
				ITLSLDNPL	
			Staphylococcus aureus	EGNIDFITL	
				LAEEKGLAL	
				LSDLGKSIY	
		51BR-1	Brachybacterium alimentarium	ALGGVILNV	
				GLGAGAVVL	
				LAFLGDISV	
				TLSIDIEAV	
			Eikenella corrodens	ALASGLLLY	
				ITTAEVVAV	
				VAISLVTAT	
			Fusobacterium nucleatum	SVVVDELFEV	
				VLTDTYLTL	
			Staphylococcus aureus	ALGVDALLLL	+
				ITDFIDPNQY	
				LLSGVFATV	
				LREVDIFNL	+
				TITKLTPTY	
				TVEAIATAV	
				VIATLVIIL	
				VLILVRPTI	
		51BR-2		ALDPISTLAI	
				AVDASVVAVL	
				FIAQRAEVL	
				GLLGVVAAV	

			Brachybacterium alimentarium	IAAANALSDIY	
				LALETGVLVL	
				LALTALTTA	
				MGLLPGALL	
				SADGDFLTVL	+
				TLSIDIEAV	
				VAAGVAGLV	
				VTDLSTQF	
			Eikenella corrodens	ALSSAVLVV	
				GSAVSAFVL	
				GVSAAAVAI	
				IPETVLTAL	
				ITDADLEKAY	
				LDAVDLNSY	
				QILLPDWQL	
			Fusobacterium nucleatum	ALTLAITTA	
				FSDSLPDSL	
				GATAGVANTY	
				GVDLGTLTY	
				IAVCSITAL	
				ITNTGAVTV	
				SLTDKISII	
				VLTDTYLTL	
			Staphylococcus aureus	AILTAITTA	
				EGNIDFITL	
				ITDGDIFDI	
				LSDLGKSIY	
				TITKLTPTY	
				TLTGFILGV	
				YIESNELYF	
55	55A3		Kingella denitrificans	TVDTLIDTL	+
			Staphylococcus capitis	AEVRTLTAV	
				GELATRSGT	
				LEGTVLDTL	
				LGADQIAEV	
				QLQTLTQQL	
				TAAALVMSI	
				TVIAELRAL	
			Streptococcus gordonii	KTDKVQALK	
		55A7-1	Clostridium clostridioforme	KVDKLELVLK	
				LKIGNIISY	+
				TVFDAELRGI	
			Dialister invisus	ITFTGGEPL	
			Dialister pneumosintes	ILYFKTQQR	
				LVRTPDLLL	
			Enterobacter cancerogenus	ASITKSIFY	
				IPSSGAATLI	
				LPSVGICLVI	
			Enterobacter hormaechei	GELERALAS	
			Veillonella dispar	VDRVLITNI	
			Veillonella parvula	KLIDKDISL	+
		55B3	Gardnerella vaginalis	AETNIPLTA	
				APAEAAEAA	
				AVSGVALAR	
				GLTNLIDGI	
				RDGVEKILSK	
				VERSEGVVA	
				VTASDLQAI	
			Sphingomonas dokdonensis	AAAARVAAI	
				AEAPAAEAA	
				ALSDMSLAL	
				GTAAAVAAL	
				GVAPSSATI	
				LASERALAL	
			Sphingomonas melonis	AAAARVAAI	
				AAAPAAAAAAA	
				AEGAAAIAA	
				ALSDMSLAL	
				GEAIAADAA	
				LIQVDIASI	
				TDLQTAEAI	
			Staphylococcus capitis	APPAPPVAV	
				GELGVALSA	
				LEGTVLDTL	
				LGASAPVAL	
				SPLDRDSIL	
				TPGSSALTL	
				TVACVLTAL	
				VVIDGLGAI	

		Staphylococcus caprae	LEGTVLDTL	
			SISIGVVELL	
70	70.1	Staphylococcus capitis	AIGVAASILY	
			ALGAAVITLY	
			FAGELSIVL	
			FRVDVVVGK	
			GLAASLIVY	
			LILDPWLETY	
			TPIVAVNAL	
			VALRRAIGL	
			VTSGVTAAY	
		Streptococcus gordonii	AAADKIVASF	
			ETLDKTIEF	
			GALLAIVTY	
86	86B	Corynebacterium humireducens	AAAPAGSAL	
			AATPAAAQY	
			AVMLPAAVL	
			AVSPARGIL	
			HAALAVIAL	
			LPAAVNGVL	
			SPGTAAAAL	
			VAGGLGALV	
			VGGTVVVSY	
			VQDGKQVAVL	
		Staphylococcus lugdunensis	ALAVITISY	
			IIGGPPSTTY	
			LKSTENTTY	
			LVRNDVIVY	
			NLQQIGLNI	
			VEYGKVFNI	
	86B2	Acinetobacter cumulans	AQSLIRHSA	
			LVTGIVVLL	+
			NVGLGVTAY	+
			SSISSVQEI	+
		Bacteroides vulgatus	AVEKNGLEI	
			ELDRRIAKY	
			KQRFTNALQK	
			LATAYANIAY	
			LQERKQIII	
			VSAGAVLLV	
			YASAGVLSI	
		Staphylococcus capitis	AIGVAASILY	+
			AMATVITAL	
			ATARGTGPFL	
			EYLLKDINI	+
			GQTLVTSRL	
			IRANIKDLI	
			LAAQTSLSL	
			LGAAVITLY	
			LGVDSLLAL	
			SLIKEPSLL	
			TPIVAVNAL	
			TTLQNGAVI	
			VTSGVTAAY	+
92	92B3	Fusobacterium nucleatum	GILSNIGIEI	+
			IISEKPITI	
			ITELNSPVL	
			ITNTGAVTV	+
			LENETNAII	+
			SLTDKISII	+
			VENLIIASI	
112	112	Staphylococcus aureus	DAVSRQAEL	
			EEVKTVGSL	
			ELKEILTL	
			GEIPTLDPI	
			KEIKNIEKEL	
			LENKEEIAL	
			LESNFITEI	
			NQKELEIIL	
			QLESVITQL	
			SVLVGTLIGF	
			TVKEVLIGF	
			VTNINTVTF	
152	152A2	Bacteroides dorei	ELLSNVNLY	
			GGSTITQQLAK	
			GISILPESIK	
			KLADPGTVLQK	
			LIYFYPPRK	
			RISATELLK	

				RTLMKVNEEK	
				SLLSIISLGK	
			Bacteroides ovatus	ALKELYTRK	
				ELLSNVNLY	
				LVDLKRMGM	
				LVKIREAIK	
				NPVELAKIY	
				TLDPTIQTY	
				VSWLNSLQLK	+
			Bacteroides vulgatus	DIEGLLQTY	
				GIDWSKLTTK	
				ILKVKELAL	
				IVLTLRLNK	
				KLIDLYGKR	
				LGDILTSSY	
				LSKDRKLSL	
				LVDLADISF	
			Campylobacter concisus	GLAPSLKTR	
				ILPDSIIKI	
				LKISIPIIQR	
				RLLQSIGWK	
			Eikenella corrodens	AIIGRLQAK	
				GETPVAPVY	
				LPESQGPLL	
				NLLKPALSK	
				VVWPQEVAAK	
			Fusobacterium nucleatum	DLIERKDSI	
				IERRKLREI	
				KLDNISSRIK	
				KQEPLSVSY	
				LLIEEIDRY	
				TLTATLLKY	
				VQLKGLLSY	
				YLADLIKIY	
			Veillonella dispar	LWDSIITRL	
				NISKVQVL	
				VPLSEVATI	
HLA-II	19	19T	Enterococcus faecium	TETIMQTSDQWEEL	+
			Staphylococcus capitis	AGNEVIEGDEL	
				YEYSVSAEL	
	51	27	Actinomyces odontolyticus	VTLGSRWRA	
			Fusobacterium nucleatum	LDEYIAERIQHIFN	
			Schaalia odontolytica	VTLGSRWRA	
		51BR-1	Eikenella corrodens	DKAAVEMFIATVFQ	+
			Fusobacterium nucleatum	DNNYSNSLEAINFLQ	
			Staphylococcus aureus	IEAGHLVEDDQTI	
				LYSAGQIVENAEIK	+
		51BR-2	Brachybacterium alimentarium	DSLDAAVVSIGA	
	55	55A3	Staphylococcus capitis	LGDWLARIPMAALVA	+
				SAGAAIAGL	
			Streptococcus gordonii	LGAASGAAVAI	+
		55A7-1	Enterobacter cancerogenus	IAALSATMAV	
				VDLINAGKA	
			Enterobacter hormaechei	AQVIILNHPGQISQ	+
				DKKSNIRNMSVIAHVD	
				DKKSNIRNMSVIAHVDHG	
				KKSNIRNMSVIAHVDH	
				KKSNIRNMSVIAHVDHG	+
				VIILNHPGQISQ	+
			Veillonella dispar	MDIIFSSLPVEKINE	+
			Veillonella parvula	MDIIFSSLPIDKINE	+
		55A7-2	Enterobacter hormaechei	AQVIILNHPGQISQ	+
				ARSRLREITTLQGGLE	+
		55B3	Gardnerella vaginalis	QGEAVISFDGVDAADA	+
				TGSAVSVLSMFAATVAAAGA	+
			Sphingomonas dokdonensis	DVVGANRAQL	
			Sphingomonas melonis	EAVASSEAI	
				LLFKDTILN	
	70	70.1	Staphylococcus capitis	LTHIRGRSL	
	86	86B	Corynebacterium humireducens	NVLRIVNEPTAAALAYG	+
			Staphylococcus lugdunensis	NPTEGEVVSPIEGK	
		86B2	Acinetobacter cumulans	VTMLLTYEATKDLP	+
			Staphylococcus capitis	DVAHQVEVL	
				TNIASLNTQRNLSSS	
				VARIEALMSDELGK	
	92	92B3	Fusobacterium nucleatum	LEAAISAGAKKA	
				QVLFYITPV	

To conclude, we analyzed the HLA-peptidomics data using 5% FDR, followed by conducting the filtration steps summarized in Extended Data Figure 4: First, we removed bacterial peptides exhibiting an ambiguous I/L position or harboring the same sequence as potential non-coding sequences or pseudogenes. Then, we used a linear regression model of the peptides' experimental retention time and predicted hydrophobicity, and removed the outliers (Supplementary Figure 6). We further removed peptides with low MS/MS coverage, as described above, keeping only those predicted to bind the patient's HLA alleles. This allowed us to achieve a peptidome that conforms to the expected length distribution (Supplementary Figure 2) and that was clustered to the patient's HLA allele motifs (Supplementary Figure 1). These filtration steps enabled us to retain true HLA ligands, while avoiding a stricter filtration using 1% FDR.

References

1 Kalaora, S. *et al.* Combined Analysis of Antigen Presentation and T-cell Recognition Reveals Restricted Immune Responses in Melanoma. *Cancer Discov* **8**, 1366-1375, doi:10.1158/2159-8290.CD-17-1418 (2018).

Supplementary Videos

Supplementary Video 1-51AL Staphylococcus caprae:

http://weizmann.ac.il/mcb/uploads/Samuels/Supplementary_Video_1-51AL_Staphylococcus_caprae.avi

Supplementary Video 2-55A3 Staphylococcus caprae:

http://weizmann.ac.il/mcb/uploads/Samuels/Supplementary_Video_2-55A3_Staphylococcus_caprae.avi

Supplementary Video 3-58A Staphylococcus capitis:

http://weizmann.ac.il/mcb/uploads/Samuels/Supplementary_Video_3-58A_Staphylococcus_capitis.avi

Supplementary Video 4-51AL Staphylococcus succinus:

http://weizmann.ac.il/mcb/uploads/Samuels/Supplementary_Video_4-51AL_Staphylococcus_succinus.avi

Supplementary Video 5-55A3 Staphylococcus succinus:

http://weizmann.ac.il/mcb/uploads/Samuels/Supplementary_Video_5-55A3_Staphylococcus_succinus.avi

Supplementary Video 6-51AL Actinomyces odontolyticus:

http://weizmann.ac.il/mcb/uploads/Samuels/Supplementary_Video_6-51AL_Actinomyces_odontolyticus.avi

Supplementary Video 7-55A3 Actinomyces odontolyticus:

http://weizmann.ac.il/mcb/uploads/Samuels/Supplementary_Video_7-55A3_Actinomyces_odontolyticus.avi

Supplementary Video 8-51AL Fusobacterium nucleatum:

http://weizmann.ac.il/mcb/uploads/Samuels/Supplementary_Video_8-51AL_Fusobacterium_nucleatum.avi

Supplementary Video 9-55A3 Fusobacterium nucleatum:

http://weizmann.ac.il/mcb/uploads/Samuels/Supplementary_Video_9-55A3_Fusobacterium_nucleatum.avi

Supplementary Video 10-51AL Staphylococcus caprae:

http://weizmann.ac.il/mcb/uploads/Samuels/Supplementary_Video_10-51AL_Staphylococcus_caprae.mp4

Supplementary Video 11-55A3 Staphylococcus caprae:

http://weizmann.ac.il/mcb/uploads/Samuels/Supplementary_Video_11-55A3_Staphylococcus_caprae.mp4

Supplementary Video 12-58A Staphylococcus capitis:

http://weizmann.ac.il/mcb/uploads/Samuels/Supplementary_Video_12-58A_Staphylococcus_capitis.mp4

Supplementary Video 13-51AL Staphylococcus succinus:

http://weizmann.ac.il/mcb/uploads/Samuels/Supplementary_Video_13-51AL_Staphylococcus_succinus.mp4

Supplementary Video 14-55A3 Staphylococcus succinus:

http://weizmann.ac.il/mcb/uploads/Samuels/Supplementary_Video_14-55A3_Staphylococcus_succinus.mp4

Supplementary Video 15-51AL Actinomyces odontolyticus:

http://weizmann.ac.il/mcb/uploads/Samuels/Supplementary_Video_15-51AL_Actinomyces_odontolyticus.mp4

Supplementary Video 16-55A3 Actinomyces odontolyticus:

http://weizmann.ac.il/mcb/uploads/Samuels/Supplementary_Video_16-55A3_Actinomyces_odontolyticus.mp4

Supplementary Video 17-51AL Fusobacterium nucleatum:

http://weizmann.ac.il/mcb/uploads/Samuels/Supplementary_Video_17-51AL_Fusobacterium_nucleatum.mp4

Supplementary Video 18-55A3 Fusobacterium nucleatum:

http://weizmann.ac.il/mcb/uploads/Samuels/Supplementary_Video_18-55A3_Fusobacterium_nucleatum.mp4