NIST Human Plasma Glycopeptide Libraries (June 2025)

This document describes the three varieties of NIST human plasma glycopeptide libraries as presented at the 2015 ASMS meeting in Baltimore by Yi Liu of NIST.

1. A tryptic glycopeptide MSMS library (nist-plasma-proteins-msms). This is in the format of earlier NIST peptide libraries available on the NIST library website. The spectra selected are the highest scoring and highest purity spectra for each reliably identified glycopeptide ion. Each is fully annotated by MS_Piano (https://pubs.acs.org/doi/full/10.1021/acs.jproteome.1c00324).
2. A library containing Glycopeptide Abundance Distribution Spectra, GADS, (nist-plasma-proteins-gads) is presented in the format of previous GADS libraries. These libraries are available in the same section of the NIST downloadable libraries site as this one. Methods for validating GADS are described in publications associated with these libraries. The number of spectra used for creating for each GADS is reported in the comments as nSpec=<>, and number of different peptide sequences as nPep=<>.
3. A glycan fragmentation library (nist-plasma-proteins-glycan), derived from selected spectra in MSMS library above. It contains a novel variety of 'spectra' that represent all glycan containing fragments, including oxonium and glycopeptide fragment ions. The latter are represented as their mass loss (difference in glycopeptide fragment ion and their precursor). Searched and lookup is done by specific Glycans, using x as the prefix in the Name search window. For example, all G4H5FS containing glycopeptides can be found by entering xG4H5FS in the Name search tab in the NISTMS program (note G=GlcNac or GalNac, H=Hexose, F=Fucose, S=Sialyl). Individual glycopeptides and their position in the protein follow in alphabetical order. Peptide sequence can also be employed.

Libraries 1 and 2 are described in the document 'nist-gads-glycopeptide-user-guide.pdf' in the same folder as this file. Also, each of these libraries is available in NIST '.msp' format – a simple text (ASCII) format.

A unique feature of this GADS library is that each GADS is derived from multiple individual GADS, sometimes from different proteases and peptide sequences, all using combined charge states. They are most easily viewed in the Name tab, by either protein symbol first (UNIPROT style, A1AG for example) followed by sequon position, and cluster number. A prefix of zg substitutes gene name for UNIPROT abbreviated protein name.

Library 3 is created from both high-resolution ion trap and stepped energy spectra in MSMS Library 1. The x-axis corresponds to the glycan mass of the oxonium fragment ion or the mass of the LOSS of glycan from the precursor glycopeptide (i.e., the difference in mass of the precursor glycopeptide ion and product glycopeptide ion). The charge(s) of the latter are given in the annotation of each peak are these two varieties of fragment ions are aligned.

To access this library, run the accompanying 'nistms-glyco.exe' program and select the default folder (it contains the 3 libraries above). Documentation concerning other features of the software and libraries are available in other PDF files. Library contents are most directly examined in the 'Names' tab after selecting one of the 3 libraries.