Ryan Tal Fellers, Joseph Brent Greer, Bryan Patrick Early, AJ van Nispen

<u>Guide for Uploading Proteoforms to the Proteoform Repository</u>

Part 1: Defining Terms

Part 2: Labeling Convention Examples

Part 1: Defining Terms

Entry Terms

Uniprot ID: The Uniprot ID is a useful means of identifying a sequence, but it is not a stable identifier as is the accession number. The Uniprot ID often contains biologically relevant information.

Entry Accession: The stable identifier used to cite UniProtKB entries. Accession numbers consist of 6 or 10 alphanumerical characters.

Entry Description: The Description lines contain general descriptive information about the sequence stored. This information is generally sufficient to identify the protein precisely. Use the FULL Recommended name in the description.

Isoform Terms

Isoform (as related to proteomics): denotes related forms of protein molecules arising from the same gene by alternative splicing or variable promoter usage.

Isoform Accession: The isoform accession is a unique identifier that consists of the primary accession number of the entry, followed by a dash and a number. It describes the alternate sequence with respect to the canonical sequence displayed by default in the entry.

Isoform Description: Description of the alternative splicing, alternative promoter usage, and alternative translation initiation that generate the isoform.

Isoform Sequence: the sequence of naturally occurring alternative protein isoform(s). The changes in the amino acid sequence may be due to alternative splicing, alternative promoter usage, alternative initiation, or ribosomal frameshifting.

Proteoform Terms

Start Position: **The observed sequence's start position with respect to the full protein sequence's amin**o acid number. The first amino acid position for the full protein is ONE.

End Position: **The observed sequence's end position with respect to the full protein sequence's amino acid number.** The last amino acid position is equivalent to the length of the observed sequence.

Observed Sequence: The portion of the full protein sequence observed, listed by amino acids.

Sequence Modifications: Amino Acid Modifications. Any modification that changes the amino acid sequence of the observed sequence.

PTM: Characterization of Post Translational Modifications. A post-translational modification is a covalent processing event resulting from a proteolytic cleavage or from the addition of a modifying group to one amino acid. PTMs modulate the function of most eukaryote proteins by altering their activity state, localization, turnover, and interactions with other proteins. Post translational modifications complement the information provided at the sequence level or describes modifications for which position-specific data is not yet available.

N-Terminal Modification: Protein N-termini can be modified co - or posttranslationally. Modifications include the removal of initiator methionine (iMet) by aminopeptidases, attachment of small chemical groups such as acetyl, propionyl and methyl, and the addition of membrane anchors, such as palmitoyl and myristoyl groups

C-Terminal Modification: The description of the modification at the C-terminus of the protein that occurred postranslationally.

Part 2: Labeling Conventions/Examples

Entry Terms

Uniprot ID: S100A7_HUMAN

Entry Accession: P31151

Entry Description: Protein S100-A7

Isoform Terms

Isoform Accession: P31151

Isoform Description: S10A7_HUMAN Protein S100-A7 OS=Homo sapiens GN=S100A7 PE=1 SV=4

Isoform Sequence:
MSNTQAERSIIGMIDMFHKYTRRDDKIEKPSLLTMMKENFPNFLSACDKKGTNYLADVF
EKKDKNEDKKIDFSEFLSLLGDIATDYHKQSHGAAPCSGGSQ

Proteoform Terms

Start Position: 2

End Position: 101

Observed Sequence:
SNTQAERSIIGMIDMFHKYTRRDDKIDKPSLLTMMKENFPNFLSACDKKGTNYLADVFE
KKDKNEDKKIDFSEFLSLLGDIATDYHKQSHGAAPCSGGSQ

Sequence Modifications: E28D (E amino acid at position 28 of the isoform sequence modified to D amino acid.)

PTM: Write the chemical formula of the modification: Acetylation = CH_3CO

N-Terminal Modification: Write the chemical formula of the modification: Acetylation = CH_3CO

C-Terminal Modification: Write the chemical formula of the modification: If none, leave blank.
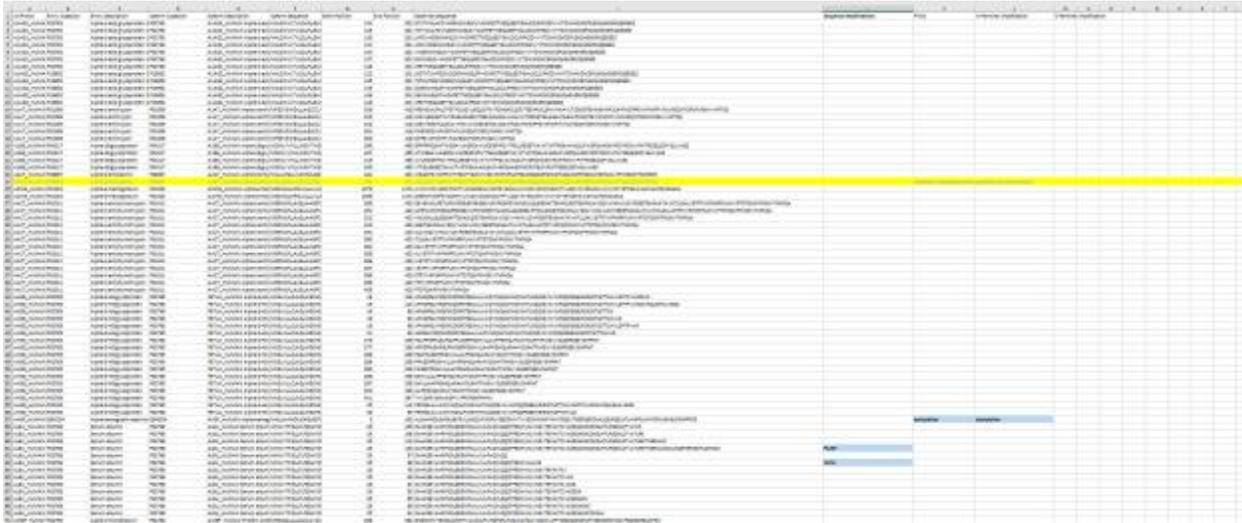
Figure 1: Example of Excel spreadsheet where data will be inputted.