

# Metadata Specifications: Proteins

**Importance** 1: Required, 2: Required if available, 3: Optional  
**Common Fields** Fields that are common across all LINC metadata standards  
**Custom Fields** Fields that are unique to a single LINC metadata standard or common across only a subset of them

Common Fields

LINC Field Name	Related to	Description	Comments	Importance
PR_LINC_ID	Canonical	Unique LINC identifier for the protein	-	1
PR_Name	Canonical	The primary name of the protein	Proposed standard: Use the UniProt short protein name.	1
PR_Alternative_Name	Canonical	List of synonymous protein names	Proposed standard: Include only synonyms listed in the UniProt database.	2
PR_Alternative_ID	Canonical	List of other alternative protein IDs	-	2
PR_Center_Canonical_ID	Canonical	Center-specific protein ID	LINC DSGC-specific canonical ID. This will be assigned by a given LINC DSGC according to its protein registration scheme.	1
PR_Relevant_Citations	Batch	Appropriate literature reference(s) for reagent derivation, production, and/or validation (not information about the endogenous function of a protein)	-	2
PR_Center_Name	Batch	LINC center using the protein	-	1
PR_Center_Batch_ID	Batch	LINC DSGC-specific batch ID. This will be assigned by a given LINC DSGC according to its protein registration scheme.	-	1
PR_Provider_Name	Batch	Vendor or lab that supplied a protein reagent	-	1
PR_Provider_Catalog_ID	Batch	Batch ID or catalog number assigned to the protein by the vendor or provider	-	1
PR_Provider_Batch_ID	Batch	Batch or lot number assigned to the protein by the vendor or provider	-	1

Custom Fields

PR_PLN	Canonical	- Protein line notation (PLN) provides a unique identifier that obviates the need for the individual fields: PR_UniProt_ID (which includes isoform information when relevant), PR_Mutations, and PR_Modifications - If PLN is not in use by a DSGC, those individual fields will need to be included and populated instead.	The PLN standard and associated tools for users are under development by HMS and DCIC.	2
PR_UniProt_ID	Canonical	The UniProt ID of the specific protein and, if relevant, isoform	-	1
PR_Mutations	Canonical	List of known amino acid substitutions	Proposed standard: No controlled vocabulary is proposed. Adoption of PLN is encouraged instead.	2
PR_Modifications	Canonical	List of known posttranslational or chemical modifications	Proposed standard: No controlled vocabulary is proposed. Adoption of PLN is encouraged instead.	2
PR_Protein_Complex_Known_Component_LINC_IDs	Canonical	The LINC IDs of each known protein subunit of the complex	For registration of complexes only	2
PR_Protein_Complex_Known_Component_UniProt_IDs	Canonical	The UniProt ID of each known protein subunit of the complex	For registration of complexes only	2
PR_Protein_Complex_Known_Component_Center_Protein_IDs	Canonical	The center-specific protein IDs of each known protein subunit of the complex	For registration of complexes only	2
PR_Protein_Complex_Details	Canonical	A free text description of the protein complex	For registration of complexes only	2
PR_Protein_Complex_Stoichiometry	Canonical	The stoichiometry of subunits of the complex, if known	For registration of complexes only	3
PR_Amino_Acid_Sequence	Batch	The amino acid sequence of the reagent as supplied by the vendor or provider	Proposed standard: This field should only be populated when sequence information is supplied by the vendor or provider. This field should not be populated using reference sequence from UniProt or a similar database.	2
PR_Production_Source_Organism	Batch	The organism from which the reagent was isolated	Proposed standard: Use NCBI Taxonomy as the controlled vocabulary (e.g. "Escherichia coli").	2
PR_Production_Method	Batch	A controlled vocabulary describing the method of protein synthesis (e.g. chemically synthesized, recombinantly expressed in E. coli, etc.)	Proposed standard: When possible, use the BAO controlled vocabulary for "Protein preparation method" as the controlled vocabulary ( <a href="http://bioportal.bioontology.org/ontologies/BAO/?p=classes&amp;conceptid=http%3A%2F%2Fwww.bioassayontology.org%2Fbao%23BAO_0000356">http://bioportal.bioontology.org/ontologies/BAO/?p=classes&amp;conceptid=http%3A%2F%2Fwww.bioassayontology.org%2Fbao%23BAO_0000356</a> ).	2
PR_Protein_Purity	Batch	A description of a protein's level of purity (e.g., if it was partially purified, purified, unpurified, etc.) as stated by the vendor or provider	-	2