Protein Specification vr. 0.0.3

Bioschemas specification describing the usage of BiologicalEntity for the “protein” biological type

**Description.** This protein specification presents the usage of the generic type BiologicalEntity by the biological type “protein”. Please be aware “protein” is NOT a schema.org type but a BiologicalEntity profile.

**Supporting information**

* [Use cases](https://docs.google.com/document/d/1MNU-r7CMAvKdGav1VT3XeyQXr2XwRA6JXDsoWoA0R1E)
* [Bioschemas mapping for proteins](https://docs.google.com/spreadsheets/d/1QQH4AkzdwPT1Qt5OLmH5HosLpkFU7khwE4Ql9_Cb9ZQ)
* [UniProt protein example](https://docs.google.com/document/d/1SWHXYr4Hf2Sy7Oc5Z_oveGwOUlNhJmmLigWN2cPTqOk)
* InterPro protein example
* [PDBe protein example](https://drive.google.com/open?id=0B7CgVeE2Tg5LdnQ3RF9Lcnh3R2s)

# Properties

Based on schema.org [CreativeWork]

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Property** | **Expected Type** | **Description** | **CN** | **MG** | **CV** |
| citation | CreativeWork or URL | A citation or reference to a creative work, such as a publication, web page, scholarly article, etc. | many | R | no |
| dateCreated | Date or DateTime | The date on which the CreativeWork/BiologicalEntity was created or the item was added to a DataFeed/Dataset/DataRepository. | one | O | no |
| dateModified | Date or DateTime | The date on which the CreativeWork/BiologicalEntity was created or the item was added to a DataFeed/Dataset/DataRepository. | one | O | no |
| hasPart | BiologicalEntity | Indicates a CreativeWork/BiologicalEntity that is (in some sense) a part of this BiologicalEntity e.g. Domains  Inverse property: isPartOf. | many | O | no |
| isPartOf | BiologicalEntity | Indicates a CreativeWork/BiologicalEntity that this BiologicalEntity is (in some sense) part of.  Inverse property: hasPart. | many | O | no |

Based on schema.org [Dataset]

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Property** | **Expected Type** | **Description** | **CN** | **MG** | **CV** |
| distribution | DataDownload | A downloadable form of this dataset, at a specific location, in a specific format. From structures could be a mmCIF file or a PDB File etc | many | O | no |

Based on schema.org [Thing]

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Property** | **Expected Type** | **Description** | **CN** | **MG** | **CV** |
| [additionalProperty](https://schema.org/additionalProperty) | [PropertyValue](https://schema.org/PropertyValue) | A property-value pair representing an additional characteristics of the entity, e.g. a product feature or another characteristic for which there is no matching property in schema.org.  For protein structures, the below properties would be highly recommended:  resolution  boundMolecule  biochemicalFunction  biologicalProcess  **The below could be optional:**  expressionSystem  assemblyComposition  assemblyComposition | many | O | no |
| [alternateName](http://schema.org/alternateName) | [Text](http://schema.org/Text) | An alias for the item. | many | R | no |
| [description](http://schema.org/description) | [Text](http://schema.org/Text) | A description of the item. | many | R | no |
| [identifier](http://schema.org/identifier) | PropertyValue or  Text or  URL | The identifier property represents any kind of identifier for any kind of Thing, such as ISBNs, GTIN codes, UUIDs etc. Schema.org provides dedicated properties for representing many of these, either as textual strings or as URL (URI) links. See background notes for more details. Recommendation: identifiers.org whenever possible | one | M | no |
| [image](http://schema.org/image) | ImageObject or  URL | An image of the item. This can be a URL or a fully described ImageObject. | many | R | no |
| [name](http://schema.org/name) | [Text](http://schema.org/Text) | The name of the item. | one | R | no |
| [sameAs](http://schema.org/sameAs) | [URL](http://schema.org/URL) | URL of a reference Web page that unambiguously indicates the item's identity. E.g. the URL of the item's Wikipedia page, Wikidata entry, or official website. | many | R | no |
| [url](http://schema.org/url) | [URL](http://schema.org/URL) | URL of the item. Usually the URL of the page being marked up. | one | R | no |

New properties for [Thing]

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Property** | **Expected Type** | **Description** | **CN** | **MG** | **CV** |
| isMentionedIn | Thing | CreativeWork, Dataset, collection mentioning this entity  Inverse of:mentions | one | R | no |
| isBasedOn | Thing | Biological Entity of type “protein” that the structure is based on. | many | R | no |

New properties for [BiologicalEntity]

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Property** | **Expected Type** | **Description** | **CN** | **MG** | **CV** |
| biologicalType | [Text](http://schema.org/Text) | **“protein structure”** | many | M | no |
| crossReference | [Thing](http://schema.org/Thing) | A pointer to another, somehow related entity.  Usage: Whenever isBasedOn/isBasisFor, isPartOf/hasPart, citation or any other more specific does not work. | many | O | no |
| taxon | URL | A url pointing to NCBI Taxonomy or a taxonomic resource | one | O | yes |

**Legend:** *CN: Cardinality (one, many)*

*MG: Marginality (M: minimum; R: recommended; O: optional)*

*CV: Suggested controlled vocabularies (yes, no)*

# Examples

Schema.org [suggests](http://schema.org/docs/gs.html) implementing metadata using JSON-LD, RDFa or Microdata. JSON-LD is the recommended format by Google, but any of these formats can be used for embedding information about tools in a web page or other online resource.

|  |
| --- |
| Example 1 - [JSON-LD]. [Example description: UniProt protein] |
| Full example available at [**P00519**](https://docs.google.com/document/d/1SWHXYr4Hf2Sy7Oc5Z_oveGwOUlNhJmmLigWN2cPTqOk) |

|  |
| --- |
| Example 2 - [JSON-LD]. [Example description: InterPro protein] |
| Full example available at <https://github.com/BioSchemas/BiologicalEntity/Examples/InterPro/P00519.json> |

|  |
| --- |
| Example 3- [JSON-LD]. [Example description: PDBe protein] |
| Full example available at <TODO> |

# 

# Controlled Vocabularies

This section contains a list of fields that require a controlled vocabulary or enumeration and suggests what is acceptable for each.

## [biologicalType]

From the [BiologicalEntity enumeration](https://docs.google.com/document/d/1XASuESIHU3bi1aXMxQS5-rCOQX0ugjMNkh68VF4co4Q/edit#bookmark=id.jtedzau00aty) only “protein” is used here

## [Taxon]

Ontology: Any well-known taxonomic resource. NCBI Taxon or UniProt Taxon are the recommended ones.