NA5.1
Standards, Ontology and Database Schema for Ligand Binders Information

Deliverable 5.1.2
Ligand-binders ontology

INTRODUCTION

Development of the ProteomeBinders (PB) information resource implies standardization of the information to be exchanged. Standardization of the data will assure comprehensiveness and quality of the information as well as facilitate its sharing via the information system. As part of the standardization process, we have undertaken the development of ligand-binders ontology to represent the domain knowledge of binder/protein pairs and associated experimental information.

In this document is presented the scope, set of concepts (and the relationships between them) and structure of the ligand-binders ontology (version 1).

The ligand-binders ontology was developed on the basis of the definition of the minimal information for the unambiguous description of binder/protein pairs (Deliverable 5.1.1 - Specifications of minimal information to unambiguously describe binder/protein pairs) as well as on the consultation of experts during an ontology working session (NA5.1 Workshop, “Standards and Ontology”, Saint Emilion, France July 3-4, 2006) and via a survey (February, 2008).

1. What is an ontology?

Derived from a philosophical concept, an ontology in computer sciences is a representation of a set of concepts within a domain and the relationships between these concepts. This representation is usually encoded in a specific ontology language which can be understood by both human beings and computer programs. It generally includes classes, representing concepts, which may contain other more restrictive classes. Properties are used to unambiguously describe objects of a given class. An important use of properties is to define the relationships (also known as relations) between objects in the ontology.

As such an ontology serves several purposes. First of all, it allows describing the domain in a standardized way. Moreover it can be used to reason about the characteristics of the domain.

2. Why an ontology for ProteomeBinders?

One of the aims of the ProteomeBinders information resource is the standardization of the information regarding the characterization of binder/target couples. One of the major elements to be standardized in this context is the vocabulary: we have to ensure that a given term is a consensus tag for a specific concept. The descriptive dimension of an ontology provides such a controlled vocabulary.

Another aim of the ProteomeBinders information resource is to provide assistance to binder producers, quality control centres and binder users in their everyday tasks. For binder producers and quality control centres it may be prioritization of the production and the quality assessment of binders according to the importance in biological pathways of the molecules to detect. For binder users, it might be the choice of a binder to use according to a given experimental goal and specific experimental settings. As a basis for reasoning an ontology is an invaluable resource to develop computer programs performing these assistance tasks.

THE LIGAND-BINDERS ONTOLOGY

1. Ontology scope and goal

The scope of the ligand-binders ontology covers the knowledge domain of protein binder technologies applied to human proteome analysis.

The ontology should provide information on protein binder performance, quality and applicability. The granularity of the ontology must allow research and technical specifications of the binding tools. Furthermore, ligand-binders ontology should enable the integration of new ligand-binder information within the state-of-the-art knowledge of the field. In this respect, the intended ontology will be a knowledge tool in which concepts are connected through relationships enabling the development of computer reasoning application. It differs from an ontology like Gene Ontology (GO) [Gene Ontology: tool for the unification of biology. Nature Genet. (2000) 25: 25-29.] where controlled terms are provided for the unique goal of vocabulary standardization. The ligand-binder ontology will propose more than a hierarchy of consensus terms, since its further network of relationships between concepts (i.e. classes) will allow knowledge inference (or computer reasoning). Such a knowledge tool can be used by software applications in order to assist scientists for the exploitation of the available electronic information.

2. Overview of the main concepts and their relationships

The ligand-binders ontology was developed based on the vocabulary used by the scientists in the field and their current practice in protein identification, quantification and capture technologies. Five leading concepts were identified (see table 1): (i) the molecular species involved in a binding event (i.e. binders, targets and cross-reacting molecules), represented by the *"Binder"* and *"Bound Entities"* classes and subclasses., (ii) their associated biophysical properties, represented by *"Molecular Characteristics"*, *"Structural Characteristics"* and *"Binding Characteristics"* classes and subclasses, (iii) the experimental context, represented by the "*Experimental Workflow*" and *"Experiment"* classes and subclasses (iv) the associated evidences which are represented by the "*Experimental data*" class and subclasses and finally (v) the data producers represented by the "*Experimentalist*" class and subclasses.

Table 1 – Main concepts represented in the ligand-binders ontology

Concepts	Classes (and subclasses)
Binders, targets and cross-reacting molecules	• *"Binder"* • *"Bound Entities"*
Binders and targets properties	• *"Molecular Characteristics"*, • *"Structural Characteristics"* • *"Binding Characteristics"*
The experiments demonstrating binders and targets properties, as well as their association into a binder-target couple	• "*Experimental Workflow*" • *"Experiment"*
The evidences	• "*Experimental data*"
The data producers	• "*Experimentalist*"

These different concepts are linked by several fundamental relationships. The molecular species (i.e. binder and target) are generated through experiments which confer them specific properties. These specific properties, including binder ability to recognize its designated target, are characterized also through experiments. Experiments are associated in sets called experimental workflow which are goal-oriented combinations of several experiment types (e.g. binder production). Each experiment produces data whose nature is linked to the experimental method used and the assessment made. Experimental workflows are performed by experimentalists invested with specific tasks: binder (or target) production, binder (or target) quality control or binder application.

Besides these main classes (or concepts), the ontology include several classes such as "*Binder (or Target) Type*", "*Molecular Type*", "*Chemical Structure*", "*modification*" or "*Entity of Reference*"

which are needed for the description of the binders and targets within a general biochemical and biological context. On the same note, the class “Reference” which is linked to all the main classes is an essential bond between the ligand-binders information and other knowledge domains.

3. Ontology structure

The Ligand-Binders Ontology structure has been designed according to the logic and principles of the protein binding field. Currently, 25 concepts have been identified and represented under classes and subclasses which contain general or specific properties. These properties can be represented under different format:

- Free text
- Numerical value(s),
- Boolean expression (i.e. “yes” or “no”),
- List(s) of allowed terms
- Class(es)
- Instance(s) of class(es)

A general principle in ontology structures is that a subclass inherits the properties of the parent class it belongs to. However, the subclass might include new specific properties.

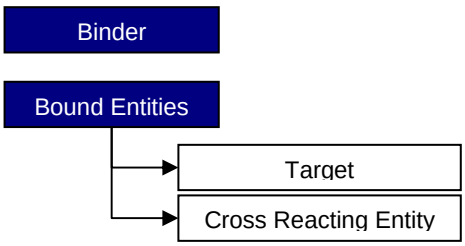
3.1. Representation of binder, target and other molecular species

3.1.1. “*Binder”* and *“Bound Entities*”

- Concept

The main actors of a binding event are the binder and its binding partner. Among the bound molecules, one would distinguish the binder designated target from other cross-reacting molecules. The target is the molecular species for which the binder has been designed (or selected) and demonstrates specific protein binding activities. On the other hand, cross-reacting entities are molecular species which are not structurally identical to the designated target. However, they might share identical or highly similar binding site(s) (or epitope(s)) with the designated target. Therefore, using structure homology analysis, identified cross reacting species may bring insights about binding site (or epitope) structure of the binder. The different binding partners are represented by the “*Binder*” and “*Bound Entities*” classes and subclasses (see figure 1).

Figure 1 - *“Binder”* and *“Bound Entities”* class and subclasses



- Class properties

The class and subclasses properties (see Appendix 1 - Table 1) provide information on molecules identity (i.e. name), origin (instance of “*Source*” class), biochemical (“*Binder Type*” - or “*Target Type*”, “*Molecular Type*” classes and instance of “*Molecular State*” class) and biological (instance of “*Entity of Reference*” class) category, structure (“*Chemical Structure*” class and instances of “*Chemical Structure Characteristics*” and “*Modification*” classes) and

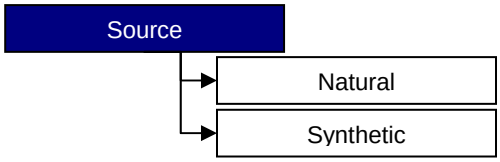
biophysics characteristics (instances of *"Molecular Characteristics"* and "*Binding Characteristics*" classes). The class properties also link the binder and bound entities to the experimental (instance of "*Experimental workflow*") and literature (or molecular database) (instance of "*Reference*") context which they are associated with, including authors of experiments.

3.1.2. "*Source*"

- Concept

Origin of the binding molecules is represented by the "Source" class (see figure 2). This class is further split into "*Natural*" source and "*Synthetic*" (i.e. non natural) source. In this context, "*Synthetic*" source applies to any molecule whose molecular structure has been artificially designed in whole (e.g. random peptide) or partially (e.g. fusion or engineered protein) and therefore has not been characterized as naturally occurring in any organism. This does not take into account the method of synthesis: a human protein produced in E. Coli is of natural origin. However, the related fusion protein corresponding to the human protein sequence fused to a protein tag and also produced in E. Coli, is of synthetic origin since its sequence does not correspond to any natural protein. (See figure 2).

Figure 2 - "*Source*" class and subclasses



- Class properties

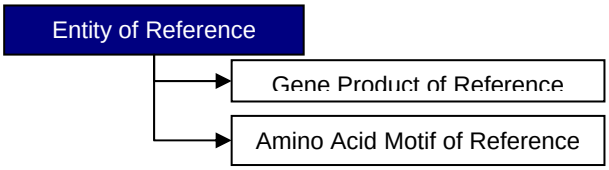
The class properties provide information on the source identity (i.e. name), the molecular species it is source of (instance of "*Binder*" or *"Bound entities"*) and literature (or molecular database) context (instance of "*Reference*"). (See Appendix 1 - Table 2)

3.1.3. "*Entity of Reference*"

- Concept

To represent the characterized molecules which binders or bound entities can be structurally related to (i.e. identical to or derived from), the class "*Entity of Reference*" has been created. For instance, it allows to express that a target is related to a specific gene expression product and then to integrate the target information within a biological context. (See figure 3).

Figure 3 - "*Entity of Reference*" class and subclasses



- Class properties

The class properties provide information on the identity (i.e. name) of the reference entity and the literature context (instance(s) of "Reference"). Additional properties about gene name and symbol are included in the "Gene Product Reference Entity" subclass. (See Appendix 1 - Table 3)

3.1.4. "*Binder Type*", "*Target Type*" and "*Molecular Type*"

- Concept

To classify the different categories of binder and target molecules, the classes *"Binder Type*" and "*Target Type*" have been created (see respectively figure 4 and 5). This classification integrates the molecular species description within a broader biochemical context. It also allows to navigate through group of molecules which possess similar biophysics properties, and therefore, similar technical or research applications.

Figure 4 - "*Binder Type*" class and subclasses

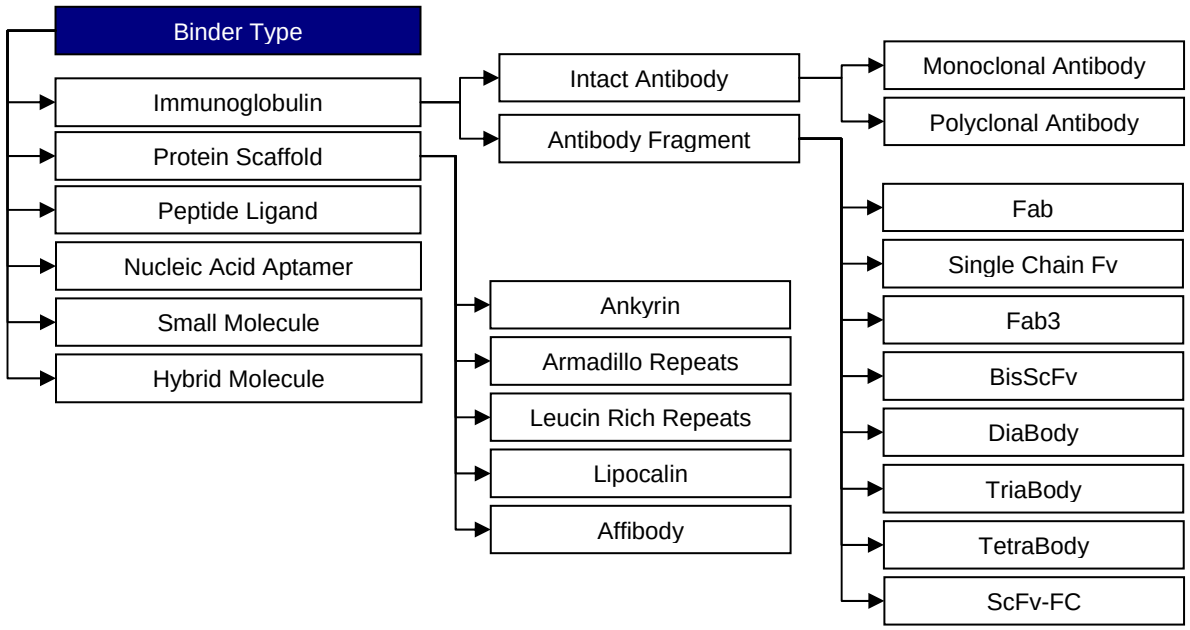
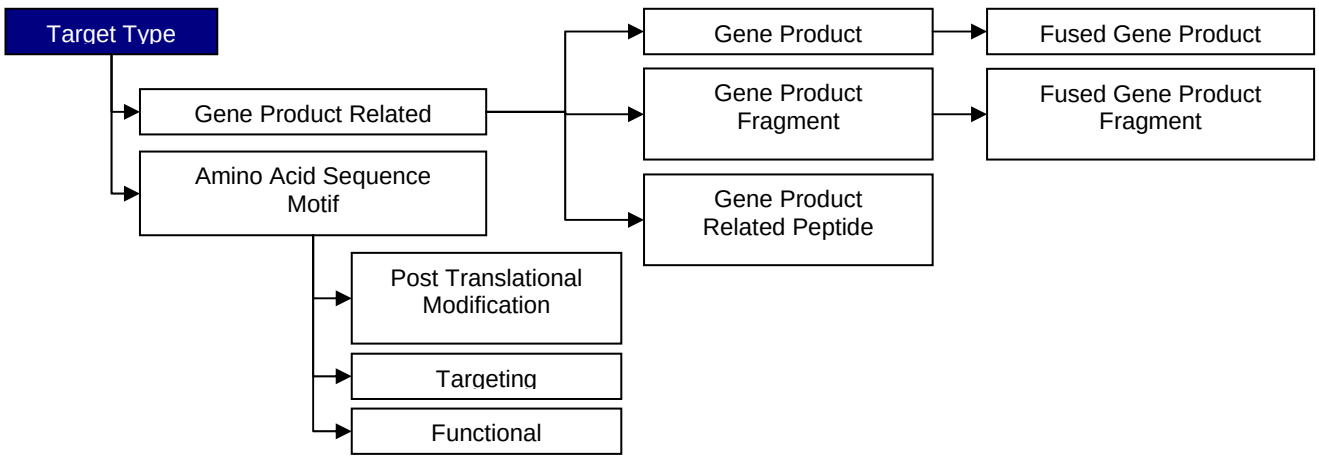
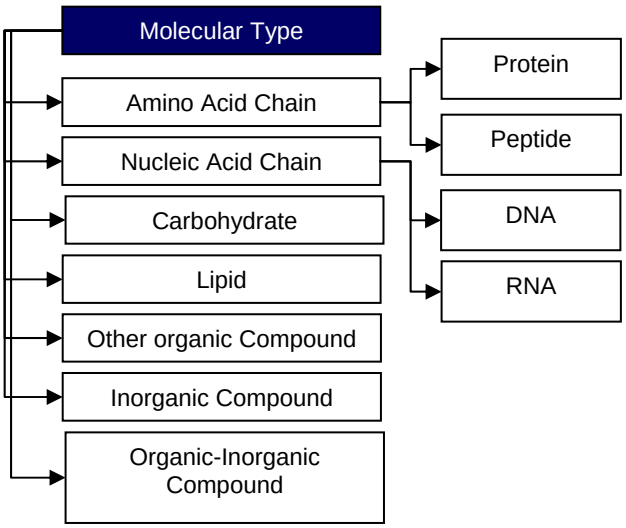


Figure 5 - "*Target Type*" class and subclasses



For the same purposes of molecular species classification and information integration, the class "*Molecular Type*" has been created (see figure 6).

Figure 6 - "*Molecular Type*" class and subclasses



- Class properties

The "*Binder Type*", "*Target Type*" "*Molecular Type*" class properties provide information on the biochemical category it belongs to ("*Molecular Type*" class) or/and the nature of chemical structure it possesses (instance of "*Chemical structure*" class). (See Appendix 1 - Table 4, 5 and 6)

3.1.5. "*Molecular State*"

- Concept

To indicate the structural state of target molecules used in experiments, a "Molecular State" class has been created. This information was considered as critical since it conditions the molecular recognition by the binders.

- Class properties

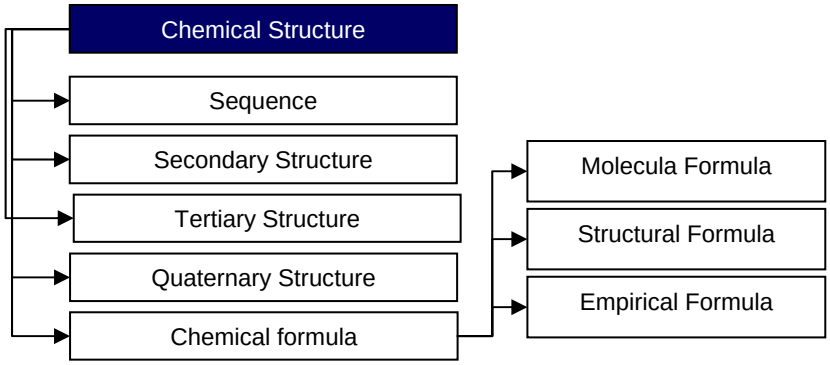
The class properties include a set of Boolean expression allowing the description of the native (or denatured), the folded (or unfolded) and the aggregated (or not) state of the molecules. It also links the "*Molecular State*" to the actual "*target*" it refers (instance of "*Target*"). (See Appendix 1 - Table 7)

3.1.6. "*Chemical Structure*"

- Concept

According to their biochemical category, molecular species can display several level of structure organization. For instance, proteins and nucleic acids can be described in a structure hierarchy of four levels. Hence, this class allows the description of the general structure associated to each "*Molecular Type*". (See figure 7).

Figure 7 - "*Chemical Structure*" class and subclasses



- Class properties

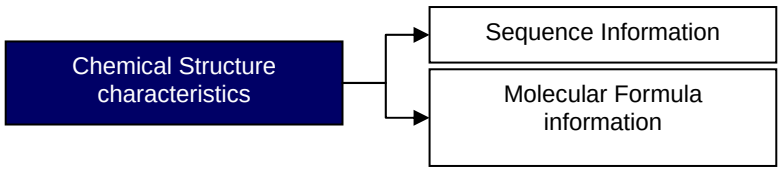
Currently, no particular properties have been identified for this class.
(See Appendix 1 - Table 8)

3.1.7. "*Chemical Structure Characteristics*"

- Concept

On the contrary to the previous class, the "*Chemical Structure Characteristics*" class allows the description of the specific structure associated with each "*Binder"* or *"Bound Entities*". This information will be represented by the sequence (primary structure) for proteins and nucleic acids, and by the molecular formula for smaller organic compounds. (See figure 8).

Figure 8 - "*Chemical Structure*" class and subclasses



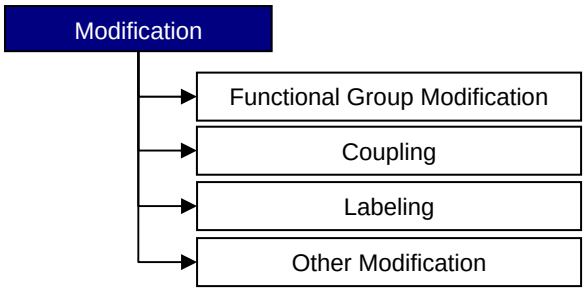
- Class properties

The class properties provide information on the molecular entity it refers to (instance of "*Binder"* or *"Bound Entities*") and the structure evidences (instance of "*Molecular Sequence Data*" or "*Molecular Formula Data"*). (See Appendix 1 - Table 9)

3.1.8. "*Modification*"

Initial chemical structure of molecules can be modified after synthesis for technical application purposes such as protein labelling, capture or immobilization. To represent these post-synthesis alterations of the initial molecular structure, the "*Modification*" class and subclasses have been created (See figure 9).

Figure 9 - "*Modification*" class and subclasses



- Class properties

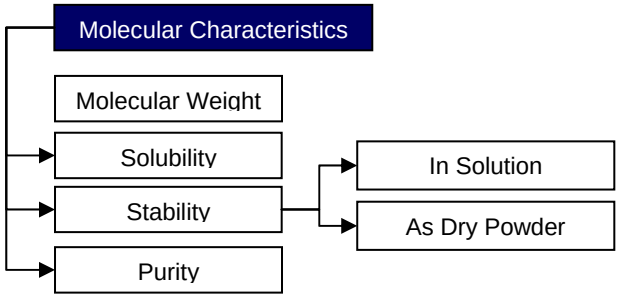
According to subclasses, the properties provide information on either type of chemical modification (for "*Functional Group Modification*" subclass), name of the coupled molecule (or support) (for "*Coupling*" subclass) or type and name of the label (for "*Labelling*" subclass). For "*Other Modification"* subclass, the properties include description and name of the reported modification. (See Appendix 1 - Table 10)

3.1.9. "*Molecular Characteristics*"

- Concept

Each molecular species is defined by specific molecular characteristics which are conditioned by their molecular structure. These molecular characteristics are expressed by the "*Molecular characteristics*" class. Based on the binder/target minimum information requirements, 4 molecular properties are currently expressed: "*Molecular Weight*", "*Solubility*", "*Stability* and "*Purity*". (See figure 10).

Figure 10 - "*Molecular Characteristics*" class and subclasses



- Class properties

The class properties provide information on the molecular species that the molecular characteristics refer to (instance of "*Binder"* or *"Bound Entity*") and the evidences linked to each "*Molecular Characteristics*" (instance of "*Experimental Data*"). In addition, subclass properties allow the detailed description of each type of "*Molecular Characteristics*". (See Appendix 1 - Table 11)

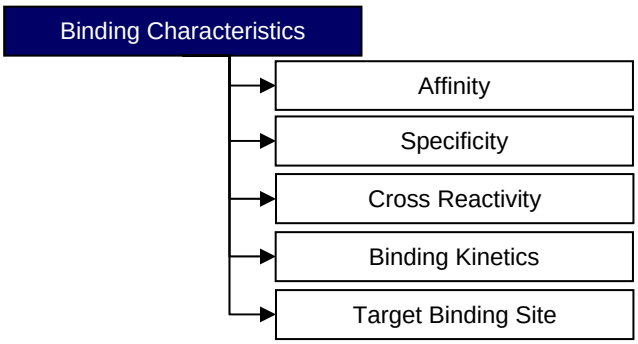
3.1.10. "*Binding Characteristics*"

- Concept

Binders are characterized by their ability to recognize unique molecular species (i.e. targets) in order to form a stable complex. These characteristics are expressed by the "Binding Characteristics" class which is further divided into subclasses to represent the following binding

parameters: "*Affinity*", "*Specificity*", "*Cross-reactivity*", "*Binding Kinetics*" and "*Target Binding Site*". (See figure 11).

Figure 11 - "*Binding Characteristics*" class and subclasses



- Class properties

The class properties provide information on the binder that the binding characteristics refer to (instance of "*Binder*") and the evidences linked to each "*Binding Characteristics*" (instance of "*Experimental Data*"). In addition, subclass properties allow the detailed description of each type of "*Binding Characteristics*". (See Appendix 1 - Table 12)

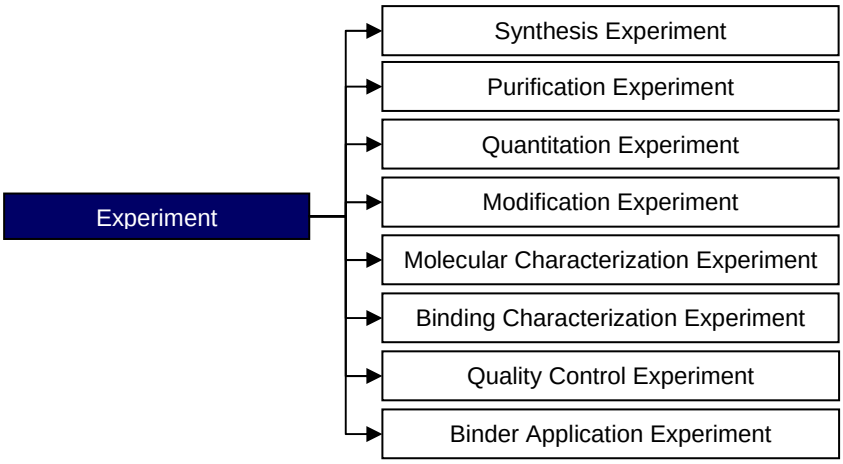
3.2. Representation of the experimental context

3.2.1. "*Experiment*"

- Concept

Each molecular entity is associated with a series of experiments represented by the "*Experiment*" class. This class is further divided in subclasses according to the different experiment goals (e.g. synthesis, purification, characterization, etc…). (See figure 12).

Figure 12 - "*Experiment*" class and subclasses



- Class properties

The class properties provide the information generally associated with an experiment (e.g. approach, method, setting, etc…) and indicate the nature of the results.

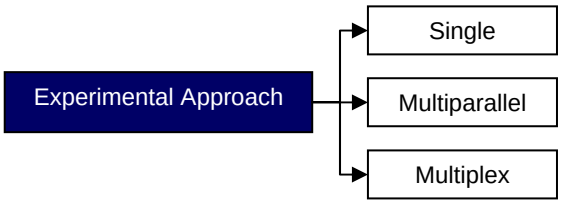
(See Appendix 1 - Table 13)

3.2.2. "*Experiment Approach*"

- Concept

To distinguish single experiments from experiments conducted in parallel or multiplex, the "Experimental approach" class has been created. (See figure 13).

Figure 13 - "*Experiment Approach*" class and subclasses



- Class properties

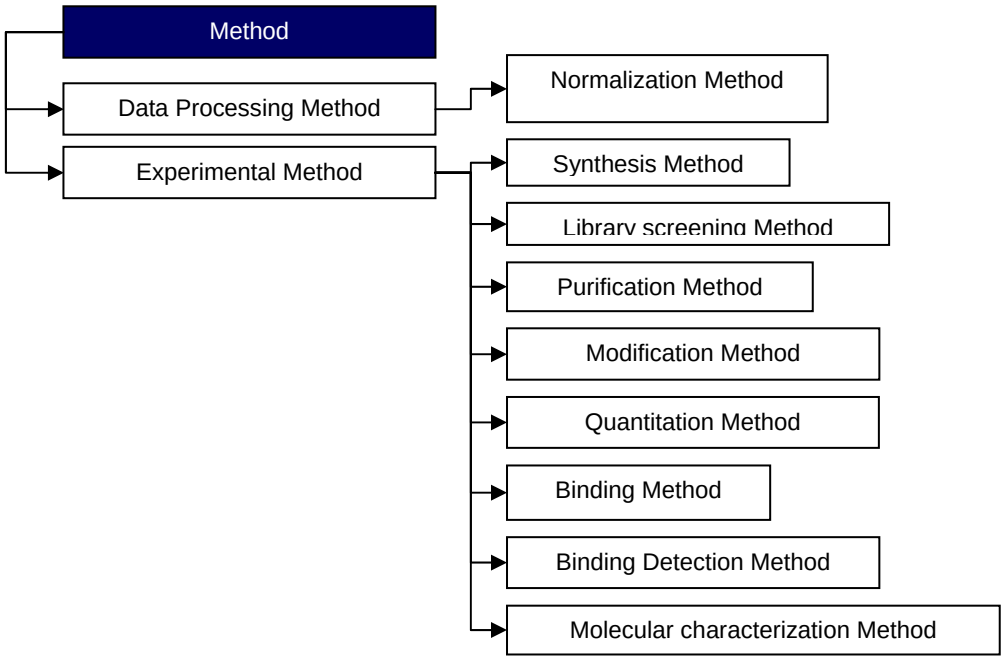
Currently, no particular properties have been identified for this class. (See Appendix 1 - Table 14)

3.2.3. "*Method*"

- Concept

The methodologies, applied in experiments or data processing, are represented in the "*Method*" class. (See figure 14).

Figure 14 - "*Method*" class and subclasses



- Class properties

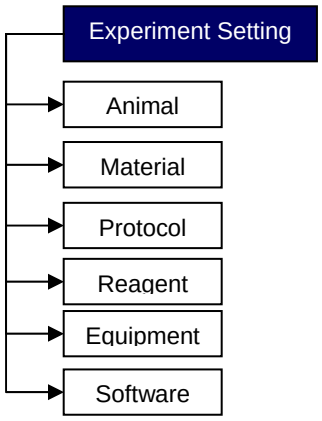
The class properties provide information on method name, principle and reference(s).
(See Appendix 1 - Table 15)

3.2.4. "*Experiment Setting*"

- Concept

The detailed information about animal model, material(s), protocol, reagent(s), software(s) and equipment(s) used in experiments is expressed in the "Experiment Setting" class. (See figure 15).

Figure 15 - "*Experiment Setting*" class and subclasses



- Class properties

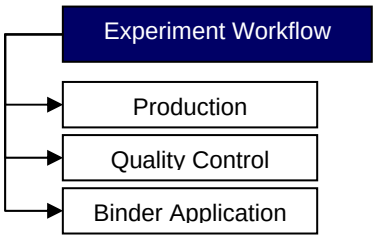
According to the subclass, the properties provide information details on (i) animal model species and strain, (ii) material name, source and reference, (iii) protocol title and reference, (iv) reagent name and reference, etc…(See Appendix 1 - Table 16)

3.2.5. "*Experimental workflow*"

- Concept

The combination of several experiments in a sequential order is represented by the "*Experimental workflow*" class. This class allows the description of experimental information about binder and target production, quality control and application. (See figure 16).

Figure 16 - "*Experimental Workflow*" class and subclasses



- Class properties

The class properties provide information on the experiments included in the workflow and link the experimental set to a type of data producer. (See Appendix 1 - Table 17)

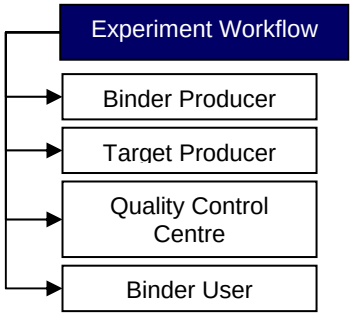
3.2.6. "*Experimentalist*"

- Concept

Experimentalists are categorized based on the task that they perform. Three types of experimentalist have been identified: the binder (or target) producers, the quality controllers and

the binder users. The association of a particular experimental set (i.e. experiment and resulting data) with a specific experimentalist gives the flexibility to report different data for the same assessment using the same method and experimental setting. (See figure 17).

Figure 17 - "*Experimentalist*" class and subclasses



- Class properties

The class properties provide information on experimentalist name, organization name and the experimental workflow performed. (See Appendix 1 - Table 18)

3.3. Representation of data

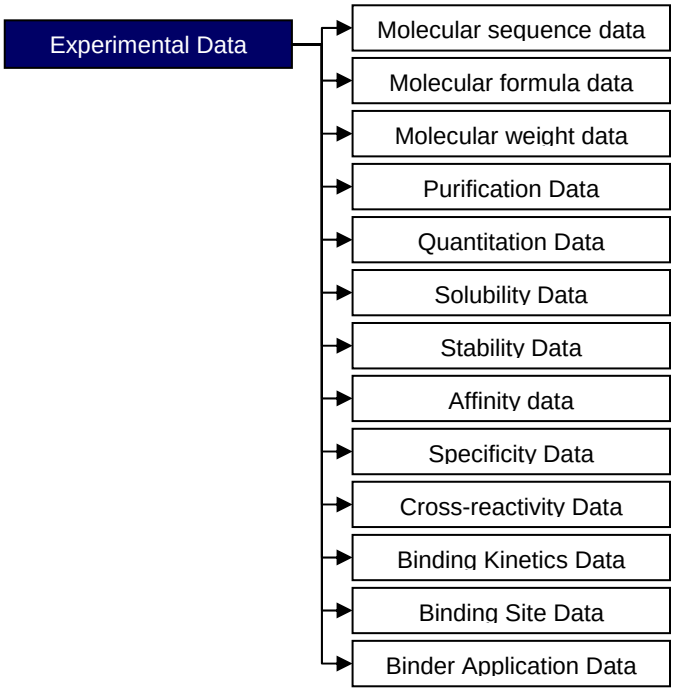
Several aspects need to be considered when representing data collected from experiments: the nature of the assessment, the analysis state of the reported data and the data display.

3.3.1. "*Experimental Data*"

- Concept

Data collected from experiments are represented in the class "*Experimental Data*". The class is further divided according to the nature of the assessment (e.g. sequence, quantitation, purity, affinity, etc….). (See figure 18).

Figure 18 - "*Experimental Data*" class and subclasses



- Class properties

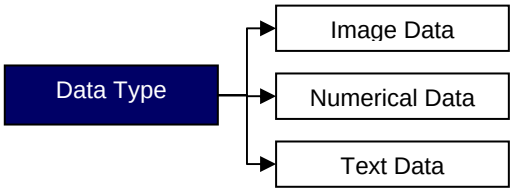
The class properties provide information on the corresponding processed (i.e. summary or processed data) or non processed (i.e. raw data) versions of the data set (instance of "*Data Report*"). They also link the data to the associated experiment (instance of "*Experiment*"). (See Appendix 1 - Table 19)

3.3.2. "*Data Type*"

- Concept

Experimental data set can be displayed under text, numerical or image format. To specify the nature of the available data, the "*Data Type*" class has been created. (See figure 19).

Figure 19 - "*Data Type*" class and subclasses



- Class properties

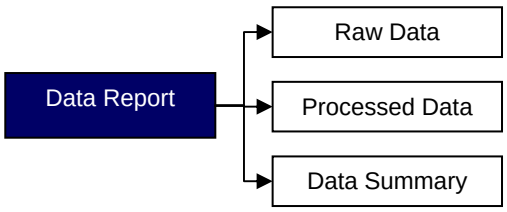
Currently, no particular properties have been identified for this class. (See Appendix 1 - Table 20)

3.3.3. "*Data Report*"

- Concept

For each experiment, the different data analysis states are represented in the "*Data report*" class. The direct report of harvested data without further data processing is represented by the subclass "*Raw Data*". Report corresponding to further processing of the raw data is represented by the "*Processed Data*" subclass. Finally, the "*Data summary*" subclass represents the experimentalist conclusive report based on raw or processed data. (See figure 20).

Figure 20 - "*Data Report*" class and subclasses



- Class properties

The general class properties provide information on the corresponding data file (i.e name and file format) and the nature of the data ("*Data Type*" class). Additional properties are included in the "*Processed Data*" subclass which informs on the processing method (instance of "*Data Processing Method*") and the initial data set used (instance of "*Raw data*"). The "Data Summary" class properties also include information on the initial data set used (instance of "*Raw data*") but also allow to alternatively provide the summary content as free text instead of data file. For instance, this can be particularly useful when only one measurement value is to be reported as data summary (e. g. molecular weight or affinity constant value). (See Appendix 1 - Table 21)

3.4. Representation of references

3.4.1. "*Reference*"

- Concept

An important link to other knowledge domain is the reference information. This information will link the ligand-binder knowledge base to other information resources such as literature or molecular database.

- Class properties

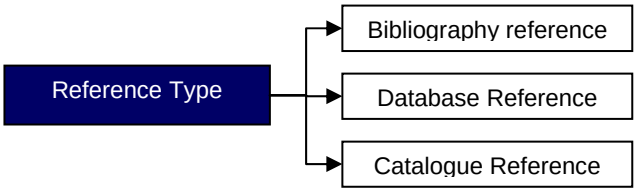
The class properties provide information on the nature of the reference (instance of "*Reference type*") and the instance of the class or subclass it pertains to. (See Appendix 1 - Table 22)

3.4.2. "*Reference Type*"

- Concept

Reference is a very general term which can describe various type or source of information. Therefore, the nature of the references has been categorized in the "Reference Type" class based on the information source type: literature, database or commercial catalogue. (See figure 21).

Figure 21 - "*Reference Type*" class and subclasses



- Class properties

For each type of information source, the corresponding references can be represented under a general set of properties. The "*Bibliography Reference*" subclass properties include the standard information found in literature references (e.g. authors, page numbers, journal, publication year, etc…). The "*Database Reference*" subclass properties include identifier, database name and URL. The "Catalogue Reference" subclass properties include catalogue number, batch number and organization name. (See Appendix 1 - Table 23)

4. Ontology application

The first application of the ligand-binders ontology will be the development of a ligand-binders knowledge base to support the ProteomeBinders community activities. The knowledge base will provide a structured and standardized information environment for experimental data collection and retrieval.

The ontology will also serve as a knowledge representation standard for the development of exchange format (i.e. PB-MI exchange format) to facilitate database interoperability. Indeed the ontology terms, their definitions and part of their hierarchical organisation will be used as controlled vocabulary for the exchange format.

Further application of the ontology will include the development of application for data mining, data comparison, experimental design (i.e. binder selection) and binder production prioritization.

CONCLUSION

The ligand binder ontology has been developed to provide a structured and standard representation of ligand binder experimental knowledge domain. The current version was derived from the minimal information for the unambiguous description of binder/protein pairs

(Deliverable 5.1.1 - Specifications of minimal information to unambiguously describe binder/protein pairs). It includes all the main concepts (and associated relationships) required for the description of experiments, the molecular species involved and collected data. It integrates ligand-binders information within general biochemical and biological context and provides links (i.e. references) to other knowledge domain as previously defined (Deliverable 5.1.5 - List of external sources of information to be linked to the ligand binder database). Finally, the ontology allows differentiating the experimental set (i.e. experiments and resulting data) according to the category of experimentalists: binder (or target) producer, quality control centre or binder user.

Future development steps will include the extension of some classes (e.g. “Method” and “Experiment Setting”) with more subclass divisions. In addition, the relationship network will be supplemented with more transversal relationships (such as intra-class relationships) which will be critical for computer reasoning applications.

NOTE

The actual ontology is available in two formats:

1. Protégé files. Protégé is a tool developed by the University of Stanford.

The tool can be dowloaded from the following URL:
http://protege.stanford.edu/download/release/full/. Current version is 3.3.1. It is a JAVA software so an installed JAVA virtual machine is required.
The ontlogy consists in 3 files: *.pprj files are Protégé project files and are the ones to open, the *.pins and *.pont files contain the actual data.

2. OWL file.

OWL is one of the standard ontology formats which is developed by the W3C and which is becoming more and more widely used.

The ontology in OWL format is the *.owl file. Protégé can be used to open the file, but a wide variety of other softwares are freely available to edit OWL files. Any OWL compatible ontology editor can be used.

ACKNOWLEDGEMENTS

We would like to express our thanks to the NA5.1 workgroup experts for their contribution to this work.
Our particular thanks to the participants of the Ligand-binder Ontology survey (see Appendix 2):
Ronald Frank
Zoltàn Konthur
Sylvia Krobitsch
Silvere van der Maarel
Serge Muyldermans
Martin Schlapshy
Oda Stoevesandt
Christer Wingren

APPENDIX 1

Ligand-Binders Ontology – Classes and subclasses description and properties tables

Table 1 - Description and properties of the "*Binder*" and "*Bound entities*" class and subclasses.

Class name	Description	Properties
Binder	A binder is a single molecular entity or a complex mixture of molecular entities which specifically recognizes a target.	• Name • Belong to Binder Type • Binds to Cross-Reacting Entity • Binds to Target • Belong to molecular type • Has chemical structure • Has modification • Has reference • Has source • Result from production • Structurally related to other molecular entity or entity of reference • Is characterized by Affinity • Is characterized by Binding Kinetics • Is characterized by Cross-Reactivity • Is characterized by Solubility • Is characterized by Specificity • Is characterized by Stability • Is characterized by Target Binding Site • Is characterized by chemical structure characteristics • Is characterized by molecular weight

Table 1 - Description and properties of the "*Binder*" and "*Bound entities*" class and subclasses. (Continued)

Class name	Description	Properties
Bound Entities	Bound entities are single molecular entities or complex mixtures of molecular entities which are recognized by a binder.	• Name • Is bound by Binder • Belong to molecular type • Has chemical structure • Has modification • Has reference • Has source • Is characterized by chemical structure characteristics • Is characterized by molecular weight • Result from production • Structurally related to other molecular entity or entity of reference
Subclass name	**Description**	**Supplementary properties**
Target	A target is a single molecular entity or a complex mixture of molecular entities which is specifically recognized by a binder.	• Belong to Target Type • Has Molecular State • Is characterized by Purity
Cross-Reactant	A cross-reactant is a single molecular entity or a complex mixture of molecular entities which is recognized by a binder in addition to its designed target.	(None)

Table 2 - Description and properties of the "*Source*" class and subclasses.

Class name	Description	Properties
Source	A source defines the natural or synthetic origin of a molecular entity or complex mixture of molecular entities.	• Has Reference • Name • Is the source of Molecular Entity
Subclass name	**Description**	**Supplementary properties**
Natural	A natural source is any part of micro-organisms, plants, or animals that has been processed, extracted or in whole	(None)
Synthetic	A synthetic source corresponds to the design and creation of artificial molecular entities or the re-design of existing, natural molecular entities.	(None)

Table 3 - Description and properties of the "*Entity of reference*" class and subclasses.

Class name	Description	Properties
Entity of Reference	An entity of reference is a characterized entity which a molecular entity is structurally related to.	• Has reference • Name
Subclass name	**Description**	**Supplementary properties**
Gene Product Reference Entity	A gene product reference entity is a characterized gene product which a molecular entity is structurally related to.	• Gene name • Gene symbol
Amino Acid Sequence Motif Reference Entity	An amino acid sequence motif reference entity is a characterized amino acid sequence motif which a molecular entity is structurally related to.	(None)
Chemical Reference Entity	A chemical reference entity is a characterized chemical which a molecular entity is structurally related to.	(None)

Table 4 - Description and properties of the "*Binder Type*" class and subclasses.

Class name	Description	Properties
Binder Type	A binder type corresponds to a category of binders.	• Belong to molecular type • Has chemical structure
Subclass name	**Description**	**Supplementary properties**
Immunoglobulin	An immunoglobulin is any of the structurally related proteins that function as antibodies. They are divided into five classes (IGA, IGD, IGE, IGG, IGM) on the basis of structure and biological activity..	(None)
Subclass name	**Description**	**Supplementary properties**
Intact Antibody	An intact antibody corresponds to a whole unmodified immunoglobulin.	• Belong to Target Type
Subclass name	**Description**	**Supplementary properties**
Monoclonal Antibody	A monoclonal antibody is an antibody produced by a clone or genetically homogenous fast-growing cells i.e., hybridoma. Hybridoma cells are cloned to establish cell lines producing a specific antibody that is chemically and immunologically homogeneous.	(None)

Table 4 - Description and properties of the "*Binder Type*" class and subclasses. (Continued)

Subclass name	Description	Supplementary properties
Polyclonal Antibody	A polyclonal antibody is a heterogeneous antibody derived from a host organism, typically produced by immunization of a suitable mammal with an antigen which induces B-lymphocytes to produce IgG immunoglobulins specific for the antigen. The resulting collection of antibodies has specificity for different epitopes of the antigen.	(None)
Antibody Fragment	An antibody fragment is a part of an immunoglobulin molecule or an association of parts of immunoglobulin molecules in a new stable complex.	• oligomeric status
Fab	A Fab fragment is a part of an immunoglobulin antibody that consists of both a light chain and part of a heavy chain. By comparison, natural antibodies consist of two heavy and two light chains. An FAB (fragment antibody) offers the advantages of smaller size and lower cross-reactivity compared to the complete antibody.	
Fab2	A Fab2 is a genetically engineered antibody derivate consisting in two Fab fragments joined together by a flexible peptide linker.	
Fv	A Fv is a genetically engineered antibody derivate consisting of both the variable heavy chain (VH) and the light chain (VL) of an immunoglobulin.	
Single Chain Fv	A single chain Fv is a genetically engineered antibody derivate consisting of both the variable heavy chain (VH) and the light chain (VL) of an immunoglobulin. These entities are joined together by a flexible peptide linker.	
Fab3	A Fab3 is a genetically engineered antibody derivate consisting in three Fab fragments joined together by a flexible peptide linker.	
Minibody	A minibody is a genetically engineered antibody derivate consisting in the VL and VH domains of a native antibody fused to the hinge region and CH3 domain of the immunoglobulin molecule.	
BisScFv	A BiScFv is a genetically engineered antibody derivate consisting in the two ScFv joined together by a flexible peptide linker.	
Diabody	A diabody is a genetically engineered antibody derivate consisting in two minibodies joined together in a dimer which can be either monovalent or divalent.	
Triabody	A triabody is a genetically engineered antibody derivate consisting in thee minibodies joined together in a trimer which can be from monovalent to trivalent.	

Table 4 - Description and properties of the "*Binder Type*" class and subclasses. (Continued)

	Subclass name	Description	Supplementary properties
	Tetrabody	A tetrabody is a genetically engineered antibody derivate consisting in four minibodies joined together in a quadrimer which can be from monovalent to quadrivalent.	
	scFv-Fc	A scFv-Fc is a genetically engineered antibody derivate consisting in a fusion molecule between an scFv fragment and an Fc region from an antibody	
Protein Scaffold		A protein scaffold is an engineered amino acid sequence based on a protein framework chosen for its specific biophysical properties.	(None)
	Subclass name	**Description**	**Supplementary properties**
	Ankyrin Repeats	Ankyrin Repeats are tandem modules of about 33 amino acids. The conserved domain structure has been described as side-by-side anti-parallel alpha helices connected by intervening beta hairpin motifs or as beta, alpha, alpha, beta secondary structures or as an L-shaped beta-hairpin and two alpha-helices. The repeats associate to form a higher order structure. Despite sequence variation, the domain core maintains a stable surface of contact residues to mediate protein-protein interactions. The Ankyrin cytoskeleton protein is composed almost entirely of these repeats.	• Belong to Target Type • Has Molecular State • Is characterized by Purity
	Armadillo Repeats	Armadillo repeat proteins are abundant eukaryotic proteins involved in several cellular processes, including signalling, transport, and cytoskeleton regulation. These sequences constitute the starting point for the generation of designed armadillo repeat protein libraries for the selection of peptide binders, exploiting their modular structure and their conserved binding mode.	(None)
	Leucine Rich Repeats	Leucine rich repeats are right-handed beta-alpha super helix used in the design of protein scaffolds.	(None)
	Zinc Finger	Zinc finger (Znf) domains are relatively small protein motifs that bind one or more zinc atoms, and which usually contain multiple finger-like protrusions that make tandem contacts with their target molecule. They are used in the design of protein scaffolds.	(None)
	Lipocalin	Some proteins of the Lipocalin family present a beta-barrel structure similar to the one which was first identified in the retinol-binding protein (RBP) and which provides them with interesting binding properties. They are used in the design of protein scaffolds.	(None)
	Affibody	Affibodies are engineered binding proteins for which the three-helix bundle Z domain derived from Staphylococcus aureus protein A (SPA), is used as a scaffold.	(None)

Table 4 - Description and properties of the "*Binder Type*" class and subclasses. (Continued)

Subclass name	Description	Supplementary properties
Binding Peptide	A binding peptide is a peptide whose amino acid sequence is artificially designed or gene encoded.	• Chain conformation (i.e. linear, cyclic, branched)
Nucleic Acid Aptamer	Nucleic acid aptamers are double-stranded DNA or single-stranded RNA oligonucleotide sequences which were selected from random pools because of their capacity to recognize target molecules. Fundamentally distinct from antibodies, aptamers mimic their properties.	(None)
Small Molecule	A small molecule is a molecule with a low molecular weight that is not determined by a genome sequence.	(None)
Hybrid Molecule	A hybrid molecule is a molecule containing structure parts corresponding to at least two different biochemical categories (e.g. glycopeptides).	(None)

Table 5 - Description and properties of the "*Target Type*" class and subclasses.

Class name	Description	Properties
Target Type	A target type corresponds to a category of targets.	• Belong to molecular type • Has chemical structure
Subclass name	**Description**	**Supplementary properties**
Gene Product Related	A gene product related entity is a protein entity whose sequence (in whole or partial) is encoded by a gene.	(None)
Subclass name	**Description**	**Supplementary properties**
Gene Product	A gene product is a protein whose sequence corresponds to the whole sequence of a gene expression product.	(None)
Subclass name	**Description**	**Supplementary properties**
Fused Gene Product	A fused gene product is a protein whose sequence is fused to another amino acid chain, encoded or not by a gene, for instance for labelling or immobilisation purposes.	• Fusion protein name • Fusion type (i.e. label or tag)
Gene Product Fragment	A gene product fragment is a protein whose sequence is a contiguous section of a gene product sequence.	(None)
Fused Gene Product Fragment	A fused gene product fragment is a protein whose sequence is fused to another amino acid chain, encoded or not by a gene, for instance for labelling or immobilisation purposes.	• Fusion protein name • Fusion type (i.e. label or tag)

Table 5 – Description and Properties of the *"Target Type"* class and subclasses. (Continued)

	Subclass name	Description	Supplementary properties
	Gene Product Related Peptide	A gene product related peptide is a peptide whose sequence is identical to a contiguous part of a gene product sequence.	(None)
Amino Acid Sequence Motif		An amino acid sequence motif is a highly conserved sequence pattern that has, or is conjectured to have, a biological significance.	(None)
	Subclass name	**Description**	**Supplementary properties**
	Post Translational Modification	Post translational modification (PTM) motifs are amino acid sequence motifs recognized by enzyme(s) for specific chemical modifications of protein covalent structure. Examples of PTM include glycosylation, acylation, limited proteolysis, phosphorylation, isoprenylation.	(None)
	Localization	A localization motif is an amino acid sequence motif recognized for specific translocations of protein to particular subcellular compartment, cell machinery or extra cellular compartment.	(None)
	Functional	Functional motifs are amino acid sequence motif which confer to the protein its biological activity (e.g. catalytic activity) or specific binding properties.	(None)

Table 6 - Description and properties of the "*Molecular Type*" class and subclasses.

Class name		Description	Properties
Molecular Type		The molecular type corresponds to the biochemical class of a molecule.	• Has Chemical Structure
	Subclass name	**Description**	**Supplementary properties**
	Amino Acid Chain	An amino acid chain is a compound of two or more amino-acids (organic compounds containing amino (-NH2) and carboxyl (-COOH) groups) belonging to the twenty-plus L-alpha-amino acids found in proteins, where the alpha carboxyl group of one is bound to the alpha amino group of another	(None)
	Subclass name	**Description**	**Supplementary properties**
	Protein	A protein is an amino acid chain compound encoded by a gene.	• number of polypeptide chains
	Peptide	A peptide is a short amino acid chain which length do not excess 100 residues long.	(None)

Table 6 - Description and properties of the "*Molecular Type*" class and subclasses. (Continued)

Subclass name	Description	Supplementary properties
Nucleic Acid Chain	A nucleic acid chain is a complex compound of high molecular weight occurring in living cells which consist of nucleotides (nucleoside phosphates linked together by phosphate bridges).	• Number of strands
Subclass name	**Description**	**Supplementary properties**
DNA	DNA is a high molecular weight, linear polymers, composed of nucleotides containing deoxyribose and linked by phosphodiester bonds.	(None)
RNA	RNA is a high molecular weight, linear polymers, composed of nucleotides containing ribose and linked by phosphodiester bonds.	(None)
Carbohydrate	Carbohydrates are organic compounds, including starches, glycogens, cellulose, gums, and simple sugars. Carbohydrates are composed of carbon, hydrogen, and oxygen in a ratio of Cn(H2O)n.	(None)
Lipid	Lipids are hydrocarbon-containing organic compounds. Lipids play important roles in living organisms: these roles include functioning as energy storage molecules, serving as structural components of cell membranes, and constituting important signalling molecules. Lipids can be subdivided into 2 groups: fatty acids and glycerides.	(None)
Other Organic Compound	Other organic compounds are any organic compounds which do not belong to the biochemical classes of amino acid chains, nucleic acid chains, carbohydrates and lipids.	(None)
Inorganic Compound	Inorganic compounds are substances, principally from mineral sources of non-biological origin, encompassing all those that do not include carbon and its derivatives as their principal elements.	(None)
Organic-Inorganic Compound	Organic inorganic compounds are hybrid compounds which exhibit organic and inorganic molecular structures.	(None)

Table 7 - Description and properties of the "*Molecular State*" class and subclasses.

Class name	Description	Properties
Molecular State	The state of a protein is characterized by covalent or non covalent modifications of the native chain secondary, tertiary or quaternary structure	• Aggregated (yes/no) • Denaturated (yes/no) • Molecular state of Target • Unfolded (yes/no)

Table 8 - Description and properties of the "*Chemical Structure*" class and subclasses.

Class name	Description	Properties
Chemical Structure	The chemical structure corresponds to the components and their relationship and manner of arrangement in constituting a whole molecule.	(None)
Subclass name	**Description**	**Supplementary properties**
Sequence	The sequence or primary structure of polypeptides, proteins or nucleic acids is the sequence of amino acids in the polypeptide chain with reference to the locations of any disulfide bonds or the order of nucleotides in a nucleic acid chain. The primary structure may be thought of as a complete description of all of the covalent bonding in a polypeptide chain or protein.	(None)
Secondary Structure	The secondary structure is the general three-dimensional form of local segments of biopolymers such as proteins and nucleic acids. Regular hydrogen-bond interactions within contiguous stretches of polymer chain give rise to alpha helices, beta strands (which align to form beta sheets) or other types of coils. This is the first folding level of protein polymer chain conformation.	(None)
Tertiary Structure	The tertiary structure of a biopolymer describes the folding of the biopolymer chain to assemble the different secondary structure elements in a particular arrangement.	(None)
Quaternary Structure	The quaternary structure is the arrangement of multiple folded polymer molecules in a multi-subunit complex.	(None)
Chemical Formula	The chemical formula represents the elemental composition of a compound.	(None)
Subclass name	**Description**	**Supplementary properties**
Molecular Formula	The molecular formula indicates the actual number of atoms of each element present in a single molecule of a chemical compound.	(None)
Structural Formula	The structural formula is a graphical representation of the molecular structure showing how the atoms are arranged.	(None)
Empirical Formula	The empirical formula is a simple expression of the relative number of each type of atom in a compound.	(None)

Table 9 - Description and properties of the "*Chemical Structure Characteristics*" class and subclasses.

Class name	Description	Properties
Chemical Structure Characteristics	The chemical structure characteristics correspond to experimental determination of the chemical structure of a molecular entity.	• Chemical structure characteristics of Molecular Entity • Demonstrated by Experimental Data
Subclass name	**Description**	**Supplementary properties**
Sequence information	The sequence information corresponds to the actual description of the sequence of a molecular entity as experimentally defined.	• Starts (i.e. residue number) • Ends (i.e. residue number) • Numbering from sequence of reference (yes/no) • Sequence standard format (e.g. FASTA) • Value (i.e. sequence)
Molecular formula information	The molecular formula information corresponds to the actual description of the molecular formula of a molecular entity as experimentally defined.	• Value (i.e. formula)

Table 10 - Description and properties of the "*Modification*" class and subclasses.

Class name	Description	Properties
Modification	A modification consists in the artificial alteration of the covalent structure of a molecule after its synthesis, for instance for labelling or immobilisation purposes.	(None)
Subclass name	**Description**	**Supplementary properties**
Functional Group Modification	A functional group modification is a post synthesis modification of any of the molecule functional groups.	• Chemical modification type (e.g. phosphorylation, glycosylation, etc.…)
Coupling	Coupling is a post synthesis modification which results in the conjugation between a molecule and another molecule or solid phase.	
Labelling	Labelling is a post synthesis modification which results in the conjugation between a molecule and a fluorescent, enzymatic, biotin or radioactive label.	• Label name • Label type (e.g. fluorescence, radioactive, etc…)
Other Modification	Other modification is a post synthesis modification excluding labelling, coupling or functional group modification.	• Description • Name

Table 11 - Description and properties of the "*Molecular Characteristics*" class and subclasses.

Class name	**Description**	**Properties**
Molecular Characteristics	Molecular characteristics are the distinguishing qualities of a molecule.	• Demonstrated by Experimental Data
Subclass name	**Description**	**Supplementary properties**
Molecular Weight	The molecular weight is the he sum of the relative atomic masses of the constituent atoms of a molecule. It may be only available as an experimentally acquired estimate.	• Molecular Weight of Molecular Entity • Unit • Value
Solubility	The solubility corresponds to the ability of a particular molecule to dissolve in a particular solution at a given temperature.	• Solubility of Binder • Solution volume • Solution type • Temperature • Weight
Stability	The stability corresponds to the quality of being free from change or variation over time or not taking part readily in chemical change.	• Stability of Binder • Temperature • Time
Purity	The purity corresponds to a quantitative assessment of the homogeneity or uniformity of a mixture. It refers to the degree of being free of contaminants or heterogeneous components.	• Purity of Target • Percentage

Table 12 - Description and properties of the "*Binding Characteristics*" class and subclasses.

Class name	**Description**	**Properties**
Binding Characteristics	Binding characteristics are the distinguishing qualities of a binder.	• Demonstrated by Experimental Data
Subclass name	**Description**	**Supplementary properties**
Affinity	The affinity corresponds to the strength of non-covalent chemical binding between two molecules as measured by the dissociation or association constant of the complex.	• Affinity constant (i.e KD value and unit) • Affinity of Binder
Specificity	The specificity corresponds to the ability of a binder to bind to its defined target and only its target.	• Specificity of Binder • Value (i.e. Target name)
Cross-reactivity	The cross reactivity correspond to the observation that a binder binds to characterized or uncharacterised molecular entities whose molecular characteristics differ from its target.	• Cross-reactivity of Binder • Cross-reactivity statement (yes/no) • Value (i.e. cross-reacting entity name(s))
Binding Kinetics	The binding kinetics corresponds to the measurement of the association and dissociation rates of a binder-target complex.	• Binding kinetics of Binder • Association constant (i.e. kon value and unit) • Dissociation constant (i.e. koff value and unit)
Target binding site	The binding site corresponds to the exact part of a target which is recognised by a binder. It is also called epitope.	• Target binding site of Binder
Subclass name	**Description**	**Supplementary properties**
Target Linear Binding Site	A target linear binding site consists in a continuous sequence of amino acids recognized by a binder.	(None)
Target Non Linear Binding Site	A target non linear binding site consists in several non contiguous sequences of amino acids. In the tertiary or quaternary structure of the target related molecular entity, these amino acids are close together and form a structure which is recognized by the binder.	• Residue number

Table 13 - Description and properties of the "*Experiment*" class and subclasses.

Class name	Description	Properties
Experiment	An experiment is a set of actions and observations, performed to verify or falsify a hypothesis or research a causal relationship between phenomena or obtain a product.	• Has description • Has Experimental Approach • Has Experimental Method • Has Experiment Setting • Has reference • Title
Subclass name	**Description**	**Supplementary properties**
Synthesis Experiment	A synthesis experiment consists in a series of steps intended to synthesize a specific molecular entity.	• Synthesizes molecular entity • Produces Experimental Data
Purification Experiment	A purification experiment consists in a series of steps intended to isolate a single molecular entity or a subset of molecular entities to (at least) apparent homogeneity from a complex mixture.	• Produces Experimental Data
Quantitation Experiment	A quantitation experiment consists in a series of steps which aims at measuring or estimating the quantity of a molecular entity present in a solution.	• Produces Experimental Data
Molecular Characterization Experiment	A molecular characterization experiment consists in a series of steps intended to measure or estimate molecular characteristics of molecules.	• Produces Experimental Data
Modification Experiment	A modification experiment consists in a series of steps intended to modify the chemical structure of a molecule after its synthesis.	(None)
Binding Characterization Experiment	A binding characterization experiment consists in a series of steps intended to measure or estimate binding characteristics of a binder.	• Produces Experimental Data
Binder Application Experiment	A binder application experiment consists in a series of steps which includes the use of a binder.	• Produces Experimental Data
Library Screening Experiment	A library screening experiment consists in a series of steps intended to select binders according to some biophysical properties of interest.	• Number of round • Size of the screen

Table 14 - Description and properties of the "*Experimental Approach*" class and subclasses.

Class name		Description	Properties
Experimental Approach		The experimental approach provides information regarding the dimensionality of an experiment.	(None)
	Subclass name	**Description**	**Supplementary properties**
	Single	A single experimental approach corresponds to an experimental setting where one single sample is being assessed at once for one parameter.	(None)
	Multi-parallel	A multi-parallel experimental approach corresponds to an experimental setting where different samples are assessed in parallel (different wells) for the same parameter or different parameters are assessed in parallel (different wells) in the same sample.	(None)
	Multiplex	A multiplex experimental approach corresponds to an experimental setting where different parameters are assessed in the same sample or well.	(None)

Table 15 - Description and properties of the "*Method*" class and subclasses.

Class name	Description	Properties
Method	A method is a means, manner of procedure, or systematic course of actions which have to be performed in order to accomplish a particular goal.	• Has reference • Name • Principle
Subclass name	**Description**	**Supplementary properties**
Data Processing Method	A data processing method describes series of operations on data made by a computer in order to retrieve or transform or classify information.	• Data processing method of Processed Data
Subclass name	**Description**	**Supplementary properties**
Normalization Method	A normalization method aims at making something conform to a standard; at removing useless or extraneous entries from a data set.	(None)
Experimental Method	An experimental method describes the general approach chosen to perform an experiment.	(None)
Subclass name	**Description**	**Supplementary properties**
Synthesis Method	A synthesis method describes the general procedure used to produce a biochemical or chemical compound, usually by the union of simpler compounds.	(None)
Library Screening Method	A screening method describes the general procedure used to select molecular entities based on particular properties they exhibit.	(None)
Binding Method	A binding method describes the general procedure used to assess the binding or measure or estimate binding characteristics of a binder.	(None)
Purification Method	A purification method describes the general procedure used to isolate a specific entity or a subset of entities from a complex mixture.	(None)
Modification Method	A modification method describes the general procedure used to modify a molecular entity after its synthesis.	(None)
Binding Detection Method	A binding detection method describes the general procedure used to reveal the binding of a binder to a molecular entity.	(None)
Molecular Characterization Method	A molecular characterization method describes the general procedure used to measure or estimate molecular characteristics of a molecular entity.	(None)
Quantitation Method	A quantitation method describes the general procedure used to measure or estimate the quantity of a molecular entity present in a solution.	(None)

Table 16 - Description and properties of the "*Experimental Setting*" class and subclasses.

Class name		**Description**	**Properties**
Experimental Setting		Experimental settings describe the conditions in which an experiment has been conducted.	(None)
	Subclass name	**Description**	**Supplementary properties**
	Animal	The animal describes the species and strain used in the experiment.	• Species • Strain
	Material	A material is the specific biological (or chemical) components with which a particular experiment is performed (e.g. cell lines, protein extracts, tissue slides, protein chips, antibodies, purified proteins, etc…).	• Has reference • Has source • Manufacturer • Name
	Protocol	A protocol is the formal plan of an experiment or research activity, including the objective, rationale, design, materials and methods for the conduct of the study; intervention description, and method of data analysis.	• File format • File name • Title
	Reagent	A reagent is a chemical that has general laboratory application and that is not labelled or otherwise intended for a specific experiment. General purpose reagents include antibody second reagents, detection reagents, cytological preservatives, fixatives and adhesives, tissue processing reagents, isotonic solutions, and pH buffers.	• Has reference • Manufacturer • Name
	Equipment	A piece of equipment is an object, or item of electrical or electronic equipment, which is designed to carry out a specific function or set of functions.	• Has reference • Model • Manufacturer
	Software	Software is a set of coded instructions, which a computer follows in processing data, performing an operation, or solving a logical problem, upon execution of the program.	• Name

Table 17 - Description and properties of the "*Experimental Workflow*" class and subclasses.

Class name		Description	Properties
Experimental workflow		An experimental workflow is a reliably repeatable pattern of experimental activity enabled by a systematic organization of resources, into a work process that can be documented and learned.	• Is performed by Experimentalist
	Subclass name	**Description**	**Supplementary properties**
	Production	The production is the workflow which leads to the physical availability of a specific molecular entity.	• includes-Binding Characterization Experiment • includes-Synthesis Experiment • includes-Modification Experiment • includes-Molecular Characterization Experiment • includes-Purification Experiment • includes-Quantitation Experiment • includes-Screening Experiment
	Quality Control	The quality control is the workflow which leads to the assessment of a binder ability to recognize its designated target.	• includes-Binding Characterization Experiment • includes-Molecular Characterization Experiment
	Binder Application	A binder application is a workflow which implies using a particular binder.	• includes-Molecular Characterization Experiment • includes-Binding Characterization Experiment • includes-Binding Application Experiment

Table 18 - Description and properties of the "*Experimentalist*" class and subclasses.

Class name	Description	Properties
Experimentalist	The experimentalist is the person/ laboratory/institution who performs an experiment.	• Name • Organization Name • Performs Experimental Workflows
Subclass name	**Description**	**Supplementary properties**
Binder Producer	A binder producer is the person/laboratory/institution who produces a binder through a production experimental workflow.	(None)
Target Producer	A target producer is the person/laboratory/institution who produces a target through a production experimental workflow.	(None)
Quality Controller	A quality controller is the person/laboratory/institution who assesses experimentally the quality of a binder towards its target through a quality control workflow.	(None)
Binder user	A binder user is a person/laboratory/institution who uses the binder in a binder application experiment.	(None)

Table 19 - Description and properties of the "*Experimental Data*" class and subclasses.

Class name	Description	Properties
Experimental Data	Experimental data correspond to the actual results collected from an experiment.	• Experimental data of Experiment • Has data summary • Has processed data • Has raw data
Subclass name	**Description**	**Supplementary properties**
Molecular Sequence Data	The molecular sequence data correspond to the data associated to the experimental determination of the numbered amino acid or nucleotide sequence.	(None)
Molecular Formula Data	The molecular sequence data correspond to the data associated to the experimental determination of the molecular formula.	(None)
Molecular Weight Data	The molecular weight data correspond to the data associated to the experimental estimate of the molecular weight.	(None)
Quantitation Data	The quantitation data correspond to the data associated to the experimental estimate of the molecular entity production yield.	(None)

Table 19 - Description and properties of the "*Experimental Data*" class and subclasses. (Continued)

	Subclass name	Description	Supplementary properties
	Purification Data	The purification data correspond to the data associated to the experimental estimate of the percentage of purity.	(None)
	Solubility Data	The solubility data correspond to the data associated to the experimental estimate of the solubility.	(None)
	Stability Data	The stability data correspond to the data associated to the experimental estimate of the stability.	(None)
	Affinity Data	The affinity data correspond to the data associated to the experimental estimate of the affinity.	(None)
	Specificity Data	The specificity data correspond to the data associated to the experimental estimate of the specificity.	(None)
	Cross Reactivity Data	The cross reactivity data correspond to the data associated to the experimental estimate of the cross reactivity.	(None)
	Binding Kinetics Data	The binding kinetics data correspond to the data associated to the experimental estimate of the binding kinetics.	(None)
	Binding Site Data	The binding site data correspond to the data associated to the experimental determination of the binding site of a binder.	(None)
	Binder Application Data	The binding application data correspond to the data associated to the experimental use of a binder.	(None)

Table 20 - Description and properties of the "*Data Type*" class and subclasses.

Class name	Description	Properties
Data Type	The data type describes the general category of the data format.	(None)
Subclass name	**Description**	**Supplementary properties**
Image Data	Image data are data presented as images.	(None)
Numerical Data	Numerical data are data presented as numbers, be it one single number or number lists.	(None)
Text Data	Text data are data presented as free text.	(None)

Table 21 - Description and properties of the "*Data Report*" class and subclasses.

Class name	Description	Properties
Data Report	A data report corresponds to the file containing the data which are the result of an experiment.	• File format • File name • Has Data Type
Subclass name	**Description**	**Supplementary properties**
Raw Data	A raw data report corresponds to the data which were directly acquired during the course of an experiment.	(None)
Processed Data	A processed data report corresponds to the data resulting from any transformation of the raw data acquired during the course of an experiment.	(None)
Summary Data	A data summary report corresponds to the conclusive report of the raw or processed data acquired during the course of an experiment.	(None)

Table 22 - Description and properties of the "*Reference*" class and subclasses.

Class name	Description	Properties
Reference	A reference is a note acknowledging a source of information.	• Reference of • Has Reference Type

Table 23 - Description and properties of the "*Reference Type*" class and subclasses.

Class name		**Description**	**Properties**
Reference Type		A reference type defines a category of references categories and the actual data included in each entry.	(None)
	Subclass name	**Description**	**Supplementary properties**
	Bibliography Reference	A bibliography reference is a reference to a journal publication.	• Author • Doi • First page number • Last page number • Journal name • Issue number • Journal volume number • Publication year
	Database Reference	A database reference is a reference to an entry in a public database.	• Database name • Identifier • Url
	Catalogue Reference	A catalog reference is the reference of a product in the catalog of a producer, a manufacturer or a supplier.	• Catalogue number • Organization name • Batch/lot number

APPENDIX 2

Topic 6 – Ligand-Binder Ontology

Ontology overview

What is an ontology? An ontology is a representation of some domain knowledge in terms of concepts and relationships which are understandable both by a human being and a computer program.

Domain and scope of the Ligand Binder Ontology. The ontology will cover the domain of protein binder technologies applied to human proteome analysis.

For what are we going to use the ontology? The terms in the ontology will be used for accessing protein binder information and associated reagents in order to perform proteome analysis experiments. It will also serve as a basis for more advanced reasoning such as identification of relationships among binders and targets or inference of the best binders for an experimental setting.

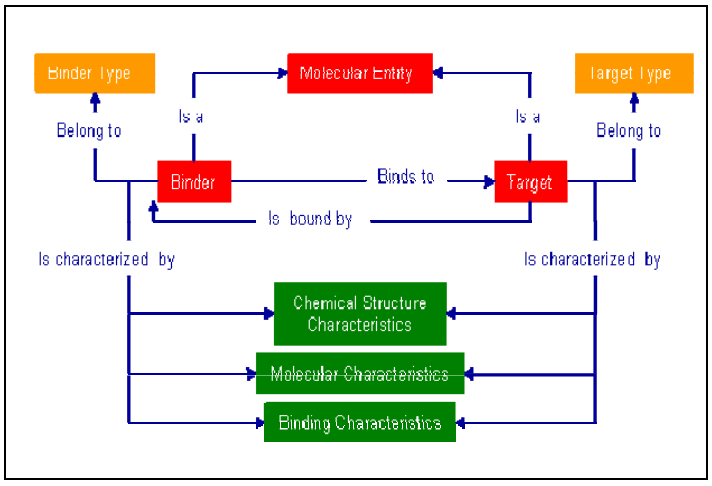
What kind of information the ontology should provide? The ontology should provide information on protein binder performance, quality and applicability. Furthermore, the level of detail of the information provided should enable the distinction between two protein binders based on their properties and application specification.

How is the ontology organized? The ontology is organized into a hierarchy of concepts, represented by classes and subclasses. Each class (or subclass) has properties which describe the various features of the concept and establish the relationships between classes (or subclasses) (see figure.1 below).

What are the ontology main concepts? The main concepts represented in the ontology are:

- Binders and targets represented by the *"Molecular Entity"* class and subclasses.
- Their properties represented by the *"Molecular Characteristics"*, *"Structural Characteristics"* and *"Binding Characteristics"* classes and subclasses.
- The experiments demonstrating binders and targets properties, as well as their association into a binder-target couple, represented by the "*Experimental Workflow*" and *"Experiment"* classes and subclasses.
- The experiment resulting data represented by the "*Experimental data*" class and subclasses.
- The data producers represented by the "*Experimentalist*" class and subclasses.

Figure 1 – Representation of the binder/target couple and their characteristics



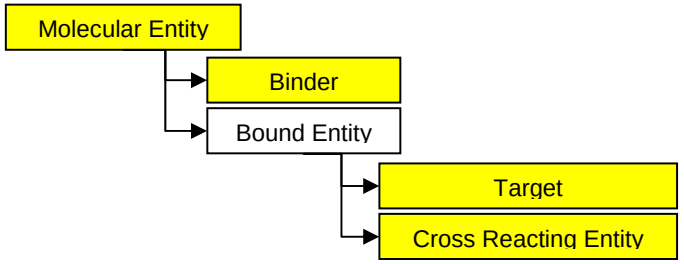
NA5.1 Standards, Ontology and Database Schema for Ligand Binders Information
March2008 - S. Palcy, J. Bourbeillon, D. Sherman and A. de Daruvar

Ontology Survey

The survey is divided into 12 tables, each corresponding to a main class requiring your review. Your approval (and suggestion if desired) will be requested about the description of classes or subclasses highlighted in yellow in the ontology schema presented above each table. A brief summary of the context in which the descriptions were written is given next to the schema. In this particular context, you will be asked to approve or reject the description presented in the table.

Table 1 – Representation of the binding partners

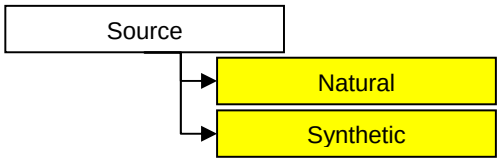
To represent the molecular species involved in a binding experiment, the class “molecular entity” and its subclasses “binder”, “target”, and “cross reacting entity” were created



In this context, do you totally agree with the following descriptions?	Yes	No	Corrected or new description (optional)
A **molecular entity** is any constitutionally or isotopically distinct atom, molecule, identifiable as a separately distinguishable entity.	☐	☐	[ADD YOUR REPLY HERE]
A **binder** is a molecule which recognizes a unique amino acid sequence (or molecular structure).	☐	☐	[ADD YOUR REPLY HERE]
A **target** is a unique amino acid sequence (or molecular structure) specifically recognized by a binder molecule.	☐	☐	[ADD YOUR REPLY HERE]
A **cross-reacting entity** is a unique amino acid sequence (or molecular structure) which is recognized by a binder in addition to its designed target	☐	☐	[ADD YOUR REPLY HERE]

Table 2 – Representation of molecule origin

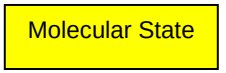
To represent the origin of the molecular species involved in a binding experiment, the class “source” and its subclasses “natural” and “synthetic” were created.



In this context, do you totally agree with the following descriptions?	Yes	No	Corrected or new description (optional)
A **natural source** is any part of micro-organisms, plants, or animals that has been processed, extracted or in whole.	☐	☐	[ADD YOUR REPLY HERE]
A **synthetic source** corresponds to the design and construction of artificial biologically active molecular entities or the re-design of existing, natural molecular entities.	☐	☐	[ADD YOUR REPLY HERE]

Table 3 – Representation of the molecule state

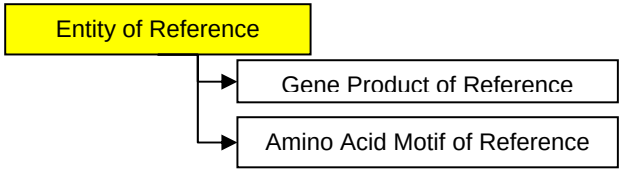
To indicate that a molecule is either native or denatured, folded or unfolded or aggregated, a “Molecular State” class was created.



In this context, do you totally agree with the following descriptions?	Yes	No	Corrected or new description (optional)
The **state** of a molecule is characterized by covalent or non covalent modifications of the native chain tertiary or quaternary structure.	☐	☐	[ADD YOUR REPLY HERE]

Table 4 – Representation of the reference molecules

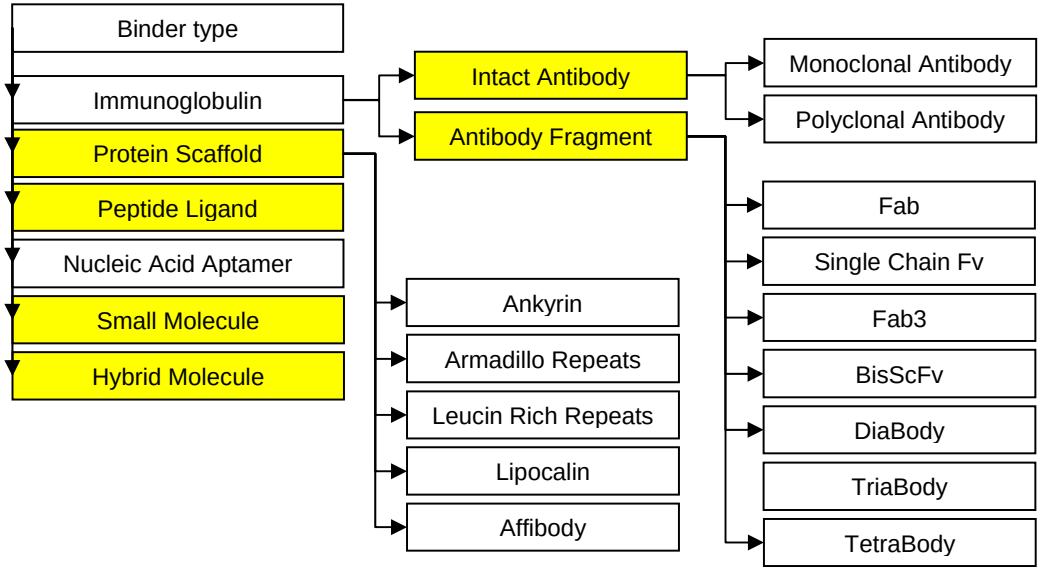
To represent the biological entity of reference which a given molecular species is related to, the class “Entity of reference” and subclasses were created.



In this context, do you totally agree with the following descriptions?	Yes	No	Corrected or new description (optional)
An **entity of reference** is the biological entity (molecule or molecule fragment) which a molecular entity is structurally referenced by or related to.	☐	☐	[ADD YOUR REPLY HERE]

Table 5 – Representation of the different types of binders

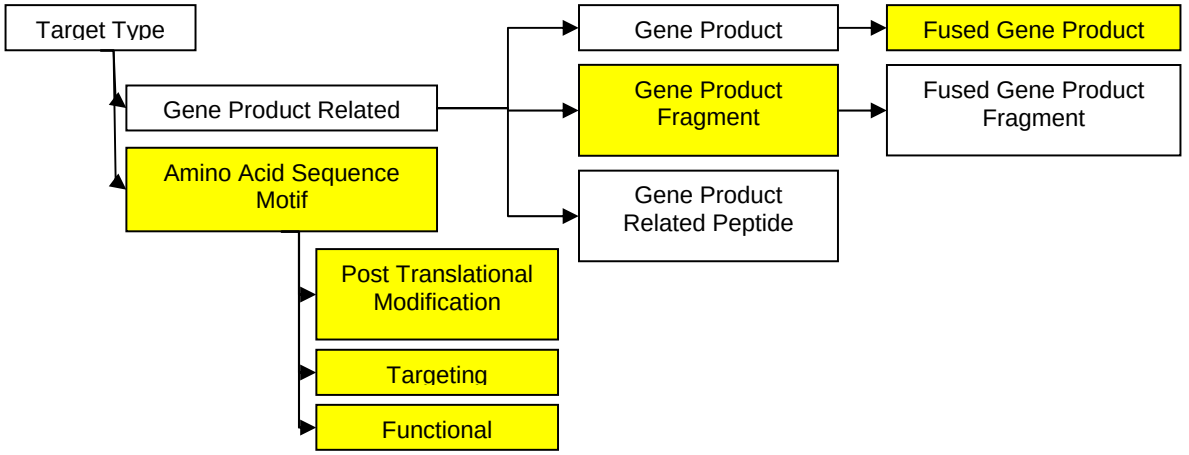
To represent the various types of binders considered, the class “binder type” and a list of subclasses were created. Among these subclasses, the “immunoglobulin” and “protein scaffold” subclasses were further decomposed to describe the various subtypes.



In this context, do you totally agree with the following descriptions?	Yes	No	Corrected or new description (optional)
An **intact antibody** corresponds to a whole unmodified immunoglobulin.	☐	☐	[ADD YOUR REPLY HERE]
An **antibody fragment** is a part of an immunoglobulin molecule or an association of parts of immunoglobulin molecules in a new stable complex.	☐	☐	[ADD YOUR REPLY HERE]
A **protein scaffold** is an engineered molecule based on a protein framework chosen for its specific biophysical properties.	☐	☐	[ADD YOUR REPLY HERE]
A **peptide ligand** is a peptide which binds to a specific target.	☐	☐	[ADD YOUR REPLY HERE]
A **small molecule** is a molecule with a low molecular weight that is not determined by a genome sequence and has been selected for its capacity to recognize target molecules.	☐	☐	[ADD YOUR REPLY HERE]
A **hybrid molecule** is a molecule presenting structure parts corresponding to at least two different biochemical categories (e.g. glycopeptides).	☐	☐	[ADD YOUR REPLY HERE]

Table 6 – Representation of the different types of target

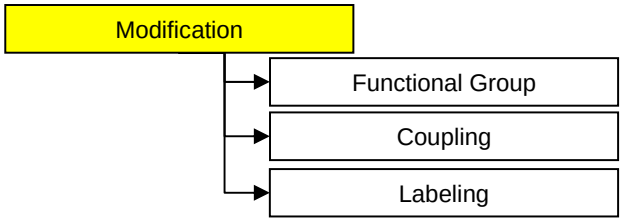
To represent the various types of targets considered, the class “Target type” and a list of subclasses were created. Two main subclasses were created to distinguish targets which are specific of gene product sequences and targets which are specific of consensus sequence motifs (e.g. protein phosphorylation motif), present in many gene product sequences.



In this context, do you totally agree with the following descriptions?	Yes	No	Corrected or new description (optional)
An **amino acid sequence motif** is a highly conserved sequence pattern that has, or is conjectured to have, a biological significance.	☐	☐	[ADD YOUR REPLY HERE]
A **post translational modification** (PTM) motif is an amino acid sequence motif recognized by enzyme(s) for specific chemical modifications of the protein covalent structure. Examples of PTM include glycosylation, acylation, limited proteolysis, phosphorylation, isoprenylation.	☐	☐	[ADD YOUR REPLY HERE]
A **targeting motif** is an amino acid sequence motif recognized for specific translocations of protein to particular membrane-bound subcellular organelles.	☐	☐	[ADD YOUR REPLY HERE]
A **functional motif** is an amino acid sequence motif which confers to the protein its biological activity (e.g. catalytic activity) or specific binding properties.	☐	☐	[ADD YOUR REPLY HERE]
A **gene product fragment** is a protein whose sequence is a contiguous section of a gene product sequence.	☐	☐	[ADD YOUR REPLY HERE]
A **fused gene product** is a protein whose sequence is fused to another amino acid chain, encoded or not by a gene, for instance for labelling or immobilisation purposes.	☐	☐	[ADD YOUR REPLY HERE]

Table 7 – Representation of structure modification

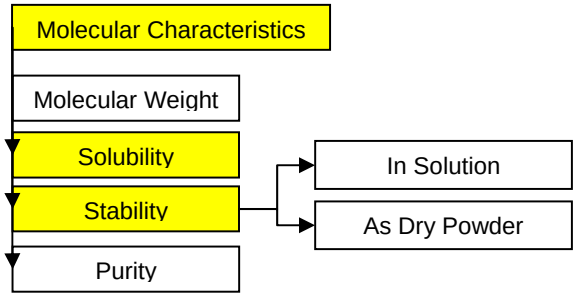
To represent the different modifications which can be performed on a molecular species after its synthesis, a “Modification” class was created.



In this context, do you totally agree with the following descriptions?	Yes	No	Corrected or new description (optional)
A **modification** consists in the alteration of a molecule covalent structure after its synthesis, for instance for labelling or immobilisation purposes.	☐	☐	[ADD YOUR REPLY HERE]

Table 8 – Representation of the molecule characteristics

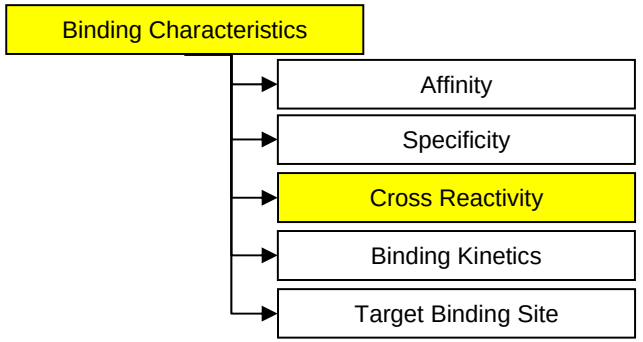
To represent the various characteristics of a molecular entity, a “molecular characteristics” class was created.



In this context, do you totally agree with the following descriptions?	Yes	No	Corrected or new description (optional)
Molecular characteristics are the distinguishing properties of a molecular entity as a molecule.	☐	☐	[ADD YOUR REPLY HERE]
The **solubility** corresponds to the ability of a particular substance to dissolve in a particular solution at a given temperature.	☐	☐	[ADD YOUR REPLY HERE]
The **stability** corresponds to the quality of being free from change or variation or not taking part readily in chemical change.	☐	☐	[ADD YOUR REPLY HERE]

Table 9 – Representation of the characteristics of the binder

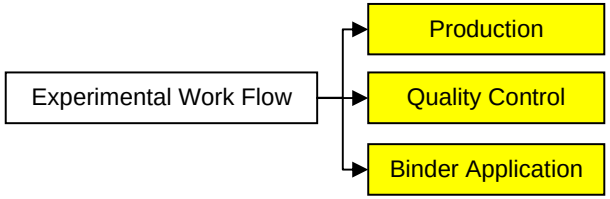
To represent the protein binding characteristics of a binder, a “binding characteristics” class was created.



In this context, do you totally agree with the following descriptions?	Yes	No	Corrected or new statement (optional)
Binding characteristics are the distinguishing properties which enable a binder to recognize its designated target.	☐	☐	[ADD YOUR REPLY HERE]
The **cross reactivity** corresponds to the observation that a binder binds to characterized or uncharacterised molecular entities whose molecular characteristics differ from its designated target.	☐	☐	[ADD YOUR REPLY HERE]

Table 10 – Representation of experiment series

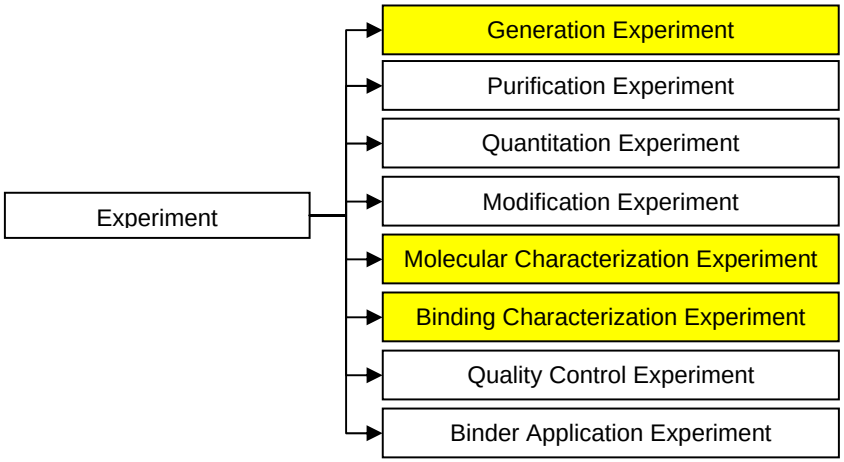
To represent sets of experiments which are performed in series in order to achieve particular goal, an “experimental work flow” class was created.



In this context, do you totally agree with the following descriptions?	Yes	No	Corrected or new statement (optional)
The **production** is the workflow which leads to the availability of a specific molecular entity.	☐	☐	[ADD YOUR REPLY HERE]
The **quality control** is the workflow which leads to the quality assessment of a binder ability to recognize its designated target.	☐	☐	[ADD YOUR REPLY HERE]
A **binder application** is a workflow which implies the use of a particular binder.	☐	☐	[ADD YOUR REPLY HERE]

Table 11 – Representation of the different experiments

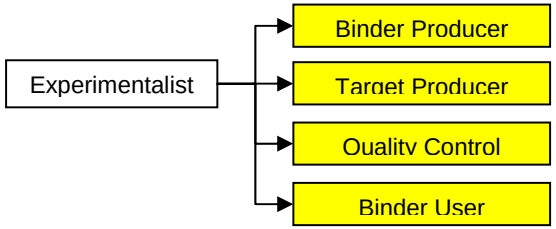
To represent different categories of experiments, an “experiment” class was created.



In this context, do you totally agree with the following descriptions?	Yes	No	Corrected or new statement (optional)
A **generation experiment** consists in a series of steps intended to synthesize a specific molecular entity.	☐	☐	[ADD YOUR REPLY HERE]
A **molecular characterization experiment** consists in a series of steps intended to measure or estimate molecular characteristics of molecular entities (e.g. molecular weight).	☐	☐	[ADD YOUR REPLY HERE]
A **binding characterization experiment** consists in a series of steps intended to measure or estimate binding characteristics of a binder.	☐	☐	[ADD YOUR REPLY HERE]

Table 12 – Representation of the data producers

To represent the different categories of persons who perform experiments, an “experimentalist” class was created.



In this context, do you totally agree with the following descriptions?	Yes	No	Corrected or new description (optional)
A **binder producer** is the person who produces a binder through a production experimental workflow.	☐	☐	[ADD YOUR REPLY HERE]
A **target producer** is the person who produces a target through a production experimental workflow.	☐	☐	[ADD YOUR REPLY HERE]
A **quality control center** is the person who assesses experimentally the quality of a binder towards its target through a quality control workflow.	☐	☐	[ADD YOUR REPLY HERE]
A **binder user** is a person who uses the binder in a binder application experiment.	☐	☐	[ADD YOUR REPLY HERE]

Any comment that you wish to make about the Ligand-Binder Ontology

- **In general**

[ADD YOUR REPLY HERE]

- **About classes of concept**

[ADD YOUR REPLY HERE]

- **About classes and subclasses organization**

[ADD YOUR REPLY HERE]

Thank you very much for your collaboration!