IPSJ Transactions on Bioinformatics Vol.6 9–17 (May 2013)

[DOI: 10.2197/ipsjtbio.6.9]

Original Paper

Protein Experimental Information Management System (PREIMS) Based on Ontology: Development and Applications

Junko Sato[1] Kouji Kozaki[2] Susumu Handa[3] Takashi Ikeda[3] Ryotaro Saka[3]
Kohei Tomizuka[3] Yugo Nishiyama[3] Toshiyuki Okumura[4] Shinichi Hirai[4]
Tadashi Ohno[4] Mamoru Ohta[5] Susumu Date[6] Haruki Nakamura[1,a)]

Received: November 19, 2012, Accepted: January 23, 2013, Released: May 28, 2013

Abstract: We developed a new information management system, *Protein Experimental Information Management System* (*PREIMS*), which has the ontology-based functions for quality control, validation, scalability, and information sharing. Its contents are mainly experimental protocols for the analyses of protein structures and functions, and their results. They are stored separately in the PREIMS database (DB), as the ontology based protocol data and the result data. The synchrotron experimental information was stored as the latter result data in Extensible Markup Language (XML). Furthermore we converted those protocols in the format of Resource Description Framework (RDF) for integration with other biological information resources.

Keywords: ontology, experimental protocol, RDF, HOZO

1. Introduction

In the research field of life science, management of "big-data" is now a crucial issue, because high-throughput technologies in the post-genomic era have been producing a huge amount of experimental data [1]: Genome sequences in DDBJ [2], as a member of the international organization, INSDC [3], protein structures in PDBj [4], as a member of international organization, wwPDB [5], Microarray gene expression information such as GEO [6], and any other omics data.

During the previous structural genomics programs, PSI [7] in USA and Protein 3000 [8] in Japan, some information management systems were developed so as to distribute and share the experimental information produced in the programs. However, in those management systems, the majority of the protocols involved in protein structural analysis were mainly described in the plain text style. Likewise, the words or terms in those protocols were not well organized or not well standardized. It is clear that those situations prevented us from reusing the information efficiently.

After Protein 3000 program, another program named Targeted Proteins Research Program (TPRP) [9] started in Japan. Its mission was to reveal the structures and functions of biologically and medically important proteins. However, in most cases, there were many difficulties in protein preparation, purification, and crystallization for structural analysis, and researchers had to develop their own new experimental methods and protocols to overcome the difficulties. Therefore, those new protocols should be recorded and accumulated in a database to be reused and shared within the wide research society. In addition, efficient management of the image data of X-ray diffraction from protein crystals collected by synchrotron diffraction experiments is crucial for high-throughput structure analysis. However, an archive having only the image data is useless without precise experimental conditions with the protocols, which should be in machine readable style.

On the basis of these backgrounds, we aim to construct a new PRotein Experimental Information Management System named PREIMS, so as to share the various experimental protocols and the results by synchrotron diffraction experiments in machine readable style. In order to maintain quality of data and promote reuse of them, data validation is necessarily required in PREIMS. Entities of experimental protocols should also be standardized. For these purposes, we conceptualized the experimental information based on ontology which provides us explicit terminology and a methodology to describe experimental data consistently. In this paper, we discuss construction of ontology for experimental protocols with implementation of PREIMS. The Resource Description Framework (RDF) format for the semantic web applications is also incorporated.

2. Construction of Ontology

In order to share protein experimental information on the Web,

1 Institute for Protein Research, Osaka University, Suita, Osaka 565–0871, Japan
2 Institute of Scientific and Industrial Research, Osaka University, Ibaraki, Osaka 567–0047, Japan
3 NEC Informatec Systems, Ltd., Kawasaki, Kanagawa 211–8666, Japan
4 Mitsui Knowledge Industry Co., Ltd., Minato, Tokyo 105–6215, Japan
5 Enegate Co., Ltd., Kita Osaka 531–0077, Japan
6 Cybermedia Center, Osaka University, Ibaraki, Osaka 567–0047, Japan
a) harukin@protein.osaka-u.ac.jp

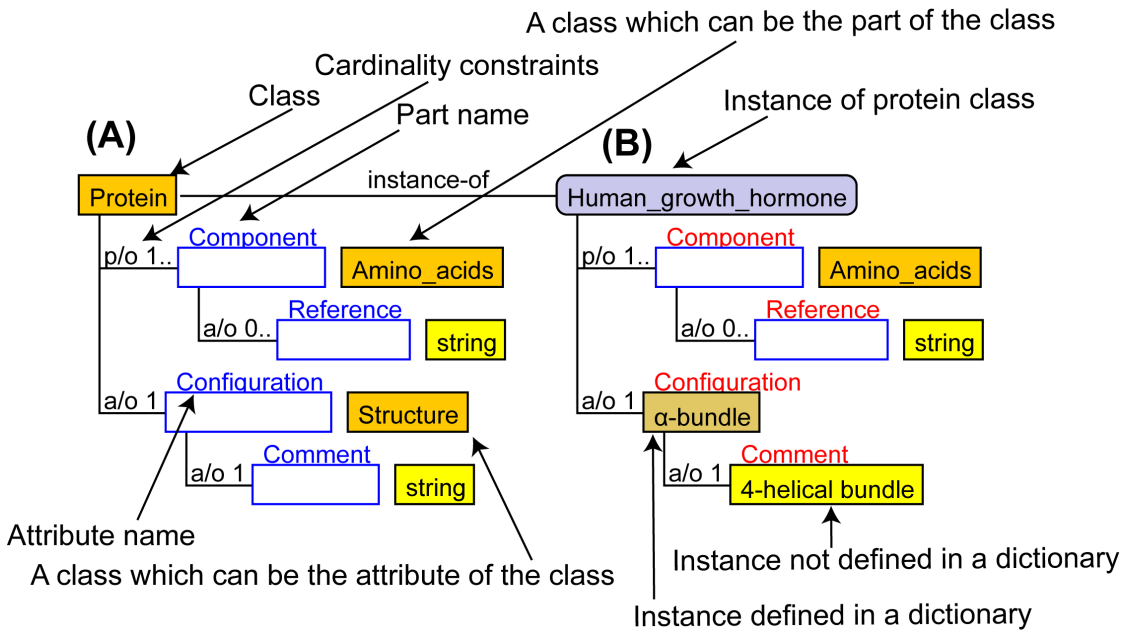

Fig. 1 (A) Hozo representation [10] of the Class *Protein*, and (B) its instance examples. In this example, a *Protein* is composed of one or more *Amino_acids* and has a unique *Structure*. In Hozo, each orange box shows the concept class, and any class is defined using two types of relations: "part-of" relation indicated by "p/o" and "attribute-of" relation indicated by "a/o." In the case in (A), a part of *Protein* is a "Component," which corresponds to a particular composition of *Amino_acids*, and an attribute of *Protein* is a "Configuration," which corresponds to a particular shape of *Structure*. The number next to "p/o" or "a/o" shows the Cardinality constraints. It represents the number of instances which can be the part/attribute of instances of the class. For example, "1" means "exactly one" and "1.." means "one or more than 1." A relationship between class and its instance is represented by an "instance-of" relation. Since any instances cannot have subclasses, "is-a" relationship is not appeared between instances. In (B), "Human_growth_hormone" is an instance of the class *Protein*. An instance of its configuration is "α-bundle" defined in the dictionary, which is shown in a brown box. When the instance is not defined in the dictionary, as "4-helical bundle" in the current case, it is indicated in a yellow box. "string" in a yellow box represents that its instance is not defined in the dictionary and described in natural language (words or sentences) for supplemental explanation.

it requires machine readable data formats with consistent vocabulary. Although the World Wide Web is mainly based on documents written in Hypertext Markup Language (HTML), it is only able to represent structures of documents. That is, HTML itself has no capacity to express the meanings of documents unambiguously. In order to represent meaning of documents, technologies of Semantic Web, such as Resource Description framework (RDF), Web Ontology Language (OWL) are widely used in many domains. The basic idea of Semantic Web is to add machine readable meta-data on web documents. Since the meta-data provide descriptions about meanings of Web documents, the contents themselves could be descriptive data stored in Web-accessible databases.

Here we took advantage of ontology, which is one of traditional terminology-based approaches. Ontology provides common vocabularies for representing and sharing knowledge about various topics and a set of relationships between them. In order to construct ontology of protein experimental information target, most of the experimental information given by the TPRP [9] was conceptualized, and representation scheme to describe the information was built in frame structures. Consequently, the new ontology was named PREIMSOntology (PREIMSOn). PREIMSOn was constructed in an ontology editor, HOZO [10] by using three built-in relations: subclass-of (denoted *is-a*), part-of (denoted p/o) and attribute-of (denoted a/o) relations. The representation by HOZO is briefly explained in **Fig. 1**. The details of PREIMSOn are discussed in the followings sections.

3. PREIMSOntology (PREIMSOn)

In general, materials in protein experiments are prepared and operated in the most assured and reliable method. An experimental method shows what should be prepared and how researchers should operate. That is, a protocol can be represented by a sequence of operations on target materials. Though a protocol consists of many operations, it can be divided into several sets of operation. Therefore, the experimental method class (*Protocol* class), the experimental operation class (*Action* class), and the experimental material class (*Object* class) are considered as the top level classes of PREIMSOn. The common requirement for the all classes defined in PREIMSOn is that they should be formalized so that researchers can reuse them by browsing and querying protein experimental information. The three top level classes and unique terms necessary to understand and manage PREIMS are described in details below.

3.1 *Protocol* Class

Protocol class is defined as the experimental specification which describes successive experimental procedures and exper-

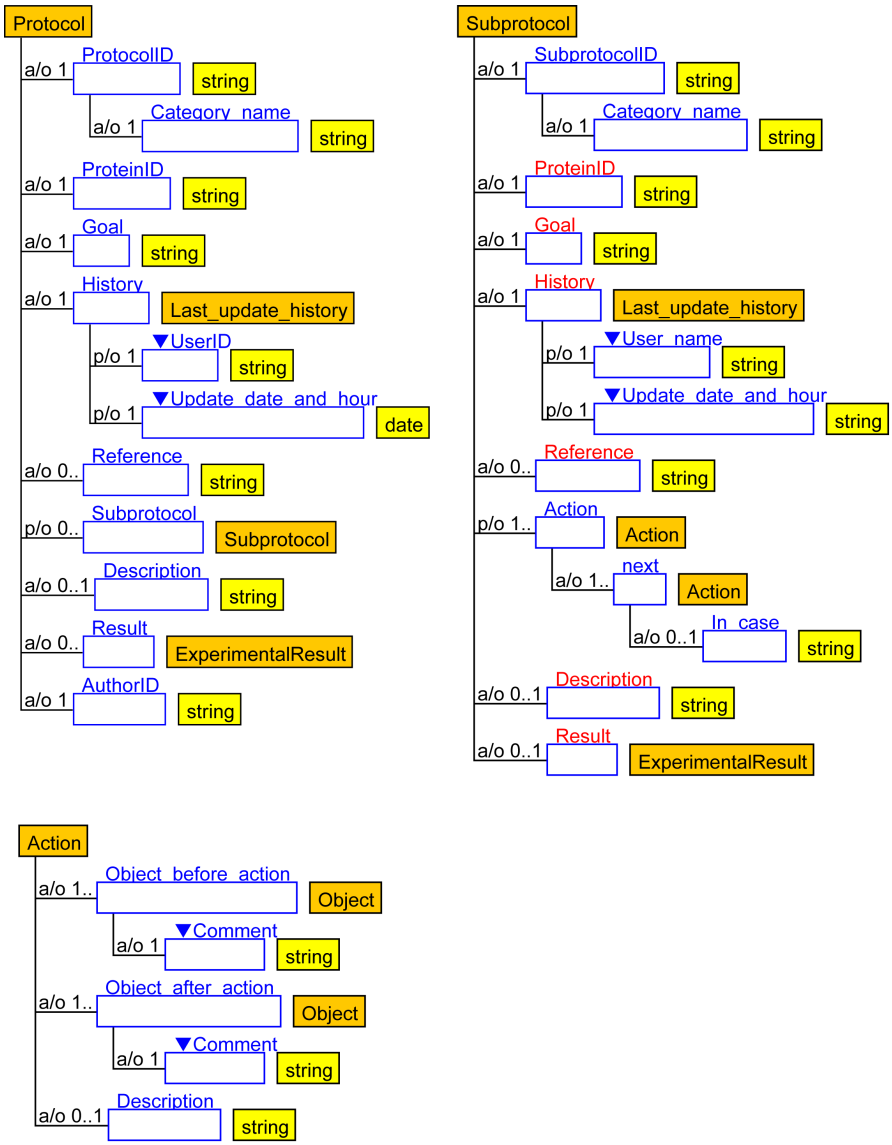

Fig. 2 Class definition of *Protocol*, *Sub-protocol*, and *Action* classes in PREIMSOntology. Each of these classes is defined using basic and mandatory classes located at lower hierarchy with the "a/o" or "p/o" relation. Since a *Protocol* is composed of one or more *Sub-protocols*, *Protocol* class should have the p/o relation with *Sub-protocol* class, and its cardinality constraints is "1..". The *Protocol* class is also characterized by a set of mandatory attributes, for example, "Protocol ID," "Category," "Goal" and so on. The attributes of *Protocol* class, "ProteinID," "Goal," "History," "Reference," "Description" and "Result," are inherited to those of the lower "*Sub-protocol* class."

imental materials which is necessary for the experiment. It has a set of essential attributes such as goal, Protein ID, category, update date and time and so on. Although a protocol can be represented as a sequence of operation, some combinations of operations are commonly observed in several protocols. Such combinations of operations are defined as *Sub-protocol* classes. Since the *Sub-protocol* class is defined as a sub-class of the *Protocol* class, it inherits all attributes from the *Protocol* class other than Sub-protocol ID. A *Sub-protocol* class is consists of more than one *Action* class (operation) and a *Protocol* class is composed of more than one *Sub-protocols*. That is, a *Protocol* class has several *part-of* relations with *Sub-protocol* classes as shown in **Fig. 2**. According to any protein experiments, template models of the *Sub-protocol* class are designed and instantiated by appropriate experimental data values. It enables us well organized definition of protocols. Considering these requirements, the *Protocol* class and the *Sub-protocol* class are defined as shown in Fig. 2. Experimental protocols are described as instances (instantiated models) of classes defined in PREIMSOn. Relationships between a class and its instances are represented by "instance-of" relation as shown in Fig. 1.

3.2 *Action* Class

Action class is defined as the simplest experimental specification, which shows how experimentalists operate experimental objects towards the experimental goal (Fig. 2). An Action is conceptualized according to a change of target object from before the action to after it. So any *Action* class has two kinds of *attribute* relations with *Object* classes before and after the experimental operation. "is-a" hierarchy of *Action* classes is based on the change between before/after *objects* (**Fig. 3**). Thanks to these attributes, compositions of more than one *Action* classes such as

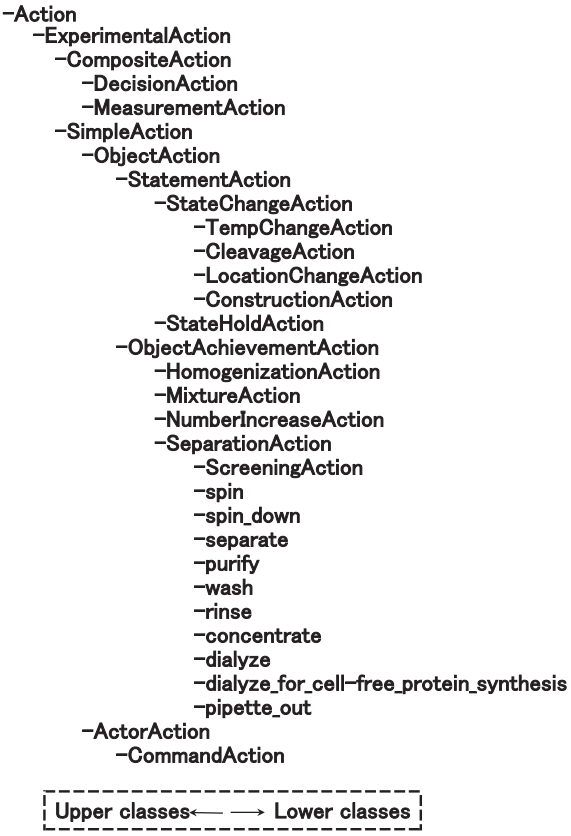

Fig. 3 An "is-a" hierarchy example of the *Action* class. An "is-a" hierarchy of *Action* classes is constructed based on the change between the *object*s before and after the *Action*. In an example of the *Separation action* class, the change between the *objects* before and after the *Action* is a separation caused by a particular experimental operation with instruments, such as "spin," "purify," "dialyze," and so on.

branched or combined experimental procedures.

In addition to the above definition, an *Action* class has *attribute-of* relations with some necessary parameters such as equipment, experimental instrument, and so on as well as two types of experimental objects. In the case of the successive actions, it stands to reason that *object after action* of the prior action shall be equal to *object before action* of the later one. An example of the *Action* class, spin, describing the centrifugation operation is shown in **Fig. 4**.

3.3 *Object* Class

Object class denotes all experimental materials which are used in protein experimental procedures. Their names are used as experimental values in *Sub-protocol* instances. The class hierarchy of *Object* class according to *is-a* relation is shown in **Fig. 5**.

3.4 *PREIMS Library*

PREIMS Library is a collection of experimental *Sub-protocol* models and the instances in PREIMS server.

3.5 *PREIMS Dictionary*

PREIMS Dictionary is a glossary of experimental information standardized based on PREIMSOn. To upload protocols onto the PREIMS server, they have to be described using PREIMS Dictionary.

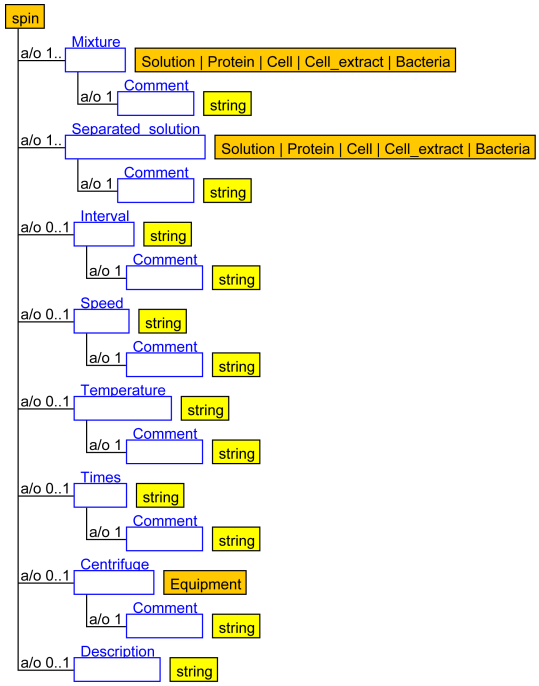

Fig. 4 Example of *Action* class, *spin,* which is constructed following the Class definition of *Action*. "Mixture," "Separated solution," and "Description" are inherited from their upper class *Action*, which is "SeparationAction" as shown in Fig. 3. Another five necessary classes are added for detailed representation of the experimental procedure, the names of which are "Interval," "Speed," "Temperature," "Times" and "Centrifuge," in this case. In this example, "|" means "or." That is, "Solution |Protein |Cell |Cell_extract |Bacteria" means Solution, Protein, Cell, Cell_extract or Bacteria can be Mixture attributes of spin.

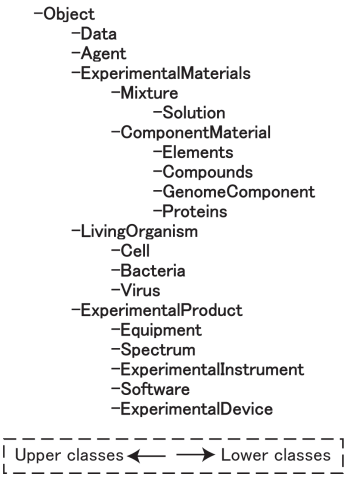

Fig. 5 An "is-a" hierarchy of the *Object* class. Sub-classes of the *Object* class are defined using the "is-a" relation.

4. Curation of Protein Experimental Protocols

There are many kinds of experimental information in the protocol description. They should be standardized based on PREIMSOn before they are uploaded to PREIMS server [11] since they are not machine readable while experimental protocols described in text style are human readable. Such machine readable information requires comprehensible and clear curation, especially when *Sub-protocol* models are created according to the *protocol* description.

Here, we show a curation example for a protocol of protein

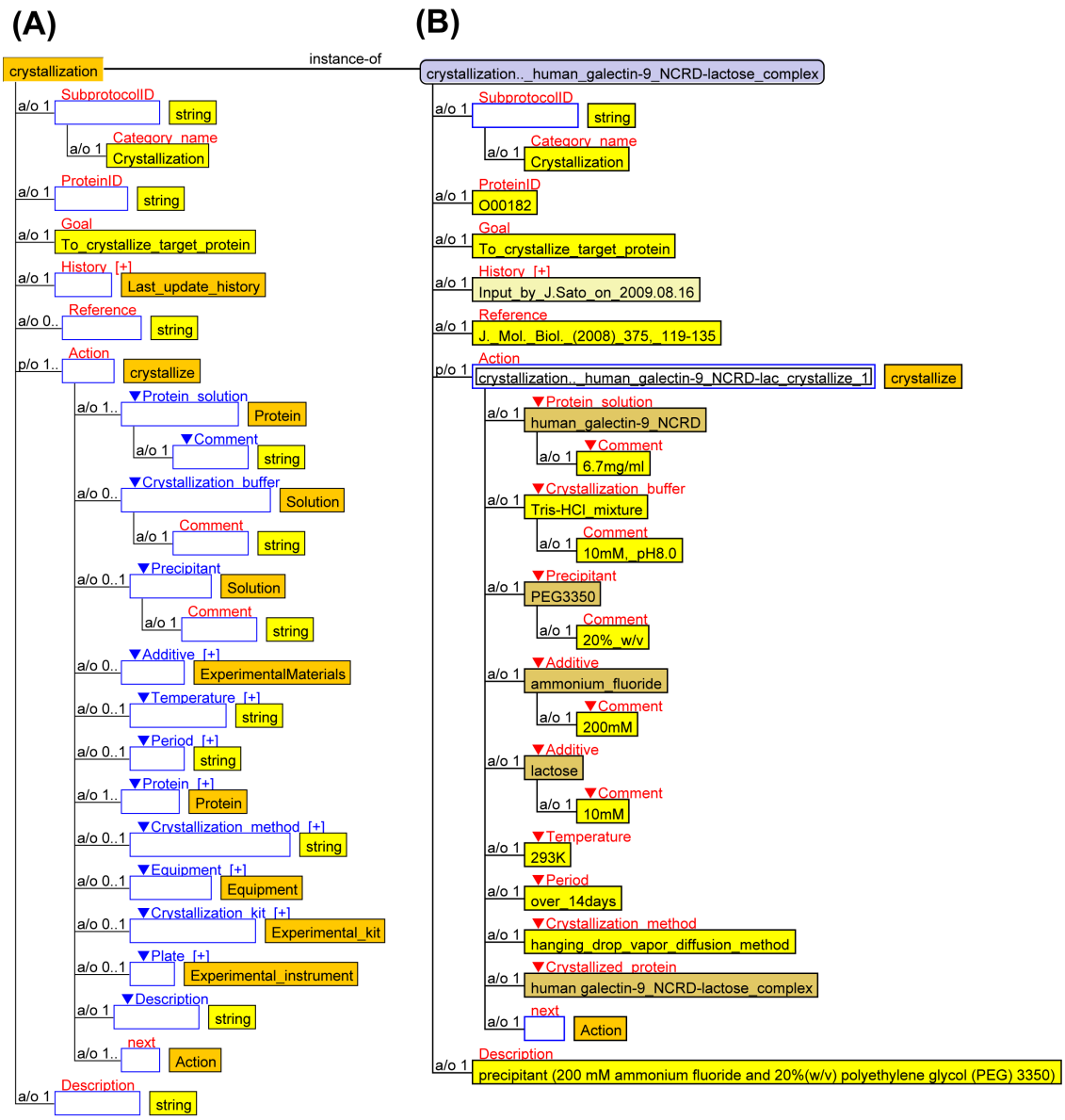

Fig. 6 *Sub-protocol* model and the instances of protein crystallization. Left slot (A) is the *Sub-protocol* model *crystallization*, which is defined to represent the experimental objects and optimized conditions of protein crystallization using *Action* class *crystallize*. As a *Sub-protocol* class, "*crystallization"* model has several lower classes. For example, the *object before action* class is inherited as the *Protein solution* name, and the *object after action* class is inherited as the *Protein* name. Moreover, "temperature," "crystallization buffer," "additive" and so on, are newly added as the mandatory attributes. Right slot (B) shows the instances of the *Sub-protocol crystallization*, in which experimental data values are assigned to the attribute parameters of the *Action* class *crystallize*. For example, "crystallization..._human_galectin_9_NCRD-lac_crystallize_1," which is located next to the *crystallize* orange box, specifies the *crystallize* action, as the first action of the instance *crystallization..._human_galectin-9_NCRD-lactose_complex*. These slots are designed by the Ontology editor *HOZO* [10].

crystallization. The experimental information below was contributed by members of TPRP. Its text description is "Optimization of crystallization conditions was performed manually by the hanging drop vapor diffusion method. Crystals of the human galectin-9 NCRD-lactose complex were grown from drops containing equal volumes of 6.7 mg/ml protein in sample buffer (10 mM Tris-HCl (pH 8.0), 100 mM NaCl, and 1 mM DTT) containing 10 mM lactose, and precipitant (200 mM ammonium fluoride and 20%(w/v) polyethylene glycol (PEG) 3350) at 293 K over 14 days [12]".

This experiment is one of the protein crystallization techniques, known as the hanging drop vapor diffusion method. First, *Sub-protocol* models were defined so as to describe the protocol of protein crystallization along to PREIMSOn with the experimental objects and the optimized conditions which were written above in the text style. We built a *Sub-protocol* model to be allowed to represent them of which the name is *crystallization.* The experimental information was represented as the corresponding attributes of Action *crystallize* class. The metadata are also added to them. For example, protein solution, buffer, precipitant, additive, temperature, experimental instrument, and so on (**Fig. 6**-A blue terms) were added as the

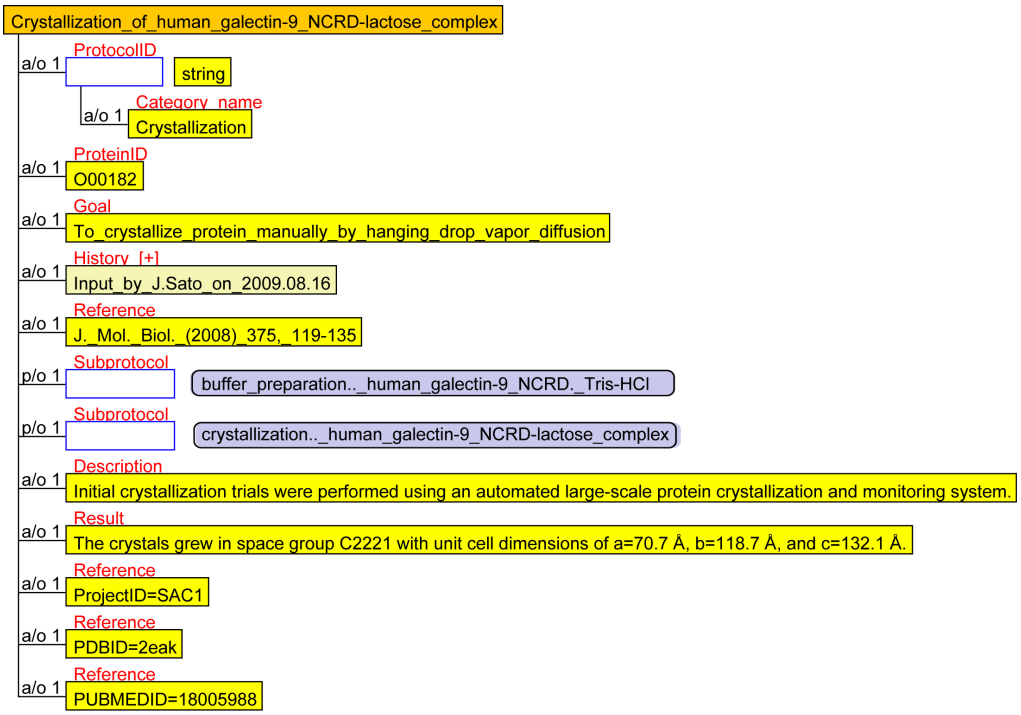

Fig. 7 An example of *Protocol* class composed of one more *Sub-protocol* instances. The *Protocol Crystallization of human galectin-9 NCRD-lactose complex* has two *Sub-protocol* classes, *buffer_preparation_1.._human_galectin-9_NCRD-lactose_complex*, and *crystallization.._human_galectin-9_NCRD-lactose_complex*. Other experimental information is added as the attributes of a *Protocol* class.

metadata. Finally, each value of the experimental data was assigned as the instance to the right attribute parameter, as shown in Fig. 6-B. The *Sub-protocol* instance is named *crystallization.._human_galectin-9_NCRD-lactose_complex*. We made a general rule to name *Sub-protocol* instances as follows: "*Sub-protocol name.._target_protein name*." All experimental data values were defined in PREIMS dictionary. *Comment* classes were used for the quantitative or qualitative supplementary explanation.

Furthermore, in order to finish up the crystallization protocol (**Fig. 7**), we built another *Sub-protocol* instance, which is named *buffer_preparation_1.._human_galectin-9_NCRD-lactose_complex*. Necessary experimental information for preparation of the crystallization buffer is set as the attribute of the *Action* class *mix* and annotated in the *Sub-protocol* model *buffer_preparation_1*, for example, added reagent, pH, concentration, temperature and so on. (**Fig. 8**-A and B)

5. Implementation of PREIMS

PREIMS exploited an eXtensible Markup Language (XML)-database which stores the experimental protocols and their results information. The protocol information is formalized using the ontology editor Hozo [9], and then it is converted to data in the XML format for PREIMS so that the information is managed as a machine readable meta-data. As described above, PREIMSOn consists of *PREIMS Library* and *PREIMS Dictionary*. They are described in the XML format for Hozo. Since PREIMSOn provides class schemas for instances, it is used to validate instances of experimental protocol information in the XML format. **Figure 9** shows the validation procedure in PREIMS.

PREIMS also implemented a database for experimental results data in order to store results information collected by synchrotron diffraction experiments. Experimental log-data file provided by the diffractometers of SPring-8 and Photon Factory is converted into those in the XML format for PREIMS using the ontology of each synchrotron diffraction experimental protocols. Then, those results data with experimental parameters were stored in PREIMS as the results of synchrotron diffraction experiments. The system was made fully secure with the user's ID of SPring-8 and Photon Factory so that every user can access and browse only his/her own data.

By accessing PREIMS Linux portal server [11] controlled by PostgreSQL, any users can query, browse, and download protocol information which is open and public. For integration with the linked data world, the protocol and sub-protocol information are provided also in RDF format (*http://preims.pdbj.org/preims/rdf/*) for the semantic web applications. On the contrary, only the privilege members are able to query, browse, download and upload the information of private experimental protocols and their results, since PREIMS portal is also designed so that the access could be restricted with user authentication.

6. Conclusion

During development of PREIMS, the validation function of the contents was designed as the most important one because only such validated experimental information should be shared and reused by any other researchers in principle. Thus, PREIMSOn was constructed using the ontology editor *HOZO* [10], and the experimental information was formalized based on PREIMSOn. At the same time, *PREIMS dictionary* and *PREIMS library* were

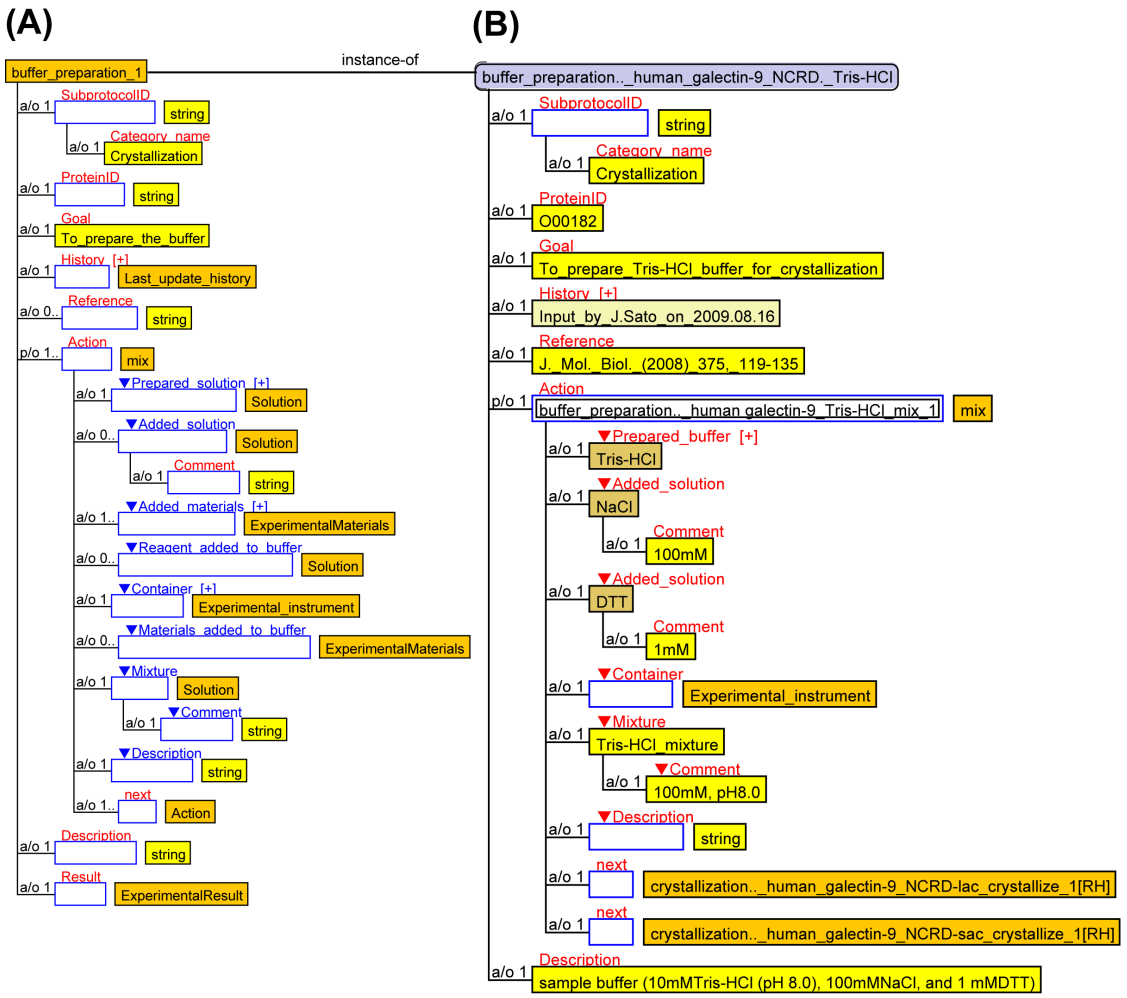

Fig. 8 An example of *Sub-protocol* model to prepare a crystallization buffer (A) and the corresponding *Sub-protocol* instances. Left slot (A) is the *Sub-protocol* model *buffer_preparation,* which is defined to represent necessary experimental information (added reagent, pH, concentration, temperature and so on). Right slot (B) is the *Sub-protocol* instance *buffer_preparation_1..._human_galectin-9_NCRD-lactose_complex.* Here, the prepared buffer "Tris-HCl" is assigned to the *Preparation buffer* name as the *object before action*. The other solutions, "NaCl" and "DTT," correspond to the *added solution* name "Tris-HCl mixture," which is assigned as the *object after action* used in the next sub-protocol, "*crystallization..._human_galectin-9_NCRD-lactose_complex.*" That is indicated by the *next* action, "*crystallization..._human_gal-9_NCRD-lac_crystallize_1*."

developed, and they were utilized to validate experimental protocols based on PREIMSOn. Thus, consistency in PREIMS is kept based on the ontology, and the database with a machine readable format was developed for experimental protocols of various analyses of protein structures and functions. Consequently, the excellent new protocols can be widely reused and shared by the protein researchers. Furthermore, providing the database also in the standard format for the Semantic Web, RDF format data could enhance the integration of the PREIMS with other bioinformatics resources as one of Linked open data, as well as PDB/RDF developed at PDBj [4]. Thus, integrated queries become possible by crossing multiple different databases, for example, a specific crystallization method in PREIMS/RDF format can be searched by linking to particular protein structures in PDB/RDF, which is now commonly used in the protein sequence database, UniProt [13].

Acknowledgments This work was supported by the Targeted Proteins Research Program (TPRP) from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan to S.Y. S.M. is a research fellow of the Japan Society for the Promotion of Science (JSPS). We would like to thank all the TPRP members, who contribute not only to give new procol information, but also to improve PREIMS and to construct PREIMSOn.

References

[1] Mervis, J.: U.S. Science, Policy: Agencies Rally to Tackle Bid Data, *Science*, Vol.336, p.22 (2012).
[2] Kodama, Y., Mashima, J., Kaminuma, E., Gojobori, T., Ogasawara, O., Takagi, T., Okubo, K. and Nakamura, Y.: The DNA Data Bank of Japan launches a new resource, the DDBJ Omics Archive of functional genomics experiments, *Nucl. Acids Res.*, Vol.40, pp.D38–D42 (2012), available from ⟨http://www.ddbj.nig.ac.jp/⟩.
[3] Karsch-Mizrachi, I., Nakamura, Y. and Cochrane, G.: The International Nucleotide Sequence Database Collaboration, *Nucl. Acids Res.*, Vol.40, pp.D33–D37 (2012).
[4] Kinjo, A.R., Suzuki, H., Yamashita, R., Ikegawa, Y., Kudou, T., Igarashi, R., Kengaku, Y., Cho, H., Standley, D.M., Nakagawa, A. and Nakamura, H.: Protein Data Bank Japan (PDBj): Maintaining a structural data archive and resource description framework format, *Nucl. Acids Res.*, Vol.40, pp.D453–D460 (2011). PDBj, available from ⟨http://pdbj.org/⟩.
[5] Berman, H., Henrick, K., Nakamura, H. and Markley, J.L.: The world-

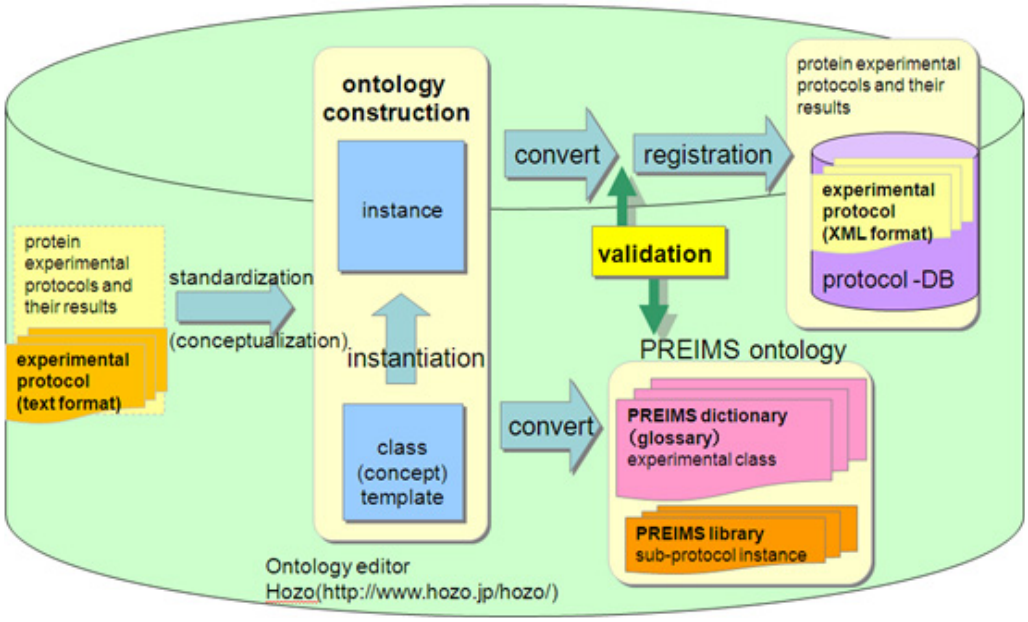

Fig. 9 Flow of the protein experimental protocols in PREIMS. A protein experimental protocol in the text format is converted to the ontology-based one in the XML format. After mapped to the PREIMSOn-based schema, it is registered to the protocol database in PREIMS.

wide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data, *Nucl. Acids Res.*, Vol.35, pp.D301–D303 (2007).

[6] Barrett, T., Troup, D.B., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Muertter, R.N., Holko, M., Ayanbule, O., Yefanov, A. and Soboleva, A.: NCBI GEO: Archive for functional genomics data sets — 10 years on, *Nucl. Acids Res.*, Vol.39, pp.D1005–D1010 (2011).

[7] Chandonia, J.M., Kim, S.H. and Brenner, S.E.: Target selection and deselection at the Berkeley Structural Genomics Center, *Proteins*, Vol.62, pp.356–370 (2006). PepcDB, available from ⟨http://pepcdb-dev.rutgers.edu/⟩.

[8] Standley, D.M., Toh, H. and Nakamura, H.: Functional annotation by sequence-weighted structure alignments: Statistical analysis and case studies from the Protein 3000 structural genomics project in Japan, *Proteins*, Vol.72, pp.1333–1351 (2008).

[9] Targeted Proteins Research Program (TPRP), available from ⟨http://tanpaku.org/e_index.php⟩.

[10] Mizoguchi, R., Sunagawa, E., Kozaki, K. and Kitamura, Y.: A Model of Roles within an Ontology Development Tool: Hozo, *J. Appl. Ontology*, Vol.2, pp.159–179 (2007), available from ⟨http://www.hozo.jp/⟩

[11] PREIMS server, available from ⟨http://preims.pdbj.org/preims/index_e.jsp⟩.

[12] Nagae, M., Nishi, N., Nakamura-Tsuruta, S., Hirabayashi, J., Wakatsuki, S. and Kato, R.: Carbohydrate Recognition Domain Reveals Unexpected Properties that Differ from the Mouse Orthologue, *J. Mol. Biol.*, Vol.375, pp.119–135 (2008).

[13] Redaschi, N. and the UniProt Consortium: UniProt in RDF: Tackling Data Integration and Distributed Annotation with the Semantic Web, *Nature Prec.* (2009), DOI: 10.1038/npre.2009.3193.1: ftp://ftp.uniprot.org/pub/databases/uniprot/current_release/rdf/.

Junko Sato is an annotator of the Protein Data Bank Japan. She graduated from the Department of Macromolecular Science, School of Science, Osaka University in 1981. From 2007 to 2012 she took part in the development of PRotein Experimental Information Management System (PREIMS, http://preims.pdbj.org/preims/).

Kouji Kozaki is Associate Professor at the Department of Knowledge Systems, the Institute of Scientific and Industrial Research, Osaka University. He received his Bachelor, Master and Doctoral degrees of Engineering in 1997, 1999 and 2002 from Osaka University, respectively. His research interest includes ontology building, fundamental theory of ontological engineering, ontology development system, ontology-based application. He is a member of JSAI and IEICE.

Susumu Handa is currently a manager of image sensing group at NEC Informatec Systems, Ltd. He worked on developing scientific visualization and high performance computing systems especially related to biomolecular science. He received his Ph.D. degree from Kyushu Institute of Technology in 2001.

Takashi Ikeda has been working for NEC Informatec Systems, Ltd. and currently working as system engineer for the ground system in a satellite. He received his bachelor degree from Toyama University in 1986.

Ryotaro Saka has been working for NEC Informatec Systems, Ltd. and developing software systems for high performance computing related to biological research fields. His current interests include developing software for large scale computing.

Kohei Tomizuka received his B.E. degree in Information Science from Toho University in 1997, and joined NEC Informatec Systems, Ltd. He is currently working as an engineer in bioinformatics, and his research interests include big data analysis derived from Next Generation Sequencing.

Yugo Nishiyama has been working for NEC Informatec Systems, Ltd. and developing software systems. He received his M.E. degree from Kyushu Institute of Technology in 2003. His current interests include developing software for image recognition.

Toshiyuki Okumura is a chief manager of Bioscience department at Mitsui Knowledge Industry Co., Ltd. He received his Doctoral degree of Informtion Science in 2006 from Osaka University. His research interests include computational analysis of genomic data, comparative genomics and gene evolution. He is a member of JSBi and SESJ.

Shinichi Hirai is a system architect of Bioscience department at Mitsui Knowledge Industry Co., Ltd. He received his master's degree of Genetic Resources Technology in 1998 from Kyushu University. His research interests include database of lifescience, sales force automation of pharmaceutical drug.

Tadashi Ohno is a system architect of Customer solution department at Mitsui Knowledge Industry Co., Ltd. He received his master's degree of science and technology in 2002 from Tokyo University of Science. His research interests include core system of MBK, internet information technology.

Mamoru Ohta received his B.E. and M.E. degrees from Kyoto Institute of Technology in 1993 and 1995, and then Ph.D. degree from Osaka University in 2011, respectively. He has been working for Enegate Co., Ltd. since 1995. He is interested in knowledge management systems, ontology engineering tools and so on.

Susumu Date received his B.E., M.E., and Ph.D. from Osaka University in 1997, 2000, and 2002, respectively. He was an Assistant Professor at the Graduate School of Information Science and Technology, Osaka University from 2002 to 2005. He also worked as a Visiting Scholar in University of California, San Diego in 2005. He worked as a Specially-appointed Associate Professor for the Internationalization of Education in the Graduate School of Information Science and Technology, Osaka University through the MEXT-funded educational program from 2005 to 2008. From 2008 he is working as an Associate Professor of the Cybermedia Center at Osaka University. His current research interests include application of Grid computing and related information technologies. He is a member of IEEE and IPSJ.

Haruki Nakamura is a Professor of Laboratory of Protein Informatics, Institute for Protein Research, Osaka University. He is also a Head of Protein Data Bank Japan (PDBj, http://pdbj.org/), one of the four members of the wwPDB (http://wwpdb.org/) as the international organization to look after PDB. He was born on 7 April 1952, in Tokyo, Japan. He graduated from the Department of Physics, Faculty of Science, the University of Tokyo, in March 1975, and took the Doctor of Science in March 1980, there. His research fields are structural bioinformatics, biophysical studies about protein architecture, electrostatic properties and enzymatic functions, protein modeling, protein design, structure guided drug development, and molecular and electronic simulations.

(Communicated by *Toshinori Endo*)