A Community Standard Format for the Representation of Protein Affinity Reagents*[S]

David E. Gloriam,[a,b,c] Sandra Orchard,[a] Daniela Bertinetti,[d] Erik Björling,[e] Erik Bongcam-Rudloff,[f] Carl A. K. Borrebaeck,[y] Julie Bourbeillon,[g,h] Andrew R. M. Bradbury,[i] Antoine de Daruvar,[h,j] Stefan Dübel,[k] Ronald Frank,[l] Toby J. Gibson,[m] Larry Gold,[n] Niall Haslam,[m] Friedrich W. Herberg,[d] Tara Hiltke,[o] Jörg D. Hoheisel,[p] Samuel Kerrien,[a] Manfred Koegl,[q] Zoltán Konthur,[r] Bernhard Korn,[p] Ulf Landegren,[s] Luisa Montecchi-Palazzi,[a] Sandrine Palcy,[h,j] Henry Rodriguez,[o] Sonja Schweinsberg,[d] Volker Sievert,[r] Oda Stoevesandt,[t] Michael J. Taussig,[t] Marius Ueffing,[u] Mathias Uhlén,[e] Silvère van der Maarel,[v] Christer Wingren,[w] Peter Woollard,[x] David J. Sherman,[g,h] and Henning Hermjakob[a]

Protein affinity reagents (PARs), most commonly antibodies, are essential reagents for protein characterization in basic research, biotechnology, and diagnostics as well as the fastest growing class of therapeutics. Large numbers of PARs are available commercially; however, their quality is often uncertain. In addition, currently available PARs cover only a fraction of the human proteome, and their cost is prohibitive for proteome scale applications. This situation has triggered several initiatives involving large scale generation and validation of antibodies, for example the Swedish Human Protein Atlas and the German Antibody Factory. Antibodies targeting specific subproteomes are being pursued by members of Human Proteome Organisation (plasma and liver proteome projects) and the United States National Cancer Institute (cancer-associated antigens). ProteomeBinders, a European consortium, aims to set up a resource of consistently quality-controlled protein-binding reagents for the whole human proteome. An ultimate PAR database resource would allow consumers to visit one online warehouse and find all available affinity reagents from different providers together with documentation that facilitates easy comparison of their cost and quality. However, in contrast to, for example, nucleotide databases among which data are synchronized between the major data providers, current PAR producers, quality control centers, and commercial companies all use incompatible formats, hindering data exchange. Here we propose Proteomics Standards Initiative (PSI)-PAR as a global community standard format for the representation and exchange of protein affinity reagent data. The PSI-PAR format is maintained by the Human Proteome Organisation PSI and was developed within the context of ProteomeBinders by building on a mature proteomics standard format, PSI-molecular interaction, which is a widely accepted and established community standard for molecular interaction data. Further information and documentation are available on the PSI-PAR web site. ***Molecular & Cellular Proteomics 9:1–10, 2010.***

From the [a]European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, United Kingdom, [b]Medicinal Chemistry, Pharmaceutical Faculty, Copenhagen University, 2100 Copenhagen, Denmark, [d]Department of Biochemistry, University of Kassel, 34132 Kassel, Germany, [e]Royal Institute of Technology, AlbaNova University Center, SE-106 91 Stockholm, Sweden, [f]Department of Animal Breeding and Genetics and The Linnaeus Centre for Bioinformatics, Swedish University of Agricultural Sciences, 751 24 Uppsala, Sweden, [g]Institut National de Recherche en Informatique et en Automatique Bordeaux-Sud Ouest, Talence 33405, France, [h]University of Bordeaux, Laboratoire Bordelais de Recherche en Informatique, CNRS 5800, Talence 33405, France, [i]Los Alamos National Laboratory, Los Alamos, New Mexico 87545, [j]Centre de Bioinformatique de Bordeaux, Université de Bordeaux, Bordeaux 33000, France, [k]Institute of Biochemistry and Biotechnology, Technical University Braunschweig, D-38106 Braunschweig, Germany, [l]Helmholtz Center for Infection Research, D-38124 Braunschweig, Germany, [m]European Molecular Biology Laboratory, D-69117 Heidelberg, Germany, [n]SomaLogic Inc., Boulder, Colorado 80301, [o]NCI, National Institutes of Health, Bethesda, Maryland 20892, [p]German Cancer Research Center, D-69120 Heidelberg, Germany, [q]Resource Center for Genome Research, D-69120 Heidelberg, Germany, [r]Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany, [s]Department of Genetics and Pathology, Uppsala University, 751 85 Uppsala, Sweden, [t]Babraham Bioscience Technologies Ltd., Babraham Research Campus, Cambridge CB22 3AT, United Kingdom, [u]Department of Protein Science, Helmholtz Zentrum München, Research Center for Environmental Health, D-85764 Neuherberg, Germany, [v]Leiden University Medical Center, 2300 RC Leiden, The Netherlands, [w]Department of Immunotechnology, Lund University, SE-221 84 Lund, Sweden, [x]GlaxoSmithKline Pharmaceuticals, New Frontiers Science Park, Harlow, Essex CM19 5AW, United Kingdom, and [y]Department of Immunotechnology, CREATE Health, BioMedical Center (BMC) D13, Lund University, Lund, Sweden

Received, April 14, 2009, and in revised form, June 17, 2009
Published, MCP Papers in Press, August 14, 2009, DOI 10.1074/mcp.M900185-MCP200

Protein affinity reagents (PARs),[1] most commonly antibodies, are essential and ubiquitous reagents in academic and

[1] The abbreviations used are: PAR, protein affinity reagent; HUPO, Human Proteome Organisation; PSI, Proteomics Standards Initiative; MI, molecular interaction; CV, controlled vocabulary; mAb, monoclonal anti-

applied research. They have a wide use in functional characterization of proteins (expression levels, modifications, protein-protein interactions, and localization at the tissue and cellular level), purification of specific proteins and protein complexes, diagnostics (1), and therapeutics (2). PARs are used in standard laboratory techniques such as ELISA, Western blot, immunofluorescence, and immunohistochemistry and *in vivo* for imaging and therapy. Furthermore, they are increasingly used in highly multiplexed formats on microarrays both as immobilized capture reagents and as detection reagents. Research in the proteomics era has an unprecedented demand for specific PARs. Minimally, specific reagents are needed for a representative product of each open reading frame within entire genomes. Ideally, even wider sets of reagents should also distinguish diverse protein forms resulting from differential splicing and posttranslational modifications. However, the majority of human proteins lack a specific affinity reagent, and many proteins are represented by large numbers of different PARs of uncertain quality. Moreover, many of the existing PARs have not been adequately validated with regard to epitope site, specificity, affinity for different protein forms (splice variants and native/denatured form) or applicability in experimental techniques (*e.g.* immunohistochemistry *versus* Western blot). This hampers rational choices by PAR users who also lose time and money if purchased affinity reagent proves inadequate for their needs. Thus, increased throughput in PAR production and quality control are essential to avoid a bottleneck in proteomics research depending on these reagents and misinterpretation of data generated when using them.

Several initiatives for the systematic generation and validation of antibodies have been launched worldwide (3). The Swedish Human Protein Atlas (4, 5) catalogues protein distribution in healthy and diseased tissues and subcellular localization data in various cell types. For this purpose monospecific antibodies (affinity-purified polyclonal antibodies) are manufactured, quality-controlled, and applied in house. It is also possible for academic and commercial sources to submit antibodies for validation and use in the atlas. Release 4 of the Human Protein Atlas (6) contains more than five million images corresponding to ~5,000 human genes, and the antibodies can be ordered on line. The resource has also been embraced by HUPO in the form of a Human Antibody Initiative, which will incorporate antibodies developed in other HUPO projects such as the plasma and liver proteome projects.

The German Antibody Factory is a national collaboration developing automated *in vitro* methods for recombinant antibody production. Array- and bead-based systems are applied in selection protocols optimized toward minimum step numbers. The Antibody Factory also aims to integrate antibody selection into a pipeline with an enhanced rate of antigen production and high throughput specificity and cross-reactivity testing (7). Targets of interest include selected subproteomes, *e.g.* human transcription factors and signal molecules.

The Clinical Proteomic Technologies for Cancer Reagents and Resources component within the United States National Cancer Institute aims to establish a resource of highly characterized monoclonal antibodies directed against human proteins associated with cancer. This program aims to cover multiple target epitopes, to be applicable to a multitude of affinity platforms, and to generate standard operating procedures that are freely accessible to the public. Although hybridoma generation is performed by several contractors, final quality control is centralized. New antibodies selected for their relevance by literature mining and community feedback are released on a monthly basis (8).

The European ProteomeBinders consortium is the most ambitious initiative in the field of PAR resources, envisioning a resource of consistently quality-controlled affinity reagents for the entire human proteome, including functional protein variants. A particular focus is on replenishable reagents (recombinant binders selected by *in vitro* methods or monoclonal antibodies) as only these guarantee a sustainable resource (9). The consortium includes the Swedish Human Protein Resource, the German Antibody Factory, and more than 20 other leading academic and commercial laboratories in protein affinity reagent production, quality control, and applications. Currently funded by the European Commission for the planning of a future affinity reagent resource, a major activity of ProteomeBinders is the development of a bioinformatics infrastructure for large volumes of PAR data.

The aim of a PAR database resource would be to allow consumers to visit one on-line warehouse and find all available affinity reagents from different providers together with documentation that facilitates easy comparison of their cost, quality, and fields of application. However, in contrast to nucleotide databases among which data are synchronized between the major data providers, current PAR producers, quality control centers, and commercial companies all use incompatible formats, hindering the data exchange necessary to develop comprehensive resources. As part of the ProteomeBinders effort, a new publicly available portal was recently launched, called Antibodypedia (10), to allow sharing of information regarding validation of antibodies. In this pilot database, contributors are expected to provide experimental evidence and a validation score for each antibody, and the users can subsequently provide feedback and comments on the use of the antibody. The work to develop a PAR portal has resulted in an urgent need for a standard format for affinity reagent validation data. As an initial step to improve this situation, here we propose a global community standard for the representation and exchange of protein

body; xref, cross-reference; HSA, human serum albumin; DARPin, designed ankyrin repeat protein; Her2, human epidermal growth factor receptor 2.

affinity reagent data. The PAR format is maintained by the HUPO Proteomics Standards Initiative (PSI) and has consequently been assigned the acronym PSI-PAR. The PSI-PAR format was developed within the context of ProteomeBinders and has undergone the PSI document review process (11) in which several experts have provided criticism of the representation.

PAR-target protein binding is a type of molecular interaction, and for this reason the PSI-PAR format has been produced by adapting an existing format for MI, the PSI-MI format. PSI-MI, which is a mature proteomics standard, was developed in 2004 (12) by the HUPO PSI (13) and has been released in a new version, PSI-MI XML2.5 (14). It is a community standard for molecular interaction data, and currently several databases export data in this format. Building on an already existing format has the advantages that it has a thoroughly tested basis, software tools have already been developed that facilitate its use, and the maintenance effort is significantly reduced.

PSI-PAR FORMAT

The PSI-PAR format for data representation of protein affinity reagents presented here consists of the following.

- The PSI-MI XML2.5 schema for molecular interactions.
- The PSI-PAR controlled vocabulary.
- Documentation and user manual.

These parts are described in the sections below followed by three examples of published data represented in the PSI-MI XML2.5 schema.

Use Cases and Scope of PSI-PAR Format

The *inner* section of Fig. 1 illustrates how the PSI-PAR format is planned to be used for data exchange within the ProteomeBinders consortium (9). Member centers will share and exchange some data directly or via a central repository of accumulated data. For targeted proteins, the optimal (*e.g.* unique) epitopes are suggested by a bioinformatics pipeline (EpiC, the ProteomeBinders Epitope Choice Resource), and these data will be shared with target protein and PAR production centers. Data on produced target proteins are a prerequisite for many techniques generating affinity reagents; for example, immunization depends on a suitable (*e.g.* pure) immunogen. Thus, data will be transferred from protein to PAR production centers. Information about the produced proteins, affinity reagents, and procedures used to generate them will then be transferred and stored in the central data repository. Standardized protocols in the Molecular Methods Database (MolMeth) can be referenced to describe methods, reagents, and equipment used.

Quality control and characterization of affinity reagents and proteins will be conducted by member centers having complementary expertise in different experimental techniques. Importantly, the data they generate will be used to assess the quality of the affinity reagents and their suitability for certain purposes, *e.g.* application in an experimental technique such as ELISA. A public "warehouse" of protein affinity reagents will present to "customers" (*i.e.* members of the scientific community using the reagent resource) a summary of the key production and characterization information from the central repository. Finally, external sources, commercial and nonprofit, that are interested in making their affinity reagents available could be invited to do so after ensuring that their products meet a reference for quality control standard.

The use cases for types of data exchange can be summarized into three broad categories: 1) affinity reagent and target protein production data, 2) characterization/quality control results, and 3) complete summaries of end products. The first category, PAR and target protein production data, is a new scope as it was not previously represented in the PSI-MI format. This has necessitated new types of molecules, specifically varieties of affinity reagents, as well as production methods to be added to the representation. Also, the representation of characterization data demands more in-depth descriptions of, for example, experimental materials, binding sites, and non-interacting molecules, which are typically controls in experiments that assess cross-reactivity. The complete summaries of end products span, for each affinity reagent, the generation information and characterization/quality control results as well as marketing information, such as price and supplied form. A formal document of the minimum information about a protein affinity reagent is currently in preparation.[2] This will serve as a guideline for the community that defines the information that needs to be disclosed to unambiguously describe a protein affinity reagent.

PSI-MI XML2.5 Schema

Standard formats provide a common structure for data representation. This section gives a brief overview of the PSI-MI XML2.5 schema. More information can be found in the original publication (14) and the on-line documentation (PSI-PAR and PSI-MI web pages). The key elements of the PSI-MI XML2.5 are outlined in Fig. 2 and written in italic the text below. The root, *entrySet*, can hold several *entry* elements, each typically containing the information from one publication or study. The *entry* has six child elements that provide the overall coverage of data: *source* (groups, institutes, companies, etc.), *availabilityList* (availability restrictions, *e.g.* copyrights or intellectual properties), *experimentList* (experiments), *interactorList* (interacting molecular species), *interactionList* (experimental outcomes such as produced molecules or char-

[2] J. Bourbellion, S. Orchard, I. Benhar, C. Borrebaeck, A. de Daruvar, S. Dübel, R. Frank, F. Gibson, D. Gloriam, N. Haslam, I. Humphrey-Smith, M. Hust, D. Junker, M. Koegl, K. Konthur, B. Korn, S. Krobitsch, S. Muyldermans, P. Å. Nygren, S. Palcy, B. Polic, H. Rodriguez, A. Sawyer, M. Schlapshy, M. Snyder, O. Stoevesandt, M. Taussig, M. Templin, M. Uhlen, S. van der Maarel, C. Wingren, H. Hermjakob, and D. Sherman, manuscript in preparation.

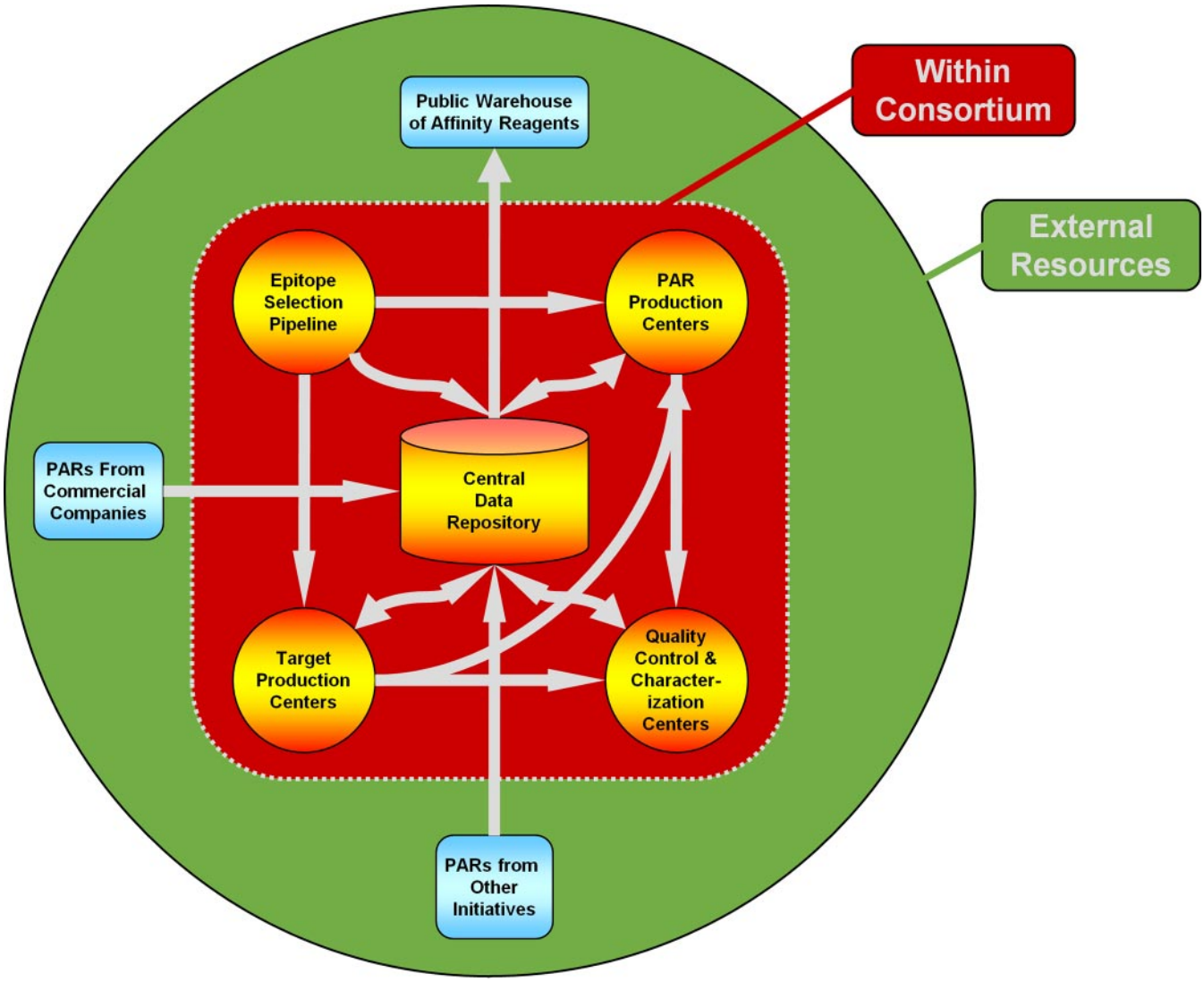

FIG. 1. **Use cases for HUPO PSI-PAR format.** Each data exchange or sharing event is illustrated with an *arrow* in the diagram. The common means of PAR representation facilitate the building of integrated networks of PAR-producing and -characterizing centers, here exemplified by ProteomeBinders. This will be of tremendous benefit for the scientific community as it allows for centralized and standardized sources of information on quality and availability of PARs ("*Public Warehouse of Affinity Reagents*" in the figure).

acterization results), and *attributeList* (additional attributes). These six elements have further branches (not shown in Fig. 2) that give in-depth representation. For example, *experimentDescription* has child elements that specify the experimental methods used and the organism in which the experiment has been performed (can be *in vitro*).

Proteins and affinity reagents (as well as other molecules) are represented at two levels in the XML schema. The *interactor* element, at the generic level, captures basic information about the identity such as name, reference to a public database entry, sequence, and/or chemical structure. The *participant* element (child of *interaction*) describes the specific version of the molecule in the given experiment detailing, for example, its preparation, sequence features (labels, tags, binding sites, etc.), and role in the experiment. Furthermore, the *participant* has a child element, *parameter*, which represents quantitative properties of molecules such as weight or percent purity. The *parameter* element is also found under the *interaction* element where it captures quantitative experimental results such as affinity and kinetic data. The PSI-MI XML2.5 schema has built-in extendibility in the form of the *attributeList*. The *attributeList* gives a semistructured extension as the names of *attributes* are defined by the controlled vocabulary whereas the information they hold is free text. It is available for the descriptions for molecules (*interactor* and *participant*), experiments (*experimentDescription*), and experimental outcomes (*interaction*), and new attributes created here include, for example, "protocol," "equipment," and "results comment."

To conserve compatibility with existing software tools and infrastructures, no new elements have been added to the

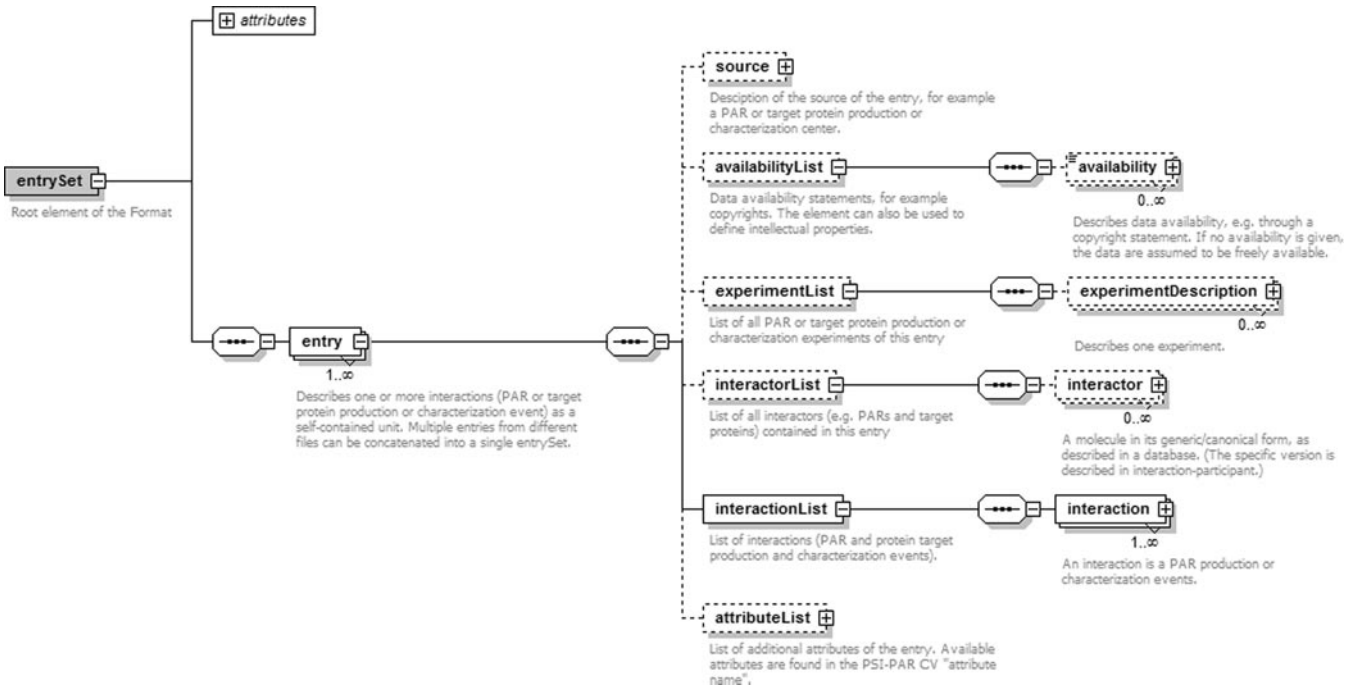

FIG. 2. **Graphical representation of PSI-MI XML2.5 schema.** Some elements have been collapsed for clarity (indicated by a "+" in a *rectangular box*). The figure is derived from the publication of the PSI-MI 2.5 format (14).

PSI-MI XML2.5 schema. Using the existing rather than constructing a new XML schema reduces the maintenance effort and prevents users from having to develop their own software and tools. Although no new elements have been added, new functionality has in two examples been added by using cross-references to new controlled vocabulary (CV) subtypes/branches. First, in the *feature* element, a cross-reference of the type PSI-PAR term for experimental scope to the experimental scope CV subtype can be used to describe the scope of the experiment as being either molecule production or one of a range of characterization objectives (see below). Second, in the *experimentDescription*, a cross-reference of the type BioSapiens Annotations term for secondary structure to the branch polypeptide secondary structure in the BioSapiens Annotations CV can be used to describe secondary structures of polypeptides. Furthermore, the *interaction* element has adopted a slightly modified use as its scope has expanded to encompass the outcome of molecule production experiments, *i.e.* physical products such as antibodies. Moreover, in the description of "more traditional" molecular interactions, *interactions* labeled with a *negative* element have adopted a more central role. In MIs, *interactions* with a *negative* element are relatively rare and indicate that the given interaction does not occur under the specified experimental conditions. In the representation of PAR data, these are used to capture nonbinding relationships typically when affinity reagents are tested for cross-reactivity against controls.

PSI-PAR Controlled Vocabulary

XML schemas, such as the PSI-MI XML2.5 schema, standardize the structure, but not the semantics, of data representation. To ensure common terminology, elements of the schema are populated with terms from controlled vocabularies, which outline lists of standardized terms. Each CV term has a standardized name, a definition, and one or more aliases. CVs have the advantage that a richer representation can be obtained solely by adding new terms and thus without the requirement to change XML schema structure, which needs to remain stable to conserve compatibility with software. As described above, the scopes of molecular interactions and protein affinity reagents are largely overlapping but are also partially unique. This fact is reflected in the PSI-PAR CV that contains the majority of the terms from the PSI-MI CV and in addition ~200 new terms. The PSI-PAR CV can be browsed using the Ontology Lookup Service (16) as can also most of the external CVs/ontologies used together with the PSI-MI XML2.5 schema (including the Gene Ontology, NCBI taxonomy ontology, BioSapiens Annotations, and Unit Ontology). The maintenance of the PSI-PAR and PSI-MI CVs is performed by an elected editorial board of the PSI-MI work group that keeps them in one common master, and users may request new terms via an on-line tracker (SourceForge).

One CV subtype, molecule production method (see Fig. 3), is specific to the PSI-PAR CV and encompasses experimental methods used for affinity reagent and target protein production, for example expression from cDNA, immunization, and

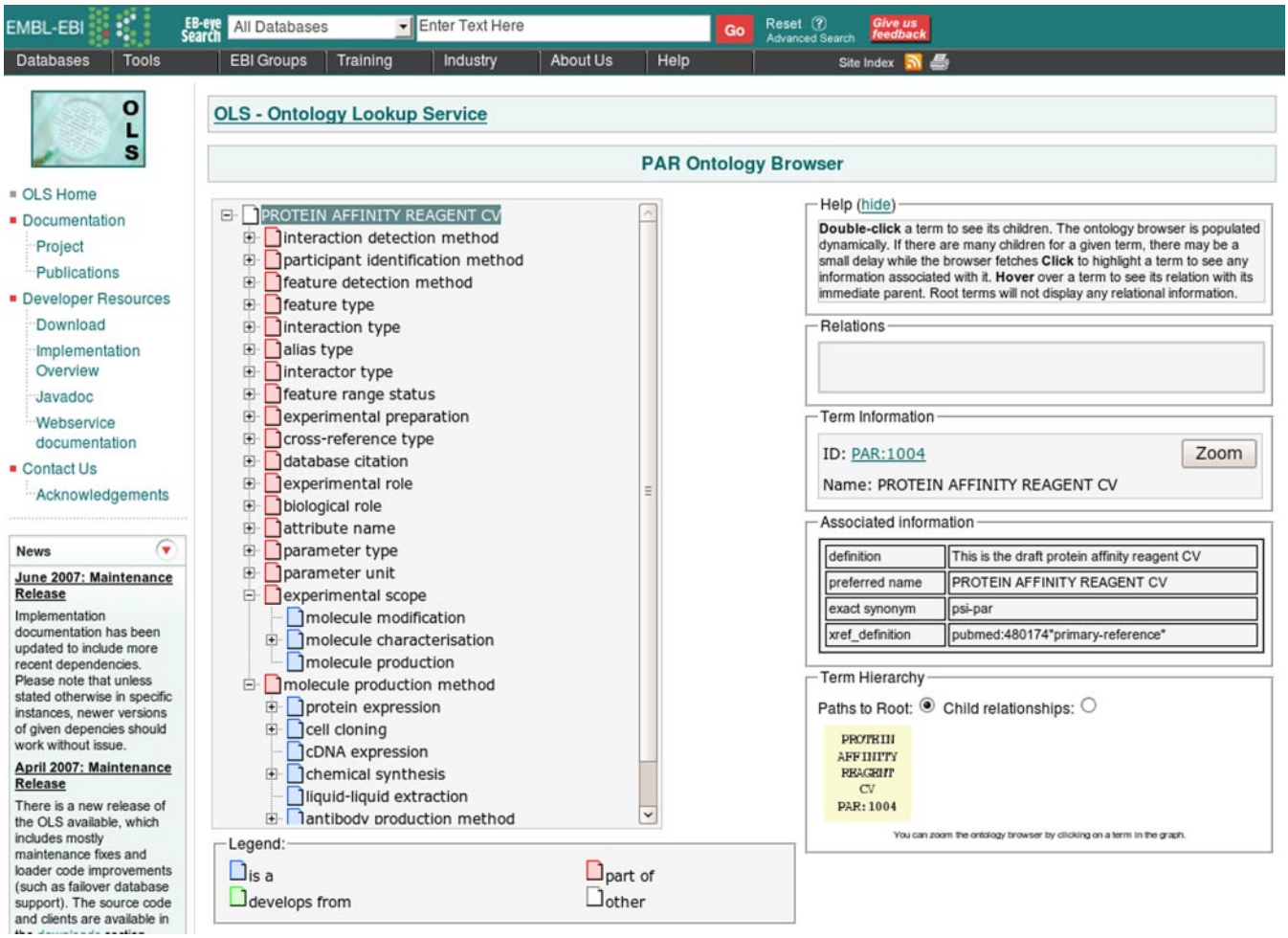

FIG. 3. **Representative section of PSI-PAR controlled vocabulary displayed in Ontology Lookup Service (16).** The two new CV subtypes "experimental scope" and "molecule production method" can be seen at the *bottom*.

chemical synthesis. Note that the molecule production method CV subtype is used by the *interaction detection method* element, which utilizes the interaction detection method CV subtype for all other types of methods. The new CV subtype experimental scope describes the scope of the experiment. The scope can be defined as either molecule production or one of a range of molecule characterization objectives such as binding site assessment, quantification, and kinetics determination. Three existing CV subtypes have been equipped with new relatively large branches. In the attribute name CV subtype, the branch experimental material attribute name describes experimental materials such as microarrays, tissues, molecule libraries, and cDNA expression vectors. In the experimental preparation CV subtype, the molecular state branch lists aggregation, folding, and purity states for molecules (*participants*). In the interactor type CV subtype, two branches, antibody and engineered protein scaffold, have been added. In addition, a significant number of individual and spread new CV terms have been added, for example experimental control, equipment, and protocol.

Documentation, User Manual, and Examples of Protein Affinity Reagent Data in PSI-PAR Format

Documentation aiming to support new users of the PSI-PAR format has been developed and made available on the PSI-MI web page. The use of schema elements is explained in words in a user manual that does not require previous knowledge about the PSI-MI format. It is also explained graphically in an autogenerated documentation in which the schema structure can be browsed and the composition of individual elements can be examined. To further illustrate the representation of PAR data in the PSI MI XML2.5 schema, we have captured sample data from three relevant published articles that were carefully chosen to cover a diversity of PAR production and characterization data. Complete XML and HTML files are available in the supplemental data, which also include a spreadsheet overview of their representation. Below, selected features from these examples are described with an emphasis on how the representation has been adapted to protein affinity reagent and target protein production and characterization data.

Example 1 of Protein Affinity Reagent Data in PSI-PAR Format—The first example is a representation of data in the article "Characterization of monoclonal antibodies to human group B rotavirus and their use in an antigen detection enzyme-linked immunosorbent assay" by Burns *et al.* (17). This study describes the production of three monoclonal antibodies (mAbs) and design of a capture antigen detection ELISA intended to be used as diagnostic tools for the human group B rotavirus, an agent implicated in epidemic outbreaks of diarrhea in China.

For each PAR or protein production experiment the methods, for example immunization and hybridoma production, have been specified using terms from the new molecule production method CV subtype. The production of the interactors is described in a number of experiments (shown in bold) that have been labeled with the experimental scope CV term molecule production. Liquid-liquid extractions (liquid-liquid extraction) of nine stool samples containing human group B rotavirus were prepared (**burns-1989-1**). One sample, J-1, was further purified by banding in a CsCl gradient (solution sedimentation) (**burns-1989-2**). The mAbs, B5C9, B5E4, and B10G10, were produced by immunization (animal immunization) of the purified J-1 virus in mice (**burns-1989-3**) followed by the generation of hybridomas (hybridoma generation) (**burns-1989-4**) and isolation of mAbs from a polyclonal mixture (ascites fluid) with HPLC (chromatography technology) (**burns-1989-5**).

Each experiment above has resulted in one or several interactions, one for each product. Note that in this case the interaction does not describe characterization/assay results but rather a physical product. The antibodies and viruses produced here represent two new types in the interactor type CV, which has been extended with a variety of protein affinity reagents. The identities of the virus *interactors* have been defined by assigning them references to a human group B rotavirus entry in the UniProt taxonomy resource "newt." On the *participant* level, the *biologicalRole* of the products has been defined as either protein affinity reagent or protein affinity reagent target, and this terminology is used throughout the representation to selectively label these. Starting materials (not annotated for this example), intermediate products (*e.g.* hybridomas), and end products (*e.g.* antibodies) are all represented as *interactors*/*participants* and have the *experimentalRole* starting material, intermediate product, and generation product, respectively.

Work flows of sequential production steps have been captured by the successive linking of *interactions* to the preceding using a cross-reference (*xref*) of the type preceding interaction. The referencing starts from the last *interaction*, which is the one generating the end product (defined with the *experimentalRole* CV term generation product), and continues until the first *interaction* in the work flow. For example "mab_b5e4_isolation" (isolation of the mAb B5E4) references "antij1_b5e4_hybridom" (hybridoma generation) that in its turn references "immunization_w_j1" (animal immunization).

The first characterization experiment (**burns-1989-6**) was a competitive binding ELISA mapping the relative epitopes (experimental scope: binding site determination) for the three antibodies. To facilitate easy matching of captured and original data (in this case binding curves), the figure numbers in the original article (Fig. 3, a–c) have been added to the respective *interactions* in free text *attributes* of the type figure legend. The results from the epitope mapping have been captured in another *attribute*, results comment, describing which antibodies were found to compete, *i.e.* have overlapping binding sites. The status of the mAbs, *e.g.* competing or non-competing, is also captured by their *experimentalRoles* (experimental role). The *experimentalRole* was defined as competitor for competing antibodies and neutral component for non-competing antibodies. Two antibodies had been added, in increasing concentrations, to the solutions to assess whether their binding of the target virus in solution (*experimentalRole*: prey) was competing with that of the immobilized antibody (*experimentalRole*: bait). Distinguishing affinity reagents with unique binding sites is of high significance because it implies that the reagents can bind simultaneously to the same protein, which is a requisite for utilization in sandwich assays and *in situ* confirmation in *e.g.* diagnostic experiments.

A second characterization experiment (**burns-1989-7**) assessed the affinity (experimental scope: affinity determination) of the three mAbs to each of the nine virus stool samples. This was a "capture antigen detection ELISA," and this is here reflected by the *experimentalRoles* of the *participants*; *i.e.* immobilized antibodies were defined as bait, and virus samples in solution were defined as prey. Each antibody-virus sample measurement (Table 1 in Burns *et al.* (17)) has been captured in its own *interaction*, and the results in the form of absorbance values (direct and normalized, respectively) are represented as *parameters*.

The third characterization experiment (**burns-1989-8**) assessed the specificity of the antibodies in a similar ELISA experiment where they were again used as capture reagents. Here, the virus samples were not presented separately but in two groups of 1) 15 group B rotavirus-containing samples and 2) 57 other samples whereof 37 have been shown to contain group A rotavirus. The three antibodies were found to bind to the first group but not the second (Table 2 in Burns *et al.* (17)), and thus, they could successfully distinguish group B viruses. The binding of the antibodies to the 15 binding group B rotavirus-containing samples was not captured because their ability to bind such viruses has already been captured in more detail in previous *interactions*. The non-binding of the mAbs to the 57 non-group B rotavirus samples was represented in *interactions* being assigned a "true" *negative* element to specify that they are not binding/interacting. Because the identities of the specific samples are unknown, they could not be captured individually; instead, the non-binding 57 non-group B

samples were simply represented as one *interactor*. The identities of these viruses were specified using a reference (*xref*) to the UniProt taxonomy resource newt, and their source was captured by the *participant attribute* supplier.

Example 2 of Protein Affinity Reagent Data in PSI-PAR Format—The second example describes the representation of data from the article "A proteomics-based approach for monoclonal antibody characterization" by Weiler *et al.* (18). The objective of this study was to develop a proteomics methodology that can be utilized as a generic approach for the assessment of the specificity of antibodies without the need for pure antigens/proteins.

Six anti-human serum albumin (HSA; UniProtKB P02768) monoclonal antibodies were produced (experimental scope: molecule production) by immunization (**weiler-2003-1**), hybridoma production (**weiler-2003-2**), and harvesting directly from the growth medium (**weiler-2003-3**). In the immunization *interaction* ("immunization_w_hsa"), HSA is represented as a *participant* with the *experimentalRole* immunogen. The *interactions* representing the productions of the hybridomas do not contain any *participants* as the representation of intermediate products is optional (see above). The harvesting procedure is not described in the article but is implicit from the hybridoma generation. In the representation of these data, however, it is necessary to have a separate *experimentDescription* (**weiler-2003-3**) for this step because the antibodies should not be captured as *participants* of the hybridoma generation *interactions* (*e.g.*"mab_6g11_hybridoma") that have only generated the intermediates (hybridomas). As described for the first example, production work flows have been captured by successively linking *interactions* in reverse chronological order using an *xref* of the type preceding interaction.

The first two characterization experiments had the objective to assess whether the six antibodies work as affinity reagents (experimental scope: affinity determination) in the ELISA (**weiler-2003-4**) and Western blot (**weiler-2003-5**) techniques. All six antibodies produced positive signals in ELISA, whereas two of them failed in binding to HSA in the blotting experiment, and consequently their *interactions* ("7b3-hsa_wb_neg" and "11g9-hsa_wb_neg") were labeled with the *negative* element with the value true to indicate that the *participants* (mAb-HSA) do not bind. *Interactions* containing the *negative* element have also been used to capture the non-interacting status of the negative controls (*experimentalRole*: negative control) used: medium, mouse serum, and an irrelevant antibody.

The next three characterization experiments (**weiler-2003-6**) assessed the selectivity (experimental scope: cross-reactivity assessment) of the six mAbs by testing whether they could selectively capture HSA from an artificial protein mixture (Table 2 in Weiler *et al.* (18)), synovial fluid (Fig. 3 in Weiler *et al.* (18)), and a cell lysate (Table 3 in Weiler *et al.* (18)). Here, the *interaction detection method* was defined as pulldown, the *participant identification method* was defined as peptide mass fingerprinting, and the capture mAbs were assigned the *experimentalRole* bait. Successful captures/pulldowns were represented in *interactions* in which the HSA *participant* was assigned the *experimentalRole* prey. Non-HSA proteins in the artificial mixture were assigned the *experimentalRole* negative control and captured in separate *interactions* labeled with the *negative* element. The synovial fluid and cell lysate are too complex to allow for the specification of non-binding constituents, and for this reason, they were assigned the experimentalRole neutral component and captured in the same *interaction* as the mAb and HSA. The fact that the antibody could selectively capture human serum albumin from synovial fluid and the cell lysate was captured in an *attribute* with the name results comment.

One more characterization experiment (**weiler-2003-7)**, a competitive ELISA, was captured that confirmed that 10C9 binds to human serum albumin but not to human haptoglobin (UniProtKB P00738) (Fig. 4 in Weiler *et al.* (18)). Another antibody, 11G9, was also assayed but with the opposite results. These data were not captured because the focus was restricted to affinity reagents for the intended target protein (HSA), and 11G9 had been captured before (*interaction*"11g9-mix").

Example 3 of Protein Affinity Reagent Data in PSI-PAR Format—The last example provided is a representation of selected data from the article "A designed ankyrin repeat protein evolved to picomolar affinity to Her2" by Zahnd *et al.* (19). The affinity reagents in this study, "designed ankyrin repeat proteins" (DARPins), against human epidermal growth factor receptor 2 (Her2) had previously been generated from large synthetic libraries *in vitro* using ribosome display (20). This study aimed to mature the affinities of these DARPins by modifying their structures using error-prone PCR of their cDNA clones and a new round of ribosome display to select for mutated versions with increased affinity.

As the title of the article states ("A designed ankyrin repeat protein evolved to picomolar affinity…"), its primary focus is on the one DARPin with the highest affinity, and in this representation, only this molecule, H10-2-G3, has been represented from the first experiment and throughout. The first experiment in this study, the creation of the cDNA library containing the mutated DARPins (by error-prone PCR), has not been captured because the representation of experimental materials has been restricted to their use excluding production. Thus, the first experiment captured here (**zahnd-2007-1**) describes how high affinity DARPins were selected from this library using ribosome display. In this experiment, the number of screening rounds (three) has been defined in an *attribute*, screening rounds. The outcome of the experiment is represented in an *interaction* ("lib_screen") containing two *participants*, H10-2-G3 and Her2, that have been assigned the *experimentalRoles* generation product and prey (two roles) and bait, respectively. Two *features* were added to the *participant*, "biotinylated Her2 (UniProtKB P04626) extracellular domain," and these were of the *feature types* biotin tag and extracellular

domain, respectively. After identification, the highest affinity DARPin, H10-2-G3, was produced by cDNA expression in *Escherichia coli* (**zahnd-2007-2**) and purification with immobilized metal ion affinity chromatography (**zahnd-2007-3**). The name of the expression vector, pAT224, was captured in an *attribute* of the cDNA expression, *experimentDescription*.

Also, the capturing of the protein affinity reagent characterization experiments has been restricted, and the construction of point mutants and structural analyses were considered too peripheral. However, the determined structure of the highest affinity DARPin has been captured with an *xref* for this *interactor* to its entry in the Protein Data Bank (code 2jab). The first captured characterization experiment (**zahnd-2007-4**) involved the determination of affinities (experimental scope: affinity determination) with kinetic surface plasmon resonance. Kinetic parameters of the *interaction* determined (Table 1 and Fig. 2 in Zahnd *et al.* (19)) have been captured as *parameters*, and the name of the machine used ("Biacore 3000") was added in an *attribute*, equipment, of the *experiment*. Another characterization experiment (**zahnd-2007-5**) represented here was a sensitivity determination in a Luminex assay. In this case, three *participants* were captured, each with different *experimentalRoles*: H10-2-G3 (bait), Her2 extracellular domain (prey), and a detection antibody, sp185_ab (secondary protein affinity reagent). The capturing of detection antibodies should be considered optional but can be useful in many cases. H10-2-G3 was fused via a biotin tag to avian protein D, and this fact was represented in two *features* of the *feature types* biotin tag and fusion protein, respectively. The sensitivity (signal to noise ratio; Fig. 6a in Zahnd *et al.* (19)) of the highest affinity DARPin, H10-2-G3, was captured as an *attribute* of the *interaction* with the type results comment.

How PAR Users Would Benefit from a Standard Format—The main benefit of a standard format is that it can be used to gather information from different sources into one comprehensive resource, such as is described in Fig. 1, which shows how production and quality control data from many institutions can be collated in one public warehouse. For a PAR user, this means access to more affinity reagents and quality control data. The three examples above comprise different PARs and target proteins, but even these limited data would allow a PAR consumer to do the following.

- Search for different types of PARs: mAbs (examples 1 and 2) and DARPin (example 3).
- Assess compatibility with an experimental technique: only two of the six antibodies in example 2 could be used in Western blot, whereas they all worked in ELISA format.
- Compare affinity data: affinities as well as ELISA readouts for the mAbs and virus samples in example 1.
- Inspect cross-reactivity data: five of six mAbs in example 2 can selectively pull down the target protein from an artificial protein mixture, a cell lysate, and synovial fluid, whereas the last mAb had affinity for another protein.
- Identify PAR pairs with sandwich assay compatibility: mAbs B5C9 and B10G10 recognize different epitopes on the analyte (example 1).
- List PARs with known three-dimensional structure: the DARPin in example 3.

Software Tools

A number of tools have been developed for the PSI-MI XML2.5 schema, and below is a summary of the ones most relevant to PSI-PAR. A complete description can be found on the PSI-MI web site (14). Validating tools have been developed that can interface with ontologies and perform semantic validation on PSI-MI XML2.5 files. Validation can be carried out by 1) a Java API that enables the embedding of the validator into any third party application, 2) a command line interface, and 3) a web application (PSI-MI 2.5 Validator) that allows the uploading of a PSI-MI data file and reporting of both syntactic and semantic discrepancies. The Ontology Lookup Service is an ontology viewer with browsing and search functionality (16). It comprises the PSI-PAR CV and a number of additional CVs that are used in conjugation with the PSI-MI XML2.5 schema such as the Gene Ontology, BioSapiens Annotations, and Unit Ontology. A Java XML parser has been developed that allows for import and export of PSI-MI XML2.5 files to and from databases. It comprises a Java library and may also be used to develop any type of software reading and/or writing PSI-MI XML2.5 data (SourceForge). XML style sheets are available that can convert PSI-MI XML2.5 data files to HTML, thus providing user-friendly human-readable representation. Finally, a complete, open source database implementation providing reading, writing, and interactive editing of data in PSI-MI XML2.5 schema exists, the IntAct molecular interaction database (15). A front end tailored to the semantics and use cases of protein affinity reagents would, for example, allow customers in the PAR warehouse described under "Use Cases and Scope of PSI-PAR Format" to browse the database for availability and quality control information.

CONCLUSIONS

We have provided the first framework for a common representation of protein affinity reagents, a very complex domain of proteomics, spanning a range of biological and chemical entities and numerous experimental techniques on a proteome-wide scale. This has been achieved through the complementing of an existing, mature proteomics standard, the PSI-MI format, with a modified controlled vocabulary and extensive documentation including a user manual and example files. The acceptance of the PSI-PAR format as a standard will be determined by its usability and reliability, which have here been achieved by conserving compatibility with existing software tools and the incorporation into the PSI community effort to ensure long term maintenance. Molecular interaction databases have exported in the PSI-MI format for several years, and the ProteomeBinders consortium intends to use the PSI-PAR format in a future production phase to connect

multiple partners involved in reagent generation, quality control, and application. The use cases described here, based on the infrastructure of ProteomeBinders, apply to the general community of protein affinity reagent producers.

Non-profit initiatives, commercial vendors, and users could be interested in a common standard for different reasons. Non-profit initiatives would benefit from the free access to the format and the associated tools. Commercial vendors would be attracted by increased market exposure of their products. Researchers wishing to purchase protein affinity reagents would benefit from the possibility of establishing centralized stores of quality-controlled PARs with larger choice, higher quality, and lower cost. The PSI effort actively seeks input and advice from the wider community. Anyone wishing to become involved in the technical representation of proteomics data is invited to visit the PSI web site to participate in the discussion groups listed and to contribute to the further development of the PSI-PAR and other proteomics standards.

AVAILABILITY

The PSI-MI XML2.5 schema, PSI-PAR CV, documentation, and tools are maintained by the PSI-MI work group. The freely available documents are posted and updated on the web pages of PSI-PAR and PSI-MI.

* This work was authored, in whole or in part, by National Institutes of Health staff. This work was supported by the ProteomeBinders European Commission (EC) FP6 research infrastructures coordination action (Grant RI-CA 026008) and the EC FP7 Biobanking and Biomolecular Resource Infrastructure (Grant Agreement 21211) within WP4.

[S] The on-line version of this article (available at http://www.mcponline.org) contains supplemental data 1–7.

[c] Additionally supported by the European Science Foundation and its functional genomics program in the form of a postdoctoral fellowship. To whom correspondence should be addressed. Tel.: 46703148390; Fax: 4535336040; E-mail: par@psidev.info.

REFERENCES

1. Borrebaeck, C. A. (2000) Antibodies in diagnostics—from immunoassays to protein chips. *Immunol. Today* **21,** 379–382
2. Piggee, C. (2008) Therapeutic antibodies coming through the pipeline. *Anal. Chem.* **80,** 2305–2310
3. Stoevesandt, O., and Taussig, M. J. (2007) Affinity reagent resources for human proteome detection: initiatives and perspectives. *Proteomics* **7,** 2738–2750
4. Persson, A., Hober, S., and Uhlén, M. (2006) A human protein atlas based on antibody proteomics. *Curr. Opin. Mol. Ther.* **8,** 185–190
5. Uhlén, M., Björling, E., Agaton, C., Szigyarto, C. A., Amini, B., Andersen, E., Andersson, A. C., Angelidou, P., Asplund, A., Asplund, C., Berglund, L., Bergström, K., Brumer, H., Cerjan, D., Ekström, M., Elobeid, A., Eriksson, C., Fagerberg, L., Falk, R., Fall, J., Forsberg, M., Björklund, M. G., Gumbel, K., Halimi, A., Hallin, I., Hamsten, C., Hansson, M., Hedhammar, M., Hercules, G., Kampf, C., Larsson, K., Lindskog, M., Lodewyckx, W., Lund, J., Lundeberg, J., Magnusson, K., Malm, E., Nilsson, P., Odling, J., Oksvold, P., Olsson, I., Oster, E., Ottosson, J., Paavilainen, L., Persson, A., Rimini, R., Rockberg, J., Runeson, M., Sivertsson, A., Sköllermo, A., Steen, J., Stenvall, M., Sterky, F., Strömberg, S., Sundberg, M., Tegel, H., Tourle, S., Wahlund, E., Waldén, A., Wan, J., Wernérus, H., Westberg, J., Wester, K., Wrethagen, U., Xu, L. L., Hober, S., and Pontén, F. (2005) A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol. Cell. Proteomics* **4,** 1920–1932
6. Berglund, L., Björling, E., Oksvold, P., Fagerberg, L., Asplund, A., Szigyarto, C. A., Persson, A., Ottosson, J., Wernérus, H., Nilsson, P., Lundberg, E., Sivertsson, A., Navani, S., Wester, K., Kampf, C., Hober, S., Pontén, F., and Uhlén, M. (2008) A genecentric Human Protein Atlas for expression profiles based on antibodies. *Mol. Cell. Proteomics* **7,** 2019–2027
7. Konthur, Z. (2007) Automation of selection and engineering, in *Handbook of Therapeutic Antibodies* (Dübel, S., ed) Wiley-VCH, Weinheim, Germany
8. Haab, B. B., Paulovich, A. G., Anderson, N. L., Clark, A. M., Downing, G. J., Hermjakob, H., Labaer, J., and Uhlen, M. (2006) A reagent resource to identify proteins and peptides of interest for the cancer community: a workshop report. *Mol. Cell. Proteomics* **5,** 1996–2007
9. Taussig, M. J., Stoevesandt, O., Borrebaeck, C. A., Bradbury, A. R., Cahill, D., Cambillau, C., de Daruvar, A., Dübel, S., Eichler, J., Frank, R., Gibson, T. J., Gloriam, D., Gold, L., Herberg, F. W., Hermjakob, H., Hoheisel, J. D., Joos, T. O., Kallioniemi, O., Koegl, M., Konthur, Z., Korn, B., Kremmer, E., Krobitsch, S., Landegren, U., van der Maarel, S., McCafferty, J., Muyldermans, S., Nygren, P. A., Palcy, S., Plückthun, A., Polic, B., Przybylski, M., Saviranta, P., Sawyer, A., Sherman, D. J., Skerra, A., Templin, M., Ueffing, M., and Uhlén, M. (2007) ProteomeBinders: planning a European resource of affinity reagents for analysis of the human proteome. *Nat. Methods* **4,** 13–17
10. Björling, E., and Uhlén, M. (2008) Antibodypedia, a portal for sharing antibody and antigen validation data. *Mol. Cell. Proteomics* **7,** 2028–2037
11. Vizcaíno, J. A., Martens, L., Hermjakob, H., Julian, R. K., and Paton, N. W. (2007) The PSI formal document process and its implementation on the PSI website. *Proteomics* **7,** 2355–2357
12. Haslam, N., and Gibson, T. (2009) EpiC: A Resource for Integrating Information and Analyses to Enable Selection of Epitopes for Antibody Based Experiments (Istrail, S., Pevzner, P., and Waterman, M., ed), Springer, Berlin/Heidelberg, Germany, pp. 173–181
13. Hermjakob, H., Montecchi-Palazzi, L., Bader, G., Wojcik, J., Salwinski, L., Ceol, A., Moore, S., Orchard, S., Sarkans, U., von Mering, C., Roechert, B., Poux, S., Jung, E., Mersch, H., Kersey, P., Lappe, M., Li, Y., Zeng, R., Rana, D., Nikolski, M., Husi, H., Brun, C., Shanker, K., Grant, S. G., Sander, C., Bork, P., Zhu, W., Pandey, A., Brazma, A., Jacq, B., Vidal, M., Sherman, D., Legrain, P., Cesareni, G., Xenarios, I., Eisenberg, D., Steipe, B., Hogue, C., and Apweiler, R. (2004) The HUPO PSI's molecular interaction format—a community standard for the representation of protein interaction data. *Nat. Biotechnol.* **22,** 177–183
14. Hermjakob, H. (2006) The HUPO Proteomics Standards Initiative—overcoming the fragmentation of proteomics data. *Proteomics* **6,** 34–38
15. Kerrien, S., Orchard, S., Montecchi-Palazzi, L., Aranda, B., Quinn, A. F., Vinod, N., Bader, G. D., Xenarios, I., Wojcik, J., Sherman, D., Tyers, M., Salama, J. J., Moore, S., Ceol, A., Chatr-Aryamontri, A., Oesterheld, M., Stümpflen, V., Salwinski, L., Nerothin, J., Cerami, E., Cusick, M. E., Vidal, M., Gilson, M., Armstrong, J., Woollard, P., Hogue, C., Eisenberg, D., Cesareni, G., Apweiler, R., and Hermjakob, H. (2007) Broadening the horizon—level 2.5 of the HUPO-PSI format for molecular interactions. *BMC Biol.* **5,** 44
16. Kerrien, S., Alam-Faruque, Y., Aranda, B., Bancarz, I., Bridge, A., Derow, C., Dimmer, E., Feuermann, M., Friedrichsen, A., Huntley, R., Kohler, C., Khadake, J., Leroy, C., Liban, A., Lieftink, C., Montecchi-Palazzi, L., Orchard, S., Risse, J., Robbe, K., Roechert, B., Thorneycroft, D., Zhang, Y., Apweiler, R., and Hermjakob, H. (2007) IntAct—open source resource for molecular interaction data. *Nucleic Acids Res.* **35,** D561–D565
17. Côté, R. G., Jones, P., Apweiler, R., and Hermjakob, H. (2006) The Ontology Lookup Service, a lightweight cross-platform tool for controlled vocabulary queries. *BMC Bioinformatics* **7,** 97
18. Burns, J. W., Welch, S. K., Nakata, S., and Estes, M. K. (1989) Characterization of monoclonal antibodies to human group B rotavirus and their use in an antigen detection enzyme-linked immunosorbent assay. *J. Clin. Microbiol.* **27,** 245–250
19. Weiler, T., Sauder, P., Cheng, K., Ens, W., Standing, K., and Wilkins, J. A. (2003) A proteomics-based approach for monoclonal antibody characterization. *Anal. Biochem.* **321,** 217–225
20. Zahnd, C., Wyler, E., Schwenk, J. M., Steiner, D., Lawrence, M. C., McKern, N. M., Pecorari, F., Ward, C. W., Joos, T. O., and Plückthun, A. (2007) A designed ankyrin repeat protein evolved to picomolar affinity to Her2. *J. Mol. Biol.* **369,** 1015–1028
21. Zahnd, C., Pecorari, F., Straumann, N., Wyler, E., and Plückthun, A. (2006) Selection and characterization of Her2 binding-designed ankyrin repeat proteins. *J. Biol. Chem.* **281,** 35167–35175