Open Access

Development of a use Case for Chemical Resource Description Framework for Acquired Immune Deficiency Syndrome Drug Discovery

Talapady Narayana Bhat*,[1] and John Barkley[2]

[1]*Biochemical Science Division (831), NIST, Gaithersburg, MD 20899 USA and* [2]*Software Diagnostics and Conformance Testing Division (897), NIST, Gaithersburg, MD 20899 USA*

Abstract: There is considerable interest in RDF (Resource Description Framework) as a data representation standard for the growing information technology needs of drug discovery. Though several efforts towards this goal have been reported, most of the reported efforts have focused on text-based data. Structural data of chemicals are a key component of drug discovery and molecular images may offer certain advantages over text-based representations for them. Here we discuss the steps that we used to develop and search chemical Resource Description Framework (RDF) using text and image for structures of relevant to Acquired Immune Deficiency Syndrome (AIDS). These steps are (a) acquisition of the data on drugs, (b) definition of the framework to establish RDF on drugs using commonly asked questions during a drug discovery effort, (c) annotation of the structural data on drugs into RDF using the framework established in step (b), (d) validation of the annotation methods using Semantic Web concepts and tools, (e) design and development of public Web to distribute data to the public, (f) generation and distribution of data using OWL (Web Ontology Language). This paper describes this effort, discusses our observations and announces the availability of the OWL model at the W3C Web site (http://esw.w3.org/topic/HCLS/ChemicalTaxonomiesUseCase). The style of this paper is chosen so as to cover a broad audience including structural biologists, medicinal chemists, and information technologists and at times may appear say to the obvious for certain experts. A full discussion of our method and its comparison to other published methods is beyond the scope of this publication.

Key Words: Semantic Web, RDF, Chemical Taxonomy, HIV, AIDS, Structural data, Chemical ontology, Chem-BLAST, Chemical Genomics, Intelligent Web, XML, RDF, OWL, URL: AIDS database http://bioinfo.nist.gov/SemanticWeb_hiv3d/chemblast.do and http://xpdb.nist.gov/hivsdb/hivsdb.html Contact: Bhat@nist.gov.

INTRODUTION

The size of chemical databases has grown exponentially in recent years and there is an urgent need for novel data representation techniques and user-friendly search engines that enable the effective use of these data resources. There is considerable interest in the RDF and the Semantic Web [1] as a possible solution to the growing information technology needs of biological [2] healthcare [3-14] and drug discovery [15-17] related data. The Semantic Web is proposed by W3C (http://ww.w3.org/) as a vision for the future of the Web in which information is given explicit meaning, making it easier for humans to find what they want by enabling machines to process information available on the Web. Designing and populating RDF to provide answers to commonly asked questions by users of a Web is a key step towards the establishing a Semantic Web. Recently, there have been several Web-based efforts reported (e.g. ChEBI, http://www.ebi.ac.uk/chebi/; SMID http://smid.blueprint.org/smid_about.php) towards this goal for biological and chemical data. Here we discuss our efforts towards this goal in the context of the structural data represented either as text-based or image-based data and distributed by an AIDS structural database (http://bioinfo.nist.gov/SemanticWeb_hiv3d/chemblast.do). The database is developed by NIST in collaboration with NIAID, NCI and Rutgers University.

A full discussion on Semantic Web efforts world-wide is beyond the scope of this paper and here we refer readers to several references provided above and to a quote from BusinessWeek report dated April 2007 -'The Project10X study found that semantic tools are being developed by more than 190 companies, including Adobe (ADBE), AT&T (T), Google (GOOG), Hewlett-Packard (HPQ), Oracle (ORCL), and Sony (SNE)' (http://www.businessweek.com/technology/content/apr2007/tc20070409_248062.htm) .

Chemical structures are complicated to be intuitively defined using text-based descriptors. For this reason, scientists and textbooks have been using images (schematic drawing of the bonds that connect atoms of the chemical compound) to present them and to describe their relationships to other structures. Weaning through a large number of such images over a Web to locate structures of interest may become overwhelming for users. Organization of the information into a taxonomy tree [18] is one of the familiar concepts used to streamline the text-based information presented to a user over the Web. Here we illustrate how this technique could be extended to present images of components of chemical structures. First we describe the rules that we used to model chemical structures and their components (elements of RDF) into a taxonomical tree using

*Address correspondence to this author at the Biochemical Science Division (831), NIST, Gaithersburg, MD 20899 USA; Tel: 301 975 5448;
E-mail: bhat@nist.gov

commonly known concepts and apply them to AIDS inhibitors. Then we model the elements of taxonomy using OWL, RDF and design and develop a public Web page. We also illustrate the limited use of Protégé for this modeling.

METHOD

The key elements of the method that we used are (a) acquisition of the structural data on drugs from the AIDS database, (b) definition of the rules to establish RDF for drugs, (c) annotation of the drugs into RDF, (d) validation of the annotation process using Semantic Web tools, (e) designing and development of public Web site using conventional databases such as Oracle 9i to distribute data to the public, (f) generation and distribution of OWL and modeling using Protégé.

Steps (a-d) define the rules and processes of establishing RDF and steps (e-f) establish query interface, modeling and distribution of the data to the public.

STEP (A)- ACQUISITION OF THE DATA

All the data used in the work described here were obtained from the HIV structural database (HIVSDB http://xpdb.nist.gov/hivsdb/hivsdb.html) [19, 20]. HIVSDB groups structures into two main categories, one that was determined by X-ray or NMR studies, and the other generated from chemical connectivity information. Results from X-ray or NMR studies come with three-dimensional coordinate information of the molecules (3-D data), whereas the ones generated from chemical connectivity have only two-dimensional co-ordinate information (2-D data). The 3-D data itself has two components, one that describes the AIDS protein such as the HIV -1 protease and the other that describes an inhibitor of a protein such as inhibitor of the HIV -1 protease. This inhibitor is called an AIDS drug and presentation of both its structure and its components in a Web using elements of RDF is the subject of this paper. For the sake of simplicity, we chose to ignore the structural information on proteins and chose to process both the 3-D and 2-D structural data on drugs in the same manner. Also, we chose to ignore the non-structural text-based data such as the binding constants of these drugs to their target enzymes.

STEP (B) – DEFINE THE FRAMEWORK TO ESTABLISH RDF

Within the context of the work described here, an RDF describes relationships between a pair of structural data using concepts derived from the questions commonly asked by users of the database. Some of these commonly asked questions are listed by us in the W3C (HCLS) Wiki Web site (http://esw.w3.org/topic/HCLS/ChemicalTaxonomiesUseCase). These questions may be grouped into two main types (1) structural biology and modeling view, and (2) a medicinal chemist's view.

1) A structural biology and modeling view postulates questions asked by a structural biologist and/or a modeler during a drug design effort involving the structural data.

2) A medicinal chemist view postulates questions asked by a medicinal chemist during a drug design cycle.

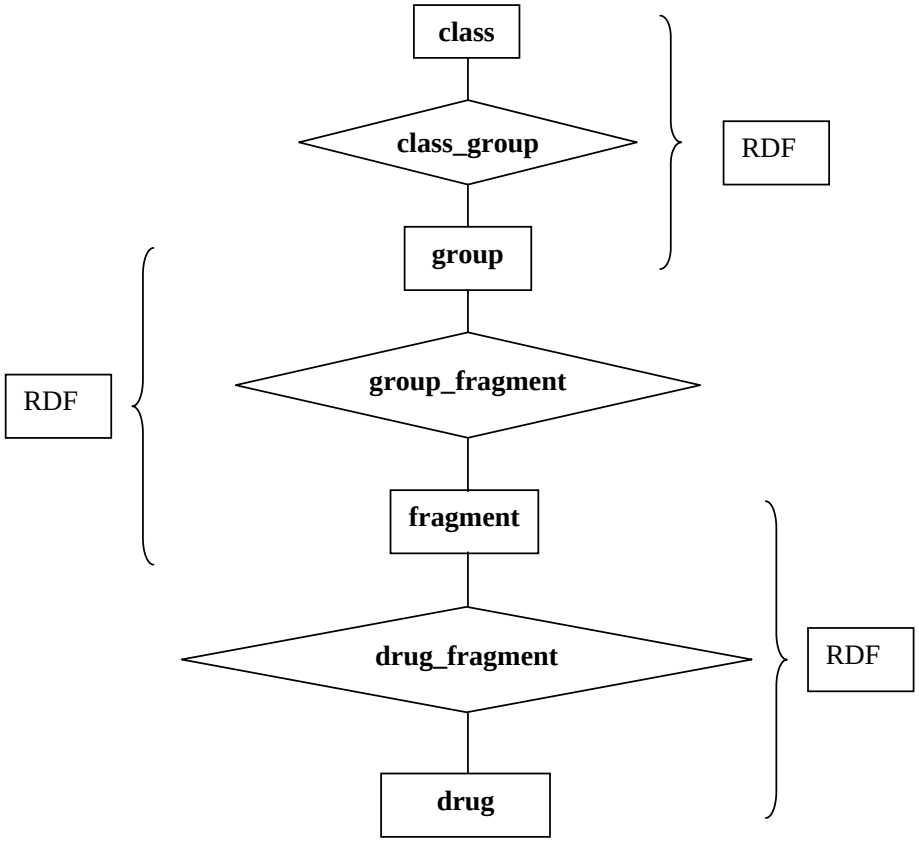

Fig. (1a). An RDF is made up of a subject related to an object by a predicate. Two successive rectangular shaped boxes show a pair of subject and object related by a predicate shown with a diamond shaped box between them. Inter-connected RDFs form a taxonomical tree.

These commonly asked questions are aimed at getting the most pertinent information from the database for the purpose of designing a drug on the basis of a use case. These questions form the framework for establishing the RDF in step (c).

STEP (C) – ANNOTATION OF THE DATA INTO RDF

Commonly asked questions are analyzed and all the data in the database are annotated and grouped [19] into RDF so as to obtain answers for these questions. An RDF relates a subject to an object through a predicate. An RDF in the context of this paper defines relationships (predicate) between the fragments (subject, element) and the drugs (object) held in the database to facilitate answers to commonly asked questions mentioned in step (c). For instance, to facilitate an answer to a question - what fragments are available for designing a drug, RDFs arrange fragments of each drug using the RDF syntax (drug name -> has fragment -> fragment name). In step (c) all the structures acquired in step (a) are decomposed into fragments (Fig. (**1a**)) and then to RDF. The process of establishing the RDF is repeated for as many commonly asked questions as possible.

STEP (D) – VALIDATING THE RDF USING PROTÉGÉ

Protégé is the de facto standard for validating Semantic Concepts. However, Protégé is featured to handle textual data whereas the chemical structures that are focus of this paper are more intuitive when represented as images. Further the taxonomy that we use allows classes to have multiple overlapping super classes (directed acyclic graph [21], a collection of multiple semi-joint trees) that may not be allowed by Protégé but they are commonly found among elements of a drug. In spite of this we would like to illustrate in part the modeling of chemical data (Figs. **1-3**) using Protégé for the sake of presenting the structural RDF using the familiar Semantic Web text-based tool - Protégé. Also, the modeling described here is sort of an exercise and a use-case modeling (searching (Fig. **4**) the database for structural elements of interest) is done by individual users using the Web tool - Chem-BLAST [19].

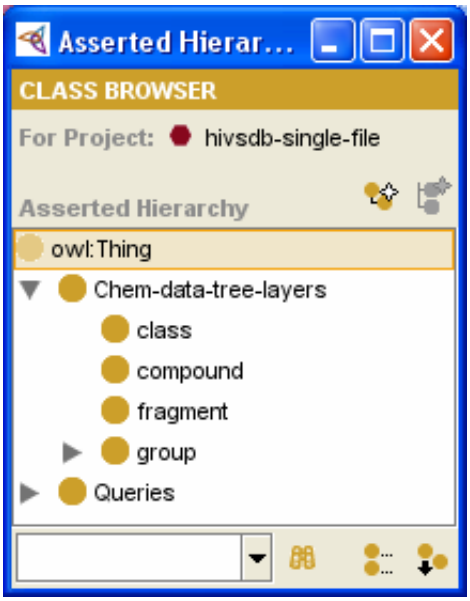

Fig. (1b). Chemical-data-tree layers as OWL Classes. The four OWL Classes (class, compound, fragment, and group) correspond to the four layers (class, compound or drug, fragment, and group) of a chemical taxonomy (chemical-data-tree-layers).

This hierarchy shows relationships between chemical-data-tree-layers defined as OWL properties. The class-group relation is defined by the has-group and member-of-class

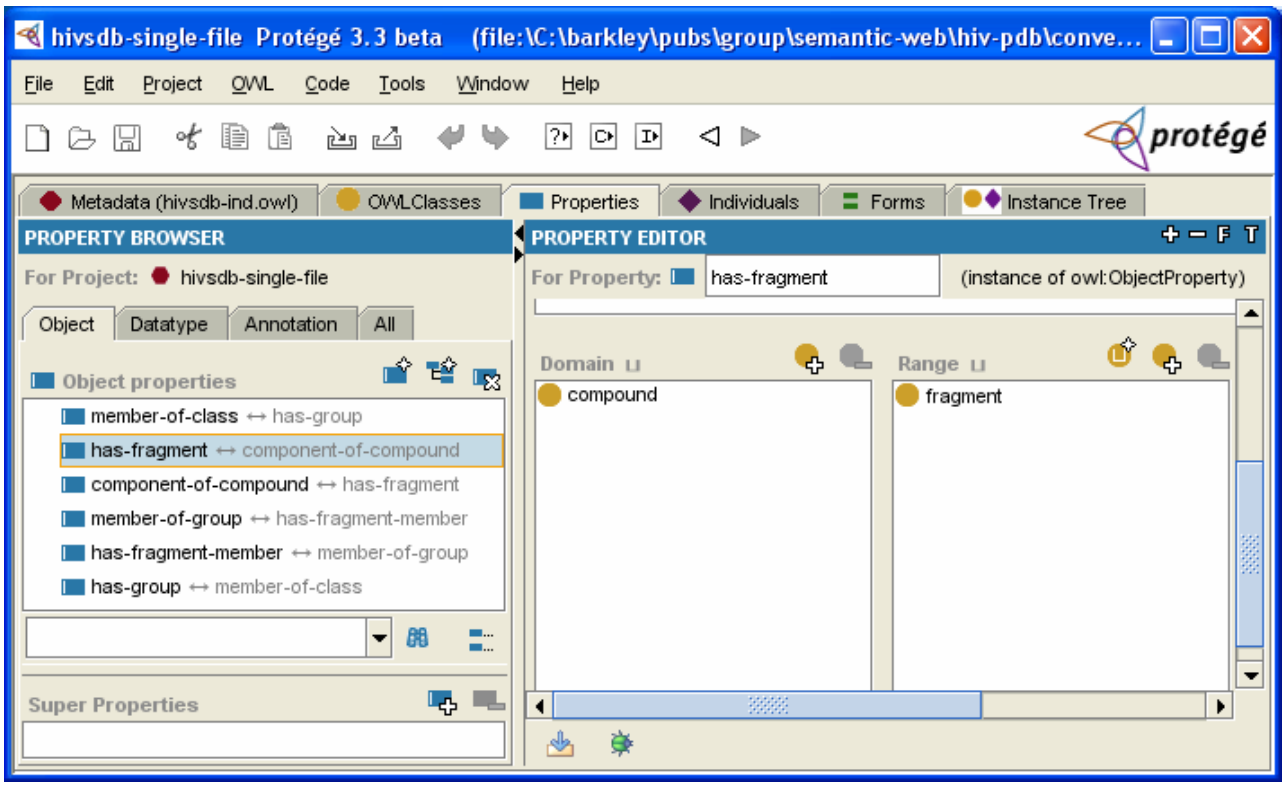

Fig. (2). Chemical-data-tree layer relations defined as OWL properties using Protégé.

properties; the group-fragment relation defined by the has-fragment-member and member-of-group properties; and the compound-fragment relation is defined by the has-fragment and component-of-compound properties.

Alternatively, a user may view (Fig. **4**) all the classes (left-most column of Fig. **4**) using the Chem-BLAST Web tool. A user may click on the seven-member-ring and then on its subgroup to view the pictures of these rings. On click-

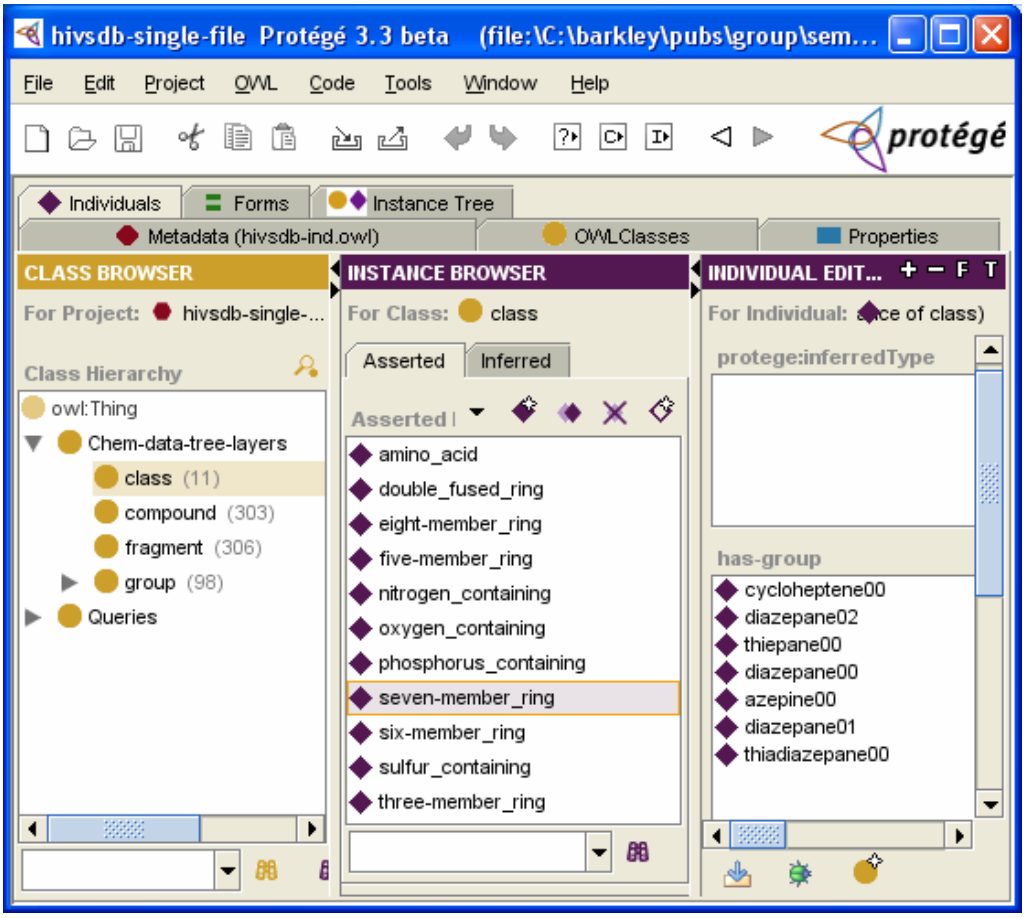

Fig. (3). Shows display of all the OWL Classes, their objects and associated members of one of the class object. The left-most column lists the four OWL Classes shown in Fig. **1**. The middle column lists the subjects belonging to the OWL Class 'class'. The right column lists the groups that are members of the class 'seven-member-ring'. It shows the seven groups in the seven-member-ring class of the chemical-data-tree. This screen is produced by Protégé and it is the equivalent of a query to find all groups of the seven-member-ring class.

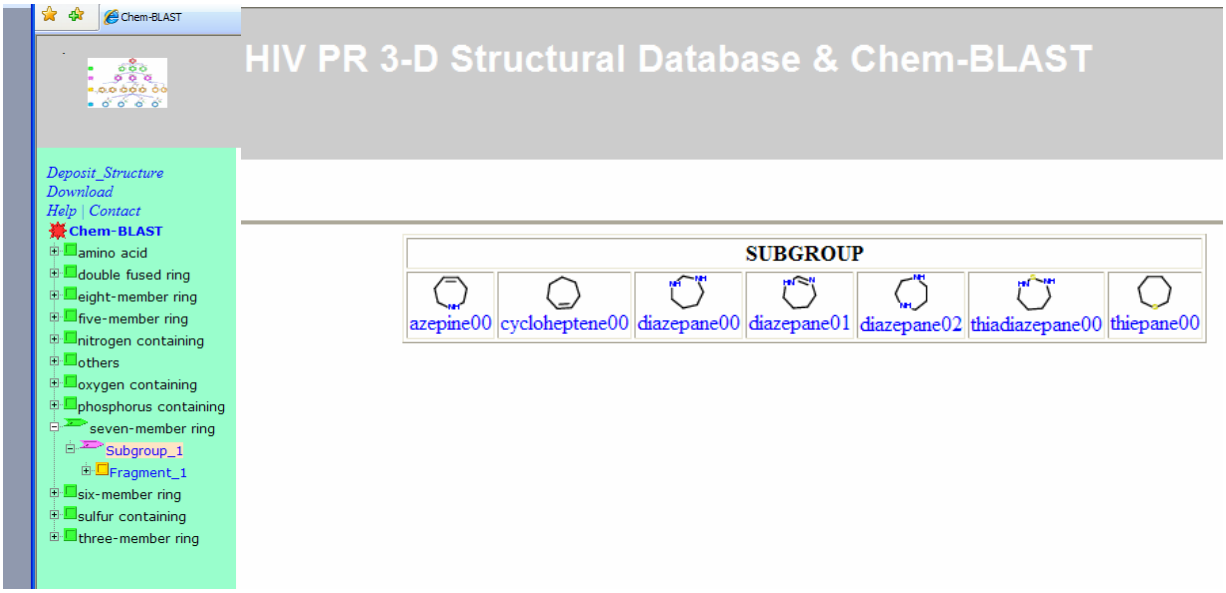

Fig. (4). Shows the Web tools – Chem –BLAST (http://bioinfo.nist.gov/SemanticWeb_pr3d/chemblast.do) for searching through the elements of RDF.

ing any one of the rings the Web tool will display the structure (structures) that contain the chosen subgroup.

Though Protégé allows many other advanced modeling on text-based identifiers, since the focus of this paper is the modeling on images, additional modeling using Protégé will not be presented.

STEP (E) – ESTABLISHING THE DATABASE

This step stores and catalogues each of the elements of the OWL representation in database tables. We chose two types of database tables, one that we call the relationship table, and the other the object table. The relationship table stores all the relationships between objects represented as OWL properties in a vertical stack and it provides simplicity to add, delete, and modify object relationships as needed. This table also provides a means of making simple queries about relationships among the objects. The object table has one column for each object in the chemical-data-tree. These tables allow the Web tool (Chem-BLAST) to display results from a query on objects using their images.

STEP (F) – GENERATION AND DISTRIBUTION OF OWL (WEB ONTOLOGY LANGUAGE)

In our implementation, OWL identifiers are the commonly used names of substructures of chemicals and the International Chemical Identifier of drugs (InChI as implemented in [20]) recommended by the International Union of Pure and Applied Chemistry (IUPAC). The InChI are assigned by automated procedures. InChI uses chemical connectivity of the atoms to assign an identifier and thus they are rule based. We replace 'space' and '/' of InChI by '_' and '_2FC' respectively to arrive at the Uniform Resource Identifier (URI). These features of the Web Ontology Language (OWL) presented here makes it possible both to assign new and to interpret existing URI in a distributed environment. The URIs are grouped into two types;) a semi-invariant ontologically defined URI (OURI) and invariant URI [22]; OURI denotes the ontologically defined elements (such as class, subgroup and fragment) of the RDF, whereas a URI identifies a complete drug with links to their biological properties. A semi-invariant URI may mean different things in different drugs and it may have different effects on the property of two drugs. For instance, a six-member-ring may lead to different properties depending on its relative position in the structure of the drug. The software used to assign these InChIs may be freely downloaded from the IUPAC Web site. The OWL for the AIDS drugs is downloadable from the Wiki site (http://esw.w3.org/topic/HCLS/ChemicalTaxonomiesUseCase) described above.

DISCUSSION

Here we describe the method that we used to develop the information model for HIVSDB – AIDS structural resource which was in part described previously by us [23]. Some of the important aspects in the development of this information model with emphasis on drug design are as follows. a) Develop a rule based on a dictionary of fragments and linkers (a chemical entity that connects two fragments) that may have specific semantics from the view point of structure-based drug discovery (a use case). b) Use this dictionary to develop all individual entities, class, class axioms relevant to the use case. c) Extend the dictionary, entities, OWL Class and OWL Class axioms to cover all compounds of interest for the use case. For instance, in a use case of drug-design, include both three-dimensional (X-ray or NMR) and two-dimensional structures (chemical structure described by atomic connectivity but no three-dimensional co-ordinates). d) Establish rules to define relationships between the entities to serve the commonly asked questions for the use case, and use these rules to arrange entities in a chemical-data-tree. e) Verify the integrity of the dictionary, data, and taxonomy by information and model design, Semantic Web tools, and metrics; make changes to the model if necessary. f) Transform the taxonomy into a database table and use database and Web tools to query and display data in a Semantic context.

This method modifies the focus of the traditional data uniformity step [24] to include both the generation of name synonyms and rule-based data organization and OWL with machine enforceable cross-indexing between the elements to form a taxonomical tree This annotation step emphasizes an additional aspect of data annotation - the organization of the data and cross-indexing of elements common to drugs in a format that is amenable to machine reasoning using a logic that is directly applicable to user's queries. This method is different from the methods [20, 25] of cross-indexing data using a unique but universal property of the data that may or may not be applicable to the use case. The method of generating fragment-based chemical-data-tree focused on a use case outlined above is one of the important differences to other approaches (ChEBI, http://www.ebi.ac.uk/chebi/; SMID http://smid.blueprint.org/smid_about.php)

RDF (Resource Description Framework, http://www.w3.org/RDF/) and OWL (Web Ontology Language, http://www.w3.org/2004/OWL/) are commonly used in Semantic Web technology. They enable information resources to interoperate on the Semantic Web. Here, we illustrate the use of OWL to represent the biological and structural data resource HIVSDB [19, 20] (http://xpdb.nist.gov/hivsdb/hivsdb.html), that is the Standard Reference Database 102 at the National Institute of Standards and Technology. The HIVSDB facilitates query and comparison of inhibitors of an AIDS target enzyme, HIV-1 protease, which is targeted by about half of the clinically used AIDS drugs. Finding the cure for AIDS is still a work in progress and thus resources such as the HIVSDB play a critical role in fighting this global epidemic health issue. HIVSDB has over 2000 compounds and has the largest collection of 3-D structures of HIV-1 protease complexes available in the public domain. HIVSDB also has information on biological and antiviral data related to most of these drugs.

Establishing an OWL representation begins with design of a use case. A use case (http://esw.w3.org/topic/HCLS/ChemicalTaxonomiesUseCase) is developed by analyzing the needs of users of the Web resource. Some of these needs are discussed below.

A structural biologist and a modeler design drugs by studying enzyme drug interactions using a computer generated model built using experimental X-ray or NMR data. The popular paradigm in drug discovery seeks to collect, compare, and test many chemically similar compounds built by linking several building blocks generally known as fragments. This approach to modern drug discovery [26] is rational and knowledge-based with a defined hypothesis on the functional role of individual fragments that make up a drug.

The process of drug design, therefore, begins with a lead compound followed by a hypothesis on how different fragments of this compound interact with the amino acid residues of the target protein molecule [27]. Following this, database searches are performed to gather 'structural neighbors' of the lead compounds using what is commonly known as 'mix and match method of building the functional fragments' of a lead compound. The suitability of these fragments that make up a lead compound is established using fragment-based *in silicio* modeling or X-ray crystallographic screening techniques [28, 29]. The structural biologist and modeler would like to know: what fragments are available for designing a new drug; what fragment of a given drug has been used in other drugs; and which other known drugs are structurally related to a given structure? The RDF presented here facilitate answers to such questions.

Often medicinal chemists and structural biologists may have different views of a drug. While structural biologists may be thinking of drugs in terms of fragments that provide complementary interaction to the enzyme at specific structural pockets, medicinal chemists may think of compounds in terms of the type of rings involved such as a phenyl group. These types of questions form the basis of the rules of establishing a use case for drug design. The implementation of the semantics for this use case for chemical-data-tree was first done using RDB (relational database – ORACLE) tools and then with Semantic Web tools.

Search engines were developed using Perl to present the chemical-data-tree and related structural and biological data in a series of steps. In each of these steps a user chooses a structural feature (chosen from classes, subgroups or fragments) of his interest from the many possibilities. Different data relationships established by the chemical-data-tree are used by the search engines to produce answers to questions related to drug discovery (http://bioinfo.nist.gov/Semantic Web_pr3d/chemblast.do).

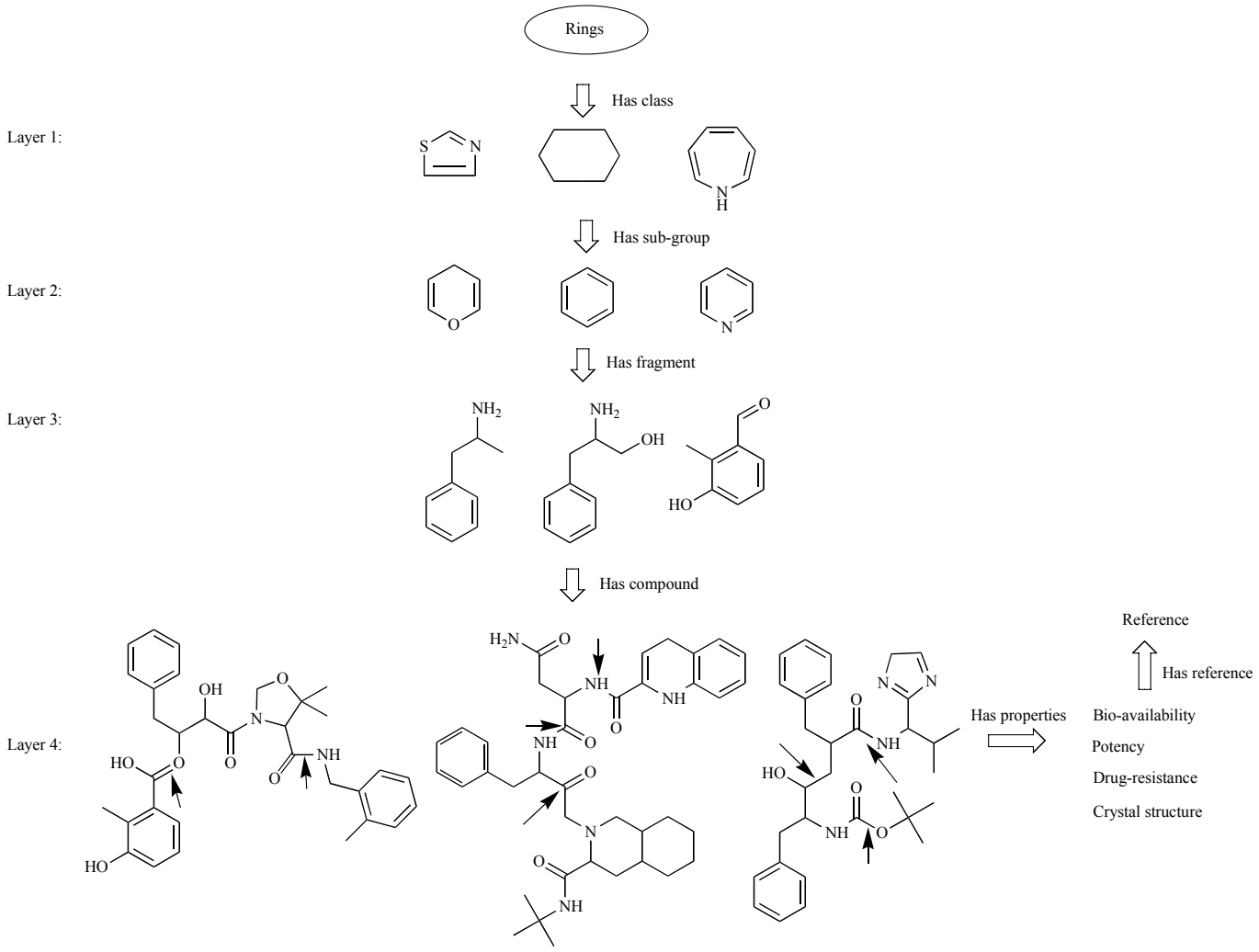

Fig. (5). Chemical-data-tree from HIV Structural Database. Elements of adjacent layers are used to generate OWL. Compounds are cleaved into fragments using the cleavage points (shown by arrows) defined in a dictionary. These fragments are then arranged into layers 1-3 using another dictionary of sub-group and classes for these fragments. Having developed the chemical-data-tree information model, the model and associated data are stored in relational database tables using ORACLE or MySQL.

CONCLUSION

In this paper we describe the steps that we used to build chemical structural RDF for inhibitors of interest to AIDS research. We illustrated how the structural data may be modeled using Protégé and queried by Chem-BLAST – a publicly available Web tool developed by us that uses features of Oracle 9i for storing and querying the data. The data is also downloadable from the Website using OWL. Chem-BLAST is designed for chemical structures and their components that are represented as images. In our application we use OWL and relational database (ORACLE, MYSQL). This choice in no way may be used to imply that we consider these choices superior to others such as Unified Modeling Language (UML) or a commercial database for RDF.

The method described here for representing chemical structures in RDF and then querying is different from the SMILES [30] technology for querying chemical structures. The SMILES technology (used by a most of chemical databases) establishes entity relationships at the time of query on the Web, whereas Chem-BLAST works on pre-defined relationships coded as elements of RDF. These pre-defined relationships make it easier for Web tools to allow users to step through elements of RDF to query the contents of the database while visualizing the elements through pre-defined images of the structures denoted by the element. However, it can be argued that the capacity to define entity relationships at run time using SMILES has its own advantage. Ideally, it is desirable to provide both SMILES and Chem-BLAST based features in a Web environment.

Elements of chemical structures may have overlapping structural features (such as in phenyl and naphthalene), a feature that is permitted by OWL and RDF. Though Protégé does not permit this relationship, we chose Protégé to illustrate the general concepts of modeling chemical structural elements. However, this work should not be interpreted to imply that Protégé permits complete modeling of chemical structural elements. In fact, our web page represents structures using their images and Protégé is designed to view only text-based elements and thus the use of Protégé is limited for those who are able to mentally map text-based names to their structures.

We chose a conventional database system instead of RDF based ones for two reasons; (1) our database design and Web development pre-dates the availability of databases specifically designed to handle RDF; (2) the representation of the structural data as images is more intuitive than their representation as text. Therefore we developed special Web tools to enable query on these images that work on ontology-based conventional database system. We distribute data using OWL; instead, the data could have been distributed using entity/relationship model or a UML model.

ACKNOWLEDGEMENTS

This work was in part supported by the SIMA program and some of the Web tools were developed in part by Anh-Dao Nguyen. We thank Anne Plant for her interest and support.

DISCLAIMER

Certain trade and company products are identified in this article to specify adequately the computer products used in the example. In no case does such identification imply endorsement by the National Institute of Standards and Technology (NIST), nor does it imply that the products are necessarily the best available for the purpose.

REFERENCES

[1] H. J. Berners-Lee T, O. Lassila, "The Semantic Web," *Scientific American,* vol. 284, pp. 34 - 43, may 2001.

[2] T. Clark, "Looking Ahead: The Semantic Web in Life Science," *Genome Technology,* vol. 12, pp. 23 -24, Jan/Feb 2005.

[3] M. Boniface, E. R. Watkins, A. Saleh, A. Dogac, and M. Eichelberg, "A secure semantic interoperability infrastructure for inter-enterprise sharing of electronic healthcare records," *Stud Health Technol Inform,* vol. 120, pp. 225-35, 2006.

[4] M. Boniface and P. Wilken, "ARTEMIS: towards a secure interoperability infrastructure for healthcare information systems," *Stud Health Technol Inform,* vol. 112, pp. 181-9, 2005.

[5] D. Budgen, M. Turner, I. Kotsiopoulos, F. Zhu, M. Russell, M. Rigby, K. Bennett, P. Brereton, J. Keane, and P. Layzell, "Managing healthcare information: the role of the broker," *Stud Health Technol Inform,* vol. 112, pp. 3-16, 2005.

[6] H. Gu, L. M. Liu, M. Halper, J. Geller, and Y. Perl, "Converting an integrated hospital formulary into an object-oriented database representation," *Proc AMIA Symp,* vol. pp. 770-4, 1998.

[7] V. Jagannathan, "Experimenting with semantic web services to understand the role of NLP technologies in healthcare," *AMIA Annu Symp Proc,* vol. pp. 966, 2006.

[8] M. N. Kamel Boulos, F. E. Harvey, A. V. Roudsari, R. Bellazzi, M. E. Hernando, T. Deutsch, D. G. Cramp, and E. R. Carson, "A proposed semantic framework for diabetes education content management, customisation and delivery within the M2DM project," *Comput Methods Programs Biomed,* vol. 83, pp. 188-97, Sep 2006.

[9] T. Moller, H. Schuldt, A. Gerber, and M. Klusch, "Next-generation applications in healthcare digital libraries using semantic service composition and coordination," *Health Informatics J,* vol. 12, pp. 107-19, Jun 2006.

[10] M. Muller, K. Marko, P. Daumke, J. Paetzold, A. Roesner, and R. Klar, "Biomedical data mining in clinical routine: expanding the impact of hospital information systems," *Medinfo,* vol. 12, pp. 340-4, 2007.

[11] E. Neumann and L. Prusak, "Knowledge networks in the age of the Semantic Web," *Brief Bioinform,* vol. 8, pp. 141-9, May 2007.

[12] I. Roman, L. M. Roa, L. J. Reina, and G. Madinabeitia, "Standardization of demographic service for a federated healthcare environment," *Stud Health Technol Inform,* vol. 114, pp. 259-69, 2005.

[13] L. A. Slaughter, D. Soergel, and T. C. Rindflesch, "Semantic representation of consumer questions and physician answers," *Int J Med Inform,* vol. 75, pp. 513-29, Jul 2006.

[14] C. Weng, J. H. Gennari, and D. B. Fridsma, "User-centered semantic harmonization: a case study," *J Biomed Inform,* vol. 40, pp. 353-64, Jun 2007.

[15] J. S. MacNeil, "Diving Deep Into The Chemical Genome," *Genome Technology,* pp. 26 - 35, Jan/Feb 2005.

[16] C. P. Austin, L. S. Brady, T. R. Insel, and F. S. Collins, "NIH Molecular Libraries Initiative," *Science,* vol. 306, pp. 1138-9, Nov 12 2004.

[17] H. I. Feigenbaum L, T. Hongsermeier, E. Neumann, S. Stephens, "The Semantic Web in Action," *Scientific American,* vol. 297, pp. 90 - 97, Dec 2007 2007.

[18] D. L. Wheeler, T. Barrett, D. A. Benson, S. H. Bryant, K. Canese, V. Chetvernin, D. M. Church, M. DiCuccio, R. Edgar, S. Federhen, L. Y. Geer, Y. Kapustin, O. Khovayko, D. Landsman, D. J. Lipman, T. L. Madden, D. R. Maglott, J. Ostell, V. Miller, K. D. Pruitt, G. D. Schuler, E. Sequeira, S. T. Sherry, K. Sirotkin, A. Souvorov, G. Starchenko, R. L. Tatusov, T. A. Tatusova, L. Wagner, and E. Yaschenko, "Database resources of the National Center for Biotechnology Information," *Nucleic Acids Res,* vol. 35, pp. D5-12, Jan 2007.

[19] M. D. Prasanna, J. Vondrasek, A. Wlodawer, H. Rodriguez, and T. N. Bhat, "Chemical compound navigator: a web-based chem-BLAST, chemical taxonomy-based search engine for browsing compounds," *Proteins,* vol. 63, pp. 907-17, Jun 1 2006.

[20] M. Prasanna, J. Vondrasek, A. Wlodawer, TN. Bhat, "Application of InChI to curate, index and query 3-D structures," *Proteins, Structure, Function, and Bioinformatics,* vol. 60, pp. 1-4, 2005.

[21] M. Graham and J. Kennedy, "Exploring multiple trees through DAG representations," *IEEE Trans Vis Comput Graph,* vol. 13, pp. 1294-301, Nov-Dec 2007.

[22] T. N. Bhat, J. Barkley, "Semantic Web for the Life Sciences - Hype, Why, How and Use Case for AIDS Inhibitors," *Proceedings of the IEEE Congress on Services,* vol. pp. 87 - 91, July 7 2007.

[23] M. D. Prasanna, J. Vondrasek, A. Wlodawer, H. Rodriguez, and T. N. Bhat, "Chemical Compound Navigator: A Web-based Chem-BLAST Search Engine for Browsing Compounds," *PROTEINS: Structure, Function, and Bioinformatics,* vol. 63, pp. 907-917, 2006.

[24] T. N. Bhat, P. E. Bourne, Z. Feng, G. L. Gilliland, S. Jain, V. Ravichandran, B. Schneider, K. Schneider, N. Thanki, H. Weissig, J. Westbrook, and H. M. Berman, "The PDB data uniformity project," *Nucleic Acids Res,* vol. 29, pp. 214-8, Jan 1 2001.

[25] P. Murray-Rust, H. S. Rzepa, S. M. Tyrrell, and Y. Zhang, "Representation and use of chemistry in the global electronic age," *Org Biomol Chem,* vol. 2, pp. 3192-203, Nov 21 2004.

[26] J. Drews, "Drug discovery: a historical perspective.," *Science,* vol. 287, pp. 1960 -1964, 2000.

[27] I. Kuntz, "Structure-based strategies for drug design and discovery," *Science,* vol. 257, pp. 1078 - 1082, 1992.

[28] T. L. Blundell, B. L. Sibanda, R. W. Montalvao, S. Brewerton, V. Chelliah, C. L. Worth, N. J. Harmer, O. Davies, and D. Burke, "Structural biology and bioinformatics in drug design: opportunities and challenges for target identification and lead discovery," *Philos Trans R Soc Lond B Biol Sci,* vol. 361, pp. 413-23, Mar 29 2006.

[29] C. Rummey, S. Nordhoff, M. Thiemann, and G. Metz, "In silico fragment-based discovery of DPP-IV S1 pocket binders," *Bioorg Med Chem Lett,* vol. 16, pp. 1405-9, Mar 1 2006.

[30] D. Weininger, "SMILES, a Chemical Language and Information System 1. Introduction to Methodology and Encoding Rules,," *J.Chem. Inf. Comput. Sci.,* vol. 28, pp. 31 - 36, 1988.

Received: April 7, 2008 Revised: May 23, 2008 Accepted: May 26, 2008

© Bhat and Barkley; Licensee *Bentham Open.*
This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.5/), which permits unrestrictive use, distribution, and reproduction in any medium, provided the original work is properly cited.