IMMUNOME RESEARCH

RESEARCH Open Access

A Model for Collaborative Curation, The IEDB and ChEBI Curation of Non-peptidic Epitopes

Randi Vita[1*], Bjoern Peters[1], Zara Josephs[2], Paula de Matos[2], Marcus Ennis[2], Steve Turner[2], Christoph Steinbeck[2], Emily Seymour[1], Laura Zarebski[1], and Alessandro Sette[1]

Abstract

The Immune Epitope Database (IEDB) recently expanded and enhanced its non-peptidic epitope related data utilizing a collaboration with Chemical Entities of Biological Interest (ChEBI), resulting in the first resource that brings together published immunological data with the expertise of the ChEBI database. This procedure took advantage of the distinct expertise of the IEDB and ChEBI databases to improve content and enhance interoperability of both databases. This project has resulted in the comprehensive inventory and curation of immune epitope data related to non-peptidic structures and serves as a model for successful collaborative curation between established resources.

Introduction

The important discoveries that make up the scientific literature are hidden within published manuscripts and laboratory notebooks. In order to be fully utilized, this data needs to be converted into the standardized formats of databases and these formats need to be compatible. There exists a need for experts in each field of curation in order to understand the data being curated. This specialized training can be beneficial across different research specialties if the individual databases collaborate and cooperate. Currently, several collaborative standards exist, such as the Open Biomedical Ontologies (OBO) Foundry [1] and Minimum Information for Biological and Biomedical Investigations (MIBBI) [2]. Here we present a novel collaborative curation method that takes advantage of the specialized curators of two very different databases; the IEDB and ChEBI. This collaboration resulted in enhancements to both datasets and provided new depth and insights to the data being curated.

The IEDB contains data related to antibody and T cell epitopes for humans, non-human primates, rodents, and other animal species [3]. To date, curation of peptidic epitope data relating to infectious diseases, allergens, and autoimmunity is complete; resulting in the manual curation of 11,641 published manuscripts. Throughout this experience, the IEDB has undergone major revision and expanded its database fields, features, and curation guidelines to ensure accurate, thorough, and consistent curation of peptidic epitope data. However, data related to non-peptidic epitopes such as carbohydrates, lipids, and chemicals were largely uncurated. In order to expand the content of the IEDB to include non-peptidic epitopes we reached out to the experts studying these epitopes and began a fruitful collaboration with the ChEBI resource.

Results

Inventory of the literature describing non-peptidic immune epitopes

As a first step towards making non-peptidic epitope data available to the immunological community and scientific community in general, papers potentially containing data that could be curated and entered in the IEDB were inventoried. To this end we adapted an approach already described and validated elsewhere [4]. Briefly, the PubMed database was searched utilizing a query, purposely designed to be very broad and thus inclusive. The results of this query were then narrowed to select potentially relevant papers, by the use of an automated text classifier. This automated classifier was trained on 20,910 abstracts that were manually assigned to be curatable or not by a domain expert. The automated classifier is used to discard all references that with 95% confidence do not contain curatable information. The remaining references are manually reviewed by a domain expert to select the curatable ones. These manual assignments are used to continuously update the automated classifier, resulting in its continuous improvement [5]. As of October 1st 2010,

[1]La Jolla Institute for Allergy and Immunology, Vaccine Discovery, La Jolla, California, USA. [2]Chemoinformatics and Metabolism Team, European Bioinformatics Institute, Hinxton, Cambridge, UK
[*]Corresponding author: rvita@liai.org

Bjoern Peters bpeters@liai.org
Zara Josephs zjosephs@ebi.ac.uk
Paula de Matos pmatos@ebi.ac.uk
Marcus Ennis mennis@ebi.ac.uk
Steve Turner steve.turner@ebi.ac.uk
Christoph Steinbeck christoph.steinbeck@ebi.ac.uk
Emily Seymour eseymour@liai.org
Laura Zarebski laura@liai.org
Alessandro Sette alex@liai.org

a total of 27,636 potentially relevant references were identified and categorized as a function of subject matter. Of those, 2,642 related to HIV research were not scrutinized further, as HIV is currently outside of the scope of the IEDB. The remaining 24,994 references categorized as described in Davies et al [4] were separated based on whether they described peptidic or non-peptidic epitopes. As shown in Figure 1, it was found that 20% of the references identified in PubMed describe non-peptidic epitopes. While the majority (80%) of the literature describes peptidic epitopes, this large proportion of non-peptidic epitopes was notable and unexpected.

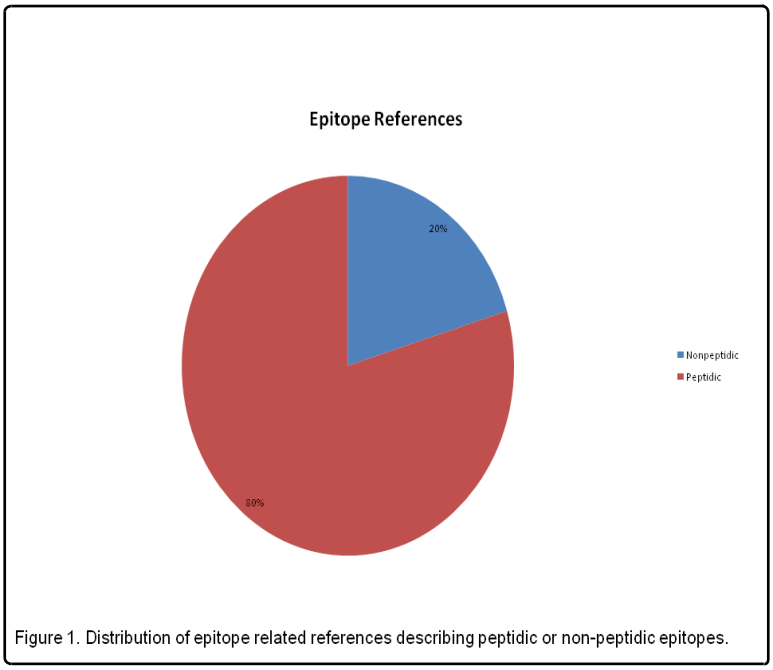


Figure 1. Distribution of epitope related references describing peptidic or non-peptidic epitopes.

Next, we probed in more detail the nature of the non-peptidic epitopes described in these references [Figure 2]. As done previously for peptidic epitopes, the references were classified first and foremost on the basis of their association to particular diseases and biological processes, and then also on the basis of the chemical structure of the epitope itself. Approximately half (56%) of all non-peptidic epitope references relate to specific diseases, such as allergic diseases (776 references; 15%), autoimmunity (648; 13 %), infectious diseases (517; 10%), transplantation (341; 7%) or cancer (542; 11%). Non-peptidic epitopes related to allergic reactions mostly include molecules involved in contact dermatitis, such as nickel or haptens used in experimental models of allergy. Non-peptidic epitopes identified in autoimmunity research are mainly lipids and glycolipids. Example structures include cardiolipin and phosphatidic acid. Infectious disease references mostly describe carbohydrate and glycolipid epitopes such as LPS and glycosylated proteins. A number of non-peptidic epitopes are discussed in references related to transplantation, including for example complex carbohydrate blood sugar antigens and galactose residues involved in xenotransplantation. Epitopes predominantly recognized in cancer research include the Tn, Thomsen-Friedenreich, and Lewis antigens, which are also implicated in autoimmunity and transplantation.

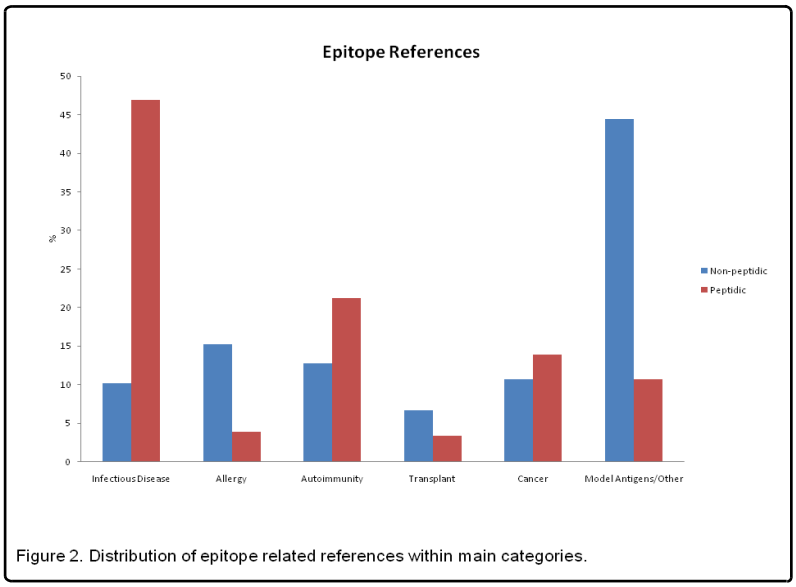


Figure 2. Distribution of epitope related references within main categories.

However, the largest category of non–peptidic references (44%) describe haptens and carbohydrate moieties not directly associated with a particular disease. These include model epitopes/antigens that are used to study fundamental mechanisms of immunity (TNP, DNP, ABA, NIP) and other small molecules such as natural hormones, narcotics, drugs and their metabolites, pollutants, poisons, and other assorted small molecules for which detection assays are being developed.

The relative distribution of references relating to non-peptidic epitopes in comparison to references describing peptidic epitopes also revealed some striking differences. Most notably about 46% of peptidic references are related to infectious diseases, compared to about 10% for non-peptidic. Conversely, only about 13% of the peptidic references are related to model antigens and other molecules, compared to about 44% for non-peptidic. These large differences in distribution are likely the reflection of the widespread use of well-defined small haptens to study immune responses, especially in early immunological literature, combined with the technical challenges associated with the exact definition of non-peptidic epitopes recognized in infectious diseases. In

conclusion, a large fraction of the immunological literature describes epitopes that are non-peptidic.

Representing non-peptidic data in the IEDB

Having identified references within the IEDB scope and their relative subject matter, we next determined to what extent the existing database structure was amenable to the representation of non-peptidic epitopes and what changes or modifications might be necessary. By comparison, describing peptide epitopes is relatively straightforward, with the ability to describe both continuous and discontinuous epitopes using the standard single letter code for amino acid sequences. In terms of nomenclature, in the case of peptidic epitopes simple BLAST searches allow assignment of source proteins, taking advantage of existing web resources, such as Genbank.

We began by contacting approximately 50 researchers who had authored publications describing non-peptidic epitopes. Two main points emerged from their collective feedback. First, since the types of molecules studied included a wide variety of chemical entities such as polysaccharides, drugs, haptens, metals, and glycolipids, they recommended using classification systems including generic names and IUPAC. Second, search options could include generic names, InChI strings, CAS numbers, and the drawn structure. No modifications were deemed necessary to the existing database structure to represent the immune responses recognizing the epitopes, the host, immunization and assay-related fields.

Indeed, by far the most challenging aspect relating to the curation of non-peptidic epitopes is the representation of their molecular structure and identification through free text and alpha-numeric identifiers. The commonly used nomenclature is not uniform. In different references the same chemical entity can be described by its molecular formula, its simplified molecular input line entry specification (SMILES), its 3-D rendering, a common name, a commercial name or a variety of chemical nomenclatures such as IUPAC. Fortunately and coincidentally, the ChEBI initiative had already been generating a framework where IUPAC names, 3-D structures, a structural hierarchy, and a list of synonyms for non-peptidic structures can be provided for small molecules [6]. Our interest in linking and integrating the IEDB data to this type of information mirrored an interest from the ChEBI project, in linking immunological data to its database. Based on the strong potential for synergy, a formal collaboration began in June of 2009.

Establishment of an Effective Curation Process

The next task was to put in place a process for curation of the nonpeptidic epitope data. Figure 3 presents the work flow with initial review by the IEDB staff of a given reference to identify the immunogens, antigens, and epitopes. This information is then transferred to a dedicated ChEBI curator to generate new ChEBI entries, utilizing a shared spreadsheet format. The ChEBI curator then locates the

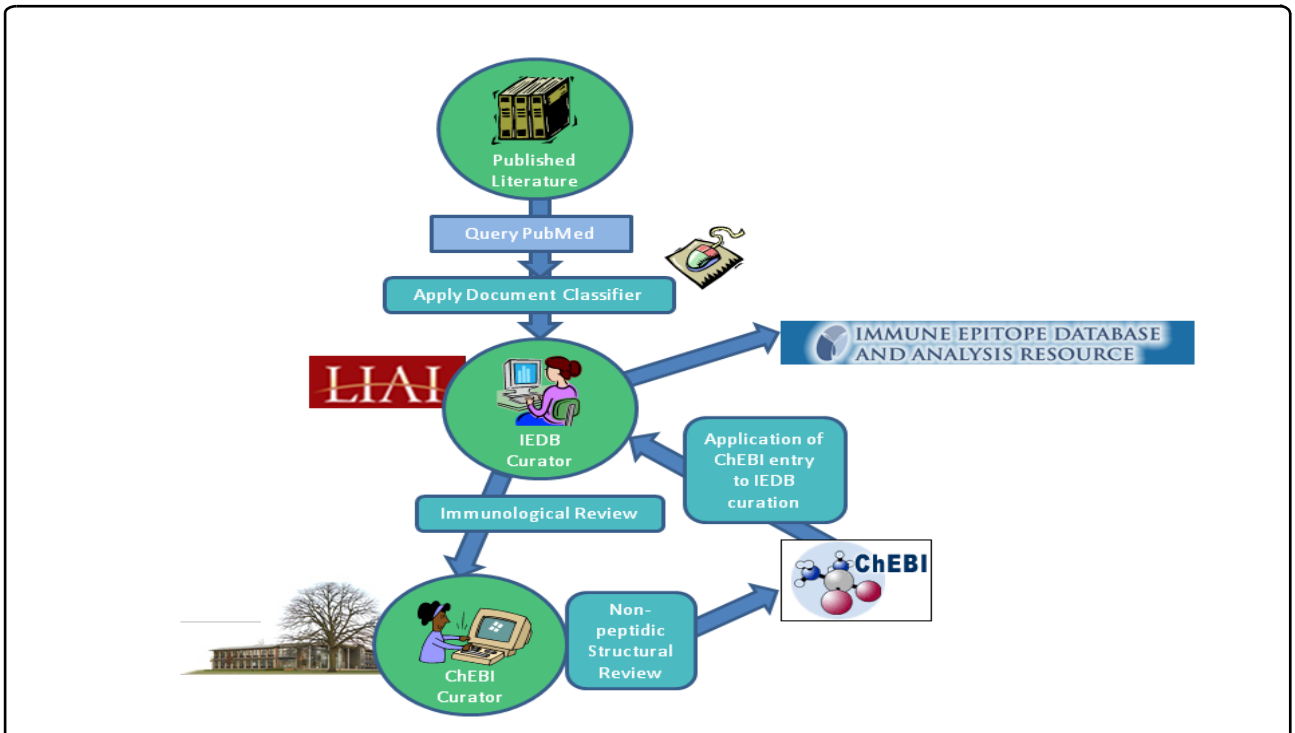


Figure 3. Work flow for IEDB:ChEBI collaboration.

structures the IEDB curator identified, and determines if it already exists on the ChEBI website. If it does, the existing entry is enhanced by including any new names or synonyms to which the manuscript referred, citation to the particular manuscript, the role(s) played by that entity in that manuscript, and updates any other information in the entry as needed. Alternatively, if the structure does not already exist on the ChEBI site, an entirely new entry is produced. The new information that has been added to ChEBI is released with each scheduled build of the ChEBI website. The ChEBI curator then supplies the IEDB curator with the correct ChEBI ID per requested structure via the shared spreadsheet. An example of a ChEBI entry curated on behalf of the IEDB is shown in Figure 4. In this figure, amoxicillin, a commonly recognized allergy epitope, is represented in the following formats: 3-D drawing, IUPAC International Chemical Identifier (InChI), SMILES, and chemical formula. Additional information present on the ChEBI website such as synonyms, brand names, and ontological relationships, as well as numerous external links is shown in Figures 5 and 6.

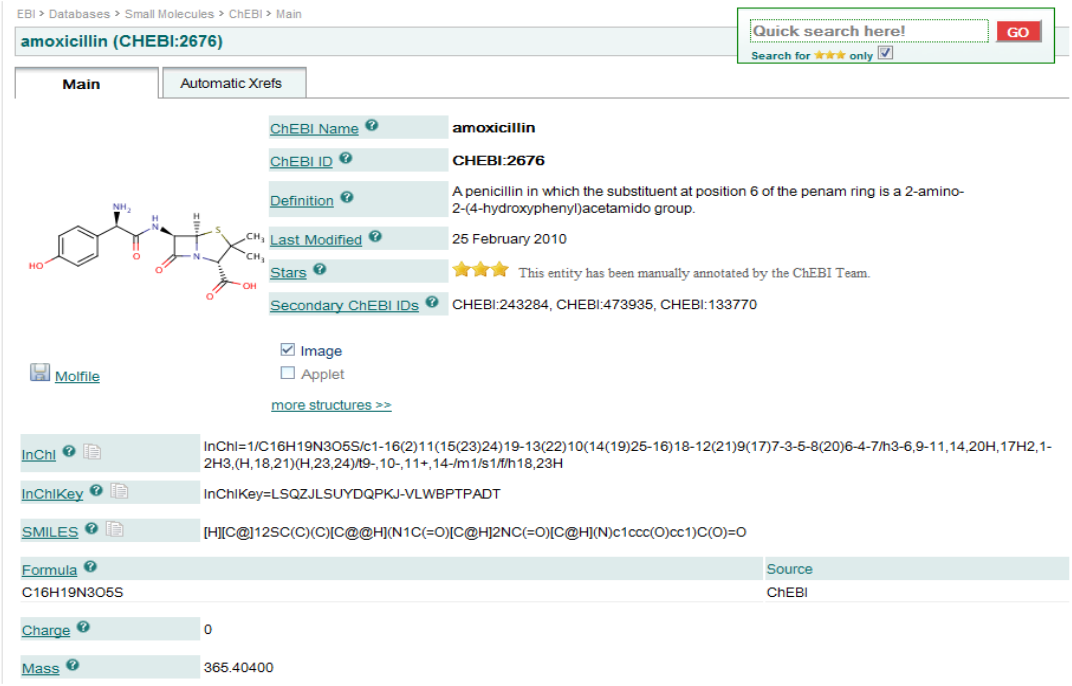


Figure 4. An Immune epitope as presented on the ChEBI website.

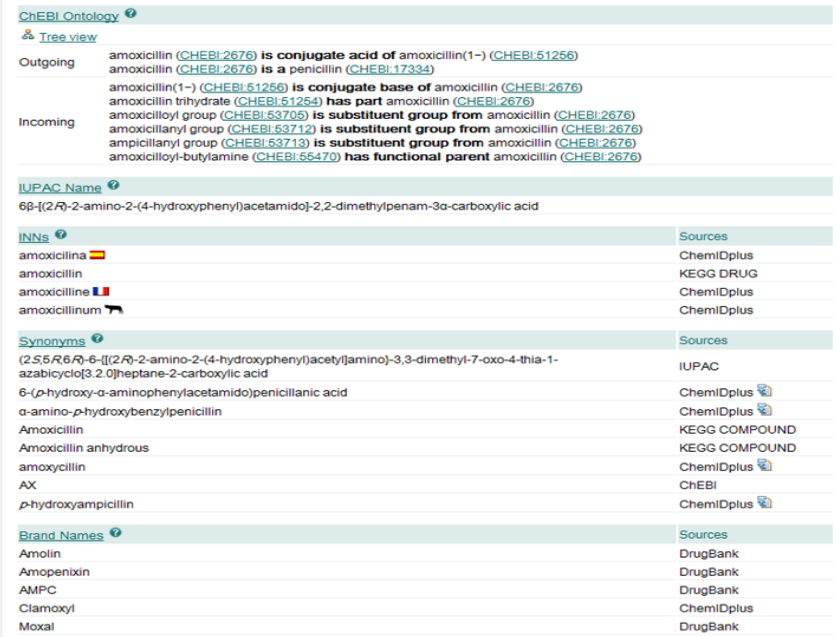


Figure 5. Synonyms, ontology, and alternate names provided on the ChEBI website.

Database Links	Databases
C06827	KEGG COMPOUND
D07452	KEGG DRUG
DB01060	DrugBank
DE1942693	Patent
GB1241844	Patent
GB978178	Patent
US3192198	Patent
243284	ChEMBL

Registry Numbers	Types	Sources
26787-78-0	CAS Registry Number	ChemIDplus
4274654	Beilstein Registry Number	Beilstein

Citations

Kogen H,Tago K,Arai M,Minami E,Masuda K,Akiyama T (1999)
A highly stereoselective synthesis of plaunotol and its thiourea derivatives as potent antibacterial agents against Helicobacter pylori.
Bioorganic & medicinal chemistry letters **9,** *1347-50 (Source: ChEMBL)* [PubMed:10360733]
[show Abstract]

Katsura Y,Tomishi T,Inoue Y,Sakane K,Matsumoto Y,Morinaga C,Ishikawa H,Takasugi H (2000)
Anti-Helicobacter pylori agents. 4. 2-(Substituted guanidino)-4-phenylthiazoles and some structurally rigid derivatives.
Journal of medicinal chemistry **43,** *3315-21 (Source: ChEMBL)* [PubMed:10966750]
[show Abstract]

Danelian E,Karlen A,Karlsson R,Winiwarter S,Hansson A,Lofas S,Lennernas H,Hamalainen MD (2000)
SPR biosensor studies of the direct interaction between 27 drugs and a liposome surface: correlation with fraction absorbed in humans.
Journal of medicinal chemistry **43,** *2083-6 (Source: ChEMBL)* [PubMed:10841786]

von Greyerz S,Bultemann G,Schnyder K,Burkhart C,Lotti B,Hari Y,Pichler WJ (2001)
Degeneracy and additional alloreactivity of drug-specific human alpha beta(+) T cell clones.
International immunology **13,** *877-85* [PubMed:11431418]
[show Abstract]

Ando R,Kawamura M,Chiba N (2001)
3-(Arylacetylamino)-N-methylbenzamides: a novel class of selective anti-Helicobacter pylori agents.
Journal of medicinal chemistry **44,** *4468-74 (Source: ChEMBL)* [PubMed:11728192]
[show Abstract]

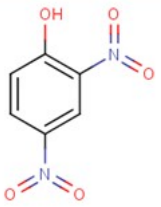
Zhao Z,Baldo BA,O'Brien RM,Plomley RF (2000)
Reaction with, and fine structural recognition of polyamines by human IgE antibodies.
Molecular immunology **37,** *233-40* [PubMed:10930630]
[show Abstract]

Figure 6. External links provided on the ChEBI website.

Epitope Information

Distinct Epitope	
Epitope ID:	112424
Name:	2,4-dinitrophenol
Accession:	CHEBI:42017
SMILES Structure:	Oc1ccc(cc1N(=O)=O)N(=O)=O

ChEBI Image



Reference (24) | B Cell Assay (14) | T Cell Assay (37)

24 item(s) found, displaying 1 to 24 (Click the column headers to adjust the sorting)
Export all results: (compact | full)

Reference ID	Author	Title	Abstract	Summary
322	Leo C James; Pietro Roversi; Dan S Tawfik	Antibody multispecificity mediated by conformational diversity.	A single antibody was shown to adopt different binding-site conformations and thereby bind unrelated...	Science, 2003
1005582	O Moriya; Y Ichikawa	Elicitation of delayed type hypersensitivity in chicks after in ovo sensitization with different molecular forms of the same hapten.	Lymphocyte sensitization, which participates in delayed type hypersensitivity (DTH) in chick embryos...	Microbiol Immunol, 1983
1005626	M C Ray; M D Tharp; T J Sullivan; R E Tigelaar	Contact hypersensitivity reactions to dinitrofluorobenzene mediated by monoclonal IgE anti-DNP antibodies.	Several studies have suggested a possible role for IgE antibodies in the pathogenesis of cutaneous h...	J Immunol, 1983
1007912	H Miyauchi; T Horio	A new animal model for contact dermatitis: the hairless guinea pig.	Allergic and irritant contact reactions were evaluated in the recently identified hairless guinea pi...	J Dermatol, 1992
1012055	Y Niwa; H Niwa; J Kohmura; D W Ou; M M Yokoyama	Studies on the dinitrochlorobenzene (DNCB) sensitization test.	One hundred and ninety-nine patients were divided into three groups and examined for their skin reac...	Ann Allergy, 1982
1012059	R J Dearman; J M Hegarty; I Kimber	Inhalation exposure of mice to trimellitic anhydride induces both IgG and IgE anti-hapten antibody.	The development of antibody responses resulting from inhalation exposure to chemical allergens has b...	Int Arch Allergy Appl Immunol, 1991
1012204	M Uehara; T Sawai	A longitudinal study of contact sensitivity in patients with atopic dermatitis.	Sensitization trials with dinitrochlorobenzene (DNCB) were performed in 150 patients with atopic der...	Arch Dermatol, 1989

Figure 7. Immune epitope as presented on the IEDB.

Once available, the IEDB reviews each structure and links the IEDB website with the ChEBI resource to download the required fields (generic name, IUPAC name, when available, synonyms, molecular structure (SMILES), parent classes), making the newly generated ChEBI structure selectable by IEDB curators via an internal 'Molecule Finder' application. Once the IEDB curation of the reference is completed, it appears on the IEDB's external website. An example of a ChEBI structure utilized by the IEDB is shown in Figure 7. Here 2,4-dinitrophenol (DNP), a commonly studied hapten epitope, is presented with its ChEBI image, SMILES, name, and ChEBI link, together with the IEDB curated immunological data.

Current Status of Reference curation

As of October 15th 2010, 770 references describing non-peptidic epitopes have been curated on the IEDB throughout this process. This involved 1,220 non-peptidic structures, the replacement of over 300 previously IEDB curated non-peptidic structures with ChEBI curated structures and the assignment of ChEBI parents to all IEDB structures to allow placement in ChEBI's ontological tree, as described in more detail below.

As seen in Table 1, the curation of allergy-related records is essentially complete, and approximately 30% of the references related to autoimmunity and infectious diseases have been curated. After completion of the references included in these two categories, transplant related references will be addressed followed by model antigen references. The curation of cancer epitope references is not currently within the scope of the IEDB. We envision that these activities will be essentially complete by the end of 2011.

Query and Display

The previous sections describe how a process for capturing non-peptidic epitope data was put into place and how the curation of references containing such information is in its advanced stages. In parallel, it became necessary to redesign the IEDB external site search capabilities, to allow searching for this newly available information. Previously, non-peptidic epitopes could be searched by a molecule finder which only allowed users to search for non-peptidic structures by name. This was problematic as the nomenclature of these structures is complex, variable, and full of synonyms. This issue was resolved by a newly developed non-peptidic molecule finder, where all synonyms and abbreviations are now searchable.

A further advance is the incorporation of a non-peptidic molecule tree that makes use of ChEBI's formal ontology. The ChEBI website provides a structural ontology, grouping together similar structures, such as all carbohydrates or all lipids in a hierarchical fashion. The IEDB incorporated this information in the form of a non-peptidic tree, where all structures are organized and grouped by type. For example, all carbohydrates can be found under the grouping of 'carbohydrate' with additional branches such as 'oligosaccharide', "carbohydrate phosphate', and 'glucosamine'. For example, in the penicillin group while previously each member of the penicillin family had to be searched individually by name, such as 'ampicillin', in the new tree, one may select the entire group of penicillins or individual members.

In this respect the tree allows a search strategy that is analogous and complementary to searches using the NCBI taxonomy tree, already in use in the IEDB, where searches can be performed at the taxonomical level of species, genus, family, and so on as desired. Figure 8 shows this non-peptidic tree structure on the IEDB's search interface, where the lipopolysaccharide (LPS), a commonly recognized infectious disease epitope, is presented as a child of carbohydrate, lipid, and polysaccharide. Accordingly, end users can search for data related specifically to LPS or on all carbohydrate or polysaccharide epitopes. Additionally, newly available information on each non-peptidic structure is now provided via direct links to each structure's ChEBI webpage where more detailed information on each structure, including a variety of nomenclatures, links, and citations can be found.

Discussion

Herein we present an account of the inventory and curation of immunological information related to well-defined non-peptidic epitopes. To the best of our knowledge, this is the first time this task has been undertaken in a comprehensive and systematic manner. Accordingly, an original solution had to be developed. Finally, the result of this effort highlights a case where considerable synergy can be obtained

Number of Relevant Non-Peptidic References			
Main Class	**Total**	**Number Curated**	**% Curated**
Infectious Disease	517	155	30
Allergy	776	700	90
Autoimmunity	648	244	38
Transplant	341	8	2
Model Antigens/Other	2259	15	1
Total Potentially Relevant	**4541**	**1122**	25

Table 1. Curation status of non-peptidic references.

by integrating different database resources, each focusing on clearly distinct yet related subject matters.

The inventory of all references related to non-peptidic epitopes revealed that close to a quarter of all references describing immune epitopes was indeed related to non-peptidic epitopes. This result was surprising and might be reflective of the fact that the relatively low visibility of these references might be related to the absence of comprehensive repositories where this information is cataloged, and most importantly, easily retrieved. In terms of the type of references, as compared to peptidic references, several observations were made. First, model antigens are predominant in the non-peptidic epitopes, reflecting much of the seminal work utilizing model haptens to define basic immune mechanisms. Second, the relatively large number of references relating to allergies, reflect the well-recognized importance of small molecules in terms of causing allergic reactions. Finally, relatively lower numbers of references were found describing non-peptidic epitopes related to infectious diseases and transplantation. This is not necessarily a reflection of the lack of importance of non peptidic epitopes in these settings, but rather is likely due to the technical challenges associated with the considerable chemical complexity of the structures recognized. Thus this might highlight an important opportunity for future research, as the exact definition of these molecules might facilitate the development of new diagnostic or therapeutic approaches.

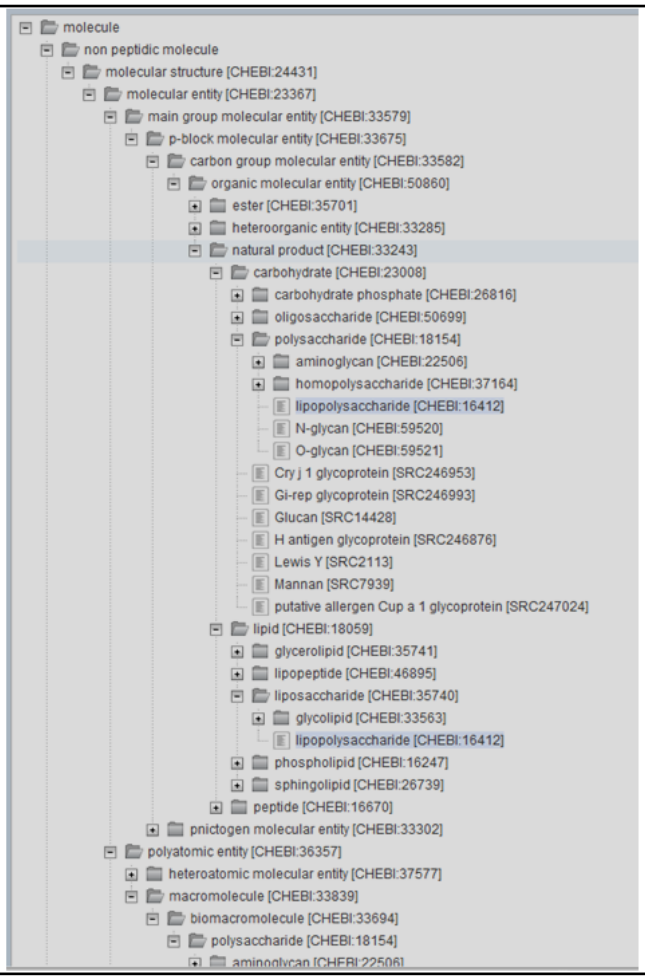
The main challenges to curation and display of immune epitope non-peptidic data were addressed by developing a curation process that integrated the processes already in place in the IEDB with processes independently developed by the ChEBI initiative. The ChEBI database was first released in 2004 as a database and ontology for chemical entities of biological interest that is a freely available dictionary of molecular entities focused on 'small' chemical compounds. ChEBI is unique in its content as it was developed to fill a niche that it describes as "long neglected by the computational biology/ bioinformatics community." ChEBI contains by far the most data available on non-peptidic structures, however, prior to the IEDB:ChEBI collaboration their immunological relevance as epitopes, immunogens or antigens was not described. In this respect the processes described herein represent an important example of how two different yet related resources can be integrated, realizing significant synergies.

Since the collaboration with ChEBI begun in 2009, curation of non-peptidic references has been swift, with complete curation of these references expected by the end of 2011. As a result, the IEDB now is a comprehensive resource for non-peptidic epitope information. Data that was previously embedded in the literature and described utilizing a wide variety of disparate nomenclature is now collected in the IEDB in a standardized and easily searchable manner. In addition, these structures are also present in the ChEBI resource with additional metadata such as citations and roles the structures can play. This represents a significant accomplishment, as only few web-based, freely available resources for non-peptidic structures are available [7]. Available online resources have historically been skewed towards peptidic/protein-based research with a variety of publically available resources such as the GenBank and Uni-prot databases.

Following the development of an efficient curation process that allows simplified curation and query of non-peptidic structures, the IEDB plans to further enhance the database to include new features. One such enhancement could allow searching for non-peptidic structures by way of the roles that they play in chemistry, immunology, and medicine. The ChEBI database provides this information for every entry and includes such roles as pesticide, antimicrobial, acid, catalyst, analgesic, etc. This information could be added to the IEDB query interface to enhance the database content and increase flexibility of search parameters. Additionally, interoperability between the ChEBI and the IEDB will be increased by linking ChEBI structures utilized by the IEDB to its epitope information on the IEDB's website later this year.

In conclusion, a curation process has been designed to specifically address non-peptidic epitopes. This process capitalized on the distinct knowledge and expertise of the IEDB and ChEBI databases and resulted in improved content with newly curated ChEBI structures and IEDB non-peptidic epitopes, as well as enhanced interoperability via direct links between the websites and the integration of ChEBI's standardized nomenclature and ontology into the IEDB's search interface. The IEDB and ChEBI initiative has resulted in a comprehensive inventory and curation of immune epitope data related to non-peptidic structures. This integration allows rigorous yet user-friendly retrieval of the data and it is envisioned that it will ultimately facilitate both basic and applied research related to fields as diverse as model antigens, infectious disease, allergies, autoimmunity and transplantation.

Figure 8. Integration of ChEBI ontology into the IEDB. Query by use of the molecule tree represents non-peptidic structures in an hierarchical fashion.



COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

ES, LZ, ZJ, PdM, ME, ST, and CS contributed to acquisition of data, drafting and revising the manuscript, and gave final approval of the version to be published.

BP and AS contributed to the conception and design of the manuscript, analysis and interpretation of the data, drafting and revising the manuscript, and gave final approval of the version to be published.

RV is the corresponding author and contributed to the conception and design of the manuscript, acquisition of data, analysis and interpretation of the data, drafting and revising the manuscript, and gave final approval of the version to be published.

ACKNOWLEDGMENTS AND FUNDING

The IEDB is funded by The National Institutes of Health/National Institute of Allergy and Infectious Diseases, Immune Epitope Database (contract number: HHSN2662004000 0 6C), under the Immune Epitope Database and Analysis Program.

We would like to thank these experts in the study of non-peptidic epitopes: Brian A Baldo, Andrea Cavani, Betty Diamond, Uri Galili, Eon N Harris, Christoph Helma, Kim D. Janda , Dennis L. Kasper, Stefan F. Martin, Werner J. Pichler, Liise-anne Pirofski, Mireille Sebbag, Rene Toes, and Dirk Zajonc.

REFERENCES

1. Smith B, Ashburner M, Rosse C, Bard C, Bug W, Ceusters W, Goldberg L J, Eilbeck K, Ireland A, Mungall C J, The OBI Consortium, Leontis N, Rocca-Serra P, Ruttenberg A, Sansone S-A, Scheuermann R H, Shah N, Whetzel P L and Lewis S: The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nature Biotechnology 2007, 25, 1251-1255. http://www.ncbi.nlm.nih.gov/sites/ppmc/articles/PMC2814061/

2. Taylor CF, Field D, Sansone SA, Aerts J, Apweiler R, Ashburner M, Ball CA, Binz PA, Bogue M, Booth T, Brazma A, Brinkman RR, Michael Clark A, Deutsch EW, Fiehn O, Fostel J, Ghazal P, Gibson F, Gray T, Grimes G, Hancock JM, Hardy NW, Hermjakob H, Julian RK Jr, Kane M, Kettner C, Kinsinger C, Kolker E, Kuiper M, Le Novère N, Leebens-Mack J, Lewis SE, Lord P, Mallon AM, Marthandan N, Masuya H, McNally R, Mehrle A, Morrison N, Orchard S, Quackenbush J, Reecy JM, Robertson DG, Rocca-Serra P, Rodriguez H, Rosenfelder H, Santoyo-Lopez J, Scheuermann RH, Schober D, Smith B, Snape J, Stoeckert CJ Jr, Tipton K, Sterk P, Untergasser A, Vandesompele J, Wiemann S: **Promoting coherent minimum reporting guidelines for biological and biomedical investigations: the MIBBI project**. *Nat Biotechnol* 2008 Aug;26(8):889-96.

3. Vita R, Zarebski L, Greenbaum JA, Emami H, Hoof I, Salimi N, Damle R, Sette A, Peters B: **The immune epitope database 2.0.** *Nucleic Acids Res*. 2010 Jan;38 (Database issue):D85462. http://www.ncbi.nlm.nih.gov/sites/ppmc/articles/PMC2808938/

4. Davies V, Vaughan K, Damle R, Peters B, Sette A: **Classification of the universe of immune epitope literature: representation and knowledge gaps.** *PLoSOne* 2009 Sep 14;4(9):e6948. http://www.ncbi.nlm.nih.gov/sites/ppmc/articles/PMC2747625/

5. Wang P, Morgan AA, Zhang Q, Sette A, Peters B: **Automating document classification for the Immune Epitope Database.** *BMC Bioinformatics* 2007 Jul 26;8:269. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1965490/

6. de Matos P, Alcántara R, Dekker A, Ennis M, Hastings J, Haug K, Spiteri I, Turner S, Steinbeck C: **Chemical entities of biological interest: an update.** *Nucleic Acids Res* 2010 Jan;38 (Database issue):D249-54. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2808869/

7. Kawasaki T, Nakao H, Tominaga T: **GlycoEpitope: A Database of Carbohydrate Epitopes and Antibodies.** *Experimental Glycoscience* 2008 Part 4, Section XXIII, 429-431.