ChEBI, IEDB and their relationship.

What ChEBI is: Chemical Entities of Biological Interest (ChEBI) is a curated database of small chemical entities important in bio-systems [1]. Its focus is on entities of no more than 1,500 atomic mass units.

ChEBI evolution: Since its inception in 2004, ChEBI has evolved from an illustrated dictionary of terms into a semantically rich knowledge base with an internal hierarchy that organises entities by their molecular structure types and potential roles. Its 2009 acquisition of the BioFocus drug discovery dataset [2] exponentially increased the number of entities from 20,000 to 500,000. ChEBI continues to develop automatic curation protocols to maintain the high standards characteristic of the smaller dataset, while the number of manually curated entities continues to rise.

ChEBI-IEDB collaboration: The Immune Epitope and Analysis Resource (IEDB) is a project supported by contract from the National Institute of Allergy and Infectious Diseases (NIAID) with the aim of making epitope-related data on infectious diseases and immune disorders freely available to researchers worldwide [3]. In addition, the IEDB houses cutting-edge analytical tools [1]. In June 2009, ChEBI began working with the IEDB on a joint project incorporating into ChEBI, by manual curation, a pilot subset of immunologically important chemical substances that had been identified as immune epitopes.

IEDB epitopes in ChEBI – the curation process: A list of 1,200 entities was provided by the IEDB, each with one or more PMIDs (PubMed unique Identifiers) for publications describing contexts in which the immune epitope status of the structure had been assayed. These chemical units were treated in one of three ways:

1. Where the entity already existed in ChEBI, the PMID was entered into the ChEBI database and automatically hyperlinkd to the PubMed site.
2. Entites automatically downloaded as part of the BioFocus drug discovery dataset required definitions, ontology specifications, JUPAC names, synonyms and registry numbers for the various databases linked to ChEBI.
3. New ChEBI identification numbers were generated for entities that did not already exist. In addition to the data items mentioned in (1) and (2) above, structures were drawn and generally cataloged using ChEBI in-house tools. The rôles played by these structures (e.g. immunogen, antigen, epitope and hapten) were also curated.

The status quo: In the year since the ChEBI-IEDB collaboration began, the pilot 1,200 entities have all been fully curated. Of this number, 50% were newly created for this project, and included a number of new rôles related to immunology. 250 new citations were added to ChEBI from this subset of IEDB data, covering epitopes relating to a variety of entities including antibiotics, organic radicals, oligosaccharides, norticins, transition metal complexes, alkaloids and amino acids. The IEDB continues to request that new structures be processed by ChEBI as new categories of immune epitope-related publications are curated. Approximately 100 structures per month are processed by ChEBI for the IEDB.

The significance of the project: The increasing global prevalence of immune-related diseases, and the complexity of the contributing factors [4, 5], underscore an ever-increasing need for cross-talk among the various scientific disciplines, and makes ChEBI involvement in this project particularly relevant. The ChEBI-IEDB teamwork here described illustrates the multiple advantages to be gained from such joint endeavours: while the incorporation of IEDB items into ChEBI has effectuated a significant enrichment of the latter’s database content and ontology, IEDB has profited from the ChEBI team’s expertise in describing non-peptidic epitopes and from the structure tree lay-out and the multiplicity of synonyms used in ChEBI, which facilitate a simplified search process. The IEDB has recently updated its search interface to make the most of its collaboration with ChEBI.

Conclusion: In identifying and curating these non-classical epitopes, we have highlighted the broad array of non-peptidic epitopes in existence. To date, we have identified over 20 categories of such entities, including antibiotics, resins, food additives and nucleic acid derivatives, in addition to the eight categories depicted in figures 1-8 of this poster. In so doing, we have demonstrated how collaboration among curators working on different databases can lead to reciprocal benefits combined with an enhanced service to our users.

References:
[1] ChEBI home page: www.ebi.ac.uk/chembl
[2] ChEBI press release, http://www.ebi.ac.uk/chembl/press/2008/07/01/rd
[3] Lita R. et al. (2010). The immune epitope database 2.0. Nucleic Acids Res. 38: (Database issue):CB54-62. Epub 2009, Nov. 11.
[4] Zhang Q et al. (2008). "Immune epitope database analysis resource (IEDB-Ar)." Nucleic Acids Res. May 31, PMID: 18515843.
[5] Warner J et al. (2006). Allergy practice worldwide: A report by the World Allergy Organization Specialty & Training Council, Allergy Clin. Immunol. Int. 18(1), http://www.aci-net/pdf/ACI_106_004-010_Forum.pdf
[6] Romagnani, S (2004): The increased prevalence of allergy and the hygiene hypothesis: missing immune deviation, reduced immune suppression, or both? Immunology 112(3): 352-363; doi: 10.1111/j.1365-2567.2004.01925.x; pdf: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1883556/pdf/nem0111-0531.pdf

Acknowledgements

ChEBI is supported by grants from the European Commission and the National Institutes of Health.

Fig.1: Haptens. Haptens are molecules that stimulate the immune response. They are usually small molecules that can be bound to proteins or other larger molecules.

Fig.2: Carbohydrate epitope. Carbohydrate epitopes are found on glycoproteins and glycolipids, and can be recognized by carbohydrate-specific antibodies.

Fig.3: Alanine-derived epitope. Alanine is a common amino acid that can be found in many proteins. Alanine-derived epitopes can be recognized by antibodies specific to alanine.

Fig.4: Fully acylated epitope. Fully acylated epitopes are found in proteins that are modified by acylation. These epitopes can be recognized by antibodies specific to the acylated region.

Fig.5: Steroidal epitope. Steroidal epitopes are found in steroids, which are a class of lipids that include hormones and vitamins. Steroidal epitopes can be recognized by antibodies specific to the steroid moiety.

Fig.6: Aromatic hydroxyepoxide epitope. Aromatic hydroxyepoxide epitopes are found in aromatic hydroxyepoxides, which are formed during the metabolism of aromatic compounds. These epitopes can be recognized by antibodies specific to the aromatic hydroxyepoxide.

Fig.7: Organosilicon epitope. Organosilicon epitopes are found in organosilicon compounds, which are a class of compounds that contain silicon and other elements. Organosilicon epitopes can be recognized by antibodies specific to the organosilicon moiety.

Fig.8: Glucosamine epitope. Glucosamine epitopes are found in glucosamine-containing compounds, which are a class of compounds that contain glucose. Glucosamine epitopes can be recognized by antibodies specific to the glucosamine moiety.

Fig.9: View of ChEBI ontology relating to a mimotope (molecule whose structure mimics that of an epitope).

Fig.10: ChEBI-IEDB collaboration diagram. The diagram illustrates the collaboration between ChEBI and IEDB, showing how data from IEDB is incorporated into ChEBI.