CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Nature Precedings

EMBL-EBI

European Bioinformatics Institute
is an Outstation of the
European Molecular Biology Laboratory.

Epitopes in ChEBI – A Collaboration with the IEDB

Zara Josephs[1], Marcus Ennis[1], Steve Turner[1] Gareth Owen, Björn Peters[2], Randi Vita[2] & Christoph Steinbeck[1]

[1]European Bioinformatics Institute, Wellcome Trust Genome Campus, Cambridge, UK
[2]IEDB Project, La Jolla Institute for Allergy and Immunology, California, USA

Broad range of non-peptidic epitopes unearthed by IEDB and represented in ChEBI

Fig.1: Nickel ion hapten – Ni^{2+} (seen here in ChEBI:53001) represents the most common contact sensitiser in the industrialised world. It acts as a hapten (partial antigen), capable of inducing an immune response upon binding to MHC-associated proteins.

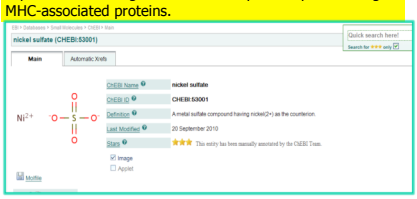


Fig.2: Saccharide antigen (ChEBI: 59384) - a branched tetrasaccharide typical of many bacterial cell wall-derived antigens.

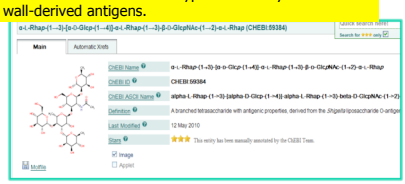


Fig.3: Amino acid-based epitope (ChEBI: 59565) – an alanine derivative.

Fig.4: Fatty acid-based epitope (ChEBI: 45573) - a butyric (butanoic) acid-derived epitope.

Fig.5: Steroidal epitope (ChEBI: 8884) - an epitope with a 3α-hydroxy steroid skeleton.

Fig.6: Organic heterocycle epitope (ChEBI: 59525) - a ladder-like polycyclic ether characteristic of certain marine toxins.

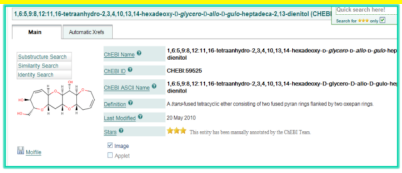


Fig.7: Organosilicon epitope (ChEBI: 59555) - an organosilicon compound with a piperidine *N*-oxide core.

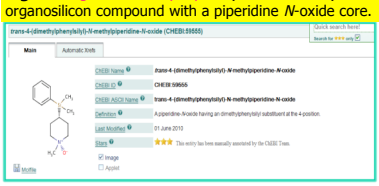
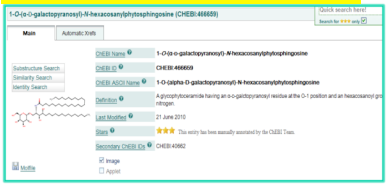


Fig.9: View of ChEBI ontology relating to a mimotope (molecule whose structure mimics that of an epitope).

Fig.8: Glycosphingolipid epitope (ChEBI: 466659) - an α-D-galactose-derived glycophytoceramide, the most potent agonistic antigen of the T cell receptor of natural killer T cells.



ChEBI, IEDB and their relationship.

What ChEBI is: Chemical Entities of Biological Interest (ChEBI) is a curated database of small chemical entities important in biosystems [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 rôles. 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.

The 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 [4]. In June 2009, ChEBI began working with the IEDB on a project aimed at 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 hyperlinked to the PubMed site;

(2) Entities automatically downloaded as part of the BioFocus dataset required definitions, ontology specifications, IUPAC 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 or generated using ChEBI's 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, norligans, 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 [5, 6], 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 mutual advantages to be gained from such joint endeavours: while the incorporation of IEDB items into ChEBI has effected 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/chebi
[2] EMBL-EBI press release, http://www.ebi.ac.uk/Information/News/pdf/Press23July08
[3] Vita R. et al. (2010). The immune epitope database 2.0. Nucleic Acids Res. 38: (Database issue):D854-62. 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/PMC1782506/pdf/imm0112-0352.pdf

Acknowledgements

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

The ChEBI Team

European Bioinformatics Institute (EMBL-EBI)
Wellcome Trust Genome Campus
Hinxton, Cambridge CB10 1SD

T +44 (0) 1223 494 444
F +44 (0) 1223 494 468
http://www.ebi.ac.uk