D906–D910 *Nucleic Acids Research, 2007, Vol. 35, Database issue* *Published online 7 November 2006*
doi:10.1093/nar/gkl849

SuperHapten: a comprehensive database for small immunogenic compounds

Stefan Günther*, Dorothea Hempel, Mathias Dunkel, Kristian Rother and Robert Preissner

Institute of Molecular Biology and Bioinformatics, Charité—University Medicine Berlin, Arnimallee 22, 14195 Berlin, Germany

Received August 15, 2006; Revised October 6, 2006; Accepted October 9, 2006

ABSTRACT

The immune system protects organisms from foreign proteins, peptide epitopes and a multitude of chemical compounds. Among these, haptens are small molecules, eliciting an immune response when conjugated with carrier molecules. Known haptens are xenobiotics or natural compounds, which can induce a number of autoimmune diseases like contact dermatitis or asthma. Furthermore, haptens are utilized in the development of biosensors, immunomodulators and new vaccines. Although hapten-induced allergies account for 6–10% of all adverse drug effects, the understanding of the correlation between structural and haptenic properties is rather fragmentary. We have developed a manually curated hapten database, SuperHapten, integrating information from literature and web resources. The current version of the database compiles 2D/3D structures, physicochemical properties and references for about 7500 haptens and 25,000 synonyms. The commercial availability is documented for about 6300 haptens and 450 related antibodies, enabling experimental approaches on cross-reactivity. The haptens are classified regarding their origin: pesticides, herbicides, insecticides, drugs, natural compounds, etc. Queries allow identification of haptens and associated antibodies according to functional class, carrier protein, chemical scaffold, composition or structural similarity. SuperHapten is available online at http://bioinformatics.charite.de/superhapten.

INTRODUCTION

Today, many aspects of the immune system like receptor-mediated signalling or induction of autoimmune diseases are only partly understood (1–3). Nevertheless, the substantial progress in the development of new vaccines (4,5) or therapeutic antibodies (6) demonstrates the great impact of immunological research on drug design and immune therapy. Although not antigenic by themselves, haptens interact with T-cell receptors or specific antibodies when conjugated to a larger antigenic molecule, usually a carrier protein. T-cells recognize haptens which are bound to the major histocompatibility complex (MHC) presented on the surface of various cell types. In contrast, hapten recognition by B-cells is mediated by receptors located on their membranes. Once they are activated, B-cells differentiate into Ig-secreting plasma cells. Such cells are able to produce highly specific antibodies that have the capability to bind haptens without carrier conjugation (7).

Specific antibody recognition is utilized in the development of biosensors (8), catalytic antibodies (9), immunomodulators (10) or new vaccines (11). However, an immune response is not always favourable. Natural and synthetic compounds in food or cosmetic products can cause skin inflammation (12), asthma and other allergic symptoms (13). Even though several mechanisms are discussed to be operative in the pathogenesis, immunogenic compounds are capable of eliciting autoimmune diseases like hepatocellular hepatitis (14) or systemic lupus erythematosus (15). It is estimated that drug allergies account for 6–10% of all adverse drug effects (16).

Many databases exist that provide diverse information for immunology. The International Immunogenetics Information System (17) consists of databases, online-tools and Web resources regarding various immunological aspects. Extensive information is also provided by the HIV Molecular Immunology Database (http://www.hiv.lanl.gov/content/immunology). Data relating to antigenic epitopes are collected in the databases JenPep (18), AntiJen (19) and Epitome (20). Furthermore the databases Kabat (21), MHCBN (22), Bcipep (23) and FIMM (24) provides comprehensive information about MHC binding peptides, B-cell epitopes or antibody sequences.

Haptens have rarely been subject to an extensive description. Recently, a large manually curated immune epitope database (IEDB) was published (25). Although the database focuses on antigenic peptides, it contains 91 haptens. Much more haptens are described within the HaptenDB (26), a collection of approximately 1000 haptens with its main focus on the underlying immunochemical assays. It contains information about the assay method, conjugation ratio of haptens and carrier molecules, sensitivity and specificity of the system or

*To whom correspondence should be addressed. Tel: +49 30 450 528375; Fax: +49 30 450 528942; Email: stefan.guenther@charite.de

cross reactivity of the antibodies used. Although the database is very helpful to get an overview of the performed immunologic experiments, we noticed that many important immunologic compounds are missing. For example, drugs such as aspirin or beta adrenergic antagonists are capable of eliciting severe autoimmune diseases like bronchial asthma (27) or Psoriasis (28). Chlorated benzenes present in pesticides, solvents and lubricants are frequently described as haptens but apparently not incorporated within the HaptenDB. While the authors processed in detail the data of approximately 250 hapten-related articles, a large proportion of described haptens and associated immunologic articles is ignored. Furthermore, information about the availability of the described haptens and antibodies will be helpful in case a described immunoassay is to be reproduced.

To overcome these problems, we have performed a more extensive semi-automatic literature screening, resulting in a comprehensive dataset.

DATABASE DESCRIPTION

SuperHapten currently contains 7257 different haptens and references to about 10,000 immunologic articles. The basic application of the database is the identification of compounds that elicit immune response. The user interface provides diverse query types, like searching by hapten name, physicochemical properties, industrial use or associated carrier protein. Figure 1 depicts various query types. A built-in molecule editor allows the user to draw a molecule and to screen the database for compounds with similar molecular structure. This feature is of particular interest even if the specific requested molecule is not part of the database. Since compounds with high structural similarity frequently exhibit similar activities (29), this method provides an informative basis for the identification of new immunogenic compounds. Identified haptens are displayed by 3D structure and may act as starting points for further similarity searches. Each compound is provided with a full info page containing synonyms, molecular properties and references to the underlying literature. Another important aspect to assist immunologists is the availability of the haptens for further experiments. We have checked whether the haptens are commercially obtainable and provided 6279 compounds with ordering codes and external links to suppliers. Similarly, the ordering information of 453 hapten related antibodies was determined and included within the database. This procedure may avoid exhausting *in vitro* antibody generations if specific antibodies are required. Carrier proteins are annotated with UNIPROT-ID, PDB-structure and underlying references to literature. Furthermore, we have checked whether the annotated haptens are structurally resolved in complex with antibodies, carriers or other proteins. The corresponding PDB cross references are included on the hapten related info pages. Some haptens were identified, which are assured to be MHC-mediated. These 46 haptens are specified separately and directly retrievable. Table 1 specifies some key numbers of the database content.

Another intention was to allow immunologists an overview of known immunogenic compounds. For that purpose, the haptens were compared all against all and clustered by structural similarity. The resulting 154 clusters containing at least eight similar compounds were manually ordered to obtain a hierarchical compound tree. Each node was named according to the chemistry of the included compounds. For example, halogenated chemicals are frequently responsible for environmental skin diseases (30). About 800 references to articles are retrievable, which describe an immunologic response to halogenated compounds. By the same procedure, chemical classes like sulfurones or triazines are easily accessible. The presented cluster tree could potentially allow researchers to analyse the relationship between molecular structure and immunogenic effect. Exemplarily, Table 2 shows seven chemicals according to different compound groups.

METHODS

Haptens were collected from literature and various web resources. The abstracts of the literature database PubMed were filtered for relevant immunologic articles, using specific keywords. The 15,000 abstracts thus obtained were screened against names and synonyms of 3 million chemical compounds as well as a distinct set of substrings of IUPAC names.

The text passages containing matches were manually curated by a scientific team that confirmed the matching and analysed whether the identified compound was recognized by antibodies or T-cell receptors. About 200 full text review articles were subjected to the same procedure. The resulting 1600 haptens formed the confirmed basic dataset. Two web resources were checked, the DIMDI contact allergen database (http://www.dimdi.de/static/en/db/dbinfo/dbkurz/ka00.htm) and the HaptenDB (26). Additional 400 compounds were detected and included within the SuperHapten. Each compound of the basic dataset was translated into a structural fingerprint, using the chemistry development kit (http://almost.cubic.uni-koeln.de/cdk/). The fingerprint algorithm follows the approach taken by Daylight (http://www.daylight.com/dayhtml/doc/theory/). Up to six connected atoms all connectivity paths within the molecule are determined. For each path, a hash function calculates the location of a representing bit in a Boolean array. The number of possible paths is huge and it is not possible to assign a particular bit to each path, thus a particular bit represents several paths. The combination of all representing bits of all paths yields in a specific pattern or structural fingerprint.

The similarity index used is the Tanimoto coefficient, which is the number of bit positions set to 1 in both fingerprints divided by the number of bit positions set to 1 in at least one of the fingerprints. If a set bit is considered as a feature present in the molecule, the Tanimoto coefficient is a measure of the number of common features in both molecules (31). A Tanimoto coefficient of >0.85 indicates that two molecules have similar activities (32). Calculated fingerprints were used for a further structural screening against 4 million compounds. In this way, 5248 compounds were detected that resulted in Tanimoto coefficients of >0.9. These putative haptens are highly probable candidates for eliciting the same kind of immune response, binding to the same carrier and being recognised by the same antibodies or T-cells as the haptens found in literature (29).

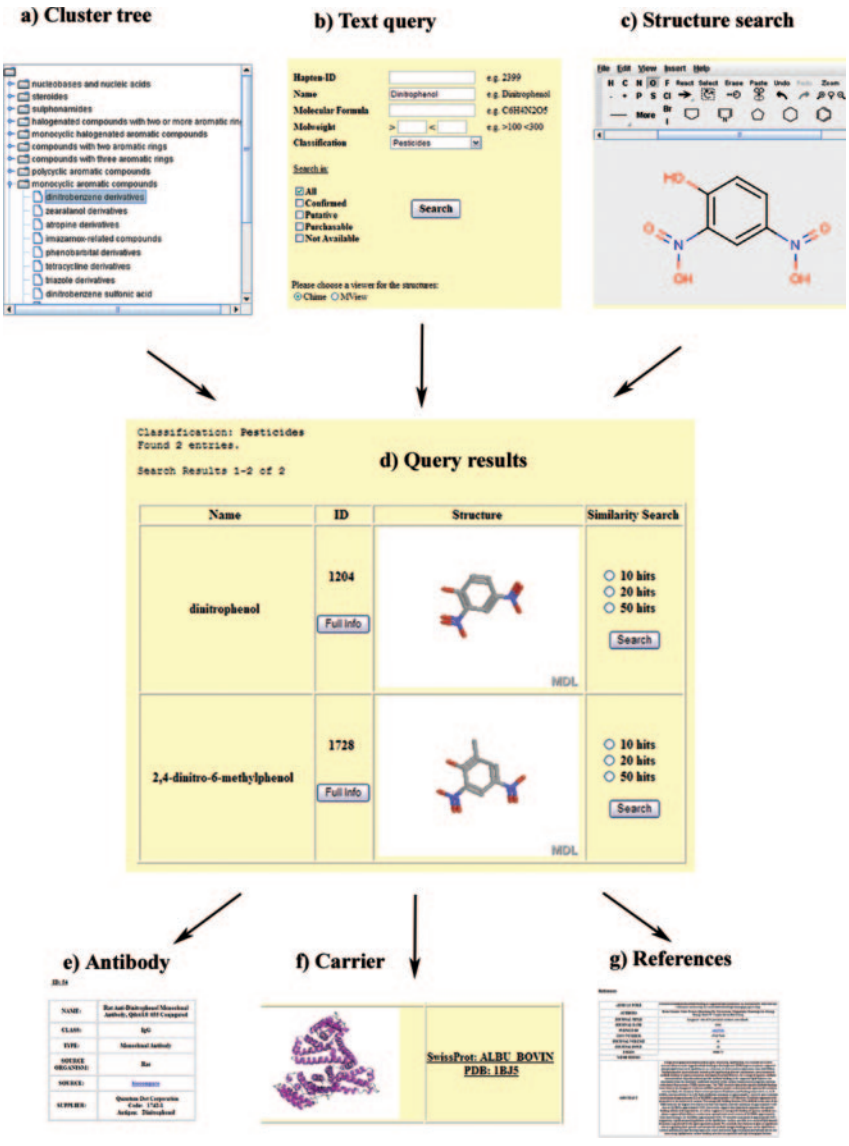


Figure 1. Queries and results of the SuperHapten web-interface. Query types: (**a**) Part of the cluster-tree, showing the structure of the scaffold classification. Clicking on the nodes expands the branch: monocyclic aromatic compounds → dinitrobenzene derivatives etc. Clicking on leafs triggers a database search in a separate window. (**b**) Text query options including a classification regarding the origin of the haptens. (**c**) Screenshot of the java applet Marvin which allows upload or drawing of own structures for similarity searches in the SuperHapten database showing Dinitrophenol. Results (**d**) Query results with search options for 2D similarity: Dinitrophenol (**C6H4N2O5**) and 2,4-dinitro-6-methylphenol (C7H6N2O5). The compounds can be rotated (left mouse button), different display styles are available, structures can be saved (right mouse button) and more detailed information, such as synonymous names and formula or supplier information can be obtained by use of the FULL INFO button. Further information (**e**) Antibody and suitable information like supplier, ID and source organism are displayed. (**f**) Information on hapten specific carrier protein(s), references, UNIPROT-, PDB-ID etc. (**h**) Recent scientific references confirming the haptenic effect including author information and abstract.

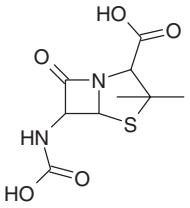
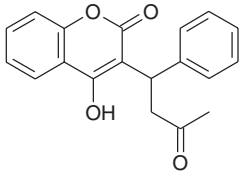
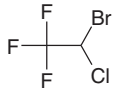
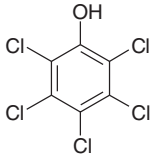
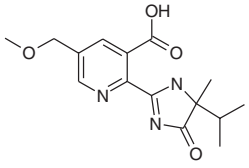
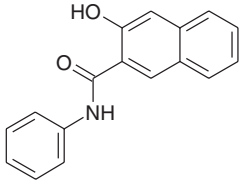
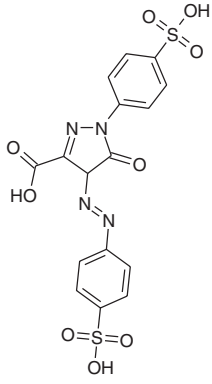
Table 1. Key numbers of the database content

Entities	
Haptens	7257
Antibodies	453
Carrier proteins	24
References to literature	9670
Hapten subgroups	
Confirmed	2009
Putative	5248
Commercially available	6285
Drugs	392
Natural compounds	1397
Pesticides	227

3D representations of all haptens were generated with Discovery Studio, (Accelrys Inc., http://www.accelrys.com/dstudio) and visualized with the free Chime-Plugin from MDL (available for Windows, SGI, Mac). Linux compatibility is warranted by a second visualizer, MarvinView (ChemAxon). The same tool was also used for the built-in molecule editor that allows structural screening with self-edited molecules.

For compatibility, our data model are directed to the schema implemented in the IEDB (25). A diagram of the database schema is shown at the website. The data are implemented as a relational database on a MySQL server and publicly available at http://bioinformatics.charite.de/superhapten.

Table 2. Examples of haptens classified as drugs, contact allergens or pesticides

Hapten cluster	Sample structure	Cluster size	Properties, industrial use
Drugs			
Penicillin and derivatives		92	used in the treatment of bacterial infections
Warfarin derivatives		10	used for the prophylaxis of thrombosis and embolism in many disorders
Pentahalogenated ethanes (halothane)		16	its vapour is an inhalational general anaesthetic
Pesticides			
Tetra- and pentahalogenated phenols		29	synthetic fungicides
Imazamox-related compounds		119	used as herbicides
Contact allergens			
Naphthol derivatives		8	colourless, crystaline solid, used in the production of dyes
Tartrazines		8	lemon yellow, azo dye, used for food colouring

The cluster size specifies the number of highly similar haptens grouped together.

CONCLUSIONS AND FUTURE DIRECTIONS

By now, the presented database has become a useful tool to retrieve information about specific immunogenic compounds, or to get an overview of the known substances that elicit immune responses. The KEGG-pathway analysis of single haptenic compounds revealed that their synthesis is exclusively performed by bacterial enzymes (33). SuperHapten enables systematic approaches on the position of haptens in metabolic or signalling networks. The included data on purchasability of haptens and related antibodies will enable systematic experimental approaches on the relation between structural similarity and cross-reactivity.

Furthermore, structure comparisons of self-edited molecules to the annotated haptens may allow a first rough estimation of the potential immunogenicity of new chemicals.

ACKNOWLEDGEMENTS

This work is supported by the Deutsche Forschungsgemeinschaft (DFG). Funding to pay the Open Access publication charges for this article was provided by DFG-Sonderforschungsbereich 449.

Conflict of interest statement. None declared.

REFERENCES

1. Rioux,J.D. and Abbas,A.K. (2005) Paths to understanding the genetic basis of autoimmune disease. *Nature*, **435**, 584–589.
2. O'Neill,L.A. (2006) How Toll-like receptors signal: what we know and what we don't know. *Curr. Opin. Immunol.*, **18**, 3–9.
3. Mahoney,J.A. and Rosen,A. (2005) Apoptosis and autoimmunity. *Curr. Opin. Immunol.*, **17**, 583–588.
4. Linhart,B. and Valenta,R. (2005) Molecular design of allergy vaccines. *Curr. Opin. Immunol.*, **17**, 646–655.
5. Houghton,M. and Abrignani,S. (2005) Prospects for a vaccine against the hepatitis C virus. *Nature*, **436**, 961–966.
6. Kim,S.J., Park,Y. and Hong,H.J. (2005) Antibody engineering for the development of therapeutic antibodies. *Mol. Cells*, **20**, 17–29.
7. Singh,K.V., Kaur,J., Varshney,G.C., Raje,M. and Suri,C.R. (2004) Synthesis and characterization of hapten-protein conjugates for antibody production against small molecules. *Bioconjug. Chem.*, **15**, 168–173.
8. Cho,A.Y., Yi,K.S., Rhim,J.H., Kim,K.I., Park,J.Y., Keum,E.H., Chung,J. and Oh,S. (2006) Detection of abnormally high amygdalin content in food by an enzyme immunoassay. *Mol. Cells*, **21**, 308–313.
9. Hanson,C.V., Nishiyama,Y. and Paul,S. (2005) Catalytic antibodies and their applications. *Curr. Opin. Biotechnol.*, **16**, 631–636.
10. Holzer,A.M., Kaplan,L.L. and Levis,W.R. (2006) Haptens as drugs: contact allergens are powerful topical immunomodulators. *J. Drugs Dermatol.*, **5**, 410–416.
11. Berd,D., Sato,T., Maguire,H.C., Jr, Kairys,J. and Mastrangelo,M.J. (2004) Immunopharmacologic analysis of an autologous, hapten-modified human melanoma vaccine. *J. Clin. Oncol.*, **22**, 403–415.
12. Divkovic,M., Pease,C.K., Gerberick,G.F. and Basketter,D.A. (2005) Hapten-protein binding: from theory to practical application in the *in vitro* prediction of skin sensitization. *Contact Dermatitis*, **53**, 189–200.
13. Vanoirbeek,J.A., Tarkowski,M., Vanhooren,H.M., De Vooght,V., Nemery,B. and Hoet,P.H. (2006) Validation of a mouse model of chemical-induced asthma using trimellitic anhydride, a respiratory sensitizer, and dinitrochlorobenzene, a dermal sensitizer. *J. Allergy. Clin. Immunol.*, **117**, 1090–1097.
14. Ju,C. (2005) Immunological mechanisms of drug-induced liver injury. *Curr. Opin. Drug. Discov. Devel.*, **8**, 38–43.
15. Pelizza,L., De Luca,P., La Pesa,M. and Minervino,A. (2006) Drug-induced systemic lupus erythematosus after 7 years of treatment with carbamazepine. *Acta. Biomed Ateneo Parmense*, **77**, 17–19.
16. Ju,C. and Pohl,L.R. (2005) Tolerogenic role of Kupffer cells in immune-mediated adverse drug reactions. *Toxicology*, **209**, 109–112.
17. Lefranc,M.P. (2005) IMGT, the international ImMunoGeneTics information system(R): a standardized approach for immunogenetics and immunoinformatics. *Immunome Res.*, **1**, 3.
18. McSparron,H., Blythe,M.J., Zygouri,C., Doytchinova,I.A. and Flower,D.R. (2003) JenPep: a novel computational information resource for immunobiology and vaccinology. *J. Chem. Inf. Comput. Sci.*, **43**, 1276–1287.
19. Toseland,C.P., Clayton,D.J., McSparron,H., Hemsley,S.L., Blythe,M.J., Paine,K., Doytchinova,I.A., Guan,P., Hattotuwagama,C.K. and Flower,D.R. (2005) AntiJen: a quantitative immunology database integrating functional, thermodynamic, kinetic, biophysical, and cellular data. *Immunome Res.*, **1**, 4.
20. Schlessinger,A., Ofran,Y., Yachdav,G. and Rost,B. (2006) Epitome: database of structure-inferred antigenic epitopes. *Nucleic Acids Res.*, **34**, D777–D780.
21. Johnson,G. and Wu,T.T. (2000) Kabat database and its applications: 30 years after the first variability plot. *Nucleic Acids Res.*, **28**, 214–218.
22. Bhasin,M., Singh,H. and Raghava,G.P. (2003) MHCBN: a comprehensive database of MHC binding and non-binding peptides. *Bioinformatics*, **19**, 665–666.
23. Saha,S., Bhasin,M. and Raghava,G.P. (2005) Bcipep: a database of B-cell epitopes. *BMC Genomics*, **6**, 79.
24. Schonbach,C., Koh,J.L., Flower,D.R., Wong,L. and Brusic,V. (2002) FIMM, a database of functional molecular immunology: update 2002. *Nucleic Acids Res.*, **30**, 226–229.
25. Sathiamurthy,M., Peters,B., Bui,H.H., Sidney,J., Mokili,J., Wilson,S.S., Fleri,W., McGuinness,D.L., Bourne,P.E. and Sette,A. (2005) An ontology for immune epitopes: application to the design of a broad scope database of immune reactivities. *Immunome Res.*, **1**, 2.
26. Singh,M.K., Srivastava,S., Raghava,G.P. and Varshney,G.C. (2006) HaptenDB: a comprehensive database of haptens, carrier proteins and anti-hapten antibodies. *Bioinformatics*, **22**, 253–255.
27. Pfaar,O. and Klimek,L. (2006) Aspirin desensitization in aspirin intolerance: update on current standards and recent improvements. *Curr. Opin. Allergy Clin. Immunol.*, **6**, 161–166.
28. O'Brien,M. and Koo,J. (2006) The mechanism of lithium and beta-blocking agents in inducing and exacerbating psoriasis. *J. Drugs Dermatol.*, **5**, 426–432.
29. Martin,Y.C., Kofron,J.L. and Traphagen,L.M. (2002) Do structurally similar molecules have similar biological activity? *J. Med. Chem.*, **45**, 4350–4358.
30. Yamamoto,O. and Tokura,Y. (2003) Photocontact dermatitis and chloracne: two major occupational and environmental skin diseases induced by different actions of halogenated chemicals. *J. Dermatol. Sci.*, **32**, 85–94.
31. Delaney,J.S. (1996) Assessing the ability of chemical similarity measures to discriminate between active and inactive compounds. *Mol. Divers*, **1**, 217–222.
32. Patterson,D.E., Cramer,R.D., Ferguson,A.M., Clark,R.D. and Weinberger,L.E. (1996) Neighborhood behavior: a useful concept for validation of 'molecular diversity' descriptors. *J. Med. Chem.*, **39**, 3049–3059.
33. Honda,W., Kawashima,S. and Kanehisa,M. (2006) Metabolite antigens and pathway incompatibility. *Genome Informatics*, **17**, 184–194.