





DATA NOTE

Hepitopes: A live interactive database of HLA class I epitopes in hepatitis B virus [version 1; referees: 2 approved, 1 approved with reservations]

Sheila Lumley¹, Howard Noble², Martin J. Hadley², Liz Callow³, Amna Malik⁴, Yi Yi Chua⁵, Owen J. Duffey⁵, Natalia Grolmusova⁵, Arvind Kumar⁵, Samuel Ravenscroft⁵, Jonathan I. Spencer⁵, Christoph Neumann-Haefelin⁶, Robert Thimme⁶, Monique Andersson¹, Paul Klenerman  ^{1,7,8}, Eleanor Barnes^{7,8}, Philippa C. Matthews  ^{1,7}

¹Department of Infectious Diseases and Microbiology, Oxford University Hospitals NHS Foundation Trust, John Radcliffe Hospital, Oxford, UK

²IT Services, University of Oxford, Oxford, UK

³Bodleian Health Care Libraries, University of Oxford, John Radcliffe Hospital, Oxford, UK

⁴Department of Paediatrics, University of Oxford, Oxford, UK

⁵University of Oxford Medical School, John Radcliffe Hospital, Oxford, UK

⁶Department of Medicine II, University Hospital Freiburg, Freiburg, Germany

⁷Nuffield Department of Experimental Medicine, University of Oxford, Oxford, UK

⁸NIHR Biomedical Research Centre, John Radcliffe Hospital, Oxford, UK




v1 First published: 15 Nov 2016, 1:9 (doi: [10.12688/wellcomeopenres.9952.1](https://doi.org/10.12688/wellcomeopenres.9952.1))
Latest published: 15 Nov 2016, 1:9 (doi: [10.12688/wellcomeopenres.9952.1](https://doi.org/10.12688/wellcomeopenres.9952.1))

Abstract

Increased clinical and scientific scrutiny is being applied to hepatitis B virus (HBV), with focus on the development of new therapeutic approaches, ultimately aiming for cure. Defining the optimum natural CD8+ T cell immune responses that arise in HBV, mediated by HLA class I epitope presentation, may help to inform novel immunotherapeutic strategies. Therefore, we have set out to develop a comprehensive database of these epitopes in HBV, coined 'Hepitopes'. This undertaking has its foundations in a systematic literature review to identify the sites and sequences of all published class I epitopes in HBV. We also collected information regarding the methods used to define each epitope, and any reported associations between an immune response to this epitope and disease outcome. The results of this search have been collated into a new open-access interactive database that is available at <http://www.expmndm.ox.ac.uk/hepitopes>. Over time, we will continue to refine and update this resource, as well as inviting contributions from others in the field to support its development. This unique new database is an important foundation for ongoing investigations into the nature and impact of the CD8+ T cell response to HBV.

Open Peer Review

Referee Status:   

| | Invited Referees | | |
|--|---|---|---|
| | 1 | 2 | 3 |
| version 1 published 15 Nov 2016 |  report |  report |  report |

- Antonio Bertoletti**, Duke-National University of Singapore Graduate Medical School Singapore, **Anthony Tanoto Tan**, Duke-National University of Singapore Graduate Medical School Singapore
- Upkar S Gill**, Barts and The London School of Medicine & Dentistry, Queen Mary University of London UK
- Louise Corti**, University of Essex UK
- Joerg Timm**, Heine-University, University Hospital Germany

Discuss this article[Comments \(0\)](#)

Corresponding author: Philippa C. Matthews (p.matthews@doctors.org.uk)

How to cite this article: Lumley S, Noble H, Hadley MJ *et al.* **Hepitopes: A live interactive database of HLA class I epitopes in hepatitis B virus [version 1; referees: 2 approved, 1 approved with reservations]** Wellcome Open Research 2016, 1:9 (doi: [10.12688/wellcomeopenres.9952.1](https://doi.org/10.12688/wellcomeopenres.9952.1))

Copyright: © 2016 Lumley S *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Grant information: Wellcome Trust Intermediate Fellowship [110110/Z/15/Z]; NIHR (Research Capability Funding).
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: No competing interests were disclosed.

First published: 15 Nov 2016, 1:9 (doi: [10.12688/wellcomeopenres.9952.1](https://doi.org/10.12688/wellcomeopenres.9952.1))

Introduction

Hepatitis B virus (HBV) is the prototype virus in the *Hepadnaviridae* family. It has unique features that make it interesting and distinct from other related viruses. Its unusual, partially double-stranded circular genomic structure, comprising several overlapping reading frames¹, represents a potential barrier to variability – any nonsynonymous nucleic acid substitution potentially has to be tolerated in the resulting sequences of more than one protein product². However, its error-prone reverse transcriptase enzyme, responsible for the generation of nucleic acid intermediates, is a source of diversity that is unusual in a DNA virus^{3–5}.

HBV is estimated to infect 240 million people globally, and to cause over half a million deaths each year (<http://www.who.int/mediacentre/factsheets/fs204/en/>). At present, therapy consists of peginterferon alfa-2a, tenofovir disoproxil or entecavir (<https://www.nice.org.uk/guidance/cg165>). Although successful in mediating viral suppression, these drugs, either alone or in combination, do not commonly mediate cure, and many patients are committed to life-long therapy. However, the striking recent success of direct acting antiviral therapy for HCV has sparked new enthusiasm for the vision of a HBV cure⁶.

To optimize our chances of HBV eradication, there are a number of important public health interventions that must be pursued in parallel, including diagnostics, treatment and prevention. A crucial component of this strategy is the development and testing of new drug agents, and the investigation of immunotherapeutic interventions^{7,8}. There is growing interest in the idea of a therapeutic vaccine that could boost or mimic the immune responses that correlate best with HBV clearance in natural infection^{6,9,10}.

There are several strands of evidence for an important role of the CD8+ T cell response in HBV. Associations have been documented between HLA genotype and disease outcome^{11–13}, including responses to, and recovery from, acute infection^{14–16}. In patients that clear HBV infection, interferon (IFN)-gamma producing CD8+ T cells are enriched, compared to those with chronic infections¹⁷. The recent literature demonstrates an emerging interest into investigations of this CD8+ mediated influence on HBV, including the identification of HLA class I epitopes (for examples, see references^{13,18,19}; for a full dataset, see ‘Data availability’ section), and reports of HLA-driven selection pressure driving viral diversity^{20,21}.

This dynamic between host and virus is well described for HIV^{22–24}, in which the HLA class I restricted T cell response is a major influence on disease control^{25–27}. An on-line database of HIV epitopes (https://www.hiv.lanl.gov/content/immunology/ctl_search) is an important resource in pooling knowledge and resources, and driving collaboration, in the field of HIV T cell immunology. Although there are several useful on-line repositories for HBV sequence data (including HBV-specific databases <https://hbvdb.ibcp.fr/HBVdb/> and <http://www.hepseq.org>, as well as HBV sequences that can be found in broader sequence repositories, such as <http://www.ncbi.nlm.nih.gov>), and a recent publication collates the most robustly

reported HBV class I epitopes²⁸, there is no current publically-available database of HBV epitopes to parallel the successful resource that has grown and developed for HIV (https://www.hiv.lanl.gov/content/immunology/ctl_search).

Therefore, we have begun the process of collating HBV class I epitopes. Our primary aim was to provide a detailed and comprehensive database of epitopes that have been identified to date, but with an ongoing remit to develop this into a sustainable open-access online research resource that evolves over time.

Materials and methods

Systematic literature review

We searched two online bibliographic databases, Medline and Embase, in January 2016, using the OVID search interface (<http://ovidsp.ovid.com/>) made available by the University of Oxford. No date restrictions were imposed; Medline was searched from 1946–2016 and Embase from 1974–2016. The relevant subject headings for Epitopes, HLA Antigens, CD8 Antigens and Hepatitis B from the thesauri (MESH and EMTREE) were exploded and searched. In addition, the terms ‘epitope*’ (to pick up singular and plural) and ‘Hepatitis B’ and ‘HBV’ were searched in the title and abstract fields. The terms of the search strategy and the Boolean operators used are presented in [Table 1](#). We also identified three additional references by searching the bibliographies of articles that had been identified in the primary search. These can be identified within our dataset (by searching the column entitled ‘database’ which specifies the origin of the citation).

Using this approach, we identified 447 papers in total. We reviewed each paper to ensure it met our search criteria, excluding duplicates (n=113), those not relevant or missing essential data (n=177), publications not written in English (n=26) and those that we were unable to access (n=19), leading to the inclusion of final data from 112 papers. From these, we recorded the following details: citation, epitope location and sequence, HLA restriction, experimental approach used to confirm the epitope, and any documented association with stage of HBV disease. From the selected 112 manuscripts, we collated a database of HBV epitopes using Google Sheets (<https://docs.google.com/spreadsheets>) each of which has documented optimal and/or overlapping peptide sequence and defined HLA-class I restriction.

Sequence numbering

In order to unify our approach to numbering amino acid position within HBV proteins, we used a reference sequence published by Liu *et al.*²⁹. In order to support consistent alignment, and to identify potential genotype-specific differences that may affect epitope presentation or recognition, we have also defined consensus sequences for each protein based on all available sequences downloaded from <https://hbvdb.ibcp.fr/HBVdb/>. The reference sequence and aligned genotype-specific consensus sequences are available on-line (DOI: <http://dx.doi.org/10.6084/m9.figshare.4040700.v1>³⁰).

Table 1. Search strategy for systematic literature review to identify HLA class I epitopes in hepatitis B virus (HBV). This table represents the two databases searched by our literature review, the dates of publication included in the search, and the search terms used to identify relevant citations. 'Exp' denotes that the term has been selected from the database thesaurus and exploded to pick up its narrower terms. The / sign denotes that the term is a subject heading from the thesaurus. 'Epitope*' includes search for 'epitope' and 'epitopes'. '.mp' denotes a search of the title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, and unique identifier.

| Database | Embase | Medline |
|-----------------|---|---|
| Date searched | January 1974 – January 2016 | January 1946 – January 2016 |
| Search strategy | <ol style="list-style-type: none"> 1. exp epitope/ 2. epitope*.mp 3. 1 or 2 4. exp HLA antigen/ 5. exp CD8 antigen/ 6. 4 or 5 7. exp Hepatitis B/ 8. (Hepatitis B or HBV).mp. 9. 7 or 8 10. 3 and 6 and 9 | <ol style="list-style-type: none"> 1. exp epitopes/ 2. epitope*.mp 3. 1 or 2 4. exp HLA Antigens/ 5. exp Antigens, CD8/ 6. 4 or 5 7. exp Hepatitis B/ 8. (Hepatitis B or HBV).mp. 9. 7 or 8 10. 3 and 6 and 9 |

Interactive database

The results of the literature search, which will form the basis of an expanding data resource, can be viewed at <http://www.expmendm.ox.ac.uk/hepitopes> as an interactive web element to facilitate other researchers in exploring and querying the dataset (Figure 1A). This web element utilizes the widely used JavaScript library Datables (SpryMedia Limited v. 1.10.12; datatables.net) to allow the data to be filtered and sorted interactively. Hosting is provided by RStudio webservice (shinyapps.io) and the web element was developed using the R framework Shiny (Shiny: Web Application Framework for R, v. 0.14.1.; <https://CRAN.R-project.org/package=shiny>)³¹.

The interactive web element was developed as a case study of the Live Data project led by Research Support Services and funded by IT Services at University of Oxford. The project was initiated to investigate whether a central interactive visualization service would increase engagement with datasets underlying research projects, and foster additional research impact. Hepitopes provided a template case study for researchers interested in making creating interactive interfaces to databases and flat files.

We will continue to curate the Hepitopes on-line database, adding amendments and updates over time. To facilitate wider contributions from the scientific community, our website includes a contact portal (Figure 1B); we welcome contributions of new data and citations, corrections to existing information, or general feedback about the site and resource, with the aims of refining the quality of the dataset, encouraging dialogue and collaboration, and adding new tools, including links to other relevant resources.

Dataset validation

The results of our primary literature search are limited by our ability to identify all reports that may contain pertinent data; it may be that we have overlooked relevant citations due to the data being

presented in manuscripts that did not contain our specific search terms. Furthermore, we were unable to access the complete report for 19 citations, which may contain relevant data (e.g. conference abstracts, for which the entire dataset has not been published electronically). We recognize that the nature and quality of the data presented in different reports is heterogeneous, due to different genotypes of HBV studied presented with varying alignments, and a diverse range of methodological approaches ranging from *in silico* prediction to detailed *in vitro* elucidation of epitopes in natural infection or in animal models.

It is important to recognize that the predominance of any particular HLA class I molecule may not reflect a true biological immunodominance hierarchy, but rather is a result of the bias towards identification, reporting and further investigation of epitopes restricted by the HLA class I alleles that occur at the highest phenotypic frequency in the majority of human populations. This is illustrated by HLA-A*02 epitopes, which - at the point of releasing the database - account for 42% of the total and are referenced in 107 of the cited manuscripts.

We also acknowledge that the HBV literature is skewed towards investigation of certain populations, most specifically in the Far East, where HBV genotypes B and C are endemic³², and potentially also in Western Europe and North America where better resources are available for these studies. On these grounds, the results of our literature search over-represent certain host/virus interactions, specifically between HBV genotype-B/C and host HLA class I alleles represented at the highest phenotypic frequency in these populations. An alternative source of bias in the existing data is the peptide sets used for *in vitro* confirmation of optimal epitopes, among which genotype D is anecdotally over-represented (although this methodological bias is difficult to quantify in any systematic way). The aim is that, as the resource develops over time, we will build

A

Find the on-line hepitopes database with this URL

Use this toolbar to select the columns you want to visualise (click on an existing column title and hit delete to remove, or click on the toolbar and select new columns from the drop-down menu to add)

Use this scrollbar to look through all the results of your search

Sort the data in any column using the arrows at the side of the column header

Enter a specific search term here

In this row, the citation reported only a 17-mer OLP sequence, not the optimal epitope, so the optimal amino acid sequence is 'NS' (not specified).

In this row, the citation reported an 11-mer optimal epitope, but not the flanking sequences within an OLP, so the OLP amino acid sequence is 'NS' (not specified).

B

Hepitopes Feedback Form

Thank you for your interest in the Hepitopes Database.

We are keen to be contacted with amendments, suggestions and feedback; please use the form below to get in touch.

If you provide us with an email address, we will get back to you following your enquiry.

Q1. Please specify the nature of your enquiry

- I would like to suggest a correction to the database (if this pertains to a specific epitope, please fill in specific details in Q2 below)
- I have a new epitope to add to the database (if you select this option, please fill in specific details in Q3 below)
- I would like to offer feedback / suggestions about the database (if you select this option, please fill in specific details in Q4 below)

Q2. Please enter specific details of corrections to any existing epitope listed by the database (please be as specific as possible, including the location and sequence of the epitope(s), and providing any relevant citations).

Your answer

Q3. Please enter specific details of any new epitope you would like to be added to the database (please be as specific as possible, including the location and sequence of the epitope(s), and providing any relevant citations).

Your answer

Q4. Please enter any other comments / feedback into the free text box below

Your answer

Q5. Please provide your clinical or academic affiliation (optional)

Your answer

Q6. Please provide a contact email address (optional)

Your answer

SUBMIT

Figure 1. Screenshots of the Hepitopes database (<http://www.expmendnm.ox.ac.uk/hepitopes>). (A) Appearance of the data visualization on-line, with notes on interrogation of the database (green boxes) and features of the output (purple boxes). (B) On-line contact portal to develop dialogue and collaborations and to feed updates into the database.

a more comprehensive picture of the true spectrum of HLA class I presented epitopes, and gather data from populations that have been under-represented to date, including our own particular focus of interest in southern Africa^{33,34}.

Data and software availability

On-line visualization of the Hepitopes database

The live interactive database containing the collated results of our literature review is available via the Hepitopes project website: <http://www.expmedndm.ox.ac.uk/hepitopes>. This visualisation is represented by the screenshots in [Figure 1](#), together with some notes about use of this interface. The code for interactive elements is made available under an MIT license and is hosted on GitHub, <https://github.com/ox-it/hepitopes> (archived version at the time of publication: DOI: <http://dx.doi.org/10.5281/zenodo.16310635>).

Raw data

Data is ingested into the web element via a Figshare repository (DOI: <http://dx.doi.org/10.6084/m9.figshare.401582436>), which we will continue to update over time. Additionally, a snapshot of the data at the time of submission has been deposited on the Oxford Research Archive (ORA; DOI: <http://dx.doi.org/10.5287/bodleian:zr0VAr78q37>; <https://ora.ox.ac.uk/objects/uuid:9aa65287-5027-4422-abd1-f697aa3f7dc9>).

Aligned reference and consensus sequences for HBV proteins

Our sequence alignment, including a description for the way this was generated, is available from Figshare (DOI: <http://dx.doi.org/10.6084/m9.figshare.4040700.v130>).

Access to citations represented in the Hepitopes database

The DOI (digital object identifier, issued for articles published after the year 2000) listed for each citation within the Hepitopes database is in the public domain. Access to the abstract and full text is subject to institutional or individual subscription to the journal, or open access availability.

Ethics

Patient consent was not required for this project, and no specific ethics approval was required. The original manuscripts cited in our database should be consulted individually for details of specific ethics approvals and consent, if required.

Author contributions

SL, MA and PM conceived the study. LC advised on the bibliographic search to identify relevant citations. SL, AM, JS, NG, OD, YYC, SR, and AK performed the original review of all the citations and collected relevant data. PM re-reviewed all the references in the final dataset. HN and MH developed the software to support on-line visualization of the database in consultation with PM. SL and PM wrote the manuscript, with methods contributed by HN, MH and LC, and expert input from CNH, RT, PK and EB. All authors were involved in the revision of the draft manuscript and have agreed to the final content.

Competing interests

No competing interests were disclosed.

Grant information

Wellcome Trust Intermediate Fellowship [110110/Z/15/Z]; NIHR (Research Capability Funding).

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements

We are grateful to John Rowlands in the Nuffield Department of Medicine for his support in setting up and maintaining the Hepitopes website.

A report of this project and a link to the on-line Hepitopes database have been presented at the international HBV meeting, Seoul, South Korea, September 2016. The poster can be reviewed here: <https://f1000research.com/posters/5-2357>

References

- Gurtsevitch VE: **Human oncogenic viruses: hepatitis B and hepatitis C viruses and their role in hepatocarcinogenesis.** *Biochemistry (Mosc)*. 2008; **73**(5): 504–513. [PubMed Abstract](#) | [Publisher Full Text](#)
- Hayer J, Jadeau F, Deléage G, *et al.*: **HBVdb: a knowledge database for Hepatitis B Virus.** *Nucleic Acids Res*. 2013; **41**(Database issue): D566–570. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Echevarría JM, Avellón A: **Hepatitis B virus genetic diversity.** *J Med Virol*. 2006; **78**(Suppl 1): S36–42. [PubMed Abstract](#) | [Publisher Full Text](#)
- Haijun D, Yong H, Ailong H, *et al.*: **Distinct quasispecies characteristics and positive selection within the core gene in chronic hepatitis B virus infected child and adult patients.** *Yi Chuan*. 2015; **37**(5): 465–472. [PubMed Abstract](#) | [Publisher Full Text](#)
- Kramvis A: **Genotypes and genetic variability of hepatitis B virus.** *Intervirology*. 2014; **57**(3–4): 141–150. [PubMed Abstract](#) | [Publisher Full Text](#)
- Barnes E: **Therapeutic vaccines in HBV: lessons from HCV.** *Med Microbiol Immunol*. 2015; **204**(1): 79–86. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kapoor R, Kottitil S: **Strategies to eliminate HBV infection.** *Future Virol*. 2014; **9**(6): 565–585. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Godon O, Evlachev A, Bourguine M, *et al.*: **Recognition of core-derived epitopes from a novel HBV-targeted immunotherapeutic by T-cells from patients infected by different viral genotypes.** *Vaccine*. 2015; **33**(36): 4548–4553. [PubMed Abstract](#) | [Publisher Full Text](#)

9. Elvidge S: **Blockbuster expectations for hepatitis B therapeutic vaccine.** *Nat Biotechnol.* 2015; **33**(8): 789.
[PubMed Abstract](#) | [Publisher Full Text](#)
10. Grimm D, Heeg M, Thimme R: **Hepatitis B virus: from immunobiology to immunotherapy.** *Clin Sci (Lond).* 2013; **124**(2): 77–85.
[PubMed Abstract](#) | [Publisher Full Text](#)
11. Singh R, Kaul R, Kaul A, *et al.*: **A comparative review of HLA associations with hepatitis B and C viral infections across global populations.** *World J Gastroenterol.* 2007; **13**(12): 1770–1787.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
12. Tan AT, Sodsai P, Chia A, *et al.*: **Immunoprevalence and immunodominance of HLA-Cw*0801-restricted T cell response targeting the hepatitis B virus envelope transmembrane region.** *J Virol.* 2014; **88**(2): 1332–1341.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
13. Comber JD, Karabudak A, Shetty V, *et al.*: **MHC Class I Presented T Cell Epitopes as Potential Antigenes for Therapeutic Vaccine against HBV Chronic Infection.** *Hepat Res Treat.* 2014; **2014**: 860562.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
14. Dunn C, Peppas D, Khanna P, *et al.*: **Temporal analysis of early immune responses in patients with acute hepatitis B virus infection.** *Gastroenterology.* 2009; **137**(4): 1289–1300.
[PubMed Abstract](#) | [Publisher Full Text](#)
15. Fisicaro P, Valdatta C, Boni C, *et al.*: **Early kinetics of innate and adaptive immune responses during hepatitis B virus infection.** *Gut.* 2009; **58**(7): 974–982.
[PubMed Abstract](#) | [Publisher Full Text](#)
16. Thimme R, Wieland S, Steiger C, *et al.*: **CD8⁺ T cells mediate viral clearance and disease pathogenesis during acute hepatitis B virus infection.** *J Virol.* 2003; **77**(1): 68–76.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
17. Webster GJ, Reignat S, Maini MK, *et al.*: **Incubation phase of acute hepatitis B in man: dynamic of cellular immune mechanisms.** *Hepatology.* 2000; **32**(5): 1117–1124.
[PubMed Abstract](#) | [Publisher Full Text](#)
18. Desmond CP, Bartholomeusz A, Gaudieri S, *et al.*: **A systematic review of T-cell epitopes in hepatitis B virus: identification, genotypic variation and relevance to antiviral therapeutics.** *Antivir Ther.* 2008; **13**(2): 161–175.
[PubMed Abstract](#)
19. Kefalakes H, Budeus B, Walker A, *et al.*: **Adaptation of the hepatitis B virus core protein to CD8⁺ T-cell selection pressure.** *Hepatology.* 2015; **62**(1): 47–56.
[PubMed Abstract](#) | [Publisher Full Text](#)
20. Lin YY, Liu C, Chien WH, *et al.*: **New insights into the evolutionary rate of hepatitis B virus at different biological scales.** *J Virol.* 2015; **89**(7): 3512–3522.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
21. Desmond CP, Gaudieri S, James IR, *et al.*: **Viral adaptation to host immune responses occurs in chronic hepatitis B virus (HBV) infection, and adaptation is greatest in HBV e antigen-negative disease.** *J Virol.* 2012; **86**(2): 1181–1192.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
22. Kawashima Y, Pfafferott K, Frater J, *et al.*: **Adaptation of HIV-1 to human leukocyte antigen class I.** *Nature.* 2009; **458**(7238): 641–645.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
23. Matthews PC, Leslie AJ, Katzourakis A, *et al.*: **HLA footprints on human immunodeficiency virus type 1 are associated with interclade polymorphisms and intraclade phylogenetic clustering.** *J Virol.* 2009; **83**(9): 4605–4615.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
24. Payne RP, Matthews PC, Prado JG, *et al.*: **HLA-mediated control of HIV and HIV adaptation to HLA.** *Adv Parasitol.* 2009; **68**: 1–20.
[PubMed Abstract](#) | [Publisher Full Text](#)
25. Kiepiela P, Ngumbela K, Thobakgale C, *et al.*: **CD8⁺ T-cell responses to different HIV proteins have discordant associations with viral load.** *Nat Med.* 2007; **13**(1): 46–53.
[PubMed Abstract](#) | [Publisher Full Text](#)
26. Leslie A, Matthews PC, Listgarten J, *et al.*: **Additive contribution of HLA class I alleles in the immune control of HIV-1 infection.** *J Virol.* 2010; **84**(19): 9879–9888.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
27. Matthews PC, Koyanagi M, Kløverpris HN, *et al.*: **Differential clade-specific HLA-B*3501 association with HIV-1 disease outcome is linked to immunogenicity of a single Gag epitope.** *J Virol.* 2012; **86**(23): 12643–12654.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
28. Nitschke K, Luxemburger H, Kiraithe MM, *et al.*: **CD8⁺ T-Cell Responses in Hepatitis B and C: The (HLA-) A, B, and C of Hepatitis B and C.** *Dig Dis.* 2016; **34**(4): 396–409.
[PubMed Abstract](#) | [Publisher Full Text](#)
29. Liu WC, Lin CP, Cheng CP, *et al.*: **Aligning to the sample-specific reference sequence to optimize the accuracy of next-generation sequencing analysis for hepatitis B virus.** *Hepatol Int.* 2016; **10**(1): 147–157.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
30. Matthews P: **Hepatitis B Virus (HBV) amino acid alignments and genotype-specific consensus sequences.** *Figshare.* 2016.
[Data Source](#)
31. Chang W, Cheng J, Allaire JJ, *et al.*: **shiny: Web Application Framework for R version 0.14.1.** 2016.
[Reference Source](#)
32. Sunbul M: **Hepatitis B virus genotypes: global distribution and clinical importance.** *World J Gastroenterol.* 2014; **20**(18): 5427–5434.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
33. Matthews PC, Beloukas A, Malik A, *et al.*: **Prevalence and Characteristics of Hepatitis B Virus (HBV) Coinfection among HIV-Positive Women in South Africa and Botswana.** *PLoS One.* 2015; **10**(7): e0134037.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
34. Matthews PC, Geretti AM, Goulder PJ, *et al.*: **Epidemiology and impact of HIV coinfection with hepatitis B and hepatitis C viruses in Sub-Saharan Africa.** *J Clin Virol.* 2014; **61**(1): 20–33.
[PubMed Abstract](#) | [Publisher Full Text](#)
35. Martin JH: **Hepitopes Database [Data set].** *Zenodo.* 2016.
[Data Source](#)
36. Lumley S, Noble H, Hadley M, *et al.*: **'Hepitopes': A Database of HLA Class I Epitopes in Hepatitis B Virus.** *Figshare.* 2016.
[Data Source](#)
37. Matthews P: **'Hepitopes': A Database of HLA Class I Epitopes in Hepatitis B Virus (HBV).** Oxford University Research Archive.
[Data Source](#)

Open Peer Review

Current Referee Status:   

Version 1

Referee Report 05 December 2016

doi:[10.21956/wellcomeopenres.10727.r17597](https://doi.org/10.21956/wellcomeopenres.10727.r17597)



Joerg Timm

Institute of Virology, Heine-University, University Hospital, Düsseldorf, Germany

The authors describe a novel and publically available database of CD8+ T cell epitopes in HBV. Our knowledge of the targeted HBV epitopes is still limited and until now no comprehensive collection of all described epitopes is available. The database is therefore of great interest to researchers working in the field of HBV T cell immunology and potentially is an important tool for the design of therapeutic vaccination strategies.

The manuscript is well written and gives all relevant details about the strategy of data mining from the literature. The title is appropriate and the abstract provides an adequate summary of the article. Although the database is of great value, there are a few limitations that need to be addressed.

I have the following comments:

- It is unclear to me how the search criteria were selected. It seems to me that epitopes from relevant publications have not been included in the database (e.g. doi: 10.1002/hep.27771), despite having been mentioned in the manuscript. I understand that the existing data have not yet been validated. But the authors should reconsider the search strategy to include more epitopes.
- The authors state that they will continue to curate the database in the future. As I understand, this will be done based on information provided by users and by additional literature searches. In line with my first comment, the literature search should be broadened in order to avoid missing relevant publications. More importantly, will there be a quality check of entries provided by users? If so, the authors should describe how this is done.
- I believe, the database would be more helpful, if the epitopes were listed as a single entry with all relevant references linked to it. Some epitopes have been described in multiple publications (e.g. the A*02-restricted core epitope FLPSDFFPSV is listed 56 times). This complicates the search for individual epitopes.
- I believe, it would strongly help if additional information about the assay used for identification of the epitope was included in the database. It is quite different if an epitope was experimentally confirmed by a functional assay or HLA class I-multimers, if HLA class I-peptide-binding was determined or if the epitope was only predicted *in silico*.
- Do the authors consider including CD4+ T cell epitopes as well? Is there a rational why the database is limited to CD8+ T cell epitopes?

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: No competing interests were disclosed.

Referee Report 01 December 2016

doi:[10.21956/wellcomeopenres.10727.r17596](https://doi.org/10.21956/wellcomeopenres.10727.r17596)



Louise Corti

UK Data Archive, University of Essex, Colchester, UK

Cataloguing efforts

This retrospective collation and cataloguing work of sites and sequences of all published class I epitopes in HBV (Hepitopes) is a useful contribution to the knowledge base in this field. This systematic literature review has clearly been a heavily manual effort as information is so dispersed. The two online bibliographic databases that were selected to undertake targeted searches, Medline and Embase, may not offer 100% coverage and what about other resources in other languages? The 'universe' is not explicitly mentioned in the paper and needs to be added.

Technology and sustainability

I found the website easy to use, but at times, unstable, with multiple crashes on the site/ disconnection from the server. It is good to see the use of open source tools, for example, Shiny R. The interactivity is quite basic, based on column sorting (effectively an online spreadsheet) but nevertheless useful. I found the feedback and contributor process to be manual and quite primitive, through what is essentially a minimally structured paper-based form. It would nice to see a more automated upload process with structured fields available for population and validation on the fly; and some local vetting once an entry is submitted.

In terms of future proofing and keeping data up to date, the team note that they have an ongoing remit to develop this into a sustainable open-access online research resource that evolves over time. While the code is usefully for the datatable available on [Github](#), it is not that clear how any maintenance of the project might be supported and under what budget. The team have promised to continue to refine and update this resource, yet what happened if key staff leave?

Future proofing is the main concern of so many local resources like this, which is why the Research Council's typically try not to fund on-going longer-term information resources under the auspices of smaller one-off research projects due to the lack of future accessibility. ESRC have set up a centralised web based ReStore to ensure that their web-based research methods resources created under one-time research awards are maintained in the longer term – either as maintained resources or static snapshot web-based resources. So may resources created with the very best intentions, are not kept up to date and may become out of date quickly, and this become a relic

While I really embrace the idea of crowd sourcing information to maintain future, in my opinion the physical resource /database (the catalogue) should be looked after by an official long standing trusted repository e.g. Genbank and not an individual university. The fact that the authors mention another almost identical on-line database of epitopes, but for HIV ((

https://www.hiv.lanl.gov/content/immunology/ctl_search), and other online repositories in the field of HIV T cell immunology (sequence data including HBV-specific databases <https://hbvdb.ibcp.fr/HBVdb/2> and <http://www.hepseq.org>, and in broader sequence repositories, such as <http://www.ncbi.nlm.nih.gov>) makes me wonder why a unified database has not been proposed with a single interface! The last thing the community needs is a proliferation of locally managed lists so why not combine them into one infrastructure? This professionalised future proofing of resources like this really needs to be addressed in the paper, even if they are aspirational.

Finally, if the web resource/front end was not to be retained in the future this snapshot database is usefully backed up in Figshare. The paper says the team will keep this updated, but it is not clear how future dumps of the dataset might be formally archived, persistently cited, and referenced in Figshare and how formal versioning might work. Will Oxford and Figshare sync their versions? This area needs to be addressed to make it clear they have really considered future plans for this useful table of information.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Competing Interests: No competing interests were disclosed.

Referee Report 24 November 2016

doi:[10.21956/wellcomeopenres.10727.r17603](https://doi.org/10.21956/wellcomeopenres.10727.r17603)



Antonio Bertoletti¹, Anthony Tanoto Tan², Upkar S Gill³

¹ Emerging Infectious Diseases Program, Duke-National University of Singapore Graduate Medical School, Singapore, Singapore

² Emerging Infectious Diseases Research Program, Duke-National University of Singapore Graduate Medical School, Singapore, Singapore

³ Hepatology, Centre for Immunobiology, Blizard Institute, Barts and The London School of Medicine & Dentistry, Queen Mary University of London, London, UK

This manuscript describes the creation of a database cataloging epitopes derived solely from Hepatitis B virus. The authors need to be congratulated for their effort to set up a clear and user friendly resource that provides all the current scientific information known on HBV epitopes at present.

What I found particularly nice is that it details how the different epitopes were experimentally determined and this allow the user to understand whether the epitope was able to induce a T cell response in a subject with natural infection or was only the result of *in silico* determination and binding assays.

Such a resource will allow easy progression of future research in T cell immunology. Researchers will have access to the information on relevant HBV epitopes in one dataset, which will be both convenient and time-saving. In addition as noted, the experimental technique of epitope determination is noted which can further direct experimental design.

As noted by the authors, the resource, will hopefully, can continue to grow, certainly considering the number of people who are screening HBV patients, this might not be a trivial task. In addition it is also of value that comments from readers can be added and thus allow authors to update the resource as

appropriate. Thus the importance being that the resource will evolve over time and therefore be an invaluable asset in HBV research.

The authors of the manuscript/resource have clearly outlined their methods for the literature search in developing the interactive database which are robust. I also note the limitations of dataset validation, whereby citations could be overlooked as they are not search specific. However, accepting this I have a couple of suggestions, that I hope can be useful to improve the utility of the site:

1. The data gathered do not appear to be consolidated and as such do not generate a “final” state of knowledge on HBV epitopes landscape. The website provides all the information but they are listed as a list of papers. This can create potential confusion, since for example if you are interested to know whether the core 18-27 FLPSDFFPSV epitope comprises a polymorphism (which it does) you need to ask this information when you are reading the relevant paper, where such polymorphism was demonstrated. If you are asking this identical question starting from a paper where such polymorphism was not analyzed, the answer is that there is no polymorphism in this region. I think this is an aspect that could be improved since it might generate some misinterpretation.
2. A similar problem is also present in relation to the impact of HLA-A02 subtypes. Some epitopes are presented as pan A2 based on old papers but more recent publication (Tan AT *et al.* 2008) have show that some epitopes are exclusively A0201. For example the envelope epitopes FLLTRILT1 or FLLTKILT1 were demonstrated exclusively on HLA-A0201 subjects and not on HLA-A0203, 0206, 0207 subjects.) The progressive evolution of information is somehow lost. Thus at places where the term ‘pan’ is used this may need to be verified as new data may not corroborate this. Having said that I appreciate the limitations, as noted in the manuscript regarding data validation.
3. When asking for a list of, for example, HBV epitopes within the core region, the web site provides all the papers where core epitopes were demonstrated but there is a lot of redundancy. There are 6 pages listing all the papers published using the core 18-27 epitope. I am wondering whether there could be a method to first list all the different epitopes with the indication of the number of publications where such epitope was reported as a measure of the evidence supporting the validity of the epitope. This may be a method to improve the resource.
4. The description given in the database for the method of epitope characterization is very detailed, but what may not immediately be clear is whether the method used is accurate enough to give you real HBV epitopes. For example, if the epitopes were characterized using the T cells from naturally infected patients and targets where HBV proteins were endogenously processed, that would be ideal. If the epitope was only characterized by predictive algorithms, then it would be less so. Perhaps a scale (Patient T cells with endogenously processed epitopes > Patient T cells with peptide pulsed targets > binding assays > predictive algorithms) indicating the validity of the consolidated epitopes might be clearer for the users of the database.
5. As some citations may have been missed due to the search criteria, it may be a possibility to widen the search. This may be more feasible once readers have provided feedback. This will allow the search to be increased possibly to include all the relevant data for the resource.

References

1. Tan AT, Loggi E, Boni C, Chia A, Gehring AJ, Sastry KS, Goh V, Fiscaro P, Andreone P, Brander C,

Lim SG, Ferrari C, Bihl F, Bertoletti A: Host ethnicity and virus genotype shape the hepatitis B virus-specific T-cell repertoire. *J Virol.* 2008; **82** (22): 10986-97 [PubMed Abstract](#) | [Publisher Full Text](#)

We have read this submission. We believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: No competing interests were disclosed.
